STRATEGIC TESTING IN GAMES

Anonymous authors
Paper under double-blind review

ABSTRACT

When chess world champion Magnus Carlsen accused Hans Niemann of cheating in September 2022, it spotlighted a challenge that extends well beyond chess: how can one audit a competitor's play while protecting honest players from false accusations? We propose a framework for auditing a monitored player relative to known reference strategies in two-player normal-form games. In particular, we introduce (ϵ, δ)-strategic testing, which limits the flag rate under honest reference play to δ and ensures that no deviating strategy achieves an expected accepted winning payoff more than ϵ above the honest benchmark, uniformly over opponent actions. We characterize the optimal acceptance policy through a linear program, derive a direct threshold rule, and compare it with uniform acceptance. We extend testing to zero-payoff outcomes through separate flag budgets and exposure scores, and connect the framework to deviation costs and incentive compatibility, showing that payoff control alone need not preserve honest play as a best response. We also formulate the auditing problem as an auxiliary zero-sum Stackelberg game. Finally, we evaluate the framework through synthetic experiments comparing optimal and uniform auditing across random games and examining the trade-off between payoff control, incentives, and honest reward retention.

1 INTRODUCTION

Auditing strategic behavior requires understanding how actions affect rewards. Departing from a reference strategy can leave payoff unchanged, while shifting probability toward a valuable action can increase it. We study how an auditor can use this payoff structure to control credited rewards.

In our model, two players interact in a normal-form game. The auditor knows the payoff table and reference strategies, observes both players' actions, and decides whether to accept or flag the outcome. An accepted win credits its positive payoff. A flagged outcome credits zero. The goal is to bound the accepted winning payoff available to any deviating strategy while limiting the probability of flagging honest reference play. A flag is an audit decision and does not by itself establish misconduct.

The budget δ bounds flagging under honest reference play, while ϵ bounds accepted winning payoff above the honest benchmark W^+. We illustrate these conditions with a two-action game and compare uniform acceptance with an optimal policy.

The auditing mechanism is visible to the players. A policy that minimizes the largest accepted payoff need not make honest play a best response. We therefore also consider constraints on deviation gains and on the rewards retained by honest players. These constraints lead to a linear program. An auxiliary Stackelberg game describes the auditor's commitment against an evaluator who selects the worst deviation and opponent action.

Related work. The relevant literature addresses several distinct inferential questions. Behavioral classification uses features of observed play to identify cheating, for example from mouse movements and keystrokes (Pinto et al., 2021). Chess-specific work studies engine analysis and move preferences (Barnes & Hernandez-Castro, 2015; Laarhoven & Ponukumati, 2022). Distribution testing asks whether a sampled distribution has a specified property, such as uniformity or equality to a reference (Canonne, 2020; Paninski, 2008; Valiant & Valiant, 2017). Classical hypothesis testing studies how to distinguish statistical hypotheses while controlling error rates and seeking high power (Neyman & Pearson, 1933). Robust estimation seeks accurate estimates despite contamination or model misspecification (Diakonikolas & Kane, 2023). Our objective is to design an intervention: acceptance

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.

probabilities determine which rewards are credited, subject to an honest flag budget and a bound on accepted payoff.

The work of Gauthier et al. (2026) is particularly close to our strategic framing. They develop sequential tests for violations of strategic benchmarks. In repeated normal-form games, these tests use payoff evaluations and Nash, correlated, or coarse correlated equilibrium conditions. In stochastic games, they monitor compliance with a target policy. The tests accumulate evidence while controlling statistical errors over time. Our auditor instead commits before play to a policy for withholding rewards, bounding the payoff available under any deviation. This distinguishes detection of benchmark violations from design of the rewards credited after an outcome.

Other related approaches include strategic classification (Hardt et al., 2016), inspection games (Avenhaus et al., 2002), auditing strategic agents (Estornell et al., 2021; 2023; Das et al., 2026), and strategic hypothesis testing (Bates et al., 2026; Chen et al., 2025). Models of human play and skill estimation motivate our reference strategies (McIlroy-Young et al., 2020; Tang et al., 2024; 2025; Regan & Haworth, 2011). Our main analysis treats these strategies as known. Appendix C examines reference estimation empirically, with fixed payoffs and a fixed opponent reference.

Contributions. We formalize (ϵ, δ)-strategic testing and illustrate its geometry (Section 3), then derive an optimal threshold policy (Section 4). We extend the framework to zero-payoff outcomes, exposure scores, and deviation costs, and distinguish payoff control from incentive compatibility (Section 5). Our primal–dual LP characterization also accommodates honest reward protection, best responses, correlated obedience, and prescribed Nash equilibria (Section 6). Finally, we interpret the auditor's optimization as an auxiliary zero-sum Stackelberg game (Section 7). Proofs, algorithms, and synthetic experiments are given in the appendix.

2 Preliminaries

For a finite set S, let $\Delta(S)$ denote the probability simplex on S. A two-player game has finite action sets $\mathcal{P}_1$ and $\mathcal{P}_2$ and a payoff function $U : \mathcal{P}_1 \times \mathcal{P}_2 \to \mathbb{R}$ for Player 1. Payoffs are normalized so that $U(p_1, p_2) \in [-1, 1]$. Write $U^+(p_1, p_2) := \max\{U(p_1, p_2), 0\}$ for the positive part. A mixed strategy for Player i is a distribution in $\Delta(\mathcal{P}_i)$. When both players' incentives matter, write $U_1 = U$ and let U_2 denote Player 2's payoff table.

3 The (ϵ, δ)-Strategic Testing Framework

3.1 Auditing model and honest reference

We work first in a full-information setting: the auditor knows the action sets $\mathcal{P}_1, \mathcal{P}_2$, the payoff table U, and fixed reference strategies $\mu_1 \in \Delta(\mathcal{P}_1)$ and $\mu_2 \in \Delta(\mathcal{P}_2)$. The auditor observes play and chooses a decision from $\mathcal{A} := \{\textsc{Accept}, \textsc{Flag}\}$. The goal is to limit the accepted winning payoff available under any player strategy, while limiting the probability of flagging honest play. The framework applies to general-sum games, including draws, and initially counts only positive payoffs. Section 5.2 extends this score to other outcomes and action costs. The honest expected winning payoff is

$$W^+ := \sum_{p_1, p_2} \mu_1(p_1)\mu_2(p_2)U^+(p_1, p_2). \tag{1}$$

A policy assigns an acceptance probability $\alpha(p_1, p_2) \in [0, 1]$ to each action pair. The auditor chooses this policy using the payoff table, the reference strategies, and the budget δ. When we add incentive constraints, pairs with equal payoffs may receive different acceptance probabilities because they affect the player's choices differently.

