THE STRATEGIC COST OF RATIONALITY

Anonymous authors
Paper under double-blind review

ABSTRACT

When predictions influence the outcomes they forecast, learning can become impossible: under adversarial responses, regret against an oracle that knows the response is $\Omega(T)$. We show that learnability is restored when the world is rational, meaning it best-responds to each published prediction by minimising a convex cost under which acting more becomes relatively cheaper as the prediction rises. Rationality then makes the honest prediction self-certifying: the outcome at any prediction reveals its residual, whose square is the excess loss, and the rational response makes that residual monotone, so it points to a unique honest prediction identified from the data at that prediction alone. This gives regret $\Theta(\log T)$ at every level of responsiveness. Self-certification needs an honest prediction to exist, but when the world consists of agents of discrete types it may not: no prediction is calibrated, every learner's loss stays a constant above the noise, and zero bias and accuracy can no longer both be achieved. The same rationality that makes the world learnable also makes it predictable enough to steer. A learner with a stake in the outcome gains by steering it: at the honest prediction its marginal incentive splits into conflict of interest and manipulation in the ratio $1 : \varepsilon$, where ε is the elasticity of the world's response, and manipulation adds harm to users, beyond what a learner ignoring its influence would cause, exactly when $\varepsilon > 0$. But steering moves the learner away from the one prediction that certifies itself. The strategic target depends on the slope of the world's response, whereas the outcome at that target reveals only its level, so when the fundamental is unknown, learning the strategic target costs $\Theta(\sqrt{T})$ rather than $\Theta(\log T)$. Rationality thus makes honest prediction self-certifying, but leaves a strategic learner needing information its own outcomes do not reveal.

1 INTRODUCTION

When a forecast is published it can move the outcome it predicts. A demand forecast shifts inventory, a congestion forecast shifts traffic, a price target shifts order flow. Performative prediction (Perdomo et al., 2020) formalises this: the learner publishes a forecast $\hat{p}_t$, the world responds, and the outcome y_t depends on $\hat{p}_t$.

If the response is adversarial, or merely random, nothing can be learned: a world that flips a fair coin between two response intensities forces linear regret, against an oracle that knows the intensity, on every algorithm (Proposition B.9). Real respondents are rational agents who minimise their own costs; when acting becomes relatively cheaper as the forecast rises, their response is a fixed, monotone function of what they see. This paper asks what that rationality buys a learner, and what it costs the people who rely on the forecast.

The answer is a sequence of questions, drawn in Fig. 1; each section ends on the question the next one answers. Under this condition the response is monotone (Section 2). The outcome then reveals the residual of the current forecast, and monotonicity makes that residual point to a unique honest forecast, which can be learned at logarithmic regret at every level of responsiveness (Section 3). If the response jumps, as it does when agents of discrete types switch on, no forecast is calibrated; every learner pays linear regret against the calibrated benchmark, and zero bias and accuracy trade off along a straight frontier (Section 4). Where a calibrated forecast exists, only an objective other than accuracy moves a learner away from it. A stake in the outcome does, and at the honest forecast its marginal incentive splits into conflict of interest and manipulation in the ratio $1 : \varepsilon$, where ε is the elasticity of the world's intensity (Section 5). The strategic forecast depends on the slope of the

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

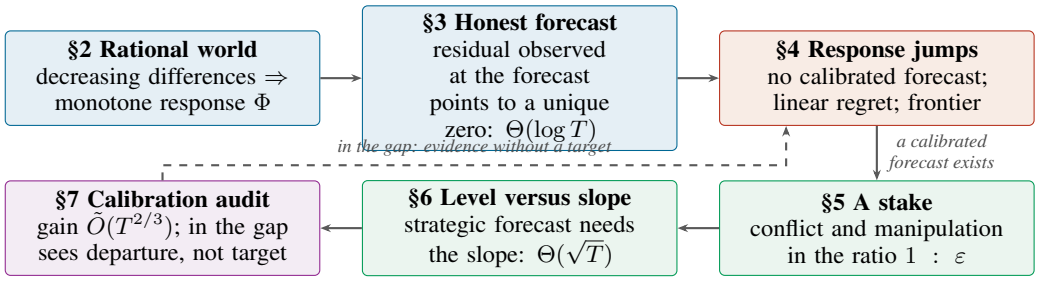

Figure 1: The argument as a loop of questions. Blue: what rationality buys an honest learner. Red: where the equilibrium disappears. Green: what it buys a learner with a stake. Purple: what outsiders can see.

response, which the outcome at that forecast does not reveal, so learning it costs $\Theta(\sqrt{T})$ instead of $\Theta(\log T)$ (Section 6). A strategic learner is therefore miscalibrated, and a binned audit caps its gain at $\tilde{O}(T^{2/3})$; in the gap, though, the audit sees a departure from calibration but not the objective behind it (Section 7).

In one sentence: *learning the honest forecast and learning the strategic one are different statistical problems: the honest target is certified by the residual generated at the target itself, while the strategic target depends on response geometry that observations at the target do not identify.*

What is classical and what is new. With a monotone response, learning the honest forecast is the multiperiod control problem of Lai and Robbins; the $\sigma^2 \ln T$ rate, the slope-tuned gain, its adaptive version and the Bayes lower bound are theirs (Lai & Robbins, 1979; 1981; 1982), and we credit them as such (Remark B.23). New are: the reduction from a rational world to that problem, with the decreasing-differences condition that makes it valid (Section 3); the gap regime, its bias–accuracy frontier, and learning the accuracy-optimal side at excess $O(T^{2/3})$ over the best attainable forecast (Section 4); the $1 : \varepsilon$ decomposition of the marginal incentive and the sign theorem for harm relative to simultaneous play (Section 5); the $\Theta(\sqrt{T})$ versus $\Theta(\log T)$ separation between learning the strategic and the honest forecast, proved through an uninformative curve at the strategic target (Section 6); and the binned calibration audit with its $T^{2/3}$ limit (Section 7). A learner with a stake in the decision taken from its forecast was studied by Boutilier (2012) with an exogenous outcome; new here is the stake together with a performative outcome. Main-text results are stated with their assumptions and name the appendix result that proves them; Appendices A–F follow the order of Sections 2 to 7.

Rationality enables $\log T$ learning, but only for an honest learner.

2 A RATIONAL WORLD

At rounds $t = 1, \dots, T$ the learner publishes a forecast $\hat{p}_t \in \mathcal{P} := [0, P]$; the world observes it and chooses an intensity $\lambda_t = \lambda^*(\hat{p}_t)$; and the outcome is

$$y_t = \mu - \lambda_t \hat{p}_t + \eta_t = \mu - \Phi(\hat{p}_t) + \eta_t, \qquad \Phi(p) := \lambda^*(p)\, p, \tag{2.1}$$

with $\mu > 0$ unknown and η_t a martingale difference of conditional variance σ^2 (Assumptions A.1 to A.3). The timing is Stackelberg, with the learner leading (Appendix A, Fig. 4). The learner knows neither μ nor λ^*. The world is *rational*: it chooses λ to minimise its own cost. With the tracking cost

$$C_W(\lambda; \hat{p}) = \alpha_W \lambda^2 + \beta_W (\lambda - \kappa_W \hat{p})^2, \tag{2.2}$$

which charges for acting (α_W) and for missing a target response proportional to the forecast (β_W, κ_W), the best response is $\lambda^*(\hat{p}) = \gamma \hat{p}$ with responsiveness $\gamma = \beta_W \kappa_W / (\alpha_W + \beta_W)$ (Proposition A.7), so $\Phi(p) = \gamma p^2$. We call this the *quadratic world*. More generally, if the world's cost has decreasing differences in $(\lambda, \hat{p})$ and $\lambda \geq 0$, its best response is nondecreasing (Topkis, 1978), hence Φ is nondecreasing; that, not the quadratic form, is what the learning results use. Convexity alone does not suffice (Example B.5). The world need not know μ to best-respond: it needs its own cost and the published forecast.

Table 1: Main-text notation. Table 3 extends it to the appendix.

$\mathcal{P} = [0, P]$	forecast set	$\hat{p}_t, y_t, \eta_t$	forecast, outcome, noise (variance σ^2)
μ	fundamental	$\lambda^*(p)$, $\Phi(p) = \lambda^*(p)p$	intensity, response
γ	responsiveness (quadratic world: $\Phi = \gamma p^2$)	$\varepsilon(p) = p\lambda^{*\prime}(p)/\lambda^*(p)$	elasticity of the intensity
$r(p) = p + \Phi(p) - \mu$	residual	p^*, $R' = r'(p^*)$	honest forecast, slope there
Regret_T	regret against p^*	$\mathsf{RSA}(a)$, $g(x) = \frac{x^2}{2x-1}$	residual SA, its constant
x_0, r^-, r^+	jump point, one-sided residuals	$\bar{r}_T, \overline{r^2}_T$	average bias, average excess
$\alpha, 1-\alpha$	accuracy weight, stake	$V(p) = U_A(p, \lambda^*(p))$	learner's objective
p^N, p^S	Nash forecast, strategic forecast	$\Delta_V = V(p^*) - V(p^S)$	gain from abandoning honesty
S_T	strategic regret	Σ_T, w_T, ℓ_T	pooled statistic, bin width, $\ln T \ln\ln T$

Discrete types. If the world is a population of agents who pay a fixed cost to participate, type k switches on only when the forecast exceeds a threshold, and the aggregate intensity is $\lambda^*(p) = G_k p$ on segment k, with $G_1 < G_2 < \cdots$ (Appendix A.3). This response is rational and nondecreasing, but it jumps at every threshold. Section 4 shows what the jump does.

Residual, calibration and regret. The *residual* of a forecast is $r(p) := p + \Phi(p) - \mu = p - \mathbb{E}[\, y \mid \hat{p} = p\,]$. A forecast is calibrated when the outcome it induces has mean equal to the forecast, that is, when $r(p) = 0$. We call such a forecast *honest* and write p^*. The loss is $\ell_t = (\hat{p}_t - y_t)^2$, and regret is measured against the oracle that plays p^*: $\mathrm{Regret}_T := \sum_{t\le T} \mathbb{E}\,\ell_t - T\sigma^2 = \sum_{t\le T} \mathbb{E}\, r(\hat{p}_t)^2$. In the language of Perdomo et al. (2020), the honest forecast is performatively stable, whereas a learner who optimises its own objective through the world's response seeks a performatively optimal forecast (Proposition D.2). For a learner with a stake, the Nash forecast p^N differs from p^* (Section 5). Table 1 collects the notation. *When is the honest forecast learnable, and how fast?*

3 Honest targets certify themselves

Lemma 3.1 (Residual identity; Lemma B.1)**.** *Let $z_t := \hat{p}_t - y_t$. Then $z_t = r(\hat{p}_t) - \eta_t$, and with homoscedastic noise $\mathbb{E}[\ell_t \mid \mathcal{F}_{t-1}] = r(\hat{p}_t)^2 + \sigma^2$. If Φ is nondecreasing on $\mathcal{P}$, then $r(q) - r(p) \ge q - p$ for all $p < q$, so r has at most one root.*

The outcome reveals the residual at the current forecast for any fixed response: z_t is an unbiased measurement of $r(\hat{p}_t)$, and the excess loss is exactly $r(\hat{p}_t)^2$. Monotonicity adds what learning needs: the residual has a unique zero, and its sign points toward it. So whether p is the honest forecast is decided by the distribution of y at p alone, with no knowledge of μ, Φ or γ (Theorems E.13 and E.16). This is what we mean by *self-certification*. It suggests residual stochastic approximation, which uses only the observed errors:

$$\hat{p}_{t+1} = \Pi_{\mathcal{P}}\big(\hat{p}_t - \tfrac{a}{t}\, z_t\big) \qquad (\mathsf{RSA}(a),\ \text{Algorithm A.6}).$$

What rationality buys. Monotonicity of r, not the size of the response, decides learnability. Negative feedback of any strength (Φ nondecreasing) and weak positive feedback (Φ Lipschitz with constant below one, the regime of Mendler-Dünner et al., 2020) both give a residual with a positive modulus (Proposition B.6). Strong positive feedback ($\Phi' < -1$ somewhere) breaks monotonicity and can create several self-fulfilling equilibria (Proposition B.7), and a random response forces linear regret against the oracle that knows each round's intensity (Proposition B.9). On the intensity, the right condition is a *lower* bound on its elasticity, $\varepsilon \ge -1$ (Proposition B.3).

Theorem 3.2 (Logarithmic learning at every level of responsiveness)**.** *Assume r has a root $p^* \in \mathcal{P}$ with $(p - p^*)\, r(p) \ge m(p - p^*)^2$ and $|r(p)| \le L|p - p^*|$ on $\mathcal{P}$, for some $m, L > 0$.*

(i) If $2am > 1$, then RSA(a) has $\mathrm{Regret}_T \le C(1+\ln T)$ for all T, with C explicit in (a, m, L, σ, P) (Theorem B.13).

(ii) If p^* is interior, r is differentiable at p^* with slope $R' := r'(p^*)$, the noise is Gaussian and $2aR' > 1$, then $\mathrm{Regret}_T = g(aR')\,\sigma^2 \ln T\,(1+o(1))$, where $g(x) := x^2/(2x-1)$ is minimised at $x = 1$ with $g(1) = 1$ (Theorem B.21).

(iii) With Gaussian noise, every algorithm, even one that knows Φ, has $\sup_{\mu\in I} \mathbb{E}\,\mathrm{Regret}_T \ge \sigma^2 \ln T - C_I$ on any interval I of fundamentals (Theorem B.29).

The rate does not depend on responsiveness; the constant of a fixed gain does (Proposition B.22). The gain $a = 1/R'$ attains σ^2 but uses the slope at the answer; a least-squares slope estimate recovers it almost surely for a clipped-publication variant (Lai & Robbins, 1981, Proposition B.24). *What if the world is rational but its response jumps?*

4 When the equilibrium disappears

Take a rational response that is nondecreasing and continuous except at one threshold x_0, where it jumps (the jump setting, Assumption C.1); the discrete-type model is the leading case. Write $r^- := r(x_0-)$ and $r^+ := r(x_0+)$.

Theorem 4.1 (No calibrated forecast)**.** *Suppose the jump crosses zero, $r^- < 0 < r^+$.*

(i) r has no root: no forecast is calibrated. In the discrete-type model this is the gap-window condition (A.2) (Proposition C.3).

(ii) The myopic dynamics, which publish the forecast that would be calibrated if the world kept its last intensity, $\hat{p}_{t+1} = \mu/(1+\lambda^*(\hat{p}_t))$, converge to a 2-cycle straddling x_0, in closed form for discrete types (Theorems C.4 and C.5).

(iii) Every algorithm has excess loss at least $T\min(r^{-2}, r^{+2})$ relative to the calibrated benchmark, which does not exist here (Proposition C.7).

Against the best attainable forecast, the two-phase algorithm of Theorem 4.2(iv) has excess $O(T^{2/3})$ (Theorem C.23). Smoothing the jump over width h restores a calibrated forecast for every $h > 0$. For a fixed gain the regret constant grows like $1/h$ as $h \to 0$, while the slope-tuned gain attains σ^2 for each fixed h; the relation between the limits $h \to 0$ and $T \to \infty$ is open (Proposition C.12, Remark C.13).

Which linear cost to pay is then a choice. Write $\bar{r}_T := \frac{1}{T}\sum_{t\le T} r(\hat{p}_t)$ for the average bias and $\overline{r^2}_T := \frac{1}{T}\sum_{t\le T} r(\hat{p}_t)^2$ for the average excess.

Theorem 4.2 (Bias–accuracy frontier)**.** *Assume $r^- < 0 < r^+$.*

(i) Every algorithm, on every sample path, satisfies $\overline{r^2}_T \ge (r^- + r^+)\,\bar{r}_T - r^- r^+$ (Theorem C.10).

(ii) If Φ is Lipschitz on each side of x_0 and the noise has six moments, every point of this line with $\bar{r} \in [r^-, r^+]$ is attained, up to $O(T^{-1/2})$, by RSA run on a shifted residual (Theorem C.10(ii)).

(iii) Plain RSA converges to x_0, its average bias vanishes, and $\mathrm{Regret}_T/T \to |r^- r^+|$ (Theorem C.8).

(iv) The accuracy optimum $\min(r^{-2}, r^{+2})$ lies at an endpoint of the line. A two-phase algorithm that identifies the cheaper side from the moments of RSA's errors attains it, with excess $O(T^{2/3})$ when σ is known, and consistently when σ is unknown (Theorems C.19, C.21 and C.23).

Since $|r^- r^+| > \min(r^{-2}, r^{+2})$ whenever $|r^-| \ne r^+$, no learner has both zero average bias and optimal accuracy (Fig. 2). With no calibrated forecast available, the data fix r^- and r^+ and hence the line; the learner's objective picks the point on it. *Where a calibrated forecast does exist, what would make a learner leave it?*

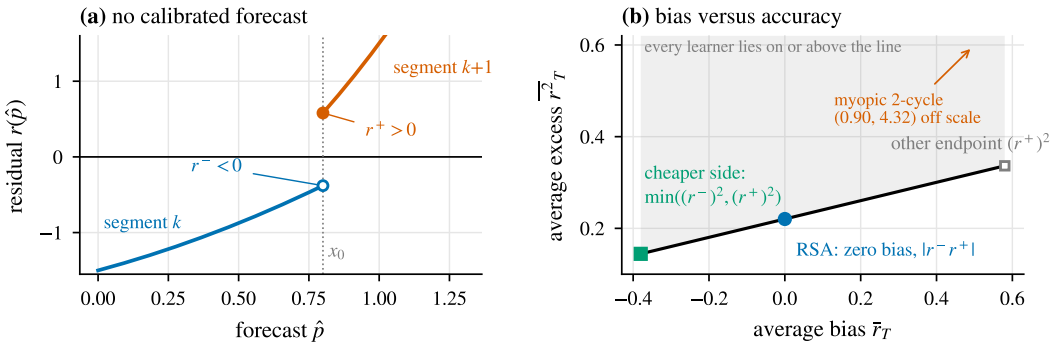

Figure 2: The gap, for discrete types with $(\mu, G_k, G_{k+1}, x_0) = (1.5, 0.5, 2, 0.8)$. (a) The residual jumps over zero at the activation threshold, so no forecast is calibrated (Theorem 4.1). (b) Every learner lies on or above the line $(r^- + r^+)\bar{r} - r^- r^+$ (Theorem 4.2). RSA sits at zero bias and pays $|r^- r^+|$; the accuracy optimum sits at an endpoint; the myopic 2-cycle lies far above the line.

5 A LEARNER WITH A STAKE

An accuracy-only learner never leaves the honest forecast: its excess loss is $r(p)^2$, and r has at most one root (Lemma 3.1). A persistent departure needs another objective, and the simplest is a stake in the outcome. The learner minimises

$$U_A(p, \lambda) = \alpha\big(p(1+\lambda) - \mu\big)^2 + (1-\alpha)\lambda^2 p^2,$$

a weighted sum of its accuracy and the expected loss of a position held at the fundamental, since $\mathbb{E}(\mu - y)^2 = \lambda^2 p^2 + \sigma^2$; α is the accuracy weight and $1 - \alpha \in (0, 1)$ the stake (the stake model, Assumption D.19). Users of the forecast bear the harm $L(p, \lambda) = (p(1+\lambda) - \mu)^2$, which equals $r(p)^2$ when the world best-responds. We compare the Nash forecast p^N, where the learner best-responds to the world's current intensity; the strategic forecast p^S, which minimises $V(p) = U_A(p, \lambda^*(p))$; and the honest forecast p^* as benchmark.

Proposition 5.1 (A stake splits into conflict and manipulation)**.** *(i) The learner's preferred forecast differs from the users' at every $\lambda > 0$ if and only if $\alpha < 1$: a stake is exactly a conflict of interest (Proposition D.26).*

(ii) Let λ^* be a regular rational response (Assumption D.21). At the honest forecast the marginal incentive V' is the direct effect D_A plus the response wedge W_A, and

$$W_A(p^*) = \varepsilon(p^*)\, D_A(p^*), \qquad V'(p^*) = \big(1 + \varepsilon(p^*)\big) D_A(p^*) = 2(1-\alpha)\, \Phi(p^*)\, \Phi'(p^*)$$

(Proposition D.27). Conflict and manipulation stand in the ratio $1 : \varepsilon$; in the quadratic world ($\varepsilon \equiv 1$) they are equal.

If $\Phi'(p^*) > 0$, equivalently $\varepsilon(p^*) > -1$, every learner with a stake shades its forecast below the honest forecast (Theorem E.14(i)); without that condition it need not (Example D.29). The marginal wedge need not dominate the resulting distortion. To first order in the stake, $(p^S - p^*)/(p^N - p^*) = (1+\varepsilon)(1+\lambda)/(1+\lambda+\varepsilon\lambda)$ (Proposition D.30), which tends to 1 as responsiveness grows in the quadratic world.

Theorem 5.2 (When manipulation adds harm; Theorem D.34)**.** *Assume Assumptions D.19 and D.21 and single crossing (Assumption D.31), meaning the leader's first-order condition has one zero below p^*, which holds for every convex response (Lemma D.32), in particular the quadratic world. Then for every stake and every Nash forecast,*

$$\operatorname{sgn}(p^S - p^N) = -\operatorname{sgn}\varepsilon(p^N), \qquad \operatorname{sgn}(H^S - H^N) = \operatorname{sgn}\varepsilon(p^N),$$

where H^S and H^N are the harms at the strategic and Nash outcomes. In every convex world, leading by the learner weakly adds harm.

So the part of the distortion due to manipulation, as distinct from conflict of interest, harms users exactly when the intensity rises with the forecast ($\varepsilon > 0$).

Theorem 5.3 (The cost of a stake grows with responsiveness; quadratic world)**.** *In the quadratic world with $\alpha \in (0,1)$, let $\Delta_V := V(p^*) - V(p^S)$ be the gain from abandoning honesty and $\alpha\, r(p^S)^2$ the rent, the part of the harm the learner bears through its accuracy weight. The harm $r(p^S)^2$, the rent and Δ_V are all strictly increasing in $\gamma\mu$, and $r(p^S)^2 \to (1-\alpha)^2\mu^2$ as $\gamma\mu \to \infty$ (Theorem D.17). Users lose strictly more than the learner gains, $r(p^S)^2 > \Delta_V$ (Proposition D.13), and the world also gains when its cost is scored against the published forecast (Proposition D.15).*

Theorem 5.3 measures total distortion relative to honesty, which includes conflict of interest. Relative to simultaneous play, the added harm from manipulation is nonmonotone in responsiveness and vanishes as it grows (Remark D.35); at high responsiveness almost all the harm is conflict of interest. Beyond the quadratic world these statements hold under precise conditions. Manipulation destroys value whenever the response is concave between the two forecasts, and more generally under a secant-over-tangent condition (Theorem D.36). A failure of this condition needs convex curvature, and a failure of the pooled audit of Section 7 needs concave curvature, so both fail together only where the response has an inflection, as it does where a subpopulation switches on (Corollary D.39 and Theorem D.41). *The strategic learner must learn its target. Can it do so as cheaply as the honest learner?*

6 Level versus slope

The strategic target is statistically different from the honest target. The honest target is certified by the residual generated at the target itself. The strategic target depends on the slope of the response, but observations generated at that target do not identify it. Information about the target must be obtained by moving away from it, and those moves incur regret.

Consider the scaled setting $\Phi_\theta = \theta\varphi$, with φ increasing and (μ, θ) unknown in a box (Assumption E.11). The honest forecast is the root of $r(p) = p - m(p)$, where $m(p)$ is the mean outcome at p; it depends on the *level* of the response. The strategic forecast solves $V'(p) = 0$, where $\frac{1}{2}V' = \alpha\, r\,(1 + \theta\varphi') + (1-\alpha)\theta^2\varphi\varphi'$; it depends on the *slope* $\theta\varphi'$. We measure the strategic learner by its strategic regret $S_T := \sum_{t\le T} \mathbb{E}\, V(\hat{p}_t) - T \min_{\mathcal{P}} V$.

Theorem 6.1 (What the outcome at the target reveals)**.** *In the scaled setting:*

(i) two worlds with the same mean outcome at p have the same residual at p;

(ii) at $p = p^S(\mu_0, \theta_0)$, the worlds $(\mu_0 + u\varphi(p),\ \theta_0 + u)$ all induce the same distribution of y at p, yet p is the strategic forecast only at $u = 0$ among worlds with $\theta > 0$

(Theorem E.16(i)–(ii), Theorem E.1).

Theorem 6.2 (Learning the honest and the strategic forecast)**.** *In the scaled setting, with the honest learner measured by Regret_T and the strategic learner by S_T:*

(i) learning the honest forecast costs $\Theta(\log T)$ (Theorem 3.2);

(ii) with μ unknown, learning the strategic forecast costs $\Theta(\sqrt{T})$: explore-then-commit attains $S_T = O(\sqrt{T})$, and every algorithm has $\sup S_T \ge c\sqrt{T} - C$ on a curve of worlds through any interior point (Theorem E.16(iii), Theorems E.2 and E.3);

(iii) with μ known, it costs $\Theta(\log T)$, attained by greedy least squares (Theorem E.16(iv), Theorem E.8).

The mechanism is the uninformative curve of Theorem 6.1(ii): the observation at the strategic forecast carries no information about where that forecast is, so information can only be bought by playing away from it, and the regret charges for that. The lower-bound technique is that of Broder & Rusmevichientong (2012); the same level-versus-slope dichotomy appears in Keskin & Zeevi (2014). We prove it in this model (Lemmas E.5 and E.6). Knowing μ pins the curve and removes the $\sqrt{T}$ term. The honest learner never faces it, because its target is visible at the target (Fig. 3). By contrast, a heuristic calculation (Remark E.17) gives explore-then-commit a cost of order $\mu\sigma\sqrt{T}$ for the honest learner too. Both results carry over to vector forecasts with monotone response operators (Theorems G.4 and G.7). *A strategic learner is miscalibrated. Can outsiders see it, and does that discipline it?*

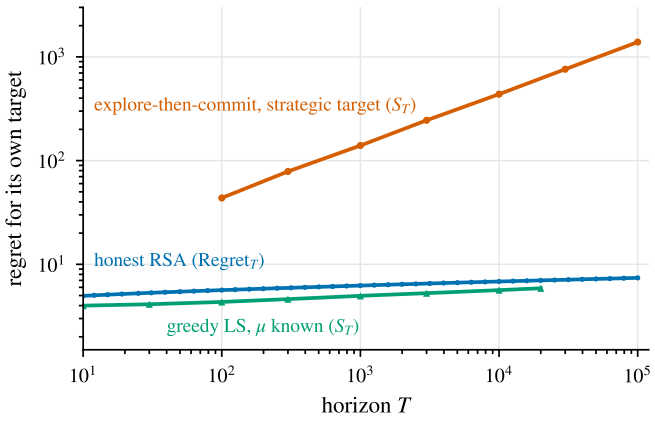

Figure 3: Level versus slope in the quadratic world with $\mu = 1$, $\sigma = 0.5$, $\gamma = 2$. Honest RSA with the tuned gain, measured by Regret_T, grows like $\ln T$. For the strategic target with $\alpha = 0.5$, measured by S_T: explore-then-commit with μ unknown grows like $\sqrt{T}$; greedy least squares with μ known grows like $\ln T$ (Theorem 6.2). Marks are simulation; the explore-then-commit curve has no theoretical constant.

7 Calibration as Evidence

An auditor who knows nothing about μ or Φ can still test calibration. The pooled audit flags the learner when $|\Sigma_T| > z\sigma\sqrt{T}$, where $\Sigma_T := \sum_{t \le T}(y_t - \hat{p}_t)$, at a cost Π_T to the learner (Assumption F.1). For an open-loop, fully informed learner, stakes that grow at least linearly, with κ above the threshold of Theorem F.5, make the optimal bias $O(1/T)$ and keep the gain from manipulating bounded uniformly in T; with $\Pi_T = o(T)$ the harm is linear in T (Theorem F.5 and Proposition F.7). An adaptive, closed-loop learner is limited only to $O(\sqrt{T \ln T})$ (Theorem F.18, Table 11). Honest RSA with $c = aR' \ge 1$ passes (Propositions F.9 and F.12). Pooling fails outside a tangent condition, which every convex response satisfies: a strategic learner alternating between biases of opposite sign keeps Σ_T at noise level while gaining linearly (Theorem F.20). The repair is the forecasting notion of calibration (Dawid, 1982; Foster & Vohra, 1998): test the bias conditional on the forecast.

Theorem 7.1 (The binned audit)**.** *Bin the forecasts at width $w_T = T^{-1/3}\ell_T^{1/3}$ with $\ell_T = \ln T \ln\ln T$, test every bin holding at least ℓ_T/w_T rounds, and charge κT when the learner is flagged, with $\kappa > \Delta_V$ (Definition F.22).*

(i) In every world with a monotone Lipschitz response, every strategic learner, adaptive and knowing the bins, gains at most $O(T^{2/3}\ell_T^{1/3})$ over the full-information honest learner, and causes harm of the same order (Theorem F.23).

(ii) Among binned audits that accept honest RSA(1) in a gap world, none does better than $cT^{2/3}$ (Theorems F.41 and F.46).

(iii) Honest RSA with any stable gain is flagged with probability tending to zero, at rate $O((\ln T)^{-1/2})$ in smooth worlds (Theorem F.39) and $O((\ln T \ln\ln T)^{-1/2})$ in the gap for $a \ge 1$ (Theorem F.51); in the gap, however, $T \cdot \mathbb{P}(\text{flag}) \to \infty$ (Theorem F.36).

Honest learning is then the cheapest policy for a learner that must learn, provided the audit level is raised by one factor of T inside the logarithm: this is proved in linear worlds (Remark F.29, Theorem F.30); the smooth nonlinear case is Theorem F.33. This fails in the gap for every binned audit (Theorem F.36).

In the gap. The audit's verdict turns on the objective, which it cannot see. Any learner whose average bias stays away from zero is flagged by the audit with probability tending to one; that includes the accuracy-optimal learner of Theorem 4.2(iv), while honest RSA, which pays $|r^- r^+|$, passes (Corollaries F.17 and F.52). Where calibration is possible, the audit separates honest from strategic

forecasts. Where it is not, it separates zero-bias forecasts from accurate ones: it sees the departure from calibration, not the objective behind it.

8 Related work and limits

Performative prediction. Perdomo et al. (2020) introduce the framework and its stable and optimal points, and Mendler-Dünner et al. (2020) give stochastic rates under a sensitivity condition that our monotonicity condition extends to negative feedback of any strength (Proposition B.6). Miller et al. (2021) and Izzo et al. (2021) optimise performative risk, Jagadeesan et al. (2022) study regret for performative optima, and Brown et al. (2022) and Ray et al. (2022) let the distribution depend on its history. Narang et al. (2023) treat several learners, Hardt et al. (2022) study the power of those who deploy predictors, and Hardt & Mendler-Dünner (2025) survey the area. We add a micro-founded rational response, the level–slope separation between stable and optimal points, and the gap regime.

Adaptive control, pricing and Stackelberg learning. Learning the honest forecast is the control problem of Lai & Robbins (1979; 1981; 1982). The $\sqrt{T}$ versus $\log T$ separation for the strategic target mirrors dynamic pricing with unknown versus known demand at an incumbent price (Keskin & Zeevi, 2014). Learning optimal commitments costs $\sqrt{T}$ in Stackelberg security games (Balcan et al., 2015), and who leads can depend on update speeds (Zrnic et al., 2021).

Strategic classification and calibration. In strategic classification the follower moves its features in response to a published rule (Brückner & Scheffer, 2011; Hardt et al., 2016; Dong et al., 2018; Kleinberg & Raghavan, 2019); here the follower moves the outcome in response to a published forecast, and the leader may itself be strategic. Both are Stackelberg problems with a rational follower. Milli et al. (2019) show a social cost of the leader's advantage in classification; Theorem 5.3 is its analogue for forecasting. Calibration follows Dawid (1982) and Foster & Vohra (1998). Oesterheld et al. (2023) study proper scoring rules that incentivise honest performative predictions. The separation between belief and publication relates to persuasion and cheap talk (Kamenica & Gentzkow, 2011; Crawford & Sobel, 1982), and the intensity to Kyle's price impact (Kyle, 1985); Appendix H.3 gives the mappings.

Limits. Self-certification and the sign results need decreasing differences, not rationality alone. Single crossing can fail for concave responses (Example D.33); in every violating case we tested, the sign conclusion still held. The gap results measure loss against a calibrated benchmark that does not exist; against the best attainable forecast the problem is learnable. Every strategic result concerns self-negating forecasts, where the response lowers the outcome; self-fulfilling forecasts are covered by the learning results only (Proposition B.6). We treat one learner, or several homogeneous honest ones (Proposition G.12), and the learner always leads. A world that leads, a world that steers an unaware learner, consensus between a strategic learner and the world, and competition between learners are the subject of a companion paper and are not claimed here. Open problems are listed in Appendix I.

AI use statement

We used a generative AI assistant to edit the writing and check grammar, and to assist with formatting. The authors reviewed all AI-assisted changes and take full responsibility for the content of this paper.

Ethics statement

This is a theoretical contribution. It involves no human subjects, personal data, or released models or datasets. Sections 5 and 6 show that learners whose forecasts move outcomes can gain by misreporting; we present this, together with the audit results of Section 7, to inform forecast regulation and the design of audits.

Reproducibility statement

All assumptions are stated in Appendix A and in each result, and every main-text result names the appendix result that proves it. The supplementary package contains this source and bibliography; every experiment script (E1–E13, E15–E18) and every verification script (the T series), with fixed seeds; the result files (JSON) behind every figure and table; the scripts that regenerate every figure from those files; the interval-arithmetic certificates of Remark D.37; and a README that maps each figure and table to its script and result file.

References

Maria-Florina Balcan, Avrim Blum, Nika Haghtalab, and Ariel D. Procaccia. Commitment without regrets: Online learning in Stackelberg security games. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pp. 61–78, 2015. doi: 10.1145/2764468.2764478.

Roland Bénabou and Guy Laroque. Using privileged information to manipulate markets: Insiders, gurus, and credibility. *The Quarterly Journal of Economics*, 107(3):921–958, 1992. doi: 10.2307/2118369.

Philip Boeken, Onno Zoeter, and Joris M. Mooij. Conditional forecasts and proper scoring rules for reliable and accurate performative predictions. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-0547. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/17da6ad7f0821c9b62a9116ff9c1cf1b-Abstract-Conference.html.

Craig Boutilier. Eliciting forecasts from self-interested experts: Scoring rules for decision makers. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*, pp. 737–744, 2012. URL https://www.cs.toronto.edu/~cebly/Papers/Boutilier_aamas12.pdf. arXiv:1106.2489.

Josef Broder and Paat Rusmevichientong. Dynamic pricing under a general parametric choice model. *Operations Research*, 60(4):965–980, 2012. doi: 10.1287/opre.1120.1057.

Gavin Brown, Shlomi Hod, and Iden Kalemaj. Performative prediction in a stateful world. In *International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 6045–6061. PMLR, 2022.

Michael Brückner and Tobias Scheffer. Stackelberg games for adversarial prediction problems. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 547–555, 2011. doi: 10.1145/2020408.2020495.

Vincent P. Crawford and Joel Sobel. Strategic information transmission. *Econometrica*, 50(6):1431–1451, 1982. doi: 10.2307/1913390.

A. Philip Dawid. The well-calibrated Bayesian. *Journal of the American Statistical Association*, 77 (379):605–610, 1982. doi: 10.2307/2287720.

Jinshuo Dong, Aaron Roth, Zachary Schutzman, Bo Waggoner, and Zhiwei Steven Wu. Strategic classification from revealed preferences. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pp. 55–70, 2018. doi: 10.1145/3219166.3219193.

Dean P. Foster and Rakesh V. Vohra. Asymptotic calibration. *Biometrika*, 85(2):379–390, 1998. URL http://www.jstor.org/stable/2337364.

Bruce Hajek. Hitting-time and occupation-time bounds implied by drift analysis with applications. *Advances in Applied Probability*, 14(3):502–525, 1982. URL https://resolve.cambridge.org/core/journals/advances-in-applied-probability/article/abs/hittingtime-and-occupationtime-bounds-implied-by-drift-analysis-with-applications/69266B522601C6C5160A2728590AE4FD.

David L. Hanson and Farroll T. Wright. A bound on tail probabilities for quadratic forms in independent random variables. *The Annals of Mathematical Statistics*, 42(3):1079–1083, 1971. doi: 10.1214/aoms/1177693335.

Moritz Hardt and Celestine Mendler-Dünner. Performative prediction: Past and future. *Statistical Science*, 40(3):417–436, 2025. doi: 10.1214/25-STS986. arXiv:2310.16608 (2023).

Moritz Hardt, Nimrod Megiddo, Christos Papadimitriou, and Mary Wootters. Strategic classification. In *Proceedings of the ACM Conference on Innovations in Theoretical Computer Science*, pp. 111–122, 2016. doi: 10.1145/2840728.2840730.

Moritz Hardt, Meena Jagadeesan, and Celestine Mendler-Dünner. Performative power. In *Advances in Neural Information Processing Systems*, volume 35, pp. 22969–22981, 2022.

N. Hemachandra, Kishor Patil, and Sandhya Tripathi. Equilibrium points and equilibrium sets of some GI/M/1 queues. *Queueing Systems*, 96(3–4):245–284, 2020. doi: 10.1007/s11134-020-09677-5.

Yuri I. Ingster and Irina A. Suslina. *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*, volume 169 of *Lecture Notes in Statistics*. Springer, 2003.

Zachary Izzo, Lexing Ying, and James Zou. How to learn when data reacts to your model: Performative gradient descent. In *International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 4641–4650. PMLR, 2021.

Meena Jagadeesan, Tijana Zrnic, and Celestine Mendler-Dünner. Regret minimization with performative feedback. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 9760–9785, 2022. URL https://proceedings.mlr.press/v162/jagadeesan22a.html.

Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101(6): 2590–2615, 2011. doi: 10.1257/aer.101.6.2590.

Navin Kartik. Strategic communication with lying costs. *The Review of Economic Studies*, 76(4): 1359–1395, 2009. doi: 10.1111/j.1467-937X.2009.00559.x.

Navin Kartik, Marco Ottaviani, and Francesco Squintani. Credulity, lies, and costly talk. *Journal of Economic Theory*, 134(1):93–116, 2007. doi: 10.1016/j.jet.2006.04.003.

N. Bora Keskin and Assaf Zeevi. Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies. *Operations Research*, 62(5):1142–1167, 2014. doi: 10.1287/opre.2014.1294.

Jon Kleinberg and Manish Raghavan. How do classifiers induce agents to invest effort strategically? In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pp. 825–844, 2019. doi: 10.1145/3328526.3329584.

Albert S. Kyle. Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1335, 1985.

Tze Leung Lai. Stochastic approximation. *The Annals of Statistics*, 31(2):391–406, 2003. doi: 10.1214/aos/1051027873.

Tze Leung Lai and Herbert Robbins. Adaptive design and stochastic approximation. *The Annals of Statistics*, 7(6):1196–1221, 1979. doi: 10.1214/aos/1176344840.

Tze Leung Lai and Herbert Robbins. Consistency and asymptotic efficiency of slope estimates in stochastic approximation schemes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 56(3):329–360, 1981. doi: 10.1007/BF00536178.

Tze Leung Lai and Herbert Robbins. Adaptive design and the multi-period control problem. In Shanti S. Gupta and James O. Berger (eds.), *Statistical Decision Theory and Related Topics III*, volume 2, pp. 103–120. Academic Press, New York, 1982.

Celestine Mendler-Dünner, Juan C. Perdomo, Tijana Zrnic, and Moritz Hardt. Stochastic optimization for performative prediction. In *Advances in Neural Information Processing Systems*, volume 33, pp. 4929–4939, 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/33e75ff09dd601bbe69f351039152189-Paper.pdf.

John P. Miller, Juan C. Perdomo, and Tijana Zrnic. Outside the echo chamber: Optimizing the performative risk. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 7710–7720, 2021. URL https://proceedings.mlr.press/v139/miller21a.html.

Smitha Milli, John Miller, Anca D. Dragan, and Moritz Hardt. The social cost of strategic classification. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 230–239, 2019. doi: 10.1145/3287560.3287576.

Adhyyan Narang, Evan Faulkner, Dmitriy Drusvyatskiy, Maryam Fazel, and Lillian J. Ratliff. Multiplayer performative prediction: Learning in decision-dependent games. *Journal of Machine Learning Research*, 24(202):1–56, 2023. URL https://jmlr.org/papers/v24/22-0131.html.

Caspar Oesterheld, Johannes Treutlein, Emery Cooper, and Rubi Hudson. Incentivizing honest performative predictions with proper scoring rules. In *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*, volume 216 of *Proceedings of Machine Learning Research*, pp. 1564–1574, 2023. URL https://proceedings.mlr.press/v216/oesterheld23a.html.

Juan C. Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 7599–7609, 2020. URL https://proceedings.mlr.press/v119/perdomo20a.html.

Mitas Ray, Lillian J. Ratliff, Dmitriy Drusvyatskiy, and Maryam Fazel. Decision-dependent risk minimization in geometrically decaying dynamic environments. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 8081–8088, 2022. doi: 10.1609/aaai.v36i7.20780.

Herbert Robbins and David Siegmund. A convergence theorem for non negative almost supermartingales and some applications. In Jagdish S. Rustagi (ed.), *Optimizing Methods in Statistics*, pp. 233–257. Academic Press, New York, 1971. doi: 10.1016/B978-0-12-604550-5.50015-8.

Mark Rudelson and Roman Vershynin. Hanson–Wright inequality and sub-Gaussian concentration. *Electronic Communications in Probability*, 18(82):1–9, 2013. doi: 10.1214/ECP.v18-2865. Paper no. 82.

Joel Sobel. A theory of credibility. *The Review of Economic Studies*, 52(4):557–573, 1985. URL https://ideas.repec.org/a/oup/restud/v52y1985i4p557-573..html.

Jeremy C. Stein. Efficient capital markets, inefficient firms: A model of myopic corporate behavior. *The Quarterly Journal of Economics*, 104(4):655–669, 1989. doi: 10.2307/2937861.

Donald M. Topkis. Minimizing a submodular function on a lattice. *Operations Research*, 26(2):305–321, 1978. doi: 10.1287/opre.26.2.305.

Ching-Zong Wei. Multivariate adaptive stochastic approximation. *The Annals of Statistics*, 15(3):1115–1130, 1987. doi: 10.1214/aos/1176350496.

Tijana Zrnic, Eric Mazumdar, S. Shankar Sastry, and Michael I. Jordan. Who leads and who follows in strategic classification? In *Advances in Neural Information Processing Systems*, volume 34, pp. 15257–15269, 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/812214fb8e7066bfa6e32c626c2c688b-Paper.pdf.

APPENDIX

Appendices A–F follow Sections 2 to 7 of the main text; each main-text result names the result here that proves it. Appendix G collects extensions, Appendix H the experiments, the verification index and extended related work, and Appendix I the open problems. Every result below states its assumptions. Where a proof uses a published theorem, the theorem is stated where it is used.

CONTENTS OF THE APPENDIX

A The model in full **13**

A.1 Assumptions, the residual map and the learner . . . 13

A.2 The world's best response . . . 14

A.3 Discrete agent types . . . 15

A.4 A second instance: strategic queues . . . 15

B Honest targets certify themselves: proofs **16**

B.1 The residual identity . . . 16

B.2 The sign of the feedback . . . 16

B.3 Upper bounds . . . 17

B.4 Tuning the gain from levels . . . 23

B.5 The lower bound . . . 23

C When the equilibrium disappears: proofs **24**

C.1 Unbiasedness versus accuracy . . . 27

C.2 The optimum in the gap: learning the cheaper side . . . 28

D A learner with a stake: proofs **32**

D.1 the stake model with the learner leading and the strategic forecast . . . 32

D.2 The learner's incentive and its welfare cost . . . 34

D.3 The Stackelberg structure: who gains, who pays, and how the cost scales . . . 35

D.4 Orders of play: conflict and manipulation . . . 36

D.5 When manipulation destroys value . . . 40

D.6 Where the conditions fail . . . 41

D.7 Near an activation threshold . . . 42

E Level versus slope: proofs **44**

E.1 The quadratic world: identification and the cost of learning to lead . . . 44

E.2 Scaled rational worlds . . . 46

E.3 An explore-then-commit baseline for the honest forecast . . . 48

F Calibration as evidence: proofs **49**

F.1 The pooled audit . . . 49

F.2 When the pooled audit fails . . . 54

F.3 The binned audit 55
F.4 Summary of the scaled setting 67

G Extensions **67**
G.1 Beyond the scalar model 67
G.2 Robustness 70

H Experiments, verification index and extended related work **70**
H.1 Experiments 70
H.2 Verification index 78
H.3 Extended related work: persuasion, cheap talk and price impact 81

I Open problems **82**

Table 2: Figure index.

Figure	Where	What it shows	Result	Experiment
1	Section 1	the argument as a loop of questions	all	none
2	Section 4	residual jump; bias–accuracy frontier	Theorems 4.1 and 4.2	exact
3	Section 6	regret for the honest and strategic targets	Theorem 6.2	E2, E11
4	Appendix A	timing of the game	model	none
5	Appendix H.1	when the equilibrium is learnable	Propositions B.3, B.7 and B.9	E5, E6
6	Appendix H.1	the sharp constant	Theorem B.21	E1
7	Appendix H.1	lower bound; price of not tuning	Theorem B.29 and Proposition B.22	E2, E3
8	Appendix H.1	the gap regime	Theorem C.8 and Proposition C.7	E7, E9
9	Appendix H.1	the cheaper side of the gap	Theorem C.19	E8
10	Appendix H.1	geometry of the rent	Theorem D.7 and Proposition D.9	E10
11	Appendix H.1	learning the strategic forecast	Theorems E.2 and E.3 and Proposition D.11	E11, E12
12	Appendix H.1	the pooled audit	Theorem F.5 and Proposition F.9	E15, E16
13	Appendix H.1	adaptive gain; closed loop	Proposition B.24 and Theorem F.18	E4, E17
14	Appendix H.1	tuning the gain from levels	Remark B.27	E18

A The model in full

A.1 Assumptions, the residual map and the learner

Assumption A.1 (Timing)**.** At rounds $t = 1, \dots, T$ the learner publishes $\hat{p}_t \in \mathcal{P} := [0, P]$, $P > 0$, measurable with respect to $\mathcal{F}_{t-1}$, the σ-field generated by $y_1, \dots, y_{t-1}$ and by the learner's own randomisation U, which is independent of everything else. The world responds with $\lambda_t = \lambda^*(\hat{p}_t)$ and the outcome is

$$y_t = \mu - \Phi(\hat{p}_t) + \eta_t, \qquad \Phi(p) := \lambda^*(p)\, p. \tag{A.1}$$

Assumption A.2 (Noise)**.** $\mathbb{E}[\eta_t \mid \mathcal{F}_{t-1}] = 0$, $\mathbb{E}[\eta_t^2 \mid \mathcal{F}_{t-1}] = \sigma^2(\hat{p}_t) \le \bar{\sigma}^2$, and $\mathbb{E}[|\eta_t|^j \mid \mathcal{F}_{t-1}] \le M_j$ for every j up to the order required by the result (at most $j \le 6$). In the base model $\sigma^2(\cdot) \equiv \sigma^2$. **(A2-G)** Where stated: η_t are i.i.d. $N(0, \sigma^2)$ and independent of U.

Assumption A.3 (Response)**.** $\Phi : \mathcal{P} \to \mathbb{R}$ is bounded and measurable.

Definition A.4 (Residual map)**.** $r(p) := p + \Phi(p) - \mu$. An *equilibrium* is a point $p^* \in \mathcal{P}$ with $r(p^*) = 0$.

Table 3: Appendix notation, extending Table 1.

$\mathcal{F}_{t-1}$, U	history before round t; the learner's randomisation
$z_t = \hat{p}_t - y_t$, $e_t = \hat{p}_t - p^*$, $u_t = \mathbb{E}e_t^2$	observed error; distance to the honest forecast; its mean square
(M1)–(M3)	root, modulus m, local Lipschitz bound L (Assumption A.5)
(A2-G)	i.i.d. Gaussian noise (Assumption A.2)
$c = aR'$	normalised gain; stable when $c > \frac{1}{2}$
$s = \sqrt{1+4\gamma\mu}$	in the quadratic world, $p^* = 2\mu/(s+1)$ and $R' = s$
the jump setting (G1)–(G6)	response with one jump at x_0 (Assumption C.1); $u = -r^-$, $v = r^+$, $B = \sup_{\mathcal{P}} \lvert r \rvert$
G_k, δG_{k+1}	slopes of the discrete-type intensity and their jumps (Appendix A.3)
ζ	target shift of shifted RSA (Theorem C.10); $\omega = u/(u+v)$ occupation fraction
Ψ, ψ, h	smoothing distribution, its density, smoothing width (Proposition C.12)
the stake model	U_A, U_W, L (Assumption D.19); D_A, W_A direct effect and wedge (Definition D.23)
the scaled setting	$\Phi_\theta = \theta\varphi$, (μ, θ) in a box $\mathfrak{B}$ (Assumption E.11); $F = V'/2$
ETC-S, B_V, C_S	explore-then-commit for the strategic forecast and its constants (Theorem E.2)
Σ_T, z, z_T, β, q_z	pooled statistic, level, level $\sqrt{2\ln T}$, false-alarm level, flagging curve (Assumption F.1 and Lemma F.3)
$W(\rho)$, W_1, ρ_s	learner's cost as a function of its bias, its slope at 0, its minimiser (Lemma F.2)
w, $n_{\min}$, M, Γ_T	bin width, minimum count, number of bins, gain bound (Definition F.22 and Theorem F.23)
$\Phi_{\mathcal{N}}$, $\bar{\Phi}_{\mathcal{N}}$, $\phi_{\mathcal{N}}$	standard normal distribution, tail, density
$\mathbf{1}$	indicator
the vector setting	vector forecasts with a monotone response operator (Assumption G.1)
E#, T#	experiments and verification checks (Appendices H.1 and H.2)

Assumption A.5 (Root, modulus, local Lipschitz bound)**.** **(M1)** there is $p^* \in \mathcal{P}$ with $r(p^*) = 0$; **(M2)** $(p - p^*)\, r(p) \geq m\,(p - p^*)^2$ for all $p \in \mathcal{P}$, for some $m > 0$; **(M3)** $|r(p)| \leq L\,|p - p^*|$ for all $p \in \mathcal{P}$.

Loss and regret. The loss is $\ell_t = (\hat{p}_t - y_t)^2$ and

$$\mathrm{Regret}_T := \sum_{t=1}^{T} \mathbb{E}\big[r(\hat{p}_t)^2\big].$$

In the base model this equals $\sum_t \mathbb{E}\ell_t - T\sigma^2$, the excess over the oracle that plays p^* every round (Lemma B.1).

Quadratic the quadratic world. $\Phi(p) = \gamma p^2$ with $\gamma, \mu > 0$ and $P \geq \mu$. Write $s := \sqrt{1+4\gamma\mu}$. Then $p^* = 2\mu/(s+1) \in (0, \mu)$ and, for $p \in \mathcal{P}$,

$$r(p) - r(p^*) = (p - p^*)\big(1 + \gamma(p + p^*)\big), \qquad 1 + \gamma(p + p^*) \in \big[1 + \gamma p^*,\ 1 + \gamma(P + p^*)\big].$$

Hence in the quadratic world, (M2) holds with $m = 1 + \gamma p^* = (s+1)/2$, (M3) holds with $L = 1 + \gamma(P + p^*)$, and $R' := r'(p^*) = 1 + 2\gamma p^* = s$.

Algorithm A.6 (Residual stochastic approximation, RSA(a))**.** Choose $\hat{p}_1 \in \mathcal{P}$. At each round observe y_t, set $z_t := \hat{p}_t - y_t$, and update

$$\hat{p}_{t+1} := \Pi_{\mathcal{P}}\Big(\hat{p}_t - \frac{a}{t}\, z_t\Big),$$

where $\Pi_{\mathcal{P}}$ is the Euclidean projection onto $\mathcal{P}$. The algorithm uses no knowledge of μ, Φ, γ or σ.

Notation. $e_t := \hat{p}_t - p^*$ and $u_t := \mathbb{E}e_t^2$.

A.2 The world's best response

Proposition A.7 (Best response under the world's cost)**.** *Let* $C_W(\lambda) = \alpha_W \lambda^2 + \beta_W(\lambda - \kappa_W \hat{p})^2$ *with* $\alpha_W, \beta_W, \kappa_W > 0$*, minimised over* $\lambda \geq 0$*. The minimiser is* $\lambda^*(\hat{p}) = \max(0, \gamma\hat{p})$ *with* $\gamma := \beta_W\kappa_W/(\alpha_W + \beta_W)$.

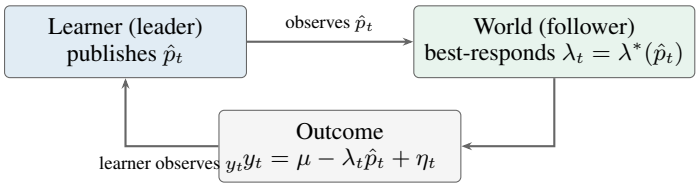

Figure 4: Performative prediction with a rational world. The learner (leader) publishes $\hat{p}_t$; the world (follower) observes it and best-responds with $\lambda_t = \lambda^*(\hat{p}_t)$; the outcome $y_t = \mu - \lambda_t \hat{p}_t + \eta_t$ is realised and the learner observes it. Under the cost of Proposition A.7, $\lambda^*(\hat{p}) = \gamma\hat{p}$; under discrete types (Appendix A.3), $\lambda^*(\hat{p}) = G_k\hat{p}$ on segment k.

Proof. C_W is a quadratic in λ with leading coefficient $\alpha_W + \beta_W > 0$, hence strictly convex. Setting the derivative $2\alpha_W\lambda + 2\beta_W(\lambda - \kappa_W\hat{p})$ to zero gives the unconstrained minimiser $\lambda_0 = \gamma\hat{p}$. A strictly convex function of one variable decreases on $(-\infty, \lambda_0]$ and increases on $[\lambda_0, \infty)$, so its minimiser over $[0, \infty)$ is λ_0 if $\lambda_0 \geq 0$ and 0 otherwise. □

Remark A.8. On $\mathcal{P} = [0, P]$ this gives $\Phi(p) = \gamma p^2$. For $\hat{p} < 0$ the constraint $\lambda \geq 0$ gives $\lambda^* = 0$ and $y = \mu + \eta$, so the formula $y = \mu - \gamma\hat{p}^2$ does not extend to negative forecasts. We restrict to $\hat{p} \geq 0$.

A.3 Discrete agent types

When the world is a population of agents with participation thresholds, the smooth response of Proposition A.7 is replaced by a piecewise-linear intensity. The world is a population of K agent types. Type k has payoff $U_W^{(k)}(\lambda_k, \hat{p}) = \pi_k\lambda_k\hat{p} - \alpha_k\lambda_k^2$, fixed participation cost $F_k > 0$ and population mass $w_k > 0$. Its best response is $\lambda_k^*(\hat{p}) = \pi_k\hat{p}/(2\alpha_k)$, with maximal payoff $\pi_k^2\hat{p}^2/(4\alpha_k)$, so type k participates when this exceeds F_k:

$$\hat{p} \geq \hat{p}_k^* := \frac{2\sqrt{\alpha_k F_k}}{\pi_k}.$$

With types ordered by threshold ($\hat{p}_1^* < \cdots < \hat{p}_K^*$), the aggregate intensity on $[\hat{p}_k^*, \hat{p}_{k+1}^*)$ is

$$\lambda^*(\hat{p}) = G_k\hat{p}, \qquad G_k = \sum_{j \leq k} \frac{w_j\pi_j}{2\alpha_j}, \qquad G_1 < G_2 < \cdots < G_K,$$

so $\Phi(p) = G_k p^2$ on segment k. At the threshold $\hat{p}_k^*$ the intensity jumps by $w_k\sqrt{F_k/\alpha_k}$. Each agent optimises its own payoff given the published forecast and needs no knowledge of μ; the response is rational and nondecreasing, but discontinuous. At a threshold the marginal type is indifferent; we assume its tie-break is deterministic, so λ^* takes one of its two one-sided values there (this is (G4) of Assumption C.1).

For the threshold $x_0 := \hat{p}_{k+1}^*$, no forecast near x_0 is calibrated exactly when μ lies in the *gap window*

$$x_0 + G_k x_0^2 \;<\; \mu \;<\; x_0 + G_{k+1}x_0^2, \tag{A.2}$$

of width $(G_{k+1} - G_k)x_0^2$. In the notation of Assumption C.1 this is $r^- < 0 < r^+$ (Proposition C.3(iii)): the candidate equilibrium of segment k overshoots into segment $k+1$, and that of segment $k+1$ undershoots back.

