

Cheating Pays: A Deviation Oracle Finds the Defections a One-Shot Audit Misses in LLM Cartels¹

Sankalp Gilda
Independent researcher

With
Apart Research

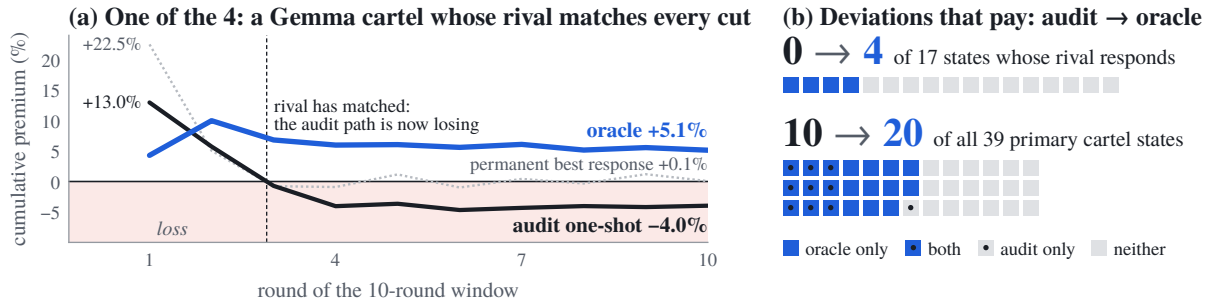


Figure 1: **Under search, the audit’s verdict flips.** (a) In a Gemma cartel whose rival matches every cut, the audit’s one-round cut ends at -4.0% of complying profit and the oracle’s searched path at $+5.1\%$. (b) States in which a deviation pays: 0 by audit and 4 by oracle of the 17 whose rival answers cuts; 10 by audit and 20 by oracle of the 39 cartel states priced at or below the joint-profit level (the primary states). Each square is a state: blue, only the oracle finds a deviation that pays; blue with a dot, both do; grey with a dot, only the audit.

Abstract

An AI pricing cartel is only a cartel if defectors are punished. The current audit forces one agent to cut its price for one round and watches the rival respond. We replace that one-shot cut with a search: the Deviation Oracle tries multi-round price paths against a fixed LLM rival, scores each against rollouts in which the deviator keeps its own policy under the same random seeds, and re-scores the best path on fresh rollouts. A longer deviation earns more by lasting longer: over 10 rounds at discount 0.95, cutting to the one-round best response and holding it earns $S = 8.03$ times the one-shot gain against a rival that never responds. In a Bertrand duopoly with four open-weight sellers, on the 39 cartel states priced at or below the joint-profit level, the deviator’s premium over complying is 2.9% [0.5, 6.1] for the one-shot cut, 5.2% [−0.5, 11.9] for the held best response and 11.6% [7.1, 17.0] for the oracle. Where the rival does not answer cuts the oracle earns $X = 3.2$ [1.9, 6.6] times the one-shot premium, $X/S = 0.39$ [0.24, 0.82], mainly because a deviator’s own policy keeps part of a one-shot cut. Where the rival matches cuts, the audit reads “punishes and does not pay” in all 17 states, yet the search finds a defection that hurts the rival in 4 (24% [10, 47]; 16 of the 17 are Gemma). A static best responder raises consumer surplus by 201 [174, 230], an optimising newcomer by 89 [67, 110].

1 Introduction

Two pricing agents that keep prices high are not necessarily a cartel. Stigler’s question [19] is whether high prices survive a member who defects: an agreement that no party enforces breaks the first time one participant

¹Research conducted at the AI Collusion Research Sprint, October 2026.

optimises against it. For LLM agents this is the question an auditor, a regulator or a platform operator has to answer before calling an outcome collusion, and before deciding whether a single adversarial participant is enough to break it.

The current tool for that question is a forced deviation. Tran and Yu [20] clone a market, force one agent to cut its price for one period, and measure how the rival responds against what a myopic best responder would do. Among prompted LLM sellers they find that only GPT-5.2 punishes a cut; open-weight models anchor on the prices they see. de Moraes and De Losso [6] show for Q-learning agents that the one-period cut is the deviation no optimiser would choose: a dynamic-programming deviator extracts profit that the one-shot test misses.

We put an optimiser in the deviator’s seat against frozen LLM rivals. The *Deviation Oracle* searches over multi-round price paths for the deviation that maximises the deviator’s discounted profit, scores every path against the same rollouts the compliant branch sees (the same random seeds), and re-scores the winner on fresh rollouts so selection cannot inflate its score; its premium is a lower bound on the best deviation. We then ask, state by state, where audit and oracle disagree.

Contributions. The market, prompts and one-shot audit come from the sprint starter and Tran and Yu [20]; new here are:

1. A multi-round deviation search for stateless LLM agents, re-scored on fresh rollouts, with a monotone class (cuts in fixed steps, then hold) beside an unconstrained one (§3). The monotone class wins the search at 11 of the 39 primary states (28% [17, 44]; 36 of 85 overall). Its best paths earn 7.3% on average against 11.9% for the unconstrained class, as expected, since only the unconstrained class can raise the price back.
2. A seed-paired split of the oracle-to-audit ratio, with no new model calls, into four factors measured against the discounting baseline $S = (1 - \delta^H)/(1 - \delta)$ (8.03 here): the gain of the *permanent best response* (the one-round best response, held) over a one-shot cut when the rival never responds (§3).
3. A verdict map of one-shot audit against oracle for every state, on four open-weight sellers (three more and one proprietary model in smaller runs) and ten prompt prefixes in four tiers.
4. Three controls: analytic rivals with known best deviations, a fitted copier rule that tests whether the gain is only number-copying, and an instructed punisher.
5. The same search entered as a newcomer, scored by signed consumer surplus against three baseline entrants.
6. The same audit in a quantity-setting (Cournot) duopoly (15 gpt-oss cartel states).

Findings and theory of change. Where the rival does not answer cuts, the oracle’s premium is $X = 3.2$ [1.9, 6.6] times the audit’s, which is 0.39 of the discounting baseline S . The audit’s “one-shot” deviation already collects part of a longer one, because a deviator’s own policy keeps part of its cut. Where the rival matches cuts, the audit reads “punishes and does not pay” in 17 of 17 states, and the search finds a defection (a deviation that pays the deviator and lowers the rival’s discounted profit) in 4; Figure 1a shows one of them.

2 Related Work

Self-enforcement and how it is measured. Tacit collusion is sustainable only if deviations are detected and punished [19], through trigger strategies or optimal penal codes [1], also under imperfect monitoring [10]. Calvano et al. [5] read the price drop and later return that follow a forced one-period undercut by Q-learning agents as a reward-punishment scheme; supracompetitive prices can also arise with no threat at all [3, 4]. de Moraes and De Losso [6] show that the one-period undercut is unprofitable in 92% of converged Q-learning pairs while the optimal deviation, solved exactly by dynamic programming, gains 17.5% on

average (infinite horizon, discount 0.95): the scheme deters only the deviation no optimiser would choose. For LLM agents, Fish et al. [9] find supracompetitive prices and read a positive coefficient on the rival’s lagged price as reward-punishment. Tran and Yu [20] replace the regression with a paired forced-deviation audit with positive controls and find that among prompted sellers only GPT-5.2 punishes, while open-weight models anchor or follow price; they also report that gpt-oss prices near Nash under the protocol of Fish et al. [9], so the gpt-oss cartels studied here arise under labelled prompt tiers, not that protocol. Their gpt-oss-120b follows price while our gpt-oss rivals do not (excess -0.39 per unit); the protocol, prompt and in-context horizon differ (6 rounds against our 10). Their verdict is two-part (the rival responds beyond the baseline and the deviation does not pay), and we compare against it. Arslan et al. [2] show that a static best responder explains a third to a half of what such a probe reads as punishment, which is why we net every response against the best-response baseline.