The auditor publicly commits to α before players choose their strategies. Players know the acceptance probabilities and may respond to them. The audit's random draw occurs only after play. After observing both actions (p_1, p_2), the auditor accepts with probability $\alpha(p_1, p_2)$ and flags otherwise. Thus the same monitored-player action can receive different acceptance probabilities against different opponent actions. Flagging gives zero immediately credited payoff in the model. Any later review or repayment is outside its scope. For a feasible policy, the retained honest probability on a pair

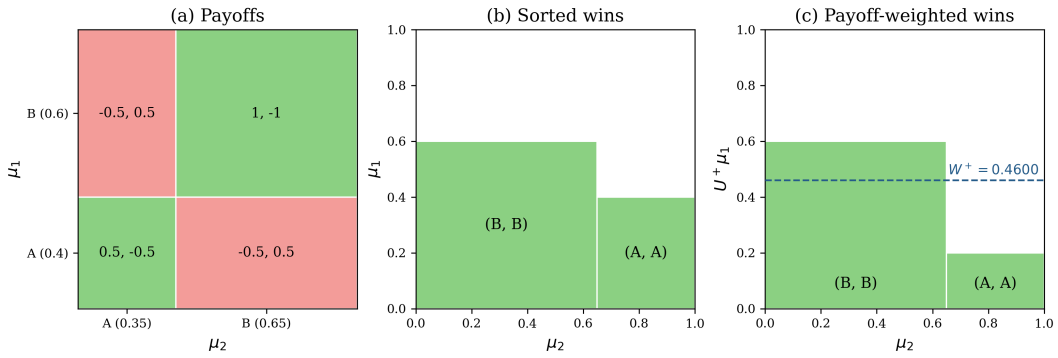

Figure 1: Geometry of the two-action example. (a) Joint action regions, with both players' payoffs. Green denotes Player 1's wins and red losses. (b) Winning regions sorted by area. (c) The same regions with payoff-weighted heights. Widths remain opponent probabilities. The dashed line marks $W^+ = 0.46$ as a numerical reference. The total green area in (c), not its maximum height, equals W^+.

is $\mu_1(p_1)\mu_2(p_2)\alpha(p_1, p_2)$, and its total is at least $1 - \delta$. The references remain fixed when players deviate: they specify the behavior that the audit is designed to protect.

To illustrate the model, consider auditing Player 1 in the two-action game in Table 1. Its unequal winning rewards make the difference between reference probability and payoff visible.

	Player 2	
Player 1	A	B
A	$0.5, -0.5$	$-0.5, 0.5$
B	$-0.5, 0.5$	$1, -1$

Table 1: A two-action game with unequal winning rewards. Entries give Player 1's and Player 2's payoffs, respectively.

For $\mu_1 = (0.4, 0.6)$ and $\mu_2 = (0.35, 0.65)$, each action pair occupies a rectangle of width $\mu_2(p_2)$ and height $\mu_1(p_1)$ in the joint-action square (Figure 1(a)). The winning regions have total area $0.39 + 0.14 = 0.53$ in panel (b). Weighting their heights by the positive payoffs gives total area $W^+ = 0.39 + 0.07 = 0.46$ in panel (c). The audit may remove at most δ of the unweighted area. The corresponding payoff-weighted area measures the reward lost under honest play. The largest payoff available to a deviating player is a separate quantity, which motivates the second condition below.

The product $\mu_1(p_1)\mu_2(p_2)$ specifies the reference used in the example and the displayed formulas. A known correlated honest reference can be used by replacing this product everywhere in the honest benchmark and mass budget by its joint probabilities. The per-opponent payoff constraints are unchanged, and the direct threshold rule still applies to payoff control. Marginal strategies alone do not specify a correlated reference. Its joint law must also be supplied. The opponent-averaged incentive comparisons below use the displayed product reference. Conditional obedience under correlation is treated in Section 6.

3.2 Strategic testing guarantee

The following definition bounds the removed reference mass and the expected accepted winning payoff for every deviating strategy $\mu \in \Delta(\mathcal{P}_1)$ and every opponent action p_2.

Definition 1 ((ϵ, δ)-strategic tester)**.** Let $\epsilon \geq 0$ and $\delta \in [0, 1]$. An acceptance policy $\alpha : \mathcal{P}_1 \times \mathcal{P}_2 \to [0, 1]$ is an (ϵ, δ)-strategic tester for $(\mu_1, \mu_2, \mathcal{P}_1, \mathcal{P}_2, U)$ if:

(i) *Honest-player protection:*

$$\sum_{p_1,p_2} \mu_1(p_1)\mu_2(p_2)\big(1-\alpha(p_1,p_2)\big) \le \delta.$$

(ii) *Payoff control:*

$$\max_{\mu\in\Delta(\mathcal{P}_1)} \mathbb{E}_{\substack{p_1\sim\mu \\ a\sim\mathrm{Bernoulli}(\alpha(p_1,p_2))}} \big[a\cdot U^+(p_1,p_2)\big] \le W^+ + \epsilon, \qquad \forall p_2 \in \mathcal{P}_2.$$

where $a = 1$ means ACCEPT and $a = 0$ indicates FLAG.

An instance is (ϵ, δ)-*strategically testable* if such a policy exists.

Condition (i) averages over the honest reference. The auditor removes joint reference mass $\mu_1(p_1)\mu_2(p_2)(1-\alpha(p_1,p_2))$ from each action pair, with total removal at most δ. This total is the honest false-flag probability. The budget averages over honest opponent actions drawn from μ_2, so it need not hold conditionally in every column.

Condition (ii) averages over the player's action and the audit coin, without conditioning on acceptance. It enforces a common payoff ceiling $W^+ + \epsilon$ against every opponent action, including actions outside the support of μ_2.

The payoff is linear in the deviating strategy, so for every p_2,

$$\max_{\mu\in\Delta(\mathcal{P}_1)} \sum_{p_1} \mu(p_1)\alpha(p_1,p_2)U^+(p_1,p_2) = \max_{p_1} \alpha(p_1,p_2)U^+(p_1,p_2).$$

Condition (ii) is therefore equivalent to the pairwise constraints $\alpha(p_1,p_2)U^+(p_1,p_2) \le W^+ + \epsilon$ for all pairs. We call this an *external payoff ceiling*: the benchmark averages unaudited honest play over μ_2, but the bound holds separately for every opponent action. Thus ϵ measures excess over a common benchmark and can reflect favorable opponent actions as well as changes in the monitored player's strategy.