A.4 A second instance: strategic queues

The abstract gap mechanism (Theorem C.4) is not specific to forecasting. In an M/M/1 queue with service rate $\mu_s > 0$ the waiting time is $W(\lambda) = 1/(\mu_s - \lambda)$ for $\lambda < \mu_s$, increasing and convex, and an operator sets a price $p \geq 0$. There are two customer types; type k has valuation v_k, waiting cost c_k and potential arrival rate Λ_k, and joins if and only if $v_k - p - c_k W \geq 0$. Type 2 is assumed to join in all cases below. For type 1 the condition is that the arrival rate be at most $\lambda_1^{\text{crit}} := \mu_s - c_1/(v_1 - p)$. If $\Lambda_1 + \Lambda_2 \leq \lambda_1^{\text{crit}}$ both types join, and if $\lambda_1^{\text{crit}} < \Lambda_2$ only type 2 joins; both are stable equilibria. If $\Lambda_2 < \lambda_1^{\text{crit}} < \Lambda_1 + \Lambda_2$ there is no static equilibrium: with both types present the arrival rate exceeds

λ_1^{crit} and type 1 leaves; without type 1 it falls below and type 1 joins. Myopic dynamics toggle between Λ_2 and $\Lambda_1 + \Lambda_2$, a 2-cycle with gap Λ_1. Since λ_1^{crit} decreases in p, the toggling region is the price interval (p_L, p_H) with

$$p_L = v_1 - \frac{c_1}{\mu_s - \Lambda_1 - \Lambda_2}, \qquad p_H = v_1 - \frac{c_1}{\mu_s - \Lambda_2}, \qquad p_H - p_L = \frac{c_1 \Lambda_1}{(\mu_s - \Lambda_1 - \Lambda_2)(\mu_s - \Lambda_2)},$$

and the gap Λ_1 is the same throughout it. This is the structure of Theorem C.4: a monotone map with one discontinuity straddling the diagonal. Equilibrium sets of this kind in queues are characterised by Hemachandra et al. (2020).

B Honest targets certify themselves: proofs

B.1 The residual identity

Lemma B.1 (Residual identity). *Under Assumptions A.1 to A.3:*

(i) $z_t = r(\hat{p}_t) - \eta_t$;

(ii) $\mathbb{E}[\ell_t \mid \mathcal{F}_{t-1}] = r(\hat{p}_t)^2 + \sigma^2(\hat{p}_t)$;

(iii) if $\Phi(q) - \Phi(p) \geq -(1-m)(q-p)$ *for all* $p \leq q$ *in* $\mathcal{P}$, *with* $m > 0$, *then* $r(q) - r(p) \geq m(q-p)$. *Consequently* r *is strictly increasing, has at most one root, and satisfies (M2) whenever (M1) holds.*

Proof. (i) $z_t = \hat{p}_t - y_t = \hat{p}_t - \mu + \Phi(\hat{p}_t) - \eta_t = r(\hat{p}_t) - \eta_t$.

(ii) Since $\ell_t = z_t^2 = r(\hat{p}_t)^2 - 2r(\hat{p}_t)\eta_t + \eta_t^2$ and $r(\hat{p}_t)$ is $\mathcal{F}_{t-1}$-measurable (because $\hat{p}_t$ is), taking conditional expectations and using Assumption A.2 gives $r(\hat{p}_t)^2 - 0 + \sigma^2(\hat{p}_t)$.

(iii) For $p \leq q$: $r(q) - r(p) = (q-p) + \Phi(q) - \Phi(p) \geq (q-p) - (1-m)(q-p) = m(q-p)$. Since $m > 0$, r is strictly increasing and has at most one root. For (M2), take $p \in \mathcal{P}$. If $p \geq p^*$, then $r(p) = r(p) - r(p^*) \geq m(p - p^*)$; multiplying by $p - p^* \geq 0$ gives $(p-p^*)r(p) \geq m(p-p^*)^2$. If $p < p^*$, then $r(p^*) - r(p) \geq m(p^* - p)$, so $r(p) \leq m(p - p^*)$; multiplying by $p - p^* < 0$ reverses the inequality and gives the same conclusion. □

Remark B.2. In the base model $\sum_t \mathbb{E}\ell_t - T\sigma^2 = \sum_t \mathbb{E}r(\hat{p}_t)^2 = \text{Regret}_T$, and the oracle that plays p^* has expected loss σ^2 in every round.

B.2 The sign of the feedback

Proposition B.3 (The correct elasticity condition). *Let* $\lambda^* \geq 0$ *be differentiable on* $(0, \infty)$ *and write* $\varepsilon_\lambda(p) := p\,\lambda^{*\prime}(p)/\lambda^*(p)$ *wherever* $\lambda^*(p) > 0$. *Then* $\Phi(p) = p\lambda^*(p)$ *is nondecreasing on* $(0, \infty)$ *if and only if* $\varepsilon_\lambda(p) \geq -1$ *at every* p *with* $\lambda^*(p) > 0$.

Proof. $\Phi'(p) = \lambda^*(p) + p\lambda^{*\prime}(p)$. Where $\lambda^*(p) > 0$ this equals $\lambda^*(p)(1 + \varepsilon_\lambda(p))$, which is nonnegative iff $\varepsilon_\lambda(p) \geq -1$. Where $\lambda^*(p) = 0$ at some $p > 0$, the point p is an interior minimum of the differentiable function $\lambda^* \geq 0$, so $\lambda^{*\prime}(p) = 0$ and $\Phi'(p) = 0$. A differentiable function on an interval is nondecreasing iff its derivative is nonnegative everywhere (mean value theorem). □

Remark B.4. Monotonicity needs a *lower* bound on elasticity, not an upper bound. For $\lambda^*(p) = c/p^2$, $\varepsilon_\lambda \equiv -2$, so $\Phi(p) = c/p$ is decreasing and $r(p) = p + c/p - \mu$ is decreasing on $(0, \sqrt{c})$.

Example B.5 (Convexity alone does not give monotonicity). Let the world's cost be $C(\lambda, p) = (\lambda - 2(1-p))^2$ on $\mathcal{P} = [0, 1]$ with $\lambda \geq 0$. It is convex in λ, but it has increasing differences in (λ, p), so its best response $\lambda^*(p) = 2(1-p)$ decreases. Then $\Phi(p) = 2p(1-p)$ and, with $\mu = 1.05$, $r(p) = 3p - 2p^2 - 1.05$ has two roots, 0.5563 and 0.9436: two calibrated forecasts, and r is decreasing on $(3/4, 1]$ (T62). Decreasing differences, not convexity, is what Lemma B.1(iii) needs.

Proposition B.6 (One condition for both feedback signs). *The condition* $\Phi(q) - \Phi(p) \geq -(1-m)(q-p)$ *of Lemma B.1(iii) covers negative feedback of any size (*Φ *nondecreasing,* $m = 1$*) and positive feedback that is Lipschitz with constant* $L_{\text{MD}} < 1$ *(*$m = 1 - L_{\text{MD}}$*). The second case*

is exactly the regime of Mendler-Dünner et al. (2020): their greedy-deploy SGD converges when the sensitivity L_{MD} (in Wasserstein-1 distance) satisfies $L_{\mathrm{MD}} < \gamma_{\mathrm{MD}}/\beta_{\mathrm{MD}}$, with effective strong convexity $\gamma_{\mathrm{MD}} - L_{\mathrm{MD}}\beta_{\mathrm{MD}}$ (as stated in the survey of Hardt & Mendler-Dünner, 2025). For squared loss $\gamma_{\mathrm{MD}} = \beta_{\mathrm{MD}} = 2$, and here $W_1(\mathcal{D}(p), \mathcal{D}(p')) = |\Phi(p) - \Phi(p')|$, so their condition is $\mathrm{Lip}(\Phi) < 1$ with effective modulus $2(1 - L_{\mathrm{MD}})$, matching $m = 1 - L_{\mathrm{MD}}$. The first case (nondecreasing Φ of any slope) lies outside their condition.

Proof. If Φ is nondecreasing, $\Phi(q) - \Phi(p) \geq 0 \geq -(1-m)(q-p)$ with $m = 1$. If $|\Phi(q) - \Phi(p)| \leq L_{\mathrm{MD}}|q - p|$, then $\Phi(q) - \Phi(p) \geq -L_{\mathrm{MD}}(q - p) = -(1 - m)(q - p)$ with $m = 1 - L_{\mathrm{MD}}$. □

Proposition B.7 (Strong positive feedback breaks monotonicity)**.** *If Φ is differentiable at p_0 with $\Phi'(p_0) < -1$, then $r'(p_0) < 0$, so r is not monotone.*

Proof. $r'(p_0) = 1 + \Phi'(p_0) < 0$. □

Remark B.8 (Example). Take $\Phi(p) = -3\tanh(p - 2)$ and $\mu = 2$, so that $y = \mu + 3\tanh(p - 2) + \eta$ rises with the prediction. Then $r(2) = 0$ exactly and $r'(2) = 1 - 3 = -2$. Numerically (T12), r has roots at -0.9847, 2 and 4.9847: several self-fulfilling equilibria, one of them unstable.

Proposition B.9 (A random response forces linear regret)**.** *Let λ_t be i.i.d. uniform on $\{0, \Lambda\}$ with $\Lambda > 0$, independent of $\mathcal{F}_{t-1}$ and of η_t, with $y_t = \mu - \lambda_t \hat{p}_t + \eta_t$ and $\eta_t \sim N(0, \sigma^2)$ i.i.d. The per-round oracle, which knows λ_t, plays $\mu/(1 + \lambda_t)$ and has expected loss σ^2. For every algorithm,*

$$\sum_{t=1}^{T} \mathbb{E}\ell_t - T\sigma^2 \;\geq\; T \cdot \frac{\mu^2\Lambda^2}{2(\Lambda^2 + 2\Lambda + 2)}.$$

Proof. The cross terms with η_t have zero mean, and λ_t is independent of $\hat{p}_t$, so

$$\mathbb{E}[\ell_t \mid \mathcal{F}_{t-1}] = F(\hat{p}_t) + \sigma^2, \qquad F(p) := \tfrac{1}{2}(p - \mu)^2 + \tfrac{1}{2}\big(p(1 + \Lambda) - \mu\big)^2.$$

For numbers $a_1, \ldots, a_n$ not all zero, $\min_p \sum_i (a_i p - \mu)^2 = \mu^2\big(n - (\sum a_i)^2/\sum a_i^2\big)$: setting the derivative $2\sum_i a_i(a_i p - \mu)$ to zero gives $p_0 = \mu \sum a_i / \sum a_i^2$, and substituting gives $\mu^2(n - 2(\sum a_i)^2/\sum a_i^2 + (\sum a_i)^2/\sum a_i^2)$. With $n = 2$, $a = (1, 1+\Lambda)$: $\sum a_i = 2+\Lambda$, $\sum a_i^2 = \Lambda^2 + 2\Lambda + 2$, and $2 - (2 + \Lambda)^2/(\Lambda^2 + 2\Lambda + 2) = \Lambda^2/(\Lambda^2 + 2\Lambda + 2)$. Multiplying by $\frac{1}{2}$ gives $\min F$; summing over t completes the proof. □

Remark B.10. The world here is random, not adversarial, and the mean outcome $\mathbb{E}[y \mid p] = \mu - \Lambda p/2$ is a fixed function of the forecast. The linear regret is measured against the oracle that knows λ_t in each round; against the best fixed forecast the problem is learnable.

B.3 Upper bounds

Lemma B.11 (Recursion)**.** *Let $v_t \geq 0$ with $v_t \leq V_0$ for all $t \geq 1$. Let $k > 0$ be real, $c > k$, $C \geq 0$, $D \geq 0$, and let $t_1 \geq \max(c, 2C/(c - k))$ be an integer. Suppose that for every $t \geq t_1$*

$$v_{t+1} \leq \Big(1 - \frac{c}{t} + \frac{C}{t^2}\Big) v_t + \frac{D}{t^{k+1}}.$$

Then $v_t \leq M/t^k$ for all $t \geq 1$, where $M := \max\big(V_0 t_1^k,\ 2D/(c - k)\big)$.

Proof. For $t \leq t_1$: $v_t \leq V_0 = V_0 t_1^k / t_1^k \leq V_0 t_1^k / t^k \leq M/t^k$. Induction for $t \geq t_1$: assume $v_t \leq M/t^k$. Since $t \geq c$, the coefficient $1 - c/t + C/t^2 \geq 1 - c/t \geq 0$, so substituting the bound is valid:

$$v_{t+1} \leq \frac{M}{t^k} - \frac{cM}{t^{k+1}} + \frac{CM}{t^{k+2}} + \frac{D}{t^{k+1}}.$$

The map $x \mapsto (1 + x)^{-k}$ is convex on $x > -1$, so it lies above its tangent at 0: $(1 + x)^{-k} \geq 1 - kx$. With $x = 1/t$, $M/(t+1)^k \geq M/t^k - kM/t^{k+1}$. It therefore suffices that $(c - k)M - D - CM/t \geq 0$. Since $t \geq t_1 \geq 2C/(c - k)$, $CM/t \leq (c - k)M/2$, so it is enough that $(c - k)M/2 \geq D$, i.e. $M \geq 2D/(c - k)$. □

Lemma B.12 (Moment bounds). *Under Assumptions A.1 and A.2 with conditional moments up to order $2k$, and (M1)–(M3), with $2am > 1$, and let $k \geq 1$ be an integer. Then RSA(a) satisfies $\mathbb{E}e_t^{2k} \leq M_k' \, t^{-k}$ for all $t \geq 1$, with constants given in Step 5. (The induction below runs for every k. Gaussian noise has all moments, so all k are available there.)*

Proof. Write $\mathcal{F} = \mathcal{F}_{t-1}$ and $y_t^u := \hat{p}_t - (a/t)z_t$ for the unprojected point.

Step 0 (projection). $\Pi_{\mathcal{P}}$ is nonexpansive and $\Pi_{\mathcal{P}}(p^*) = p^*$, so $|e_{t+1}| \leq |y_t^u - p^*| = |w_t + v_t|$, where $w_t := e_t - (a/t)r(\hat{p}_t)$ is $\mathcal{F}$-measurable and $v_t := (a/t)\eta_t$, using Lemma B.1(i).

Step 1 (bound on w). $w_t^2 = e_t^2 - 2(a/t)e_t r(\hat{p}_t) + (a/t)^2 r(\hat{p}_t)^2$. By (M2), $e_t r \geq m e_t^2$; by (M3), $r^2 \leq L^2 e_t^2$. Hence $w_t^2 \leq q_t e_t^2$ with

$$q_t := 1 - \frac{2am}{t} + \frac{a^2L^2}{t^2} = \Big(1 - \frac{am}{t}\Big)^2 + \frac{a^2(L^2 - m^2)}{t^2}.$$

Since $me^2 \leq e\,r \leq |e||r| \leq Le^2$ for $p \neq p^*$, $L \geq m$ and $q_t \geq 0$. Also $|w_t| \leq (1 + aL)|e_t|$.

Step 2 (binomial expansion). Because $2k$ is even, $e_{t+1}^{2k} \leq (w_t + v_t)^{2k} = \sum_{j=0}^{2k} \binom{2k}{j} w_t^{2k-j} v_t^j$. Taking conditional expectations: the $j = 0$ term is w_t^{2k}; the $j = 1$ term is $2k\, w_t^{2k-1}(a/t)\mathbb{E}[\eta_t \mid \mathcal{F}] = 0$; for $j \geq 2$ the absolute value is at most $\binom{2k}{j}(1 + aL)^{2k-j}|e_t|^{2k-j}(a/t)^j M_j$.

Step 3 (main term). $w_t^{2k} \leq q_t^k e_t^{2k}$. Put $b_t := 2am/t - a^2L^2/t^2$, so $q_t = 1 - b_t$. For $t \geq t_a := \lceil aL^2/(2m) \rceil$ we have $0 \leq b_t \leq 2am/t$, and $b_t \leq 1$ because $q_t \geq 0$. We use $(1-b)^k \leq 1 - kb + \binom{k}{2}b^2$ for $b \in [0, 1]$: for $k = 1$ it is an equality; for $k \geq 2$ let $f(b) := 1 - kb + \binom{k}{2}b^2 - (1-b)^k$, then $f(0) = 0$, $f'(b) = -k + k(k-1)b + k(1-b)^{k-1}$ gives $f'(0) = 0$, and $f''(b) = k(k-1)\big(1 - (1-b)^{k-2}\big) \geq 0$ on $[0, 1]$, so $f' \geq 0$ and $f \geq 0$. Hence for $t \geq t_a$,

$$q_t^k \leq 1 - kb_t + \tbinom{k}{2}b_t^2 \leq 1 - \frac{c}{t} + \frac{C}{t^2}, \qquad c := 2amk, \quad C := ka^2L^2 + 2k(k-1)a^2m^2.$$

Step 4 (remainder, by induction on k). Let $v_t^{(k)} := \mathbb{E}e_t^{2k}$. Then

$$v_{t+1}^{(k)} \leq \Big(1 - \frac{c}{t} + \frac{C}{t^2}\Big)v_t^{(k)} + \sum_{j=2}^{2k} \binom{2k}{j} (1 + aL)^{2k-j} a^j M_j \, t^{-j} \, \mathbb{E}|e_t|^{2k-j}.$$

For $k = 1$ only $j = 2$ appears, with $\mathbb{E}|e_t|^0 = 1$; the remainder is D_1/t^2 with $D_1 := a^2M_2$. For $k \geq 2$ assume $\mathbb{E}e_t^{2k-2} \leq M_{k-1}' t^{-(k-1)}$. For $0 < q \leq 2k - 2$, Lyapunov's inequality gives $\mathbb{E}|e_t|^q \leq (\mathbb{E}e_t^{2k-2})^{q/(2k-2)} \leq \max(1, M_{k-1}')\, t^{-q/2}$. With $q = 2k - j$ each term is of order $t^{-j}t^{-(2k-j)/2} = t^{-k-j/2} \leq t^{-k-1}$; for $j = 2k$ it is of order $t^{-2k} \leq t^{-k-1}$. So the remainder is at most D_k/t^{k+1} with $D_k := \sum_{j=2}^{2k} \binom{2k}{j}(1 + aL)^{2k-j} a^j M_j \max(1, M_{k-1}')$.

Step 5. Apply Lemma B.11 with $V_0 = P^{2k}$ (since $|e_t| \leq P$), $c = 2amk > k \iff 2am > 1$, the C of Step 3, $D = D_k$ and $t_1 := \max\big(t_a, \lceil c \rceil, \lceil 2C/(c-k) \rceil\big)$. This gives $M_k' := \max\big(P^{2k}t_1^k, 2D_k/(c-k)\big)$. □

Theorem B.13 (Logarithmic regret). *Under Assumptions A.1 to A.3, (M1)–(M3) and $2am > 1$, RSA(a) satisfies, for all $T \geq 1$,*

$$u_t \leq M_1'/t \qquad \textit{and} \qquad \mathrm{Regret}_T \leq L^2 M_1'(1 + \ln T),$$

with $M_1' = \max\big(P^2 t_1,\, 2a^2\bar{\sigma}^2/(2am - 1)\big)$ and $t_1 = \max\big(\lceil aL^2/(2m) \rceil, \lceil 2am \rceil, \lceil 2a^2L^2/(2am - 1) \rceil\big)$.

Proof. The first claim is Lemma B.12 with $k = 1$. For the second, by (M3) and $\sum_{t \leq T} 1/t \leq 1 + \int_1^T dx/x$, $\mathrm{Regret}_T = \sum_t \mathbb{E}r(\hat{p}_t)^2 \leq L^2 \sum_t u_t \leq L^2 M_1' \sum_{t \leq T} t^{-1} \leq L^2 M_1'(1 + \ln T)$. □

Remark B.14. The constant here depends on L, m, a and $\bar{\sigma}$; the sharp constant σ^2 is asymptotic (Theorem B.17). In the quadratic world, $m = (s+1)/2$, so Theorem B.13 holds for every $a > 1/(s+1)$, a range that contains both $a = 1$ and the tuned gain $a = 1/s$.

Corollary B.15 (Almost-sure convergence)**.** *Under the assumptions of Lemma B.12 with $k = 2$, $\hat{p}_t \to p^*$ almost surely.*

Proof. By monotone convergence, $\mathbb{E}\sum_t e_t^4 = \sum_t \mathbb{E}e_t^4 \le M_2' \sum_t t^{-2} < \infty$. A nonnegative random variable with finite expectation is finite a.s., so $\sum_t e_t^4 < \infty$ a.s. and $e_t \to 0$ a.s. □

Lemma B.16 (Chung-type limit)**.** *Let $u_t \ge 0$ with $u_t \le M/t$, and suppose that for all large t, $u_{t+1} = (1 - c/t)u_t + d/t^2 + \varepsilon_t$ with $c > 1$ and $t^2\varepsilon_t \to 0$. Then $tu_t \to d/(c-1)$.*

Proof. Let $s_t := tu_t$ (bounded) and $x_t := s_t - d/(c-1)$. Multiplying the recursion by $t+1$ and using $(t+1)(1-c/t) = t + 1 - c - c/t$,

$$s_{t+1} = s_t\Big(1 + \frac{1-c}{t} - \frac{c}{t^2}\Big) + \frac{d}{t} + \frac{d}{t^2} + (t+1)\varepsilon_t.$$

Substituting $s_t = x_t + d/(c-1)$ in the terms of order $1/t$ gives $-(c-1)s_t/t + d/t = -(c-1)x_t/t$. Hence $x_{t+1} = (1 - b/t)x_t + \zeta_t$ with $b := c - 1 > 0$ and $\zeta_t := -cs_t/t^2 + d/t^2 + (t+1)\varepsilon_t$, so $t\zeta_t \to 0$. For $t > b$, $a_t := |x_t|$ satisfies $a_{t+1} \le (1 - b/t)a_t + \beta_t/t$ with $\beta_t := t|\zeta_t| \to 0$. Fix $\varepsilon > 0$ and $t_2 > b$ with $\beta_t \le \varepsilon$ for $t \ge t_2$. With $a_t' := a_t - \varepsilon/b$,

$$a_{t+1}' \le (1 - b/t)a_t + \varepsilon/t - \varepsilon/b = (1 - b/t)a_t'.$$

If $a_t' \le 0$ for some $t \ge t_2$ then $a_s' \le 0$ for all later s, since the coefficient is positive. Otherwise $a_T' \le a_{t_2}' \prod_{t=t_2}^{T-1}(1 - b/t) \to 0$ because $\sum b/t = \infty$. In either case $\limsup a_t \le \varepsilon/b$; ε was arbitrary. □

Theorem B.17 (Sharp constant)**.** *Assume Assumption A.1; Assumption A.2 with moments up to order 6; (M1) with p^* in the interior of $\mathcal{P}$ and $D_0 := \min(p^*, P - p^*) > 0$; (M2); (M3); $2am > 1$; **(B1)** r is differentiable at p^* with derivative R'; **(B2)** $\sigma^2(\cdot)$ is continuous at p^*. Then $R' \ge m$ and*

$$tu_t \to \frac{a^2\sigma^2(p^*)}{2aR' - 1}, \qquad \frac{\mathrm{Regret}_T}{\ln T} \to \frac{a^2R'^2\sigma^2(p^*)}{2aR' - 1}.$$

Proof. **Step 0.** By (M2), $r(p)/(p - p^*) \ge m$ for $p \ne p^*$; letting $p \to p^*$ gives $R' \ge m$, so $c := 2aR' \ge 2am > 1$. Define $\rho(e) := r(p^* + e) - R'e$ and $\omega(\delta) := \sup\{|\rho(e)|/|e| : 0 < |e| \le \delta,\ p^* + e \in \mathcal{P}\}$; $\omega(\delta) \to 0$ by (B1), and $|\rho(e)| \le (L + R')|e|$ by (M3). Lemma B.12 gives $u_t \le M_1'/t$, $\mathbb{E}e_t^4 \le M_2'/t^2$, $\mathbb{E}e_t^6 \le M_3'/t^3$.

Step 1 (unprojected second moment). Let $f_t := w_t + v_t = y_t^u - p^*$. Since $\mathbb{E}[v_t \mid \mathcal{F}] = 0$ and $\mathbb{E}[v_t^2 \mid \mathcal{F}] = (a/t)^2\sigma^2(\hat{p}_t)$, and $e_t r = R'e_t^2 + e_t\rho(e_t)$,

$$\mathbb{E}f_t^2 = \Big(1 - \frac{2aR'}{t}\Big)u_t - \frac{2a}{t}\mathbb{E}[e_t\rho(e_t)] + \frac{a^2}{t^2}\mathbb{E}r(\hat{p}_t)^2 + \frac{a^2}{t^2}\mathbb{E}\sigma^2(\hat{p}_t).$$

Step 2 (cost of projection). If $y_t^u \in \mathcal{P}$ then $e_{t+1} = f_t$; otherwise $|e_{t+1}| \le |f_t|$ and $|f_t| > D_0$. Hence $0 \le \mathbb{E}f_t^2 - u_{t+1} \le \mathbb{E}[f_t^2\mathbf{1}\{|f_t| \ge D_0\}] \le \mathbb{E}f_t^6/D_0^4$, and $(x+y)^6 \le 32(x^6 + y^6)$ gives $\mathbb{E}f_t^6 \le 32\big((1 + aL)^6M_3't^{-3} + a^6M_6t^{-6}\big)$. The projection gap is at most C_Pt^{-3}.

Step 3 (recursion). Combining, $u_{t+1} = (1 - c/t)u_t + d/t^2 + \varepsilon_t$ with $d := a^2\sigma^2(p^*)$ and $\varepsilon_t = -\frac{2a}{t}\mathbb{E}[e\rho(e)] + \frac{a^2}{t^2}\mathbb{E}r^2 + \frac{a^2}{t^2}\big(\mathbb{E}\sigma^2(\hat{p}_t) - \sigma^2(p^*)\big) - (\mathbb{E}f_t^2 - u_{t+1})$.

(a) Splitting on $|e| \le \delta$ and using $e^2\mathbf{1}\{|e| > \delta\} \le e^4/\delta^2$: $|\mathbb{E}[e\rho(e)]| \le \omega(\delta)M_1'/t + (L + R')M_2'/(\delta^2t^2)$.

(b) $(a^2/t^2)\mathbb{E}r^2 \le a^2L^2M_1'/t^3$.

(c) $\hat{p}_t \to p^*$ a.s. (Corollary B.15) and (B2) give $\sigma^2(\hat{p}_t) \to \sigma^2(p^*)$ a.s.; since $\sigma^2 \le \bar{\sigma}^2$, bounded convergence gives $\mathbb{E}\sigma^2(\hat{p}_t) \to \sigma^2(p^*)$, so this part is $o(t^{-2})$.

(d) At most $C_P t^{-3}$. Together, $\limsup_t t^2|\varepsilon_t| \le 2aM_1'\omega(\delta)$ for every $\delta > 0$, so $t^2\varepsilon_t \to 0$.

Step 4. Lemma B.16 gives $tu_t \to d/(c-1)$.

Step 5 (from e to the excess). $r^2 = R'^2e^2 + 2R'e\rho + \rho^2$. On $|e| \le \delta$, $|2R'e\rho + \rho^2| \le (2R'\omega(\delta) + \omega(\delta)^2)e^2$; on $|e| > \delta$, it is at most $C_\rho e^4/\delta^2$. Hence $t\,|\mathbb{E}r(\hat{p}_t)^2 - R'^2u_t| \le (2R'\omega(\delta) + \omega(\delta)^2)M_1' + C_\rho M_2'/(\delta^2 t)$. Letting $t \to \infty$ then $\delta \to 0$: $t\,\mathbb{E}r(\hat{p}_t)^2 \to \kappa := R'^2 d/(c-1)$.

Step 6 (summation). With $A_T := \sum_{t\le T} \mathbb{E}r(\hat{p}_t)^2$ and $B_T := \sum_{t\le T} 1/t \uparrow \infty$, Stolz–Cesàro gives $A_T/B_T \to \lim (T+1)\mathbb{E}r(\hat{p}_{T+1})^2 = \kappa$, and $B_T/\ln T \to 1$. □

Corollary B.18 (Optimal gain)**.** *Let $g(a) := a^2R'^2/(2aR'-1)$ on $a > 1/(2R')$. Then $g'(a) = 2aR'^2(aR'-1)/(2aR'-1)^2$, so g has a unique minimiser at $a = 1/R'$ with $g(1/R') = 1$. In the quadratic world, $a = 1/s$ satisfies $2am = (s+1)/s > 1$, so Theorem B.17 applies and gives leading constant exactly σ^2.*

Proof. Quotient rule: $g'(a) = [2aR'^2(2aR'-1) - 2a^2R'^3]/(2aR'-1)^2 = 2aR'^2(aR'-1)/(2aR'-1)^2$, negative for $a < 1/R'$ and positive for $a > 1/R'$; and $g(1/R') = 1$. □

Remark B.19 (Scope of the gain condition). Theorem B.17 as stated assumes $2am > 1$. For general Φ with $R' \ge 2m$ the tuned gain violates this. Theorem B.21 below removes the condition: the formula holds in expectation for every gain with $2aR' > 1$, so the tuned gain $a = 1/R'$ is always covered. Experiment E1 (and T52) shows that when the global slope is weak the finite-T transient is large, although the asymptotic slope is correct.

Lemma B.20 (Moments without the global gain condition)**.** *Under Assumptions A.1 and A.2 with conditional moments up to order $2k_0$, and (M1)–(M3), with any gain $a > 0$. Fix $\rho \in (0, \min(2am, 1))$.*

(i) For every integer $k \le k_0$, $\mathbb{E}e_t^{2k} \le M_k t^{-\rho k}$ for all $t \ge 1$.

(ii) If $\rho k_0 > 1$, then $\hat{p}_t \to p^*$ almost surely.

(iii) If moreover p^* is interior to $\mathcal{P}$, r is differentiable at p^* with $2aR' > 1$, and $\rho k_0 \ge 4$, then $\mathbb{E}e_t^{2k} \le M_k' t^{-k}$ for $k = 1, 2, 3$: the conclusions of Lemma B.12 hold.

Proof. Write $r_t := r(\hat{p}_t)$, let ξ_t be the observation noise in the update (conditional mean 0, conditional moments S_j of order $j \le 2k_0$), and let $w_t := e_t - (a/t)(r_t + \xi_t)$ be the unprojected error. Projection onto $\mathcal{P}$ is nonexpansive and $p^* \in \mathcal{P}$, so $|e_{t+1}| \le |w_t|$.

Expansion. $w_t^{2k} = \sum_{j=0}^{2k} \binom{2k}{j} e_t^{2k-j}\big(-\frac{a}{t}(r_t+\xi_t)\big)^j$. The $j=1$ term has conditional expectation $-(2ka/t)e_t^{2k-2}(e_t r_t)$. For $j \ge 2$, $|r_t + \xi_t|^j \le 2^{j-1}(L^j|e_t|^j + |\xi_t|^j)$ since $|r_t| \le L|e_t|$ by (M3) and $r(p^*) = 0$. Hence

$$\mathbb{E}[e_{t+1}^{2k} \mid \mathcal{F}_t] \le e_t^{2k}\Big(1 + \frac{C_k}{t^2}\Big) - \frac{2ka}{t}e_t^{2k-2}(e_t r_t) + C_k \sum_{j=2}^{2k} t^{-j}|e_t|^{2k-j}. \tag{$*$}$$

(i) Induction on k. By (M2), $e_t r_t \ge m e_t^2$, so the middle term is at most $-(2kam/t)e_t^{2k}$. For $2 \le j \le 2k$, Lyapunov's inequality and the induction hypothesis give $\mathbb{E}|e_t|^{2k-j} \le (\mathbb{E}e_t^{2k-2})^{(2k-j)/(2k-2)} \le Ct^{-\rho(2k-j)/2}$ (for $k=1$ only $j=2$ occurs, with $\mathbb{E}|e_t|^0 = 1$). The exponent satisfies $j + \rho(2k-j)/2 = \rho k + j(1-\rho/2) \ge \rho k + 2 - \rho \ge \rho k + 1$, because $j \ge 2$ and $\rho \le 1$. So $v_t := \mathbb{E}e_t^{2k}$ satisfies $v_{t+1} \le (1 - 2kam/t + C_k/t^2)v_t + Dt^{-(\rho k+1)}$, with $v_t \le P^{2k}$. Since $\rho < 2am$, $2kam > \rho k$, and Lemma B.11 with exponent ρk gives $v_t \le M_k t^{-\rho k}$.

(ii) Take $k \le k_0$ with $\rho k > 1$. Then $\mathbb{E}\sum_t e_t^{2k} < \infty$, so $\sum_t e_t^{2k} < \infty$ a.s. and $e_t \to 0$ a.s.

(iii) Localisation. Choose $R_- < R'$ with $2aR_- > 1$, and $\delta > 0$ with $[p^* - \delta, p^* + \delta] \subset \mathcal{P}$ and $e\,r(p^* + e) \ge R_- e^2$ for $|e| \le \delta$ (differentiability at p^*). Since $e_t r_t \ge 0$ everywhere, $e_t r_t \ge R_- e_t^2 - R_- e_t^2 \mathbf{1}\{|e_t| > \delta\}$. In $(*)$ this gives the middle term at most $-(2kaR_-/t)e_t^{2k} + (2kaR_-/t)P^{2k}\mathbf{1}\{|e_t| > \delta\}$. By (i) with k_0 and Markov's inequality, $\mathbb{P}(|e_t| > \delta) \le M_{k_0}\delta^{-2k_0}t^{-\rho k_0} \le Ct^{-4}$, so the tail term contributes at most Ct^{-5}. Induction on $k = 1, 2, 3$:

with $\mathbb{E}e_t^{2k-2} \le Ct^{-(k-1)}$, Lyapunov gives $\mathbb{E}|e_t|^{2k-j} \le Ct^{-(2k-j)/2}$, and $j + (2k-j)/2 = k + j/2 \ge k+1$. So $v_{t+1} \le (1 - 2kaR_-/t + C/t^2)v_t + Dt^{-(k+1)}$, using $t^{-5} \le t^{-(k+1)}$ for $k \le 3$. Since $2kaR_- > k$, Lemma B.11 gives $v_t \le M'_k t^{-k}$. □

Theorem B.21 (Sharp constant for every stable gain)**.** *The conclusion of Theorem B.17,* $\mathrm{Regret}_T = \sigma^2(p^*)\,\frac{a^2R'^2}{2aR'-1}\ln T\,(1+o(1))$*, holds with the hypotheses "*$2am > 1$*" and "moments up to order 6" replaced by "*$2aR' > 1$*" and "conditional moments up to order* $2k_0$ *with* $\rho k_0 \ge 4$ *for some* $\rho \in (0, \min(2am, 1))$*" (for Gaussian noise, all orders are available). In particular the tuned gain* $a = 1/R'$ *attains* $\sigma^2(p^*)\ln T(1+o(1))$ *in every world satisfying (M1)–(M3), (B1), (B2), whatever the global modulus* $m > 0$*.*

Proof. The proof of Theorem B.17 uses $2am > 1$ in exactly two places, both in Step 0: to obtain $c := 2aR' > 1$, and to invoke Lemma B.12 for $\mathbb{E}e_t^{2k} \le M'_k t^{-k}$, $k = 1, 2, 3$. The first is now assumed. The second is Lemma B.20(iii). The almost-sure convergence used for the variance term (c) is Lemma B.20(ii) in place of Corollary B.15. Steps 1–6 use only these moment bounds, (M1)–(M3), (B1) and (B2), and the projection bound of Step 2 uses the moment bounds of order up to 6, which (iii) provides. The rest of the proof is unchanged. □

Proposition B.22 (Closed forms in the quadratic world; price of not tuning)**.** *With* $s = \sqrt{1+4\gamma\mu}$*:* $p^* = 2\mu/(s+1)$, $\gamma p^* = (s-1)/2$, $R' = s$, $h := R'p^* = 2\mu s/(s+1)$, $m = (s+1)/2$. *(i)* $h/\mu = 2s/(s+1)$ *increases strictly on* $(1,\infty)$ *with range* $(1,2)$*, so* $R'^2 p^{*2} = h^2 \in (\mu^2, 4\mu^2)$*. (ii) Using* $a = 1$ *instead of the tuned gain costs the factor* $g(1)/g(1/R') = s^2/(2s-1) = s/2 + s/(2(2s-1)) \ge s/2 \ge \sqrt{\gamma\mu}$*, which is unbounded in* $\gamma\mu$*.*

Proof. The root of $\gamma p^2 + p - \mu = 0$ is $(s-1)/(2\gamma) = 2\mu(s-1)/(s^2-1) = 2\mu/(s+1)$, using $2\gamma = (s^2-1)/(2\mu)$. Then $\gamma p^* = (s-1)/2$, $R' = 1 + 2\gamma p^* = s$, $h = 2\mu s/(s+1)$, and $2s/(s+1) = 2 - 2/(s+1)$ increases from 1 to 2. $m = 1 + \gamma p^* = (s+1)/2$ as in Appendix A. For (ii), $s^2/(2s-1) - s/2 = s/(2(2s-1)) > 0$ and $s \ge 2\sqrt{\gamma\mu}$. □

Remark B.23 (Lai and Robbins: credit for this section). Lai & Robbins (1979) study $x_{n+1} = x_n - y_n/(nb_n)$ with $y_n = M(x_n) + \varepsilon_n$, i.i.d. errors with mean 0 and variance $\sigma^2 < \infty$, x_1 independent of the errors, b_n $\mathcal{F}_{n-1}$-measurable with $b_n \to b > 0$ a.s., and a regression function on all of $\mathbb{R}$ satisfying

$$\text{(1.2) } M(\theta) = 0 \text{ uniquely and } M'(\theta) = \beta > 0; \quad \text{(1.3) } \inf_{\delta < |x-\theta| < 1/\delta} M(x)(x-\theta) > 0 \text{ for all } \delta \in (0,1); \quad \text{(1.4) } |M(x)| \le A|x| + D.$$

For $b < 2\beta$ (their Theorems 2 and 4(i)) they prove, almost surely, $\sum_{n \le N}(x_n - \theta)^2 \sim (\sigma^2/\beta^2) f(b/\beta)\log N$ with $f(t) = 1/(t(2-t))$; for random b_n their Theorem 3(ii) adds $\mathbb{E}|\varepsilon|\log|\varepsilon| < \infty$ or symmetric errors. With $M = r$, $\theta = p^*$, $\beta = R'$ and $b = 1/a$, $f(1/c) = c^2/(2c-1)$ with $c = aR'$ is exactly the constant of Theorem B.17. Lai & Robbins (1981, Theorems 2 and 3) show that a truncated least-squares slope estimate b_n converges to β a.s., under (1.2)–(1.4) and $M \in C^1$ near θ, so the adaptive scheme attains $(\sigma^2/\beta^2)\log N$ a.s.; Theorem 2 truncates with known bounds $b < \beta < B$, and Theorem 3 uses a data-driven truncation needing no bounds. Their only in-expectation statement is (1.10) of the 1979 paper, for the linear model with β known. The Bayes lower bound $\sigma^2 \log N + O(1)$ with normal errors is due to Lai & Robbins (1982), via dynamic programming in the linear multiperiod control model (as stated in Lai, 2003).

So Theorems B.13, B.17 and B.29 are self-contained versions of classical results and are credited to Lai and Robbins. Relative to the classical statements they cover projection onto a bounded set, martingale-difference noise, and (in Theorem B.17) a sharp constant *in expectation* for nonlinear residuals. The almost-sure adaptive result, Proposition B.24, is taken from the classical papers.

Proposition B.24 (Almost-sure theory via Lai and Robbins, for the clipped-publication variant)**.** *Assume Assumption A.1 with i.i.d. noise of mean* 0 *and variance* $\sigma^2 < \infty$*, (M1), (M2) and (M3) on* $\mathcal{P}$*,* p^* *interior to* $\mathcal{P}$*, and* r *differentiable at* p^* *with* $R' = r'(p^*) > 0$*. Consider the* clipped-publication *scheme: keep an internal state* $x_n \in \mathbb{R}$*, publish* $\hat{p}_n := \Pi_{\mathcal{P}}(x_n)$*, observe* $z_n = \hat{p}_n - y_n$*, and update* $x_{n+1} = x_n - z_n/(nb_n)$*.*

(i) With a fixed gain $b_n \equiv 1/a$ and any $c := aR' > \frac{1}{2}$ (no global modulus condition), Regret *realised along the path satisfies $\sum_{n\le N} r(\hat{p}_n)^2 \sim \sigma^2 \frac{c^2}{2c-1} \log N$ almost surely.*

(ii) If moreover $r \in C^1$ near p^* and the noise is symmetric or has $\mathbb{E}|\eta| \log |\eta| < \infty$, and b_n is the truncated least-squares slope of Lai & Robbins (1981), with known bounds $0 < \underline{b} < R' < \bar{b}$ (their Theorem 2) or with their data-driven truncation (their Theorem 3), then $\sum_{n\le N} r(\hat{p}_n)^2 \sim \sigma^2 \log N$ almost surely: the adaptive gain attains the optimal constant without knowing R'.

Proof. **Reduction.** The scheme is exactly the Lai–Robbins recursion with errors $\varepsilon_n := -\eta_n$ and regression function $M(x) := r(\Pi_{\mathcal{P}}(x))$ on $\mathbb{R}$.

(1.2) $M(p^*) = r(p^*) = 0$; the root is unique because $M(x) \neq 0$ for $x \neq p^*$ (next step); and since p^* is interior, $M = r$ near p^*, so $M'(p^*) = R' > 0$.

(1.3) Let $\delta \in (0, 1)$ and $\delta < |x - p^*| < 1/\delta$. If $x \in \mathcal{P}$, (M2) gives $M(x)(x - p^*) \ge m\delta^2$. If $x > P$, then $M(x) = r(P) \ge m(P - p^*) > 0$ and $x - p^* > \delta$, so the product is at least $m(P - p^*)\delta$. If $x < 0$, then $M(x) = r(0) \le -mp^* < 0$ and $x - p^* < -\delta$, so the product is at least $mp^*\delta$.

(1.4) $|M| \le \max_{\mathcal{P}} |r| < \infty$.

Conclusion. The cited theorems give $x_n \to p^*$ a.s. and $\sum_{n\le N}(x_n - p^*)^2 \sim (\sigma^2/R'^2) f(b/R') \log N$ a.s., with $b = 1/a$ in (i) and $b = R'$ in (ii) (so $f = 1$). Since p^* is interior, $\hat{p}_n = x_n$ for all n beyond an a.s. finite time, and $r(\hat{p}_n) = R'(x_n - p^*)(1 + o(1))$ by differentiability; so $\sum r(\hat{p}_n)^2 \sim R'^2 \sum (x_n - p^*)^2$ a.s. In (ii), $M \in C^1$ near p^* because $M = r$ there. □

Proposition B.25 (Almost-sure constant for the state-projected RSA)**.** *Under the assumptions of Proposition B.24 with Gaussian noise, the state-projected RSA(a) of Algorithm A.6 with any gain $c := aR' > \frac{1}{2}$ satisfies $\sum_{n\le N} r(\hat{p}_n)^2 \sim \sigma^2 \frac{c^2}{2c-1} \log N$ almost surely.*

Proof. **Convergence.** By Lemma B.20(ii), $\hat{p}_n \to p^*$ a.s. Since $\mathbb{E}|\eta| < \infty$, $\eta_n/n \to 0$ a.s. by Borel–Cantelli, so the step $(a/n)(r(\hat{p}_n) - \eta_n) \to 0$ a.s. As p^* is interior, the projection is active only finitely often, almost surely.

Restart. Fix $\delta > 0$ with $[p^* - \delta, p^* + \delta] \subset \operatorname{int} \mathcal{P}$. For $k \ge 1$ let $\nu_k := \inf\{n \ge k : |\hat{p}_n - p^*| \le \delta\}$, a stopping time, finite a.s. Define the clipped-publication process started at ν_k: $X_1 := \hat{p}_{\nu_k}$ and $X_{j+1} := X_j - \frac{1}{j\, b'_j}\big(r(\Pi_{\mathcal{P}}(X_j)) - \eta_{\nu_k+j-1}\big)$ with $b'_j := \frac{\nu_k+j-1}{a\, j}$. The errors $\{\eta_{\nu_k+j-1}\}_{j\ge1}$ are i.i.d. and independent of $\mathcal{F}_{\nu_k-1}$ (strong Markov property of an i.i.d. sequence), X_1 is $\mathcal{F}_{\nu_k-1}$-measurable, and b'_j is measurable at stage $j-1$ with $b'_j \to 1/a$ a.s. The regression function $M(x) = r(\Pi_{\mathcal{P}}(x))$ satisfies (1.2)–(1.4) (proof of Proposition B.24), and $1/a < 2R'$. So Lai & Robbins (1979, Theorem 4(i)) gives $\sum_{j\le J}(X_j - p^*)^2 \sim (\sigma^2/R'^2)\frac{c^2}{2c-1} \log J$ a.s.

Coincidence. Let A_k be the event that the projection is never active after time k. On A_k, induction gives $\hat{p}_{\nu_k+j-1} = X_j$ for all j: the two recursions agree whenever the unprojected update stays in $\mathcal{P}$, where $\Pi_{\mathcal{P}}$ is the identity. So the asymptotic holds on each A_k, and $\mathbb{P}(\bigcup_k A_k) = 1$ by the first step. Finally $r(\hat{p}_n) = R'(\hat{p}_n - p^*)(1 + o(1))$ and $\log(J + \nu_k) \sim \log J$. □

Remark B.26 (What is proved for the adaptive gain). Almost surely: every stable fixed gain is covered for both the clipped-publication variant (Proposition B.24(i)) and the state-projected RSA (Proposition B.25); in expectation, every stable fixed gain is covered for the state-projected RSA (Theorem B.21). The adaptive gain is proved almost surely for the clipped-publication variant (Proposition B.24(ii)), which is therefore the algorithm this paper recommends when R' is unknown. Two statements remain open: the adaptive gain for the state-projected RSA (the Lai–Robbins consistency proof depends on their recursion from time 1 and on the form of their estimator, and does not transfer by restarting), and the adaptive gain *in expectation*. For the latter the obstacle is genuine: the least-squares slope converges only at rate $(\log n)^{-1/2}$, so paths on which the gain stays poor for a long time are not obviously rare enough for the expectation. Experiment E4 is consistent with the almost-sure statement at $\gamma\mu \in \{2, 10\}$.

B.4 TUNING THE GAIN FROM LEVELS

Remark B.27 (What the levels say about the slope). Write $\Phi(p) = \lambda^*(p)\, p$ with $\lambda^* \geq 0$ bounded and differentiable at p^*, so that $0 < p^* \leq \mu$. Differentiating and using $p^* + \Phi(p^*) = \mu$ gives the identity

$$p^* R' = p^* + \Phi(p^*) + {p^*}^2 \lambda^{*\prime}(p^*) = \mu + {p^*}^2 \lambda^{*\prime}(p^*),$$

and when $\lambda^*(p^*) > 0$ the last term equals $\varepsilon(p^*)(\mu - p^*)$. Three consequences follow.

(i) *The quadratic world.* There $\lambda^*(p) = \gamma p$ and $\gamma {p^*}^2 = \mu - p^*$, so the levels determine the slope: $R' = 2\mu/p^* - 1$. The gain $a = p^*/(2\mu - p^*) = 1/R'$ therefore attains the constant σ^2 of Theorem B.21.

(ii) *The level gain in general.* The gain $a = p^*/\mu$ has $aR' = 1 + \varepsilon(p^*)(1 - p^*/\mu)$. If $0 \leq \varepsilon(p^*) \leq E$, then $1 \leq aR' \leq 1 + E$, and since $g(x) = x^2/(2x - 1)$ is nondecreasing on $[1, \infty)$, its constant is at most $g(1 + E)\, \sigma^2$ (Theorem B.21); in the quadratic world ($E = 1$) it lies in $(\sigma^2, \frac{4}{3}\sigma^2)$. More generally this gain is stable whenever ${p^*}^2 \lambda^{*\prime}(p^*) > -\mu/2$, and since g is minimised at 1, a slightly negative elasticity still gives a constant close to σ^2.

(iii) *What the data identify.* Without knowledge of Φ, μ is identified only by plays at $p = 0$, where $\Phi(0) = 0$ and $y = \mu + \eta$. The gains in (i) and (ii) are oracles: they use p^*, which is the unknown being learned.

Appendix H.1.10 reports simulations of the level gain and of a plug-in version.

B.5 THE LOWER BOUND

Lemma B.28 (Van Trees inequality). *Let π be a density on $[\mu_1, \mu_2]$, absolutely continuous, positive on (μ_1, μ_2), vanishing at both endpoints, with finite Fisher information $I(\pi) := \int (\pi')^2/\pi$. Let $X = (X_1, \ldots, X_n)$ be i.i.d. $N(\mu, \sigma^2)$ given μ and let $\hat{\mu}(X)$ satisfy $\mathbb{E}\hat{\mu}^2 < \infty$. Then*

$$\int \mathbb{E}_\mu\big[(\hat{\mu} - \mu)^2\big]\, \pi(\mu)\, d\mu \;\geq\; \frac{1}{n/\sigma^2 + I(\pi)}.$$

Proof. Let $f(x \mid \mu)$ be the joint density and $q := f\pi$. For fixed x, integrating by parts in μ, $\int (\hat{\mu}(x) - \mu)\, \partial_\mu q\, d\mu = [(\hat{\mu} - \mu)q]_{\mu_1}^{\mu_2} + \int q\, d\mu$, and the boundary term vanishes because $\pi(\mu_1) = \pi(\mu_2) = 0$. Integrating over x: $\iint (\hat{\mu} - \mu)\partial_\mu q = 1$. By Cauchy–Schwarz, writing $(\hat{\mu} - \mu)\partial_\mu q = [(\hat{\mu} - \mu)\sqrt{q}][\partial_\mu q/\sqrt{q}]$, $1 \leq \iint (\hat{\mu} - \mu)^2 q \cdot \iint (\partial_\mu q)^2/q$. On the support, $\partial_\mu q/q = \partial_\mu \log f + \partial_\mu \log \pi$, so

$$\iint q(\partial_\mu \log f + \partial_\mu \log \pi)^2 = \int \pi I_X + 2 \int \pi(\partial_\mu \log \pi)\Big(\int \partial_\mu f\, dx\Big) d\mu + I(\pi),$$

and the middle term vanishes because $\int \partial_\mu f\, dx = \partial_\mu \int f\, dx = 0$ (differentiation under the integral is valid for the Gaussian family). For n i.i.d. $N(\mu, \sigma^2)$ observations $I_X(\mu) = n/\sigma^2$. □

Theorem B.29 (Matching lower bound). *Fix a known Φ on $[0, \infty)$, let $\psi(p) := p + \Phi(p)$, assume (A2-G), and let $I = [\mu_1, \mu_2] \subset (0, \infty)$ have width w. For every (possibly randomised) algorithm with $\hat{p}_t \geq 0$,*

$$\sup_{\mu \in I} \mathbb{E}_\mu \mathrm{Regret}_T \;\geq\; \sigma^2 \ln T - \sigma^2 \max\big(0, \ln(4\pi^2\sigma^2/w^2)\big).$$

This holds a fortiori when Φ is unknown.

Proof. **Step 1 (i.i.d. reduction).** Define $x_s := y_s + \Phi(\hat{p}_s) = \mu + \eta_s$. Conditional on $U = u$, $\hat{p}_1$ is a constant and $\hat{p}_s$ is a function of $(y_1, \ldots, y_{s-1})$, so $(y_1, \ldots, y_{t-1})$ and $(x_1, \ldots, x_{t-1})$ determine each other recursively: $x_1 = y_1 + \Phi(\hat{p}_1)$; given $x_{1:s-1}$ one recovers $y_{1:s-1}$, then $\hat{p}_s$, then x_s from y_s, and conversely. Given (μ, u) the x_s are i.i.d. $N(\mu, \sigma^2)$.

Step 2 (excess loss is estimation error). $\hat{\mu}_t := \psi(\hat{p}_t)$ is a function of $x_{1:t-1}$ given u, and $r(\hat{p}_t)^2 = (\psi(\hat{p}_t) - \mu)^2 = (\hat{\mu}_t - \mu)^2$ exactly.

Step 3. Use the raised-cosine prior $\pi(\mu) = (2/w)\cos^2(\pi(\mu - \mu_c)/w)$, $\mu_c = (\mu_1 + \mu_2)/2$, which satisfies Lemma B.28 with $I(\pi) = 4\pi^2/w^2$ (direct integration; also checked by quadrature, T5).

With $b := \sigma^2 I(\pi)$, Lemma B.28 conditionally on u, then averaged over u, gives $\int \mathbb{E}_\mu(\hat{\mu}_t - \mu)^2 \pi \geq \sigma^2/(t-1+b)$ (trivial if $\mathbb{E}\hat{\mu}_t^2 = \infty$). Summing, $\int \mathbb{E}_\mu \mathrm{Regret}_T \, \pi \geq \sigma^2 \sum_{k=0}^{T-1} 1/(k+b)$.

Step 4. Since $1/(k+b) \geq \int_k^{k+1} dx/(x+b)$, the sum is at least $\ln((T+b)/b)$, which is $\geq \ln T - \ln b$ if $b \geq 1$ and $\geq \ln(T+b) \geq \ln T$ if $b < 1$. The supremum over I dominates the π-average. □

Remark B.30. In the quadratic world with $a = 1/s$, Theorem B.17 gives $\mathrm{Regret}_T = \sigma^2 \ln T(1+o(1))$ at each fixed μ, and the next proposition makes the $o(1)$ uniform over I. This is *not* a matching upper bound for the learners of Theorem B.29: the tuned gain $a(\mu) = 1/s(\mu)$ uses the slope at the answer, and in the quadratic world the slope reveals the answer ($p^* = 2\mu/(s+1)$). Without that information, the constant σ^2 is attained almost surely by the adaptive clipped-publication learner (Proposition B.24(ii)); attaining it in expectation is open. Experiment E2 (Bayes version, same prior) gives slopes $0.98\sigma^2$ (tuned RSA), $1.01\sigma^2$ (oracle knowing γ) and $1.00\sigma^2$ (exact bound).

Proposition B.31 (Uniform sharp constant for the oracle-tuned gain in the quadratic world)**.** *In the quadratic world with γ fixed, constant σ^2, μ ranging over a compact $I = [\mu_1, \mu_2] \subset (0, \infty)$ with $p^*(\mu)$ interior to $\mathcal{P}$ for all $\mu \in I$, and the tuned gain $a(\mu) = 1/s(\mu)$,*

$$\sup_{\mu \in I} \left| \frac{\mathrm{Regret}_T(\mu)}{\sigma^2 \ln T} - 1 \right| \to 0.$$

Proof. Every constant in the proof of Theorem B.17 depends continuously on $(\mu, \gamma, a, \sigma, P)$ through $m = (s+1)/2$, $L = 1+\gamma(P+p^*)$, $R' = s$ and $D_0 = \min(p^*, P-p^*)$, all continuous and bounded away from 0 and ∞ on I. So M_1', M_2', M_3' and the projection constant C_P are bounded uniformly. In the quadratic world the remainder is exactly $\rho(e) = \gamma e^2$, so $\omega(\delta) = \gamma\delta$ uniformly, and with constant σ^2 the variance term (c) of Step 3 vanishes. Hence $t^2|\varepsilon_t| \leq \bar{\beta}_t$ for a sequence $\bar{\beta}_t \to 0$ not depending on μ. With the tuned gain $c = 2aR' = 2$, so $b = c - 1 = 1$ in Lemma B.16 for every μ. The proof of that lemma bounds $|tu_t - d/(c-1)|$ by a quantity depending only on b, $\sup_t |s_t|$ and the sequence β_t, all uniform in μ; so $tu_t \to d/(c-1)$ uniformly. Steps 5 and 6 (the comparison of $\mathbb{E}r^2$ with $R'^2 u_t$, and Stolz–Cesàro) preserve uniform convergence, since their error terms are bounded by the same uniform quantities. □

C When the equilibrium disappears: proofs

Assumption C.1 (The jump setting)**.** **(G1)** $x_0 \in (0, P)$, and write $D_0 := \min(x_0, P - x_0)$; **(G2)** Φ is bounded and nondecreasing on $\mathcal{P}$; **(G3)** Φ is continuous on $\mathcal{P} \setminus \{x_0\}$, with one-sided limits $\Phi(x_0-)$ and $\Phi(x_0+)$; **(G4)** $\Phi(x_0) \in \{\Phi(x_0-), \Phi(x_0+)\}$ (the segment convention); **(G5)** the gap condition $r^- < 0 < r^+$, where $r^- := x_0 + \Phi(x_0-) - \mu$ and $r^+ := x_0 + \Phi(x_0+) - \mu$. Where stated: **(G6)** Φ is Lipschitz with constant L_Φ on $[0, x_0)$ and on $(x_0, P]$. Write $B := \sup_{\mathcal{P}} |r| < \infty$, $u := -r^- > 0$ and $v := r^+ > 0$. The discrete-type model has $\Phi(p) = G_j p^2$ on segment j, with $G_{j+1} > G_j \geq 0$.

Lemma C.2 (Fact F)**.** *Under (G2), $r(q) - r(p) \geq q - p$ for all $p < q$ in $\mathcal{P}$. Consequently $r^- - r(p) \geq x_0 - p$ for $p < x_0$ and $r(p) - r^+ \geq p - x_0$ for $p > x_0$.*

Proof. $r(q) - r(p) = (q-p) + \Phi(q) - \Phi(p) \geq q - p$. Let $q \uparrow x_0$ (resp. $p \downarrow x_0$) and use (G3). □

Proposition C.3 (The gap is a jump over zero)**.** *(i) Under (G2), (G3), (G4) and (G5), r has no root in $\mathcal{P}$. (ii) If $r(0) < 0 < r(P)$ and either $r^- > 0$ or $r^+ < 0$, then r has a root in $\mathcal{P}$. (iii) In the discrete-type model with threshold $x_0 = \hat{p}^*_{k+1}$, $r^- < 0 < r^+$ is equivalent to the gap-window condition* (A.2).

Proof. **(i)** By Lemma C.2, $r(p) < r^- < 0$ for $p < x_0$ and $r(p) > r^+ > 0$ for $p > x_0$; and $r(x_0) \in \{r^-, r^+\}$ by (G4).