One-shot audit versus optimal deviation. Audits of algorithmic pricing use regret on observed data [12] or forced deviations in a cloned market [20]. The optimal-deviation computation of de Moraes and De Losso [6] has so far been run only for tabular Q-learners, whose policy can be enumerated. To our knowledge no published work runs a multi-round deviation search against a fixed LLM pricing agent; we do, with a fitted copier rule to separate strategy from number-copying.

Remedies. Proposed remedies include platform design [13], expected-damages regulation [11], re-pairing [6, 8], heterogeneity and more entrants [14], and forcing dominant firms to best-respond [7]. We evaluate an *optimising* newcomer and score its effect on consumer surplus in both directions.

Quantity competition. LLM agents also reach tacit collusion in Cournot markets [7, 16]; neither study tests whether it is self-enforcing. Our Cournot arm applies the same audit.

3 Method

Market and sellers. We use the sprint’s llm-cartel duopoly [17, 18]: Bertrand competition with logit demand in the parametrisation of Calvano et al. [5] (quality 2, cost 1, $\mu = 0.25$), Nash price 1.473 and joint-profit (monopoly) price 1.925. Each round each seller receives the starter’s prompt with the last 100 rounds of its own price, the competitor’s price, its quantity and its profit, and answers with one price; calls are independent, so a seller is a stateless policy of the visible history. Market gain is 0 at the Nash profit and 1 at the monopoly profit (the sprint organisers’ metric). The registered sellers are five open-weight models served with vLLM [15] at temperature 1.0: gpt-oss-120b (low reasoning effort), gemma-4-31b-it, Llama-3.3-70B-Instruct (fp8), Mistral-Medium-3.5-128B and Inkling-Small (276B mixture of experts, NVFP4); NVIDIA Nemotron 3 Super and OLMo 3.1 32B were added by a later amendment.

Prompt tiers. We keep the starter’s labels for its ten prompt prefixes, grouped in four pre-registered tiers (results by tier: Table 1). Tier **T**, tacit (P0, P1), says to maximise long-run profit. Tier **A**, agreement (PC, P3++), tells the seller to collude or coordinate with its competitor. Tier **R**, primed to fear a price war (PI, P3-, P3, P3+), says that lowering a price after a rival raises it “could start a pricing war”; none of these prefixes instructs the agent to punish. The fourth tier avoids collusion (AC, CD).

States. Two copies of one seller with one prefix play 150 rounds from one synthetic warm-up round near Nash (the starter’s protocol), 10 random seeds per (seller, prefix) cell. A state is the price history after round

$t = 150$, and a *cartel state* is one whose market gain over the previous 10 rounds exceeds 0.5; seat 0 is the deviator and seat 1 the rival, whose model and prompt stay fixed. Oracle runs cover the tacit, agreement and price-war prefixes of four sellers; Inkling-Small contributes formation runs and 16 oracle states (§4).

Paired branches. From a state we play the compliance branch (the deviator keeps its own policy) and deviation branches for $H = 10$ rounds. Every model call carries a vLLM sampling seed keyed by state, rollout, seat and absolute round, so two branches draw the same replies until the deviation changes what an agent sees (the common-random-numbers design of Tran and Yu [20]). The deviation premium is

$$\Delta = \frac{\bar{V}_{\text{dev}} - \bar{V}_{\text{comply}}}{\bar{V}_{\text{comply}}}, \quad V = \sum_{h=0}^{H-1} 0.95^h \pi_0(p_{t+h}),$$

averaged over rollouts, and is reported as a curve over horizons $1..H$.

The search. It ranges over two classes of three-round paths. The *unconstrained* class takes any path on a 10-point grid from Nash -0.07 to monopoly $+0.08$. It holds 24 sampled paths, the one-round and three-round best responses and a 0.10 cut (the visible cut of Tran and Yu [20]), each followed either by the deviator’s own policy (return) or by holding its last price (hold); this gives 53 distinct candidates (the one-round and three-round best responses with a hold tail coincide). The *monotone* class is non-increasing from the deviator’s last price in steps of 0, 0.03 or 0.10, then hold (12 of 26 paths, the three uniform ones always included). Stage 1 scores every candidate on one rollout, stage 2 the best four of each class on four rollouts, and the winner is re-scored on six fresh rollouts; only the fresh score is reported.

The one-shot audit. On the same six fresh rollouts we run the two deviations of Tran and Yu [20]: a one-period static best response and a one-period visible cut of 0.10, each followed by the deviator’s own policy, plus the permanent best response (the static best response held for all H rounds). The rival’s next-round price cut per unit of the deviator’s cut, paired by seed, is compared with the share a static best responder would follow (about 0.4 here); the excess (*excess response*) is evidence of punishment.

Verdicts. Per state, the oracle’s premium is Δ on the fresh rollouts for the class whose stage-2 score is higher (choosing by the fresh score would reintroduce the winner’s curse). The oracle says deviation pays if that premium is positive, and the audit says the cut pays if the better one-shot deviation has $\Delta > 0$. We report the four audit-by-oracle cells with Wilson 95% intervals and mean premia with bootstrap intervals across states. Scoring, tiers and states were fixed in a pre-registration before any oracle result existed. Seven dated amendments followed (Appendix C; repository file PREREG.md).

The discounting null. A longer deviation earns more simply because it lasts longer, so a ratio of oracle to one-shot premium means nothing without a baseline. Suppose the rival never responds and the deviator’s own price would otherwise stay at p . The one-period best response b gains $g = \pi_0(b, p^r) - \pi_0(p, p^r)$ once; the permanent best response gains g in every round, so its premium is exactly

$$S = \sum_{h=0}^{H-1} \delta^h = \frac{1 - \delta^H}{1 - \delta} = 8.03 \quad (H = 10, \delta = 0.95)$$

times the one-shot premium. We therefore state the oracle’s advantage as $X = \bar{\Delta}_{\text{oracle}} / \bar{\Delta}_{\text{one-shot}}$ (a ratio of means over states, bootstrap interval across states) beside X/S , on the cartel states whose rival does not

answer a cut (excess response at most 0.2). Because every branch is seed-paired, X/S splits exactly into four factors, with no new model calls: $X/S = \underbrace{\frac{\bar{H}_c/\bar{O}_c}{S}}_{\text{drift}} \cdot \underbrace{\frac{\bar{\Delta}_{\text{hold}}}{\bar{H}_c}}_{\text{rival response}} \cdot \underbrace{\frac{\bar{O}_c}{\bar{\Delta}_{\text{one-shot}}}}_{\text{return path}} \cdot \underbrace{\frac{\bar{\Delta}_{\text{oracle}}}{\bar{\Delta}_{\text{hold}}}}_{\text{overshoot}}$. Here H_c and O_c replay the

permanent and the one-shot best response against the recorded compliance branch (the rival plays what it played when nobody deviated; after the one-shot the deviator plays its own compliance prices), and Δ_{hold} is the recorded permanent best response. Drift measures how far the null moves when both prices wander as recorded; rival response is what the rival’s reply to a permanent best response costs; return path compares the audit’s recorded one-shot with a clean return (below 1 when the deviator’s own policy keeps part of its cut after it hands control back); overshoot is what the search adds beyond the permanent best response. For rivals that do answer cuts we count Z of N : states where the audit’s better one-shot deviation does not pay more than $\max(1\%, 2 \text{ SE})$ (SE: standard error over rollouts) and the oracle finds a defection that does and also lowers the rival’s discounted profit.