For example, suppose every row pays one against A and zero against B. With uniform μ_2 and full honest support, $W^+ = 1/2$ and $\delta = 0$ requires tolerance $1/2$, although changing the player's strategy cannot improve payoff. At $\delta = 1/4$, accepting against A with probability $1/2$ and against B surely achieves $\epsilon = 0$. The audit reduces the largest credited reward even though all player strategies remain equally profitable.

A column-specific benchmark $\sum_{a_1} \mu_1(a_1)U^+(a_1,p_2)$ would instead compare payoffs under the same opponent action. To study incentives, we must also compare honest and deviating play under the same audit (Section 5.1). We first solve the common-ceiling problem, then add these incentive requirements.

4 TESTING POLICIES

Having established the framework, we now turn to the computation of the acceptance policy. A natural baseline is uniform acceptance: $\alpha^{\mathrm{unif}}(p_1,p_2) = 1-\delta$ for every pair. It flags honest play with probability δ and has tolerance

$$\epsilon_{\mathrm{unif}}(\delta) = \left[(1-\delta)\max_{p_1,p_2} U^+(p_1,p_2) - W^+\right]_+.$$

This follows directly from the pairwise payoff constraints. The uniform policy treats every outcome alike, regardless of its payoff. An optimal policy instead chooses acceptance probabilities to minimize the largest accepted winning payoff subject to the same flag budget.

To impose a raw accepted-payoff cap $t \ge 0$, the largest permitted acceptance on a pair with $U^+(p_1,p_2) > 0$ is $\min\{1, t/U^+(p_1,p_2)\}$. Pairs with $U^+ = 0$ can be accepted surely. The minimum honest reference mass that must be flagged is therefore

$$D(t) := \sum_{p_1,p_2:\, U^+(p_1,p_2)>0} \mu_1(p_1)\mu_2(p_2)\left[1-\frac{t}{U^+(p_1,p_2)}\right]_+. \tag{2}$$

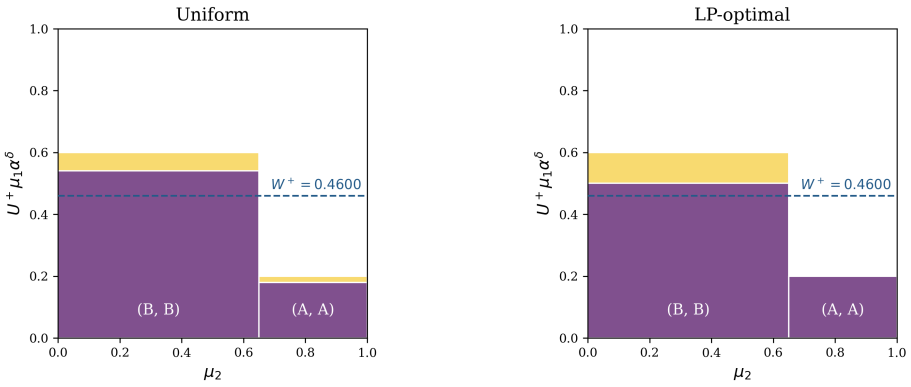

Figure 2: Testing policies for Table 1 at $\delta = 0.1$. Widths and original heights match the winning regions in Figure 1(c). Purple and yellow areas are retained and removed honest positive payoff. Here H_α denotes the retained honest payoff. Both policies flag total honest mass 0.1, but the optimal pair-dependent policy removes more honest payoff: 0.100 versus 0.046. It achieves a smaller external tolerance while retaining less honest payoff. The areas are payoff quantities, not flag probabilities. The payoff ceiling is determined by the largest αU^+. Losing pairs are accepted with probability 0.9 under the uniform rule and one under the optimal rule.

Proposition 2 (Optimal payoff ceiling)**.** *Let* $t^* := \min\{t \geq 0 : D(t) \leq \delta\}$. *The minimum tolerance is* $\epsilon^*(\delta) = [t^* - W^+]_+$. *Among policies attaining that tolerance, the pointwise-largest one accepts each pair with probability*

$$\alpha^*(p_1, p_2) = \begin{cases} \min\{1, (W^+ + \epsilon^*(\delta))/U^+(p_1, p_2)\}, & U^+(p_1, p_2) > 0, \\ 1, & U^+(p_1, p_2) = 0. \end{cases}$$

It retains the most honest reference mass among tolerance minimizers.

This threshold structure is a specialization of classical separable minimax resource allocation (Luss, 1987; Kuno et al., 1991). Algorithm 1 in Appendix A computes the threshold by sorting the positive payoffs and scanning the resulting intervals. Sorting takes $O(|\mathcal{P}_1||\mathcal{P}_2| \log(|\mathcal{P}_1||\mathcal{P}_2|))$ time and the scan is linear. The proof is given in the same appendix. This rule solves the problem when the honest flag budget is the only constraint coupling acceptance across pairs. We introduce additional strategic requirements below.

To see the rule geometrically, return to the winning regions in Figure 1. The audit reduces their accepted payoff while removing at most δ of unweighted reference mass. For Table 1 at $\delta = 0.1$, it accepts all pairs surely except (B, B), where $\alpha(B, B) = 29/39$. It removes mass 0.1 and gives tolerance $29/39 - 0.46 \approx 0.284$, compared with 0.44 when every pair is accepted with probability 0.9. Figure 2 shows both policies: lowering the largest accepted payoff can remove more honest reward, even at the same flag rate.

Once the policy is computed, the auditor accepts an observed pair with probability $\alpha(p_1, p_2)$ and flags otherwise. The algorithm controls payoff for every deviation. The next question is whether the resulting rewards also make honest play attractive.

5 The Strategic Environment

So far, we have bounded accepted winning payoff against every opponent action. We now compare the player's available returns with honest play under the same audit, and allow the auditor to account for other outcomes and deviation costs.