(ii) If $r^- > 0$, r is continuous on $[0, x_0)$ with $r(0) < 0$ and $\lim_{p \uparrow x_0} r(p) = r^- > 0$, so it has a zero there by the intermediate value theorem; the case $r^+ < 0$ is symmetric.

(iii) Dividing by $x_0 > 0$: $r^- < 0 \iff G_k x_0 < \mu/x_0 - 1$ and $r^+ > 0 \iff \mu/x_0 - 1 < G_{k+1} x_0$. □

Theorem C.4 (Abstract gap mechanism). *Let $J = [j_0, j_1]$ be compact with $x_0 \in (j_0, j_1)$ and $f : J \to J$ with (H1) f continuous on $J \setminus \{x_0\}$; (H2) f nonincreasing; (H3) $f^- := \lim_{x \uparrow x_0} f(x) > x_0 > \lim_{x \downarrow x_0} f(x) =: f^+$; (H4) $f(x_0) \in \{f^-, f^+\}$. Then (i) f has no fixed point; (ii) for every $x_1 \in J$ the iterates $x_{t+1} = f(x_t)$ converge to a 2-cycle $\{x_L, x_H\}$ with $x_L \le f^+ < x_0 < f^- \le x_H$, $x_H = f(x_L)$, $x_L = f(x_H)$. Without (H4) the statement is false: setting $f(x_0) := x_0$ keeps (H1)–(H3) and creates a fixed point. In the forecasting model (H4) is the segment convention (G4): at the threshold the world's tie-break selects one of its two extreme actions, which is what a world indifferent between them and choosing deterministically does.*

Proof. **(i)** For $x < x_0$ and any $x < y < x_0$, (H2) gives $f(x) \ge f(y)$; letting $y \uparrow x_0$, $f(x) \ge f^- > x_0 > x$. Symmetrically, for $x > x_0$, $f(x) \le f^+ < x_0 < x$. At x_0, (H4) and (H3) give $f(x_0) \ne x_0$.

(ii)

Alternation. $x > x_0 \Rightarrow f(x) \le f^+$, $x < x_0 \Rightarrow f(x) \ge f^-$, and $f(x_0) \in \{f^-, f^+\}$. So every iterate with $t \ge 2$ lies in $[j_0, f^+] \cup [f^-, j_1]$ and consecutive iterates lie in different pieces.

Double map. $f^2 = f \circ f$ is nondecreasing.

Monotone subsequences. Let $\tau \in \{2, 3\}$ be the first index with $x_\tau \le f^+$. The subsequence $x_{\tau+2i}$ satisfies $x_{\tau+2i+2} = f^2(x_{\tau+2i})$; if $x_{\tau+2} \ge x_\tau$ then by induction and monotonicity of f^2 it is nondecreasing, otherwise nonincreasing. It is bounded in $[j_0, f^+]$ and converges to some x_L. The interleaved subsequence is also an f^2-orbit and converges to some $x_H \in [f^-, j_1]$. J is closed.

2-cycle. $x_L < x_0 < x_H$, so by (H1) f is continuous at both limits; passing to the limit in $x_{\tau+2i+1} = f(x_{\tau+2i})$ gives $x_H = f(x_L)$, and similarly $x_L = f(x_H)$. □

Theorem C.5 (Closed-form orbit). *In the discrete-type model let $f(p) := \mu/(1 + \lambda^*(p))$ with $\lambda^*(p) = G_j p$ on segment j, let $x_0 := \hat{p}^*_{k+1}$ satisfy (A.2), and let $J := [\mu/(1 + \lambda^*(\mu)), \mu]$. Assume **(O1)**: J contains no activation threshold other than x_0, and (H4). Then the myopic iterates converge to $\{\hat{p}_L, \hat{p}_H\}$, the unique solution with $\hat{p}_L < x_0 < \hat{p}_H$ of*

$$\hat{p}_L(1 + G_{k+1}\hat{p}_H) = \mu, \qquad \hat{p}_H(1 + G_k\hat{p}_L) = \mu. \tag{C.1}$$

With $A := G_k + G_{k+1}$ and $B := G_{k+1} - G_k$: if $G_k > 0$, then $q := \hat{p}_L\hat{p}_H$ is the smaller root of $G_kG_{k+1}q^2 - (\mu A + 1)q + \mu^2 = 0$, $S := \hat{p}_L + \hat{p}_H = 2\mu - AP$, and $\hat{p}_{L,H} = (S \mp \sqrt{S^2 - 4q})/2$; if $G_k = 0$, then $\hat{p}_H = \mu$ and $\hat{p}_L = \mu/(1 + G_{k+1}\mu)$.

Proof. **Invariance.** For $p \in J \subset (0, \mu]$, $0 \le \lambda^*(p) \le \lambda^*(\mu)$, so $f(p) \in J$. *Apply Theorem C.4.* By (O1) f is continuous on $J \setminus \{x_0\}$; f is nonincreasing; (A.2) is equivalent to (H3) because $f^- = \mu/(1 + G_kx_0) > x_0 \iff r^- < 0$ and similarly for f^+. So the iterates converge to a 2-cycle with $x_L < x_0 < x_H$, and by (O1) x_L is on segment k and x_H on segment $k + 1$, which gives (C.1). *One quadratic* (checked symbolically, T2). Subtracting the equations: $\hat{p}_H - \hat{p}_L = BP$. Adding: $S = 2\mu - AP$. Substituting into $(\hat{p}_H - \hat{p}_L)^2 = S^2 - 4q$ gives $(A^2 - B^2)q^2 - 4(\mu A + 1)q + 4\mu^2 = 0$, and $A^2 - B^2 = 4G_kG_{k+1}$. The discriminant is $\mu^2B^2 + 2\mu A + 1 > 0$. *Root selection* ($G_k > 0$). By AM–GM $A^2 \ge 4G_kG_{k+1}$, and the larger root satisfies $q_+ \ge (\mu A + 1)/(2G_kG_{k+1})$, so $S_+ = 2\mu - AP_+ \le 2\mu - \mu A^2/(2G_kG_{k+1}) - A/(2G_kG_{k+1}) \le -A/(2G_kG_{k+1}) < 0$, impossible for a sum of positive numbers. The cycle exists, so its q is the smaller root, and $\{\hat{p}_L, \hat{p}_H\}$ is the root pair of $x^2 - Sx + q$.

$G_k = 0$. The second equation gives $\hat{p}_H = \mu$ and the first gives $\hat{p}_L$. □

Proposition C.6 (Residuals on the orbit). *With $\Delta_p := \hat{p}_H - \hat{p}_L$: $r(\hat{p}_L) = -(1 + G_k\hat{p}_L)\Delta_p$ and $r(\hat{p}_H) = (1 + G_{k+1}\hat{p}_H)\Delta_p$. The long-run average excess of the myopic dynamics is $\frac{1}{2}\Delta_p^2[(1 + G_k\hat{p}_L)^2 + (1 + G_{k+1}\hat{p}_H)^2]$ (4.32 at $(\mu, G_k, G_{k+1}, x_0) = (1.5, 0.5, 2, 0.8)$).*

Proof. $r(\hat{p}_L) = \hat{p}_L(1 + G_k\hat{p}_L) - \mu$ and $\mu = \hat{p}_H(1 + G_k\hat{p}_L)$ by (C.1); similarly $r(\hat{p}_H) = \hat{p}_H(1 + G_{k+1}\hat{p}_H) - \mu$ with $\mu = \hat{p}_L(1 + G_{k+1}\hat{p}_H)$. The iterates alternate and converge to the two orbit points, so the Cesàro average of $r(x_t)^2$ converges to the mean of the two squares. □

Proposition C.7 (Every learner pays linear regret)**.** *Under (G2)–(G5),* $\inf_{p\in K} r(p)^2 = \min(r^{-2}, r^{+2})$*, and every algorithm has*

$$\mathrm{Regret}_T \geq T \min(r^{-2}, r^{+2}).$$

Proof. By the proof of Proposition C.3(i), $r(p)^2 > r^{-2}$ for $p < x_0$, $r(p)^2 > r^{+2}$ for $p > x_0$, and $r(x_0)^2 \in \{r^{-2}, r^{+2}\}$. The infimum is attained or approached at x_0 from the relevant side by one-sided continuity. Sum over t. □

Theorem C.8 (RSA in the gap)**.** *Under Assumptions A.1 and A.2 with conditional variance at most* $\bar{\sigma}^2$*, and (G1)–(G5). For any gain* $a > 0$*, RSA*(a) *satisfies: (i)* $\hat{p}_t \to x_0$ *a.s.; (ii)* $\frac{1}{T}\sum_{t\leq T} I_t \to \omega := u/(u+v)$ *a.s., where* $I_t = 1$ *iff* $\hat{p}_t$ *lies on the right piece; (iii)* $\mathrm{Regret}_T/T \to |r^- r^+| = uv$*. Since* $uv = \sqrt{u^2 v^2}$*, the limit lies between* $\min(u^2, v^2)$ *and* $\max(u^2, v^2)$*.*

Cited theorems. Robbins–Siegmund (Robbins & Siegmund, 1971): if $V_t, \beta_t, \xi_t, \zeta_t \geq 0$ are adapted with $\mathbb{E}[V_{t+1} \mid \mathcal{F}_t] \leq (1+\beta_t)V_t + \xi_t - \zeta_t$, $\sum \beta_t < \infty$ and $\sum \xi_t < \infty$ a.s., then V_t converges a.s. to a finite limit and $\sum \zeta_t < \infty$ a.s. L^2 **martingale convergence:** an L^2-bounded martingale converges a.s. **Kronecker:** if $\sum x_t/t$ converges then $\frac{1}{T}\sum_{t\leq T} x_t \to 0$.

Proof. Let $e_t := \hat{p}_t - x_0$.

Step 1 (drift). Since $x_0 \in \mathcal{P}$ and projection is nonexpansive, $\mathbb{E}[e_{t+1}^2 \mid \mathcal{F}_{t-1}] \leq e_t^2 - (2a/t)e_t r(\hat{p}_t) + a^2(B^2 + \bar{\sigma}^2)/t^2$.

Step 2. With $c_0 := \min(u, v) > 0$: if $\hat{p}_t < x_0$ then $r(\hat{p}_t) \leq r^- < 0$ and $e_t < 0$, so $e_t r \geq |e_t| u$; if $\hat{p}_t > x_0$ then $e_t r \geq e_t v$. So $e_t r(\hat{p}_t) \geq c_0 |e_t|$.

Step 3. Robbins–Siegmund with $V_t = e_t^2$, $\beta_t = 0$, $\xi_t = a^2(B^2 + \bar{\sigma}^2)/t^2$, $\zeta_t = (2ac_0/t)|e_t|$: e_t^2 converges a.s. and $\sum |e_t|/t < \infty$; since $\sum 1/t = \infty$, $\liminf |e_t| = 0$, so the limit is 0. This is (i).

Step 4 (noise). $N_T := \sum_{t\leq T} \eta_t/t$ is a martingale with $\sum \mathbb{E}\eta_t^2/t^2 < \infty$, so it converges a.s.; hence $\eta_t/t \to 0$ and, by Kronecker, $\frac{1}{T}\sum_{t\leq T} \eta_t \to 0$ a.s.

Step 5 (projection switches off). $y_t^u = \hat{p}_t - (a/t)(r(\hat{p}_t) - \eta_t) \to x_0 \in (0, P)$ a.s., so there is an a.s. finite τ with $\hat{p}_{t+1} = y_t^u$ for $t \geq \tau$.

Step 6 (telescoping). For $t \geq \tau$, $te_{t+1} = te_t - ar(\hat{p}_t) + a\eta_t$; with $te_{t+1} = (t+1)e_{t+1} - e_{t+1}$, $(t+1)e_{t+1} - te_t = e_{t+1} - ar(\hat{p}_t) + a\eta_t$. Summing from τ to T and dividing by T, using (i), the Cesàro convergence of e_{t+1}, Step 4, and $O(1/T)$ for $t < \tau$: $\frac{1}{T}\sum_{t\leq T} r(\hat{p}_t) \to 0$ a.s.

Step 7. By (G3) and (i), $r(\hat{p}_t) - [r^- + (r^+ - r^-)I_t] \to 0$ a.s.; with Step 6, $\frac{1}{T}\sum [r^- + (r^+ - r^-)I_t] \to 0$, so $\frac{1}{T}\sum I_t \to u/(u+v)$.

Step 8. Since $I_t \in \{0, 1\}$, $[r^- + (r^+ - r^-)I_t]^2 = r^{-2}(1 - I_t) + r^{+2} I_t$, and $|x^2 - y^2| \leq 2B|x - y|$; so $\frac{1}{T}\sum r(\hat{p}_t)^2 \to u^2 \frac{v}{u+v} + v^2 \frac{u}{u+v} = uv$ a.s.

Step 9. $\frac{1}{T}\sum r^2 \leq B^2$; dominated convergence gives $\mathrm{Regret}_T/T \to uv$. □

Proposition C.9 (The myopic orbit is never better than RSA)**.** *In every gap world satisfying the assumptions of Proposition C.6, the myopic orbit's long-run average excess is at least* $|r^- r^+|$*, the long-run average excess of RSA (Theorem C.8). Equality requires* $r(\hat{p}_L) = r^-$*,* $r(\hat{p}_H) = r^+$ *and* $|r^-| = r^+$*.*

Proof. By Proposition C.6 the average excess is $\frac{1}{2}\Delta_p^2[(1 + G_k \hat{p}_L)^2 + (1 + G_{k+1}\hat{p}_H)^2] = \frac{1}{2}\big(r(\hat{p}_L)^2 + r(\hat{p}_H)^2\big)$. The orbit points satisfy $\hat{p}_L < x_0 < \hat{p}_H$ and r is nondecreasing, so $r(\hat{p}_L) \leq r^- < 0$ and $r(\hat{p}_H) \geq r^+ > 0$. Hence the excess is at least $\frac{1}{2}(r^{-2} + r^{+2}) \geq |r^-| r^+$, the last step by the inequality of arithmetic and geometric means, with equality in both steps exactly under the stated conditions. □

C.1 UNBIASEDNESS VERSUS ACCURACY

In the gap, a learner can aim for zero average bias or for small squared error, but not both. The next theorem makes this exact. It is stated for the learner's own sequence of predictions, with averages over time, so no mixed strategies or convexified dynamics are needed.

Theorem C.10 (Bias–accuracy frontier). *Assume (G1)–(G5) and write* $\bar{r}_T := \frac{1}{T}\sum_{t\le T} r(\hat{p}_t)$ *and* $\overline{r^2}_T := \frac{1}{T}\sum_{t\le T} r(\hat{p}_t)^2$ *for the average bias and average excess of any sequence of predictions in* $\mathcal{P}$.

(i) ***Lower bound, for every algorithm and every sample path:***

$$\overline{r^2}_T \;\ge\; (r^- + r^+)\,\bar{r}_T - r^- r^+.$$

Consequently $\mathrm{Regret}_T/T \ge (r^- + r^+)\,\mathbb{E}\bar{r}_T + uv$ *for every algorithm.*

(ii) ***Attainment.*** *Assume in addition (G6) and Assumption A.2 with moments up to order 6, and let* $a > \frac{1}{2}$. *There is* C *such that for every* $\zeta \in [r^-, r^+]$, *every start in* $\mathcal{P}$ *and every* n, **RSA**(a) *run on* $z_t - \zeta$ *satisfies*

$$\left|\mathbb{E}\bar{r}_n - \zeta\right| \le \frac{C}{\sqrt{n}}, \qquad \mathbb{E}\,\overline{r^2}_n \le (r^- + r^+)\zeta - r^- r^+ + \frac{C}{\sqrt{n}}.$$

(iii) ***The two ends.*** *The line* $b \mapsto (r^- + r^+)b - r^- r^+$ *on* $[r^-, r^+]$ *takes the value* $uv = |r^- r^+|$ *at* $b = 0$ *and the values* r^{-2}, r^{+2} *at the endpoints. Hence: among learners with asymptotically zero average bias, the smallest achievable average excess is* uv, *attained by* **RSA** ($\zeta = 0$, *Theorem C.8); without a bias constraint, the smallest is* $\min(r^{-2}, r^{+2})$, *attained at an endpoint (Proposition C.7 and Theorem C.19); and since* $uv > \min(u^2, v^2)$ *whenever* $u \ne v$, *no learner achieves both.*

Proof. **(i)** Every attainable residual lies outside the open gap: by Lemma C.2, $r(p) \le r^-$ for $p < x_0$ and $r(p) \ge r^+$ for $p > x_0$, and $r(x_0) \in \{r^-, r^+\}$. For any real x with $x \le r^-$ or $x \ge r^+$,

$$x^2 - \left[(r^- + r^+)x - r^- r^+\right] = (x - r^-)(x - r^+) \ge 0,$$

since both factors have the same sign (or one is zero). Apply this to $x = r(\hat{p}_t)$ and average over t. The regret statement follows by taking expectations, since the right-hand side is affine.

(ii) For $\zeta \in [r^-, r^+]$, Lemma C.15 gives modulus 1 with target $c = x_0$, at distance D_0 from $\{0, P\}$, and $|r - \zeta| \le B + \max(u, v)$. Steps (i)–(iv) of the proof of Lemma C.16 apply verbatim with $c = x_0$ and give $\mathbb{E}|\bar{r}_n - \zeta| \le C/\sqrt{n}$ and $\mathbb{E}\,\frac{1}{n}\sum_t |\hat{p}_t - x_0| \le 2\sqrt{M_1/n}$. Step (v) of that proof gives, for every $p \in \mathcal{P}$, $r^2 = r^{-2} + (r^+ + r^-)(r - r^- - \varepsilon) + \varepsilon'$ with $|\varepsilon|, |\varepsilon'| \le C'|p - x_0|$. Averaging and taking expectations,

$$\mathbb{E}\,\overline{r^2}_n \le r^{-2} + (r^+ + r^-)(\zeta - r^-) + |r^+ + r^-|\frac{C}{\sqrt{n}} + C''\sqrt{M_1/n},$$

and $r^{-2} + (r^+ + r^-)(\zeta - r^-) = (r^- + r^+)\zeta - r^- r^+$.

(iii) Direct evaluation of the line; for $u \ne v$, $uv - \min(u^2, v^2) = \min(u, v)\,|u - v| > 0$. □

Remark C.11 (Interpretation and scope). Theorem C.10 is the precise form of the gap result. Two points of language matter. First, zero is reachable only as an *average* residual, never pointwise: every realised prediction has $r(\hat{p}_t) \notin (r^-, r^+)$. Second, RSA's predictions are deterministic given the past; the zero-bias statement concerns the empirical occupation of the two sides, which converges to the unique convex combination with zero mean. Interpreting x_0 as a sliding (Filippov) equilibrium of the discontinuous drift $\dot{p} = -a\,r(p)$ is natural, but it is an interpretation, not a separate theorem here. The price of zero average bias relative to the accuracy optimum is $uv - \min(u^2, v^2) = \min(u, v)|u - v|$: zero for a symmetric gap and growing with its asymmetry. The myopic 2-cycle (Theorem C.5) spends equal time on both sides, so its average bias is $(r(\hat{p}_L) + r(\hat{p}_H))/2$ and it lies strictly above the frontier (Proposition C.9). RSA converges to x_0 while its time average splits between the two sides in the unique proportions that make the average bias zero; it is therefore optimal *subject to zero*

average bias, not suboptimal. Two scope remarks: (a) if heterogeneous thresholds smooth the jump, an equilibrium reappears; Proposition C.12 shows it is learnable by RSA with a constant growing like $1/h$, and repelling for myopic dynamics; (b) the gap is a property of the scalar prediction interface; a predictor that conditions on features identifying the groups may remove the aggregate discontinuity (open).

Proposition C.12 (Smoothing the jump)**.** *In the discrete-type model with the gap at x_0, replace the jump by heterogeneous thresholds: $\Phi_h(p) := p^2\big[G_k + (G_{k+1} - G_k)\,\Psi((p - x_0)/h)\big]$, where Ψ is a continuous strictly increasing distribution function with a continuous positive density ψ, and $h > 0$. Then:*

(i) Φ_h is continuous and nondecreasing, so $r_h(p) := p + \Phi_h(p) - \mu$ has modulus 1 and a unique root p_h^*, and $p_h^* \to x_0$ as $h \to 0$;

(ii) $R_h' := r_h'(p_h^*) = \frac{(u+v)\,\psi(\xi^*)}{h}(1 + o(1)) \to \infty$, where $\xi^* := \Psi^{-1}(u/(u+v))$;

(iii) RSA(a) with any fixed $a > \frac{1}{2}$ has logarithmic regret, with $\mathrm{Regret}_T / \ln T \to a^2 R_h'^2 \sigma^2/(2aR_h' - 1)$, which grows like $a\sigma^2 R_h'/2 \propto 1/h$;

(iv) the myopic map $f_h(p) = \mu/(1 + \Phi_h(p)/p)$ has $|f_h'(p_h^*)| \to \infty$; once $|f_h'(p_h^*)| > 1$ the equilibrium is repelling, and no myopic orbit converges to it unless it lands on it exactly.

So smoothing restores an equilibrium that RSA learns, but the learning constant grows like $1/h$ as $h \to 0$ and myopic dynamics remain non-convergent. The gap results describe the limit $h \to 0$.

Proof. **(i)** $\Phi_h'(p) = 2p[G_k + (G_{k+1} - G_k)\Psi] + p^2(G_{k+1} - G_k)\psi/h \geq 0$, so Lemma B.1(iii) applies with $m = 1$. Since $r_h(0) < 0 < r_h(P)$ and r_h is continuous and strictly increasing, there is a unique root. For fixed $\delta > 0$, $r_h(x_0 - \delta) \to r(x_0 - \delta) < r^- < 0$ and $r_h(x_0 + \delta) \to r(x_0 + \delta) > r^+ > 0$, so $p_h^* \in (x_0 - \delta, x_0 + \delta)$ for small h.

(ii) Let $\xi_h := (p_h^* - x_0)/h$. From $r_h(p_h^*) = 0$, $\Psi(\xi_h) = (\mu - p_h^* - G_k p_h^{*2})/((G_{k+1} - G_k)p_h^{*2}) \to (\mu - x_0 - G_k x_0^2)/((G_{k+1} - G_k)x_0^2) = u/(u+v)$, since $(G_{k+1} - G_k)x_0^2 = r^+ - r^- = u + v$. So $\xi_h \to \xi^*$ and $R_h' = 1 + \Phi_h'(p_h^*) = (u+v)\psi(\xi^*)/h\,(1 + o(1))$.

(iii) r_h satisfies (M2) with $m = 1$, (M3), (B1) and (B2) with p_h^* interior for small h, so Theorem B.17 applies for $a > \frac{1}{2}$.

(iv) With $\lambda_h(p) = \Phi_h(p)/p$, at the fixed point $\mu/(1 + \lambda_h) = p_h^*$, so $f_h'(p_h^*) = -p_h^* \lambda_h'(p_h^*)/(1 + \lambda_h(p_h^*))$, and $\lambda_h' \geq p(G_{k+1} - G_k)\psi/h$ gives $|f_h'| \to \infty$. If $|f'(p^*)| = \rho_0 > 1$, there is a neighbourhood where $|f(x) - p^*| \geq \frac{1+\rho_0}{2}|x - p^*|$; an orbit converging to p^* would eventually stay in it with $x_t \neq p^*$, and its distance to p^* would grow, a contradiction. □

Remark C.13. T28 (Gaussian Ψ): $R_h' = 6.5, 15.1, 39.7, 126$ for $h = 0.1, 0.03, 0.01, 0.003$; the measured RSA slope with $a = 1$ matches $R_h'^2/(2R_h' - 1)$ within 1%, and the myopic orbit stays in the same 2-cycle $\{0.433, 1.233\}$ as in the unsmoothed model. The locally tuned gain $a = 1/R_h'$ fails at these horizons for $h \leq 0.01$. It is covered asymptotically by Theorem B.21, which needs only $2aR' > 1$; but the global slope is small ($2am = 2/R_h' < 1$), so the transient before the asymptotic regime is very long (compare T52). The time scale on which smoothed behaviour takes over from gap behaviour is open.

C.2 The optimum in the gap: learning the cheaper side

Theorem C.8 shows that RSA balances the *signed* residual and so mixes both sides. A learner that cares about squared error should instead stay on the cheaper side. This subsection proves that this is achievable.

Lemma C.14 (Moments with a bounded residual)**.** *Let RSA(a) be run on the shifted residual $\rho(p) := r(p) - \zeta$, i.e. $\hat{p}_{t+1} = \Pi_{\mathcal{P}}(\hat{p}_t - (a/t)(z_t - \zeta))$. Suppose there are $c \in \mathcal{P}$, $m > 0$ and B' with $(p - c)\rho(p) \geq m(p - c)^2$ and $|\rho(p)| \leq B'$ for all $p \in \mathcal{P}$; assume Assumption A.2 with moments up to order $2k$ and $2am > 1$. Then $\mathbb{E}(\hat{p}_t - c)^{2k} \leq M_k'' t^{-k}$ for all t, with M_k'' depending only on $(a, m, B', P, M_2, \ldots, M_{2k}, k)$, and in particular not on c, ζ or $\hat{p}_1$.*

Proof. Follow Lemma B.12, replacing (M3) by $|\rho| \le B'$.

Step 1: with $e := \hat{p}_t - c$ and $w := e - (a/t)\rho(\hat{p}_t)$, $w^2 = e^2 - 2(a/t)e\rho + (a/t)^2\rho^2 \le e^2(1 - 2am/t) + a^2B'^2/t^2$, and $1 - 2am/t \in [0,1]$ for $t \ge \lceil 2am \rceil$.

Step 2: $(w^2)^k \le \sum_{i=0}^{k} \binom{k}{i} e^{2(k-i)}(1-2am/t)^{k-i}(a^2B'^2/t^2)^i$. The $i=0$ term is at most $e^{2k}(1 - 2amk/t + \binom{k}{2}(2am)^2/t^2)$ by the truncated-binomial inequality of Lemma B.12. For $i \ge 1$, by induction $\mathbb{E}e^{2(k-i)} \le Ct^{-(k-i)}$, so the term is at most $Ct^{-(k-i)-2i} \le Ct^{-k-1}$.

Step 3: for the noise terms use $|w| \le |e| + aB'/t$, so $\mathbb{E}|w|^q \le 2^{q-1}(\mathbb{E}|e|^q + (aB'/t)^q) \le Ct^{-q/2}$ for $q \le 2k-2$.

Step 4: the recursion $v_{t+1} \le (1 - 2amk/t + C/t^2)v_t + D/t^{k+1}$ holds and Lemma B.11 applies with $V_0 = P^{2k}$. □

Lemma C.15 (Every shifted residual has modulus one)**.** *Under (G1)–(G5), let $\zeta \in [r(0), r(P)]$ and define the target $c(\zeta)$: $c = x_0$ if $\zeta \in [r^-, r^+]$; c the unique $p \in [0, x_0)$ with $r(p) = \zeta$ if $\zeta \in [r(0), r^-)$; c the unique $p \in (x_0, P]$ with $r(p) = \zeta$ if $\zeta \in (r^+, r(P)]$. Then $(p-c)(r(p)-\zeta) \ge (p-c)^2$ for all $p \in \mathcal{P}$, and $|c - x_0| \le \operatorname{dist}(\zeta, [r^-, r^+])$.*

Proof. **Existence.** On $[0, x_0)$, r is continuous and strictly increasing from $r(0)$ to its limit r^-, so the intermediate value theorem gives exactly one root for $\zeta \in [r(0), r^-)$; the right case is symmetric.

Case $c = x_0$. For $p < x_0$, Lemma C.2 gives $r(p) - \zeta \le (r^- - \zeta) - (x_0 - p) \le -(x_0 - p)$; multiplying by $p - x_0 < 0$ gives $\ge (x_0 - p)^2$. For $p > x_0$, $r(p) - \zeta \ge (r^+ - \zeta) + (p - x_0) \ge p - x_0$.

Case $c < x_0$. For $p \in [0, x_0)$, Lemma C.2 with $r(c) = \zeta$ gives the inequality directly. For $p \ge x_0$: $r(p) - \zeta \ge r^+ + (p - x_0) - \zeta \ge r^- - \zeta + (p - x_0) \ge (x_0 - c) + (p - x_0) = p - c$, using $r^- - \zeta = r^- - r(c) \ge x_0 - c$.

Distance. In the left case $x_0 - c \le r^- - \zeta$ by the same inequality; the right case is symmetric. □

Lemma C.16 (Shifted RSA, uniformly in ζ)**.** *Assume (G1)–(G6) and Assumption A.2 with moments up to order 6. Let $w_0 := \min(D_0/2, (r^+ - r^-)/2)$ and $a > \frac{1}{2}$. There is C_G such that for every ζ with $|\zeta - r^-| \le w_0$, every start $\hat{p}_1 \in \mathcal{P}$ and every n, RSA(a) run on $z_t - \zeta$ for n rounds satisfies*

$$\mathbb{E}\Big[\frac{1}{n}\sum_{t \le n} r(\hat{p}_t)^2\Big] \le r^{-2} + C_G\big(|\zeta - r^-| + n^{-1/2}\big),$$

and the same with r^+ in place of r^-.

Proof. **(i) Moments.** $|\zeta - r^-| \le w_0 < r^- - r(0)$ because $r^- - r(0) \ge x_0 \ge D_0 > w_0$ (Lemma C.2). Lemma C.15 gives modulus 1 and a target c with $|c - x_0| \le |\zeta - r^-| \le D_0/2$, so c is at distance at least $D_0/2$ from $\{0, P\}$. Lemma C.14 with $m = 1$ and $B' = B + |r^-| + w_0$ gives $\mathbb{E}(\hat{p}_t - c)^{2k} \le M_k t^{-k}$, $k = 1, 2, 3$, uniformly.

(ii) Identity. Let $\tilde{e}_t := \hat{p}_t - c$ and $\pi_t := \hat{p}_{t+1} - y_t^u$. Multiplying the update by t: $(t+1)\tilde{e}_{t+1} - t\tilde{e}_t = \tilde{e}_{t+1} - a(r_t - \zeta) + a\eta_t + t\pi_t$. Summing,

$$a\sum_{t \le n}(r_t - \zeta) = \tilde{e}_1 - (n+1)\tilde{e}_{n+1} + \sum_{t \le n}\tilde{e}_{t+1} + a\sum_{t \le n}\eta_t + \sum_{t \le n} t\pi_t.$$

(iii) Projection terms. If $y_t^u \in \mathcal{P}$ then $\pi_t = 0$; otherwise $|\pi_t| = \operatorname{dist}(y_t^u, \mathcal{P}) \le |y_t^u - c| =: |f_t|$ and $|f_t| \ge D_0/2$. So $\mathbb{E}[t|\pi_t|] \le t\mathbb{E}f_t^6/(D_0/2)^5 \le Ct^{-2}$, using $\mathbb{E}f_t^6 \le 32(\mathbb{E}(\tilde{e}_t - (a/t)\rho)^6 + (a/t)^6 M_6) \le Ct^{-3}$; hence $\sum_t \mathbb{E}[t|\pi_t|] \le C_\pi$. *(iv)* With $\mathbb{E}|\tilde{e}_t| \le \sqrt{M_1/t}$ and $\mathbb{E}|\sum \eta_t| \le \bar{\sigma}\sqrt{n}$: $\mathbb{E}\big|\frac{1}{n}\sum_{t \le n}(r_t - \zeta)\big| \le \frac{1}{an}\big[P + 3\sqrt{M_1(n+1)} + a\bar{\sigma}\sqrt{n} + C_\pi\big] \le C/\sqrt{n}$.

(v) Decomposition. With $I(p) = 1$ on the right piece, (G6) gives $r(p) = r^- + (r^+ - r^-)I(p) + \varepsilon(p)$ with $|\varepsilon(p)| \le (1 + L_\Phi)|p - x_0|$, and, since $I^2 = I$, $r(p)^2 = r^{-2} + (r^{+2} - r^{-2})I(p) + \varepsilon'(p)$ with

$|\varepsilon'(p)| \le 2B(1+L_\Phi)|p-x_0|$. Eliminating I: $r^2 = r^{-2} + (r^+ + r^-)(r - r^- - \varepsilon) + \varepsilon'$. *(vi)* Averaging, taking expectations, using (iv) and $\mathbb{E}[\text{avg}|\hat{p}_t - x_0|] \le |c - x_0| + \text{avg}\,\mathbb{E}|\tilde{e}_t| \le |\zeta - r^-| + 2\sqrt{M_1/n}$ gives the claim. For r^+, eliminate $I-1$ instead: $r^2 = r^{+2} + (r^+ + r^-)(r - r^+ - \varepsilon) + \varepsilon'$. □

Lemma C.17 (Plain RSA in the gap identifies r^- and r^+)**.** *Under (G1)–(G5) and (A2-G), RSA(a) with any $a > 0$ satisfies, almost surely, $\frac{1}{n}\sum_{t\le n} z_t^2 - \sigma^2 \to uv$ and $\frac{1}{n}\sum_{t\le n} z_t^3 \to uv(v-u)$.*

Proof. By the proof of Theorem C.8, $r_t - [r^- + (r^+ - r^-)I_t] \to 0$, $\frac{1}{n}\sum I_t \to \pi = u/(u+v)$ and $\frac{1}{n}\sum r_t \to 0$ a.s. Hence $\frac{1}{n}\sum r_t^k \to (1-\pi)(-u)^k + \pi v^k$: for $k=2$ this is uv; for $k=3$ it is $(-u^3 v + uv^3)/(u+v) = uv(v-u)$. With $z = r - \eta$, $z^2 = r^2 - 2r\eta + \eta^2$ and $z^3 = r^3 - 3r^2\eta + 3r\eta^2 - \eta^3$. By the strong law, $\frac{1}{n}\sum \eta_t^2 \to \sigma^2$ and $\frac{1}{n}\sum \eta_t^3 \to 0$ (Gaussian). The averages of $r_t\eta_t$, $r_t^2\eta_t$ and $r_t(\eta_t^2 - \sigma^2)$ are averages of martingale differences with bounded conditional variance, so they tend to 0 (if X_t are martingale differences with $\mathbb{E}[X_t^2 \mid \mathcal{F}_{t-1}] \le C$, then $\sum X_t/t$ is an L^2-bounded martingale and Kronecker gives $\frac{1}{n}\sum X_t \to 0$); and $\frac{1}{n}\sum r_t\eta_t^2 = \frac{1}{n}\sum r_t(\eta_t^2 - \sigma^2) + \sigma^2 \frac{1}{n}\sum r_t \to 0$. □

Algorithm C.18 (Cheaper-side learner CS(n))**. Phase A:** run RSA(a_1), any $a_1 > 0$, for n rounds; set $\hat{m}_2 := \frac{1}{n}\sum z_t^2 - \sigma^2$, $\hat{m}_3 := \frac{1}{n}\sum z_t^3$.

Estimate: $\hat{q} := \max(\hat{m}_2, \epsilon)$ for a fixed small $\epsilon > 0$, $\hat{d} := \hat{m}_3/\hat{q}$, $\hat{u} := (-\hat{d} + \sqrt{\hat{d}^2 + 4\hat{q}})/2$, $\hat{v} := \hat{u} + \hat{d}$.

Target: $\hat{\zeta} := -\hat{u}$ if $\hat{v} > \hat{u}$, else $\hat{\zeta} := \hat{v}$.

Phase B: run RSA(a_2), $a_2 > \frac{1}{2}$, on $z - \hat{\zeta}$ for the remaining $T - n$ rounds, with the step index restarted at 1.

Theorem C.19 (The cheaper side is learnable)**.** *Assume (G1)–(G6), (A2-G) and that σ^2 is known. If $n = n_T \to \infty$ and $n_T/T \to 0$, then CS(n_T) satisfies*

$$\text{Regret}_T/T \to \min(r^{-2}, r^{+2}).$$

By Proposition C.7 no algorithm does better, so $\min(r^{-2}, r^{+2})$ is the optimal asymptotic average excess in the gap.

Proof. **Consistency.** Phase A is the first n steps of a single infinite run of RSA, so by Lemma C.17 $(\hat{m}_2, \hat{m}_3) \to (uv, uv(v-u))$ a.s. The map $(P, d) \mapsto (u, v)$ is continuous on $P > 0$, so $(\hat{u}, \hat{v}) \to (u, v)$ a.s.

Limit of $\hat{\zeta}$. Let $E^* := \{r^-, r^+\} \cap \{e : e^2 = \min(r^{-2}, r^{+2})\}$. If $u \ne v$ the sign of $\hat{v} - \hat{u}$ is eventually correct and $\hat{\zeta}$ converges to the single element of E^*; if $u = v$ both endpoints lie in E^* and every limit point of $\hat{\zeta}$ is in E^*. So $\delta_n := \text{dist}(\hat{\zeta}_n, E^*) \wedge w_0 \to 0$ a.s., hence $\mathbb{E}\delta_n \to 0$; and with $G_n := \{\text{dist}(\hat{\zeta}_n, E^*) < w_0\}$, $\mathbb{P}(G_n^c) \to 0$.

Phase B. Conditional on Phase A, Phase B is shifted RSA with fixed $\hat{\zeta}$, fresh noise and a start in $\mathcal{P}$. On G_n, Lemma C.16 at the nearest point of E^* gives $\mathbb{E}[\sum_B r_t^2 \mid \mathcal{F}_A] \le (T-n)[\min(r^{-2}, r^{+2}) + C_G(\delta_n + (T-n)^{-1/2})]$; on G_n^c use $(T-n)B^2$.

Total. $\mathbb{E}\text{Regret}_T \le n_T B^2 + (T - n_T)\big[\min + C_G\mathbb{E}\delta_{n_T} + C_G(T-n_T)^{-1/2} + B^2\mathbb{P}(G_{n_T}^c)\big]$. Divide by T. □

Lemma C.20 (Identification without knowing σ)**.** *Under the assumptions of Lemma C.17, also $\frac{1}{n}\sum_{t\le n} z_t^4 \to m_4 := uv(u^2 - uv + v^2) + 6\sigma^2 uv + 3\sigma^4$ a.s. With $m_2 := uv + \sigma^2$, $m_3 := uv(v-u)$ and $q := uv$, the number q is the unique positive root of*

$$h(q) := 2q^3 + (m_4 - 3m_2^2)q - m_3^2,$$

the root is simple, and $\sigma^2 = m_2 - q$, $v - u = m_3/q$.

Proof. **Fourth moment.** $z^4 = r^4 - 4r^3\eta + 6r^2\eta^2 - 4r\eta^3 + \eta^4$. As in Lemma C.17: $\frac{1}{n}\sum r_t^4 \to (1-\pi)u^4 + \pi v^4 = (vu^4 + uv^4)/(u+v) = uv(u^2 - uv + v^2)$; the averages of $r^3\eta$ and $r\eta^3$ are averages of martingale differences with bounded conditional variance ($\mathbb{E}\eta^3 = 0$), so they vanish; $\frac{1}{n}\sum r_t^2\eta_t^2 = \frac{1}{n}\sum r_t^2(\eta_t^2 - \sigma^2) + \sigma^2\frac{1}{n}\sum r_t^2 \to \sigma^2 uv$; and $\frac{1}{n}\sum \eta_t^4 \to 3\sigma^4$ (Gaussian).

The cubic. With $d := v - u$, $u^2 - uv + v^2 = d^2 + P$, so $m_4 - 3m_2^2 = P(d^2+P) + 6\sigma^2 P + 3\sigma^4 - 3(P+\sigma^2)^2 = Pd^2 - 2P^2$. Then $g(P) = 2P^3 + (Pd^2 - 2P^2)P - P^2d^2 = 0$.

Uniqueness and simplicity. If $m_3 \neq 0$, the coefficient signs of g are $(+, 0, \pm, -)$, with exactly one sign change either way, so by Descartes' rule g has exactly one positive root, and it is simple. If $m_3 = 0$ (then $u = v$), $g(P) = P(2P^2 + m_4 - 3m_2^2)$ with $m_4 - 3m_2^2 = -2P^2 < 0$, whose only positive root $\sqrt{(3m_2^2 - m_4)/2}$ is simple because $g' = 6P^2 + m_4 - 3m_2^2 = -2(m_4 - 3m_2^2) > 0$ there. □

Theorem C.21 (The cheaper side is learnable without knowing σ)**.** *Under the assumptions of Theorem C.19 except that σ^2 is unknown, modify Algorithm C.18 to compute also $\hat{m}_4 := \frac{1}{n}\sum z_t^4$, set $\hat{q}$ to the positive root of $2q^3 + (\hat{m}_4 - 3\hat{m}_2^2)q - \hat{m}_3^2$ (any fixed value if none exists), and $\hat{d} := \hat{m}_3/\hat{q}$, $\hat{u} := (-\hat{d} + \sqrt{\hat{d}^2 + 4\hat{q}})/2$, $\hat{v} := \hat{u} + \hat{d}$, where now $\hat{m}_2 := \frac{1}{n}\sum z_t^2$ is the raw second moment. Then* $\mathrm{Regret}_T/T \to \min(r^{-2}, r^{+2})$ *whenever $n_T \to \infty$ and $n_T/T \to 0$.*

Proof. By Lemmas C.17 and C.20, $(\hat{m}_2, \hat{m}_3, \hat{m}_4) \to (m_2, m_3, m_4)$ a.s. The population cubic has a unique simple positive root P; a simple root of a polynomial depends continuously on its coefficients, and by Descartes' rule the sample cubic has at most one positive root whenever $\hat{m}_3 \neq 0$, so $\hat{P} \to P$ a.s. Hence $(\hat{u}, \hat{v}) \to (u, v)$ a.s., and the rest of the proof of Theorem C.19 applies unchanged. □

Lemma C.22 (Rates for plain RSA in the gap)**.** *Assume (G1)–(G6) and (A2-G), and run RSA(a) with $a > \frac{1}{2}$. With $\bar{e}_n := \frac{1}{n}\sum_{t\le n}|\hat{p}_t - x_0|$ and $\bar{r}_n := \frac{1}{n}\sum_{t\le n} r(\hat{p}_t)$: $\mathbb{E}\bar{e}_n \le C(1+\ln n)/n$ and $\mathbb{E}|\bar{r}_n| \le Cn^{-1/2}$. Consequently $\mathbb{E}|\hat{m}_k - m_k| \le Cn^{-1/2}$ for $k = 2, 3$, where $\hat{m}_2, \hat{m}_3$ are the moment estimates of Algorithm C.18 and $m_2 = uv$, $m_3 = uv(v-u)$.*

Proof. **Average distance.** By Lemma F.13 (Appendix F.1), $\mathbb{E}\sum_{t\le n}|\hat{p}_t - x_0| \le C(1+\ln n)$.

Average residual. Lemma F.8 with $S_n = -\sum r_t + \sum \eta_t$ gives $a\sum_{t\le n} r_t = a\sum_{t\le n}\eta_t + \sum_{t\le n}(\hat{p}_t - \hat{p}_{n+1}) + \sum_t t\pi_t$. By Lemma F.13 and Lemma F.14 (Appendix F.1), $\mathbb{E}\sum|\hat{p}_t - x_0| \le C(1+\ln n)$, $n\mathbb{E}|\hat{p}_{n+1} - x_0| \le Cn^{1/3}$ and $\mathbb{E}\sum t|\pi_t| \le C$; so $\mathbb{E}|\bar{r}_n| \le \sigma/\sqrt{n} + C(1+\ln n + n^{1/3})/n \le Cn^{-1/2}$.

Residual moments. With I_t the side indicator and $\pi = u/(u+v)$, (G6) gives $r^k = r^{-k} + (r^{+k} - r^{-k})I + \varepsilon_k$ with $|\varepsilon_k| \le C|\hat{p} - x_0|$. Averaging, and using $\bar{r}_n = r^- + (r^+ - r^-)\bar{I}_n + \bar{\varepsilon}_{1,n}$ and $r^- + (r^+ - r^-)\pi = 0$, we get $|\bar{I}_n - \pi| \le (|\bar{r}_n| + C\bar{e}_n)/(u+v)$ and hence $|\frac{1}{n}\sum r_t^k - [(1-\pi)r^{-k} + \pi r^{+k}]| \le C(|\bar{r}_n| + \bar{e}_n)$.

Noise terms. Each noise average in the expansions of z^2 and z^3 (proof of Lemma C.17) is an average of martingale differences with bounded conditional variance, or of i.i.d. centred terms, so its L^1 norm is at most $Cn^{-1/2}$. □

Theorem C.23 (Rate for the cheaper-side learner)**.** *Under the assumptions of Theorem C.19, with $a_1, a_2 > \frac{1}{2}$ and $n_T := \lceil T^{2/3} \rceil$,*

$$\mathbb{E}\mathrm{Regret}_T - T\min(r^{-2}, r^{+2}) \le C\,T^{2/3}.$$

Proof. **Estimator error.** The map from (m_2, m_3) to the candidate targets $(-u, v)$ is smooth near the population values, because $P = uv > 0$. So there are $\delta, L_F > 0$ such that on $A := \{|\hat{m}_2 - m_2| + |\hat{m}_3 - m_3| \le \delta\}$ we have $|\hat{u} - u| + |\hat{v} - v| \le L_F(|\hat{m}_2 - m_2| + |\hat{m}_3 - m_3|)$, and, if $u \neq v$, the side decision is correct once $L_F\delta < |v - u|/2$. By Lemma C.22 and Markov's inequality, $\mathbb{P}(A^c) \le Cn^{-1/2}$. Hence $\mathbb{E}[\mathrm{dist}(\hat{\zeta}, E^*) \wedge w_0] \le L_F\,Cn^{-1/2} + w_0\mathbb{P}(A^c) \le Cn^{-1/2}$, and $\mathbb{P}(G_n^c) \le \mathbb{P}(A^c) \le Cn^{-1/2}$ if $L_F\delta < w_0$.

Total. The final display of the proof of Theorem C.19 gives $\mathbb{E}\mathrm{Regret}_T - T\min \le nB^2 + T\big[C_G C n^{-1/2} + B^2 C n^{-1/2}\big] + C_G\sqrt{T}$. With $n = \lceil T^{2/3}\rceil$ every term is $O(T^{2/3})$. □

Remark C.24. T25 finds $(\mathbb{E}\mathrm{Regret}_T - T\min)/T^{2/3}$ between 0.43 and 0.69 for $T \in [10^4, 10^6]$ on two instances. Whether $T^{2/3}$ is optimal is open (Appendix I).

Remark C.25. The estimator needs no localisation of x_0 and no probes: the telescoping identity ties the mixture weights of RSA's residuals to r^- and r^+. The rate is Theorem C.23; the case of unknown σ is Theorem C.21; the analysis of the probe-based algorithm of E8 is open. Numerical results are in Appendix H.1 (T16).

D A LEARNER WITH A STAKE: PROOFS

D.1 THE STAKE MODEL WITH THE LEARNER LEADING AND THE STRATEGIC FORECAST

In the gap there is no calibrated forecast, and where a learner lands on the frontier of Theorem C.10 is decided by what it targets: zero average bias for RSA (Theorem C.8), accuracy for the cheaper-side learner (Theorem C.19). The data fix r^- and r^+, and hence the frontier; the target picks the point on it. In a smooth world an accuracy-only learner has no such choice to make: its excess loss is $r(p)^2$ and r has at most one root, so it has no reason to depart from the calibrated forecast (Lemma B.1). A persistent departure where calibration is possible therefore needs an objective other than accuracy; the simplest is a stake in the outcome, and a stake is exactly a conflict of interest (Proposition D.26). This section gives the learner one, and asks what it then does, what it costs to learn, and whom it harms.

Assumption D.1 (The stake model with the learner leading)**.** The quadratic world with accuracy weight $\alpha \in (0,1)$ and portfolio position fixed at $b = \mu$ (this isolates the publication choice). The expected mixed loss, less σ^2, is

$$V(p) := \alpha\, r(p)^2 + (1-\alpha)\gamma^2 p^4, \qquad c(p;\alpha) := V'(p)/2.$$

Proposition D.2 (Honesty is stability; manipulation is optimality)**.** *Let $\mathcal{D}(\hat p) := N(\mu - \Phi(\hat p), \sigma^2)$ and $\ell(\phi; y) := \alpha(\phi - y)^2 + (1-\alpha)(\mu - y)^2$. (i) For $\alpha > 0$ the only performatively stable point, $\hat p \in \arg\min_\phi \mathbb{E}_{y\sim\mathcal{D}(\hat p)}\ell(\phi; y)$, is p^*. (ii) The performatively optimal points, $\arg\min_{\hat p} \mathbb{E}_{y\sim\mathcal{D}(\hat p)}\ell(\hat p; y)$, are the minimisers of V.*

Proof. **(i)** With the distribution frozen at $\mathcal{D}(\hat p)$, the second term of ℓ does not depend on ϕ, and the first is minimised at $\phi = \mathbb{E}_{\mathcal{D}(\hat p)} y = \mu - \Phi(\hat p)$. So $\hat p$ is stable iff $\hat p = \mu - \Phi(\hat p)$, i.e. $r(\hat p) = 0$.

(ii) $\mathbb{E}_{\mathcal{D}(\hat p)}(\hat p - y)^2 = r(\hat p)^2 + \sigma^2$ and $\mathbb{E}_{\mathcal{D}(\hat p)}(\mu - y)^2 = \gamma^2\hat p^4 + \sigma^2$, so the objective is $V(\hat p) + \sigma^2$. □

Proposition D.3 (The strategic forecast)**.** *(i) $c(p;\alpha) = 2\gamma^2 p^3 + 3\alpha\gamma p^2 + \alpha(1 - 2\gamma\mu)p - \alpha\mu$. (ii) c has exactly one positive root $p^S(\alpha)$, and it is simple. (iii) $p^S(\alpha) \in (0, p^*)$. (iv) $p^S(\alpha)$ is the unique minimiser of V on $[0,\infty)$, and $\partial_p c(p^S) > 0$.*

Proof. **(i)** $V'/2 = \alpha(p + \gamma p^2 - \mu)(1 + 2\gamma p) + 2(1-\alpha)\gamma^2 p^3$; expanding $(p + \gamma p^2 - \mu)(1 + 2\gamma p) = 2\gamma^2 p^3 + 3\gamma p^2 + (1 - 2\gamma\mu)p - \mu$ and collecting, the p^3 coefficient is $2\alpha\gamma^2 + 2(1-\alpha)\gamma^2 = 2\gamma^2$.

(ii) The coefficient signs are $(+, +, \mathrm{sgn}(\alpha(1 - 2\gamma\mu)), -)$: the sequences $(+,+,+,-)$, $(+,+,-)$ and $(+,+,-,-)$ each have exactly one sign change. By Descartes' rule of signs the number of positive roots with multiplicity is 1 minus an even nonnegative integer, hence exactly one simple positive root.

(iii) $c(0) = -\alpha\mu < 0$, and $c(p^*) = \alpha r(p^*)(1 + 2\gamma p^*) + 2(1-\alpha)\gamma^2 {p^*}^3 = 2(1-\alpha)\gamma^2 {p^*}^3 > 0$; apply the intermediate value theorem.

(iv) By (ii), $c < 0$ on $(0, p^S)$ and $c > 0$ on (p^S, ∞); since $V' = 2c$, V decreases on $[0, p^S]$ and increases on $[p^S, \infty)$. The root is simple and c changes sign from negative to positive, so $\partial_p c(p^S) > 0$. □

Remark D.4 (Scaling). With $p = \mu u$ and $g := \gamma\mu$, $c(\mu u) = \mu\,\tilde{c}(u;\alpha,g)$ with $\tilde{c}(u) = 2g^2u^3 + 3\alpha g u^2 + \alpha(1-2g)u - \alpha$, and $r = \mu(u + gu^2 - 1)$ (checked symbolically). Every quantity in this section is a power of μ times a function of (α, g).

Proposition D.5 (Monotonicity in α). *p^S is strictly increasing on $[0,1]$, with $p^S(0) := 0$ and $p^S(1) = p^*$.*

Proof. Write $c(p;\alpha) = 2\gamma^2p^3 + \alpha q(p)$ with $q(p) := 3\gamma p^2 + (1-2\gamma\mu)p - \mu$. Take $0 < \alpha_1 < \alpha_2 < 1$ and $p_1 := p^S(\alpha_1)$. From $c(p_1;\alpha_1) = 0$, $\alpha_1 q(p_1) = -2\gamma^2p_1^3 < 0$, so $q(p_1) < 0$ and $c(p_1;\alpha_2) = 2\gamma^2p_1^3 + \alpha_2 q(p_1) < 2\gamma^2p_1^3 + \alpha_1 q(p_1) = 0$. By Proposition D.3(iv), $c(\cdot;\alpha_2) < 0$ exactly on $(0, p^S(\alpha_2))$, so $p_1 < p^S(\alpha_2)$. At $\alpha = 1$, $c(p;1) = (2\gamma p + 1)(\gamma p^2 + p - \mu)$ with only positive root p^*; at $\alpha = 0$ the root is 0. □

Proposition D.6 (Regularity). *p^S is continuous on $[0,1]$ and C^∞ on $(0,1]$.*

Proof. At $\alpha_0 \in (0,1]$: for $\varepsilon > 0$, $c(p^S(\alpha_0) - \varepsilon;\alpha_0) < 0 < c(p^S(\alpha_0) + \varepsilon;\alpha_0)$; by joint continuity both persist near α_0, so the unique positive root stays within ε. At 0: $c(\varepsilon;\alpha) \to 2\gamma^2\varepsilon^3 > 0$ as $\alpha \to 0$, forcing $p^S(\alpha) < \varepsilon$. Smoothness: $\partial_p c(p^S) > 0$ and c is polynomial. □

Theorem D.7 (Information rent). *Let $\rho(\alpha) := \alpha\, r(p^S(\alpha))^2$. Then (i) $\rho(0) = 0$; (ii) $\rho(1) = 0$; (iii) $\rho > 0$ on $(0,1)$; (iv) ρ attains its maximum at an interior α^*; (v) the maximiser is unique. Explicitly, with $y := \gamma p^* = (s-1)/2$, let t^* be the unique positive root of*

$$-4y^4(2y+1)\,t^4 - y^2(12y^2+6y+1)\,t^3 + y^2(12y^3+24y^2+9y+1)\,t^2 + y^2(y+1)(4y^2+16y+5)\,t + 3y^2(y+1)^2;$$

then $x^* := yt^*/(1+t^*)$ and $\alpha^* = 2x^{*3}/D(x^*)$ with $D(x) := 2x^3 + (y-x)(1+x+y)(1+2x)$.

Proof. **(i)–(iii)** (i) the factor α; (ii) $p^S(1) = p^*$; (iii) $\alpha > 0$ and $p^S < p^*$ give $r(p^S) < 0$.

(iv) ρ is continuous on $[0,1]$ (Proposition D.6), so it attains its maximum (extreme value theorem), which is positive by (iii) and hence interior.

(v) Reparametrisation. By Proposition D.5 (p^S strictly increasing with $p^S(0) = 0$ and $p^S(1) = p^*$) and Proposition D.6 (continuity on $[0,1]$), $\alpha \mapsto x := \gamma p^S(\alpha)$ is a continuous strictly increasing bijection of $(0,1)$ onto $(0,y)$, with inverse $\alpha = N/(N+M) = 2x^3/D(x)$ and $|r| = (y-x)(1+x+y)/\gamma$. Hence $\gamma^2\rho = F(x) := 2x^3(y-x)^2(1+x+y)^2/D(x)$, and ρ has a unique maximiser in α iff F has a unique maximiser in $x \in (0,y)$.

(v) Derivative. A direct computation (script `t_unique.py`) gives

$$\frac{d}{dx}\log F(x) = \frac{Q(x,y)}{x\,(y-x)\,(1+x+y)\,D(x)},$$

where $Q(x,y) = 15x^4 - 12x^3y^2 - 12x^3y + 15x^3 - 18x^2y^2 - 18x^2y + 4x^2 + 4xy^4 + 8xy^3 - 3xy^2 - 7xy + 3y^4 + 6y^3 + 3y^2$. The identity $D(x) = -(3x^2 - 2xy^2 - 2xy + x - y^2 - y)$ used to write the denominator in this form was checked symbolically. The denominator is positive on $(0,y)$, so F' has the sign of Q.

(v) Sign changes. Substitute $x = yt/(1+t)$, a strictly increasing bijection of $t \in (0,\infty)$ onto $(0,y)$, and multiply by $(1+t)^4 > 0$. The result is the quartic in t displayed in the statement, whose coefficients have signs $(-,-,+,+,+)$ for every $y > 0$. By Descartes' rule it has exactly one positive root t^*, which is simple. Its constant term is positive and its leading coefficient negative, so $Q > 0$ for $x < x^*$ and $Q < 0$ for $x > x^*$. Hence F increases then decreases, and x^* is its unique maximiser. □

Remark D.8. The peak location depends on $g = \gamma\mu$: $\alpha^* = 0.00032, 0.0724, 0.160, 0.204, 0.273, 0.327$ at $g = 0.01, 0.3, 1, 2, 10, 1000$, matching a grid search to five decimals (T22), and $\alpha^* \to$ about 0.33 as $g \to \infty$.