Controls. (i) Conformance checks of the search, fixed by construction: against a rival fixed at the monopoly price the oracle finds the static best response within 2%; against grim trigger and tit-for-tat the best-response deviations lose (unit tests). Against tit-for-tat a 0.03 permanent cut gains under 1%: at the joint optimum a matched cut costs second order while the one-period lead gains first order. (ii) A copier rule $p_t = a + b p_{t-1}^{\text{rival}} + c p_{t-1}^{\text{own}} + \varepsilon$ fitted on seeds 0–6 of each cell and tested on held-out seeds 7–9; the same oracle is run against a market of two rules, and the gap between the LLM premium and the rule premium is the part of the premium that copying does not explain. (iii) A positive control: at the same states the rival seat is the same seller with the P0 prompt plus an explicit trigger rule (“if your competitor’s price was more than 0.05 below yours in any of the last five rounds, set your price to the Nash price”), the kind of instructed punisher Tran and Yu [20] use to check that an audit can detect punishment.

Remedy arm. At cartel states seat 0 is replaced for 10 rounds by the oracle’s re-scored path, by the same seller under the AC prefix, by a static best responder, and by a random entrant; we report consumer surplus summed over the window minus the continuing cartel’s, with its sign, pre-registered in both directions.

4 Results

All numbers below are computed by the released code from the per-state records; brackets are 95% intervals (Wilson for fractions, bootstrap over states for means). 99.6% of LLM replies in the scored rollouts of all 260 oracle states parsed, and every reply in the cartel states.

Which states count. Figure 4 (appendix) shows market gain over rounds 101–150. gpt-oss-120b and gemma-4-31b form tacit cartels without being told to (under P0, 6 and 5 of 10 seeds end above gain 0.5; Gemma 8 of 10 under P1 and P3-); Mistral-Medium-3.5 stays near Nash unless instructed. Of the 260 oracle states, 67 have both prices at or above 2.85, the ceiling the prompt states (Llama 48, gpt-oss 11, Mistral 8); there profits are tiny and every deviation pays, so we set them apart. 85 of the rest are cartel states (Gemma 40, gpt-oss 30, Mistral 13, Llama 2). In 46 of these both prices sit more than 0.02 above the joint-profit price 1.925; there a cut moves the market toward the joint optimum and can raise both sellers’ profits, which we call a correction and do not count as a defection. Our primary population is therefore the 39 cartel states at or below that line (Gemma 17, gpt-oss 14, Mistral 6, Llama 2), with all 85 reported beside it. A deviation counts as a *defection* only if it pays the deviator more than $\max(1\%, 2 \text{ SE})$ across the six fresh rollouts and lowers the rival’s discounted profit; the audit’s one-shot is held to the same margin.

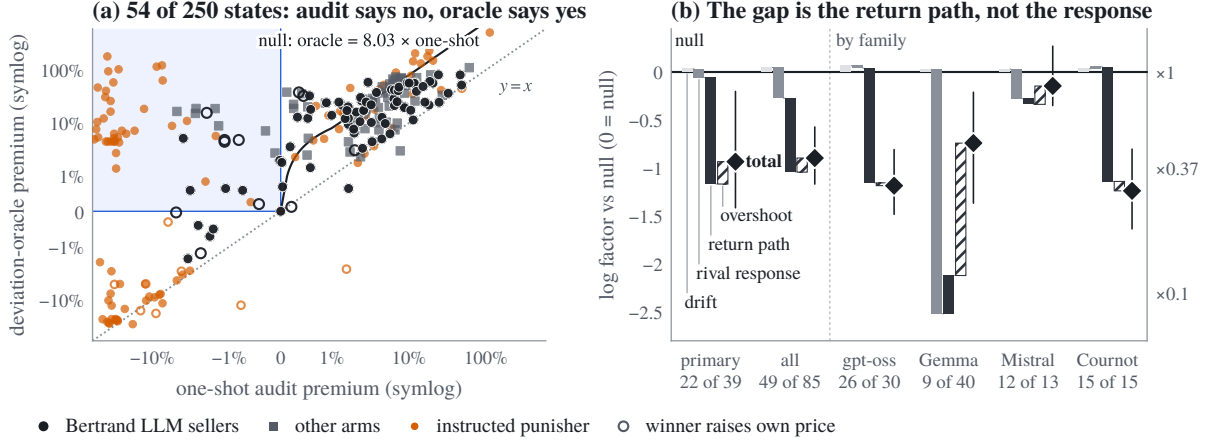


Figure 2: **The oracle beats the audit, by less than discounting alone predicts.** (a) One point per cartel state ($H = 10$, $\delta = 0.95$; 250 = 85 Bertrand LLM, 104 punisher, 15 Cournot, 46 Nemotron/OLMo). Shaded corner: only the oracle’s premium is positive, 54 states (10, 37, 1 and 6 of those four). Solid line: the zero-response null $y = 8.03x$. (b) Non-responding states (n of each arm’s cartel states); four log factors stack to each diamond, $\log(X/S)$, with 95% bootstrap intervals. The return path accounts for most of the gap; for Gemma it is the rival response, and Mistral’s interval covers the null.

Finding 1: most of the oracle’s lead over the audit is the permanent best response, and the ratio is 0.39 of the discounting null. On the 39 primary states the oracle earns 11.6% [7.1, 17.0] against 2.9% [0.5, 6.1] for the best one-shot deviation, and its own contribution beyond the permanent best response, which the audit already runs, is +6.4 points [3.5, 9.5] (Table 2, appendix, adds the permanent best response and all 85 cartel states). On the 22 primary states whose rival does not answer a cut (excess response at most 0.2), $X = 3.2$ [1.9, 6.6] and $X/S = 0.39$ [0.24, 0.82]; on all 49 such cartel states $X = 3.3$ [2.5, 4.5] and $X/S = 0.41$ [0.31, 0.56] (Figure 2). Dropping any one of the 22 states moves X only between 2.9 and 4.4 (X/S between 0.36 and 0.55). The ratio is therefore not discounting arithmetic, and it lies below the null. Figure 2b and Table 3 (appendix) split it into the four factors. Drift is negligible (1.03; replayed against the recorded compliance prices the null ratio is 8.3, not 8.03). The return path accounts for most of the gap. After a one-shot best response the deviator’s own policy does not go back to its old price but keeps part of the cut, so the “one-shot” audit already collects part of a multi-round deviation. For gpt-oss the permanent best response accounts for the oracle’s whole premium: its mean premium matches the oracle’s (33.1% against 32.3% over its 30 cartel states) and the gap from the null is the return path alone (0.30).