5.1 INCENTIVES UNDER A VISIBLE AUDIT

Suppose Player 1 maximizes accepted positive payoff, receives zero on a flag, and chooses before observing an opponent drawn from μ_2. For a committed policy α, honest audited payoff is

$$H_\alpha := \sum_{p_1,p_2} \mu_1(p_1)\mu_2(p_2)\alpha(p_1,p_2)U^+(p_1,p_2).$$

Write $v_\alpha(p_1) := \sum_{p_2} \mu_2(p_2)\alpha(p_1,p_2)U^+(p_1,p_2)$ for the return from action p_1. To allow a gain of at most $\gamma \geq 0$ from deviation, require

$$v_\alpha(p_1) - H_\alpha \leq \gamma \qquad \forall p_1. \tag{3}$$

Pure-action constraints suffice for every mixed deviation. At $\gamma = 0$, honest play is a best response: every action in its support attains the maximal audited return. Other strategies may be equally good best responses. Unlike the external ceiling, the incentive condition compares strategies against the fixed opponent distribution. Because its honest benchmark depends on the audit, it couples acceptance across actions. This model applies when the opponent is committed to μ_2, or represents a fixed randomized environment, and Player 1 chooses without observing its realized action. If the opponent adapts, this incentive guarantee need not persist, although the external ceiling does. Observing the opponent before acting requires conditional incentive constraints, given in Appendix B.

The gain constraints will be part of the joint LP in Section 6. Its two-action example shows that a policy with the smallest external payoff ceiling can make an honest reference strategy strictly suboptimal after auditing.

5.2 EXPOSURE AND DEVIATION COSTS

The auditor's score need not equal the player's utility. We can include draws and losses in the score, or deduct known action costs, while keeping the same reference-mass budget.

Testing zero-payoff outcomes. The optimal threshold rule accepts outcomes with $U = 0$ surely, since they contribute nothing to U^+. An auditor may nevertheless wish to review such outcomes, for example draws. We can split $\delta = \delta_w + \delta_t$: apply the optimal winning policy with budget δ_w, and flag zero-payoff outcomes uniformly using at most δ_t of honest reference mass. Losing outcomes remain accepted. Algorithm 2 in Appendix A.2 gives the procedure and its guarantee. It retains the winning-payoff tolerance $\epsilon^*(\delta_w)$ while keeping the total honest flag probability at most δ. When both an accepted zero-payoff outcome and a flag give the player zero utility, flagging these outcomes does not itself change their incentives.

Utility exposure. To control accepted draws as well as wins, the auditor can instead assign outcomes an exposure score $\phi : [-1, 1] \to [0, 1]$. For example, $\phi_\kappa(u) = \max\{u, 0\} + \kappa \mathbf{1}\{u = 0\}$ with $\kappa \in (0, 1]$ assigns score κ to a draw and keeps the original winning rewards. The choice $\phi(u) = \mathbf{1}\{u \geq 0\}$ gives wins and draws equal weight. Choosing $\phi(u) = \max\{u, 0\}$ recovers the original objective.

For a fixed ϕ, let $W_\phi := \mathbb{E}_{\mu_1,\mu_2}[\phi(U)]$. An (ϵ, δ)-strategic tester with respect to ϕ keeps the same honest flag budget and replaces the payoff constraints by

$$\alpha(p_1, p_2)\phi(U(p_1, p_2)) \leq W_\phi + \epsilon \qquad \forall p_1, p_2.$$

The threshold rule and LP apply to this score table as well (Appendix A.2). This allows the auditor to allocate one budget across all outcomes with positive exposure. The score specifies the auditor's objective. Incentive constraints still use the player's actual utility.

Deviation costs. Assign each action p_1 a known cost $c(p_1) \geq 0$, incurred regardless of the audit decision, and define the strategy cost by

$$C(\mu) := \sum_{p_1} \mu(p_1)c(p_1).$$

The costs may be calibrated relative to the fixed honest reference μ_1, but do not depend on the deviating strategy μ. Thus C is linear in the action probabilities. The cost difference from honest play is $C(\mu) - C(\mu_1)$, which may be negative.

When costs and exposure are expressed on a common scale, the unaudited honest benchmark is $W_{\phi,c} := W_\phi - C(\mu_1)$. Against each opponent action, the deviating cost-adjusted exposure is $\sum_{p_1} \mu(p_1)\alpha(p_1,p_2)\phi(U(p_1,p_2)) - C(\mu)$. Because this expression is linear in μ, bounding it by $W_{\phi,c} + \epsilon$ for every deviation is equivalent to

$$\alpha(p_1,p_2)\phi(U(p_1,p_2)) - c(p_1) \le W_{\phi,c} + \epsilon \qquad \forall p_1, p_2.$$

A nonlinear strategy cost need not admit a pure best response and does not directly yield this finite reduction. Cost-adjusted exposure describes net utility only when the rewards, costs, and consequences of flagging match the player's utility.

When exposure minus cost is the player's utility, the analogue of equation 3 compares each action's opponent-averaged accepted exposure minus its cost with the same score under honest play. When it is only an auditor-selected score, incentives must instead use the player's actual utility. In either case, fixed acceptance and flag utilities give affine constraints on the audit.

Honest payoff and equilibrium protection. The flag budget implies $0 \le W^+ - H_\alpha \le \delta$. Thus the initial tester only bounds same-audit gain by $\epsilon + (W^+ - H_\alpha)$, which may be positive even at $\epsilon = 0$. To protect honest rewards, one may also require $H_\alpha \ge (1-\delta)W^+$, retaining at least the honest payoff of uniform acceptance. To preserve a prescribed Nash equilibrium, impose unilateral-deviation inequalities for both players using their actual audited utilities. With fixed reference strategies these are affine requirements. Choosing an unknown equilibrium jointly with the audit is a different problem. Appendix A gives the two-player conditions, and Appendix B proves the gain bound.

6 Linear Programming and Strategic Constraints

The threshold rule solves the problem with only an honest flag budget. A linear programming formulation lets us add constraints on incentives and rewards. We first state the primal and dual for payoff control, then extend the program to these joint requirements.

6.1 Payoff-control primal and dual

Theorem 3 (Primal–dual LP characterization)**.** *The optimal payoff-control tolerance $\epsilon^*(\delta)$ is the value of the primal LP*

$$\begin{aligned} \min_{\epsilon,\alpha} \quad & \epsilon \\ \text{s.t.} \quad & U^+(p_1,p_2)\alpha(p_1,p_2) \le W^+ + \epsilon && \forall p_1,p_2, \\ & \sum_{p_1,p_2} \mu_1(p_1)\mu_2(p_2)\alpha(p_1,p_2) \ge 1-\delta, \\ & 0 \le \alpha(p_1,p_2) \le 1 && \forall p_1,p_2, \\ & \epsilon \ge 0. \end{aligned} \tag{P}$$

Its dual is

$$\begin{aligned} \max_{\lambda,\eta,\tau} \quad & \tau(1-\delta) - W^+ \sum_{p_1,p_2} \lambda(p_1,p_2) - \sum_{p_1,p_2} \eta(p_1,p_2) \\ \text{s.t.} \quad & U^+(p_1,p_2)\lambda(p_1,p_2) + \eta(p_1,p_2) \ge \tau\mu_1(p_1)\mu_2(p_2) \quad \forall p_1,p_2, \\ & \sum_{p_1,p_2} \lambda(p_1,p_2) \le 1, \\ & \lambda,\eta,\tau \ge 0. \end{aligned} \tag{D}$$

Both programs attain the same optimal value.