Proposition D.9 (The accuracy cost is monotone). *$r(p^S(\alpha))^2$, the excess prediction-error loss per round of the strategic learner, is strictly decreasing on $[0,1]$, with maximum μ^2 at $\alpha = 0$.*

Proof. r and p^S are increasing (Proposition D.5), so $r(p^S(\alpha))$ increases from $-\mu$ to 0 and stays negative on $[0, 1)$. □

Remark D.10 (Which cost?). Under prediction-accuracy regret, the strategic learner's per-round cost is $r(p^S)^2 = \rho_\alpha/\alpha$, not ρ_α. The exact identity $r(p^S)^2 = \alpha r(p^S)^2 + (1-\alpha)r(p^S)^2$ splits the harm into the part the learner internalises through reputation (ρ_α) and the part pushed onto users. "A moderate accuracy weight is worst" holds only for ρ_α.

D.2 The learner's incentive and its welfare cost

Proposition D.11 (The learner's own incentive)**.** *In the quadratic world with $\alpha \in (0, 1)$, let the honest learner run* RSA(a) *under Theorem B.13, let $H_T := \sum_t \mathbb{E}V(\hat{p}_t) - T\,V(p^S)$ and $\Delta_V := V(p^*) - V(p^S)$. Then $\Delta_V > 0$ and $H_T \geq T\Delta_V - 2V_1\sqrt{M_1' T}$ with $V_1 := \sup_{\mathcal{P}} |V'|$. Consequently, for all $T \geq T_s := \left((2\sqrt{B_V C_S} + 2B_V + 2V_1\sqrt{M_1'})/\Delta_V\right)^2$,* ETC-S *(with B_V and C_S as in Theorem E.2) has strictly smaller own-objective regret than honest* RSA*. T_s is a* sufficient incentive horizon*: an explicit upper bound on the crossover, not the crossover itself.*

Proof. $\Delta_V > 0$ because p^S is the unique minimiser and $p^* \neq p^S$. By the mean value theorem $|V(\hat{p}_t) - V(p^*)| \leq V_1|e_t|$; by Jensen and Theorem B.13, $\mathbb{E}|e_t| \leq \sqrt{M_1'/t}$; and $\sum_{t\leq T} t^{-1/2} \leq 2\sqrt{T}$. So $H_T = T\Delta_V + \sum_t[\mathbb{E}V(\hat{p}_t) - V(p^*)] \geq T\Delta_V - 2V_1\sqrt{M_1' T}$. By Theorem E.2, $S_T \leq 2\sqrt{B_V C_S T} + 2B_V$, which is below H_T once $T\Delta_V > (2\sqrt{B_V C_S} + 2V_1\sqrt{M_1'})\sqrt{T} + 2B_V$; since $\sqrt{T} \leq T$ for $T \geq 1$, this holds for all $T \geq T_s$. □

Remark D.12. In its own objective, a learner with $\alpha < 1$ prefers manipulation at every sufficiently long horizon; the opposite reading, that long horizons favour honesty, holds only under prediction-accuracy regret. Values of Δ_V at $\mu = 1$, $\gamma = 2$: 0.1213, 0.0370, 0.0012 at $\alpha = 0.2, 0.5, 0.9$. Experiment E12 locates the crossover of an untuned ETC-S at $T \approx 1.1 \times 10^3$, 1.4×10^4 and 1.8×10^7 respectively.

Proposition D.13 (Welfare comparative statics)**.** *In the stake model with the learner leading, as α increases on $(0, 1)$: (i) the learner's incentive to manipulate, $\Delta_V(\alpha) = V_\alpha(p^*) - \min_p V_\alpha(p)$, strictly decreases; (ii) the harm to users, $r(p^S(\alpha))^2$, strictly decreases (Proposition D.9); (iii) the harm to users strictly exceeds the learner's gain: with $x := \gamma p^S$ and $y := \gamma p^*$ (so $0 < x < y$),*

$$r(p^S)^2 - \Delta_V = \frac{(y-x)^3\,(1+x+y)\,(2x^2+4xy+4x+2y+1)}{\gamma^2\left[2x^3 + (y-x)(1+x+y)(1+2x)\right]} > 0.$$

Consequently the α-weighted rent ρ_α is the only one of these quantities that is nonmonotone (Theorem D.7), the sufficient horizon $T_s \propto \Delta_V^{-2}$ increases with α, and manipulation always destroys value.

Proof. **(i)** $V_\alpha(p) = \alpha r(p)^2 + (1-\alpha)\gamma^2 p^4$ is affine in α, and the minimiser is unique (Proposition D.3); by Danskin's theorem $\frac{d}{d\alpha}\min_p V_\alpha = r(p^S)^2 - \gamma^2 p^{S4}$. Hence $\Delta_V'(\alpha) = [r(p^*)^2 - \gamma^2 p^{*4}] - [r(p^S)^2 - \gamma^2 p^{S4}] = -\gamma^2(p^{*4} - p^{S4}) - r(p^S)^2 < 0$.

(iii)

Parametrisation. By Propositions D.5 and D.6, $\alpha \mapsto p^S(\alpha)$ is a continuous strictly increasing bijection of $(0, 1)$ onto $(0, p^*)$. Conversely, for $p \in (0, p^*)$ the first-order condition $\alpha r(p)(1+2\gamma p) + 2(1-\alpha)\gamma^2 p^3 = 0$ determines α uniquely:

$$\alpha = \frac{N}{N+M}, \qquad N := 2\gamma^2 p^3 > 0, \qquad M := |r(p)|(1+2\gamma p) > 0,$$

and by Proposition D.3(ii) that p is $p^S(\alpha)$. So it suffices to prove the identity for every $p \in (0, p^*)$ with $\alpha = N/(N+M)$.

Ingredients. $\mu = p^* + \gamma p^{*2}$ gives $|r(p)| = \mu - p - \gamma p^2 = (p^* - p)(1 + \gamma(p^* + p))$. With $x = \gamma p$, $y = \gamma p^*$: $|r| = (y-x)(1+x+y)/\gamma$, $N = 2x^3/\gamma$, $M = (y-x)(1+x+y)(1+2x)/\gamma$, and $\gamma^2 p^4 = x^4/\gamma^2$, $\gamma^2 p^{*4} = y^4/\gamma^2$.

Algebra. $\Delta_V = (1-\alpha)\gamma^2 p^{*4} - \alpha r^2 - (1-\alpha)\gamma^2 p^4$, so $r^2 - \Delta_V = (1+\alpha)r^2 - (1-\alpha)(\gamma^2 p^{*4} - \gamma^2 p^4)$. Substituting $\alpha = N/(N+M)$ and multiplying by $\gamma^4(N+M)$, the expression factors (computer algebra, script `t_dwl.py`; it can be checked by expanding both sides) as

$$\gamma^4(N+M)\,(r^2 - \Delta_V) = \gamma\,(y-x)^3(1+x+y)(2x^2+4xy+4x+2y+1).$$

Dividing by $\gamma^4(N+M) = \gamma^3[2x^3 + (y-x)(1+x+y)(1+2x)]$ gives the display. Each factor is positive for $0 < x < y$. □

Remark D.14. The closed form agrees with direct numerical computation to a relative error below 10^{-6} on a grid of $\gamma\mu \in [0.01, 300]$ and $\alpha \in [0.05, 0.97]$ (T21, limited by root-finding tolerance), and the ratio $r(p^S)^2/\Delta_V$ ranged from 1.0002 to 94.8 on the grid of T20. As $\alpha \to 1$ ($x \to y$) both sides vanish and their ratio tends to 1.

D.3 The Stackelberg structure: who gains, who pays, and how the cost scales

The learner is a Stackelberg leader: it commits to $\hat{p}$, and the world, a rational follower, best-responds with $\lambda^*(\hat{p})$ (Proposition A.7). Honest publication of p^* is the performatively stable outcome, in which neither side exploits the order of play; strategic publication of p^S is the Stackelberg (performatively optimal) outcome (Proposition D.2). The leader's gain from abandoning honesty is Δ_V. This subsection makes the title precise: the world's rationality costs nothing to an honest learner (Appendices B.3 and B.5), but once the learner exploits its first-mover position the cost appears, and it grows with the follower's responsiveness $\gamma\mu$. (The follower is equally rational at every responsiveness; what varies is how strongly its optimal action reacts to the forecast.)

Proposition D.15 (Incidence of the Stackelberg distortion)**.** *In the stake model with the learner leading, with the world's cost $C_W(\lambda) = \alpha_W\lambda^2 + \beta_W(\lambda - \kappa_W\hat{p})^2$ of Proposition A.7:*

(i) the leader gains $\Delta_V > 0$ per round relative to the honest forecast;

(ii) the follower's minimised cost is $c_W\hat{p}^2$ with $c_W := \alpha_W\beta_W\kappa_W^2/(\alpha_W+\beta_W)$, so the follower also gains, $c_W(p^{*2} - p^{S^2}) > 0$ per round;

(iii) the users of the forecast lose $r(p^S)^2$ per round, which exceeds the leader's gain (Proposition D.13(iii)).

So both players prefer the Stackelberg outcome; its cost falls on those who rely on the forecast being accurate.

Proof. (i) is Proposition D.11. (ii) At $\lambda^* = \gamma\hat{p}$ with $\gamma = \beta_W\kappa_W/(\alpha_W+\beta_W)$: $C_W = \alpha_W\gamma^2\hat{p}^2 + \beta_W(\gamma-\kappa_W)^2\hat{p}^2$, and $\gamma - \kappa_W = -\alpha_W\kappa_W/(\alpha_W+\beta_W)$, so $C_W = \frac{\alpha_W\beta_W^2\kappa_W^2 + \beta_W\alpha_W^2\kappa_W^2}{(\alpha_W+\beta_W)^2}\hat{p}^2 = c_W\hat{p}^2$; and $p^S < p^*$ (Proposition D.3). (iii) is Proposition D.13(iii). □

Remark D.16 (Which reading this uses). Part (ii), that the follower also gains, scores the world with $C_W = \alpha_W\lambda^2 + \beta_W(\lambda - \kappa_W\hat{p})^2$, i.e. against the *published* number. That is exact under the instrument reading (the published number is what the world acts on and is paid by). Under the forecast reading, a follower misled about the outcome should be scored against the truth, and the sign of its gain can change; the part (ii) statement is claimed only for the instrument reading.

Theorem D.17 (The strategic cost of rationality)**.** *In the stake model with the learner leading, fix $\alpha \in (0,1)$ and measure the follower's responsiveness by $g := \gamma\mu$ (larger when the follower weights matching the forecast more, β_W, or finds acting cheaper, α_W). Then, as functions of $g \in (0,\infty)$:*

(i) the harm to users $r(p^S)^2$ is strictly increasing, with $r(p^S)^2 \to 0$ as $g \to 0$ and $r(p^S)^2 \to (1-\alpha)^2\mu^2$ as $g \to \infty$;

(ii) the rent $\rho_\alpha = \alpha\, r(p^S)^2$ is strictly increasing;

(iii) the leader's gain from abandoning honesty Δ_V is strictly increasing.

Proof. By scaling take $\mu = 1$; write $u := p^S$ and $v := p^*$, so $0 < u < v < 1$, $gv^2 + v = 1$, and $|r(p^S)| = 1 - u - gu^2$.

(i) Monotonicity. Differentiating the root of $c(u;\alpha,g) = 2g^2u^3+3\alpha gu^2+\alpha(1-2g)u-\alpha$ implicitly, $u_g = -c_g/c_u$ with $c_u > 0$ (Proposition D.3), so

$$\frac{d|r|}{dg} = -\big(u_g(1+2gu)+u^2\big) = \frac{c_g(1+2gu)-u^2c_u}{c_u}.$$

Eliminating α with $c=0$ and writing $x := gu$, $y := gv$ (so $g = y+y^2$ and $0<x<y$), computer algebra (script `t_stack2.py`; checkable by expansion) gives

$$c_g(1+2gu)-u^2c_u = \frac{2x^3\,(y-x)\,(3x+2)\,(1+x+y)}{y^2(1+y)^2\,D(x)} > 0,$$

with $D(x) = 2x^3+(y-x)(1+x+y)(1+2x) > 0$ as in Theorem D.7.

(i) Limits. As $g\to 0$: $r(p^S)^2 \le V(p^S)/\alpha \le V(p^*)/\alpha = (1-\alpha)\gamma^2 p^{*4}/\alpha \le (1-\alpha)g^2/\alpha \to 0$. As $g\to\infty$: put $u = w/\sqrt{g}$; then $c/\sqrt{g} = 2w(w^2-\alpha)+(3\alpha w^2-\alpha)/\sqrt{g}+\alpha w/g$, whose positive root converges to the simple root $w=\sqrt{\alpha}$ of $2w(w^2-\alpha)$; hence $|r| = 1-w/\sqrt{g}-w^2 \to 1-\alpha$.

(ii) $\rho_\alpha = \alpha r(p^S)^2$ with α fixed.

(iii) $\Delta_V(g) = V_g(p^*(g)) - \min_p V_g(p)$. By Danskin's theorem and the chain rule, with $p^{*\prime}(g) = -v^2/(1+2gv)$, $\Delta_V'(g) = \partial_g V(p^*)+\partial_p V(p^*)\,p^{*\prime}(g)-\partial_g V(p^S)$. Eliminating α and $g=(1-v)/v^2$, computer algebra (script `t_stack3.py`) gives

$$\Delta_V'(g) = \frac{2(v-u)^2(1-v)(u+v-uv)\,Q_1(u,v)}{v^2(2-v)\,Q_2(u,v)},$$

with $Q_1 := v^4+(2-v)(u^3+u^2v+uv^2) > 0$ and $Q_2 := 3u^2(v-1)-uv^2-2uv+2u+v^2$. For $0<u<v<1$ every factor of the numerator and $v^2(2-v)$ are positive. Q_2 is concave in u (leading coefficient $3(v-1)<0$) and positive at both ends of $[0,v]$ ($Q_2(0)=v^2$, $Q_2(v)=2v(1-v)^2$), hence positive on $[0,v]$. □

Remark D.18 (Learning to lead versus learning to follow). The same pattern holds for learning. Learning the honest forecast costs regret of order $\ln T$ whatever the follower's responsiveness; the constant σ^2 of Theorem B.29 is attained by the tuned gain, which uses the slope at the answer, and almost surely by the adaptive learner (Theorem B.17 and Proposition B.24): the world's rationality is essentially free to an honest learner. Learning the Stackelberg outcome costs $\Theta(\sqrt{T})$ when μ is unknown (Theorems E.2 and E.3) and $\Theta(\log T)$ when μ is known (Theorem E.8), because the follower's response is not identifiable at the Stackelberg action (Theorem E.1). Related work: learning optimal commitments in finite Stackelberg security games also costs $\sqrt{T}$ (Balcan et al., 2015); regret for performative optima is studied by Jagadeesan et al. (2022); who leads can depend on update speeds (Zrnic et al., 2021); and the social cost of a decision-maker's strategic advantage is studied in strategic classification by Milli et al. (2019). The contribution here is the continuous-action, noisy-outcome version with an explicit mechanism and a matching $\Theta(\sqrt{T})$ versus $\Theta(\log T)$ dichotomy, together with Proposition D.15 and Theorem D.17.

D.4 Orders of play: conflict and manipulation

Appendix D.1 studies the learner (the learner) leading. This subsection places that regime among the timing structures in which both players are rational and proves the two results used in Section 5: at the honest forecast the learner's incentive splits exactly into a conflict-of-interest part and a manipulation part, in the ratio $1:\varepsilon$, where ε is the elasticity of the world's response intensity (Proposition D.27); and, under single crossing, learner leadership adds harm relative to simultaneous play if and only if $\varepsilon > 0$, for every stake (Theorem D.34). Regimes in which the world leads, or in which both players internalise each other, are left to a companion paper.

D.4.1 Model, timing and definitions

Assumption D.19 (The stake model)(R1) The learner publishes $p \in \mathcal{P} = [0,P]$ and the world chooses an intensity $\lambda \in [0,\bar{\Lambda}]$. The outcome is $y = \mu - \lambda p + \eta$, with $\mu > 0$, $P \ge \mu$, and noise η with $\mathbb{E}\eta = 0$, $\mathbb{E}\eta^2 = \sigma^2$, independent of (p,λ). This *multiplicative channel* is the micro-foundation of Proposition A.7: $\Phi(p) = \lambda p$.

(R2) Objectives (expected losses up to the constant σ^2):

$$U_A(p,\lambda) = \alpha\big(p(1+\lambda)-\mu\big)^2 + (1-\alpha)\lambda^2 p^2,$$
$$U_W(p,\lambda) = \alpha_W\lambda^2 + \beta_W(\lambda-\kappa_W p)^2, \qquad L(p,\lambda) = \big(p(1+\lambda)-\mu\big)^2,$$

with $\alpha \in (0,1]$, $\alpha_W, \beta_W, \kappa_W > 0$, and $\bar{\Lambda} \geq \kappa_W P(1+\kappa_W\mu)$. The *stake* is $1-\alpha$. L is the loss of third parties who use the forecast.

With $\lambda = \mathrm{BR}_W(p)$ and $\mathrm{BR}_W(p) = \gamma p$, $\gamma = \beta_W\kappa_W/(\alpha_W+\beta_W)$, the reduced objective $U_A(p,\gamma p) = \alpha r(p)^2 + (1-\alpha)\gamma^2 p^4$ is exactly the objective V of Assumption D.1: The stake model with the learner leading is the stake model with the learner leading.

Lemma D.20 (Best responses)**.** *Under Assumption D.19: (i)* $\mathrm{BR}_W(p) := \arg\min_\lambda U_W(p,\lambda) = \gamma p$*; (ii) for each* $\lambda \geq 0$*,* $U_A(\cdot,\lambda)$ *is strictly convex and*

$$p_\lambda := \mathrm{BR}_A(\lambda) = \frac{\alpha(1+\lambda)\mu}{\alpha(1+\lambda)^2 + (1-\alpha)\lambda^2} \in (0,\mu];$$

(iii) p_λ *is strictly decreasing in* λ*.*

Proof. (i) $\partial_\lambda U_W = 2\alpha_W\lambda + 2\beta_W(\lambda-\kappa_W p) = 0$ gives $\lambda = \gamma p$, the unique minimiser of a strictly convex quadratic; $\gamma p \leq \gamma P \leq \bar{\Lambda}$. (ii) The coefficient of p^2 in U_A is $D(\lambda) := \alpha(1+\lambda)^2 + (1-\alpha)\lambda^2 \geq \alpha > 0$, and $\partial_p U_A = 2\alpha(1+\lambda)(p(1+\lambda)-\mu) + 2(1-\alpha)\lambda^2 p = 0$ gives the formula; it is positive and at most $\mu/(1+\lambda) \leq \mu$ since $D \geq \alpha(1+\lambda)^2$. (iii) The derivative of $(1+\lambda)/D(\lambda)$ has numerator $D - (1+\lambda)D' = -\alpha(1+\lambda)^2 - (1-\alpha)\lambda^2 - 2(1-\alpha)\lambda < 0$. □

For results that hold beyond the quadratic world cost we use a general response.

Assumption D.21 (Regular rational response)**.** $\lambda(\cdot) := \mathrm{BR}_W(\cdot) : \mathcal{P} \to [0,\bar{\Lambda}]$ satisfies $\lambda(p) > 0$ for $p > 0$ and $\lambda \in C^2$ on $(0,P]$, and the response $\Phi(p) := \lambda(p)p$ is nondecreasing. Its *elasticity* is $\varepsilon(p) := p\lambda'(p)/\lambda(p)$; $\Phi' = \lambda(1+\varepsilon) \geq 0$ is equivalent to $\varepsilon \geq -1$. The intensity λ itself may decrease, so $\varepsilon < 0$ is allowed. This is the same primitive as the scaled setting (Assumption E.11): rationality constrains the response Φ, not the intensity. The residual is $r(p) := p(1+\lambda(p)) - \mu = p - \mathbb{E}[y \mid p]$.

Proposition A.7 and Lemma D.20(i) give Assumption D.21 with $\varepsilon \equiv 1$. For $p < q$, $r(q) - r(p) = (q-p) + \Phi(q) - \Phi(p) \geq q - p$, so r is strictly increasing with $r(0^+) = -\mu$ (as λ is bounded) and $r(\mu) = \Phi(\mu) \geq 0$: it has a unique root $p^* \in (0,\mu]$, the honest forecast.

Definition D.22 (Timing structures)**.** *Simultaneous play (Nash)*: $p^N = p_{\lambda^N}$ and $\lambda^N = \lambda(p^N)$. *Learner leads*: $p^S \in \arg\min_p V(p)$ with $V(p) := U_A(p,\lambda(p))$, and $\lambda^S = \lambda(p^S)$. *World leads* (studied in a companion paper): $\lambda^W \in \arg\min_\lambda J(\lambda)$ with $J(\lambda) := U_W(p_\lambda,\lambda)$, and $p^W = p_{\lambda^W}$. A player is *rational* if it minimises its own objective under the timing structure; all three regimes are rational for both players.

Definition D.23 (Direct effect, wedge, conflict, manipulation)**.** For $p \in (0,P]$, the *direct effect* is $D_A(p) := \partial_p U_A(p,\lambda)\big|_{\lambda=\lambda(p)}$ and the *strategic response wedge* is $W_A(p) := \partial_\lambda U_A(p,\lambda)\big|_{\lambda=\lambda(p)}\lambda'(p)$, so $V' = D_A + W_A$. Nash solves $D_A = 0$; learner leadership solves $D_A + W_A = 0$. The wedge is defined everywhere and is in general nonzero at the Nash forecast. Symmetrically the world's wedge is $W_W(\lambda) := \partial_p U_W(p_\lambda,\lambda)\,p'_\lambda$. The learner has a *conflict of interest* at λ if $\arg\min_p U_A(p,\lambda) \neq \arg\min_p L(p,\lambda)$. *Manipulation* is response internalisation: locally the wedge, globally the *manipulation effect* $p^S - p^N$ (or $\lambda^W - \lambda^N$).

Definition D.24 (Harm)**.** The *harm* of an outcome (p,λ) is $L(p,\lambda)$; when $\lambda = \lambda(p)$ it equals $r(p)^2$, and it is 0 at the honest forecast. The *outcome distortion* is the shift of the world's action or of the mean outcome relative to the honest benchmark, $(\lambda,\ \mathbb{E}y)$ against $(\lambda(p^*),p^*)$. Along the flagged path honest → Nash → leader, $H_{\text{leader}} = [H_N - 0] + [H_{\text{leader}} - H_N]$: the first bracket is the conflict contribution, the second the manipulation contribution. The decomposition depends on the path; Nash is the no-commitment baseline.

Definition D.25 (Awareness and the regimes)**.** A player is *aware* if its model of the data accounts for the other player's internalisation of its behaviour; otherwise it interprets the data as generated by a myopic rational opponent. In regime 1 nobody manipulates, and the outcome is the honest

forecast (Appendices B.3 and B.5). In regime 2 the learner manipulates and the world is unaware: the learner leads (Appendix D.1, Appendix D.4.3). Regime 2 with an aware world is the signalling (signal-jamming) extension. Regimes in which the world manipulates, or both players do, are not treated in this paper.

D.4.2 Conflict and manipulation at the honest forecast

Proposition D.26 (A stake is exactly a conflict of interest)**.** *Under Assumption D.19, for every $\lambda > 0$ the learner has a conflict of interest at λ if and only if $\alpha < 1$. Under Assumption D.21, $D_A(p^*) = 2(1-\alpha)\lambda(p^*)^2 p^*$.*

Proof. By Lemma D.20(ii), $\arg\min_p U_A(p,\lambda) = \alpha(1+\lambda)\mu/D(\lambda)$ and $\arg\min_p L(p,\lambda) = \mu/(1+\lambda)$; they coincide iff $D(\lambda) = \alpha(1+\lambda)^2$, i.e. iff $(1-\alpha)\lambda^2 = 0$, i.e. iff $\alpha = 1$. For the second claim, $\partial_p U_A = 2\alpha(1+\lambda)r + 2(1-\alpha)\lambda^2 p$ and $r(p^*) = 0$. □

Proposition D.27 (The wedge is the elasticity times the direct effect)**.** *Under Assumptions D.19 and D.21, at the honest forecast*

$$W_A(p^*) = \varepsilon(p^*)\, D_A(p^*), \qquad V'(p^*) = \big(1+\varepsilon(p^*)\big)\, D_A(p^*) = 2(1-\alpha)\,\Phi(p^*)\Phi'(p^*),$$

The manipulation share of the local incentive is $\varepsilon/(1+\varepsilon)$, and $V'(p^*) > 0$ iff $\varepsilon(p^*) > -1$. In the quadratic world ($\varepsilon \equiv 1$) the split is exactly half and half: $W_A(p^*) = D_A(p^*) = 2(1-\alpha)\gamma^2 {p^*}^3$.

Proof. $\partial_\lambda U_A = 2\alpha p\, r + 2(1-\alpha)\lambda p^2$, equal to $2(1-\alpha)\lambda {p^*}^2$ at p^*. Multiplying by $\lambda'(p^*)$ and using $p^*\lambda'(p^*) = \varepsilon\lambda$ gives $W_A(p^*) = 2(1-\alpha)\lambda^2 p^*\,\varepsilon = \varepsilon D_A(p^*)$ (Proposition D.26). Hence $V' = (1+\varepsilon)D_A$; and $\Phi' = \lambda(1+\varepsilon)$ gives $2(1-\alpha)\Phi\Phi' = (1+\varepsilon)D_A$. As $D_A(p^*) > 0$ for $\alpha < 1$, the sign claim follows. In the quadratic world, $\lambda = \gamma p$, so $\varepsilon = 1$ and $\lambda(p^*) = \gamma p^*$. □

Remark D.28. Proposition D.27 recovers Theorem E.14(i) and explains the elasticity condition (Proposition B.3). A linear response (λ constant, $\varepsilon = 0$) carries no wedge: the whole incentive is conflict. Under the additive channel $y = \mu - \lambda + \eta$ the direct effect at the honest forecast is zero and the whole incentive is wedge, so the split is a property of the channel.

Example D.29 (Without $\Phi'(p^*) > 0$ there need be no shading)**.** Let $\Phi(p) = \min(p^2, 0.25)$, $\mu = 1$ and $\alpha = 0.9$. The honest forecast is $p^* = 0.75$, where Φ is flat, so $V'(p^*) = 2(1-\alpha)\Phi(p^*)\Phi'(p^*) = 0$. On $[0.5, 1]$, $V(p) = 0.9(p-0.75)^2 + 0.00625$, minimised at 0.75; on $[0, 0.5]$, $V \geq V(0.5) = 0.0625 > V(0.75)$. So $p^S = p^*$: a learner with a stake does not shade (T63).

D.4.3 The learner leads

Write $N(p) := \frac{1}{2}D_A(p) = \alpha(1+\lambda)r + (1-\alpha)\lambda^2 p$ and $S(p) := \frac{1}{2}V'(p) = \alpha(1+\lambda+\varepsilon\lambda)r + (1-\alpha)(1+\varepsilon)\lambda^2 p$, with λ, ε, r at p. By Lemma D.20(ii), p is a Nash forecast iff $N(p) = 0$.

Proposition D.30 (First order in the stake)**.** *Under Assumptions D.19 and D.21, with $p^* \in (0, P)$, $\varepsilon := \varepsilon(p^*) > -1$, $\lambda := \lambda(p^*)$, $R' := r'(p^*) = 1 + \lambda + \varepsilon\lambda$ and $s := 1 - \alpha$: there are $s_0 > 0$ and a neighbourhood $\mathcal{N}$ of p^* such that for $s \in [0, s_0)$ the equations $N = 0$ and $S = 0$ have unique solutions $p^N(s), p^S(s) \in \mathcal{N}$, C^1 in s, with*

$$p^N(s) = p^* - s\frac{\lambda^2 p^*}{(1+\lambda)R'} + o(s), \qquad p^S(s) = p^* - s\frac{(1+\varepsilon)\lambda^2 p^*}{R'^2} + o(s),$$

and $H_S/H_N \to \big[(1+\varepsilon)(1+\lambda)/(1+\lambda+\varepsilon\lambda)\big]^2$ as $s \downarrow 0$, a limit that exceeds 1 iff $\varepsilon > 0$.

Proof. $R' > (1+\lambda) - \lambda = 1$ since $\varepsilon > -1$. Also $U_A(p, \lambda(p)) = (1-s)r^2 + s\Phi^2$. *Nash.* $G_N(p,s) := 2(1-s)(1+\lambda(p))r(p) + 2s\lambda(p)^2 p$ is C^1 near $(p^*, 0)$ with $G_N(p^*, 0) = 0$, $\partial_p G_N(p^*, 0) = 2\lambda'(p^*)r(p^*) + 2(1+\lambda)R' = 2(1+\lambda)R' > 0$ and $\partial_s G_N(p^*, 0) = 2\lambda^2 p^*$. The implicit function theorem gives the unique C^1 branch with slope $-\lambda^2 p^*/((1+\lambda)R')$. *Leader.* $G_S(p,s) := (1-s)rr' + s\Phi\Phi'$ is C^1 near $(p^*, 0)$ since $\lambda \in C^2$; $G_S(p^*, 0) = 0$, $\partial_p G_S(p^*, 0) = R'^2 > 0$, $\partial_s G_S(p^*, 0) = \Phi\Phi' = \lambda^2 p^*(1+\varepsilon)$, giving slope $-(1+\varepsilon)\lambda^2 p^*/R'^2$. As $\partial_p G_S > 0$ near $(p^*, 0)$, V is strictly convex there and the root is its unique local minimiser in $\mathcal{N}$. *The local roots are the equilibria.* Every global minimiser p of $V_s := (1-s)r^2 + s\Phi^2$ on $\mathcal{P}$ satisfies

$(1-s)(p-p^*)^2 \le (1-s)r(p)^2 \le V_s(p) \le V_s(p^*) = s\Phi(p^*)^2$, using $|r(p)| \ge |p-p^*|$ (modulus one); so $|p-p^*| \le \sqrt{s/(1-s)}\,\Phi(p^*)$, and for s small the minimiser lies in $\mathcal{N}$, where it is the unique root $p^S(s)$. For Nash forecasts: there is $\delta > 0$ with $|r| \ge \delta$ on $\mathcal{P}\setminus\mathcal{N}$, and $N_s = (1-s)(1+\lambda)r + s\lambda^2 p$ with $|(1+\lambda)r| \ge |r|$ and $s\lambda^2 p \le s\bar{\Lambda}^2 P$; so for $s < \delta/(\delta + \bar{\Lambda}^2 P)$, $N_s \ne 0$ on $\mathcal{P}\setminus\mathcal{N}$, and the Nash forecast is the unique root $p^N(s)$. *Harm.* $r(p(s)) = R' c\, s + o(s)$ for either slope c, so $H = R'^2 c^2 s^2(1+o(1))$ and $H_S/H_N \to (c_S/c_N)^2$ with $c_S/c_N = (1+\varepsilon)(1+\lambda)/R' > 0$. Since $(1+\varepsilon)(1+\lambda) - R' = \varepsilon$, the ratio exceeds 1 iff $\varepsilon > 0$. □

Assumption D.31 (Single crossing, learner)**.** $\varepsilon > -1$ on $(0, P]$, and S has exactly one zero in $(0, p^*)$.

In the quadratic world, Assumption D.31 holds for all parameters: S is a positive multiple of the strategic cubic of Proposition D.3, which has exactly one positive root, lying in $(0, p^*)$ (Proposition D.3(ii),(iii)).

Lemma D.32 (Convex responses satisfy single crossing)**.** *Let Φ be C^1, convex and nondecreasing on $\mathcal{P}$, with $\Phi(0) = 0$ and $\Phi > 0$ on $(0, P]$, and let $\alpha < 1$. Then Assumption D.31 holds.*

Proof. Write $S = \frac{1}{2}V' = \alpha r + \Phi'\psi$ with $\psi := \alpha r + (1-\alpha)\Phi$. *Elasticity.* Since Φ is convex with $\Phi(0) = 0$, $\lambda = \Phi(p)/p$ is nondecreasing, so $\varepsilon \ge 0 > -1$. *Where $\psi < 0$.* There $\Phi'\psi \le 0$, so $S \le \alpha r$, and $\psi < 0$ gives $\alpha r < -(1-\alpha)\Phi \le 0$; hence $S < 0$. *Where $\psi \ge 0$.* Since $\psi' = \alpha + \Phi' > 0$, this set is an interval $[p_0, P]$. On it S is strictly increasing: αr is strictly increasing, and $\Phi'\psi$ is a product of nonnegative nondecreasing functions (Φ' by convexity). *Signs.* $S(0+) \le -\alpha\mu < 0$, and $S(p^*) = (1-\alpha)\Phi'(p^*)\Phi(p^*) > 0$, because convexity and $\Phi(0) = 0$ give $\Phi'(p^*) \ge \Phi(p^*)/p^* > 0$. So S has exactly one zero, and it lies in $(0, p^*)$. □

Example D.33 (Single crossing can fail for concave responses)**.** Let $\Phi(p) = 0.39770\big(1 - e^{-p/0.0025097}\big) + 0.44846\,p$, $\mu = 0.78919$ and $\alpha = 0.38295$. Then Φ is concave and increasing, and S has three zeros in $(0, p^*)$, near 0.0036, 0.0129 and 0.1155, with $p^* \approx 0.2703$, so Assumption D.31 fails. The sign conclusion of Theorem D.34 still holds here: the Nash forecast is $p^N \approx 0.0036$ with $\varepsilon(p^N) < 0$, the strategic forecast is $p^S \approx 0.1155 > p^N$, and $H^S < H^N$.

Theorem D.34 (Global sign theorem)**.** *Under Assumptions D.19, D.21 and D.31, with $\alpha < 1$: V has a unique minimiser p^S on $\mathcal{P}$; every Nash forecast lies in $(0, p^*)$; and for every Nash forecast*

$$\operatorname{sgn}(p^S - p^N) = -\operatorname{sgn}\varepsilon(p^N), \qquad \operatorname{sgn}(H_S - H_N) = \operatorname{sgn}\varepsilon(p^N).$$

In the quadratic world, for every $\mu, \gamma > 0$ and every stake, $p^S < p^N$ and $H_S > H_N$: learner leadership always adds harm.

Proof. Nash forecasts. For $p \ge p^*$, $r \ge 0$ and $N \ge (1-\alpha)\lambda^2 p > 0$. Near 0, $r < 0$ gives $\alpha(1+\lambda)r \le \alpha r$, and $(1-\alpha)\lambda^2 p \to 0$ because λ is bounded; so $\limsup_{p\downarrow 0} N \le -\alpha\mu < 0$. Hence all zeros of N lie in $(0, p^*)$, and one exists. *Leader.* Write $S = \alpha r + \Phi'\psi$ with $\psi := \alpha r + (1-\alpha)\Phi$. As $p \downarrow 0$, $r \to -\mu$ and $\Phi = \lambda p \to 0$, so $\psi < 0$ near 0; with $\Phi' \ge 0$ this gives $S \le \alpha r < 0$ near 0. Also $S(p^*) = (1-\alpha)(1+\varepsilon)\lambda^2 p^* > 0$; with one zero $p^S \in (0, p^*)$, $S < 0$ on $(0, p^S)$ and $S > 0$ on $(p^S, p^*]$. For $p > p^*$, $r > 0$, $r' \ge 1$ and $\Phi\Phi' > 0$, so $S > 0$. Hence p^S is the unique minimiser of V. *Key identity.* At a Nash forecast, $\alpha r = -(1-\alpha)\lambda^2 p/(1+\lambda)$, so

$$S(p^N) = (1-\alpha)\lambda^2 p\Big[(1+\varepsilon) - \frac{1+\lambda+\varepsilon\lambda}{1+\lambda}\Big] = (1-\alpha)\lambda^2 p\,\frac{\varepsilon}{1+\lambda},$$

which has the sign of $\varepsilon(p^N)$. By the sign pattern of S, $p^N > p^S$ iff $\varepsilon(p^N) > 0$, $p^N < p^S$ iff $\varepsilon(p^N) < 0$, and $p^N = p^S$ iff $\varepsilon(p^N) = 0$. *Harm.* On $(0, p^*)$, r is strictly increasing and negative, so $H = r^2$ is strictly decreasing in p. The quadratic world: $\varepsilon \equiv 1$ and Assumption D.31 holds, as noted above. □

Proposition D.30 is the limit of Theorem D.34 as the stake vanishes. In words: *manipulation* (as opposed to conflict of interest) harms users only when the intensity rises with the forecast; the total distortion relative to honesty, studied in Appendix D.1, is unaffected.

Remark D.35 (Added harm relative to simultaneous play). In the quadratic world with $\mu = 1$, Table 4 reports the harm that learner leadership adds over simultaneous play, $H^S - H^N$, and the learner's gain from leading, $V(p^N) - V(p^S)$ (T64). Both are positive, as Theorem D.34 requires, but both are nonmonotone in responsiveness and vanish as it grows, while the total harm $r(p^S)^2$ increases to $(1-\alpha)^2\mu^2$ (Theorem D.17). At high responsiveness almost all the harm is therefore conflict of interest, not manipulation.

Table 4: Added harm $H^S - H^N$ / learner's gain from leading $V(p^N) - V(p^S)$, quadratic world, $\mu = 1$ (T64).

α	$\gamma\mu = 0.1$	1	10	100	1000
0.2	0.0014 / 0.00009	0.0461 / 0.00209	0.0506 / 0.00118	0.0235 / 0.00022	0.0085 / 0.00003
0.5	0.0001 / 0.00002	0.0109 / 0.00109	0.0234 / 0.00104	0.0134 / 0.00022	0.0051 / 0.00003
0.8	0.0000 / 0.00000	0.0012 / 0.00018	0.0038 / 0.00024	0.0025 / 0.00005	0.0010 / 0.00001

D.5 When manipulation destroys value

Theorem D.36 (Exact incidence). *Let $\Phi \geq 0$ be nondecreasing and differentiable on $\mathcal{P}$, let p^* be the honest forecast and let $p^S \in (0, p^*)$ be an interior minimiser of $V = \alpha r^2 + (1-\alpha)\Phi^2$ with $\Phi'(p^S) > 0$. Write $u := |r(p^S)|$ (so the harm is u^2), $\delta := \Phi(p^*) - \Phi(p^S)$, $z := \delta/u \in [0, 1)$ and*

$$q := \frac{\Phi(p^*) - \Phi(p^S)}{\Phi'(p^S)\,(p^* - p^S)} \quad \textit{(secant slope over tangent slope).}$$

Then

$$r(p^S)^2 - \Delta_V = u^2\,(1-z)\,\big[\,1 + \alpha + (1-\alpha)z - 2\alpha q\,\big].$$

Hence manipulation destroys value if and only if $2\alpha q < 1 + \alpha + (1-\alpha)z$; in particular whenever $q \leq (1+\alpha)/(2\alpha)$, and so whenever Φ is concave on $[p^S, p^*]$ ($q \leq 1$), including every linear response.

Proof. **Ingredients.** $\mu = p^* + \Phi(p^*)$ gives $u = (p^* - p^S) + \delta$, so $z < 1$. $\Delta_V = (1-\alpha)\Phi(p^*)^2 - \alpha u^2 - (1-\alpha)\Phi(p^S)^2$, so $u^2 - \Delta_V = (1+\alpha)u^2 - (1-\alpha)\delta(2\Phi(p^S) + \delta)$. The first-order condition at the interior minimiser, $\alpha r(1+\Phi') + (1-\alpha)\Phi\Phi' = 0$, gives $(1-\alpha)\Phi(p^S) = \alpha u(1+x)/x$ with $x := \Phi'(p^S) > 0$.

Algebra. With $\delta = qx(p^* - p^S)$ and $u = (p^* - p^S) + \delta$, substitution gives the displayed factorisation (checked symbolically, script `t_circle_dwl2.py`); it uses the identity $1 + \alpha - 2\alpha z - (1-\alpha)z^2 = (1-z)(1+\alpha+(1-\alpha)z)$.

Consequences. $1 - z > 0$. If $q \leq (1+\alpha)/(2\alpha)$ the bracket is at least $(1-\alpha)z \geq 0$, and it is positive unless $z = 0$, which would force $\delta = 0$ and then $u = p^* - p^S > 0$ with $q = 0$, bracket $1 + \alpha > 0$. Concavity on $[p^S, p^*]$ gives $\delta \leq x(p^* - p^S)$, i.e. $q \leq 1 \leq (1+\alpha)/(2\alpha)$. □

Remark D.37 (Computer-assisted proofs). Several existence statements below are proved by exhibiting one instance and verifying it with interval arithmetic (outward rounding, 60 digits; `mpmath.iv`). Every quantity is computed as an interval guaranteed to contain the true real number, so each inequality verified on intervals holds exactly. Roots are enclosed by certified sign changes of strictly monotone functions; that an interval contains the *global* minimiser is certified by the sign of the derivative on the rest of the domain (adaptive bisection); every hand-coded derivative is checked against a high-precision finite difference. The certificates are the scripts `certify.py` and `certify_regimes.py` in the verification record.

Proposition D.38 (Manipulation can create value). *Let $\Phi(p) = \frac{1}{2}p + 0.4/(1 + e^{-200(p-0.625)})$, $\mu = 1$, $\alpha = \frac{1}{2}$ (a linear response plus a subpopulation that switches on sharply near the honest forecast). Then $p^* \in [0.617557, 0.617558]$, the global minimiser of V lies in $[0.591829, 0.591830]$, and gain $-$ harm $\geq 0.01046 > 0$: the learner gains more than users lose. By Theorem D.36 this requires $2\alpha q > 1 + \alpha + (1-\alpha)z$, so Φ is not concave on $[p^S, p^*]$.*

Proof. The certificate encloses p^* by a certified sign change of r (strictly increasing), certifies $V' < 0$ on $[0, a]$ and $V' > 0$ on $[b, p^*]$ for an interval $[a, b]$ of width 2×10^{-8} (and $V' > 0$ on $(p^*, P]$ holds because $r > 0$ and $\Phi' \geq 0$ there), so the global minimiser p^S lies in $[a, b]$. Then gain $-$ harm $= V(p^*) - V(p^S) - r(p^S)^2 \geq V(p^*) - V(x) - \sup_{[a,b]} r^2$ for any $x \in [a, b]$, since $V(p^S) \leq V(x)$; this lower bound is evaluated on intervals. □

Numerical illustration (the size of the effect). Numerically, $p^* = 0.6176$ and $p^S = 0.5918$; the harm to users is 0.01248 and the learner's gain is 0.02295, so the learner gains 1.84 times what users lose.

D.6 Where the conditions fail

The precise conditions of Theorems D.36, E.14 and F.20 fail on opposite sides of zero curvature, which gives a structural statement at no extra cost.

Corollary D.39 (Failures need opposite curvature)**.** *Let $\Phi \geq 0$ be nondecreasing and C^2 on $\mathcal{P}$, and let $p^S \in (0, p^*)$ be an interior strategic minimiser with $\Phi'(p^S) > 0$.*

(i) If manipulation does not destroy value (harm $\leq$ gain), then $\Phi''(x_1) > 0$ for some $x_1 \in [p^S, p^*]$.

(ii) If the pooled audit's tangent condition fails (the tangent condition (S_α) fails; in particular if an alternating strategic learner as in Theorem F.20(iii) exists), then $\Phi''(x_2) < 0$ for some $x_2 \in \mathcal{P}$.

(iii) In a scaled family $\Phi = \theta\varphi$, if either growth statement of Theorem E.14 fails, then $\varphi''(x_3) > 0$ for some $x_3 \in [p^S, p^*]$.

(iv) Consequently, if value destruction and the tangent condition both fail, Φ'' takes both signs on $\mathcal{P}$, and by the intermediate value theorem Φ'' vanishes at a point strictly between x_1 and x_2: the response has an inflection. A response that is convex on $\mathcal{P}$ can break only the growth statements of Theorem E.14 and value destruction; a response that is concave on $\mathcal{P}$ can break only the tangent condition.

Proof. For a C^2 function on an interval, concavity is equivalent to $\Phi'' \leq 0$ throughout and convexity to $\Phi'' \geq 0$ throughout.

(i) By Theorem D.36, concavity of Φ on $[p^S, p^*]$ implies harm $>$ gain. So if harm $\leq$ gain, Φ is not concave there, and $\Phi''(x_1) > 0$ for some x_1 in that interval.

(ii) By Theorem F.20(ii), convexity of Φ on $\mathcal{P}$ implies (S_α). Contrapositive.

(iii) By Theorem E.14(iv), concavity of φ on $[p^S, p^*]$ implies both growth statements. Contrapositive.

(iv) Combine (i) and (ii); Φ'' is continuous, so it has a zero between a point where it is positive and a point where it is negative. The last sentence restates (i)–(iii) for convex and concave Φ. □

Proposition D.40 (Both failure types occur)**.** *Take $\mu = 1$, $P = 3$.*

(i) A concave response violates the tangent condition. For $\Phi(p) = 1 - e^{-p/0.01}$ and $\alpha = \frac{1}{20}$, the tangent condition fails at $\hat{p}_1 = 2.9$: $W(\rho_1) - W(0) - W_1\rho_1 \leq -3.62$. Value destruction holds by Theorem D.36 (Φ is concave).

(ii) A convex response lets manipulation create value. For $\Phi(p) = 0.3134\,p + \sum_{i=1}^{2} w_i\tau_i\big[\log(1+e^{(p-c_i)/\tau_i}) - \log(1+e^{-c_i/\tau_i})\big]$ with $(w_i, c_i, \tau_i) = (0.6781, 0.9366, 0.0751)$ and $(0.5114, 0.7479, 0.00195)$, and $\alpha = 0.7846$: gain $-$ harm $\geq 1.52 \times 10^{-4} > 0$, with $p^* \in [0.755203, 0.755204]$ and the global minimiser of V in $[0.743823, 0.743824]$. Φ is convex (a linear term plus positive multiples of softplus functions), so the tangent condition holds by Theorem F.20(ii).

Proof. (i) is certified as in Proposition F.21; (ii) as in Proposition D.38. □

D.7 Near an activation threshold

The failures concentrate where a subpopulation switches on. The smoothing family of Proposition C.12 makes this precise: as the switch sharpens, the strategic forecast converges to one of three regimes, and each regime has a closed-form verdict on value destruction.

Theorem D.41 (The smoothing limit)**.** *Take the discrete-type model with a gap at x_0 (segments $G_k < G_{k+1}$, $r^- = x_0 + G_k x_0^2 - \mu < 0 < r^+$), and its smoothing $\Phi_h(p) = p^2[G_k + \Delta G\,\Psi((p - x_0)/h)]$, $\Delta G := G_{k+1} - G_k$, with Ψ as in Proposition C.12. Write $u := -r^-$, $c := G_k x_0^2$, $r_L(p) := p + G_k p^2 - \mu$, and*

$$U(p,s) := \alpha\,(r_L(p) + s)^2 + (1-\alpha)\,(G_k p^2 + s)^2,$$
$$C_0 := \{(p,0) : 0 \le p \le x_0\} \cup \{(x_0,s) : 0 \le s \le u\}.$$

Let $p^S(h)$ be a strategic minimiser for Φ_h and $s_h := \Phi_h(p^S(h)) - G_k p^S(h)^2$. As $h \to 0$:

(i) $\min V_h \to m_0 := \min_{C_0} U$*, and every limit point of $(p^S(h), s_h)$ minimises U on C_0;*

(ii) if the minimiser (p^0, s^0) is unique, then harm $\to (r_L(p^0)+s^0)^2$ and gain $\to (1-\alpha)(c+u)^2 - m_0$;

(iii) ***the three regimes.*** *Put $\nu := \alpha u - (1-\alpha)c$.*

- *(I) $p^0 < x_0$, $s^0 = 0$: the learner stays on the lower branch, away from the threshold; harm $-$ gain $\to (1+\alpha) r_L(p^0)^2 + (1-\alpha) G_k^2 (p^0)^4 - (1-\alpha)(c+u)^2$.*
- *(II) $(p^0, s^0) = (x_0, 0)$: the learner sits at the foot of the switch. This requires $\nu \le 0$, and then harm $-$ gain $\to 2u\nu \le 0$: manipulation creates value in the limit (strictly if $\nu < 0$).*
- *(III) $p^0 = x_0$, $s^0 = \nu \in (0, u)$: the learner sits inside the switch, partially activating it. This requires $\nu > 0$, and then harm $-$ gain $\to 0$: value destruction is decided at the next order, by the sign test of Theorem D.43.*

Proof. **Reparametrisation.** Put $s_h(p) := p^2 \Delta G\,\Psi((p - x_0)/h) \ge 0$. Then $\Phi_h(p) = G_k p^2 + s_h(p)$, $r_h(p) = r_L(p) + s_h(p)$, and $V_h(p) = U(p, s_h(p))$ exactly, with U independent of h and Lipschitz on compact sets.

Feasible pairs. By Theorem E.14(i) the strategic minimiser lies in $[0, {p^*}_h]$, and ${p^*}_h \to x_0$ (Proposition C.12). For $p \le {p^*}_h$, $r_h(p) \le 0$, so $0 \le s_h(p) \le -r_L(p)$. For $\delta > 0$: (a) if $p \le x_0 - \delta$ then $s_h(p) \le P^2 \Delta G\,\Psi(-\delta/h) =: \varepsilon_h(\delta) \to 0$; (b) if $p \in [x_0 - \delta, {p^*}_h]$ then $0 \le s_h(p) \le u + \omega(\delta)$, where $\omega(\delta) := \max_{|p - x_0| \le \delta} |r_L(p) - r_L(x_0)| \to 0$.

Upper bound. For $p < x_0$ fixed, $s_h(p) \to 0$, so $V_h(p) \to U(p, 0)$. For $s \in (0, u)$, continuity of $\xi \mapsto s_h(x_0 + h\xi)$, which runs from near 0 to above s for small h, gives ξ_h with $s_h(x_0 + h\xi_h) = s$ and $x_0 + h\xi_h \le {p^*}_h$; then $V_h(x_0 + h\xi_h) = U(x_0 + h\xi_h, s) \to U(x_0, s)$. Hence $\limsup_h \min V_h \le \inf_{C_0} U = m_0$ (U is continuous and C_0 is compact).

Lower bound. Let L_U be a Lipschitz constant of U on $[0, P] \times [0, u + 1]$. In case (a), $V_h(p) \ge U(p, 0) - L_U \varepsilon_h(\delta) \ge m_0 - L_U \varepsilon_h(\delta)$. In case (b), $V_h(p) \ge U(x_0, \min(s_h(p), u)) - L_U(\delta + |{p^*}_h - x_0| + \omega(\delta)) \ge m_0 - L_U(\delta + |{p^*}_h - x_0| + \omega(\delta))$. Let $h \to 0$, then $\delta \to 0$: $\liminf_h \min V_h \ge m_0$. This proves the first part of (i).

Limit points. If $p^S(h_n) \le x_0 - \delta$ along a subsequence then $s_{h_n} \to 0$; otherwise $p^S(h_n) \to x_0$ and s_{h_n} has limit points in $[0, u]$ by (b). So every limit point of $(p^S(h), s_h)$ lies in C_0, and by continuity U equals m_0 there.

(ii) Harm is $r_h(p^S)^2 = (r_L(p^S) + s_h)^2$, a continuous function of the pair. The honest value is $V_h({p^*}_h) = (1-\alpha)\Phi_h({p^*}_h)^2 = (1-\alpha)(\mu - {p^*}_h)^2 \to (1-\alpha)(\mu - x_0)^2$, and $\mu - x_0 = c + u$ because $r_L(x_0) = r^- = -u$.

(iii) On the vertical part of C_0, $r_L(x_0) = -u$, so $U(x_0, s) = \alpha(s - u)^2 + (1-\alpha)(c + s)^2 = (s - \nu)^2 + U(x_0, \nu)$, since the coefficients of s^2 sum to 1 and the minimiser is $s = \alpha u - (1-\alpha)c = \nu$. (A) is (ii) with $s^0 = 0$. (B) If $(x_0, 0)$ is the minimiser then $\partial_s U(x_0, 0) \ge 0$, i.e. $\nu \le 0$; harm u^2, gain $(1-\alpha)(c+u)^2 - \alpha u^2 - (1-\alpha)c^2$, and their difference is $(1+\alpha)u^2 - (1-\alpha)(2cu + u^2) = 2u\nu$. (C) An interior minimiser on the segment is $s = \nu$, which requires $0 < \nu < u$; then harm $(\nu - u)^2$ and gain $U(x_0, u) - U(x_0, \nu) = (u - \nu)^2$ coincide. □

Corollary D.42 (All three regimes occur, and regime III's verdict is not universal)**.** *Take the instance* $\mu = \frac{3}{2}$, $G_k = \frac{1}{2}$, $G_{k+1} = 2$, $x_0 = \frac{4}{5}$ *(so* $u = \frac{19}{50}$, $c = \frac{8}{25}$*) with Gaussian* Ψ*. In the limit* $h \to 0$ *of Theorem D.41:*

(i) at $\alpha = \frac{1}{5}$ *the limit problem has a unique minimiser* $(p^0, 0)$ *with* $p^0 \in (0.746055, 0.746056)$ *(regime A), and harm* $-$ *gain converges to a limit at most* -0.0585*;*

(ii) at $\alpha = \frac{3}{10}$ *and* $\frac{9}{20}$ *the unique minimiser is* $(x_0, 0)$ *(regime B), and harm* $-$ *gain* $\to 2u\nu = -\frac{209}{2500}$ *and* $-\frac{19}{5000}$*;*

(iii) at $\alpha \in \{\frac{1}{2}, \frac{3}{5}, \frac{7}{10}, \frac{4}{5}, \frac{9}{10}\}$ *the unique minimiser is* (x_0, ν) *(regime C), and the coefficient* $-2(1-\alpha)(u-\nu)D$ *of Theorem D.43 is* -0.694, 0.0146, 0.0306, 0.0121 *and* 0.00176 *(to the digits shown).*

So in this instance manipulation creates value in regimes I and II, creates value at $\alpha = \frac{1}{2}$ *and destroys it at the other four values of* α *for small* h*: in regime III the verdict on value destruction is decided by the sign of* D*, and both signs occur.*

Proof. The limit problem minimises U over C_0. On the second segment $U(x_0, s) = \alpha(s-u)^2 + (1-\alpha)(c+s)^2$, minimised at $s = \nu = \alpha u - (1-\alpha)c$ (exact rational). (i) $\nu < 0$, so the second segment's minimum is its corner; the certificate encloses the zero of $\partial_p U(\cdot, 0)$ and certifies the derivative's sign on the rest of $[0, x_0]$, so the first segment has a unique minimiser $p^0 < x_0$ below the corner value; the regime-A limit formula is then evaluated on intervals. (ii) $\nu < 0$, and the certificate shows $\partial_p U(\cdot, 0) < 0$ on $[0, x_0]$, so the unique minimiser is the corner; $2u\nu$ is exact. (iii) $0 < \nu < u$, and the certificate shows $U(p, 0) > U(x_0, \nu)$ on $[0, x_0]$; ξ_0 and ξ^* are enclosed by certified sign changes of the normal distribution function (computed from the alternating Taylor series of erf with its rigorous remainder), and D and the coefficient are evaluated on intervals. □

Theorem D.43 (Regime III at the next order)**.** *In Theorem D.41, suppose the limit minimiser is* (x_0, ν) *with* $0 < \nu < u$ *(regime C) and is unique, and* $\Psi \in C^2$ *with a positive density. Put* $S(\xi) := x_0^2 \Delta G\, \Psi(\xi)$, $\xi_0 := S^{-1}(\nu)$, $\xi^* := S^{-1}(u)$ *and*

$$D := \frac{c+\nu}{S'(\xi_0)} - (c+u)\,\frac{\xi^* - \xi_0}{u-\nu}.$$

Then harm $-$ *gain* $= -2h(1-\alpha)(u-\nu)\,D + O(h^2)$*. So, for small* h*, manipulation destroys value if* $D < 0$ *and creates value if* $D > 0$*. In words: compare the inverse slope of the switch at the strategic point, weighted by the response there* $(c+\nu)$*, with its average inverse slope up to the honest forecast, weighted by the honest response* $(c+u)$*.*

Proof. **Expansion in the switch.** Write $p = x_0 + h\xi$. Exactly, $r_h(p) = r_L(x_0 + h\xi) + S(\xi)(1 + h\xi/x_0)^2$ and $\Phi_h(p) = G_k(x_0+h\xi)^2 + S(\xi)(1+h\xi/x_0)^2$. So with $A := S(\xi)$, $r = A - u + h\xi a_1(A) + O(h^2)$ and $\Phi = c + A + h\xi a_2(A) + O(h^2)$, where $a_2(A) := 2G_k x_0 + 2A/x_0$ and $a_1 = a_2 + 1$. Then $V_h(x_0+h\xi) = F(A) + h\xi G(A) + O(h^2)$ with $F(A) = \alpha(A-u)^2 + (1-\alpha)(c+A)^2 = (A-\nu)^2 + F(\nu)$ and $G(A) = 2\alpha(A-u)a_1 + 2(1-\alpha)(c+A)a_2$. The function $f_h(\xi) := V_h(x_0 + h\xi)$ is smooth in (ξ, h), and $f_0''(\xi_0) = 2S'(\xi_0)^2 > 0$.

Strategic point. By the implicit function theorem the local minimiser is $\xi_h = \xi_0 + O(h)$, and by Theorem D.41 it is the global one for small h. From $F'(\nu) = 0$, $\alpha(\nu - u) = -(1-\alpha)(c+\nu)$, so $G(\nu) = -2(1-\alpha)(c+\nu)$ and $G'(\nu) = 2(a_2(\nu) + \alpha)$. The envelope theorem gives $\min V_h = F(\nu) + h\xi_0 G(\nu) + O(h^2)$. Expanding the first-order condition, $A_h - \nu = -\frac{h}{2}[G(\nu)/S'(\xi_0) + \xi_0 G'(\nu)] + O(h^2)$, hence $r_h(p^S) = \nu - u + h(1-\alpha)[\xi_0 + (c+\nu)/S'(\xi_0)] + O(h^2)$.