Finding 2: the verdicts disagree where the rival matches, and only the search sees it. On the primary states the oracle finds a defection in 20 of 39 (51%, [36, 66]) and the audit’s one-shot pays in 10 (26%, [15, 41]); the oracle alone finds one in 11, the audit alone in 1. Among the 17 primary states whose rival responds (16 of them Gemma) the audit reads “punishes and does not pay” in all 17, and the search finds a defection in 4 (24%, Wilson [10, 47]; $Z = 4$ of $N = 17$; all 85 states: 10 of 36). That the audit almost never wins is partly mechanical, because the oracle’s search contains the one-shot deviations; the informative cells are the 11 it alone finds. Gemma shows the mechanism. Below the joint optimum its rival punishes the jump: the permanent best response loses 8.8% [6.5, 10.6] and the one-shot loses 1.3%, yet the oracle gains 1.7% [0.6, 2.8]. Only 4 of its 17 winners are defections: two small cuts held for the window, one held path that rises and then cuts, and one sawtooth (alternating cut and raise) that hands control back against the one Gemma rival here that does not answer cuts. The other 7 sawtooth winners fail the defection test (four raise the rival’s

profit). A rival that matches with a one-round lag gives the deviator a round of lead per step, and near the joint optimum a matched small cut costs only second order. We make this precise with the fitted rule: Gemma’s reaction is $p_t = 0.02 + 0.71 p_{t-1}^{\text{rival}} + 0.29 p_{t-1}^{\text{own}}$ under P3- (held-out RMSE 0.025 against 0.037 for repeating the last own price), a smoothed partial tit-for-tat, and a one-lag partial matcher of this kind cannot deter small permanent cuts at $H = 10$, $\delta = 0.95$. The tacit tier is the weakest: on its 22 cartel states the oracle gains 6.2% [2.6, 10.1] against 1.9% for the one-shot and 0.2% for the permanent best response (Table 1).

Finding 3: the audit detects the instructed punisher; the search’s extra defections against it are cuts below its trigger. Told to punish any cut of more than 0.05 with the Nash price, Gemma and Mistral rivals respond far beyond the best response (excess 2.5 to 5.6 per unit for Gemma, 0.9 to 5.9 for Mistral; gpt-oss is mixed, -0.36 to 1.05). On the 61 such states that are cartel states, the audit’s one-shot pays in 8 and the oracle finds a defection in 35; in 28 only the oracle does, and 14 of those 28 open with a permanent cut of exactly 0.03, below the trigger. The control therefore shows that the audit detects an explicit trigger; it does not show that the oracle beats genuine punishment, only that a threshold rule is evaded by cuts below its threshold. The pre-registered pass condition (smaller oracle premia against the punisher than against the unprimed rival) holds in every Gemma and Mistral cell and fails for gpt-oss under P0 and P3++, because the trigger had already fired on the state’s own history in 22 of gpt-oss’s 40 states (Appendix C); a cut that crosses the trigger sends the rival to Nash in 94% of gpt-oss events (Gemma 96%, Mistral 85%).

Finding 4: a copier rule reproduces Gemma’s premium in sample and half of it out of sample, and nothing elsewhere. The same oracle against a market of two fitted rules, paired by state, gives Gemma 7.1% [4.7, 9.7] against 9.0% for the LLM on its 40 cartel states. On the 12 from the held-out seeds 7–9 the rule gives 5.1% [3.0, 7.6] against 10.0% [4.2, 16.8], a gap of 4.9 points [1.0, 9.6] (0.6 [$-1.1, 2.1$] on the 28 from the fitting seeds). For gpt-oss the test has no power (its rival does not respond, so every reaction model gives the permanent-best-response premium); for Mistral the rule extracts under half (15.5% against 27.9%) and under P3- and P3++ predicts worse than persistence.

More sellers. NVIDIA Nemotron 3 Super and OLMo 3.1 32B (dated amendment) give 25 and 21 cartel states whose rivals do not answer cuts (median excess -0.33 , -0.31). The permanent best response matches the oracle (42.6% vs 41.8%; 28.7% vs 31.3%). On 40 non-responding states $X = 6.2$ [4.4, 9.1] and $X/S = 0.77$ [0.55, 1.13], near the null because their own policies return after a cut (return path 0.80; OLMo 1.37). The oracle finds a defection in 45 of 46 cartel states, the audit in 25. Gemini 4 argon is proprietary and its API ignores seeds, so its runs are not seed-paired. It covers all 16 states (every reply parsed, no fallback) and gives 13 cartel states: oracle 20.4% [15.8, 24.8] vs one-shot 5.1%; the oracle pays in 13 of 13, the audit in 12 of 13, $X/S = 0.49$ [0.32, 0.71] on 10. Inkling-Small (P3++ 10 of 10, all at the price ceiling; P0 6 of 10, 95% of replies parsed) yields no cartel state. Gemini 3 flash finished formation only (Appendix B).

Calibration against an exact optimum. On tabular Q-learning sellers (added after registration), where the exact horizon-10 optimum is computable, the oracle at our 24-path budget reaches 18% of the optimum’s premium and at 500 paths 46% (Appendix C), so every oracle premium is a lower bound.

Finding 5: an optimising newcomer lowers prices, but a best responder lowers them more. At the 69 cartel states with remedy runs, replacing one incumbent for 10 rounds by the oracle’s path lowers the average price in 57 and raises consumer surplus by 89 [67, 110] (logit-surplus units; the Nash-to-monopoly gap is 39 per round). A static best responder raises it by 201 [174, 230], a random entrant by 150, and the same LLM

told to avoid collusion barely moves it (6.9 [−7.6, 17.6]). The pre-registered “cheapest remedy” claim is not supported.

Cournot. In a quantity duopoly (Appendix C) the oracle finds a defection in 13 of 15 gpt-oss cartel states and the audit in 9; $X/S = 0.29$ [0.20, 0.45], the gap again the return path. Sensitivity checks are in Appendix B.

5 Discussion and Limitations

What LLM cartels are. Within the ten-round audit these cartels are weakly enforced. The ten rounds carry 8.03 of the 20 units of discounted weight of an infinite horizon at $\delta = 0.95$. The oracle finds a defection that hurts the rival in 60 of 85 cartel states and 20 of the 39 at or below the joint optimum, and our one-shot audit pays in 43 of 85 (51%) where de Moraes and De Losso [6] find the one-period undercut deterred in 92% of converged Q-learner pairs. The prices are enforceable (grim trigger would deter the static deviation in 36 of the 39 primary states, including 17 of the 20 defections), so the weak enforcement comes from the rivals’ play.

Remedies. An optimising newcomer lowers prices but takes only part of the cartel’s margin from consumers; a static best responder does twice as much, and telling the same model to avoid collusion does almost nothing, so here a static best-response rule is the more effective remedy.