In equation P, each action pair contributes one bound on accepted payoff. In equation D, $\lambda(p_1,p_2)$ weights pairs that support the worst accepted payoff, τ prices honest retention, and $\eta(p_1,p_2)$ prices the bound $\alpha(p_1,p_2) \le 1$. If the optimal tolerance is positive, complementary slackness gives $\sum_{p_1,p_2} \lambda(p_1,p_2) = 1$, so positive dual weights form a mixture of binding evaluator responses. The proof and the conditions at zero tolerance are in Appendix A.

6.2 COUPLED STRATEGIC REQUIREMENTS

The direct algorithm minimizes the external payoff ceiling subject only to the honest flag budget. Once players respond to a visible audit, a principal may also require that honest play remain attractive and that honest rewards are not reduced too far. These requirements compare different entries of the acceptance table, so pairwise thresholding need not solve the joint problem.

Fix an allowable same-audit gain $\gamma \geq 0$ and a floor on honest payoff $h \in [0, W^+]$. With H_α as defined above, the joint policy is obtained from the following LP:

$$\begin{aligned} \min_{\epsilon,\alpha} \quad & \epsilon \\ \text{s.t.} \quad & \alpha(p_1,p_2)U^+(p_1,p_2) \leq W^+ + \epsilon && \forall p_1,p_2, \\ & \sum_{p_1,p_2} \mu_1(p_1)\mu_2(p_2)\alpha(p_1,p_2) \geq 1-\delta, \\ & v_\alpha(p_1) - H_\alpha \leq \gamma && \forall p_1, \\ & H_\alpha \geq h, \\ & 0 \leq \alpha(p_1,p_2) \leq 1 && \forall p_1,p_2, \\ & \epsilon \geq 0. \end{aligned} \tag{4}$$

The first two lines reproduce the payoff-control problem. The third limits the gain from any mixed deviation because its return is linear in the player's action distribution. At $\gamma = 0$, μ_1 is a best response to μ_2 under the audit. The fourth protects honest accepted payoff. A natural benchmark is $h = (1-\delta)W^+$, the payoff retained by uniform acceptance. The floor can be omitted when honest payoff retention is not a design requirement. Appendix C varies both γ and h on games where honest play is already a best response before auditing. For an opponent-relative payout standard, the first LP constraint can instead use $W^+(p_2) := \sum_{p_1} \mu_1(p_1)U^+(p_1,p_2)$. Both ceilings can also be imposed together.

Example 1 (The objectives select different audits)**.** Both players have actions A, B and uniform references. Let

$$U^+ = \begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \end{pmatrix}, \qquad W^+ = 1/2, \qquad \delta = 1/8.$$

Without an audit, both rows return $1/2$, so uniform honest play is a best response. Payoff control alone accepts (A, A) with probability $1/2$ and every other pair surely. This reaches $\epsilon = 0$ but makes B strictly more attractive than A ($1/2$ versus $1/4$). With $\gamma = 0$, the two audited row returns must agree: $2\alpha(A,A) = \alpha(B,A) + \alpha(B,B)$. The optimum then has $\epsilon = 1/3$, attained by $\alpha = \left(\begin{smallmatrix} 5/6 & 1 \\ 5/6 & 5/6 \end{smallmatrix}\right)$. Adding the uniform-retention floor $h = 7/16$ raises the optimum to $\epsilon = 3/8$, which uniform acceptance $\alpha \equiv 7/8$ attains. Appendix A proves these values. Thus the LP describes three distinct design goals in the same game, using the same flag budget.

Fixed exposure scores and action costs preserve linearity when the appropriate payoff and incentive inequalities from Section 5.2 are used. Preserving a prescribed Nash equilibrium of both players adds their unilateral-deviation inequalities in their actual audited utilities (Appendix A). More generally, any finite collection of affine audit requirements can be added to equation 4. Such requirements can make the problem infeasible at a given δ.

Proposition 4 (Optimal testing with strategic constraints)**.** *For fixed references and a finite game, payoff control subject to an honest flag budget and finitely many affine constraints on the audit is an LP. Pure-action inequalities suffice for bounds on every mixed deviation when the deviator's payoff is linear in its strategy. If the policy set is nonempty, the optimum is attained and equals the dual optimum.*

The direct algorithm applies when the extra constraints are absent. The added LP constraints have their own dual multipliers. The baseline dual above certifies the payoff-only case. Two game settings illustrate how additional constraints arise.

Correlated recommendations. Suppose honest play is drawn from a known joint law $\pi \in \Delta(\mathcal{P}_1 \times \mathcal{P}_2)$, and each player observes only their recommended action. Replace $\mu_1(p_1)\mu_2(p_2)$ by $\pi(p_1,p_2)$

in the benchmark and flag budget. Correlation alone leaves Proposition 2 intact. If the audit must also preserve obedience to recommendations, write $V_i^\alpha(p_1, p_2)$ for Player i's audited utility at that pair. For fixed acceptance payoff U_i and flag payoff F_i, it is $V_i^\alpha = \alpha U_i + (1-\alpha)F_i$ at each pair. For every recommendation p_1 and alternative p_1', require

$$\sum_{p_2} \pi(p_1, p_2)\big[V_1^\alpha(p_1', p_2) - V_1^\alpha(p_1, p_2)\big] \le 0. \tag{5}$$

Player 2 has analogous inequalities. Because V_i^α is affine in α, these conditions join the payoff-control LP. They couple different entries of the audit table, so independent thresholding need not preserve obedience.

Prescribed Nash behavior. If the honest reference is a product distribution that should remain a Nash equilibrium after auditing, both players' unilateral-deviation constraints must hold. For Player 1, for example,

$$\sum_{p_2} \mu_2(p_2) V_1^\alpha(p_1, p_2) \le \sum_{a_1, a_2} \mu_1(a_1)\mu_2(a_2) V_1^\alpha(a_1, a_2) \qquad \forall p_1.$$

The analogous Player 2 constraints use μ_1. These are again affine for fixed utilities on acceptance and flagging. They address both players' actual incentives, while the external payoff ceiling remains an auditor-selected score. Appendix A gives the full two-player conditions.