Honest forecast. $V_h(p^*{}_h) = (1-\alpha)(\mu - p^*{}_h)^2 = (1-\alpha)(c+u-h\xi^*)^2 + O(h^2)$, and $(1-\alpha)(c+u)^2 = F(u) = (u-\nu)^2 + F(\nu)$.

Difference. Harm is $(u-\nu)^2 - 2h(u-\nu)(1-\alpha)[\xi_0 + (c+\nu)/S'(\xi_0)]$ and gain is $(u-\nu)^2 - 2h(1-\alpha)[(c+u)\xi^* - (c+\nu)\xi_0]$, both up to $O(h^2)$. Their difference is $2h(1-\alpha)[(c+u)(\xi^* - \xi_0) - (u-\nu)(c+\nu)/S'(\xi_0)] = -2h(1-\alpha)(u-\nu)D$. □

Remark D.44. Theorem D.41 settles what an argument based on q cannot: as the switch sharpens, q stays bounded in regime I (the secant over $[p^S, p^*]$ and the tangent at p^S both have finite limits when $G_k > 0$), while the direct limits of harm and gain settle every regime except C. Together with Corollary D.39, the picture is: value destruction and the tangent condition of the pooled audit fail where the response has an inflection, as it does where a subpopulation switches on; and in the sharp-switch limit that also governs the gap regime of Appendix C, value destruction has an exact limiting verdict in two of three regimes.

E LEVEL VERSUS SLOPE: PROOFS

E.1 THE QUADRATIC WORLD: IDENTIFICATION AND THE COST OF LEARNING TO LEAD

Theorem E.1 (Identification fails at the strategic forecast)**.** *Fix $\hat{p} > 0$. There exist two worlds $(\mu_1, \gamma_1) \neq (\mu_2, \gamma_2)$ inducing the same distribution of y at $\hat{p}$, with $\hat{p}$ the strategic optimum in world 1 but not in world 2. By contrast, whether $\hat{p} = p^*$ is determined by the mean of y at $\hat{p}$.*

Proof. Pick $\mu_1, \gamma_1 > 0$, let $\hat{p} := p^S(\alpha; \mu_1, \gamma_1)$, $c_0 := \mu_1 - \gamma_1 \hat{p}^2$ and $W := \{(\mu, \gamma) : \mu, \gamma > 0,\ \mu - \gamma\hat{p}^2 = c_0\}$. Every member of W gives $y = c_0 + \eta$ at $\hat{p}$, and the admissible γ form the interval $\gamma > \max(0, -c_0/\hat{p}^2)$. Along W, $r(\hat{p}) = \hat{p} - c_0$ is fixed and $V'(\hat{p}) = 2\alpha(\hat{p} - c_0)(1 + 2\gamma\hat{p}) + 4(1-\alpha)\gamma^2\hat{p}^3 =: Q(\gamma)$, a polynomial of degree 2 in γ because its leading coefficient $4(1-\alpha)\hat{p}^3$ is nonzero. So Q has at most two zeros, one of which is γ_1. The interval of admissible γ is infinite, so there is an admissible γ_2 with $Q(\gamma_2) \neq 0$: $\hat{p}$ is not a critical point of V in world 2. In the honest case the mean of y at $\hat{p}$ identifies c_0, hence $r(\hat{p}) = \hat{p} - c_0$, and $r(\hat{p}) = 0$ iff $\hat{p} = p^*$; this uses only the mean, so it holds whether or not σ^2 is known. □

Theorem E.2 (Learning to manipulate: upper bound)**.** *Let $\theta := (\mu, \gamma)$ lie in a known rectangle $\Theta := [\mu_{lo}, \mu_{hi}] \times [\gamma_{lo}, \gamma_{hi}] \subset (0, \infty)^2$, $\mathcal{P} = [0, P]$ with $P \geq \mu_{hi}$, and noise i.i.d. with mean 0 and variance σ^2. Define $S_T := \sum_t \mathbb{E}V_\theta(\hat{p}_t) - T\,V_\theta(p^S(\theta))$. The explore-then-commit algorithm ETC-S(n, P_0) (n even, $P_0 \in (0, P]$) plays 0 for $n/2$ rounds (mean $\hat{\mu}$), P_0 for $n/2$ rounds (mean $\bar{y}_2$), sets $\hat{\gamma} := (\hat{\mu} - \bar{y}_2)/P_0^2$ and $\hat{\theta} := \Pi_\Theta(\hat{\mu}, \hat{\gamma})$, and plays $p^S(\hat{\theta})$ thereafter. Then $S_T \leq nB_V + TC_S/n$ for every even n with $2 \leq n \leq T$, and with $n := \min(T, 2\lceil \frac{1}{2}\sqrt{TC_S/B_V} \rceil)$,*

$$S_T \leq \min\left(T\,B_V,\ 2\sqrt{B_V C_S T} + 2B_V\right).$$

Proof. **(a)** c is polynomial in (p, μ, γ) with $\partial_p c > 0$ at the root for every $\theta \in \Theta$ (Proposition D.3(iv)); by the implicit function theorem $\theta \mapsto p^S(\theta)$ is C^1 near Θ, its gradient is bounded by some Λ_s on the compact Θ, and Θ is convex, so p^S is Λ_s-Lipschitz (mean value theorem).

(b) For $p \in \mathcal{P}$, since $V'_\theta(p^S) = 0$, Taylor's theorem gives $V_\theta(p) - V_\theta(p^S) \leq \frac{1}{2}M_V(p - p^S)^2$ with $M_V := \sup_{\Theta \times K} |V''|$.

(c) $\mathbb{E}\hat{\gamma} = \gamma$; the blocks are independent, $\mathrm{Var}\,\hat{\mu} = 2\sigma^2/n$, $\mathrm{Var}\,\hat{\gamma} = 4\sigma^2/(nP_0^4)$; projection onto the convex $\Theta \ni \theta$ is nonexpansive, so $\mathbb{E}|\hat{\theta} - \theta|^2 \leq C_\theta/n$ with $C_\theta := 2\sigma^2 + 4\sigma^2/P_0^4$.

(d) Every round costs at most $B_V := \sup_{\Theta \times K}[V_\theta(p) - V_\theta(p^S(\theta))] < \infty$. Commit rounds cost at most $\frac{1}{2}M_V\Lambda_s^2 C_\theta/n =: C_S/n$ each, so $S_T \leq nB_V + TC_S/n$; with $n \leq \sqrt{TC_S/B_V} + 2$, $nB_V \leq \sqrt{TC_S B_V} + 2B_V$ and $TC_S/n \leq \sqrt{TC_S B_V}$. If exploration would use all T rounds, $S_T \leq TB_V$ trivially. □

Theorem E.3 (Learning to manipulate costs $\Omega(\sqrt{T})$)**.** *Fix $\alpha \in (0, 1)$, $\sigma > 0$, $\mathcal{P} = [0, P]$ and $\theta_0 = (\mu_0, \gamma_0)$ with $\mu_0, \gamma_0 > 0$ and $p^S(\theta_0) \in (0, P)$. In every neighbourhood of θ_0 there is a compact curve $\Theta_0 \ni \theta_0$ and constants $c, C > 0$ such that every (possibly randomised) algorithm, with Gaussian noise, satisfies*

$$\sup_{\theta \in \Theta_0} S_T(\theta) \geq c\sqrt{T} - C \qquad \text{for all } T.$$

With Theorem E.2, the minimax strategic regret with μ unknown is $\Theta(\sqrt{T})$.

Lemma E.4 (Quadratic growth). *For compact Θ_0, $c_V := \inf\{(V_\theta(p) - V_\theta(p^S(\theta)))/(p - p^S(\theta))^2 : \theta \in \Theta_0,\ p \in \mathcal{P},\ p \neq p^S(\theta)\} > 0$.*

Proof. Extend the ratio by $V''_\theta(p^S)/2$ on the diagonal. The extension is jointly continuous (Taylor with continuous second derivative, and continuity of p^S in θ) and positive everywhere: $V''(p^S) = 2\partial_p c(p^S) > 0$ and, off the diagonal, p^S is the unique minimiser. A positive continuous function on a compact set has a positive minimum. □

Lemma E.5 (The uninformative curve). *Let $\mu(\vartheta)$ solve $\mu'(\vartheta) = p^S(\mu(\vartheta), \vartheta)^2$, $\mu(\gamma_0) = \mu_0$, on $I := [\gamma_0 - w/2, \gamma_0 + w/2]$; the right-hand side is C^1, so for small w a unique solution exists and stays in the neighbourhood (Picard–Lindelöf). Set $\theta(\vartheta) := (\mu(\vartheta), \vartheta)$, $m_\vartheta(p) := \mu(\vartheta) - \vartheta p^2$, $\psi(\vartheta) := p^S(\theta(\vartheta))$. Then (E1) $\partial_\vartheta m_\vartheta(p) = \psi(\vartheta)^2 - p^2$, so $|\partial_\vartheta m_\vartheta(p)| \le 2P|p - \psi(\vartheta)|$ on $\mathcal{P}$; (E2) $\psi' < 0$ and $|\psi'| \ge \psi_{\min} > 0$ on I.*

Proof. (E1) is direct. For (E2), $\psi' = -D/\partial_p c$, where D is the derivative of c along $(\mu', \gamma') = (p^2, 1)$: $D = \partial_\mu c \cdot p^2 + \partial_\gamma c = 2p\big[(2-\alpha)\gamma p^2 + \alpha p - \alpha\mu\big]$. At the root, $(2-\alpha)\gamma p^2 + \alpha p - \alpha\mu = \alpha r(p) + 2(1-\alpha)\gamma p^2$, and the first-order condition gives $\alpha r = -2(1-\alpha)\gamma^2 p^3/(1+2\gamma p)$; so the bracket equals $2(1-\alpha)\gamma p^2(1+\gamma p)/(1+2\gamma p) > 0$, $D > 0$, and $\psi' < 0$. ψ' is continuous on the compact I. □

Lemma E.6 (Functional van Trees inequality for sequential designs). *Let π be the raised-cosine prior on I, $I(\pi) = 4\pi^2/w^2$, and let $\hat\psi_t$ be any function of $(U, y_1, \dots, y_{t-1})$. Then*

$$\int \mathbb{E}_\vartheta\big(\hat\psi_t - \psi(\vartheta)\big)^2 \pi(\vartheta)\, d\vartheta \ \ge\ \frac{\big(\int \psi'\pi\big)^2}{\int I_t(\vartheta)\pi(\vartheta)\, d\vartheta + I(\pi)}, \qquad I_t(\vartheta) := \sigma^{-2}\sum_{s<t}\mathbb{E}_\vartheta\big[(\partial_\vartheta m_\vartheta(\hat p_s))^2\big].$$

Proof. Fix $U = u$. The data have density $f(y \mid \vartheta) = \prod_{s<t}\phi_\sigma(y_s - m_\vartheta(\hat p_s))$, where $\hat p_s$ depends on $(u, y_{1:s-1})$; it is differentiable in ϑ with $\int \partial_\vartheta f = 0$. As in Lemma B.28, $\int(\hat\psi - \psi(\vartheta))\partial_\vartheta(f\pi)d\vartheta = \int \psi' f\pi\, d\vartheta$ (the boundary term vanishes); integrating over y gives $\mathbb{E}_\pi\psi'$, and Cauchy–Schwarz gives the bound with the Fisher information $\mathbb{E}(\text{score}^2)$. The score $\sum_s (y_s - m_\vartheta(\hat p_s))\partial_\vartheta m_\vartheta(\hat p_s)/\sigma^2$ is a sum of martingale differences, so $\mathbb{E}(\text{score}^2) = \sigma^{-2}\sum_s \mathbb{E}(\partial_\vartheta m)^2$. Finally, $x \mapsto 1/(x + I(\pi))$ is convex, so Jensen's inequality lets the per-u bounds average into the stated bound. □

Proof of Theorem E.3. Let $B_t := \int \mathbb{E}_\vartheta(\hat p_t - \psi(\vartheta))^2\pi\, d\vartheta$ and $A_t := \sum_{s\le t} B_s$. Lemma E.6 with $\hat\psi_t = \hat p_t$, (E1) and (E2) give $B_t \ge c_1/(c_2 A_{t-1} + c_3)$ with $c_1 := \psi_{\min}^2$, $c_2 := 4P^2/\sigma^2$, $c_3 := I(\pi)$. Since A is increasing, $B_t \ge c_1/(c_2 A_T + c_3)$ for $t \le T$; summing, $c_2 A_T^2 + c_3 A_T \ge c_1 T$, hence $A_T \ge (-c_3 + \sqrt{c_3^2 + 4c_1c_2T})/(2c_2) \ge \sqrt{c_1T/c_2} - c_3/(2c_2)$. By Lemma E.4, $S_T(\theta) \ge c_V\sum_t \mathbb{E}_\theta(\hat p_t - \psi)^2$, so $\sup_{\Theta_0} S_T \ge \int S_T\pi \ge c_V A_T \ge c_V\psi_{\min}\sigma/(2P)\sqrt{T} - C$. □

Remark E.7 (Mechanism, and the case of known μ). Along the curve of Lemma E.5, the observation at the strategic forecast is uninformative while the target moves, so information about the target can only be bought by playing away from it, which is exactly what the regret charges for. The lower-bound technique is that of Broder & Rusmevichientong (2012); the same level-versus-slope dichotomy appears in Keskin & Zeevi (2014). We prove it in this model. With μ known the curve cannot move μ, $\partial_\vartheta m = -p^2 \neq 0$ near the target, and the argument gives no $\sqrt{T}$ bound. With μ known the rate is $\Theta(\log T)$ (Theorem E.8).

Theorem E.8 (With μ known, learning to manipulate costs $\Theta(\log T)$). *Let μ be known, $\gamma \in [\gamma_{lo}, \gamma_{hi}] \subset (0, \infty)$ unknown, $\mathcal{P} = [0, P]$ with $P \ge \mu$, and Gaussian noise.*

(i) ***Upper bound.*** *The greedy learner plays $\hat p_1 := P_0 \in (0, P]$, then $\hat p_t := p^S(\mu, \hat\gamma_t)$ with $\hat\gamma_t := \Pi_{[\gamma_{lo},\gamma_{hi}]}\big(\sum_{s<t}(\mu - y_s)\hat p_s^2/\sum_{s<t}\hat p_s^4\big)$. It satisfies $S_T \le B_V + C(1 + \ln T)$.*

(ii) ***Lower bound.*** *For every γ_0 in the interior of $[\gamma_{lo}, \gamma_{hi}]$ and every algorithm, $\sup_\gamma S_T \ge c\ln T - C$ over a small interval around γ_0.*

Lemma E.9 (The target moves with γ). *At the root, $\partial_\gamma c = p^S\,(4\gamma {p^S}^2 + 3\alpha p^S - 2\alpha\mu) > 0$; hence $\partial_\gamma p^S = -\partial_\gamma c/\partial_p c < 0$.*

Proof. Differentiating Proposition D.3(i), $\partial_\gamma c = 4\gamma p^3 + 3\alpha p^2 - 2\alpha\mu p$. The first-order condition $\alpha r(1+2\gamma p) = -2(1-\alpha)\gamma^2 p^3$ with $r = p+\gamma p^2-\mu$ gives $\alpha\mu = \alpha(p+\gamma p^2)+2(1-\alpha)\gamma^2 p^3/(1+2\gamma p)$. Substituting, $4\gamma p^2 + 3\alpha p - 2\alpha\mu = \alpha p + \gamma p^2\big[(4-2\alpha) - 4(1-\alpha)\gamma p/(1+2\gamma p)\big] \geq \alpha p + 2\gamma p^2 > 0$, because $\gamma p/(1+2\gamma p) < \frac{1}{2}$. (Also checked on a grid, T26.) □

Proof of Theorem E.8. **(i) Information never dries up.** Every play satisfies $\hat{p}_t \geq p_{\min} := \min\big(P_0, \min_{\gamma\in[\gamma_{lo},\gamma_{hi}]} p^S(\mu,\gamma)\big) > 0$ (a positive continuous function on a compact interval).

(i) Estimation error. Since $\mu - y_s = \gamma\hat{p}_s^2 - \eta_s$, the unprojected estimate equals $\gamma - M_t/A_t$ with $M_t := \sum_{s<t}\eta_s\hat{p}_s^2$ (a martingale, since $\hat{p}_s$ is predictable) and $A_t := \sum_{s<t}\hat{p}_s^4 \geq (t-1)p_{\min}^4$. Projection onto an interval containing γ is nonexpansive, so $\mathbb{E}(\hat{\gamma}_t - \gamma)^2 \leq \mathbb{E}M_t^2/((t-1)^2 p_{\min}^8) = \sigma^2\mathbb{E}\sum_{s<t}\hat{p}_s^4/((t-1)^2 p_{\min}^8) \leq \sigma^2 P^4/((t-1)p_{\min}^8)$.

(i) Regret. $p^S(\mu,\cdot)$ is Λ-Lipschitz on the interval (implicit function theorem and compactness), and $V - V(p^S) \leq \frac{1}{2}M_V(p-p^S)^2$ (proof of Theorem E.2). So $\mathbb{E}[V(\hat{p}_t) - V(p^S)] \leq \frac{1}{2}M_V\Lambda^2\sigma^2P^4 p_{\min}^{-8}/(t-1)$ for $t \geq 2$; round 1 costs at most B_V; sum over t.

(ii) Apply Lemma E.6 with $\vartheta = \gamma$ on an interval I around γ_0 and μ fixed, so $m_\vartheta(p) = \mu - \vartheta p^2$ and $(\partial_\vartheta m)^2 = p^4 \leq P^4$; and $\psi = p^S(\mu,\cdot)$ has $|\psi'| \geq \psi_{\min} > 0$ on I by Lemma E.9. So $B_t \geq \psi_{\min}^2/((t-1)P^4/\sigma^2 + I(\pi))$; summing as in Theorem B.29, $\sum_{t\leq T} B_t \geq (\psi_{\min}^2\sigma^2/P^4)(\ln T - C')$, and Lemma E.4 gives $\sup S_T \geq c_V\sum_t B_t$. □

Remark E.10. Together with Theorem E.3, knowing μ is worth exactly the difference between $\Theta(\sqrt{T})$ and $\Theta(\log T)$: this is the in-model version of the Keskin–Zeevi dichotomy (unknown demand versus known demand at an incumbent price). Experiment E11 is consistent with it.

E.2 Scaled rational worlds

This subsection asks which conclusions about the quadratic world follow from rationality alone, for a family of scaled responses. For every scaled rational world, the honest forecast is learnable at logarithmic rate because it certifies itself (Theorem E.13); every learner with a stake has an incentive to manipulate, and in the cases stated the incentive grows with responsiveness (Theorem E.14); and the strategic forecast must be learned by moving away from it, which costs $\Theta(\sqrt{T})$ when μ is unknown (Theorem E.16). Value destruction and the discipline of the pooled audit hold under precise conditions (Theorems D.36 and F.20) and can fail; every failure found arises where the response switches on sharply between the strategic and the honest forecast, the structure that in Appendix C removes the calibrated forecast itself.

Assumption E.11 (The scaled setting)**.** $\mathcal{P} = [0,P]$. The response is $\Phi_\theta(p) = \theta\varphi(p)$ with $\varphi \in C^2(\mathcal{P})$, $\varphi(0) = 0$ and $\varphi' > 0$ on $(0,P]$, so $\varphi > 0$ on $(0,P]$. The parameters (μ,θ) lie in a box $B = [\mu_{lo},\mu_{hi}] \times [\theta_{lo},\theta_{hi}] \subset (0,\infty)^2$ with $P \geq \mu_{hi}$; noise is $N(0,\sigma^2)$; $\alpha \in (0,1)$. Write $r(p) = p + \theta\varphi(p) - \mu$, $m(p) = \mu - \theta\varphi(p)$ (the mean outcome at p), $V = \alpha r^2 + (1-\alpha)\theta^2\varphi^2$ and $F = V'/2 = \alpha r(1+\theta\varphi') + (1-\alpha)\theta^2\varphi\varphi'$. **(N)** For every (μ,θ) in a neighbourhood of B, V has a unique minimiser p^S on $\mathcal{P}$, it is interior, and $V''(p^S) > 0$.

Remark E.12 (Why this is "rational"). If the world's best response $\lambda^*(p)$ minimises a cost with decreasing differences in (λ,p), the monotone comparative statics of Topkis (1978) give a nondecreasing $\lambda^* \geq 0$, hence a nondecreasing $\Phi = \lambda^* p$. Scaling by θ is the responsiveness of the world: in Proposition A.7, $\theta = \gamma = \beta_W\kappa_W/(\alpha_W + \beta_W)$ and $\varphi(p) = p^2$. The stake model with the learner leading is the case $\varphi(p) = p^2$; (N) holds there by Proposition D.3.

E.2.1 The honest forecast is learnable

Theorem E.13 (Self-certification in the scaled setting)**.** *In the scaled setting:*

(i) $r(q) - r(p) \geq q - p$ *for* $p < q$*, so the honest forecast* p^* *is unique and (M2) holds with* $m = 1$*;*

(ii) the honest residual is observable: $r(p) = p - m(p)$ *is a function of the mean outcome at* p *alone; so whether* $p = p^*$ *is decided by the distribution of* y *at* p*;*

(iii) ***RSA***(a) *with* $a > \frac{1}{2}$ *has* $\mathrm{Regret}_T = O(\log T)$*, and every algorithm has* $\mathrm{Regret}_T \geq \sigma^2 \ln T - C$*; the constant is* σ^2 *with the tuned gain* $1/R'$*,* $R' = 1 + \theta\varphi'(p^*)$ *(Theorem B.21; see Remark B.19).*

Proof. (i) is Lemma B.1(iii) with Φ nondecreasing. (ii) $r(p) = p - (\mu - \theta\varphi(p)) = p - m(p)$. (iii) (M3) holds with $L = 1 + \theta \max_{\mathcal{P}} \varphi'$, (B1)–(B2) hold, and $p^* \in (0, \mu) \subset (0, P)$ because $r(0) = -\mu < 0 < \theta\varphi(\mu) = r(\mu)$; apply Theorems B.13, B.17 and B.29. □

E.2.2 The incentive grows with responsiveness

Theorem E.14 (The incentive in the scaled setting)**.** *In the scaled setting:*

(i) $V'(p^*) = 2(1-\alpha)\theta^2\varphi(p^*)\varphi'(p^*) > 0$*; hence* $p^S < p^*$ *(manipulation shades the forecast down) and* $\Delta_V := V(p^*) - V(p^S) > 0$*;*

(ii) at fixed μ, α*, with* $D := 1 + \theta\varphi'$ *and all functions at* p^S*,*

$$\frac{d\,|r(p^S)|}{d\theta} = \frac{(1-\alpha)\theta\varphi\left[2\varphi' D - \theta\varphi\varphi''\right]}{D\,\partial_p F};$$

(iii) $\dfrac{d\Delta_V}{d\theta} = 2(1-\alpha)\theta\left[g(p^*) - g(p^S)\right]$ *with* $g := \varphi^2/(1+\theta\varphi')$*;*

(iv) consequently the harm to users and the learner's gain both increase strictly with responsiveness whenever φ *is concave on* $[p^S, p^*]$*, and for every power law* $\varphi(p) = p^k$*,* $k > 0$*.*

Proof. **(i)** At p^*, $r = 0$. For $p > p^*$, $r > 0$, $1 + \theta\varphi' > 0$ and $\varphi\varphi' \geq 0$, so $V' > 0$; the minimiser is in $[0, p^*]$, and it is not p^* because $V'(p^*) > 0$.

(ii) $r(p^S) < 0$, and $\frac{d}{d\theta} r(p^S) = \varphi + D\,\frac{dp^S}{d\theta}$ with $\frac{dp^S}{d\theta} = -\partial_\theta F/\partial_p F$ (implicit function theorem; $\partial_p F = V''/2 > 0$ by (N)). Here $\partial_\theta F = \alpha\varphi D + \alpha r\varphi' + 2(1-\alpha)\theta\varphi\varphi'$ and $\partial_p F = \alpha(D^2 + r\theta\varphi'') + (1-\alpha)\theta^2(\varphi'^2 + \varphi\varphi'')$. So $\frac{d|r|}{d\theta} = [D\partial_\theta F - \varphi\partial_p F]/\partial_p F$. Substituting the first-order condition $\alpha r = -(1-\alpha)\theta^2\varphi\varphi'/D$ and simplifying (script `t_circle_resp.py`) gives the display.

(iii) Danskin's theorem at the unique minimiser and the chain rule at p^* (with $dp^*/d\theta = -\varphi(p^*)/(1+\theta\varphi'(p^*))$) give $\Delta_V' = \partial_\theta V(p^*) + V'(p^*)\frac{dp^*}{d\theta} - \partial_\theta V(p^S)$ with $\partial_\theta V = 2\alpha r\varphi + 2(1-\alpha)\theta\varphi^2$; substituting the first-order condition at p^S gives the display (symbolic check in the same script).

(iv) Concave φ: $\varphi'' \leq 0$ makes the bracket in (ii) positive; φ increasing and φ' nonincreasing make g increasing, so $g(p^*) > g(p^S)$. Power laws: the bracket is $2kp^{k-1} + \theta k(k+1)p^{2k-2} > 0$, and $g'(p) = kp^{2k-1}[2 + \theta(k+1)p^{k-1}]/(1+\theta kp^{k-1})^2 > 0$. □

Remark E.15. Both monotonicities can fail, and exactly as (ii)–(iii) predict: on 1499 random increasing responses the two sign conditions agreed with finite differences in every case, including 25 and 27 cases respectively where monotonicity fails (T33). All of these have large φ'' at p^S, i.e. a response that switches on sharply.

E.2.3 The strategic forecast is uninformative

Theorem E.16 (Identification and rates in the scaled setting)**.** *In the scaled setting:*

(i) ***Honest targets certify themselves.*** *If* $m_{\mu,\theta}(p) = m_{\mu',\theta'}(p)$ *then* $r_{\mu,\theta}(p) = r_{\mu',\theta'}(p)$*.*

(ii) ***Strategic targets do not.*** *Fix* (μ_0, θ_0) *and* $p := p^S(\mu_0, \theta_0)$*. Along* $u \mapsto (\mu_0 + u\varphi(p), \theta_0 + u)$ *the mean outcome at* p *is constant, while*

$$\frac{d}{du}F = (1-\alpha)\theta\varphi\varphi'\Big[2 - \frac{\theta\varphi'}{1+\theta\varphi'}\Big] > 0$$

at $u = 0$*. So for small* $u \neq 0$ *the worlds agree on everything observable at* p*, but* p *is the strategic forecast in only one of them. Globally, among the worlds on this line with* $\theta > 0$*,* p *is the strategic forecast only at* $u = 0$*.*

(iii) ***Learning to lead, μ unknown, costs*** $\Theta(\sqrt{T})$***.*** *For every (μ_0, θ_0) in the interior of B there is a compact curve $\Theta_0 \ni (\mu_0, \theta_0)$ with $\sup_{\Theta_0} S_T \geq c\sqrt{T} - C$ for every algorithm; and* ***ETC-S****, which plays 0 and P_0 and estimates $\hat{\mu} = \bar{y}_1$, $\hat{\theta} = (\hat{\mu} - \bar{y}_2)/\varphi(P_0)$, has $S_T = O(\sqrt{T})$.*

(iv) ***Learning to lead, μ known, costs*** $\Theta(\log T)$***.*** *The greedy learner with $\hat{\theta}_t = \Pi\big(\sum_{s<t}(\mu - y_s)\varphi(\hat{p}_s)/\sum_{s<t}\varphi(\hat{p}_s)^2\big)$ has $S_T = O(\log T)$, and every algorithm has $\sup S_T \geq c \ln T - C$ near any interior θ_0.*

(v) ***Learning to follow costs*** $\Theta(\log T)$ *(Theorem E.13).*

Proof. **(i)** $r(p) = p - m(p)$.

(ii) Along the curve $m(p) = \mu - \theta\varphi(p)$ has derivative $\varphi(p) - \varphi(p) = 0$, so $r(p)$ is constant; $F = \alpha r(1 + \theta\varphi') + (1 - \alpha)\theta^2\varphi\varphi'$ then has s-derivative $\alpha r\varphi' + 2(1 - \alpha)\theta\varphi\varphi'$. The first-order condition $\alpha r = -(1 - \alpha)\theta^2\varphi\varphi'/(1 + \theta\varphi')$ gives the display, which is positive because $\varphi(p) > 0$, $\varphi'(p) > 0$ and $\theta\varphi'/(1 + \theta\varphi') < 1$. Since $\partial_p F > 0$ at p, the root of F moves with s.

(iii) Follow the proof of Theorem E.3 with three replacements. *Quadratic growth* (Lemma E.4) uses only uniqueness, $V'' > 0$ and continuity, i.e. (N). *The curve*: $\mu'(s) = \varphi(\psi(s))$, $\theta'(s) = 1$, with $\psi(s) := p^S(\mu(s), \theta(s))$, which is C^1 by the implicit function theorem under (N); then $\partial_s m_s(p) = \varphi(\psi(s)) - \varphi(p)$, so $|\partial_s m_s(p)| \leq L_\varphi |p - \psi(s)|$ with $L_\varphi := \max_{\mathcal{P}} \varphi'$, replacing (B1); and $\psi' = -\frac{dF}{ds}/\partial_p F < 0$ by (ii), bounded away from 0 on a compact interval, replacing (B2). *Constants*: $c_2 = L_\varphi^2/\sigma^2$ in place of $4P^2/\sigma^2$. The upper bound is the proof of Theorem E.2: $\mathbb{E}\hat{\mu} = \mu$ because $\varphi(0) = 0$, $\mathbb{E}\hat{\theta} = \theta$, the variances are $O(1/n)$, p^S is Lipschitz on the box by the implicit function theorem under (N), and $V - V(p^S) \leq \frac{1}{2} M_V (p - p^S)^2$.

(iv) Upper bound as in Theorem E.8: every play satisfies $\hat{p}_t \geq p_{\min} := \min(P_0, \min_\theta p^S(\mu, \theta)) > 0$, so $\varphi(\hat{p}_t) \geq \varphi(p_{\min}) > 0$; the unprojected estimate is $\theta - M_t/A_t$ with $M_t = \sum_{s<t} \eta_s \varphi(\hat{p}_s)$ and $A_t \geq (t - 1)\varphi(p_{\min})^2$, so $\mathbb{E}(\hat{\theta}_t - \theta)^2 \leq \sigma^2\varphi(P)^2/((t - 1)\varphi(p_{\min})^4)$. Lower bound: Lemma E.6 with $\vartheta = \theta$, $\partial_\theta m = -\varphi(p)$, $|\partial_\theta m| \leq \varphi(P)$, and $\psi'(\theta) = -\partial_\theta F/\partial_p F$ with $\partial_\theta F = \alpha\varphi(1 + \theta\varphi') + (1 - \alpha)\theta\varphi\varphi'[2 - \theta\varphi'/(1 + \theta\varphi')] > 0$ (the μ-fixed version of the computation in (ii)); sum as in Theorem B.29.

(ii), globally. Along the line the residual at p is constant, $r = r_0$, because μ and $\theta\varphi(p)$ move together. Hence $F(u) = \alpha r_0(1 + (\theta_0 + u)\varphi') + (1 - \alpha)(\theta_0 + u)^2\varphi\varphi'$, and since $F(0) = 0$, $F(u) = u(B + Cu)$ with $B = \alpha r_0\varphi' + 2(1 - \alpha)\theta_0\varphi\varphi'$ and $C = (1 - \alpha)\varphi\varphi' > 0$. The only other zero is $u = -B/C$, at $\theta = \theta_0 - B/C$. From $F(0) = 0$, $\alpha r_0 = -(1 - \alpha)\theta_0^2\varphi\varphi'/(1 + \theta_0\varphi')$, so $B - \theta_0 C = \varphi'\big(\alpha r_0 + (1 - \alpha)\theta_0\varphi\big) = (1 - \alpha)\theta_0\varphi\varphi'/(1 + \theta_0\varphi') > 0$. Thus $B/C > \theta_0$ and the other zero has $\theta < 0$. For every other world on the line with $\theta > 0$, $F(p) \neq 0$, so the interior point p is not stationary and is not the strategic forecast. □

E.3 An explore-then-commit baseline for the honest forecast

Remark E.17 (A heuristic, not a theorem). A learner that does not use the residual identity must identify the response before committing. In the quadratic world, consider exploring for n rounds at design points within δ of p^* and then committing to the root of the estimated residual. Exploration costs about $n\delta^2 R'^2$, since $r(p^* \pm \delta)^2 \approx R'^2\delta^2$. The commitment error is dominated by the estimate of γ, whose regressor p^2 varies by order $p^*\delta$ across the design, so the commit phase costs about $T\sigma^2 {p^*}^2/(n\delta^2)$. Balancing the two gives a cost of order $R' p^* \sigma\sqrt{T}$. The fixed-point identity

$$R' p^* = (1 + 2\gamma p^*)p^* = 2\mu - p^* \in (\mu, 2\mu)$$

makes this of order $\mu\sigma\sqrt{T}$ for every $\gamma > 0$. We do not claim this as a bound: it is a heuristic for one design, not a lower bound for learners that identify the response. It explains why $\sqrt{T}$ is the natural rate for a learner that must buy the response, and why RSA avoids it: by Lemma B.1 the residual at the forecast is observed, so the honest learner never needs γ. The strategic learner cannot avoid it (Theorem E.3), because its target depends on the slope. Whether the constant in Theorem E.2 is independent of γ is not established; the identity does not hold at p^S, which does not solve the fixed-point equation.

F CALIBRATION AS EVIDENCE: PROOFS

F.1 THE POOLED AUDIT

Assumption F.1 (Audit model)**. (Au1)** The quadratic world with $\alpha \in (0,1)$; the learner knows μ, γ, σ. **(Au2)** Open loop: at time 0 the learner commits to a deterministic sequence $\hat{p}_1, \dots, \hat{p}_T \in \mathcal{P} = [0, P]$, $P \geq \mu$. (Relaxed in Theorem F.18.) **(Au3)** η_t i.i.d. $N(0, \sigma^2)$. **(Au4)** The auditor computes $\Sigma_T := \sum_{t \leq T}(y_t - \hat{p}_t)$ and flags the learner iff $|\Sigma_T| > z\sigma\sqrt{T}$. It needs no knowledge of μ, γ, Φ, only σ. **(Au5)** Being flagged costs the learner $\Pi_T \geq 0$, who minimises $J := \sum_t V(\hat{p}_t) + \Pi_T \mathbb{P}(\text{flag})$.

Terminology. The statistic Σ_T tests *unbiasedness*: whether the forecast errors average to zero. In the forecasting literature, *calibration* means something stronger, unbiasedness conditional on the value of the forecast (Dawid, 1982; Foster & Vohra, 1998). We therefore call this a *pooled audit*. The forecast-conditional strengthening is the binned audit of Appendix F.3 and Theorem 7.1.

Notation. $\rho_t := r(\hat{p}_t)$ (the bias); $R_{\mathcal{P}} := r(\mathcal{P}) = [-\mu, P + \gamma P^2 - \mu]$; $W(\rho) := V(r^{-1}(\rho))$; $\rho_s := r(p^S) < 0$; $\Delta_V := V(p^*) - V(p^S)$; $W_1 := W'(0)$; $q_z(m) := \bar{\Phi}_{\mathcal{N}}(z + m) + \bar{\Phi}_{\mathcal{N}}(z - m)$ with $\bar{\Phi}_{\mathcal{N}} := 1 - \Phi_{\mathcal{N}}$ the standard normal tail; $\beta := q_z(0) = 2\bar{\Phi}_{\mathcal{N}}(z)$.

Lemma F.2 (The learner's objective in terms of bias)**.** *(i)* $W(\rho) = \alpha\rho^2 + (1-\alpha)h(\rho)^2$ *with* $h(\rho) := \mu + \rho - r^{-1}(\rho) = \gamma(r^{-1}(\rho))^2$. *(ii)* W *is strictly convex on* $R_{\mathcal{P}}$. *(iii) The unique minimiser of* W *is* ρ_s*, and* $W(0) - W(\rho_s) = \Delta_V$. *(iv)* $W_1 = 4(1-\alpha)\gamma^2 {p^*}^3/s > 0$. *(v) For every* $\rho \in R_{\mathcal{P}}$*:* $W(0) - W(\rho) \leq \min\big(\Delta_V, W_1 \max(-\rho, 0)\big)$*, and* $W(\rho) \geq W(0) + W_1\rho + \alpha\rho^2$.

Proof. **(i)** With $p = r^{-1}(\rho)$, $\gamma p^2 = \mu + \rho - p = h(\rho)$.

(ii) r is strictly increasing and convex on $\mathcal{P}$ ($r'' = 2\gamma > 0$), so r^{-1} is increasing and concave: its derivative $1/(1+2\gamma p)$ decreases in ρ. Hence h is convex; also $h' = 1 - 1/(1+2\gamma p) \geq 0$ and $h \geq 0$, so $(h^2)'' = 2h'^2 + 2hh'' \geq 0$. Adding the strictly convex $\alpha\rho^2$ gives strict convexity.

(iii) r^{-1} is a bijection $R_{\mathcal{P}} \to \mathcal{P}$ and $W = V \circ r^{-1}$; V has the unique minimiser p^S (Proposition D.3), and $W(0) = V(p^*)$.

(iv) $W'(\rho) = V'(p)/r'(p)$; at $\rho = 0$, $V'(p^*) = 2c(p^*) = 4(1-\alpha)\gamma^2 {p^*}^3$ and $r'(p^*) = s$.

(v) $W - \alpha\rho^2 = (1-\alpha)h^2$ is convex with derivative W_1 at 0, so $W(\rho) \geq W(0) + W_1\rho + \alpha\rho^2 \geq W(0) + W_1\rho$; hence $W(0) - W(\rho) \leq -W_1\rho \leq W_1 \max(-\rho, 0)$. Also $W(0) - W(\rho) \leq W(0) - \min W = \Delta_V$. □

Lemma F.3 (The detection curve)**.** q_z *is even,* $q_z(0) = \beta$*, strictly increasing in* $|m|$*,* $\bar{\Phi}_{\mathcal{N}}(z - |m|) \leq q_z(m) \leq 1$*, and* $q_z''(0) = 2z\phi_{\mathcal{N}}(z) > 0$.

Proof. Evenness is immediate. For $m > 0$, $q'(m) = \phi(z - m) - \phi(z + m) > 0$ since $|z - m| < z + m$ and ϕ decreases in $|x|$. The lower bound holds because both terms are nonnegative. With $\phi'(x) = -x\phi(x)$, $q''(m) = (z-m)\phi(z-m) + (z+m)\phi(z+m)$. □

Lemma F.4 (Reduction to constant bias)**.** *For a deterministic sequence,* $\Sigma_T \sim N(-\sum_t \rho_t, T\sigma^2)$ *and* $\mathbb{P}(\text{flag}) = q_z(m)$ *with* $m := \sum_t \rho_t/(\sigma\sqrt{T})$*. Every optimal sequence is constant, and the problem is* $\min_{\bar{\rho} \in R_{\mathcal{P}}} J(\bar{\rho}) := T\,W(\bar{\rho}) + \Pi_T q_z(\sqrt{T}\bar{\rho}/\sigma)$*, which has a minimiser.*

Proof. $y_t - \hat{p}_t = -\rho_t + \eta_t$, so $\Sigma_T/(\sigma\sqrt{T}) = -m + N$ with $N \sim N(0,1)$, and $\mathbb{P}(-m + N > z) + \mathbb{P}(-m + N < -z) = \bar{\Phi}_{\mathcal{N}}(z+m) + \bar{\Phi}_{\mathcal{N}}(z-m)$. J depends on the sequence only through $\sum W(\rho_t)$ and $\sum \rho_t$; for a fixed sum, strict convexity (Lemma F.2(ii)) and Jensen give $\sum W(\rho_t) \geq TW(\bar{\rho})$ with equality only for the constant sequence. $R_{\mathcal{P}}$ is a compact interval and J is continuous. □

Theorem F.5 (Linear stakes discipline manipulation)**.** *Let* $\Pi_T = \kappa T$*, let* $M_0 > 0$ *satisfy* $\kappa(q_z(M_0) - \beta) > \Delta_V$*, and let* $c_q := \min_{0 < |m| \leq M_0}(q_z(m) - \beta)/m^2 > 0$*. Every optimal bias* $\bar{\rho}^*$ *satisfies: (i)* $\bar{\rho}^* < 0$ *and* $|\bar{\rho}^*| < M_0\sigma/\sqrt{T}$*; (ii)* $|\bar{\rho}^*| \leq W_1\sigma^2/(\kappa c_q T)$*; (iii) total harm* $T\bar{\rho}^{*2} \leq$

$W_1^2\sigma^4/(\kappa^2 c_q^2 T) \to 0$; *(iv) the learner's gain over honest play satisfies* $J(0) - J(\bar\rho^*) \le W_1^2\sigma^2/(\kappa c_q)$, *uniformly in* T.

Proof. **(i)** For $\bar\rho > 0$, $W(\bar\rho) > W(0)$ (W increases on $[\rho_s, \infty)$) and $q_z > \beta$, so $J(\bar\rho) > J(0)$. Since $q_z'(0) = 0$, the left derivative of J at 0 is $TW_1 > 0$, so J is smaller at small negative $\bar\rho$. If $|m^*| \ge M_0$ with $m^* := \sqrt{T}\bar\rho^*/\sigma$, then by Lemma F.2(v) and Lemma F.3, $J(\bar\rho^*) - J(0) \ge -T\Delta_V + \kappa T(q_z(M_0) - \beta) > 0$, a contradiction.

(ii) $J(\bar\rho^*) \le J(0)$ gives $\kappa T(q_z(m^*) - \beta) \le T(W(0) - W(\bar\rho^*)) \le TW_1|\bar\rho^*|$; the left side is at least $\kappa c_q T^2 \bar\rho^{*2}/\sigma^2$; divide by $T|\bar\rho^*| > 0$.

(iii) Square (ii) and multiply by T.

(iv) The penalty term is nonnegative, so $J(0) - J(\bar\rho^*) \le T(W(0) - W(\bar\rho^*)) \le TW_1|\bar\rho^*|$; apply (ii). □

Corollary F.6 (Vanishing false alarms)**.** *Let* $z_T := \sqrt{2\ln T}$, $\beta_T := 2\bar\Phi_{\mathcal N}(z_T)$, $\Pi_T = \kappa T$, *and suppose* $\kappa(\bar\Phi_{\mathcal N}(-1) - \beta_T) > \Delta_V$ *(if it holds at* T_0 *it holds for all* $T \ge T_0$ *since* β_T *decreases; it holds for all large* T *whenever* $\kappa\bar\Phi_{\mathcal N}(-1) > \Delta_V$, *with* $\bar\Phi_{\mathcal N}(-1) \approx 0.841$*). Then (i) an honest full-information learner pays* $\kappa T\beta_T \le 2\kappa/(\sqrt{2\pi}z_T) \to 0$*; (ii)* $|\bar\rho^*| < (z_T+1)\sigma/\sqrt{T}$, *so total harm is below* $(z_T+1)^2\sigma^2 = O(\ln T)$*; (iii) the gain is below* $W_1\sigma\sqrt{T}(z_T+1) = O(\sqrt{T\ln T})$.

Proof. **(i)** Mills: $\bar\Phi_{\mathcal N}(z) \le \phi_{\mathcal N}(z)/z$ and $\phi_{\mathcal N}(z_T) = 1/(\sqrt{2\pi}T)$.

(ii) Theorem F.5(i) with $M_0 = z_T + 1$, using $q_z(M_0) \ge \bar\Phi_{\mathcal N}(z_T - M_0) = \bar\Phi_{\mathcal N}(-1)$.

(iii) Lemma F.2(v) and (ii). □

Proposition F.7 (Sublinear stakes fail)**.** *If* $\Pi_T/T \to 0$, *every optimal bias satisfies* $\bar\rho_T^* \to \rho_s$, *and the total harm* $T\bar\rho_T^{*2} \sim T\rho_s^2$ *is linear in* T.

Proof. $TW(\bar\rho_T^*) \le J(\bar\rho_T^*) \le J(\rho_s) \le TW(\rho_s) + \Pi_T$, so $W(\bar\rho_T^*) \le \min W + \Pi_T/T \to \min W$. W is continuous on the compact $R_{\mathcal P}$ with the unique minimiser ρ_s, so every limit point of $\bar\rho_T^*$ is ρ_s. □

Lemma F.8 (A model-free identity for RSA)**.** *If* RSA(a) *never projects during rounds* $1, \dots, T$, *then for any response* Φ, $\Sigma_T = -\frac{1}{a}\sum_{t\le T}(\hat p_t - \hat p_{T+1})$. *With projection corrections* $\Delta_t := \hat p_{t+1} - y_t^u$, *in general* $-a\Sigma_T = \sum_{t\le T}(\hat p_t - \hat p_{T+1}) + \sum_{t\le T} t\Delta_t$.

Proof. $\Sigma_T = -\sum_t z_t$ and $z_t = (t/a)(\hat p_t - y_t^u) = (t/a)(\hat p_t - \hat p_{t+1} + \Delta_t)$. Summation by parts: $\sum_{t=1}^T t(\hat p_t - \hat p_{t+1}) = \sum_{t=1}^T t\hat p_t - \sum_{t=2}^{T+1}(t-1)\hat p_t = \sum_{t=1}^T \hat p_t - T\hat p_{T+1}$. □

Proposition F.9 (Honest learners pass the audit, linear model)**.** *for the linear model; for the quadratic world and other smooth worlds see Proposition F.12 Take* $r(p) = R'(p - p^*)$ *on* $\mathbb R$ *with no projection, noise as in (Au3), any* $\hat p_1$, *and* RSA(a) *with* $c := aR' > \frac12$. *Then* $\Sigma_T/(\sigma\sqrt{T}) \Rightarrow N\big(0, 1/(2c-1)\big)$. *Consequently (i) at fixed* z, $\mathbb P($*flag*$) \to 2\bar\Phi_{\mathcal N}(z\sqrt{2c-1})$, *which is at most* β *iff* $c \ge 1$*; (ii) if* $c > 1$ *and* $z = z_T$, *then* $T\,\mathbb P($*flag*$) \to 0$.

Proof. With $e_t := \hat p_t - p^*$: $e_{t+1} = (1 - c/t)e_t + (a/t)\eta_t$ and $\Sigma_T = \sum_t(\eta_t - R'e_t)$, an affine function of Gaussians, hence Gaussian.

Mean. $\mathbb E e_{t+1} = (1 - c/t)\mathbb E e_t$, so $|\mathbb E e_t| \le Ct^{-c}$ (once $u > c$ each factor is in $(0,1)$ and $\log\prod(1 - c/u) \le -c\sum 1/u$); hence $|\mathbb E\Sigma_T| \le R'C\sum_{t\le T} t^{-c} = O(T^{1-c} + \ln T)$ and $\mathbb E\Sigma_T/\sqrt{T} \to 0$ since $c > \frac12$.

Variance, $c \ne 1$**.** Multiplying the recursion by t: $te_{t+1} - (t-1)e_t = (1-c)e_t + a\eta_t$; summing, $Te_{T+1} = (1-c)\sum_t e_t + a\sum_t \eta_t$, hence, with $aR' = c$, $\Sigma_T = [\sum_t \eta_t - R'Te_{T+1}]/(1-c)$ exactly. Let $v_t := \operatorname{Var} e_t$; $v_{t+1} = (1 - c/t)^2 v_t + a^2\sigma^2/t^2$, so Lemma B.11 (rate $2c > 1$) gives $v_t = O(1/t)$ and Lemma B.16 (with $2c$ and remainder $c^2 v_t/t^2 = O(t^{-3})$) gives $tv_t \to a^2\sigma^2/(2c -$

1). Let $w_T := \mathrm{Cov}(\sum_{t\le T}\eta_t, e_{T+1})$, $w_0 = 0$; since η_T is independent of e_T and earlier noise, $w_T = (1-c/T)w_{T-1} + a\sigma^2/T$, and $d_T := w_T - a\sigma^2/c$ satisfies $d_T = (1-c/T)d_{T-1} \to 0$. Therefore $(1-c)^2\,\mathrm{Var}(\Sigma_T)/(\sigma^2 T) = 1 - 2R'w_T/\sigma^2 + R'^2 T v_{T+1}/\sigma^2 \to 1 - 2 + c^2/(2c-1) = (c-1)^2/(2c-1)$.

Variance, $c = 1$. Then $Te_{T+1} = a\sum_{t\le T}\eta_t$, so for $t \ge 2$, $R'e_t = \frac{1}{t-1}\sum_{u<t}\eta_u$ and the coefficient of η_u in Σ_T is $k_u = 1 - (H_{T-1} - H_{u-1})$ (H_n harmonic numbers). With $H_{T-1} - H_{u-1} = \ln(T/u) + O(1/u)$, $\frac{1}{T}\sum_u k_u^2 = \frac{1}{T}\sum_u (1 - \ln(T/u))^2 + O(\ln^2 T/T) \to \int_0^1 (1+\ln x)^2 dx = 1$.

(ii). For $c > 1$, $\mathbb{E}\Sigma_T = O(1)$, so the standardised mean is $m_T = O(T^{-1/2})$, and the standardised variance $v_T \to v := 1/(2c-1) < 1$, so $v_T \le v' := (1+v)/2 < 1$ for large T. Then $\mathbb{P}(\text{flag}) \le 2\bar\Phi_{\mathcal{N}}((z_T - |m_T|)/\sqrt{v_T}) \le \exp(-(z_T - |m_T|)^2/(2v'))$, and $(z_T - |m_T|)^2 \ge 2\ln T - o(1)$, so $\mathbb{P} = O(T^{-1/v'})$ and $T\mathbb{P} = O(T^{1-1/v'}) \to 0$. □

Remark F.10. The regret-optimal gain $c = 1$ sits exactly on the boundary; a slightly larger gain removes the problem at a small cost in regret (e.g. $c = 1.2$ costs the factor $c^2/(2c-1) = 1.029$).

Lemma F.11 (Negative drift gives uniform exponential moments)**.** *Let $Y_t \ge 0$ be adapted to $(\mathcal{F}_t)$ with $Y_1 \le y_1$. Suppose (D1) $\mathbb{E}[Y_{t+1} - Y_t \mid \mathcal{F}_t] \le -\varepsilon_0$ on $\{Y_t > b\}$, and (D2) $|Y_{t+1} - Y_t| \le Z_t$, where Z_t is independent of $\mathcal{F}_t$ with the law of a fixed $Z \ge 0$ satisfying $\mathbb{E}e^{\lambda_0 Z} < \infty$. Then there are $\eta > 0$ and $\rho < 1$ with*

$$\sup_t \mathbb{E}e^{\eta Y_t} \le \max\left(e^{\eta y_1},\ \frac{e^{\eta b}\,\mathbb{E}e^{\eta Z}}{1-\rho}\right) < \infty.$$

This is a special case of the drift analysis of Hajek (1982), whose conditions D1 and D2 it mirrors; we include the short proof.

Proof. **Taylor bound.** For real x, $e^x \le 1 + x + \frac{x^2}{2}e^{|x|}$. With $\Delta := Y_{t+1} - Y_t$ and $0 < \eta \le \lambda_0/2$, (D2) gives $\mathbb{E}[e^{\eta\Delta} \mid \mathcal{F}_t] \le 1 + \eta\mathbb{E}[\Delta \mid \mathcal{F}_t] + \frac{\eta^2}{2}C_2$ with $C_2 := \mathbb{E}[Z^2 e^{\lambda_0 Z/2}] < \infty$ (finite because $x^2 e^{\lambda_0 x/2} \le Ce^{\lambda_0 x}$).

Above the threshold. On $\{Y_t > b\}$, (D1) gives $\mathbb{E}[e^{\eta\Delta} \mid \mathcal{F}_t] \le 1 - \eta\varepsilon_0 + \eta^2 C_2/2 \le 1 - \eta\varepsilon_0/2 =: \rho < 1$ once $\eta \le \varepsilon_0/C_2$. So $\mathbb{E}[e^{\eta Y_{t+1}} \mid \mathcal{F}_t] \le \rho e^{\eta Y_t}$ there.

Below the threshold. On $\{Y_t \le b\}$, $e^{\eta Y_{t+1}} \le e^{\eta b}e^{\eta Z_t}$, so $\mathbb{E}[e^{\eta Y_{t+1}} \mid \mathcal{F}_t] \le e^{\eta b}\mathbb{E}e^{\eta Z} =: D'$.

Iterate. With $m_t := \mathbb{E}e^{\eta Y_t}$, $m_{t+1} \le \rho m_t + D'$, so by induction $m_t \le \max(m_1, D'/(1-\rho))$. □

Proposition F.12 (Honest learners pass the audit: nonlinear residuals)**.** *Assume Assumption A.1, (A2-G), (M1) with p^* in the interior of $\mathcal{P}$, (M2) with $2am > 1$, (M3), and $|\rho(e)| \le K_\rho e^2$ on $\mathcal{P}$ for $\rho(e) := r(p^* + e) - R'e$ (in the quadratic world, $\rho(e) = \gamma e^2$ exactly). Let $c := aR' > \frac{1}{2}$. Then (i) RSA(a) satisfies $\Sigma_T/(\sigma\sqrt{T}) \Rightarrow N(0, 1/(2c-1))$, and $\mathbb{P}(\text{flag}) \to 2\bar\Phi_{\mathcal{N}}(z\sqrt{2c-1})$ at fixed z; (ii) if moreover $c > 1$, then at the level $z_T = \sqrt{2\ln T}$, $T\,\mathbb{P}(\text{flag}) \to 0$.*

Proof. **Coupling.** Let $\tilde e_t$ be the linear process on $\mathbb{R}$ driven by the same noise: $\tilde e_1 = e_1$, $\tilde e_{t+1} = (1 - c/t)\tilde e_t + (a/t)\eta_t$. Let $\tilde\Sigma_T := \sum_{t\le T}(\eta_t - R'\tilde e_t)$; by Proposition F.9, $\tilde\Sigma_T/(\sigma\sqrt{T}) \Rightarrow N(0, 1/(2c-1))$.

The difference. With $d_t := e_t - \tilde e_t$ and projection correction π_t, $e_{t+1} = e_t - (a/t)(R'e_t + \rho(e_t) - \eta_t) + \pi_t$, so $d_{t+1} = (1-c/t)d_t - (a/t)\rho(e_t) + \pi_t$. For $t \ge c$, $\mathbb{E}|d_{t+1}| \le (1-c/t)\mathbb{E}|d_t| + (a/t)K_\rho\mathbb{E}e_t^2 + \mathbb{E}|\pi_t| \le (1-c/t)\mathbb{E}|d_t| + C/t^2$, using $\mathbb{E}e_t^2 \le M_1'/t$ (Lemma B.12) and $\mathbb{E}|\pi_t| \le \mathbb{E}f_t^6/D_0^5 \le Ct^{-3}$ (Theorem B.17, Step 2). Fix $k \in (\frac{1}{2}, \min(c,1))$; since $C/t^2 \le C/t^{k+1}$, Lemma B.11 gives $\mathbb{E}|d_t| \le Ct^{-k}$.

The statistic. $\Sigma_T = \sum_t(\eta_t - r(\hat p_t)) = \tilde\Sigma_T - R'\sum_t d_t - \sum_t \rho(e_t)$, and $\mathbb{E}\big|R'\sum_{t\le T} d_t\big| \le CT^{1-k} = o(\sqrt{T})$ and $\mathbb{E}\big|\sum_{t\le T}\rho(e_t)\big| \le K_\rho M_1'(1+\ln T)$. So $(\Sigma_T - \tilde\Sigma_T)/\sqrt{T} \to 0$ in L^1, and Slutsky's lemma gives the limit. The flagging probability converges because the limit law is continuous.

(ii) Moments of the difference. Take $q := 5$. By Lemma B.12 with $k = q$ and $k = 2q$ (Gaussian noise has all moments), $\|e_t^2\|_q \le C/t$ and, as in Theorem B.17 Step 2, $\|\pi_t\|_q \le (\mathbb{E} f_t^{4q}/D_0^{3q})^{1/q} \le Ct^{-2}$. Minkowski's inequality in the recursion for d_t gives $\|d_{t+1}\|_q \le (1 - c/t)\|d_t\|_q + C/t^2$, so by Lemma B.11 with exponent $\frac{3}{4} < \min(c, 1)$, $\|d_t\|_q \le Ct^{-3/4}$. Hence $\|\Sigma_T - \tilde{\Sigma}_T\|_q \le R' \sum_t \|d_t\|_q + K_\rho \sum_t \|e_t^2\|_q \le C(T^{1/4} + \ln T)$, and by Markov's inequality $T\,\mathbb{P}(|\Sigma_T - \tilde{\Sigma}_T| > \varepsilon z_T \sigma\sqrt{T}) \le CT \cdot T^{q/4}/(\varepsilon z_T \sigma\sqrt{T})^q = O(T^{1-q/4}) \to 0$.