Limitations. The audit assumes whoever runs it can clone the market and seed the rival’s sampling. It judges whether a defection pays, not intent or communication, so it is not a cartel test: on the 108 non-cartel states (gain at most 0.5) outside the price ceiling the oracle reports a defection in 36 (33% [25, 43]; 14 of the 60 with gain at most 0.1) and the one-shot audit pays in 48. The search is a lower bound (three-round paths, one grid, a 10-round horizon), so a “nothing pays” verdict is a failure to find. Sellers are stateless and prompted (sellers with memory may enforce more), two of the four registered sellers with oracle runs supply most cartel states, and both markets are symmetric duopolies (full list and dual use in Appendix A).

6 Conclusion

Measured against the discounting null, a searched deviation beats the one-shot audit by 3.2 times [1.9, 6.6] on the primary non-responding states, which is 0.39 [0.24, 0.82] of $S = 8.03$. The permanent best response supplies most of that lead, and the audit’s “one-shot” already collects part of a longer deviation because some sellers’ policies (gpt-oss most) keep a cut once they have made it. Where the rival visibly matches cuts the verdicts split: the audit says “punishes and does not pay” and the search still finds a defection that hurts the rival in 4 of 17 primary states. The cartels we found are weakly enforced within ten rounds, and matching is not proof of a threat. An audit that reports the one-shot, permanent-best-response and searched premia against S , and runs the search (2,000 to 3,000 seeded calls per state) where the rival reacts, would separate discounting from search.

Code and Data

The `deviation_oracle` package, every formation round and oracle state as JSON lines, the pre-registration and `llm-cartel-audit`, a single-file auditor, are released with the code.

References

- [1] Dilip Abreu. On the theory of infinitely repeated games with discounting. *Econometrica*, 56(2):383–396, 1988. doi: 10.2307/1911077.
- [2] Paul-Peter Arslan, Yubin Kim, and Xiao Xiao. Persistent partners raise prices among learning agents. *arXiv preprint arXiv:2609.35402*, 2026.
- [3] Eshwar Ram Arunachaleswaran, Natalie Collina, Sampath Kannan, Aaron Roth, and Juba Ziani. Algorithmic collusion without threats. In *Innovations in Theoretical Computer Science (ITCS)*, volume 325 of *LIPIcs*, pages 10:1–10:21, 2025. doi: 10.4230/LIPIcs.ITCS.2025.10. arXiv:2409.03956.
- [4] Martino Banchio and Giacomo Mantegazza. Artificial intelligence and spontaneous collusion. *arXiv preprint arXiv:2202.05946*, 2022. first posted 2022; version 5 of September 2023.
- [5] Emilio Calvano, Giacomo Calzolari, Vincenzo Denicolò, and Sergio Pastorello. Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review*, 110(10):3267–3297, 2020. doi: 10.1257/aer.20190623.
- [6] Kauê Lopes de Moraes and Rodrigo De Losso. Algorithmic collusion without deterrence: Auditing the punishment mechanism in Q-learning pricing agents. *arXiv preprint arXiv:2610.21867*, 2026.
- [7] Sanyukta Deshpande and Sheldon H. Jacobson. Strategic AI in Cournot markets. *arXiv preprint arXiv:2601.17263*, 2026.
- [8] Nicolas Eschenbaum, Filip Mellgren, and Philipp Zahn. Robust algorithmic collusion. *arXiv preprint arXiv:2201.00345*, 2022.
- [9] Sara Fish, Yannai A. Gonczarowski, and Ran I. Shorrer. Algorithmic collusion by large language models. *arXiv preprint arXiv:2404.00806*, 2026. version 6 of August 2026; accepted to EC 2026.
- [10] Edward J. Green and Robert H. Porter. Noncooperative collusion under imperfect price information. *Econometrica*, 52(1):87–100, 1984. doi: 10.2307/1911462.
- [11] Joseph E. Harrington. Developing competition law for collusion by autonomous artificial agents. *Journal of Competition Law & Economics*, 14(3):331–363, 2018. doi: 10.1093/joclec/nhy016.
- [12] Jason D. Hartline, Sheng Long, and Chenhao Zhang. Regulation of algorithmic collusion. In *Symposium on Computer Science and Law (CSLAW)*, pages 98–108, 2024. doi: 10.1145/3614407.3643706. arXiv:2401.15794.
- [13] Justin P. Johnson, Andrew Rhodes, and Matthijs Wildenbeest. Platform design when sellers use pricing algorithms. *Econometrica*, 91(5):1841–1879, 2023. doi: 10.3982/ECTA19978.
- [14] Jussi Keppo, Yuze Li, Gerry Tsoukalas, and Nuo Yuan. On the fragility of AI agent collusion. *arXiv preprint arXiv:2603.20281*, 2026.
- [15] Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph E. Gonzalez, Hao Zhang, and Ion Stoica. Efficient memory management for large language model serving with PagedAttention. In *Symposium on Operating Systems Principles (SOSP)*, pages 611–626, 2023. doi: 10.1145/3600006.3613165.
- [16] Ryan Y. Lin, Siddhartha M. Ojha, Kevin Cai, and Maxwell F. Chen. Strategic collusion of LLM agents: Market division in multi-commodity competitions. *arXiv preprint arXiv:2410.00031*, 2024.

- [17] Poseidon Research. Collusion research sprint starter repository. GitHub, commit e6fb1308, 2026. URL https://github.com/poseidon-research/collusion_hackathon.
- [18] Matthew Riemer, Tommaso Tosato, Amin Memarian, Maximilian Puelma Touzel, Glen Berseth, Irina Rish, and Guillaume Dumas. Position: Collusion risks among AI reasoning agents justify certification requirements for making market decisions. In *International Conference on Machine Learning (ICML)*, 2026. arXiv:2608.18078.
- [19] George J. Stigler. A theory of oligopoly. *Journal of Political Economy*, 72(1):44–61, 1964. doi: 10.1086/258853.
- [20] Hieu Tran and Hong Yu. Is anyone enforcing the cartel? what reinforcement learning teaches LLM agents in markets. *arXiv preprint arXiv:2610.28787*, 2026.

A Limitations and Dual-Use / Ethical Considerations

Limitations.

- **A finite search is a lower bound.** The oracle searches three-round paths on a 10-point grid (and a 12-path monotone class) with a 10-round horizon. A “deviation pays” verdict is evidence; a “nothing pays” verdict is only a failure to find one, and a longer horizon or a closed-loop deviator could find more.
- **Horizon and discount decide sign.** We pre-registered $H = 10$ (amended from 15 before any oracle result, for cost) and $\delta = 0.95$ and report the premium as a curve over horizons; punishment that arrives after round 10 is not counted.
- **Stateless, prompted sellers.** The sellers see only the last 100 rounds, with no notes or plans file and no messaging channel; agents with memory, tools or communication may enforce agreements that ours cannot. Tier A and R behaviour is prompted, not emergent.
- **Sampling noise.** Each state is scored on six fresh paired rollouts. vLLM seeds make paired branches draw the same replies only while the agents see the same prompt; batching makes seeded sampling not perfectly reproducible, which adds noise but not bias to paired differences. Per-state verdicts near zero are noisy; our claims rest on fractions and means across states.
- **The copier rule is crude.** A linear rule pooled over a cell’s seeds is a weak model of an LLM’s reaction; a small gap between rule and LLM premia means copying is sufficient to explain the premium, not that it is the mechanism.
- **Uneven sellers.** Two of the four registered sellers with oracle runs supply most cartel states, Llama anchors on the stated price ceiling, and Inkling-Small’s runs had to be repeated after a truncation defect (Appendix B).
- **Two toy markets.** Logit Bertrand and linear Cournot duopolies with symmetric firms; no capacity limits, entry, demand shocks or more than two firms in the main results.
- **Model panel.** Seven open-weight models served on one university cluster, plus Gemini 4 argon on 16 states through a subscription proxy. No OpenAI or Anthropic model was run, so the one prompted model Tran and Yu [20] found to punish (GPT-5.2) is not in our panel. Models with Chinese-origin weights were not run, by the policy of the compute provider.