7 Stackelberg Characterization

Public commitment has two roles here. In equation 4, a player responds to the announced audit against μ_2, and the constraints limit the gain from that response. For the external ceiling, a separate worst-case evaluator yields an auxiliary Stackelberg representation. Write $\epsilon^*(\delta)$ for the minimum external tolerance under the honest-mass budget alone (Appendix A). The auditor commits to a feasible acceptance policy before this evaluator selects a deviating strategy μ and opponent action p_2. Define the auditor's loss and the evaluator's payoff as

$$\left[\sum_{p_1} \mu(p_1)\alpha(p_1, p_2)U^+(p_1, p_2) - W^+\right]_+.$$

This is an auxiliary zero-sum Stackelberg game (Conitzer & Sandholm, 2006; Tambe, 2011). The evaluator combines the player's deviation with a worst-case opponent condition. This choice belongs to the evaluator, while the monitored player controls only its own action.

Theorem 5 (Stackelberg characterization)**.** *For any nonempty feasible audit set described above, the auxiliary game has an equilibrium whose auditor loss equals the minimum external tolerance over that set. An optimal audit and a pair maximizing accepted positive payoff form such an equilibrium. Without additional strategic constraints, the value is $\epsilon^*(\delta)$. A pure maximizing pair always exists.*

The theorem characterizes the auxiliary auditing game. Honest play is a best response in the original audited game only when the corresponding incentive constraints hold. All evaluator best responses give the same auditor loss, so favorable tie-breaking is unnecessary. The result also applies with a fixed correlated reference and its obedience constraints. Appendix B gives the proof.

8 Conclusion

We introduced (ϵ, δ)-strategic testing, a framework for auditing a player in normal-form games by controlling accepted winning payoff relative to an honest benchmark. The optimal policy admits a direct threshold rule, while a linear program incorporates constraints on incentives and honest rewards. The auxiliary Stackelberg game gives a commitment interpretation of the auditor's optimization. Our examples and experiments show that controlling accepted payoff and preserving honest incentives can require different policies. Natural extensions include repeated games and auditing when the payoff table and both reference strategies must be estimated from data.

Reproducibility statement

The main text states the auditing model, assumptions, and optimization problems. Proofs and the sorting algorithm are in Appendices A and B. Appendix C describes the synthetic game generators, policy comparisons, reference-estimation stress test, random seeds, and reported statistics. The scripts and instructions for reproducing the experiments are available at https://anonymous.4open.science/status/experiments-58D0.

AI use statement

The main framework, the definition of strategic testing, the linear programming formulation, and the connections to game-theoretic concepts originated with the authors. OpenAI Codex assisted in refining the mathematical formulations and in developing proofs of mathematical claims.

We also used OpenAI Codex to draft and polish sections of the paper, find related work, explore refinements of the authors' ideas, and assist with research execution. This included writing code for the experiments and generating synthetic game datasets. The authors retain responsibility for the paper's claims, proofs, citations, and results.

References

Rudolf Avenhaus, Bernhard von Stengel, and Shmuel Zamir. Inspection games. In *Handbook of Game Theory with Economic Applications*, volume 3, pp. 1947–1987. Elsevier, 2002.

David J Barnes and Julio Hernandez-Castro. On the limits of engine analysis for cheating detection in chess. *Computers & Security*, 48:58–73, 2015.

Stephen Bates, Michael I Jordan, Michael Sklar, and Jake A Soloff. Principal-agent hypothesis testing. *Journal of the American Statistical Association*, 2026. doi: 10.1080/01621459.2026.2724031.

C. L. Canonne. A survey on distribution testing: Your data is big. but is it blue? *Theory of Computing, Graduate Surveys*, 9:1–100, 2020.

Yatong Chen, Safwan Hossain, and Yiling Chen. Strategic hypothesis testing. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-4894.

V. Conitzer and T. Sandholm. Computing the optimal strategy to commit to. In *Proceedings of the 7th ACM Conference on Electronic Commerce (EC)*, pp. 82–90, 2006.

Sanmay Das, Fang-Yi Yu, and Yuang Zhang. Optimally auditing adversarial agents. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 16787–16794, 2026. doi: 10.1609/aaai.v40i20.38722.

Ilias Diakonikolas and Daniel M Kane. *Algorithmic high-dimensional robust statistics*. Cambridge university press, 2023.

Andrew Estornell, Sanmay Das, and Yevgeniy Vorobeychik. Incentivizing truthfulness through audits in strategic classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 5347–5354, 2021. doi: 10.1609/aaai.v35i6.16674.

Andrew Estornell, Yatong Chen, Sanmay Das, Yang Liu, and Yevgeniy Vorobeychik. Incentivizing recourse through auditing in strategic classification. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pp. 400–408, 2023. doi: 10.24963/ijcai.2023/45.

Etienne Gauthier, Francis Bach, and Michael I. Jordan. Anytime detection of strategic deviations in multi-agent systems, 2026. URL `https://arxiv.org/abs/2601.05427`.

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

Takahito Kuno, Hiroshi Konno, and Eitan Zemel. A linear-time algorithm for solving continuous maximin knapsack problems. *Operations Research Letters*, 10(1):23–26, 1991. doi: 10.1016/0167-6377(91)90082-Z.

Thijs Laarhoven and Aditya Ponukumati. Towards transparent cheat detection in online chess: An application of human and computer decision-making preferences. In *International Conference on Computers and Games*, pp. 163–180. Springer, 2022.

Hanan Luss. An algorithm for separable nonlinear minimax problems. *Operations Research Letters*, 6(4):159–162, 1987.

R. McIlroy-Young, R. Wang, S. Sen, J. Kleinberg, and A. Anderson. Aligning superhuman AI with human behavior: Chess as a model system. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 1677–1687, 2020.

J. Neyman and E. S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London, Series A*, 231:289–337, 1933.

L. Paninski. A coincidence-based test for uniformity given very sparsely sampled discrete data. *IEEE Transactions on Information Theory*, 54(10):4750–4755, 2008.

José Pedro Pinto, André Pimenta, and Paulo Novais. Deep learning and multivariate time series for cheat detection in video games. *Machine Learning*, 110(11):3037–3057, 2021.

K. W. Regan and G. M. Haworth. Intrinsic chess ratings. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence (AAAI)*, pp. 834–839, 2011.