(ii) The Gaussian part. $\tilde{\Sigma}_T$ is Gaussian with mean $O(1)$ (since $c > 1$, as in Proposition F.9) and variance $v_T\sigma^2 T$ with $v_T \to v := 1/(2c-1) < 1$. Choose $v' \in (v, 1)$ and $\varepsilon > 0$ with $(1-\varepsilon)^2 > v'$. For large T, $\mathbb{P}(|\tilde{\Sigma}_T| > (1-\varepsilon) z_T \sigma\sqrt{T}) \le \exp(-(1-\varepsilon)^2 z_T^2/(2v')) \cdot O(1) = O(T^{-(1-\varepsilon)^2/v'})$, and T times this tends to 0. Since flagging requires one of the two events, $T\,\mathbb{P}(\text{flag}) \to 0$. □

Lemma F.13 (Distance to the threshold, summed)**.** *Assume (G1)–(G5) and Assumption A.2 with moments up to order 4, and run RSA(a) with $a > \frac{1}{2}$. Then $u_t := \mathbb{E}(\hat{p}_t - x_0)^2 \le M_1/t$, $\mathbb{E}(\hat{p}_t - x_0)^4 \le M_2/t^2$, and $\mathbb{E}\sum_{t \le T} |\hat{p}_t - x_0| \le C(1 + \ln T)$.*

Proof. **Moments.** With $\kappa = 0 \in [r^-, r^+]$, Lemma C.15 gives modulus 1 and target x_0, and Lemma C.14 gives the two moment bounds.

Drift. Write $e_t := \hat{p}_t - x_0$. By Lemma C.2, $e_t r(\hat{p}_t) \ge e_t^2 + c_0|e_t|$ with $c_0 = \min(u, v)$: for $\hat{p}_t < x_0$, $r \le r^- - |e_t|$, so $e_t r \ge |e_t|(u + |e_t|)$; symmetrically on the right. By nonexpansiveness,

$$u_{t+1} \le u_t - \frac{2a}{t}\big(u_t + c_0 \mathbb{E}|e_t|\big) + \frac{C}{t^2}, \qquad \text{so} \qquad \frac{2ac_0}{t}\mathbb{E}|e_t| \le u_t - u_{t+1} + \frac{C}{t^2}.$$

Dyadic blocks. Sum over $t \in [2^j, 2^{j+1})$: $\frac{2ac_0}{2^{j+1}} \sum_t \mathbb{E}|e_t| \le u_{2^j} + C2^{1-j} \le (M_1 + 2C)2^{-j}$, so each block contributes at most $(M_1 + 2C)/(ac_0)$. There are at most $1 + \log_2 T$ blocks. □

Lemma F.14 (The squared distance decays like $t^{-4/3}$)**.** *Under the assumptions of Lemma F.13, $u_t \le M' t^{-4/3}$, hence $T\,\mathbb{E}|\hat{p}_{T+1} - x_0| \le \sqrt{M'}\, T^{1/3}$.*

Proof. **Lower bound on the first moment.** By Hölder, $\mathbb{E}e^2 = \mathbb{E}[|e|^{2/3}|e|^{4/3}] \le (\mathbb{E}|e|)^{2/3}(\mathbb{E}e^4)^{1/3}$, so $\mathbb{E}|e_t| \ge u_t^{3/2}/(\mathbb{E}e_t^4)^{1/2} \ge u_t^{3/2}\, t/\sqrt{M_2}$.

Recursion. Substituting into the drift inequality of Lemma F.13 and dropping $-(2a/t)u_t \le 0$: $u_{t+1} \le u_t - c' u_t^{3/2} + C/t^2$ with $c' := 2ac_0/\sqrt{M_2}$.

Induction. $\varphi(x) := x - c'x^{3/2}$ is increasing on $[0, x_1]$, $x_1 := 4/(9c'^2)$. By Lemma F.13, $u_t \le M_1/t$ for all t. Choose t_1 and M' with $M' \ge M_1 t_1^{1/3}$, $M' t_1^{-4/3} \le x_1$ and $c' M'^{3/2} - \frac{4}{3}M' \ge C$; this is possible, since with $M' = M_1 t_1^{1/3}$ the second condition reads $M_1/t_1 \le x_1$ and the third holds for t_1 large. For $t \le t_1$, $u_t \le M_1/t \le M_1 t_1^{1/3} t^{-4/3} \le M' t^{-4/3}$. If $u_t \le M' t^{-4/3} \le x_1$ with $t \ge t_1$, then $u_{t+1} \le \varphi(M' t^{-4/3}) + Ct^{-2} = M' t^{-4/3} - (c' M'^{3/2} - C)t^{-2} \le M' t^{-4/3} - \frac{4}{3}M' t^{-7/3} \le M'(t+1)^{-4/3}$, using $(1 + 1/t)^{-4/3} \ge 1 - \frac{4}{3t}$ and $t^{-2} \ge t^{-7/3}$. Finally $\mathbb{E}|e_{T+1}| \le \sqrt{u_{T+1}}$. □

Theorem F.15 (Unbiasedness cannot certify prediction quality: RSA in the gap)**.** *Assume (G1)–(G5) and (A2-G), and run RSA(a) with $a \ge 1$. Then (i) $\sup_t \mathbb{E}\exp(\lambda t|\hat{p}_t - x_0|) < \infty$ for some $\lambda > 0$; (ii) $\mathbb{E}|\Sigma_T| \le C(1 + \ln T)$; (iii) $\mathbb{P}(|\Sigma_T| > z\sigma\sqrt{T}) \le C(1 + \ln T)/(z\sigma\sqrt{T}) \to 0$ for fixed z and for $z = z_T$; (iv) for every gain $a > \frac{1}{2}$ (not only $a \ge 1$), $\mathbb{E}|\Sigma_T| \le C(1 + \ln T + T^{1/3})$, so $\mathbb{P}(|\Sigma_T| > z\sigma\sqrt{T}) \to 0$ for fixed z and for $z = z_T$. So a learner whose regret is necessarily linear (Proposition C.7) has smaller average bias than an honest learner in the smooth model, whose Σ_T is of order $\sqrt{T}$ (Proposition F.9).*

The proof uses Lemma F.11.

Proof. **(i)** Let $x_t := t(\hat{p}_t - x_0)$ and $Y_t := |x_t|$. By nonexpansiveness and $x_0 \in \mathcal{P}$, $|x_{t+1}| \le (1 + 1/t)|x_t - ar_t + a\eta_t|$.

Drift for $x_t > 0$. Then $\hat{p}_t > x_0$ and Lemma C.2 gives $r_t \ge r^+ + x_t/t$. Let $X := x_t - ar_t$; then $X \le x_t(1 - a/t) - ar^+$ and $X \ge x_t - aB$. If $x_t \ge aB + b$ then $X \ge b \ge 0$ and $\mathbb{E}|X + a\eta| = X + 2\mathbb{E}[(X + a\eta)^-] \le X + g(b)$ with $g(b) := 2a\mathbb{E}[|\eta|\mathbf{1}\{|\eta| > b/a\}] \to 0$. Hence $\mathbb{E}|x_{t+1}| \le (1+1/t)(x_t(1-a/t)-ar^+ + g(b)) \le x_t(1+(1-a)/t) - (ar^+ - g(b)) \le x_t - (ar^+ - g(b))$ for $a \ge 1$. The case $x_t < 0$ is symmetric with $u = -r^-$. Choosing b with $g(b) \le a\min(u, v)/2$ gives (D1) with $\varepsilon_0 = a\min(u, v)/2$.

Increments. $|\hat{p}_{t+1} - \hat{p}_t| \le (a/t)|z_t|$, so $|x_{t+1} - x_t| \le |e_{t+1}| + t|e_{t+1} - e_t| \le P + a(B + |\eta_t|)$, with Gaussian tails: (D2) holds, and Lemma F.11 gives (i).

(ii) By Lemma F.8, $a|\Sigma_T| \le \sum_{t \le T} |e_t| + T|e_{T+1}| + \sum_t t|\pi_t|$. With $M := \sup_t \mathbb{E}Y_t < \infty$: $\mathbb{E}\sum_{t \le T} |e_t| = \sum_t \mathbb{E}Y_t/t \le M(1 + \ln T)$ and $T\mathbb{E}|e_{T+1}| \le M$. $\pi_t \ne 0$ requires $|f_t| \ge D_0$ where $|f_t| \le |e_t| + (a/t)(B + |\eta_t|)$, hence $Y_t \ge tD_0/2$ or $|\eta_t| \ge tD_0/(2a) - B$, both of probability exponentially small in t by (i) and Gaussian tails; by Cauchy–Schwarz, $\sum_t t\mathbb{E}|\pi_t| < \infty$.

(iii) Markov's inequality.

(iv) In the bound of (ii), $\mathbb{E}\sum_{t \le T} |e_t| \le C(1 + \ln T)$ by Lemma F.13, $T\mathbb{E}|e_{T+1}| \le CT^{1/3}$ by Lemma F.14, and the projection term satisfies $\mathbb{E}[t|\pi_t|] \le t\,\mathbb{E}f_t^6/D_0^5 \le Ct^{-2}$ (sixth moments from Lemma C.14). Markov's inequality completes the proof. □

Remark F.16. For $a < 1$ the drift of $t|e_t|$ acquires the term $(1 - a)|e_t|$, so part (i) needs $a \ge 1$; part (iv) covers every $a > \frac{1}{2}$ by a different route. T29 ($a = 0.6, 0.8$) finds $\mathbb{E}|Te_T|$ flat at about 0.31 and 0.41, suggesting that (i) also holds for $a < 1$ (open). Numerically (T17, $a = 1$), $\mathbb{E}|t(\hat{p}_t - x_0)|$ is flat at 0.52 and 1.53 on two instances, $\mathbb{E}|\Sigma_T| \le 13.8$ at $T = 10^5$, and no run of 1000 was flagged.

Corollary F.17 (In the gap, accuracy and auditability conflict)**.** *Assume (G1)–(G5) and (A2-G). For every algorithm, $\mathbb{E}\Sigma_T = -T\,\mathbb{E}\bar{r}_T$, where $\bar{r}_T$ is the average bias of Theorem C.10. Hence:*

(i) if $\mathbb{E}|\Sigma_T| \le C\sqrt{T}$, which is the scale at which the audit operates, then $|\mathbb{E}\bar{r}_T| \le C/\sqrt{T}$ and, by Theorem C.10(i), $\mathrm{Regret}_T/T \ge uv - |r^- + r^+|\,C/\sqrt{T}$;

(ii) the accuracy-optimal learner of Theorem C.19 has $\mathbb{E}\bar{r}_T \to \zeta^* \in \{r^-, r^+\}$, which is nonzero, so $\mathbb{E}|\Sigma_T| \ge T|\mathbb{E}\bar{r}_T|$ grows linearly; since $\Sigma_T = -T\bar{r}_T + \sum_t \eta_t$ with $\frac{1}{T}\sum_t \eta_t \to 0$ a.s. and $\bar{r}_T \to \zeta^*$ in probability, it is flagged with probability tending to one for any fixed z, and for $z = z_T$.

So in the gap an auditor who enforces zero average bias forces an excess of at least uv, the RSA *value (Theorem F.15 shows* RSA *passes), and punishes the learner that is best for accuracy.*

Proof. $y_t - \hat{p}_t = -r(\hat{p}_t) + \eta_t$ and $\mathbb{E}\eta_t = 0$ give $\mathbb{E}\Sigma_T = -T\mathbb{E}\bar{r}_T$, and $|\mathbb{E}\Sigma_T| \le \mathbb{E}|\Sigma_T|$. (i) follows from Theorem C.10(i). For (ii), the proof of Theorem C.19 shows that on the events G_n, whose probability tends to one, Phase B runs shifted RSA with $\hat{\kappa}$ within $\delta_n \to 0$ of ζ^*; step (iv) of the proof of Lemma C.16, applied conditionally on Phase A, gives $\mathbb{E}[|\bar{r}_B - \hat{\kappa}| \mid \mathcal{F}_A] \le C/\sqrt{T - n}$ for the Phase B average $\bar{r}_B$, so $\bar{r}_T \to \zeta^*$ in probability, since Phase A occupies a vanishing fraction of rounds and $|r| \le B$. On that event $|\Sigma_T|/(\sigma\sqrt{T}) \ge \sqrt{T}\,(|\zeta^*| - o(1))/\sigma - |\sum_t \eta_t|/(\sigma\sqrt{T})$, which exceeds any fixed z, and exceeds $z_T = \sqrt{2\ln T}$, with probability tending to one. □

Theorem F.18 (Closed-loop manipulation)**.** *Keep Assumption F.1 except (Au2): the bias $\rho_t \in R_{\mathcal{P}}$ may be chosen after observing everything so far, possibly with randomisation. Let $G := \sum_t (W(0) - W(\rho_t))$, $D := -\sum_t \rho_t$, $N_T := \sum_t \eta_t$, and $NB := \mathbb{E}G - \kappa T\mathbb{P}(\textit{flag}) + \kappa T\beta$ (net benefit over honest play). Then: (i) for every policy, $\mathbb{E}G \le W_1(z + 1)\sigma\sqrt{T} + T\Delta_V\mathbb{P}(\textit{flag})$; (ii) $NB - \kappa T\beta \le W_1(z + 1)\sigma\sqrt{T} + (\Delta_V - \kappa)T\mathbb{P}(\textit{flag})$, so if $\kappa \ge \Delta_V$ a closed-loop strategic learner gains at most $W_1(z+1)\sigma\sqrt{T}$ beyond recovering the false-alarm penalty; (iii) if $z = z_T$ and $\kappa > \Delta_V$, every optimal policy has $\mathbb{P}^*(\textit{flag}) \le [\kappa T\beta_T + W_1(z_T + 1)\sigma\sqrt{T}]/((\kappa - \Delta_V)T) = O(\sqrt{\ln T/T})$, $\mathbb{E}G^* = O(\sqrt{T\ln T})$, and harm to users $\mathbb{E}\sum_t \rho_t^{*2} = O(\sqrt{T\ln T})$. Without an audit both gain and harm are linear in T.*

Proof. By Lemma F.2(v), $W(\rho) \geq W(0) + W_1\rho + \alpha\rho^2$, so $G \leq W_1 D - \alpha\sum \rho_t^2 \leq W_1 D$; and each term satisfies $W(0) - W(\rho_t) \leq \Delta_V$, so $G \leq T\Delta_V$. Also $\Sigma_T = D + N_T$, so on no flagging $D \leq z\sigma\sqrt{T} + |N_T|$; the noise is exogenous, so $\mathbb{E}|N_T| \leq \sigma\sqrt{T}$ whatever the policy.

(i) $\mathbb{E}G \leq \mathbb{E}[W_1(z\sigma\sqrt{T} + |N_T|)\mathbf{1}\{\text{no flag}\}] + T\Delta_V\mathbb{P}(\text{flag}) \leq W_1(z+1)\sigma\sqrt{T} + T\Delta_V\mathbb{P}(\text{flag})$.

(ii) Substitute (i) into NB.

(iii) Honest play is feasible with $NB = 0$, so $NB^* \geq 0$; rearranging (ii) bounds $\mathbb{P}^*$; substituting into (i) bounds $\mathbb{E}G^*$. For the harm, $\alpha\mathbb{E}\sum \rho^{*2} \leq W_1\mathbb{E}D^* - \mathbb{E}G^*$, with $\mathbb{E}G^* = NB^* + \kappa T\mathbb{P}^* - \kappa T\beta_T \geq -\kappa T\beta_T$ (and $\kappa T\beta_T \to 0$ by Corollary F.6) and $\mathbb{E}D^* \leq (z_T+1)\sigma\sqrt{T} + \mu T\mathbb{P}^*$ because $\rho \geq -\mu$ gives $D \leq \mu T$. □

Remark F.19. Against the exact dynamic programming of E17 (fixed $z = 1.96$, $\kappa = 0.0466 > \Delta_V = 0.0370$), bound (ii) holds at all five horizons (at $T = 10^4$: $NB - \kappa T\beta = 29.3 \leq 49.3$). The dynamic programming suggests harm of order $\ln T$; that rate is open.

F.2 When the pooled audit fails

Theorem F.20 (The pooled audit in general worlds)**.** *Let $\Phi \geq 0$ be nondecreasing and C^1 on $\mathcal{P}$, let $W(\rho) := V(r^{-1}(\rho))$ on $R_\mathcal{P} = r(\mathcal{P})$ (the learner's per-round cost as a function of its bias), and $W_1 := W'(0) = 2(1-\alpha)\Phi(p^*)\Phi'(p^*)/(1+\Phi'(p^*))$.*

(i) If there is $c > 0$ with $W(\rho) \geq W(0) + W_1\rho + c\rho^2$ for all $\rho \in R_\mathcal{P}$ (condition (S_c)), then Theorem F.18 holds verbatim with α replaced by c in the harm bound: against any closed-loop strategic learner, with linear stakes κT, $\kappa > \Delta_V$ and level z_T, the gain and the harm are $O(\sqrt{T\ln T})$.

(ii) (S_α) holds whenever Φ is convex on $\mathcal{P}$ (every power law with $k \geq 1$, including the stake model with the learner leading).

(iii) If $W(\rho_1) < W(0) + W_1\rho_1$ for some $\rho_1 \in R_\mathcal{P}$, then there are biases $\rho_- < 0 < \rho_+$ and $\lambda \in (0,1)$ with $\lambda\rho_- + (1-\lambda)\rho_+ = 0$ and $g_0 := W(0) - \lambda W(\rho_-) - (1-\lambda)W(\rho_+) > 0$. The open-loop strategic learner that plays ρ_- in $\lfloor \lambda T \rfloor$ rounds and ρ_+ otherwise has $|\sum_t \rho_t| \leq |\rho_-| + |\rho_+|$, so its flagging probability is $q_z\big(O(T^{-1/2})\big) \to \beta$, the honest rate, while its gain over honest play is at least $g_0 T - O(1)$ and its harm to users is linear in T. No pooled audit with any stakes and any fixed level disciplines it.

Proof. **(i)** The proof of Theorem F.18 uses only $W(0) - W(\rho) \leq \Delta_V$, $\rho \geq -\mu$ on $R_\mathcal{P}$ (true since $r(0) = -\mu$), and $W(\rho) \geq W(0) + W_1\rho + \alpha\rho^2$; replace the last by ($S_c$).

(ii) If Φ is convex then r is convex and increasing, so r^{-1} is concave and $h := \Phi \circ r^{-1} = \mu + \rho - r^{-1}(\rho)$ is convex, nonnegative and nondecreasing; hence h^2 is convex, and $W = \alpha\rho^2 + (1-\alpha)h^2 \geq W(0) + W_1\rho + \alpha\rho^2$, as in Lemma F.2(v).

(iii) W is differentiable at 0 because r is C^1 with $r' \geq 1$. Suppose $\rho_1 < 0$ and put $\epsilon := W(0) + W_1\rho_1 - W(\rho_1) > 0$. For small $\rho_+ > 0$ let $\lambda := \rho_+/(\rho_+ - \rho_1)$, so $\lambda\rho_1 + (1-\lambda)\rho_+ = 0$. Then $\lambda W(\rho_1) + (1-\lambda)W(\rho_+) = W(0) - \lambda\epsilon + (1-\lambda)o(\rho_+)$, using $W(\rho_+) = W(0) + W_1\rho_+ + o(\rho_+)$ and $(1-\lambda)\rho_+ = -\lambda\rho_1$. Since $\lambda \geq \rho_+/(\rho_+ + |\rho_1|)$ is of order ρ_+, the right side is below $W(0)$ for small ρ_+. The case $\rho_1 > 0$ is symmetric. For the sequence, rounding λT leaves $|\sum_t \rho_t| \leq |\rho_-| + |\rho_+|$; by Lemma F.4 the flagging probability is $q_z(-\sum_t \rho_t/(\sigma\sqrt{T})) \to \beta$; the gain is $\sum_t (W(0) - W(\rho_t)) \geq g_0 T - O(1)$; and the harm is $\sum_t \rho_t^2 \geq T\min(\rho_-^2, \rho_+^2)$. An honest learner faces the same flagging probability, so the expected penalty is the same for both, whatever the stakes. □

Proposition F.21 (The audit can be evaded)**.** *Let $\Phi(p) = \frac{1}{2}\big[(1 + e^{-40(p-0.45)})^{-1} - (1+e^{18})^{-1}\big]$ (a steep logistic response), $\mu = 1$, $P = 3$. For $\alpha = \frac{1}{5}$ the tangent condition fails at $\hat{p}_1 = 1.41454$, and for $\alpha = \frac{1}{2}$ at $\hat{p}_1 = 0.7288875$: $W(\rho_1) - W(0) - W_1\rho_1 \leq -0.148$ and ≤ -0.0142 respectively, with $\rho_1 = r(\hat{p}_1)$. By Theorem F.20(iii), in both cases an open-loop strategic learner gains linearly in T while its flagging probability tends to the honest rate: no pooled audit with any stakes and fixed level disciplines it.*

Proof. With $\rho_1 = r(\hat{p}_1)$, $W(\rho_1) = V(\hat{p}_1)$, $W(0) = V(p^*) = (1-\alpha)\Phi(p^*)^2$ and $W_1 = 2(1-\alpha)\Phi(p^*)\Phi'(p^*)/(1+\Phi'(p^*))$. The certificate encloses p^* and evaluates $V(\hat{p}_1) - W(0) - W_1 r(\hat{p}_1)$ on intervals; the result is strictly negative. The rest is Theorem F.20(iii). □

Numerical illustration (the size of the effect, at $\alpha = \frac{1}{5}$). For this world the strategic learner of Theorem F.20(iii) alternates between $\hat{p} = 0.4515$ (bias -0.291) and $\hat{p} = 1.0663$ (bias $+0.566$), with weights 0.661 and 0.339. In simulation against the pooled audit at level $z = 1.96$ (honest flagging rate 0.050), its flagging rate was 0.049–0.052 over $T \in \{10^3, 10^4, 10^5\}$; it gained 0.045 per round while users lost 0.165 per round (T35).

F.3 The binned audit

Theorem F.20(iii) shows that a pooled audit which pools all rounds cannot see a strategic learner who alternates between biases of opposite sign. The natural repair is the forecasting notion of calibration: test the bias *conditional on the forecast*. This subsection proves that such an audit disciplines manipulation in every world with a Lipschitz monotone response, without any convexity or tangent condition, and characterises the price.

Definition F.22 (Binned pooled audit)**.** Parameters: bin width $w > 0$, a minimum count $n_{\min}$, a level z, and an offset $U \sim \mathrm{Unif}[0, w)$ drawn by the auditor independently of the noise. The bins are $B_j := [U + (j-1)w, U + jw) \cap \mathcal{P}$; at most $M \le P/w + 2$ are nonempty. With $n_j := \#\{t \le T : \hat{p}_t \in B_j\}$ and $\Sigma_j := \sum_{t:\hat{p}_t \in B_j} (y_t - \hat{p}_t)$, a bin is *tested* if $n_j \ge n_{\min}$, and the audit flags the learner if some tested bin has $|\Sigma_j| > z\sigma\sqrt{n_j}$. The default choice is, with $\ell_T := \ln T \cdot \ln\ln T$,

$$w_T := T^{-1/3}\ell_T^{1/3}, \qquad n_{\min,T} := \ell_T / w_T, \qquad z_T := \sqrt{2\ln(2MT)} + \sqrt{\ell_T}.$$

The first part of the level controls the noise in every bin at once; the second, growing slightly faster than $\sqrt{\ln T}$, absorbs the bias of an honest learner with logarithmic regret (Theorem F.28).

Theorem F.23 (The binned audit disciplines manipulation in every Lipschitz world)**.** *Let $\Phi \ge 0$ be nondecreasing on $\mathcal{P} = [0, P]$ with r Lipschitz (constant L_r), let $V = \alpha r^2 + (1-\alpha)\Phi^2$ with Lipschitz constant L_V on $\mathcal{P}$, $\rho_{\max} := \max_{\mathcal{P}} |r|$ and $\Delta_V := V(p^*) - \min_{\mathcal{P}} V$. The learner knows μ, Φ, σ and the bins, and may choose $\hat{p}_t$ adaptively and at random; noise is $N(0, \sigma^2)$. Being flagged costs κT. The honest full-information learner plays p^* throughout and is flagged with probability $\beta_h \le 2\bar{\Phi}_{\mathcal{N}}(z)$. Let $G := \sum_t (V(p^*) - V(\hat{p}_t))$ and $NB := \mathbb{E}G - \kappa T\,\mathbb{P}(\text{flag}) + \kappa T\beta_h$ (the net benefit over honest play). Then, for every policy, with*

$$\Gamma_T := (z+1)\,\sigma\sqrt{MT} + L_r w T + \rho_{\max} M n_{\min},$$

(i) $\mathbb{E}[G\,\mathbf{1}\{\text{no flagging}\}] \le L_V \Gamma_T$;

(ii) $NB - \kappa T\beta_h \le L_V\Gamma_T + (\Delta_V - \kappa)\,T\,\mathbb{P}(\text{flag})$;

(iii) if $\kappa > \Delta_V$, every optimal policy has $\mathbb{P}(\text{flag}) \le (\kappa T\beta_h + L_V\Gamma_T)/((\kappa - \Delta_V)T)$, and both its gain $\mathbb{E}G$ and the harm to users $\mathbb{E}\sum_t r(\hat{p}_t)^2$ are $O(\Gamma_T + T\beta_h)$;

(iv) with the default choice, $T\beta_h \to 0$ and $\Gamma_T = O\big(T^{2/3}\ell_T^{1/3}\big) = O\big(T^{2/3}(\ln T\,\ln\ln T)^{1/3}\big)$.

No convexity of Φ and no tangent condition is needed; compare Theorem F.20(iii).

Proof. Write $\rho_t := r(\hat{p}_t)$ and $N_j := \sum_{t \le T} \eta_t \mathbf{1}\{\hat{p}_t \in B_j\}$, so $\Sigma_j = N_j - \sum_{t:\hat{p}_t \in B_j} \rho_t$.

Gain per round. By Lemma B.1(iii) with $m = 1$, $|\hat{p}_t - p^*| \le |\rho_t|$; so $V(p^*) - V(\hat{p}_t) \le \min(\Delta_V, L_V|\rho_t|)$ and $G \le L_V \sum_t |\rho_t|$.

Within a bin the bias is nearly constant. Two forecasts in the same bin differ by less than w, so their biases differ by at most $L_r w$. With $\bar{\rho}_j$ the average bias in bin j, $\sum_{t \in B_j} |\rho_t| \le n_j|\bar{\rho}_j| + L_r w n_j$.

On the event of no flagging. For a tested bin, $n_j|\bar{\rho}_j| = |N_j - \Sigma_j| \le z\sigma\sqrt{n_j} + |N_j|$. Untested bins hold at most $Mn_{\min}$ rounds in total, each with $|\rho_t| \le \rho_{\max}$. Hence, on that event,

$$\textstyle\sum_t |\rho_t| \le z\sigma \sum_j \sqrt{n_j} + \sum_j |N_j| + L_r w T + \rho_{\max} M n_{\min}.$$

Expectations. $\sum_j \sqrt{n_j} \le \sqrt{MT}$ by Cauchy–Schwarz, since $\sum_j n_j = T$. Each N_j is a sum of martingale differences ($\hat{p}_t$ is chosen before η_t, and the offset is independent of the noise), so $\mathbb{E}N_j^2 = \sigma^2 \mathbb{E}n_j$ and $\sum_j \mathbb{E}|N_j| \le \sigma \sum_j \sqrt{\mathbb{E}n_j} \le \sigma\sqrt{MT}$. This gives (i).

(ii) $\mathbb{E}G \le \mathbb{E}[G\mathbf{1}\{\text{no flag}\}] + \Delta_V T\, \mathbb{P}(\text{flag})$; substitute into NB.

(iii) Honest play is feasible with $NB = 0$, so an optimal policy has $NB \ge 0$; rearranging (ii) bounds $\mathbb{P}(\text{flag})$, and then $\mathbb{E}G \le L_V\Gamma_T + \Delta_V T\, \mathbb{P}(\text{flag})$. For the harm, $\sum_t \rho_t^2 \le \rho_{\max} \sum_t |\rho_t|$ on the event of no flagging and $\le \rho_{\max}^2 T$ otherwise.

(iv) The honest learner's T rounds lie in one bin, which is tested once $T \ge n_{\min}$, and its statistic is $\sum_t \eta_t \sim N(0, T\sigma^2)$. Since $z_T \ge z_0 := \sqrt{2\ln(2MT)}$, Mills' inequality gives $\beta_h \le 2\varphi(z_0)/z_0 = 1/(\sqrt{2\pi}\, z_0 MT)$, so $T\beta_h \to 0$. With $M \le P/w_T + 2$: $\sqrt{MT} = O(T^{2/3}\ell_T^{-1/6})$ and $z_T = O(\ell_T^{1/2})$, so the first term of Γ_T is $O(T^{2/3}\ell_T^{1/3})$; $L_r w_T T = T^{2/3}\ell_T^{1/3} L_r$; and $M n_{\min} = O(\ell_T / w_T^2) = O(T^{2/3}\ell_T^{1/3})$. □

Corollary F.24 (The binned audit catches the strategic learner the pooled audit misses)**.** *Let the world satisfy the hypotheses of Theorem F.23 and suppose the tangent condition fails, so that Theorem F.20(iii) provides the open-loop alternating strategic learner, with biases $\rho_- < 0 < \rho_+$, weight λ and $g_0 > 0$. Against the pooled audit its flagging probability tends to the honest rate (Theorem F.20(iii)). Against the binned audit with the default parameters,*

$$\mathbb{P}(\textit{flag}) \;\ge\; 1 - \frac{L_V\Gamma_T}{g_0 T - |W(\rho_+) - W(\rho_-)|} \;\longrightarrow\; 1.$$

This applies in particular to the world of Proposition F.21.

Proof. The strategic learner's forecasts are fixed in advance, so its gain $G = \sum_t \big(V(p^*) - V(\hat{p}_t)\big) = \sum_t \big(W(0) - W(\rho_t)\big)$ is deterministic: with $n_- = \lfloor \lambda T \rfloor$ rounds at ρ_-, $G = Tg_0 - (\lambda T - n_-)\big(W(\rho_+) - W(\rho_-)\big) \ge g_0 T - |W(\rho_+) - W(\rho_-)|$. By Theorem F.23(i), $G \cdot \mathbb{P}(\text{no flagging}) = \mathbb{E}[G\,\mathbf{1}\{\text{no flagging}\}] \le L_V\Gamma_T$. Dividing gives the bound, and $\Gamma_T = O(T^{2/3}(\ln T \ln\ln T)^{1/3}) = o(T)$. □

Numerical illustration (the binned audit at the default parameters). In simulation, the alternating strategic learner of Proposition F.21 was flagged by the binned audit in all 200 runs at each of $T = 10^3, 10^4, 10^5$, and by the pooled audit at the same level in none (T41, T45). Honest RSA in the smooth quadratic world, with the tuned gain and with $a = 1$, was flagged in none of 200 runs. Honest RSA ($a = 1$) in the gap was flagged in about 2.5% of runs at every horizon from 2,000 to 40,000 rounds, with no visible decline (3000 runs; T60), which is the behaviour Theorem F.36 explains; every one of those runs had the jump point close to a bin edge. The accuracy-optimal learner in the gap was flagged in all 200 runs (T45).

Remark F.25 (The price of robustness). The pooled audit achieves $O(\sqrt{T\ln T})$, but only under the tangent condition (Theorem F.20(i)); the binned audit achieves $O(T^{2/3}(\ln T \ln\ln T)^{1/3})$ in every Lipschitz world. The Lipschitz constant L_r enters linearly, so a sharp switch (large L_r, the regime of the failures in Corollary D.39) is handled but costs more: bins must eventually be narrower than the switch. Within the audit family of Definition F.22, $T^{2/3}$ is the best rate up to a logarithm for audits that accept honest learners (Theorem F.46).

Remark F.26 (The rate algebra, and the resolution trade-off). With the default parameters, $M \le P/w_T + 2$ and $z_T \sim \sqrt{\ell_T}$, the three terms of Γ_T satisfy, as $T \to \infty$ (T44, symbolic),

$$\frac{(z_T+1)\sigma\sqrt{MT}}{T^{2/3}\ell_T^{1/3}} \to \sqrt{P}\,\sigma, \qquad \frac{L_r w_T T}{T^{2/3}\ell_T^{1/3}} = L_r, \qquad \frac{\rho_{\max} M n_{\min}}{T^{2/3}\ell_T^{1/3}} \to P\rho_{\max}.$$

So no logarithm is lost: each term is of exact order $T^{2/3}\ell_T^{1/3}$. Two refinements follow. First, the default w_T does not use L_r (unknown to the auditor), so Γ_T is linear in L_r; if L_r were known, minimising the first two terms over w gives $w^\star = \frac{1}{2}\big(2Pz^2\sigma^2/L_r^2\big)^{1/3} T^{-1/3}$ and a value $\frac{3}{2}(2P)^{1/3}(L_r\sigma^2 z^2)^{1/3}T^{2/3}$, i.e. growth like $L_r^{1/3}$. Second, the minimum count trades honest against dishonest error explicitly.

Write $g_T := w_T n_{\min,T}$. Theorem F.51 bounds the honest learner's false-alarm probability in the gap by $(2 + M_\lambda/\lambda) g_T^{-1/2}$, while the untested-bin term of Γ_T is $\rho_{\max} P g_T / w_T^2$. A larger g_T protects honest learners better and leaves more room to strategic learners; the default $g_T = \ell_T$ is the largest choice that keeps all three terms of the same order. Finer bins are therefore not monotonically better: resolution beyond what the honest dynamics can support starts to flag honest learners, and the minimum count is what absorbs that.

Theorem F.28 shows that the binned audit accepts an honest learner in a smooth world, for a structural reason. The noise in each bin is controlled uniformly over bins, and by Cauchy–Schwarz the bias accumulated in any bin is at most $\sqrt{n_j}$ times the square root of the bias sum of squares inside that bin, which is bounded by the learner's regret. So any learner whose regret is small compared with σ^2 times the level squared passes.

Lemma F.27 (Noise in every bin at once)**.** *Let the forecasts be chosen adaptively ($\hat{p}_t$ decided before η_t), with Gaussian noise, and let the bins be fixed before play (random offset allowed, independent of the noise). For $z_0 := \sqrt{2\ln(2MT)}$,*

$$\mathbb{P}\big(\exists j : \ |N_j| > z_0 \sigma \sqrt{n_j}\big) \le \frac{1}{\sqrt{2\pi}\, z_0}, \qquad N_j := \sum_{t \le T} \eta_t \mathbf{1}\{\hat{p}_t \in B_j\}.$$

Proof. Fix j and let $\tau_1 < \tau_2 < \dots$ be the rounds at which the forecast lies in B_j. Each τ_k is a stopping time and $\{\tau_k = t\}$ is decided before η_t is drawn, so by optional skipping (Doob) the visit noises $\xi_k := \eta_{\tau_k}$ are i.i.d. $N(0, \sigma^2)$; extend them past the last visit by independent draws. Then $N_j = \sum_{k \le n_j} \xi_k$ with $n_j \le T$, so the event for bin j is contained in $\bigcup_{n \le T} \{|\sum_{k \le n} \xi_k| > z_0 \sigma \sqrt{n}\}$, each of probability $2\bar{\Phi}_{\mathcal{N}}(z_0) \le 2\varphi(z_0)/z_0$ (Mills). A union bound over $n \le T$ and the M bins gives $2MT\varphi(z_0)/z_0 = 1/(\sqrt{2\pi} z_0)$, since $\varphi(z_0) = (2\pi)^{-1/2}(2MT)^{-1}$. □

Theorem F.28 (The binned audit accepts honest learners with logarithmic regret)**.** *Consider any world and any learner, the binned audit with any bins, any $n_{\min}$ and level $z = z_0 + s$ with $z_0 = \sqrt{2\ln(2MT)}$ and $s > 0$. Then*

$$\mathbb{P}(\textit{flag}) \ \le \ \frac{1}{\sqrt{2\pi}\, z_0} + \mathbb{P}\Big(\max_{j \ \textit{tested}} \sum_{t : \hat{p}_t \in B_j} r(\hat{p}_t)^2 > s^2 \sigma^2 \Big) \ \le \ \frac{1}{\sqrt{2\pi}\, z_0} + \frac{\mathrm{Regret}_T}{s^2 \sigma^2}.$$

In particular, in any world satisfying Assumptions A.1 and A.2 (Gaussian) and (M1)–(M3), honest RSA(a) *with $2aR' > 1$ (any stable gain; Theorem B.21 removes the global condition $2am > 1$) has $\mathrm{Regret}_T \le C(1 + \ln T)$ (Theorem B.21), and with the default level ($s = \sqrt{\ell_T}$) its false-alarm probability is at most $1/(\sqrt{2\pi} z_0) + C(1 + \ln T)/(\sigma^2 \ln T \ln\ln T) = O(1/\ln\ln T)$. With the tuned gain of Corollary B.18, $\mathrm{Regret}_T/(\sigma^2 \ell_T) = (1 + o(1))/\ln T$ by Theorem B.17.*

Proof. Write $\rho_t := r(\hat{p}_t)$, so $\Sigma_j = N_j - \sum_{t \in B_j} \rho_t$. On the event of Lemma F.27's complement, $|N_j| \le z_0 \sigma \sqrt{n_j}$ for every bin. By Cauchy–Schwarz, $|\sum_{t \in B_j} \rho_t| \le \sqrt{n_j \sum_{t \in B_j} \rho_t^2}$, which is at most $s\sigma\sqrt{n_j}$ on the event that every tested bin has $\sum_{t \in B_j} \rho_t^2 \le s^2\sigma^2$. On both events every tested bin has $|\Sigma_j| \le (z_0 + s)\sigma\sqrt{n_j}$, so there is no flagging; this gives the first inequality. For the second, $\max_j \sum_{t \in B_j} \rho_t^2 \le \sum_{t \le T} \rho_t^2$, whose expectation is Regret_T; apply Markov's inequality. □

Remark F.29 (What the binned audit does and does not establish). Theorem F.23 disciplines manipulation relative to the *full-information* honest learner, whose false-alarm probability satisfies $T\beta_h \to 0$. An honest *learner* is accepted only with $\mathbb{P}(\text{flag}) = O((\ln T)^{-1/2})$ in smooth worlds (Theorem F.39) and $O((\ln T \ln\ln T)^{-1/2})$ in the gap (Theorem F.51). Under the linear stakes κT that the discipline results require, its expected penalty is therefore of order $\kappa T(\ln T)^{-1/2}$, far above the strategic learner's gain cap $O(T^{2/3}\ell_T^{1/3})$. So the two results do not combine into an incentive statement: that honesty is the better policy for an agent that must learn is open. The binned results are *detection* results. The pooled audit does not have this defect in smooth worlds, where $T\,\mathbb{P}_{\text{honest}}(\text{flag}) \to 0$ for $c > 1$ (Propositions F.9 and F.12). The question is settled in the linear world by Theorem F.30 (raising the noise part of the level by one factor of T inside the logarithm, and using the Hanson–Wright tail in place of Markov's inequality), and in smooth nonlinear worlds with $2am > 1$ and $aR' > 1$

by Theorem F.33 (exponential moments of the learner, Lemma F.32, plus a pathwise comparison with the linear process). In the gap the answer is negative: Theorem F.36 shows that every binned audit that could discipline manipulation flags honest RSA with probability of order $1/(\bar{n}w)$, so $T\,\mathbb{P}_{\text{honest}}(\text{flag}) \to \infty$.

Theorem F.30 (In the linear world the binned audit makes honest learning approximately optimal)**.** *Let $r(p) = R'(p - p^*)$ on $\mathbb{R}$ with $R' \geq 1$ and $p^* \in (0, P)$, so that $\Phi(p) = \mu - R'p^* + (R' - 1)p$ is nondecreasing; assume $\Phi \geq 0$ on $\mathcal{P}$ and Gaussian noise $N(0, \sigma^2)$. Use the binned audit of Definition F.22 with any bins of width w covering $\mathcal{P}$, any $n_{\min}$, and the* raised level

$$z'_T := z'_0 + \sqrt{\ell_T}, \qquad z'_0 := \sqrt{2\ln(2MT^2)}, \qquad \ell_T = \ln T \cdot \ln\ln T.$$

Let the honest learner be the unprojected RSA(a) with $c := aR' > \frac{1}{2}$ and a deterministic start (forecasts outside $\mathcal{P}$ lie in no bin). Then:

(i) $T\,\mathbb{P}_{\text{honest}}(\text{flag}) \to 0$. Precisely, for T large, $\mathbb{P}_{\text{honest}}(\text{flag}) \leq \big(\sqrt{2\pi}\,T z'_0\big)^{-1} + 3\exp(-\ell_T/K_3)$ for a constant K_3 depending on $(c, a, R', \hat{p}_1)$, and $\exp(-\ell_T/K_3) = T^{-(\ln\ln T)/K_3}$.

(ii) Theorem F.23 holds with this level, with $T\beta_h \to 0$ and $\Gamma_T = O\big(T^{2/3}(\ln T \ln\ln T)^{1/3}\big)$ for the default bins.

(iii) Let a policy's total cost be $\mathcal{C}(\mathcal{A}) := \mathbb{E}\sum_{t\leq T} V(\hat{p}_t) + \kappa T\,\mathbb{P}(\text{flag})$, with $\kappa > \Delta_V$. For every policy $\mathcal{A}$ of a full-information agent with forecasts in $\mathcal{P}$,

$$\mathcal{C}(\mathcal{A}) \;\geq\; \mathcal{C}(\text{honest RSA}) - L_V\Gamma_T - D_T, \qquad D_T = O\big(T^{\max(1-c,0)} + \ln T\big) + o(1).$$

Since $1 - c < \frac{1}{2}$, the advantage of any manipulation over honest learning *is $O(\Gamma_T) = O(T^{2/3}(\ln T \ln\ln T)^{1/3})$, sublinear, whereas without the audit a full-information strategic learner gains $\Delta_V T$ minus lower-order terms.*

Proof. **(i) Noise part.** The proof of Lemma F.27 with z'_0 in place of z_0 bounds the probability that some bin has $|N_j| > z'_0\sigma\sqrt{n_j}$ by $2MT\varphi(z'_0)/z'_0$, and $\varphi(z'_0) = (2\pi)^{-1/2}(2MT^2)^{-1}$, giving $(\sqrt{2\pi}\,Tz'_0)^{-1}$.

(i) Bias part. As in the proof of Theorem F.28, off the noise event a flagging requires some tested bin with $\sum_{t\in B_j}\rho_t^2 > \ell_T\sigma^2$, hence $Q_T := \sum_{t\leq T} r(\hat{p}_t)^2 > \ell_T\sigma^2$. By the proof of Theorem F.39, $\mathbb{E}Q_T \leq C_Q(1 + \ln T)$, so for T large $x := \ell_T\sigma^2 - \mathbb{E}Q_T \geq \ell_T\sigma^2/2$. Theorem F.39(i) gives $\mathbb{P}(Q_T \geq \mathbb{E}Q_T + x) \leq 2\exp\big(-\kappa\min\{x^2/(\sigma^4K_1(1+\ln T)),\, x/(\sigma^2K_2)\}\big) + \exp(-x^2/(K_2\sigma^2))$. Here $x^2/(\sigma^4K_1(1+\ln T)) \geq \ell_T^2/(4K_1(1+\ln T)) \geq \ell_T/(8K_1)$ once $\ln\ln T \geq 1$, and $x/(\sigma^2K_2) \geq \ell_T/(2K_2)$; the last term is smaller still. So the bias part is at most $3\exp(-\ell_T/K_3)$ with $K_3 := \max(8K_1, 2K_2)/\min(\kappa, 1)$, and $T \cdot T^{-(\ln\ln T)/K_3} \to 0$.

(ii) Γ_T depends on the level only through the factor $(z + 1)$, and $z'_T \leq \sqrt{2}\,z_T$ for the default z_T; so Γ_T keeps its order. $\beta_h \leq 2\bar{\Phi}_{\mathcal{N}}(z'_T) \leq 2\bar{\Phi}_{\mathcal{N}}(z'_0) \leq 2\varphi(z'_0)/z'_0$, so $T\beta_h \leq (\sqrt{2\pi}\,MTz'_0)^{-1} \to 0$.

(iii) Any policy. Theorem F.23(ii) gives $\mathbb{E}G - \kappa T\,\mathbb{P}_{\mathcal{A}}(\text{flag}) \leq L_V\Gamma_T + (\Delta_V - \kappa)T\,\mathbb{P}_{\mathcal{A}}(\text{flag})$, i.e. $\mathbb{E}G \leq L_V\Gamma_T + \Delta_V T\,\mathbb{P}_{\mathcal{A}}(\text{flag})$. Hence $\mathcal{C}(\mathcal{A}) = TV(p^*) - \mathbb{E}G + \kappa T\,\mathbb{P}_{\mathcal{A}}(\text{flag}) \geq TV(p^*) - L_V\Gamma_T + (\kappa - \Delta_V)T\,\mathbb{P}_{\mathcal{A}}(\text{flag}) \geq TV(p^*) - L_V\Gamma_T$.

(iii) The honest learner. In the linear world V is a quadratic on $\mathbb{R}$, so $V(\hat{p}_t) - V(p^*) = V'(p^*)e_t + \frac{1}{2}V''e_t^2$ with $e_t := \hat{p}_t - p^*$. By the representation in the proof of Theorem F.39, $\mathbb{E}e_t = m_t$ with $|m_t| \leq Ct^{-c}$, and $\mathbb{E}e_t^2 \leq m_t^2 + \sigma^2\sum_s G_{ts}^2 \leq C/t$. Summing, $\mathbb{E}\sum_t(V(\hat{p}_t) - V(p^*)) \leq C\sum_t t^{-c} + C\sum_t t^{-1} = O(T^{\max(1-c,0)} + \ln T)$, and $\kappa T\,\mathbb{P}_{\text{honest}}(\text{flag}) = o(1)$ by (i). So $\mathcal{C}(\text{honest RSA}) \leq TV(p^*) + D_T$, and combining with the previous step gives the claim. □

Remark F.31 (Scope of Theorem F.30). The rate in (i) is asymptotic in a slow way: it needs $\mathbb{E}Q_T \leq \ell_T\sigma^2/2$, and $\mathbb{E}Q_T/(\sigma^2\ell_T)$ decays only like $1/\ln\ln T$. In simulation (T58), $\mathbb{E}Q_T/\sigma^2 \approx 19$ against $\ell_T \approx 20$ at $T = 10^4$; no false alarm occurred at either level for $T \leq 4 \times 10^4$, because the Cauchy–Schwarz step in the audit proof is loose, not because the margin is active. The honest learner here is unprojected; the strategic learner is restricted to $\mathcal{P}$. Smooth nonlinear worlds are treated next (Theorem F.33); the gap regime remains open (Remark F.29).

Lemma F.32 (Exponential moments of RSA)**.** *Assume Gaussian noise $N(0,\sigma^2)$ independent of the past, (M2) with modulus m, (M3) with constant L (so $|r(p)| \le L|p-p^*|$), $p^* \in \mathcal{P}$, and $2am > 1$. Let $e_t := \hat{p}_t - p^*$ for projected RSA(a) with a deterministic start. For every $\theta \in (0, \theta_0]$, with $\theta_0 := \min\{(2am-1)/(16a^2\sigma^2),\ 1/(8a^2\sigma^2)\}$,*

$$\sup_{t\ge 1} \mathbb{E}\exp(\theta t e_t^2) \le C_\theta < \infty.$$

Consequently $\mathbb{P}\big(\exists t \le T :\ te_t^2 > (3/\theta)\ln T\big) \le C_\theta T^{-2}$.

Proof. **One step.** Let $w_t := e_t - \frac{a}{t}(r(\hat{p}_t) - \eta_t) = A_t + B_t\eta_t$ with $A_t := e_t - \frac{a}{t}r(\hat{p}_t)$ and $B_t := a/t$; projection onto $\mathcal{P} \ni p^*$ is nonexpansive, so $|e_{t+1}| \le |w_t|$. By (M2)–(M3), $A_t^2 = e_t^2 - \frac{2a}{t}e_t r(\hat{p}_t) + \frac{a^2}{t^2}r(\hat{p}_t)^2 \le e_t^2\big(1 - \frac{2am}{t} + \frac{a^2L^2}{t^2}\big)$.

Gaussian integral. For $\lambda > 0$ with $\gamma := 2\lambda B^2\sigma^2 < 1$, $\mathbb{E}\exp(\lambda(A+B\eta)^2) = (1-\gamma)^{-1/2}\exp(\lambda A^2/(1-\gamma))$. Take $\lambda := \theta(t+1)$; then $\gamma_t = 2\theta(t+1)a^2\sigma^2/t^2 \le 4\theta a^2\sigma^2/t \le \frac{1}{2}$, since $\theta \le 1/(8a^2\sigma^2)$. Using $1/(1-\gamma) \le 1+2\gamma$ and $(1-\gamma)^{-1/2} \le e^{\gamma}$ for $\gamma \le \frac{1}{2}$,

$$\mathbb{E}\big[e^{\theta(t+1)e_{t+1}^2} \,\big|\, \mathcal{F}_t\big] \le \exp\left(\theta t e_t^2\left(1+\frac{1}{t}\right)\left(1-\frac{2am}{t}+\frac{a^2L^2}{t^2}\right)\left(1+\frac{8\theta a^2\sigma^2}{t}\right) + \frac{4\theta a^2\sigma^2}{t}\right).$$

The product of the three brackets is $1 - (2am - 1 - 8\theta a^2\sigma^2)/t + O(t^{-2})$. With $\delta := 2am - 1 - 8\theta a^2\sigma^2 \ge (2am-1)/2 > 0$ for $\theta \le \theta_0$, there is t_0 such that for $t \ge t_0$ the product is at most $1 - \delta/(2t)$.

Recursion. With $X_t := \theta t e_t^2$ and $\phi_t := \ln \mathbb{E}e^{X_t}$, for $t \ge t_0$, taking expectations and using Jensen's inequality for the concave map $u \mapsto u^{1-\delta/(2t)}$: $\phi_{t+1} \le (1 - \frac{\delta}{2t})\phi_t + \frac{4\theta a^2\sigma^2}{t}$. By induction $\phi_t \le \max(\phi_{t_0}, 8\theta a^2\sigma^2/\delta)$ for $t \ge t_0$, and $\phi_t \le \theta t P^2$ for $t \le t_0$ since $|e_t| \le P$. So $C_\theta := \exp\max(\theta t_0 P^2, 8\theta a^2\sigma^2/\delta)$ works.

Consequence. By Markov's inequality $\mathbb{P}(X_t > 3\ln T) \le C_\theta T^{-3}$; a union bound over $t \le T$ gives $C_\theta T^{-2}$. □

Theorem F.33 (The incentive result in smooth nonlinear worlds)**.** *Assume Gaussian noise, (M1)–(M3) with $2am > 1$, p^* interior to $\mathcal{P}$, $r \in C^2(\mathcal{P})$ with $R' := r'(p^*)$, and projected honest RSA(a) with $c := aR' > 1$ and a deterministic start. Use the binned audit with the raised level z_T' of Theorem F.30. Then:*

(i) $T\,\mathbb{P}_{\text{honest}}(\textit{flag}) \to 0$*;*

(ii) for every policy $\mathcal{A}$ of a full-information agent with forecasts in $\mathcal{P}$, and $\kappa > \Delta_V$, $\mathcal{C}(\mathcal{A}) \ge \mathcal{C}(\text{honest RSA}) - L_V\Gamma_T - D_T$ with $D_T = O(T^{1/4} + \ln T) + o(1)$; the advantage of any manipulation over honest learning is $O(T^{2/3}(\ln T \ln\ln T)^{1/3})$.

Proof. Write $q(e) := r(p^*+e) - R'e$, so $|q(e)| \le Ke^2$ on $\mathcal{P}$ with $\mathcal{P} := \frac{1}{2}\max_{\mathcal{P}}|r''|$. Let $\tilde{e}$ be the linear comparison process of Theorem F.39: unprojected, $\tilde{e}_1 = e_1$, $\tilde{e}_{t+1} = (1-c/t)\tilde{e}_t + (a/t)\eta_t$, with the same noise; $d_t := e_t - \tilde{e}_t$; $\tilde{Q}_T := R'^2\sum_{t\le T}\tilde{e}_t^2$.

Good events. $E_1 := \{te_t^2 \le C_1 \ln T\ \forall t \le T\}$ with $C_1 := 3/\theta_0$, of probability at least $1 - CT^{-2}$ (Lemma F.32). $E_2 := \{t\tilde{e}_t^2 \le C_2 \ln T\ \forall t \le T\}$: $\tilde{e}_t$ is Gaussian with mean m_t, $|m_t| \le Ct^{-c}$, and variance at most C/t (proof of Theorem F.39), so a Gaussian tail and a union bound give probability at least $1 - 2T^{-2}$ for suitable C_2. $E_3 := \{\max_{t\le T}|\eta_t| \le \sigma\sqrt{6\ln T}\}$, of probability at least $1 - T^{-2}$.

Projection is inactive after $\tau := C_3 \ln T$. On $E_1 \cap E_3$, the unprojected point satisfies $|w_t| \le |e_t| + \frac{a}{t}(L|e_t| + |\eta_t|) \le (1+aL)\sqrt{C_1 \ln T/t} + a\sigma\sqrt{6\ln T}/t$, which is below $\operatorname{dist}(p^*, \partial\mathcal{P})$ for $t \ge \tau$ with C_3 large. So for $t \ge \tau$: $e_{t+1} = (1-c/t)e_t - \frac{a}{t}q(e_t) + \frac{a}{t}\eta_t$, and $d_{t+1} = (1-c/t)d_t - \frac{a}{t}q(e_t)$.

The gap d_t. For $t \ge \max(\tau, c)$, on E_1, $|d_{t+1}| \le (1-c/t)|d_t| + aKC_1\ln T/t^2$. With $u_t := t|d_t|$, $u_{t+1} \le (1+\frac{1}{t})(1-\frac{c}{t})u_t + \frac{2aKC_1\ln T}{t} \le (1-\frac{c-1}{t})u_t + \frac{2aKC_1\ln T}{t}$, so

$u_t \le \max(u_\tau, 2aKC_1 \ln T/(c-1))$, using $c > 1$. On $E_1 \cap E_2$, $|d_\tau| \le |e_\tau| + |\tilde{e}_\tau| \le (\sqrt{C_1} + \sqrt{C_2})\sqrt{\ln T/\tau} = O(1)$, so $u_\tau = O(\ln T)$. Hence $|d_t| \le C_4 \ln T/t$ for $t \ge \tau$, and $\sum_{t>\tau} d_t^2 \le C_4^2 (\ln T)^2/\tau = O(\ln T)$.

The squared bias. On $E_1 \cap E_2 \cap E_3$: for $t \le \tau$, $r(\hat{p}_t)^2 \le L^2 P^2$, contributing $O(\ln T)$. For $t > \tau$, $r^2 \le 2R'^2 e_t^2 + 2\mathcal{P}^2 e_t^4$, with $e_t^2 \le 2\tilde{e}_t^2 + 2d_t^2$ and $\sum_{t>\tau} e_t^4 \le \sum_{t>\tau} C_1^2 (\ln T)^2/t^2 \le C_1^2 (\ln T)^2/\tau = O(\ln T)$. Altogether $\sum_{t \le T} r(\hat{p}_t)^2 \le 4\tilde{Q}_T + C_5 \ln T$.

(i). As in the proof of Theorem F.30, off the noise event (probability at most $(\sqrt{2\pi} T z_0')^{-1}$) a flagging requires $\sum_t r(\hat{p}_t)^2 > \ell_T \sigma^2$. Since $C_5 \ln T \le \ell_T \sigma^2/2$ for T large, this requires either the complement of $E_1 \cap E_2 \cap E_3$ (probability $O(T^{-2})$) or $\tilde{Q}_T > \ell_T \sigma^2/8$, whose probability is $O(\exp(-\ell_T/K_3))$ by the argument of Theorem F.30(i) applied to $\tilde{Q}_T$ (a linear process with $c > \frac{1}{2}$). All three are $o(1/T)$.

(ii). The lower bound on $\mathcal{C}(\mathcal{A})$ is Theorem F.30(iii)'s first step, which uses only Theorem F.23. For the honest learner, $V \in C^2(\mathcal{P})$ gives $V(\hat{p}_t) - V(p^*) \le |V'(p^*)||\mathbb{E}e_t| + C\,\mathbb{E}e_t^2$ in expectation. $\mathbb{E}e_t^2 \le C/t$ (Lemma B.12); $|\mathbb{E}e_t| \le |m_t| + \mathbb{E}|d_t| \le Ct^{-c} + Ct^{-3/4}$, using $\|d_t\|_2 \le Ct^{-3/4}$ from the proof of Proposition F.12(ii) (valid for $c > 1$). Summing gives $O(T^{1/4} + \ln T)$, and $\kappa T\, \mathbb{P}_{\text{honest}}(\text{flag}) \to 0$ by (i). □

Remark F.34 (Why the nonlinear theorem needs $aR' > 1$). In the linear world, stability of the stochastic approximation ($aR' > \frac{1}{2}$) suffices, because the whole squared bias is a Gaussian quadratic form. In a nonlinear world the learner must also *track* the linear comparison process pathwise, and the tracking error d_t is driven by the nonlinear remainder, of size $e_t^2 \approx \ln T/t$. The recursion $|d_{t+1}| \le (1 - aR'/t)|d_t| + C \ln T/t^2$ keeps $t|d_t|$ bounded only when its contraction beats the $1/t$ scaling, i.e. when $aR' > 1$; for $aR' \le 1$ the tracking error decays too slowly for its squares to sum to $O(\ln T)$.

Lemma F.35 (Honest RSA puts about $1/d$ rounds at distance d)**.** *Under the assumptions of Lemma F.44, with the same constants $c_{\text{occ}}, p_{\text{occ}}$ and w_1: for every $d \in [2a\sigma/T, w_1]$,*

$$\mathbb{P}\big(\#\{t \le T : \hat{p}_t \in [x_0 + d,\, x_0 + 2d]\} \ge c_{\text{occ}}/d\big) \ge p_{\text{occ}},$$

and every round in $[x_0 + d, x_0 + 2d]$ has bias at least $r^+ > 0$.