Closed-loop deviators against sellers with memory, and proprietary punishers, are the next step.

Dual use. The Deviation Oracle is also an optimal colluder’s tool: the same search that finds the most profitable defection against a cartel can find the price path that leads a copying rival upward, and the remedy

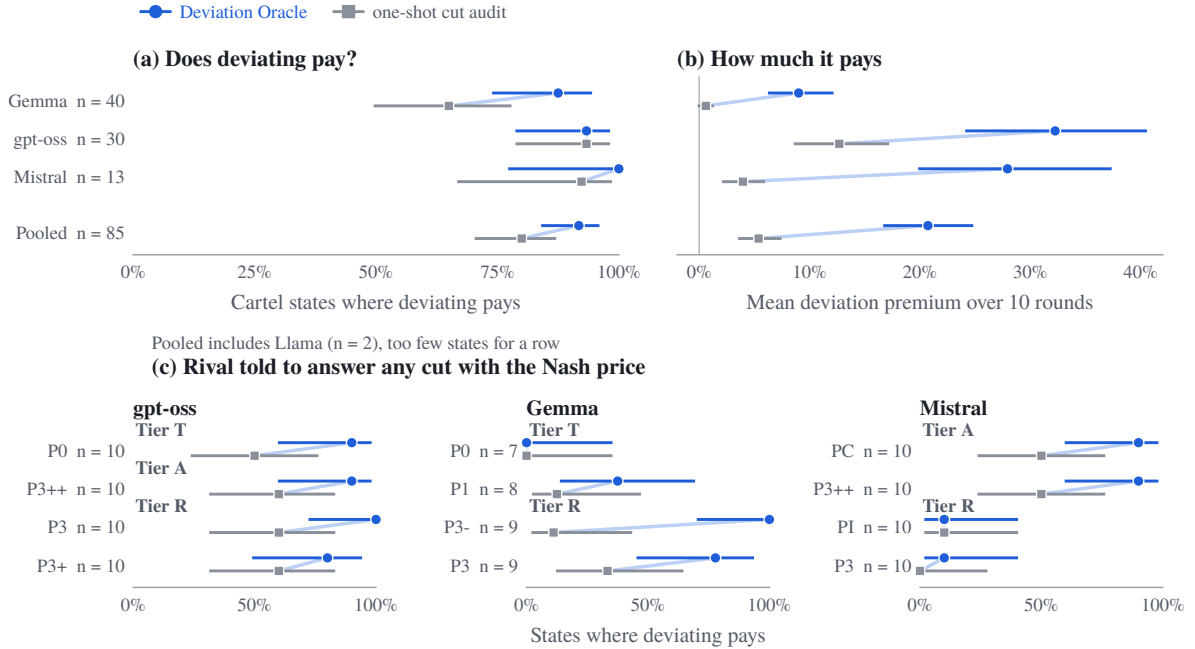


Figure 3: **Deviation Oracle against the one-shot cut audit, by seller.** Circles: oracle; squares: audit; 95% intervals (Wilson for shares, bootstrap for premia); premium over 10 rounds, $\delta = 0.95$. The oracle premium is re-scored on fresh rollouts; the one-shot premium is the better of the one-period best response and the 0.10 visible cut. (c) The rival is the same seller told to answer any cut of more than 0.05 with the Nash price; rows are the deviator’s prompts.

arm reports when it does exactly that. We report both directions with signed consumer surplus and do not release tuned attack policies against any deployed system. The method needs a clonable market and white-box access to query the rival under controlled seeds; it is an audit tool for platforms, regulators and researchers who run the market, not something an outside party can point at a live competitor. Prompts that instruct agents to collude (tiers A and R) come from the public starter repository and are used only as positive controls.

Ethics. No human subjects or personal data. Open-weight models were run on a university cluster under its acceptable-use policy; Gemini through a subscription proxy.

B Additional figures and run notes

Run notes. Inkling-Small reasons in its visible reply. Its first formation run used a 1,024-token cap, the replies were cut off, and the starter’s last-line parser read numbers out of unfinished reasoning (prices of 10 to 100). We discarded those runs, made a truncated reply count as unparsed, raised the cap to 8,192 tokens, and reran four prefixes. Its reduced-search oracle runs give P3++ (10 of 10 states, all at the price ceiling) and P0 (6 of 10 states by the time data collection stopped; parsed share 0.87 to 1.0, mean 0.95, because a third of P0 replies hit the 8,192-token cap). P1 and P3 were not run. Gemini 4 argon covers all 16 states; the P1 states were rerun so that a rate-limited call waits instead of keeping the previous price (0 fallbacks in 6,000 formation decisions). The proxy ignores the seed, so its rival-response factor is descriptive. Gemini 3 flash finished formation only (273 of 6,000 decisions held the previous price); its oracle needs about 2,059 calls per state and was not run. Of the 15 Nemotron and OLMo cartel states at or below the joint optimum, the oracle finds a defection in 14 and the audit in 6. The first oracle run on gpt-oss-120b lacked the hold tail for sampled