M. Tambe. *Security and Game Theory: Algorithms, Deployed Systems, Lessons Learned*. Cambridge University Press, 2011.

Zhenwei Tang, Difan Jiao, Reid McIlroy-Young, Jon Kleinberg, Siddhartha Sen, and Ashton Anderson. Maia-2: A unified model for human-ai alignment in chess. *Advances in Neural Information Processing Systems*, 37:20919–20944, 2024.

Zhenwei Tang, Difan Jiao, Eric Xue, Reid McIlroy-Young, Jon Kleinberg, Siddhartha Sen, and Ashton Anderson. Learning to imitate with less: Efficient individual behavior modeling in chess. *arXiv preprint arXiv:2507.21488*, 2025.

G. Valiant and P. Valiant. An automatic inequality prover and instance optimal identity testing. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 51–62, 2017.

A Policy Optimization and the Coupled Example

A.1 Optimal winning-payoff policy

Algorithm 1: Optimal pair-dependent testing

Input: Payoff table U, reference strategies μ_1, μ_2, budget δ
Output: Optimal tolerance ϵ and acceptance rule α
Compute W^+ from equation 1;
Let $\mathcal{W} = \{(p_1, p_2) : \mu_1(p_1)\mu_2(p_2) > 0,\ U^+(p_1, p_2) > 0\}$;
if $\sum_{(p_1,p_2)\in\mathcal{W}} \mu_1(p_1)\mu_2(p_2) \le \delta$ **then**
  Set $t^* = 0$;
else
  Sort the distinct values $U^+(p_1, p_2)$ in $\mathcal{W}$ as $u_1 > \cdots > u_K > 0$; set $u_{K+1} = 0$;
  for $k = 1, \ldots, K$ **do**
    Set $A_k = A_{k-1} \cup \{(p_1, p_2) \in \mathcal{W} : U^+(p_1, p_2) = u_k\}$ (with $A_0 = \varnothing$);
    Update the sums and set $t_k = \dfrac{\sum_{(p_1,p_2)\in A_k} \mu_1(p_1)\mu_2(p_2) - \delta}{\sum_{(p_1,p_2)\in A_k} \mu_1(p_1)\mu_2(p_2)/U^+(p_1, p_2)}$;
    if $u_{k+1} \le t_k \le u_k$ **then**
      Set $t^* = t_k$; **break**;
    end
  end
end
Set $\epsilon = [t^* - W^+]_+$ and $T = W^+ + \epsilon$;
For every (p_1, p_2), set $\alpha(p_1, p_2) = \min\{1, T/U^+(p_1, p_2)\}$ if $U^+(p_1, p_2) > 0$, and $\alpha(p_1, p_2) = 1$ otherwise;
return (ϵ, α);

The two sums in t_k are updated as each payoff group enters A_k, giving a linear scan after sorting.

Proof of Proposition 2 and correctness of Algorithm 1. Fix a cap $t \ge 0$. On a pair with positive payoff, the largest permitted acceptance probability is $\min\{1, t/U^+(p_1, p_2)\}$. When $U^+(p_1, p_2) = 0$, acceptance can equal one. Let $\bar{\alpha}_t$ take these largest permitted values. Every policy meeting the cap flags at least as much honest play as $\bar{\alpha}_t$. By equation 2, the flag probability of $\bar{\alpha}_t$ is exactly $D(t)$. Thus the cap is feasible if and only if $D(t) \le \delta$.

The function D is continuous and nonincreasing. If $\delta \ge D(0) = \sum_{(p_1,p_2)\in\mathcal{W}} \mu_1(p_1)\mu_2(p_2)$, then $t^* = 0$. Otherwise $D(t) = \delta$ has a root in $(0, u_1]$. For $t \in [u_{k+1}, u_k]$, only the pairs in A_k contribute to $D(t)$, so

$$D(t) = \sum_{(p_1,p_2)\in A_k} \mu_1(p_1)\mu_2(p_2) - t \sum_{(p_1,p_2)\in A_k} \frac{\mu_1(p_1)\mu_2(p_2)}{U^+(p_1, p_2)}.$$

Solving $D(t) = \delta$ on that interval gives the candidate t_k. At least one interval contains the root, including shared endpoints, so the scan returns t^*.

For any policy, the largest accepted positive payoff is $\max_{p_1,p_2} \alpha(p_1, p_2)U^+(p_1, p_2)$. Its minimum feasible value is t^*, hence the minimum tolerance is $\epsilon^*(\delta) = [t^* - W^+]_+$. Set $T = W^+ + \epsilon^*(\delta) = \max\{W^+, t^*\}$. Since $T \ge t^*$, $\bar{\alpha}_T$ is feasible. Every policy with tolerance $\epsilon^*(\delta)$ has acceptance at most $\bar{\alpha}_T$ on every pair. Hence $\bar{\alpha}_T$ retains the most honest play. Pairs with zero reference probability do not enter the scan, but the returned policy still respects their payoff bounds. Pairs with $U^+ = 0$ are accepted surely. Sorting costs $O(|\mathcal{P}_1||\mathcal{P}_2| \log(|\mathcal{P}_1||\mathcal{P}_2|))$. Updating the two sums and constructing the policy take linear time. □

A.2 Testing zero-payoff outcomes and exposure scores

Let $q_0 := \sum_{p_1,p_2:\, U(p_1,p_2)=0} \mu_1(p_1)\mu_2(p_2)$ be the honest probability of a zero-payoff outcome. To spend at most δ_t on these outcomes, use the uniform flag probability $r_t = \min\{1, \delta_t/q_0\}$ when $q_0 > 0$. If $q_0 = 0$, we set $r_t = 0$. The honest flag budget does not constrain decisions on these

zero-reference outcomes. Budgets here measure joint reference mass, so r_t is a conditional flag probability rather than the budget itself.

Algorithm 2: Strategic testing with separate winning and zero-payoff budgets

Input: Payoff table U, references μ_1, μ_2, budgets $\delta_w, \delta_t \geq 0$ with $\delta_w + \delta_t = \delta \leq 1$, observed pair (p_1, p_2)
Output: Decision in $\{\text{ACCEPT}, \text{FLAG}\}$
// Compute the policy before play
Compute (ϵ_w, α_w) with Algorithm 1 using budget δ_w;
Compute $q_0 = \sum_{a_1,a_2:\, U(a_1,a_2)=0} \mu_1(a_1)\mu_2(a_2)$;
Set $r_t = 0$ if $q_0 = 0$, and $r_t = \min\{1, \delta_t/q_0\}$ otherwise;
// Apply the policy to the observed pair
if $U(p_1, p_2) < 0$ **then**
  return ACCEPT;
else if $U(p_1, p_2) = 0$ **then**
  Set $a = 1 - r_t$;
else
  Set $a = \alpha_w(p_1, p_2)$;
end
Draw $Z \sim \text{Bernoulli}(a)$;
if $Z = 1$ **then**
  return ACCEPT;
else
  return FLAG;
end

Guarantee. Algorithm 1 accepts every nonwinning pair, so its honest flag probability comes entirely from wins and is at most δ_w. The additional zero-payoff flags have honest probability $q_0 r_t \leq \delta_t$. Thus the total is at most δ. On winning pairs, acceptance is unchanged from α_w. On all other pairs, $U^+ = 0$. The expected accepted winning payoff therefore remains at most $W^+ + \epsilon^*(\delta_w)$ for every deviation and opponent action. The procedure need not minimize winning-payoff tolerance for the total budget δ, since it reserves part of that budget for zero-payoff outcomes. It also imposes no additional incentive guarantee.