Proof. The proof of Lemma F.44 uses the bin width w only as a length scale: the target interval has length w and lies within $2w$ of x_0, and the time window W is where the step's noise has standard deviation $s_t = a\sigma/t \in [w, 2w]$. Replace w by d throughout: target $J_d := [x_0 + d, x_0 + 2d]$ (length d, within $2d$ of x_0, interior to $\mathcal{P}$ for $d \le w_1$), window $W_d := \{t : a\sigma/(2d) \le t \le a\sigma/d\} \subset [1, T]$ (as $d \ge 2a\sigma/T$), where $s_t \in [d, 2d]$. On $\{|x_t| \le \mathcal{P}\}$, $|m_t - x_0| \le \kappa d$, every point of J_d is within $(\kappa + 2)d$ of m_t, and the Gaussian density on J_d is at least $\varphi(\kappa + 2)/(2d)$; so each step lands in J_d with conditional probability at least p_1. The counting step is unchanged and gives at least $p_1|W_d|/4 \ge c_{\text{occ}}/d$ rounds with probability at least p_{occ}. The bias statement is Lemma C.2. □

Theorem F.36 (In the gap no binned audit makes honest learning approximately optimal)**.** *Assume (G1)–(G5), Gaussian noise, and honest RSA(a) with $a \ge 1$. Consider any binned audit (Definition F.22) with width $w \le w_1$, minimum count $n_{\min}$ and level z, with the offset uniform and independent of the play. Let $r_{\min} := \min(|r^-|, r^+)$ and*

$$\bar{n} := \max\big(n_{\min},\ 4z^2\sigma^2/r_{\min}^2 + 1,\ n_1\big),$$

where n_1 is a constant depending only on $(\sigma, r_{\min}, p_{\text{occ}}, c_{\text{occ}}, w_1)$. If $\bar{n} \le c_{\text{occ}} T/(4a\sigma)$, then

$$\mathbb{P}_{\text{honest}}(\textit{flag}) \ \ge\ \frac{p_{\text{occ}}}{4} \min\Big(1,\ \frac{c_{\text{occ}}}{\bar{n}\, w}\Big).$$

Consequently: (i) for the default audit, $\bar{n}w = \ell_T$ for large T, so $\mathbb{P}_{\text{honest}}(\textit{flag}) \ge c/\ell_T$ and $T\,\mathbb{P}_{\text{honest}}(\textit{flag}) \to \infty$; (ii) for every binned audit with $\bar{n}w = o(T)$, $T\,\mathbb{P}_{\text{honest}}(\textit{flag}) \to \infty$. The incentive property of Theorems F.30 and F.33 fails in the gap for every such audit.

Proof. **The edge distance.** Let J be the bin immediately to the right of the bin containing x_0, and $d \in (0, w]$ the distance from x_0 to their common edge. Since the offset is uniform on $[0, w)$ and independent of the play, d is uniform on $(0, w]$ and independent of the learner. Let $A := \{2a\sigma/T \le d \le \min(c_{\mathrm{occ}}/\bar{n}, w)\}$. As $c_{\mathrm{occ}}/\bar{n} \ge 4a\sigma/T$, $\mathbb{P}(A) \ge \frac{1}{2}\min(1, c_{\mathrm{occ}}/(\bar{n}w))$.

Occupancy. On A, $[x_0 + d, x_0 + 2d] \subset J$ (as $d \le w$), so by Lemma F.35, applied conditionally on the offset, with probability at least p_{occ} the bin J receives $n_J \ge c_{\mathrm{occ}}/d \ge \bar{n} \ge n_{\min}$ rounds: it is tested. Every round in J lies to the right of x_0 and has $r(\hat{p}_t) \ge r^+ \ge r_{\min}$ (Lemma C.2), so $\sum_{t\in J} r(\hat{p}_t) \ge n_J r_{\min}$.

Noise. As in Lemma F.27, the visit noises of J are i.i.d. $N(0, \sigma^2)$ by optional skipping. Let $C := \{\exists n \ge \bar{n} : |\sum_{k\le n} \xi_k| > n r_{\min}/2\}$; by the Gaussian tail and a union bound, $\mathbb{P}(C) \le \sum_{n\ge\bar{n}} 2e^{-n r_{\min}^2/(8\sigma^2)} \le C_0 e^{-\bar{n} r_{\min}^2/(8\sigma^2)}$ with $C_0 := 2/(1 - e^{-r_{\min}^2/(8\sigma^2)})$.

Declaration. On A, the occupancy event, and C^c: $|S_J| \ge n_J r_{\min} - n_J r_{\min}/2 = n_J r_{\min}/2 > z\sigma\sqrt{n_J}$, because $n_J \ge \bar{n} > 4z^2\sigma^2/r_{\min}^2$. So the audit flags. Hence $\mathbb{P}(\text{flag}) \ge p_{\mathrm{occ}}\mathbb{P}(A) - \mathbb{P}(C) \ge \frac{p_{\mathrm{occ}}}{2}\min(1, \frac{c_{\mathrm{occ}}}{\bar{n}w}) - C_0 e^{-\bar{n} r_{\min}^2/(8\sigma^2)}$. Choose n_1 so that $C_0 e^{-n r_{\min}^2/(8\sigma^2)} \le \frac{p_{\mathrm{occ}}}{4}\min(1, \frac{c_{\mathrm{occ}}}{n w_1})$ for all $n \ge n_1$ (possible: exponential against polynomial decay); since $w \le w_1$ the bound follows.

(i), (ii). For the default audit $z^2 = O(\ell_T)$ while $n_{\min} = \ell_T/w_T = T^{1/3}\ell_T^{2/3}$, so $\bar{n} = n_{\min}$ and $\bar{n}w = \ell_T$ for large T, and the hypothesis $\bar{n} \le c_{\mathrm{occ}}T/(4a\sigma)$ holds. In (ii), $\bar{n}w = o(T)$ and $w \le w_1$ give $\bar{n} = o(T)$, so the hypothesis holds for large T, and $T\,\mathbb{P}_{\mathrm{honest}}(\text{flag}) \ge \frac{p_{\mathrm{occ}}}{4}\min(T, c_{\mathrm{occ}}T/(\bar{n}w)) \to \infty$. □

Remark F.37 (Audits with $\bar{n}w$ of order T, and what might work instead). An audit with $\bar{n}w \ge cT$ either tests only bins holding order T/w rounds ($n_{\min} \ge cT/w$) or uses a level $z \ge c'\sqrt{T/w}$. The case analysis in the proof of Theorem F.46 (cases (a) and (c)) shows such audits admit manipulation with linear gain: a strategic learner spreads its rounds over bins each below the testing threshold, or holds a constant distortion that stays under the level. So within the family, making honesty approximately optimal in the gap and disciplining manipulation are incompatible. The obstruction is *where the bin edges fall* relative to the jump, not occupation balance within the jump's own bin: when x_0 lies within about $1/\bar{n}$ of an edge, the learner's overshoots fill the neighbouring bin with one-signed bias. Audits whose bins are placed after play, away from where forecasts concentrate, might avoid this; but Theorem F.23 assumes bins fixed before play, and such audits are open.

Theorem F.38 (The transient does not matter)**.** *In the setting of Theorem F.28 for honest RSA (moments of order 6, so Lemma B.12 holds with $k = 3$), let $L_\delta := \sup_{0<|e|\le\delta} |r(p^* + e)|/|e|$ (so $L_\delta \to R'$ as $\delta \to 0$). For every $D > 0$,*

$$\mathbb{P}(\textit{flag}) \le \frac{1}{\sqrt{2\pi}z_0} + \frac{3M_3'^{1/3}}{2D^2 n_{\min}} + \frac{L_{D+w}^2}{s^2\sigma^2}\sum_{t\le T}\min\big(\mathbb{E}(\hat{p}_t - p^*)^2, (D+w)^2\big).$$

With the default parameters and $D := \sqrt{w_T}$, the second term is $O(1/\ell_T)$ and the sum in the third is at most $M_1'(1 + \ln^+(T(D+w)^2/M_1'))$. Asymptotically (by Theorem B.17) the sum is $\frac{a^2\sigma^2}{2c-1} \cdot \frac{2}{3}\ln T\,(1 + o(1))$ with $c = aR'$, so the third term tends to $\frac{2}{3}\,g(c)/\ln\ln T$ with $g(c) = c^2/(2c-1)$, independent of the starting point.

Proof. **Tested bins lie near p^*.** Let $E_1 := \{\#\{t \le T : |\hat{p}_t - p^*| \ge D\} < n_{\min}\}$. A bin lying entirely at distance at least D from p^* collects only such rounds, so on E_1 it is untested. Every tested bin therefore meets $(p^* - D, p^* + D)$, and all its rounds satisfy $|\hat{p}_t - p^*| < D + w$.

E_1 is likely. By Chebyshev with the sixth moment, $\mathbb{P}(|\hat{p}_t - p^*| \ge D) \le M_3'/(t^3 D^6)$; summing $\min(1, \cdot)$ over t (split at $t^* = M_3'^{1/3}/D^2$) gives $\mathbb{E}\# \le \frac{3}{2}M_3'^{1/3}/D^2$, and Markov's inequality bounds $\mathbb{P}(E_1^c)$.

Bias in tested bins. On E_1, each tested bin has $\sum_{t\in B_j} r(\hat{p}_t)^2 \le L_{D+w}^2 \sum_{t\le T}\min((\hat{p}_t - p^*)^2, (D+w)^2)$. Take expectations using $\mathbb{E}\min(X, c) \le \min(\mathbb{E}X, c)$, then apply Markov's inequality and Lemma F.27 exactly as in Theorem F.28.

The sum. Rounds with $t \le M_1'/(D+w)^2$ contribute at most $(D+w)^2$ each, M_1' in total; later rounds contribute $M_1' \sum 1/t$. Asymptotically, $t\,\mathbb{E}(\hat{p}_t - p^*)^2 \to a^2\sigma^2/(2c-1)$ (Theorem B.17) and $\ln(T(D+w)^2) = \ln(Tw_T)(1+o(1)) = \frac{2}{3}\ln T(1+o(1))$. □

Theorem F.39 (Concentration of the examined bias). *(i)* ***Linear world.*** *Let $r(p) = R'(p-p^*)$ on $\mathbb{R}$ (no projection), Gaussian noise, and RSA(a) with $c := aR' > \frac{1}{2}$. Let $Q_T := \sum_{t\le T} r(\hat{p}_t)^2$. There are constants $c_{\mathrm{HW}}, K_1, K_2$ (depending on $c, a, R', \hat{p}_1$) with, for all $x > 0$,*

$$\mathbb{P}(Q_T \ge \mathbb{E}Q_T + x) \le 2\exp\Big(-c_{\mathrm{HW}}\min\Big(\frac{x^2}{\sigma^4 K_1(1+\ln T)}, \frac{x}{\sigma^2 K_2}\Big)\Big) + \exp\Big(-\frac{x^2}{K_2\sigma^2}\Big).$$

(ii) ***Smooth worlds.*** *Under the assumptions of Proposition F.12 with Gaussian noise and $c > \frac{1}{2}$, let $\tilde{Q}_T$ be the quantity of (i) for the linear comparison process. Then $\sum_t r(\hat{p}_t)^2 \le 4\tilde{Q}_T + Z_T$ with $\mathbb{E}Z_T \le C$ uniformly in T.*

Consequently, for honest RSA with any $c > \frac{1}{2}$ in any such world, the default audit's false-alarm probability is at most $1/(\sqrt{2\pi}z_0) + O(1/\ell_T) + \exp(-\Omega(\ell_T)) = O((\ln T)^{-1/2})$: the Markov rate $O(1/\ln\ln T)$ improves to the rate of the noise term.

Proof. **Representation.** Unrolling $e_{t+1} = (1-c/t)e_t + (a/t)\eta_t$ gives $e = m + G\eta$ with $m_t := e_1\prod_{u<t}(1-c/u)$ and $G_{ts} := (a/s)\prod_{u=s+1}^{t-1}(1-c/u)$ for $s < t$. Only finitely many factors (those with $u < c$) can exceed 1 in absolute value, and for $u > c$, $0 < 1 - c/u \le e^{-c/u}$; so $|G_{ts}| \le C'as^{c-1}t^{-c}$ and $|m_t| \le C't^{-c}$.

Operator norm: a weighted Schur test. With test weights $h_s := s^{-1/2}$: since $c - \frac{3}{2} > -1$, $\sum_{s<t}|G_{ts}|h_s \le C'at^{-c}\sum_{s<t}s^{c-3/2} \le C_1t^{-1/2} = C_1h_t$; and $\sum_{t>s}|G_{ts}|h_t \le C'as^{c-1}\sum_{t>s}t^{-c-1/2} \le C_2s^{-1/2} = C_2h_s$. Schur's test gives $\|G\| \le \sqrt{C_1C_2}$ for every $c > \frac{1}{2}$. This is the discrete form of Hardy's inequality for the operator $f \mapsto x^{-c}\int_0^x y^{c-1}f$, whose norm on L^2 is $1/(c-\frac{1}{2})$; T51 confirms the discrete constants are finite and of that order.

Frobenius norm and the deterministic part. $\|G\|_F^2 \le C'^2a^2\sum_t t^{-2c}\sum_{s<t}s^{2c-2} \le \mathcal{P}(1+\ln T)$ and $\|m\|^2 \le \mathcal{P}$, both because $2c > 1$.

Deviation. $Q_T/R'^2 = \|m\|^2 + 2\langle G^\top m, \eta\rangle + \eta^\top A\eta$ with $A := G^\top G$, $\|A\| = \|G\|^2$ and $\|A\|_F \le \|G\|\,\|G\|_F$. The linear term is Gaussian with variance at most $4\sigma^2\|G\|^2\|m\|^2$. The quadratic term obeys Hanson–Wright (Hanson & Wright, 1971; Rudelson & Vershynin, 2013): $\mathbb{P}(|\eta^\top A\eta - \mathbb{E}\eta^\top A\eta| > y) \le 2\exp(-c_{\mathrm{HW}}\min(y^2/(\sigma^4\|A\|_F^2), y/(\sigma^2\|A\|)))$. Split x equally. This proves (i).

(ii) Coupling. Write $e_t = \tilde{e}_t + d_t$ as in Proposition F.12 and $r(\hat{p}_t) = R'e_t + \rho(e_t)$ with $|\rho(e)| \le K_\rho e^2$. Then $r^2 \le 2R'^2e^2 + 2K_\rho^2e^4 \le 4R'^2\tilde{e}^2 + 4R'^2d^2 + 2K_\rho^2e^4$, so $Z_T := 4R'^2\sum_t d_t^2 + 2K_\rho^2\sum_t e_t^4$ works, with the factor 4 absorbed by writing $\tilde{Q}_T$ for $R'^2\sum\tilde{e}_t^2$. First, $\mathbb{E}\sum_t e_t^4 \le \sum_t M_2'/t^2$. Second, Minkowski's inequality in L^2 gives $\|d_{t+1}\|_2 \le (1-c/t)\|d_t\|_2 + (a/t)K_\rho\sqrt{M_2'}/t + \|\pi_t\|_2$ with $\|\pi_t\|_2 \le Ct^{-2}$ (moments of order 8). So $\|d_t\|_2 \le Ct^{-k'}$ for any $k' < \min(c, 1)$, and $\sum_t \mathbb{E}d_t^2 < \infty$ taking $k' > \frac{1}{2}$.

Consequence. Take $s^2 = \ell_T$ in Theorem F.28. $\mathbb{P}(\sum r^2 > \ell_T\sigma^2) \le \mathbb{P}(4\tilde{Q}_T > \ell_T\sigma^2/2) + \mathbb{P}(Z_T > \ell_T\sigma^2/2)$. The first is $\exp(-\Omega(\ell_T))$ by (i), once $\mathbb{E}\tilde{Q}_T \le \ell_T\sigma^2/16$, which holds for large T since $\mathbb{E}\tilde{Q}_T = O(\ln T)$; the second is at most $2C/(\ell_T\sigma^2)$. □

Remark F.40 (Why the crude condition fails and the audit still accepts). Three different quantities must be kept apart.

(a) *The theorem's sufficient condition*, total regret below $\ell_T\sigma^2$, is crude. In T45 it failed in 42–100% of runs up to $T = 10^5$, because realised regret includes an early transient of constant size (the transient-inflated means in T47 are 21–88 in units of σ^2, against $\ell_T = 13$–28).

(b) *The audit's actual per-tested-bin behaviour* is much tighter: the tested-bin event failed in at most 8.5% of runs, falling to 0–1% at $T = 10^5$.

(c) *The observed false-alarm rate* was 0 throughout.

There is no contradiction; the gap between (a) and (b) is itself informative. Theorem F.38 shows why: the transient's large residuals occur in bins the learner leaves quickly, so those bins are never tested, and the audit is insensitive to that regret. Only the regret accumulated near p^*, of order $\frac{2}{3}g(c)\sigma^2 \ln T$, is examined. Theorem F.39 shows that this examined quantity concentrates, in every smooth world and for every gain with $c > \frac{1}{2}$, so the remaining slack in (a) is Markov's inequality, not the audit. In T47 the standard deviation of Q_T/σ^2 did not grow with T (about 8 at $c = 1.5$, about 49 at $c = 3$).

Theorem F.41 (The $T^{2/3}$ rate is unavoidable for fixed-bin audits)**.** *Take the world with a concave kink at its honest forecast: $\Phi(p) = \gamma_1 p$ for $p \le p_0$ and $\Phi(p) = \gamma_1 p_0 + \gamma_2(p - p_0)$ for $p > p_0$, with $0 \le \gamma_2 < \gamma_1$, $\mu := p_0(1 + \gamma_1)$ (so $p^* = p_0 \in (0, P)$), Gaussian noise and $\alpha \in (0, 1)$. This world is monotone, Lipschitz and fixed (independent of T). Let the audit be any binned audit (Definition F.22) with parameters $(w, n_{\min}, z)$. There are constants $c_1, c_2, C_3, w_0 > 0$, depending only on the world, and a strategic learner who knows the bins, such that whenever $w \le w_0$ and $n_{\min}/w \le T/C_3$:*

(i) only one bin is tested, and in it the bias sum is at most $(1 + \gamma_1)w$ in absolute value, so the strategic learner is flagged with probability at most $2\bar{\Phi}_{\mathcal{N}}\big(z - (1+\gamma_1)w/(\sigma\sqrt{T/2})\big)$, essentially the honest learner's level;

(ii) its expected gain is at least $c_1 wT + c_2 n_{\min}/w - C_3$.

Consequently, for every audit of this family with $w\, n_{\min} = g$, the gain is at least $c' g^{1/3} T^{2/3} - C_3$ (minimise $wT + g/w^2$ over w). This alone gives $T^{2/3}$ only if g is bounded below; Theorem F.46 proves that accepting honest learners forces this, and so establishes the rate for the whole family.

Proof. V is piecewise quadratic, since r and Φ are piecewise linear: $V(p_0 \pm \delta) = V(p_0) \pm V'_{\pm}\delta + \frac{1}{2}V''_{\pm}\delta^2$ exactly, with $V'_{\pm} = 2(1-\alpha)\Phi(p_0)\gamma_{\pm} > 0$ ($r(p_0) = 0$), $\gamma_- = \gamma_1$, $\gamma_+ = \gamma_2$, and $V''_{\pm} = 2\alpha(1+\gamma_{\pm})^2 + 2(1-\alpha)\gamma_{\pm}^2$.

(a) Alternation inside the honest forecast's bin. Let d^* be the distance from p_0 to the nearest edge of its bin; it is uniform on $[0, w/2]$. Put $h := \min(0.9d^*, h_0)$ with $h_0 := A_1/(2A_2)$ as below. The strategic learner plays $p_0 - h$ in a fraction $\lambda := (1+\gamma_2)/(2+\gamma_1+\gamma_2)$ of these rounds and $p_0 + h$ otherwise. Then $\lambda(1+\gamma_1)h = (1-\lambda)(1+\gamma_2)h$: the biases $-(1+\gamma_1)h$ and $+(1+\gamma_2)h$ cancel on average, and after rounding the counts, the bin's bias sum is at most $(1+\gamma_1)h$. Both forecasts lie in the same bin. The gain per round is exactly

$$V(p_0) - \lambda V(p_0 - h) - (1-\lambda)V(p_0 + h) = hA_1 - h^2 A_2, \qquad A_1 := \frac{2(1-\alpha)\Phi(p_0)(\gamma_1 - \gamma_2)}{2 + \gamma_1 + \gamma_2} > 0,$$

with $A_2 := \frac{1}{2}[\lambda V''_- + (1-\lambda)V''_+]$. Since $h \le h_0$, this is at least $hA_1/2$; and $\mathbb{E}h \ge \min(0.9w/4, h_0/2)$, which is at least $0.2w$ for $w \le w_0$.

(b) Untested bins. For $0 < \delta \le V'_-/V''_-$, $V(p_0 - \delta) \le V(p_0) - V'_-\delta/2$. Fix $\varepsilon \le V'_-/(2V''_-)$ and let $I := [p_0 - 2\varepsilon, p_0 - \varepsilon]$, on which $V \le V(p_0) - g_0$ with $g_0 := V'_-\varepsilon/2$. At least $\varepsilon/w - 2$ bins lie inside I, all disjoint from p_0's bin when $w < \varepsilon$. The strategic learner places $n_{\min} - 1$ rounds in each, so they stay untested, gaining at least $g_0(n_{\min} - 1)(\varepsilon/w - 2)$.

(c) Accounting. Part (b) uses at most $(n_{\min} - 1)(\varepsilon/w + 1) \le T/2$ rounds when C_3 is large enough; the rest go to (a). Adding the gains of (a) and (b) gives (ii). For (i): the only tested bin is p_0's, with at least $T/2$ rounds, statistic $S = N - \sum \rho$, $N \sim N(0, \sigma^2 n)$ and $|\sum \rho| \le (1+\gamma_1)w$. □

Remark F.42 (What the lower bound does and does not say). Scope. The theorem is proved for the audit family of Definition F.22, with any parameters. Mechanism (a) keeps the tested bin's *sum* at pure noise, and (b) never reaches the testing threshold; but the strategic learner's bin *counts* differ from an honest learner's, so an audit whose statistic also uses the pattern of counts is not covered by this argument. Within the family, the $T^{2/3}$ barrier comes from resolution and from the exemption that protects honest learners in the gap, not from the level z. In particular, replacing the per-bin maximum by an aggregate such as a chi-square sum over bins would improve only the detection term $z\sigma\sqrt{MT}$. Against strategic learners who commit in advance, this suggests a rate near the classical nonparametric testing rate $T^{3/5}$ for the detection part (Ingster & Suslina, 2003). It does not touch

the two terms above, and against adaptive strategic learners, who can stop visiting a bin when its noise happens to favour them, aggregate statistics are less robust than the maximum over bins, which Lemma F.27 makes uniform over stopping. Audits outside this design (bins that refine over time, or tests of forecast dispersion within a bin) are not covered by Theorem F.41; whether they can beat $T^{2/3}$ while still accepting honest learners in the gap is open.

Remark F.43 (Learning the bin edges online does not help). Theorem F.23 already allows the strategic learner to know the bins exactly, and Theorem F.41's strategic learner uses that knowledge. A strategic learner who must learn the edges online is a special case of one who knows them, so it gains at most $L_V \Gamma_T$ as well. Conversely, knowledge of the edges is needed only for the within-bin alternation of Theorem F.41(a), which a strategic learner who knew only w could approximate by learning the offset from the audit's published partition. With the offset kept secret, part (b) of Theorem F.41 survives: rounds placed at points spaced exactly w apart put one point in each bin whatever the offset, so the $n_{\min}/w$ term remains. Part (a) does not obviously survive: it needs both alternation points in one bin, and a pattern whose bias cancels in *every* window of width w is impossible, because the residual changes sign only once. Whether a secret offset lets an audit beat $T^{2/3}$ is open.

The lower bound of Theorem F.41 becomes $T^{2/3}$ only if $w\, n_{\min}$ is bounded below. The next results prove that an audit which accepts honest learners in the gap must have this property, or else a level so high that a simpler manipulation succeeds.

Lemma F.44 (Honest RSA visits the neighbouring bin)**.** *Assume (G1)–(G5), Gaussian noise, RSA(a) with $a \geq 1$, and let J be the bin immediately to the right of the bin containing x_0 (for any offset). There are $w_1, c_{\mathrm{occ}}, p_{\mathrm{occ}} > 0$, depending only on the world and a, such that for all $w \in [2a\sigma/T, w_1]$,*

$$\mathbb{P}\big(\#\{t \leq T : \hat{p}_t \in J\} \geq c_{\mathrm{occ}}/w\big) \geq p_{\mathrm{occ}}.$$

Every round in J has bias at least $r^+ > 0$.

Proof. **Setup.** Let $x_t := t(\hat{p}_t - x_0)$; by Theorem F.15(i), $M := \sup_t \mathbb{E}|x_t| < \infty$. Put $\mathcal{P} := 2M$, so $\mathbb{P}(|x_t| \leq \mathcal{P}) \geq \frac{1}{2}$ for every t. Write $J = [x_0 + d, x_0 + d + w)$ with $d \in (0, w]$, and let $W := \{t : a\sigma/(2w) \leq t \leq a\sigma/w\} \subset [1, T]$, which has at least $a\sigma/(4w)$ elements for $w \leq w_1$.

One step lands in J with probability bounded below. Given $\mathcal{F}_t$, the unprojected point $y_t^u = m_t + (a/t)\eta_t$ with $m_t := \hat{p}_t - (a/t)r(\hat{p}_t)$ is $N(m_t, s_t^2)$ with $s_t := a\sigma/t \in [w, 2w]$ for $t \in W$. If $y_t^u \in J$ then $\hat{p}_{t+1} = y_t^u$, since J lies in the interior of $\mathcal{P}$ for $w \leq w_1$. Also $|m_t - x_0| \leq (|x_t| + aB)/t \leq \kappa w$ on $\{|x_t| \leq \mathcal{P}\}$, with $\kappa := 2(\mathcal{P} + aB)/(a\sigma)$. Every point of J is within $(\kappa+2)w$ of m_t, so the Gaussian density on J is at least $\varphi(\kappa+2)/s_t \geq \varphi(\kappa+2)/(2w)$. Hence $\mathbb{P}(\hat{p}_{t+1} \in J \mid \mathcal{F}_t) \geq p_1 := \varphi(\kappa+2)/2$ on $\{|x_t| \leq \mathcal{P}\}$.

Counting. Let $N^W := \#\{t \in W : \hat{p}_{t+1} \in J\}$. Then $\mathbb{E}N^W \geq (p_1/2)|W|$ and $N^W \leq |W|$. For a variable in $[0, b]$, $\mathbb{P}(X \geq \mathbb{E}X/2) \geq \mathbb{E}X/(2b)$, so $\mathbb{P}(N^W \geq p_1|W|/4) \geq p_1/4$. Take $c_{\mathrm{occ}} := p_1 a\sigma/16$ and $p_{\mathrm{occ}} := p_1/4$. The bias statement is Lemma C.2. □

Proposition F.45 (A small minimum count flags honest learners)**.** *In the setting of Lemma F.44, consider any binned audit (Definition F.22) with $w\, n_{\min} \leq c_{\mathrm{occ}}$ and $(z+z_0)^2 w < c_{\mathrm{occ}} {r^+}^2/\sigma^2$, where $z_0 := \sqrt{2\ln(2MT)}$. Then honest RSA(a) is flagged with probability at least $p_{\mathrm{occ}} - 1/(\sqrt{2\pi} z_0)$.*

Proof. On the event of Lemma F.44, J holds $N_J \geq c_{\mathrm{occ}}/w \geq n_{\min}$ rounds, so it is tested. On the complement of the event of Lemma F.27, $|S_J| \geq N_J r^+ - z_0\sigma\sqrt{N_J}$, which exceeds $z\sigma\sqrt{N_J}$ as soon as $N_J > (z + z_0)^2\sigma^2/{r^+}^2$; this holds because $N_J \geq c_{\mathrm{occ}}/w$ and $(z + z_0)^2 w < c_{\mathrm{occ}} {r^+}^2/\sigma^2$. □

Theorem F.46 (The rate for the audit family)**.** *Fix the gap world of Lemma F.44 (with honest RSA(1)) and the kink world of Theorem F.41. Let $\mathcal{A}$ be the family of binned audits (Definition F.22) with any $n_{\min}$, any level z and $w \in [2\sigma/T, w_0]$, and call an audit* accepting *if it flags honest RSA(1) in the gap world with probability at most $p_{\mathrm{occ}}/2$ for all large T. Then:*

(i) for every accepting audit in $\mathcal{A}$ there is a strategic learner in the kink world whose net benefit over honest play (gain minus extra expected penalty, stakes κT) is at least $cT^{2/3} - C$;

(ii) the default audit is accepting (Theorem F.51) and limits every strategic learner in every Lipschitz world to $O(T^{2/3}\ell_T^{1/3})$ (Theorem F.23).

So among accepting audits in this family, the best achievable worst-case manipulation is between $cT^{2/3}$ and $CT^{2/3}(\ln T \ln\ln T)^{1/3}$: the rate is $T^{2/3}$ up to a logarithmic factor.

Proof. Let $\varepsilon_0 := c_{\mathrm{occ}} {r^+}^2/\sigma^2$. Take any accepting audit and consider its parameters.

Case (a): $w\, n_{\min} > c_{\mathrm{occ}}$. By Theorem F.41, gain $\geq c_1 wT + c_2 n_{\min}/w - C_3 \geq c_1 wT + c_2 c_{\mathrm{occ}}/w^2 - C_3 \geq c' T^{2/3} - C_3$, minimising over w. If $n_{\min}/w > T/C_3$, the strategic learner instead puts all T rounds into untested bins inside the interval I of that proof, which requires only $n_{\min}(\varepsilon/w - 2) \geq T$, and gains at least $g_0 T$. The extra penalty over honest play is at most $\kappa T[2\bar{\Phi}_{\mathcal{N}}(z - \epsilon_T) - 2\bar{\Phi}_{\mathcal{N}}(z)] \leq \kappa T \epsilon_T$ with $\epsilon_T = (1+\gamma_1)w/(\sigma\sqrt{T/2})$, which is $O(w\sqrt{T}) = o(wT)$.

Case (b): $w\, n_{\min} \leq c_{\mathrm{occ}}$ and $(z+z_0)^2 w < \varepsilon_0$. By Proposition F.45 the audit flags honest RSA with probability at least $p_{\mathrm{occ}} - 1/(\sqrt{2\pi} z_0) > p_{\mathrm{occ}}/2$ for large T, so it is not accepting. This case cannot occur.

Case (c): $w\, n_{\min} \leq c_{\mathrm{occ}}$ and $(z+z_0)^2 w \geq \varepsilon_0$. If $w \geq \varepsilon_0/(32 \ln T)$, mechanism (a) of Theorem F.41 alone gains at least $c_1 wT \geq cT/\ln T$, with extra penalty $O(w\sqrt{T})$. Otherwise $z_0^2 w < \varepsilon_0/4$ for large T (since $z_0^2 = O(\ln T)$), hence $z^2 w \geq \varepsilon_0/4$ and $z^2 \geq 8 \ln T$. The strategic learner then plays the constant forecast $p_0 - \delta$ with $\delta := \min\big(h_0, z\sigma/(2(1+\gamma_1)\sqrt{T})\big)$. All rounds fall in one bin, whose bias sum is at most $z\sigma\sqrt{T}/2$ in absolute value; it is flagged only if $|N| > z\sigma\sqrt{T}/2$, with probability at most $2\bar{\Phi}_{\mathcal{N}}(z/2) \leq 2e^{-z^2/8} \leq 2/T$, so the extra penalty is at most 2κ. Its gain is $T(V'_-\delta - \frac{1}{2}V''_-\delta^2) \geq TV'_-\delta/2 \geq c\min(T, z\sigma\sqrt{T}) \geq c\min(T, \sigma\sqrt{\varepsilon_0 T/(4w)})$. Taking the better of this and mechanism (a), the gain is at least $\frac{c}{2}\big(wT + \min(T, \sqrt{T/w})\big) \geq c'' T^{2/3}$.

(ii) The default parameters satisfy the conditions of Theorem F.51(i), whose bound tends to zero, and Theorem F.23(iv). □

Remark F.47 (What is and is not established). Theorem F.46 is a lower bound for an audit *family*: every audit of Definition F.22, with any parameters, that accepts honest learners in the gap. It is not a lower bound over all conceivable audits: audits whose statistic uses the pattern of bin counts, whose bins refine over time, or which test forecast dispersion within bins are not covered, and whether any of them beats $T^{2/3}$ is open. The lower and upper bounds match up to the factor $\ell_T^{1/3} = (\ln T \ln\ln T)^{1/3}$; closing it is open. Case (b) is asymptotic in practice: at the default level, $z^2 w \approx 12.7$ at $T = 10^4$, so the default audit sits in the high-level regime at practical horizons (T50). The mechanism is nonetheless sharp: at the moderate level $z = 3$, audits with $w\, n_{\min} = 0.05, 0.2, 1$ flagged honest RSA in the gap in 100%, 69–77% and 37–40% of runs at $T = 10^4$ and 10^5, against 2.3% with the default minimum count.

In the gap, fine bins create a new difficulty: honest RSA itself alternates between biases r^- and r^+, at a scale shrinking like $1/t$ around x_0, and fine enough bins would eventually see this. The minimum count $n_{\min}$ and the random offset are what protect it.

Lemma F.48 (Occupancy of RSA away from the threshold). *Assume (G1)–(G5), Gaussian noise, and RSA(a) with $a \geq 1$, and let $M_\lambda := \sup_t \mathbb{E} e^{\lambda t |\hat{p}_t - x_0|} < \infty$. Then for every $\delta > 0$ and every T,*

$$\mathbb{E}\,\#\{t \leq T : |\hat{p}_t - x_0| \geq \delta\} \;\leq\; \frac{M_\lambda}{e^{\lambda\delta} - 1} \;\leq\; \frac{M_\lambda}{\lambda\delta}.$$

So the number of rounds RSA spends at distance at least δ from x_0 is $O(1/\delta)$, uniformly in T.

Proof. By Markov's inequality applied to $e^{\lambda t|\hat{p}_t - x_0|}$, $\mathbb{P}(|\hat{p}_t - x_0| \geq \delta) = \mathbb{P}(t|\hat{p}_t - x_0| \geq t\delta) \leq M_\lambda e^{-\lambda t\delta}$. Sum over $t \geq 1$: $\sum_t e^{-\lambda t\delta} = 1/(e^{\lambda\delta} - 1) \leq 1/(\lambda\delta)$. The exponential moments are Theorem F.15(i), which is where $a \geq 1$ is used; they are the rigorous form of the heuristic that $|\hat{p}_t - x_0|$ shrinks like $1/t$. □

Lemma F.49 (Occupancy for every stable gain). *Assume (G1)–(G5), Assumption A.2, and RSA(a) with any $a > \frac{1}{2}$. There is $C_{\mathrm{blk}} < \infty$ such that for every $\delta > 0$ and T,*

$$\mathbb{E}\,\#\{t \leq T : |\hat{p}_t - x_0| \geq \delta\} \;\leq\; \frac{C_{\mathrm{blk}}}{\delta}\Big(3 + \log_2^+ \frac{T\delta}{C_{\mathrm{blk}}}\Big).$$

Proof. The proof of Lemma F.13 shows, for every dyadic block $B_j := [2^j, 2^{j+1})$, that $\sum_{t\in B_j} \mathbb{E}|\hat{p}_t - x_0| \le C_{\mathrm{blk}} := (M_1 + 2C)/(ac_0)$; it uses only $a > \frac{1}{2}$. By Markov's inequality, block j contributes at most $\min(2^j, C_{\mathrm{blk}}/\delta)$ to the expected count. Blocks with $2^j < C_{\mathrm{blk}}/\delta$ contribute at most $2C_{\mathrm{blk}}/\delta$ in total. The remaining blocks with $2^j \le T$ number at most $1 + \log_2^+(T\delta/C_{\mathrm{blk}})$, each contributing at most C_{blk}/δ. □

Remark F.50. Lemma F.48 ($a \ge 1$) is sharper by the logarithm. Numerically the true law is $1/\delta$ for every gain tested: $\delta \times \mathbb{E}\#$ is 0.53 for $a = 1$ (T44), and 0.32 and 0.43 for $a = 0.6$ and 0.8 (T46), constant for δ from 0.03 to 0.001, while $\delta^{3/2} \times \mathbb{E}\#$ keeps falling. The logarithm in Lemma F.49 is therefore a proof artefact, but a harmless one: the default parameters absorb it.

Theorem F.51 (Honest RSA in the gap passes the binned audit)**.** *Assume (G1)–(G5), Gaussian noise, and run RSA(a). Use the binned audit with any level z and a uniform offset independent of the learner.*

(i) If $a \ge 1$ and $w_T n_{\min,T} \to \infty$, $n_{\min,T}/w_T = o(T)$, $n_{\min,T} = o(\sqrt{T})$, then $\mathbb{P}(\text{flag}) \le (2 + M_\lambda/\lambda)/\sqrt{w_T n_{\min,T}} + o(1)$, with $M_\lambda := \sup_t \mathbb{E}e^{\lambda t|\hat{p}_t - x_0|} < \infty$ (Theorem F.15(i)).

(ii) If $a > \frac{1}{2}$ and $w_T n_{\min,T}/\ln T \to \infty$, $n_{\min,T}/w_T = o(T/\ln T)$, $n_{\min,T} = o(\sqrt{T})$, then $\mathbb{P}(\text{flag}) = O\big(\sqrt{\ln T/(w_T n_{\min,T})}\big) + o(1)$.

The default parameters ($w_T n_{\min,T} = \ell_T = \ln T \ln\ln T$) satisfy both sets of conditions, so honest RSA with any gain $a > \frac{1}{2}$ is accepted, with false-alarm probability $O((\ln T \ln\ln T)^{-1/2})$ for $a \ge 1$ and $O((\ln\ln T)^{-1/2})$ for $\frac{1}{2} < a < 1$. The $\ln\ln T$ factor in ℓ_T is what makes (ii) apply.

Proof. **Distance to an edge.** Let d be the distance from x_0 to the nearest bin edge. Since the offset is uniform and independent of everything else, d is uniform on $[0, w/2]$, so $\mathbb{P}(d \le \delta) = 2\delta/w$.

Rounds away from x_0 are few. For (i), Lemma F.48 gives $\mathbb{E}\#\{t : |\hat{p}_t - x_0| \ge \delta\} \le M_\lambda/(\lambda\delta)$. For (ii), Lemma F.49 gives $\le C' \ln T/\delta$ for $\delta \ge 1/T$. We give the proof of (i); for (ii) replace $M_\lambda/(\lambda\delta)$ by $C' \ln T/\delta$ throughout and take $\delta := \sqrt{w \ln T/n_{\min}}$, which turns each bound $1/\sqrt{wn_{\min}}$ into $\sqrt{\ln T/(wn_{\min})}$ and each $o(\sqrt{T})$ requirement into $n_{\min}/w = o(T/\ln T)$. Note that $\delta \le w/2$ holds eventually, so $\mathbb{P}(d \le \delta) = 2\delta/w$ applies.

Other bins go untested. Every round outside x_0's bin has $|\hat{p}_t - x_0| \ge d$. Let n_{out} be the number of such rounds. With $\delta := \sqrt{w/n_{\min}}$, $\mathbb{P}(n_{\mathrm{out}} \ge n_{\min}) \le \mathbb{P}(d \le \delta) + \mathbb{P}(\#\{t : |\hat{p}_t - x_0| \ge \delta\} \ge n_{\min}) \le 2\delta/w + M_\lambda/(\lambda\delta n_{\min}) = (2 + M_\lambda/\lambda)/\sqrt{wn_{\min}}$. On $\{n_{\mathrm{out}} < n_{\min}\}$ no bin other than x_0's is tested.

x_0's bin passes. It holds $n_0 \ge T - n_{\min} \ge T/2$ rounds, and $S_0 = \Sigma_T - \sum_{t\ \mathrm{out}}(y_t - \hat{p}_t)$ with Σ_T the pooled statistic. On $\{n_{\mathrm{out}} < n_{\min}\} \cap \{d > \delta\}$, $|\sum_{\mathrm{out}}(y_t - \hat{p}_t)| \le Bn_{\min} + \sum_{\mathrm{out}} |\eta_t|$, and $\mathbb{E}[\sum_{\mathrm{out}} |\eta_t| ; d > \delta] \le \sigma M_\lambda/(\lambda\delta) = O(\sqrt{n_{\min}/w}) = o(\sqrt{T})$, because the event that round t is out is decided before η_t. By Theorem F.15(iv), valid for every $a > \frac{1}{2}$, $\mathbb{E}|\Sigma_T| = O(1 + \ln T + T^{1/3}) = o(\sqrt{T})$. By Markov's inequality each piece exceeds $z\sigma\sqrt{T/2}/3$ with probability $o(1)$, and $Bn_{\min} = o(\sqrt{T})$. □

Corollary F.52 (The gap conflict survives)**.** *Under the conditions of Theorem F.23(iv), with the gap response in place of a Lipschitz one, any learner whose average bias satisfies $\mathbb{P}(|\bar{r}_T| \ge b) \to 1$ for some $b > 0$ is flagged with probability tending to one. In particular the accuracy-optimal learner of Theorem C.19 is flagged, while honest RSA passes (Theorem F.51): the bias–accuracy conflict of Corollary F.17 holds for the binned audit too.*

Proof. The chain of the proof of Theorem F.23, applied to signed biases rather than absolute ones and without the within-bin step, gives on the event of no flagging $T|\bar{r}_T| = |\sum_j \sum_{t\in B_j} \rho_t| \le z\sigma\sqrt{MT} + \sum_j |N_j| + \rho_{\max} M n_{\min}$. Here a bin containing x_0 is handled like any other, since only the tested-bin bound is used. The right side is $o(T)$ in probability ($\mathbb{E}\sum_j |N_j| \le \sigma\sqrt{MT}$), while the left side is at least bT with probability tending to one. For the accuracy-optimal learner, $\bar{r}_T \to \kappa^* \in \{r^-, r^+\}$ in probability (proof of Corollary F.17). □

F.4 Summary of the scaled setting

Theorem F.53 (Summary)**.** *In the scaled setting:*

(1) rationality makes the world learnable at $\Theta(\log T)$, because the honest forecast is the root of an observable residual (Theorem E.13);

(2) the same rationality makes every learner with $\alpha < 1$ prefer to manipulate, and when φ is concave or a power law the incentive and the harm grow with the world's responsiveness (Theorem E.14);

(3) the strategic forecast depends on the slope of the response, which the outcome at the target does not reveal; learning it costs $\Theta(\sqrt{T})$, or $\Theta(\log T)$ only if μ is known (Theorem E.16);

(4) manipulation destroys value whenever $q \le (1+\alpha)/(2\alpha)$, in particular for every response concave between the two forecasts (Theorem D.36);

(5) under (S_c), in particular for every convex response, a pooled audit with linear stakes limits manipulation to $O(\sqrt{T \ln T})$ gain and harm (Theorem F.20(i)–(ii)); without the tangent condition no pooled audit can (Theorem F.20(iii)), but a binned audit limits it to $O(T^{2/3}(\ln T \ln\ln T)^{1/3})$ in every Lipschitz world, measured against the full-information honest learner (Theorem F.23), and accepts honest learners with vanishing false-alarm probability, including in the gap (Theorems F.28 and F.51). In linear and smooth nonlinear worlds it also makes honest learning approximately optimal (Theorems F.30 and F.33); in the gap no binned audit does (Theorem F.36).

Proof. Collect the cited results. □

G Extensions

G.1 Beyond the scalar model

The model so far has one scalar forecast and one scalar response. This section shows which conclusions depend on that. The learning results and the qualitative manipulation results do not need scalars or a particular cost; they need a *monotone* response operator. A cost-minimising world produces one through the additive channel of Proposition G.2; for a multiplicative vector channel this is not established (Remark G.3).

Assumption G.1 (The vector setting)**.** $\mathcal{P} \subset \mathbb{R}^d$ is compact and convex. The learner publishes $\hat{p}_t \in \mathcal{P}$ and observes $y_t = \mu - \Phi(\hat{p}_t) + \eta_t \in \mathbb{R}^d$, with η_t as in Assumption A.2 coordinatewise and $\mathbb{E}[\eta_t \eta_t^\top \mid \mathcal{F}_{t-1}] = S(\hat{p}_t) \preceq \bar{\sigma}^2 I$. Loss is $\|\hat{p}_t - y_t\|^2$ and $r(p) := p + \Phi(p) - \mu$. **Rationality**: Φ is monotone, $\langle \Phi(p) - \Phi(q), p - q\rangle \ge 0$, and L-Lipschitz; the honest forecast p^* with $r(p^*) = 0$ lies in $\mathcal{P}$.

Proposition G.2 (Where monotone responses come from)**.** *Suppose the world chooses an action $\lambda \in \mathbb{R}^k$ to minimise $c(\lambda) - \langle M\lambda, p\rangle$, with c strictly convex, differentiable and coercive, and the action shifts the outcome by $M\lambda$ ($M \in \mathbb{R}^{d\times k}$). Then $\Phi(p) = M\nabla c^*(M^\top p) = \nabla\psi(p)$ with $\psi := c^* \circ M^\top$ convex, so Φ is monotone. In one dimension every nondecreasing Φ is the derivative of a convex function, so the vector setting contains every scalar world of Appendices A and E.2.*

Proof. The first-order condition $\nabla c(\lambda) = M^\top p$ has the unique solution $\lambda^* = \nabla c^*(M^\top p)$ (convex conjugate; c^* is differentiable because c is strictly convex and coercive). The chain rule gives $\nabla\psi(p) = M\nabla c^*(M^\top p)$. The gradient of a convex function is monotone: $\langle \nabla\psi(p) - \nabla\psi(q), p - q\rangle \ge 0$ follows by adding the two supporting-hyperplane inequalities $\psi(q) \ge \psi(p) + \langle \nabla\psi(p), q - p\rangle$ and its mirror image. □

Remark G.3 (Which channel this micro-foundation uses). The shift here is $M\lambda$, the *additive* channel: the world's action moves the outcome independently of the forecast. The scalar model's main channel is multiplicative ($\Phi(p) = \lambda p$). The vector setting contains every scalar world only because every monotone scalar Φ is a gradient; for a multiplicative *vector* channel, what a cost-minimising world produces is not established here.

Theorem G.4 (Learning in vector worlds)**.** *In the vector setting:*

(i) (residual identity) $\mathbb{E}[\|\hat{p}_t - y_t\|^2 \mid \mathcal{F}_{t-1}] = \|r(\hat{p}_t)\|^2 + \operatorname{tr} S(\hat{p}_t)$*;*

(ii) (modulus one) $\langle r(p) - r(q), p - q\rangle \geq \|p - q\|^2$*, so* p^* *is unique; and* $r(p) = p - \mathbb{E}[y \mid p]$ *is observable, so the honest forecast certifies itself;*

(iii) (upper bound) vector RSA*,* $\hat{p}_{t+1} = \Pi_{\mathcal{P}}(\hat{p}_t - \frac{a}{t}(\hat{p}_t - y_t))$ *with* $a > \frac{1}{2}$*, has* $\operatorname{Regret}_T := \sum_t \mathbb{E}\|r(\hat{p}_t)\|^2 \leq L'^2 M'(1 + \ln T)$*;*

(iv) (lower bound) if Φ *is known,* $S \equiv \sigma^2 I$ *with Gaussian noise, and* μ *ranges over a box of width* w *in each coordinate, every algorithm has* $\sup_\mu \mathbb{E}\operatorname{Regret}_T \geq d\,\sigma^2 \ln T - C$*. Theorem G.5 shows the matrix gain* R'^{-1}*,* $R' := I + D\Phi(p^*)$*, attains* $d\sigma^2 \ln T(1 + o(1))$ *when* $D\Phi(p^*)$ *is symmetric, so the bound is sharp; T36 measured* $3.13\sigma^2$ *in* $d = 3$*.*

Proof. **(i)** As in Lemma B.1: $\hat{p}_t - y_t = r(\hat{p}_t) - \eta_t$, and the cross term has zero conditional mean.

(ii) $\langle r(p) - r(q), p - q\rangle = \|p - q\|^2 + \langle \Phi(p) - \Phi(q), p - q\rangle$. Two roots $p \neq q$ would give $0 \geq \|p - q\|^2$.

(iii) With $e_t := \hat{p}_t - p^*$ and $w_t := e_t - \frac{a}{t} r(\hat{p}_t)$, nonexpansiveness of $\Pi_{\mathcal{P}}$ gives $\|e_{t+1}\| \leq \|w_t + \frac{a}{t}\eta_t\|$. By (ii) and Lipschitzness of r (constant $L' := 1 + L$), $\|w_t\|^2 \leq (1 - \frac{2a}{t} + \frac{a^2 L'^2}{t^2})\|e_t\|^2$. Taking conditional expectations, the cross term vanishes and $\mathbb{E}\|e_{t+1}\|^2 \leq (1 - \frac{2a}{t} + \frac{a^2 L'^2}{t^2})\mathbb{E}\|e_t\|^2 + \frac{a^2 d\bar{\sigma}^2}{t^2}$. Lemma B.11 with $k = 1$, $c = 2a > 1$ gives $\mathbb{E}\|e_t\|^2 \leq M'/t$, and $\|r(\hat{p}_t)\| \leq L'\|e_t\|$ gives the bound, exactly as in Theorem B.13.

(iv) As in Theorem B.29: $x_s := y_s + \Phi(\hat{p}_s) = \mu + \eta_s$ are i.i.d. $N(\mu, \sigma^2 I)$ given (μ, U) and are in bijection with the data, and $\|r(\hat{p}_t)\|^2 = \|\hat{\mu}_t - \mu\|^2$ with $\hat{\mu}_t := \hat{p}_t + \Phi(\hat{p}_t)$. Put independent raised-cosine priors on the coordinates of μ. Coordinate j of the data is independent of $(\mu_i)_{i \neq j}$ and of the other coordinates' noise, so Lemma B.28 applied to coordinate j, conditionally on everything else, gives $\int \mathbb{E}(\hat{\mu}_{t,j} - \mu_j)^2 \pi \geq \sigma^2/(t - 1 + b)$. Sum over j and t as in Theorem B.29. □

Theorem G.5 (Sharp constant in vector worlds)**.** *In the vector setting, assume Assumption A.2 with conditional moments of all orders (Gaussian noise suffices),* p^* *interior to* $\mathcal{P}$*,* Φ *differentiable at* p^* *with* symmetric *Jacobian* $D\Phi(p^*)$ *(automatic when* $\Phi = \nabla\psi$ *as in Proposition G.2), and* $S(\cdot)$ *continuous at* p^**. Let* $H := R' = I + D\Phi(p^*)$*, which is symmetric with* $H \succeq I$*, and run the matrix-gain learner*

$$\hat{p}_{t+1} = \Pi_{\mathcal{P}}^H\Big(\hat{p}_t - \tfrac{1}{t} H^{-1}(\hat{p}_t - y_t)\Big),$$

where $\Pi_{\mathcal{P}}^H$ *is the projection onto* $\mathcal{P}$ *in the norm* $\|v\|_H := (v^\top H v)^{1/2}$*. Then* $\operatorname{Regret}_T = \operatorname{tr} S(p^*) \ln T\,(1 + o(1))$*; with* $S \equiv \sigma^2 I$ *this is* $d\sigma^2 \ln T(1 + o(1))$*, matching Theorem G.4(iv).*

Proof. Write $e_t := \hat{p}_t - p^*$, $r_t := r(\hat{p}_t)$, $\xi_t := \hat{p}_t - y_t - r_t$ (conditional mean 0, covariance $S(\hat{p}_t)$), $G := H^{-1}$, and $\rho(e) := r(p^* + e) - He$, so $\|\rho(e)\| \leq \omega(\|e\|)\|e\|$ with $\omega(\delta) \to 0$ (differentiability) and $\|\rho(e)\| \leq (L + 1 + \|H\|)\|e\|$.

Step 1: global moments in the H-norm. Let $A_t := \|e_t\|_H^2$ and $w_t := e_t - \frac{1}{t} G(r_t + \xi_t)$. Since $p^* \in \mathcal{P}$ and $\Pi_{\mathcal{P}}^H$ is nonexpansive in $\|\cdot\|_H$, $A_{t+1} \leq \|w_t\|_H^2 = A_t - \frac{2}{t}\langle e_t, r_t + \xi_t\rangle + \frac{1}{t^2}\|G(r_t + \xi_t)\|_H^2$, using $\langle e, HGv\rangle = \langle e, v\rangle$. By Theorem G.4(ii), $\langle e_t, r_t\rangle \geq \|e_t\|^2 \geq A_t/\lambda_{\max}(H)$. The expansion of $\|w_t\|_H^{2k}$ has exactly the structure of $(*)$ in Lemma B.20, with $|e_t|^2$ replaced by A_t, $2am$ by $2/\lambda_{\max}(H)$, and $|e_t|^{2k-j}$ by $A_t^{k-j/2}$. The proof of Lemma B.20(i) therefore gives $\mathbb{E}A_t^k \leq M_k t^{-\rho k}$ for every k and any $\rho \in (0, \min(2/\lambda_{\max}(H), 1))$, and part (ii) gives $e_t \to 0$ a.s.

Step 2: local moments. Near 0, $\langle e, r(p^* + e)\rangle = e^\top H e + \langle e, \rho(e)\rangle \geq (1 - \epsilon)A$ for $\|e\| \leq \delta$, with $2(1 - \epsilon) > 1$. Everywhere $\langle e, r\rangle \geq 0$. The localisation of Lemma B.20(iii), with the tail bound from Step 1, gives $\mathbb{E}A_t^k \leq M_k' t^{-k}$ for $k = 1, 2, 3$.

Step 3: the exact linear part. Since $GH = I$, before projection $w_t = (1 - \frac{1}{t})e_t - \frac{1}{t} G\rho(e_t) - \frac{1}{t} G\xi_t$. Write $e_{t+1} = w_t + \pi_t$ for the projection correction. Since $0 \in \mathcal{P} - p^*$, $\|\pi_t\|_H \leq \|w_t\|_H$, and $\pi_t \neq 0$ only if $\|w_t\| \geq D$, a fixed multiple of $\operatorname{dist}(p^*, \partial\mathcal{P}) > 0$. Also $\|w_t\| \leq \|e_t\| + \frac{C}{t}(\|e_t\| + \|\xi_t\|)$, so $\mathbb{E}\|w_t\|^6 \leq Ct^{-3}$ by Step 2. Hence $\mathbb{E}\|\pi_t\|^2 + \mathbb{E}\|\pi_t\|\,\|w_t\| \leq C\,\mathbb{E}\big[\|w_t\|^2 \mathbf{1}\{\|w_t\| \geq D\}\big] \leq C\,\mathbb{E}\|w_t\|^6/D^4 \leq Ct^{-3}$.

Step 4: covariance by polarisation. Fix $v \in \mathbb{R}^d$ and let $u_t := \mathbb{E}\langle v, e_t\rangle^2 \geq 0$. From Step 3,

$$u_{t+1} = \Big(1 - \frac{1}{t}\Big)^2 u_t + \frac{v^\top G\, \mathbb{E}S(\hat{p}_t)\, G^\top v}{t^2} + \varepsilon_t,$$

where ε_t collects: the cross term $-\frac{2}{t}(1 - \frac{1}{t})\mathbb{E}[\langle v, e_t\rangle\langle v, G\rho(e_t)\rangle]$, which is at most $\frac{C}{t}\mathbb{E}[\|e_t\|^2\omega(\|e_t\|)] = o(t^{-2})$ by Step 2 and dominated convergence exactly as in Theorem B.17; the term $\frac{1}{t^2}\mathbb{E}\langle v, G\rho\rangle^2 = o(t^{-2})$ similarly; the noise cross terms, which vanish because ξ_t has conditional mean 0; and the projection terms, $O(t^{-3})$ by Step 3. Also $\mathbb{E}S(\hat{p}_t) \to S(p^*)$ by a.s. convergence, continuity and boundedness ($S \preceq \bar{\sigma}^2 I$). With $(1 - 1/t)^2 = 1 - 2/t + 1/t^2$ and $u_t \leq M/t$, Lemma B.16 (with $c = 2$) gives $tu_t \to v^\top G S(p^*) G^\top v$. As v is arbitrary and the covariance is symmetric, $t\,\mathbb{E}[e_t e_t^\top] \to GS(p^*)G^\top$.

Step 5: regret. $r_t = He_t + \rho(e_t)$, so $\mathbb{E}\|r_t\|^2 = \mathrm{tr}(H\,\mathbb{E}[e_t e_t^\top]\,H) + o(1/t)$, the remainder handled as in Step 4. Since $HG = I$ and H is symmetric, $\mathrm{tr}(HGS(p^*)G^\top H) = \mathrm{tr}\, S(p^*)$. So $t\,\mathbb{E}\|r_t\|^2 \to \mathrm{tr}\, S(p^*)$, and summing (Stolz–Cesàro) gives the claim. □

Remark G.6. Symmetry of $D\Phi(p^*)$ is used once: to make H a valid inner product in which the drift $\langle e, r\rangle$ is exactly the H-norm contraction. For a non-symmetric Jacobian the matrix gain can still be locally optimal, but global convergence of the preconditioned recursion is not guaranteed, and that case is open. Like the tuned scalar gain, the learner assumes H is known; an adaptive version in the style of Wei (1987) is open.