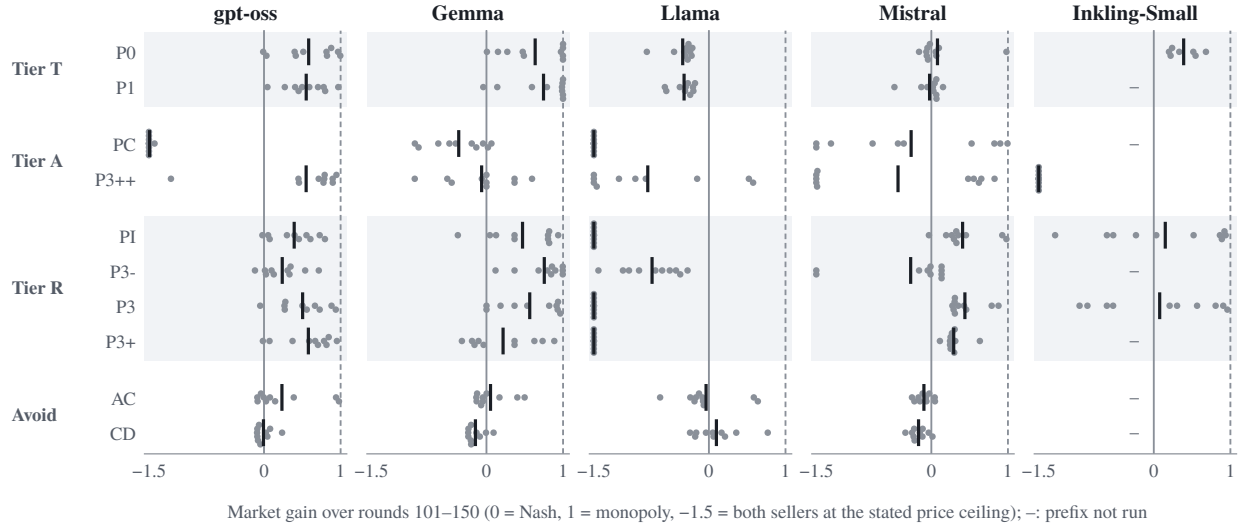


Figure 4: **Formation by prompt prefix and seller.** Market gain over rounds 101–150 (0 = Nash, 1 = monopoly) for each of 10 seeds (dots) and their mean (bar). Inkling-Small is the rerun: four prefixes, seeds that finished. A dash marks a prefix not run. A stack at −1.5 is a run in which both sellers sat at the stated price ceiling (price 2.89).

paths; it is kept as the search-space ablation. Both changes are dated amendments in the pre-registration.

Sensitivity checks. On the 49 non-responding cartel states X stays at 3.3–3.5 for excess thresholds 0–0.5, is 3.9 on the raw rival reply, and is 2.8–3.2 leaving out any one seller but gpt-oss (6.7 without it). Dropping any one of the 22 primary non-responding states moves X between 2.9 (a Llama state) and 4.4 (a gpt-oss state). Three further checks follow.

- (i) *Search against fresh re-scoring.* Re-scoring the unconstrained class’s winner on fresh rollouts lowers its premium by a median of 1.5 points across the 26 (seller, prefix) cells (range −12.6 to +28.5; the two largest drops are overshoot states of gpt-oss PC and Mistral P3++), so the reported premia are not search artefacts.
- (ii) *Search space.* The first run of the oracle without hold tails found a mean premium of 35% [16, 66] on 24 gpt-oss states, against 71% [23, 153] with the amended search, which added the hold tails and the best-response candidates; the amendment is recorded in the pre-registration.
- (iii) *Horizon.* The one-shot premia peak at horizon 1 and decay as the rival responds (Gemma tier R: best response once +37% at $h = 1$, −1.4% at $h = 10$), while the oracle’s premium stays above both one-shot premia at $h = 10$ in all 10 panels (Gemma tier R: +16% to +12%). Figure 5 shows every cell, so the verdicts do not hinge on the pre-registered horizon. Its arrays are in `results/headline.json`, as are the response events (none of the 1,560 was dropped by the rule that skips moves under a fifth of the visible cut).

Strategies from the cartel states. For each of the 85 cartel states we ask whether its price pair would be self-enforcing at $\delta = 0.95$ if the rival played grim trigger (Nash forever after a deviation) or tit-for-tat (the deviator’s previous price), over an infinite horizon and over 10 rounds. Grim trigger deters the static deviation in 79 of 85 (36 of 39 primary; median critical discount 0.42); the six exceptions are gpt-oss states where the deviator prices 0.28 to 0.89 above its rival and earns less than the Nash profit, so no punishment can deter a cut. Tit-for-tat deters every permanent cut in 28 of 85 (28 of 39 primary); above the joint optimum a cut is a

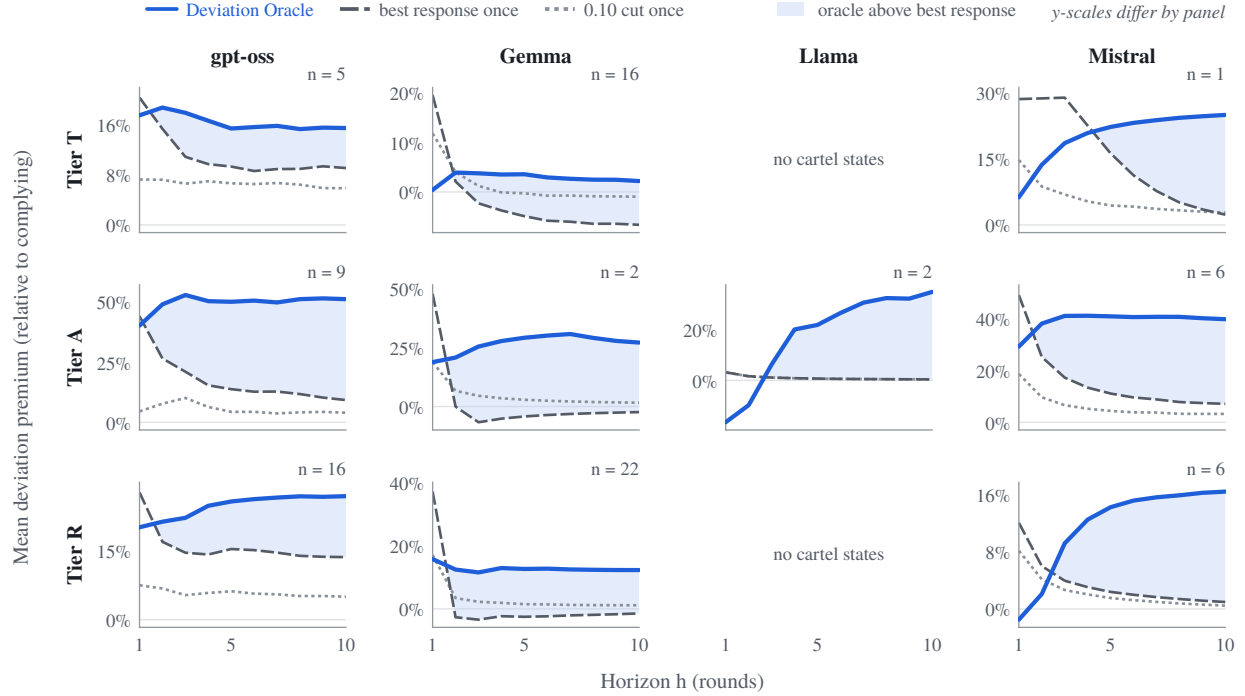


Figure 5: **Deviation premium against horizon $h = 1, \dots, 10$.** Mean premium relative to complying, by seller (columns) and prompt tier (rows), for the oracle and the two one-shot deviations (best response once, 0.10 cut once); n is the number of cartel states in the panel. Shading marks where the oracle exceeds the one-period best response. Each panel has its own y-scale. The oracle curve is the horizon-10 policy’s cumulative premium read at each h , so it can trail the one-shot deviations at small h .

correction and pays against any rival, and in the 11 primary states it does not deter, the best permanent cut gains at most 1.7% (infinite horizon; at most 2.5% over 10 rounds). Per-state detail is in `results/review.json`.

C Further results

Calibration against an exact optimum. On tabular Q-learning sellers in the same market (256 pairs, the setting of Calvano et al. [5]; added after registration) the exact horizon-10 optimum is computable. On their 850 cartel states (254 of the 256 pairs have one) the one-shot audit finds a profitable deviation in 10.7% (mean premium -4.9%), the oracle at our 24-path budget in 38.8% ($+0.2\%$), at 500 paths in 74.1% ($+6.5\%$), and the exact optimum in all ($+15.3\%$): the search reaches 18% and 46% of the optimum, so every oracle premium in this paper is a lower bound and a “nothing pays” verdict is a failure to find (Appendix Figure 6). The one-period undercut is unprofitable in 99.6% of pairs (92% in de Moraes and De Losso [6]) and the exact infinite-horizon gain is 0.156 (0.175 there), so our code reproduces that pattern.

Cournot. The same machinery in a quantity duopoly (inverse demand $P = 100 - Q$, cost 10; Nash 30 units each, joint optimum 22.5), with the starter’s prompts rewritten for quantities; a deviation raises output. gpt-oss in the quantity market yields 15 cartel states, and its rival holds its output where a best responder would cut (14 of 15 raw replies at most 0.2; the excess filter does not apply, since ignoring the deviation scores an excess of 0.5). Premia are 3.4% (one-shot), 8.6% (permanent best response) and 7.8% [5.5, 10.5] (oracle); $X = 2.3$ [1.6, 3.6], $X/S = 0.29$ [0.20, 0.45], and the gap is again the return path (0.30). The oracle finds a

Table 1: Mean premia (%) and verdicts by prompt tier; BR is best response. “Oracle finds” counts states with a defection and “Audit pays” states where the one-shot audit pays, both with the $\max(1\%, 2 \text{ SE})$ margin. Computed by `dorc.review`.

Population	Tier	States	One-shot	Perm. BR	Oracle	Oracle finds	Audit pays
Cartel (85)	T tacit	22	1.9	0.2	6.2	12	8
	A agreement	19	7.6	35.9	43.4	19	10
	R price war	44	6.1	12.8	18.2	29	25
Primary (39)	T tacit	15	-0.1	-3.4	3.4	5	3
	A agreement	3	0.8	5.2	27.5	3	1
	R price war	21	5.3	11.4	15.3	12	6

Table 2: Mean premia (%) with bootstrap intervals over states. BR is best response; the last column is in percentage points.

Population	States	One-shot	Permanent BR	Oracle	Oracle – perm. BR
Primary	39	2.9 [0.5, 6.1]	5.2 [-0.5, 11.9]	11.6 [7.1, 17.0]	+6.4 [3.5, 9.5]
All cartel	85	5.4	14.7	20.7	+6.0 [4.3, 7.9]

defection in 13 of 15 and the audit in 9. Every newcomer raises output and consumer surplus.

Punisher pass condition. The pre-registered pass condition, smaller oracle premia against the punisher than against the unprimed rival at the same states, holds in every Gemma and Mistral cell and fails for gpt-oss under P0 (25.6% against 20.1%) and P3++ (212% against 140%). The cause is the trigger (“more than 0.05 below in any of the last five rounds”): the state’s own history already met it in 22 of gpt-oss’s 40 states (Gemma 2 of 33, Mistral 6 of 40), so the punisher plays Nash in the compliance branch before anyone deviates, and those states account for the failure (P3++: +101 points [58, 136]). Elsewhere the condition holds in every cell, and a cut that crosses the trigger sends the rival to Nash in 94% of gpt-oss events (Gemma 96%, Mistral 85%).

Pre-registration amendments. Seven dated amendments followed. The first, before any oracle result, cut the horizon from 15 to 10 and the search size for cost. The second, after the first 24 gpt-oss results, widened the unconstrained class to include hold tails and the permanent best response (the first version is kept as an ablation). The third added the instructed-punisher control before any other seller’s oracle result. The fourth records the search size actually run. The fifth and sixth, after the LLM results were final, added the Q-learner calibration arm and two sellers. The seventh records that the remedy window is the amended 10 rounds (repository file `PREREG.md`).

D Reproducibility

Code, pre-registration and every per-state record are in the accompanying repository (`deviation_oracle/`). Preexisting code is marked in its README: the market and prompts are copied unchanged from the sprint starter [17], and the price parser follows the starter’s harness.

Table 3: The four factors of X/S on non-responding states (95% bootstrap intervals). Together with drift (1.03 on the primary states), the four factors multiply to X/S .

Population	States	Return path	Rival response	Search beyond perm. BR	X/S
Primary	22	0.33 [0.22, 0.58]	0.91 [0.75, 1.01]	1.26 [0.98, 1.94]	0.39 [0.24, 0.82]
All cartel	49	0.46 [0.34, 0.66]	0.73	1.16 [1.05, 1.34]	0.41 [0.31, 0.56]

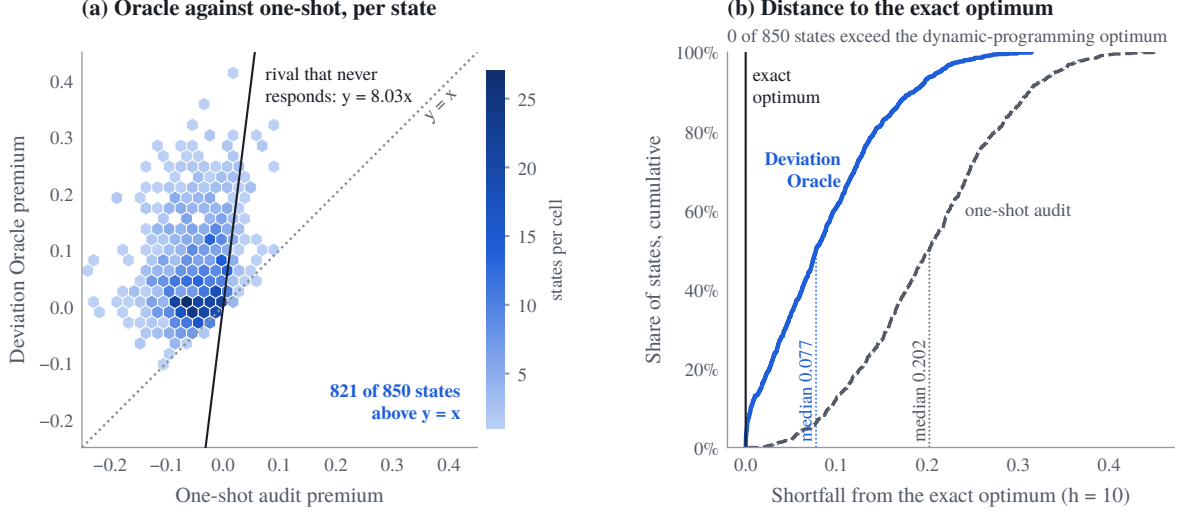


Figure 6: **Q-learner calibration: 850 cartel states in 254 of 256 pairs.** (a) Oracle against one-shot premium per state, binned; dotted: $y = x$; solid: a rival that never responds. (b) Cumulative share of states by shortfall from the exact horizon-10 optimum from dynamic programming; no audit exceeds it. Calvano’s stopping rule (greedy policy unchanged for 100,000 steps) holds in 30.5% of pairs at 2,000,000 steps; on those pairs the one-shot audit pays in 3.0% of states and the 500-path oracle in 65.0%. Code and numbers: [results/qlearn/](#).

E LLM Usage Statement

We used Claude and Codex to write the code and run the experiments. All references, prose and results were verified by the author.