Theorem G.7 (Manipulation in vector worlds)**.** *In the vector setting with Φ differentiable, let $V(p) := \alpha\|r(p)\|^2 + (1-\alpha)\|\Phi(p)\|^2$.*

(i) $\nabla V(p^*) = 2(1-\alpha)D\Phi(p^*)^\top\Phi(p^*)$ *and* $\langle \nabla V(p^*), \Phi(p^*)\rangle = 2(1-\alpha)\,\Phi(p^*)^\top D\Phi(p^*)\,\Phi(p^*) \geq 0$*, with strict inequality when the symmetric part of $D\Phi(p^*)$ is positive definite and $\Phi(p^*) \neq 0$. In that case moving the forecast from p^* against the response, along $-\Phi(p^*)$, strictly lowers V: every learner with $\alpha < 1$ has an incentive to manipulate, and $\Delta_V > 0$.*

(ii) ***The strategic forecast is not identified at the target; learning it costs*** $\Omega(\sqrt{T})$*. Any class of vector worlds that contains the embedded scalar family $\Phi(p) = (\gamma p_1^2, 0, \ldots, 0)$ on $\mathcal{P} = [0, P] \times K'$ with μ and γ unknown has $\sup S_T \geq c\sqrt{T} - C$ for every algorithm.*

(iii) ***Upper bound.*** *For linear-in-parameter families $\Phi_\theta = \sum_{i\leq k}\theta_i\varphi_i$ with (μ, θ) in a compact box, an exploration design whose regression matrix has full column rank, and a unique, interior, nondegenerate strategic minimiser throughout the box, explore-then-commit has $S_T = O(\sqrt{T})$.*

Proof. **(i)** Differentiate: $\nabla V = 2\alpha(I + D\Phi)^\top r + 2(1-\alpha)D\Phi^\top\Phi$ and $r(p^*) = 0$. The inner product is $2(1-\alpha)\Phi^\top D\Phi^\top\Phi = 2(1-\alpha)\Phi^\top \mathrm{sym}(D\Phi)\Phi$; monotonicity of a differentiable Φ makes $\mathrm{sym}(D\Phi) \succeq 0$. The directional derivative of V at p^* along $-\Phi(p^*)$ is minus this quantity.

(ii) For the embedded family, $V(p) = \alpha r_1(p_1)^2 + (1-\alpha)\gamma^2 p_1^4 + \alpha\sum_{j\geq 2}(p_j - \mu_j)^2$, and the observations split: $y_{t,1} = \mu_1 - \gamma\hat{p}_{t,1}^2 + \eta_{t,1}$, while $y_{t,j} = \mu_j + \eta_{t,j}$ for $j \geq 2$ does not depend on (μ_1, γ). So any vector algorithm, run with the coordinates $j \geq 2$ simulated from independent randomness, is a scalar algorithm for coordinate 1, and its strategic regret is at least that of coordinate 1. Apply Theorem E.3; its uninformative-curve mechanism (Theorem E.1) is inherited unchanged.

(iii) The proof of Theorem E.2 uses three facts: an estimator with $\mathbb{E}\|\hat{\vartheta} - \vartheta\|^2 \leq C/n$, which least squares on a full-rank linear design provides; Lipschitz dependence of the strategic forecast on the parameters, which the implicit function theorem gives under nondegeneracy; and quadratic growth of V near the target, from Taylor's theorem. □

Remark G.8 (What else is covered, and what is not). *Covered.* Stochastic responses with mean $\Phi(p)$ and independent zero-mean fluctuations: they add prediction-dependent noise, which Assumption A.2 already allows (as in Corollary G.9). Several homogeneous honest learners (Proposition G.12). Any responsiveness and shape: with the tuned gain the leading constant σ^2 of Theorem B.17 does not depend on Φ at all. *Not covered* (open): history-dependent responses, such as a world whose state adjusts gradually; a drifting fundamental μ_t; the vector versions of value destruction and of the audit; heterogeneous or strategic learners; responses that depend on variables other than the forecast.

Simultaneous moves change nothing for an honest learner, whose target is the Nash (stable) point, but they remove the learner's leadership and with it the manipulation of Appendix D.1.

G.2 Robustness

Corollary G.9 (Prediction-dependent noise; perception noise)**.** *Theorems B.13 and B.17 were proved under Assumption A.2, which allows $\sigma^2(\hat{p}_t)$; they hold with $\bar{\sigma}^2 = \sup_{\mathcal{P}} \sigma^2$ and leading constant $g(a)\sigma^2(p^*)$. If the world observes $\hat{p}+\xi'$ with $\xi' \sim N(0,\tau^2)$ independent, then $z_t = r(\hat{p}_t) - \eta_t + \gamma\xi'_t\hat{p}_t$ and $\sigma^2(p) = \sigma^2 + \gamma^2\tau^2p^2$, which is continuous with all conditional moments bounded on $\mathcal{P}$. In the quadratic world with $a = 1/s$ the constant is $\sigma^2(p^*) = \sigma^2 + \tau^2(s-1)^2/4$ (using $\gamma p^* = (s-1)/2$), growing linearly in $\gamma\mu$. Whether this constant is optimal under perception noise is open. Experiment E13: all 8 measured slopes lie within their confidence intervals of the prediction.*

Remark G.10 (Heavy tails). Theorem B.13 uses only the conditional variance; Theorem B.17 uses moments up to order 6. Almost surely, finite variance suffices for the sharp constant with a fixed gain, for the clipped-publication variant (Proposition B.24(i), from Lai and Robbins 1979, whose fixed-gain theorem needs only $\mathbb{E}\varepsilon^2 < \infty$); Proposition B.25 assumes Gaussian noise only to meet their convergence condition, and symmetric noise with finite variance would also do. Experiment E13 finds the Gaussian constant also for Student-t noise with 3 and 5 degrees of freedom (both with finite variance, infinite fourth and sixth moments respectively). The in-expectation version under finite variance is open.

Remark G.11 (Delay). If the update at round t uses z_{t-d}, experiment E13 finds the slope unchanged for $d \le 5$ and an additive cost of about $5d$; at $d = 50$ the finite-T slope at $a = 1/3$ is about 4.4 standard errors high, about 17% above the prediction.

Proposition G.12 (Several honest learners)**.** *Let N learners publish $p_{i,t}$, and let the world respond to the aggregate $\bar{p}_t := \frac{1}{N}\sum_i p_{i,t}$: $y_t = \mu - \Phi(\bar{p}_t) + \eta_t$. Each learner runs RSA$(a)$ on its own residual $z_{i,t} = p_{i,t} - y_t$, without projection, and assume (M1)–(M3) hold for $r(p) = p + \Phi(p) - \mu$ on all of $\mathbb{R}$ with $2am > 1$ and Assumption A.2. Then:*

(i) $\bar{p}_t$ is exactly RSA(a) on the aggregate residual, so Theorems B.13 and B.17 apply to the aggregate regret $\mathrm{Regret}_T^{\mathrm{agg}} := \sum_t \mathbb{E}r(\bar{p}_t)^2$;

(ii) the disagreement is deterministic: $p_{i,t} - p_{j,t} = (p_{i,1} - p_{j,1})\prod_{s<t}(1 - a/s)$, so $|\delta_{i,t}| \le Ct^{-a}$ for $\delta_{i,t} := p_{i,t} - \bar{p}_t$;

(iii) each learner's regret $\mathrm{Regret}_T^{(i)} := \sum_t \mathbb{E}[(p_{i,t}-y_t)^2 - \sigma^2]$ satisfies $|\mathrm{Regret}_T^{(i)} - \mathrm{Regret}_T^{\mathrm{agg}}| \le C$, uniformly in T, and exactly $\frac{1}{N}\sum_i \mathrm{Regret}_T^{(i)} = \mathrm{Regret}_T^{\mathrm{agg}} + \frac{1}{N}\sum_i\sum_t \delta_{i,t}^2$.

So a population of homogeneous honest learners behaves as one RSA learner.

Proof. **(i)** Averaging the updates, $\bar{p}_{t+1} = \bar{p}_t - (a/t)(\bar{p}_t - y_t) = \bar{p}_t - (a/t)(r(\bar{p}_t) - \eta_t)$.

(ii) Subtracting two updates, the common y_t cancels: $p_{i,t+1} - p_{j,t+1} = (1 - a/t)(p_{i,t} - p_{j,t})$; and $\prod_{s<t}|1 - a/s| \le Ct^{-a}$.

(iii) $\mathbb{E}[(p_{i,t} - y_t)^2 \mid \mathcal{F}_{t-1}] - \sigma^2 = (r(\bar{p}_t) + \delta_{i,t})^2$. So $\mathrm{Regret}_T^{(i)} - \mathrm{Regret}_T^{\mathrm{agg}} = \sum_t \big(2\delta_{i,t}\mathbb{E}r(\bar{p}_t) + \delta_{i,t}^2\big)$, and $|\mathbb{E}r(\bar{p}_t)| \le L\sqrt{M_1'/t}$ by Theorem B.13; since $a > \frac{1}{2}$ both $\sum_t t^{-a-1/2}$ and $\sum_t t^{-2a}$ converge. Averaging over i, $\sum_i \delta_{i,t} = 0$ removes the cross term exactly. □

Remark G.13. The quadratic world does not satisfy (M3) on all of $\mathbb{R}$, and without projection RSA can diverge there (observed in T30); with projection, averaging no longer commutes with the update, and that case is open. Heterogeneous gains and several *strategic* learners are open.

H Experiments, verification index and extended related work

H.1 Experiments

Common setup. Unless stated otherwise $\mu = 1$ and $\sigma = 0.5$; seeds are fixed, so every number below reproduces exactly. Regret slopes are $[\mathrm{Regret}(10^5) - \mathrm{Regret}(10^4)]/\ln 10$ with 95% con-

fidence intervals across seeds. Scripts are in `experiments/` (run order in its README) and `verification/`; raw results are in `experiments/results/*.json`. Figures are sized for the ICLR text width and use the Okabe–Ito colourblind-safe palette. Tables in this appendix are generated directly from the result files. Experiments on open items are evidence only, not proof. E# numbers experiments and T# numbers verification checks; E14 and T27, T39, T56 and T57 were retired, and the numbers are not reused. The README of the supplementary package maps every figure and table to its script and result file.

H.1.1 When the equilibrium is learnable (E5, E6)

Figure 5 reports E5 and E6. Random response (E6): the per-round excess of RSA matches the

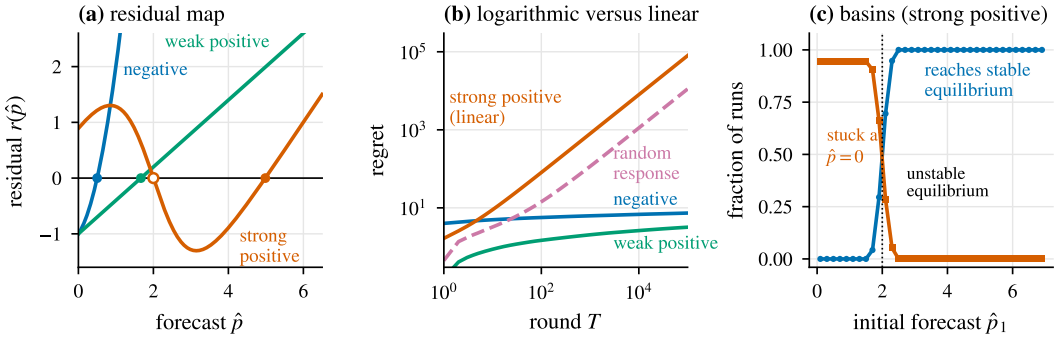

Figure 5: (a) The residual map decides learnability: under negative or weak positive feedback it is monotone with one root; strong positive feedback breaks monotonicity. (b) Monotone worlds give logarithmic regret (tuned gain; 7.4 and 3.2 at $T = 10^5$); strong positive feedback started below the unstable root and a random response ($\Lambda = 1$) give linear regret (8.0×10^4 and 1.1×10^4). (c) Basins under strong positive feedback (400 seeds per start, $T = 2 \times 10^4$): runs starting below the unstable root end at $\hat{p} = 0$. Supports Propositions B.3, B.6, B.7 and B.9.

prediction $F(2\mu/(2+\Lambda))$ to four decimals for $\Lambda \in \{0.25, 0.5, 1, 2, 4\}$ and always lies above the bound of Proposition B.9. A phase diagram over β (not plotted) gives a tuned-gain slope of 0.92 ± 0.09 times σ^2 for every $\beta > -1$ and linear regret for $\beta < -1$.

H.1.2 The sharp constant (E1)

Figure 6 and Table 5 report E1. The proved range is $c > \frac{1}{2}$ (Theorem B.21). Of the 21 points in it, 20 lie within 1.5 confidence half-widths of $c^2/(2c-1)$ (Table 5). The exception is $c = 0.6$ at $\gamma\mu = 10$, which is 9.2 half-widths high: there the transient decays only like $t^{-(2c-1)} = t^{-0.2}$, so $T \le 10^5$ is pre-asymptotic. Below $c = \frac{1}{2}$ regret is polynomial.

Table 5: E1: regret slope/σ^2 (95% CI half-width) against $c = aR'$; 1000 seeds, slope over $T \in [10^4, 10^5]$. Theory $c^2/(2c-1)$ for $c > 1/2$, the proved range. At $c = 0.6$, $T \le 10^5$ is pre-asymptotic.

c	theory	$\gamma\mu = 0.5$	$\gamma\mu = 2$	$\gamma\mu = 10$
0.35	∞	18.361 (1.345)	55.698 (2.403)	94.549 (2.937)
0.45	∞	4.547 (0.362)	7.059 (0.505)	4.165 (0.344)
0.6	1.800	1.601 (0.136)	1.621 (0.123)	5.259 (0.378)
0.8	1.067	1.073 (0.074)	1.054 (0.072)	1.171 (0.085)
1.0	1.000	1.018 (0.063)	0.984 (0.063)	1.014 (0.061)
1.5	1.125	1.110 (0.057)	1.122 (0.060)	1.078 (0.060)
2.0	1.333	1.316 (0.056)	1.328 (0.058)	1.390 (0.062)
3.0	1.800	1.742 (0.061)	1.849 (0.067)	1.771 (0.064)
5.0	2.778	2.787 (0.072)	2.758 (0.072)	2.783 (0.074)

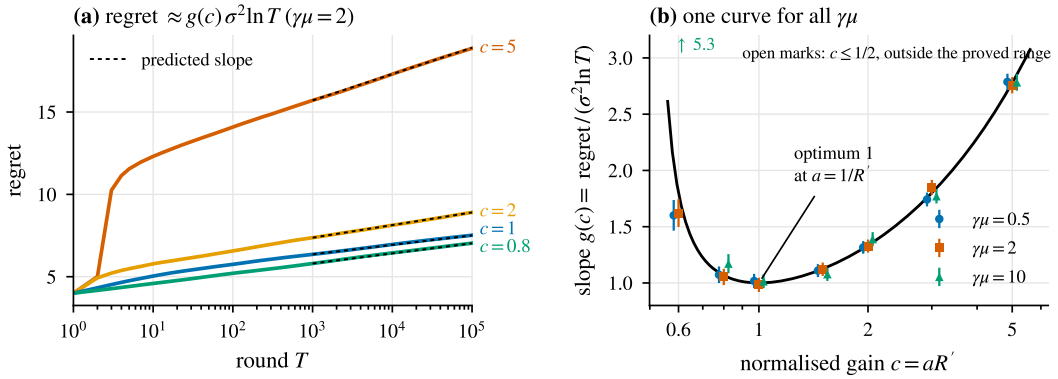

Figure 6: (a) RSA regret grows like $g(c)\sigma^2 \ln T$ (dotted: slopes from Theorem B.17; $\gamma\mu = 2$, 500 seeds). (b) Measured $g(c)$ for three levels of responsiveness collapses onto one curve, with minimum 1 at $a = 1/R'$. Open marks: $c \le \frac{1}{2}$, outside the proved range $c > \frac{1}{2}$ (Theorem B.21).

H.1.3 The lower bound, and the price of not tuning (E2, E3)

Figure 7 and Table 6 report E2 and E3.

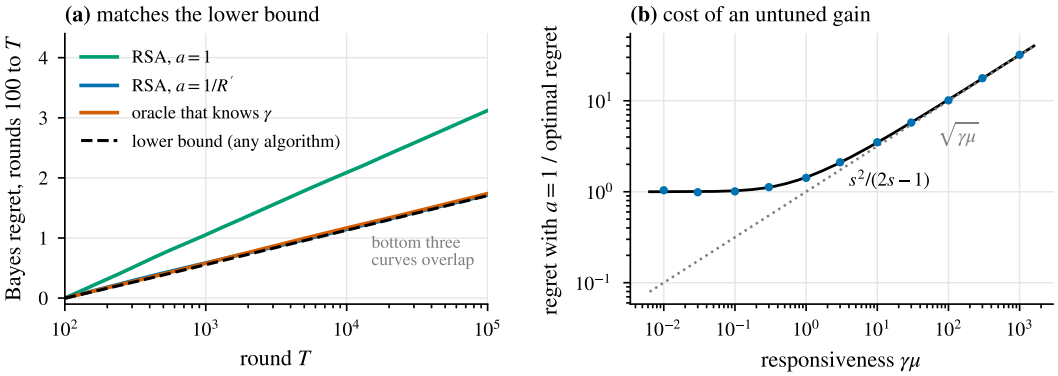

Figure 7: (a) Bayes regret from round 100 to T under the raised-cosine prior on $\mu \in [0.5, 1.5]$ used in Theorem B.29 (2000 draws, $\gamma = 2$): slopes $0.98\sigma^2$ (RSA, $a = 1/R'(\mu)$), $1.01\sigma^2$ (oracle knowing γ), $1.00\sigma^2$ (exact van Trees bound), $1.80\sigma^2$ (RSA, $a = 1$). (b) The cost of $a = 1$ follows $s^2/(2s-1)$ and grows like $\sqrt{\gamma\mu}$.

H.1.4 The gap regime (E7, E9)

Figure 8 reports E7; E9 is described below it. E9 (the closed-form orbit, Theorem C.5) is an exact computation: iteration and closed form agree to 3.7×10^{-13} on 20,000 random instances.

H.1.5 The cheaper side of the gap (E8 and T16)

Table 7 and Fig. 9 report E8. Two algorithms were tested. The probe-based variant of E8 (localise with RSA for $\sqrt{T}$ rounds, probe $\hat{x} \pm T^{-1/3}$, then run shifted RSA) identifies the side in 95–100% of runs in the four nondegenerate instances, and its excess over the minimum shrinks by a factor 2.3–2.5 per decade there, consistent with $T^{-1/3}$ (Table 7). The fifth instance, $(1, 0.5, 1, 0.62)$, is nearly degenerate: $\min(r^{-2}, r^{+2}) \approx 2 \times 10^{-5}$, the side is correct in only 88–89% of runs, and RSA's own excess (0.0008) is already below the probe learner's at every horizon (0.0015 at $T = 10^6$), so the probe learner does not help there; it is omitted from Fig. 9(b). The moment-based algorithm

Table 6: E3: price of not tuning. Slope with $a = 1$ divided by σ^2, against $s^2/(2s-1)$; 1000 seeds.

$\gamma\mu$	measured	CI	theory
0.01	1.038	0.066	1.000
0.03	0.988	0.059	1.003
0.1	1.009	0.059	1.025
0.3	1.124	0.063	1.119
1	1.421	0.057	1.440
3	2.108	0.070	2.093
10	3.502	0.081	3.473
30	5.764	0.107	5.762
100	10.106	0.131	10.269
300	17.707	0.176	17.581
1000	32.014	0.227	31.879

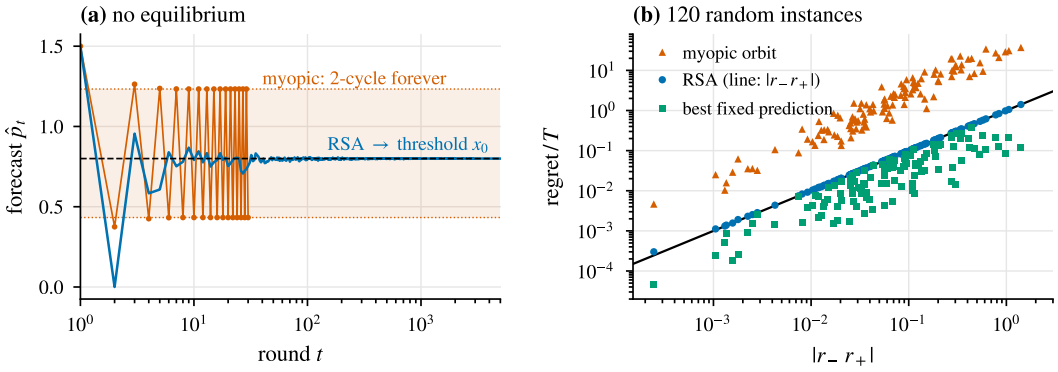

Figure 8: (a) At $(\mu, G_k, G_{k+1}, x_0) = (1.5, 0.5, 2, 0.8)$, myopic best response toggles forever between 0.43 and 1.23; RSA converges to $x_0 = 0.8$. (b) 120 random gap instances (100 seeds, $T = 2 \times 10^4$): RSA attains $|r^- r^+|$ (87.5% within 1%, 97.5% within 3%, largest absolute error 0.003); the myopic orbit pays 6.6 to over 25 times more; the best fixed prediction $\min(r^{-2}, r^{+2})$ is lower still. Supports Theorem C.8 and Propositions C.6 and C.7.

of Theorem C.19 (T16, $n_T = \sqrt{T}$) identifies the side in 92–100% of runs and reaches excess 0.009–0.018 at $T = 10^6$, against 0.076–0.195 for RSA.

Table 7: E8: probe-based target-shifted RSA in the gap. Columns: regret/T of the probe-based learner; min, the optimum $\min(r^{-2}, r^{+2})$; RSA, the value $|r^- r^+|$ that plain RSA attains; side correct, the fraction of runs that identified the cheaper side.

instance (μ, G_k, G_{k+1}, x_0)	T	regret/T	min	RSA	side correct
$(1.5, 0.5, 2, 0.8)$	10^4	0.2276	0.1444	0.2204	1.00
$(1.5, 0.5, 2, 0.8)$	10^5	0.1789	0.1444	0.2204	1.00
$(1.5, 0.5, 2, 0.8)$	10^6	0.1586	0.1444	0.2204	1.00
$(1, 0.2, 3, 0.7)$	10^4	0.1187	0.0408	0.2363	1.00
$(1, 0.2, 3, 0.7)$	10^5	0.0715	0.0408	0.2363	1.00
$(1, 0.2, 3, 0.7)$	10^6	0.0534	0.0408	0.2363	1.00
$(1, 0, 1.5, 0.9)$	10^4	0.0603	0.0100	0.1115	0.95
$(1, 0, 1.5, 0.9)$	10^5	0.0311	0.0100	0.1115	1.00
$(1, 0, 1.5, 0.9)$	10^6	0.0191	0.0100	0.1115	1.00
$(1, 0.5, 1, 0.62)$	10^4	0.0196	0.0000	0.0008	0.88
$(1, 0.5, 1, 0.62)$	10^5	0.0053	0.0000	0.0008	0.89
$(1, 0.5, 1, 0.62)$	10^6	0.0015	0.0000	0.0008	0.88
$(2, 0.3, 1.2, 1.15)$	10^4	0.3011	0.2054	0.3340	1.00
$(2, 0.3, 1.2, 1.15)$	10^5	0.2454	0.2054	0.3340	1.00
$(2, 0.3, 1.2, 1.15)$	10^6	0.2218	0.2054	0.3340	1.00

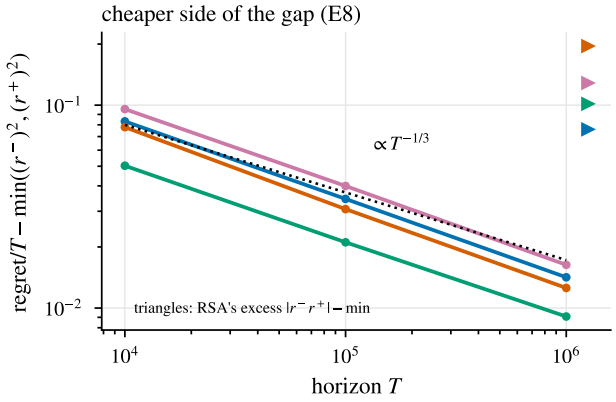

Figure 9: Probe-based target-shifted RSA: excess over $\min((r^-)^2, (r^+)^2)$ against T in the four nondegenerate instances of Table 7, with a $T^{-1/3}$ reference; triangles show RSA's excess.

H.1.6 The geometry of the rent (E10)

Figure 10 reports E10.

H.1.7 The strategic learner (E11, E12)

Figure 11 reports E11 and E12.

H.1.8 The pooled audit (E15, E16)

Figure 12 and Table 8 report E15 and E16.

H.1.9 Adaptive gain, robustness and the closed loop (E4, E13, E17)

Figure 13 and Tables 9 to 11 report E4, E13 and E17.

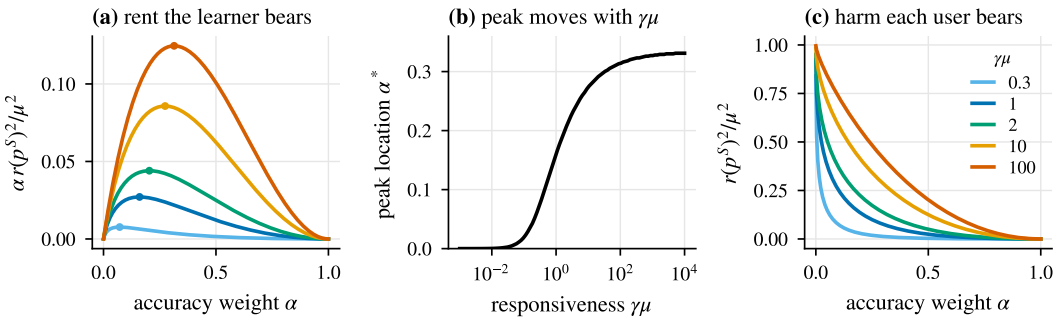

Figure 10: Exact computation. (a) The rent the learner bears peaks at an intermediate accuracy weight. (b) The peak location depends on $\gamma\mu$. (c) The harm each user bears is largest when the accuracy weight is lowest. Illustrates Theorem D.7 and Proposition D.9.

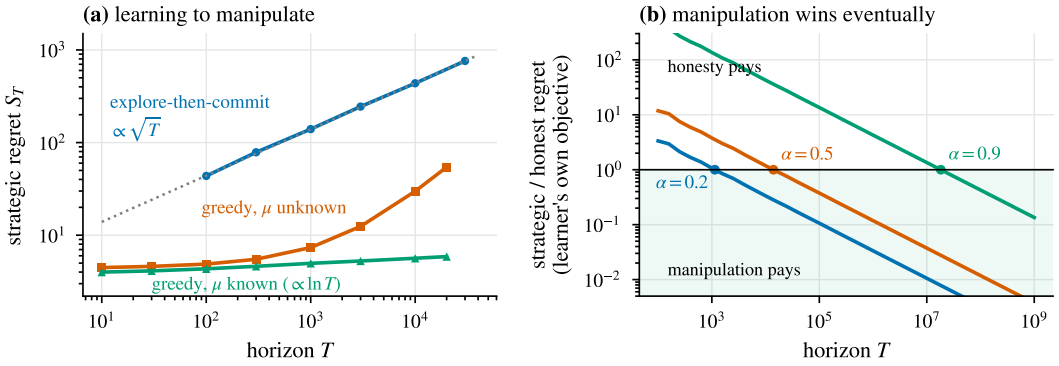

Figure 11: (a) $\alpha = 0.5$: ETC-S with μ unknown grows like $\sqrt{T}$ (factor 3.13–3.18 per decade); a greedy certainty-equivalent learner with μ unknown shows incomplete learning (1% of seeds end more than 0.05 from p^S); with μ known greedy grows like $\ln T$. (b) Ratio of strategic to honest regret in the learner's own objective; manipulation pays below 1. Interpolated crossovers $T \approx 1.1 \times 10^3$, 1.4×10^4, 1.8×10^7 for $\alpha = 0.2, 0.5, 0.9$ (untuned ETC-S, so upper estimates). Supports Theorems E.2 and E.3 and Proposition D.11; the μ-known $\ln T$ behaviour is evidence only.

Table 8: E16: honest RSA under the audit ($z = 1.96$, $T = 10^4$, 4000 seeds, starting point $\hat{p}_1 = \mu = 1$): variance of $\Sigma_T/(\sigma\sqrt{T})$ and false-alarm rate against theory. At $T = 10^4$ the $c = 0.6$ runs are pre-asymptotic: the variance is 3.66 against a limit of 5. For $c < 1$ the mean of Σ_T decays only like $T^{0.5-c}$, so the false-alarm rate at practical horizons depends on the starting point.

	linear model			the quadratic world		
c	var	rate	theory rate	var	rate	theory rate
0.6	3.660	0.4968	0.3807	3.476	0.3150	0.3807
0.8	1.608	0.1345	0.1290	1.637	0.1253	0.1290
1	1.005	0.0505	0.0500	1.025	0.0530	0.0500
1.2	0.710	0.0215	0.0204	0.716	0.0243	0.0204
1.5	0.519	0.0067	0.0056	0.521	0.0070	0.0056
2	0.331	0.0003	0.0007	0.332	0.0003	0.0007
3	0.199	0.0000	0.0000	0.199	0.0000	0.0000

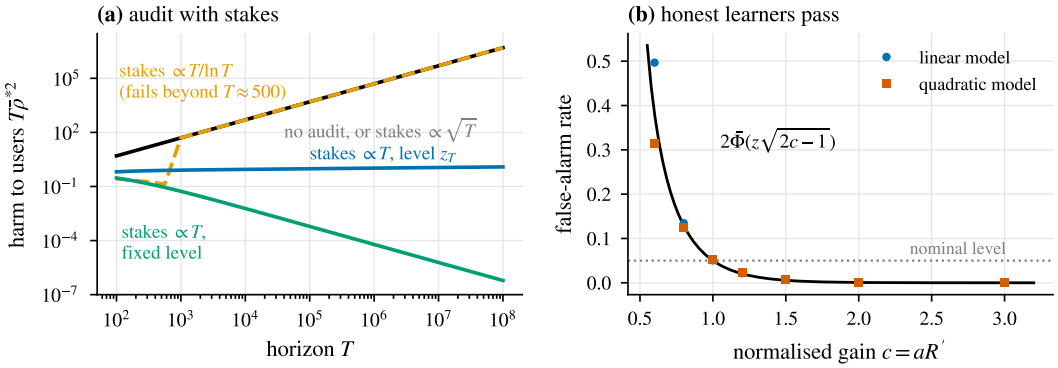

Figure 12: Here $\bar{\rho}^*$ is the optimal constant bias of Lemma F.4 and $q_z(m) = \bar{\Phi}_{\mathcal{N}}(z+m)+\bar{\Phi}_{\mathcal{N}}(z-m)$ the flagging probability at normalised mean m (Lemma F.3). (a) Total harm $T\bar{\rho}^{*2}$ from exact optimisation, T from 10^2 to 10^8, $\alpha = 0.5$: no audit or stakes $\propto \sqrt{T}$: linear; stakes $\propto T/\ln T$: disciplined until $T \approx 500$, then linear; stakes $\propto T$ at level z_T: between 0.65 and 1.22; stakes $\propto T$ at fixed level: tends to 0. A Monte Carlo check (200,000 draws) matched $q_z(m)$ to three decimals. (b) Honest RSA's false-alarm rate against $c = aR'$ follows $2\bar{\Phi}_{\mathcal{N}}(z\sqrt{2c-1})$. Supports Theorem F.5, Corollary F.6, and Propositions F.7 and F.9.

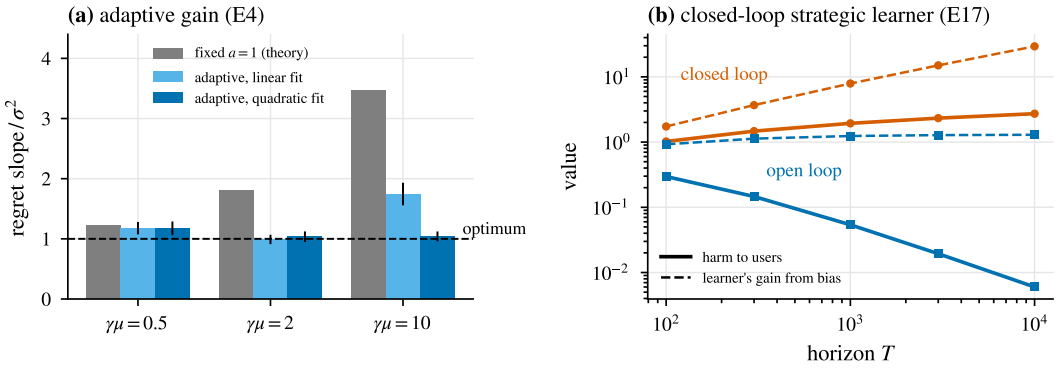

Figure 13: (a) Adaptive gain: with a quadratic local fit, RSA reaches σ^2 without knowing R' at $\gamma\mu \in \{2, 10\}$; at $\gamma\mu = 0.5$ it is indistinguishable from the fixed gain (Table 9). (b) Closed-loop strategic learner: harm grows slowly and gain grows like $\sqrt{T \ln T}$, consistent with Theorem F.18, against $O(1/T)$ harm and $O(1)$ gain in open loop (Table 11).

Table 9: E4: adaptive gain, slope/σ^2 (optimum 1). Safety factor $1+1/\ln(t+2)$; 600 seeds, $T = 10^5$.

$\gamma\mu$	fixed $a = 1$	linear fit	quadratic fit	frac. $c < \frac{1}{2}$ (quad.)
0.5	1.22	1.18 (0.10)	1.18 (0.11)	0.047
2	1.80	0.99 (0.08)	1.04 (0.09)	0.000
10	3.47	1.74 (0.19)	1.04 (0.08)	0.000

E13 (robustness). See Table 10: perception noise matches Corollary G.9 in all 8 cases; Student-t noise with 3 and 5 degrees of freedom matches the Gaussian constant; delay leaves the slope unchanged for $d \leq 5$ and adds about $5d$ to the regret level.

Table 10: E13: robustness at $\gamma\mu = 2$, slope (CI) against prediction; 1000 seeds.

experiment	a	slope	CI	prediction
perception $\tau = 0.0$	0.333	0.2580	0.0173	0.2500
perception $\tau = 0.0$	1.000	0.4444	0.0159	0.4500
perception $\tau = 0.25$	0.333	0.3204	0.0208	0.3125
perception $\tau = 0.25$	1.000	0.5536	0.0196	0.5625
perception $\tau = 0.5$	0.333	0.5120	0.0327	0.5000
perception $\tau = 0.5$	1.000	0.8889	0.0309	0.9000
perception $\tau = 1.0$	0.333	1.2820	0.0829	1.2500
perception $\tau = 1.0$	1.000	2.2382	0.0773	2.2500
Student-t, df$= 3$	0.333	0.2441	0.0158	0.2500
Student-t, df$= 3$	1.000	0.4437	0.0171	0.4500
Student-t, df$= 5$	0.333	0.2448	0.0157	0.2500
Student-t, df$= 5$	1.000	0.4531	0.0158	0.4500
Student-t, df$= 30$	0.333	0.2518	0.0159	0.2500
Student-t, df$= 30$	1.000	0.4551	0.0155	0.4500
delay $d = 0$	0.333	0.2417	0.0149	0.2500
delay $d = 0$	1.000	0.4415	0.0157	0.4500
delay $d = 1$	0.333	0.2417	0.0149	0.2500
delay $d = 1$	1.000	0.4415	0.0158	0.4500
delay $d = 5$	0.333	0.2423	0.0150	0.2500
delay $d = 5$	1.000	0.4417	0.0158	0.4500
delay $d = 50$	0.333	0.2923	0.0189	0.2500
delay $d = 50$	1.000	0.4440	0.0158	0.4500

E17 (closed loop). Exact dynamic programming over the running statistic, verified against 20,000 forward simulations (Table 11). Off-grid states are valued at $\kappa T +$ (remaining rounds) $\cdot \min W$. The closed-loop learner keeps its mean bias near $-0.9\sigma/\sqrt{T}$ and steers the statistic so that its flagging probability falls far below β.

Table 11: E17: optimal closed-loop strategic learner (exact dynamic programming) against the open-loop optimum; $\alpha = 0.5$, κ at the Theorem F.5 threshold, $z = 1.96$.

T	harm open	harm closed	gain open	gain from bias	penalty avoided	$\mathbb{P}$(flag)	DP vs sim.
100	0.2960	1.021	0.924	1.73	0.12	0.0236	10.86 / 10.88
300	0.1460	1.471	1.129	3.69	0.60	0.0073	33.92 / 33.91
1000	0.0539	1.937	1.239	7.90	2.21	0.0025	117.21 / 117.21
3000	0.0194	2.328	1.278	14.99	6.87	0.0008	360.07 / 360.13
10000	0.0060	2.723	1.294	29.36	23.22	0.0002	1220.73 / 1220.73

H.1.10 Tuning the gain from levels (E18)

Table 12 and Fig. 14 report the gains of Remark B.27 in the quadratic world with $\mu = 1$, $\sigma = 0.5$, $\hat{p}_1 = 1$ and $\mathcal{P} = [0, 5]$: 2000 seeds, regret slope over $T \in [2 \times 10^3, 2 \times 10^5]$ divided by σ^2, with 95% half-widths in parentheses. Theory is $g(aR')$ with $g(x) = x^2/(2x - 1)$. The plug-in gain is $a_t = \text{clip}(\hat{p}_t/\hat{\mu}, 0.1, 2)$, with $\hat{\mu}$ the mean of 20 plays at $p = 0$; its regret includes the cost $20\mu^2$ of those plays, a constant that does not change the slope. The lower clip is needed: without it $p = 0$ is absorbing, since the gain vanishes there. Script `e18_levelgain.py`; results `results/e18.json`. The level gain matches $g(aR')$ within Monte Carlo error at every responsiveness and stays below $4/3$, while the fixed gain grows like $\sqrt{\gamma\mu}$. The plug-in version approaches the level-gain benchmark in these runs; at $\gamma\mu = 100$ its slope, 1.337, is above $4/3$. Its analysis is open.

Table 12: E18: the level gain $a = p^*/\mu$ against the fixed gain $a = 1$, the slope-tuned gain $a = 1/R'$ and the plug-in level gain. Slopes divided by σ^2, 95% half-widths in parentheses.

$\gamma\mu$	aR', level gain	theory, level gain	level gain $a = p^*/\mu$	$a = 1$	theory, $a = 1$	$a = 1/R'$	plug-in
0.1	1.084	1.006	1.004 (0.034)	1.036 (0.034)	1.025	0.997 (0.036)	1.041 (0.038)
1	1.382	1.083	1.081 (0.031)	1.467 (0.032)	1.440	0.990 (0.035)	1.132 (0.033)
10	1.730	1.217	1.231 (0.031)	3.451 (0.041)	3.473	1.023 (0.037)	1.239 (0.030)
100	1.905	1.291	1.291 (0.030)	10.318 (0.066)	10.269	1.003 (0.036)	1.337 (0.030)

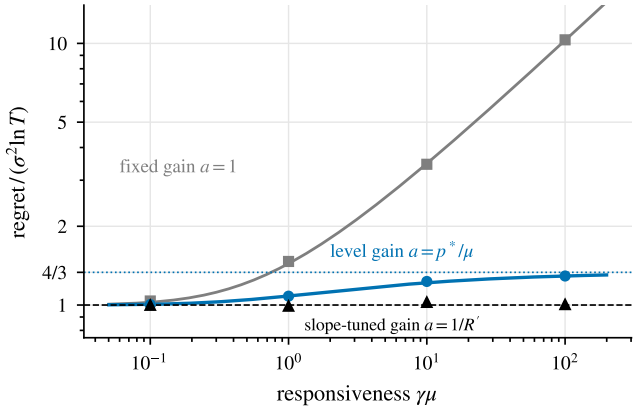

Figure 14: Regret constant of RSA against responsiveness in the quadratic world: the fixed gain $a = 1$ grows without bound, the level gain $a = p^*/\mu$ stays below $4/3$ (Remark B.27(ii)), and the slope-tuned gain attains 1. Lines are theory; marks are simulation (E18).

H.2 Verification index

Test	Script	What was checked	Result
T2	t_algebra.py	Closed forms of the quadratic world; cubic $= V'/2$; scaling; closed-form algebra; Proposition B.9 minimum	All symbolic differences exactly 0; price limit confirms $\sim \sqrt{\gamma\mu}$
T3	t_alg2.py	$c(p^*) = 2(1-\alpha)\gamma^2 {p^*}^3$; coefficient signs	Holds
T4	t_sharp.py	Theorem B.17, 4000 seeds, $T = 2 \times 10^5$	tu_t within 2.1% at 10^5 (Monte Carlo s.e. about 2.2%); slopes e.g. 0.8749 vs 0.8682
T5	t_misc.py	$I(\pi) = 4\pi^2/w^2$; sum inequality of Theorem B.29	Exact to 6 decimals; holds on all tested (b, T)
T6	t_misc.py	Theorem C.5 closed form vs iteration, 20,000 instances	Max error 3.7×10^{-13}; larger root always gives $S < 0$
T7	t_gap.py	Theorem C.8: Cesàro excess and time fraction	Within 0.0003 of $\lvert r^- r^+\rvert$; within 0.0001 of π
T10	t_strat.py	monotonicity; root location and simplicity; Δ_V; orbit vs $\lvert r^- r^+\rvert$	All reproduced; 0 violations; min $\partial_p c$ at root 0.010
T11	t_uni.py	Unimodality of ρ (log-spaced α to 10^{-12})	Unimodal for all 161 values of g
T12	t_misc.py	Positive-feedback example	Roots -0.9847, 2, 4.9847; $r'(2) = -2$
T13	t_audit.py	Lemma F.2; Lemma F.3; Jensen reduction	Convex; W_1 exact; $\Delta_W = \Delta_V$; no counterexample in 20,000 sequences

Test	Script	What was checked	Result
T14	t_audit.py	Theorem F.5, Corollary F.6, and Proposition F.7 by exact optimisation	$T\lvert\bar{\rho}^*\rvert \to 7.80$ (bound 20.1); gain $\to 2.60$ (bound 6.70)
T15	t_audit_rsa.py	Proposition F.9 in the quadratic world; Lemma F.8	Variances 1.005, 0.202, 0.499 vs 1, 0.2, 0.5; identity to 7.5×10^{-9}
T16	t_open.py	Theorem C.19	Excess over min falls with T; side correct 92–100%
T17	t_open.py	Theorem F.15	$\mathbb{E}\lvert te_t\rvert$ flat; $\mathbb{E}\lvert S_T\rvert \le 13.8$; 0 flaggings
T18	t_open2.py	Lemma E.5 algebra and curve	Symbolic differences 0; grid ratio 1.0000000000; derivative of mean along curve 2.6×10^{-7}
T19	t_open2.py	Theorem F.18(ii) vs E17	Holds at all 5 horizons
T20	t_open2.py	Proposition D.13(i) and ratio	0 violations; ratio ≥ 1.0002
T21	t_dwl.py	Proposition D.13(iii) identity	Symbolic factorisation exact; numeric agreement $< 10^{-6}$ relative
T22	t_unique.py, t_nv1.py	Theorem D.7(v): quartic derivation and α^*	Coefficient signs $(-,-,+,+,+)$ for all $y > 0$; α^* matches grid to 5 decimals at 6 values of g
T23	t_nv1.py	Theorem C.10: shifted RSA at 5 values of κ; pathwise bound on 10^5 random sequences	Points on the chord within Monte Carlo error; no violation of the bound
T24	t_nv1.py	Lemma C.20 and Theorem C.21: estimator without σ	Errors in u, v, σ^2 shrink with n ($\hat{\sigma}^2 \to 0.25$); side correct in 100% of runs
T25	t_nv2.py	Theorem C.23: $(\mathbb{E}\mathrm{Regret}_T - T\min)/T^{2/3}$	Between 0.43 and 0.69 for $T \in [10^4, 10^6]$, two instances
T26	t_nv2.py	Lemma E.9: sign of $\partial_\gamma c$ at the root	Positive on a grid of α, μ, γ (min normalised value 0.0095)
T28	t_nv3.py	Proposition C.12: R'_h, RSA slope, myopic orbit	Slope matches $R'^2_h/(2R'_h - 1)$ within 1%; myopic 2-cycle unchanged; tuned gain fails for $h \le 0.01$
T29	t_nv3.py	Theorem F.15(iv) for $a = 0.6, 0.8$	$\mathbb{E}\lvert Te_T\rvert$ flat (0.31, 0.41); $\mathbb{E}\sum\lvert e_t\rvert \approx \ln T$; 0 flaggings
T30	t_nv4.py	Proposition G.12: exact excess, $N = 1, 5, 20$	Average individual minus aggregate equals summed dispersion (0.080, 0.097); per-learner gap bounded
T31	t_stack.py, t_stack2.py, t_stack3.py	Theorem D.17: monotonicity in $g \in [10^{-3}, 10^3]$ at 5 values of α; symbolic factorisations	Harm, rent and Δ_V strictly increasing in all cases; harm $\to (1-\alpha)^2$ at large g
T32	t_circle_dwl2.py, t_circle_cex2.py	Theorem D.36: factorisation vs direct computation; counterexample	Max relative error 3×10^{-4} on 2924 responses; all 78 value-creating cases violate the q bound; explicit instance gain/harm $= 1.84$
T33	t_circle_resp.py	Theorems E.14 and E.16: derivative identities; sign conditions vs finite differences	Symbolic differences 0; agreement in 1499 of 1499 cases
T34	t_circle_audit.py	Theorem F.20: tangent condition across 7 response families	Positive margin for smooth families; fails for a steep logistic

Test	Script	What was checked	Result
T35	t_circle_alt.py	Theorem F.20(iii): alternating strategic learner vs audit	Declaration rate 0.049–0.052 vs honest 0.050; gain 0.045 per round
T36	t_vector.py	Theorem G.4: vector RSA in $d = 3$ with a nonlinear monotone response	Slope $3.13\sigma^2$ with gain R'^{-1} (prediction $3\sigma^2$); $4.59\sigma^2$ with scalar gain
T37	t_vector.py	Theorem G.7(i): incentive and direction	$\langle \nabla V(p^*), \Phi(p^*) \rangle > 0$; optimum moves against the response; harm $>$ gain at $\alpha = 0.3, 0.7$
T38	t_twofam.py, t_convex4.py	Corollary D.39: concave saturation vs the tangent condition; convex responses vs value destruction	39/140 concave cases break the tangent condition, none break value destruction; a convex counterexample to value destruction verified at 40 digits
T40	t_smoothlimit.py	Theorem D.41: regimes I, B, C as $h \to 0$	Limits matched in A and B; C tends to 0 with α-dependent sign
T41–T43	t_binaudit.py	Theorems F.23 and F.51 and Corollary F.52: binned audit vs alternating strategic learner, honest RSA (smooth and gap), accuracy-optimal learner	Strategic learner flagged 100% (pooled audit 0%); smooth RSA 0%; gap RSA 8/7/5%; cheaper-side learner 100%
T44	t_binrates.py, t_binaudit2.py	Rate algebra of Γ_T; Lemma F.48	Each term $/T^{2/3}\ell_T^{1/3}$ tends to $\sqrt{P}\sigma$, L_r, $P\rho_{\max}$; occupancy count $\times\delta \approx 0.53$ for $\delta \in [0.001, 0.02]$
T45	t_binaudit2.py, t_binaudit3.py	Theorem F.28 at the default parameters	Honest RSA never flagged (smooth, two gains); tested-bin bias event fails in $\leq 8.5\%$, falling with T; total-regret event fails in 42–100%; gap RSA flagged 2.5/2.0/0.5%; strategic learner and cheaper-side learner flagged 100%
T46	t_gapa.py	Lemma F.49; default audit, gap, $a = 0.6, 0.8$	Occupancy $\delta \times \mathbb{E}\# = 0.32, 0.43$ (law $1/\delta$); flagged 0–2.5%
T47	t_concentration.py	Theorem F.39: spread of Q_T in the linear world	Standard deviation of Q_T/σ^2 flat in T (≈ 8 at $c = 1.5$, ≈ 49 at $c = 3$); means inflated by the transient
T48	t_regimeC.py, t_regimeC2.py	Theorem D.43: next-order coefficient	Matches to four digits at $\alpha = 0.6$–0.9, converging at 0.5; signs match everywhere
T49	t_lower.py	Theorem F.41: kink world vs default audit	Never flagged; gain$/(wT + n_{\min}/w) \approx 0.04$–$0.06$, gain$/T^{2/3}\ell_T^{1/3} \approx 0.08$–$0.11$
T50	t_caseB.py, t_caseB_z3.py	Proposition F.45: small minimum counts vs honest RSA in the gap	At $z = 3$: flagged 100%, 69–77%, 37–40% for $w\, n_{\min} = 0.05, 0.2, 1$, and 2.3% at the default; at the default level, Case (b) is not yet reached at $T \leq 10^5$

Test	Script	What was checked	Result
T51	t_hwc.py	Theorem F.39: weighted Schur constants for $c = 0.6$–1.5; spread at $c = 0.75$	Schur bounds $6.1, 3.7, 2.5, 1.2$ (continuous analogue $1/(c - \frac{1}{2})$); $\|G\|_F^2 / \ln T$ bounded; standard deviation of Q_T/σ^2 grows slowly ($7.4 \to 9.6$)
T52	t_localsharp.py	Lemma B.20 and Theorem B.21: projected RSA with $2am = 0.4 < 1 < 2aR' = 2$, 4000 runs, $T \le 10^6$	Regret increment per decade $/(\sigma^2 \ln 10)$: $1.26, 1.03, 1.02$ (start 2.2) and $1.09, 1.02, 1.02$ (start 0.0), against the predicted 1; transient about 6–10
T53	t_vecsharp.py	Theorem G.5: matrix-gain learner with H-projection, gradient response in $d = 3$	$d = 3$, gradient response with symmetric non-diagonal Jacobian, 3000 runs: regret increment per decade $/(\sigma^2 \ln 10)$ 2.99 ($10^3 \to 10^4$) and 3.03 ($10^4 \to 10^5$), against the predicted 3
T54	t_signlocal.py, t_wedge.py	Propositions D.27 and D.30: wedge share and the first-order harm ratio	Wedge share $\varepsilon/(1+\varepsilon)$ to four decimals for seven power laws; harm ratio at $s = 10^{-3}$ matches the limit to four decimals for five power laws
T55	t_regimes_global.py	Theorem D.34: 3000 random draws over four response families	Sign theorem 3000/3000; single crossing held in all draws
T58	t_auditic.py	Theorem F.30: honest unprojected RSA in a linear world against the binned audit, default and raised levels	No false alarm at either level ($c \in \{0.75, 2\}$; $T \le 4 \times 10^4$; 1500–4000 runs); $\mathbb{E}Q_T/\sigma^2 \approx 19$ against $\ell_T \approx 20$ at $T = 10^4$, so the proof's margin is not yet active at these horizons
T59	t_auditic_nl.py	Lemma F.32 and Theorem F.33: The quadratic world residual, projected RSA with $c = 1.5$, 3000 runs, $T \le 5 \times 10^4$	$\mathbb{E}\exp(\theta t e_t^2) \le 1.004$ at all checkpoints ($\theta = 0.05$); $\sum_t r^2 - 4\tilde{Q}_T$ negative in every run (maximum $-0.56 \ln T$)
T60	t_gapedge.py	Theorem F.36: honest RSA ($a = 1$) in the gap against the default binned audit, 7200 runs, $T \le 4 \times 10^4$	False-alarm rate 0.024–0.025 at every horizon (no decay in T); every false alarm had x_0 within $2/n_{\min}$ of a bin edge, none otherwise
T61	certify.py, certify_regimes.py	Computer-assisted proofs (Remark D.37) of Propositions D.38, D.40 and F.21 and Corollary D.42	All certificates pass; every coded derivative checked against high-precision finite differences.

All T-scripts were rerun independently on a second machine (Windows, same seeds) with identical output. T8, T9 and T1 are analytic or covered by the rows above.

H.3 Extended related work: persuasion, cheap talk and price impact

Bayesian persuasion. In Kamenica & Gentzkow (2011) a sender commits to a signal structure to influence a receiver's action through the receiver's posterior. The learner with a stake has the same

cast: the state is the fundamental μ, the sender is the learner, the signal is the published forecast $\hat{p}_t$, the receiver is the world, its action is the intensity $\lambda_t = \gamma\hat{p}_t$, and the sender's payoff is $-V(\hat{p}_t)$. Three differences prevent a reduction. (i) There is no commitment: the learner chooses $\hat{p}_t$ after forming its belief and may change strategy at any round, which is closer to cheap talk (Crawford & Sobel, 1982); the accuracy weight α interpolates between pure cheap talk ($\alpha \to 0$) and forced honesty ($\alpha = 1$). (ii) The world responds to the forecast directly, through its cost, not through a posterior about μ. (iii) Publishing generates information for the sender: $y_t = \mu - \gamma\hat{p}_t^2 + \eta_t$ is observed, so the choice of $\hat{p}_t$ affects what the learner learns, which is what makes learning to manipulate cost $\Theta(\sqrt{T})$ (Theorem E.3). Whether an index policy combines exploration, persuasion and reputation in one score is open.

Price impact. Reading the outcome as a price, the intensity λ is a price impact per unit of forecast, as in Kyle (1985). There the impact is set by competitive market makers facing informed and noise traders; here it is chosen by a rational follower minimising its own cost. We use the analogy for interpretation only.

Credibility, reputation and costly talk. Reputation as a discipline on a sender who gains from misleading appears in Sobel (1985) and Bénabou & Laroque (1992); distortion of a publicly observed signal that the market sees through is the signal-jamming model of Stein (1989); and costs of misreporting enter through lying costs and costly talk in Kartik (2009) and Kartik et al. (2007). Boeken et al. (2025) study conditional forecasts and proper scoring rules for performative predictions.

I Open problems

1. **Scope of the strategic results.** The learning results cover weak positive feedback (Proposition B.6), but every strategic result (Appendices D.1, D.4 and E.2) concerns self-negating forecasts, where the world's response lowers the outcome. Self-fulfilling forecasts are outside the strategic analysis.
2. **The binned audit as an incentive mechanism.** Proved in linear and smooth nonlinear worlds (Theorems F.30 and F.33); impossible in the gap for binned audits with bins fixed before play (Theorem F.36); audits with bins placed after play are open.
3. **Closed-loop harm.** Theorem F.18 proves harm $O(\sqrt{T \ln T})$ under the level z_T; the dynamic programming of E17 suggests $\Theta(\ln T)$. The flagging probability at a fixed level is also open.
4. **Adaptive gain in expectation, and for the state-projected RSA.** Proved almost surely for the clipped-publication variant (Proposition B.24); fixed gains are settled in every sense (Theorem B.21 and Proposition B.25). The two adaptive statements that remain open are explained in Remark B.26.
5. **Tightness of $t|\hat{p}_t - x_0|$ for gains $a < 1$** in the gap (Theorem F.15(i)); part (iv) already covers the audit for every $a > \frac{1}{2}$.
6. **The cheaper-side rate** (Theorem C.23). $T^{2/3}$ is probably *not* optimal: once the learner is near x_0, choosing a side is a two-armed bandit in which each side's residual is observed directly, which suggests $\tilde{O}(\sqrt{T})$ or better. An algorithm achieving this, and a lower bound, are open.
7. **The time scale of smoothing**: when smoothed behaviour takes over from gap behaviour (Proposition C.12).
8. **Several learners with projection, heterogeneous gains, or strategic incentives** (Proposition G.12).
9. **Feature-conditioned predictors** in the gap (Remark C.11).
10. **One condition for all of the scaled-setting results.** By Corollary D.39, a unifying condition must be two-sided in curvature (value destruction fails only for nonconcave, the tangent condition only for nonconvex responses) and global on $\mathcal{P}$ for the tangent condition. Whether such a condition implies Theorems D.36, E.14 and F.20 at once is open, as is the analysis beyond one-parameter scaling.
11. **The smoothing family and the tangent condition**: the analogue of Theorems D.41 and D.43 for the tangent condition.

12. **Beyond the audit family.** Theorem F.46 shows $T^{2/3}$ is unavoidable, up to a logarithm, for every audit of Definition F.22 that accepts honest learners in the gap. Whether audits outside the family (statistics using the pattern of bin counts, bins that refine over time, tests of within-bin dispersion) can do better is open, as is closing the factor $\ell_T^{1/3}$ between the bounds. The lower bound when the offset is kept secret is also open.
13. **A sophisticated follower** that anticipates the leader's manipulation and discounts the forecast (a signalling or signal-jamming model); whether the rent survives and whether distortion persists without gain is open.
14. **Order of play**: if the world adapts slowly and the learner quickly, the roles may reverse (Zrnic et al., 2021); the consequences for Theorem D.17 are open.
15. **The plug-in level gain.** Remark B.27 tunes the gain from the level p^*/μ; its plug-in version with $\hat{p}_t/\hat{\mu}$ approaches that benchmark in simulation (Appendix H.1.10), but neither its expected nor its almost-sure constant is proved.