Optimizing an exposure score. For a fixed ϕ, form the nonnegative table $S(p_1, p_2) = \phi(U(p_1, p_2))$ and apply Algorithm 1 to S with the full budget δ. Its benchmark is then W_ϕ, and its output satisfies $\alpha(p_1, p_2)S(p_1, p_2) \leq W_\phi + \epsilon$ for every pair. Because $S \in [0, 1]$, the threshold proof applies unchanged and the resulting exposure tolerance is optimal under the honest flag budget. For ϕ_κ, zero-payoff outcomes of the original game have positive exposure and may therefore be flagged by this optimized rule. If incentive or honest-reward constraints are also imposed, use the LP with the exposure ceiling and the relevant utility constraints.

A.3 Linear programming and strategic constraints

Proof of Theorem 3. For fixed p_2, accepted payoff is linear in the deviating strategy, so a pure action attains the maximum over its simplex. Thus the universal payoff condition is equivalent to $U^+(p_1, p_2)\alpha(p_1, p_2) \leq W^+ + \epsilon$ for every pair. Introducing $\epsilon \geq 0$ gives the primal equation P.

Assign $\lambda(p_1, p_2) \geq 0$ to each payoff inequality, $\eta(p_1, p_2) \geq 0$ to $\alpha(p_1, p_2) \leq 1$, and $\tau \geq 0$ to the flag-budget inequality written as $1 - \delta - \sum_{p_1,p_2} \mu_1(p_1)\mu_2(p_2)\alpha(p_1, p_2) \leq 0$. The Lagrangian,

minimized over $\alpha, \epsilon \geq 0$, is

$$\begin{aligned} L = \epsilon\Big(1 - \sum_{p_1,p_2} \lambda(p_1,p_2)\Big) \\ + \sum_{p_1,p_2} \alpha(p_1,p_2)\big[U^+(p_1,p_2)\lambda(p_1,p_2) + \eta(p_1,p_2) - \tau\mu_1(p_1)\mu_2(p_2)\big] \\ + \tau(1-\delta) - W^+ \sum_{p_1,p_2} \lambda(p_1,p_2) - \sum_{p_1,p_2} \eta(p_1,p_2). \end{aligned}$$

The infimum is finite exactly when the coefficients of ϵ and every $\alpha(p_1, p_2)$ are nonnegative. These are the constraints in equation D, and the constant term is its objective. The primal is feasible at $\alpha \equiv 1$ and sufficiently large ϵ, and it is bounded below by zero. Finite-dimensional LP strong duality gives equality and attainment. Complementary slackness identifies positive $\lambda(p_1, p_2)$ with binding payoff constraints. If the optimum has $\epsilon^*(\delta) > 0$, it also forces $\sum_{p_1,p_2} \lambda(p_1, p_2) = 1$. □

Proof of Proposition 4. The additional affine conditions define a closed subset of the cube of acceptance probabilities. If nonempty, this set is compact, so the continuous objective attains its minimum. The payoff remains linear in the deviating strategy, so pure-action constraints still suffice. Introducing ϵ gives a feasible LP bounded below. Strong duality applies with the multipliers for the additional inequalities and equalities. □

Best responses and a prescribed Nash equilibrium. For equation 4, multiplying each incentive inequality by the probability of its action under any deviation and summing proves the gain bound for that deviation. Conversely, pure deviations are special cases. At $\gamma = 0$, the honest average equals the maximal action return exactly when every action in its support attains that maximum. Thus zero gain establishes a best response, not uniqueness or strict honesty.

To state the two-player extension without identifying the auditor's score with utility, write $V_i^\alpha(p_1, p_2)$ locally for Player i's expected utility after the audit. Fixed utilities on acceptance and on a flag make V_i^α affine in $\alpha(p_1, p_2)$. The prescribed product strategy (μ_1, μ_2) is a Nash equilibrium of this audited game if and only if

$$\begin{aligned} \sum_{p_2} \mu_2(p_2) V_1^\alpha(p_1,p_2) \leq \sum_{a_1,a_2} \mu_1(a_1)\mu_2(a_2) V_1^\alpha(a_1,a_2) \quad \forall p_1, \\ \sum_{p_1} \mu_1(p_1) V_2^\alpha(p_1,p_2) \leq \sum_{a_1,a_2} \mu_1(a_1)\mu_2(a_2) V_2^\alpha(a_1,a_2) \quad \forall p_2. \end{aligned}$$

Linearity in each mixed strategy proves sufficiency for all deviations. These affine constraints preserve the LP, as do honest-utility floors. Preservation of an original equilibrium additionally requires that the prescribed pair be an equilibrium before auditing. Feasibility and uniqueness are not guaranteed. Jointly optimizing the reference strategies would introduce products of decision variables.

Solution of Example 1. The incentive equality and mass budget imply $3\alpha(A, A) + \alpha(A, B) \geq 7/2$. Since $\alpha(A, B) \leq 1$, we obtain $\alpha(A, A) \geq 5/6$, hence $\epsilon \geq 5/6 - 1/2 = 1/3$. The policy

$$\alpha = \begin{pmatrix} 5/6 & 1 \\ 5/6 & 5/6 \end{pmatrix}$$

attains the bound: its flag probability is $1/8$, both row returns are $5/12$, and its largest accepted pair payoff is $5/6$. Under the additional floor $H_\alpha \geq 7/16$, the incentive equality implies $H_\alpha = \alpha(A, A)/2$, so $\alpha(A, A) \geq 7/8$ and $\epsilon \geq 3/8$. Uniform acceptance $\alpha \equiv 7/8$ satisfies the flag, incentive, and floor constraints with equality and attains this bound.

B Incentive Comparisons and Commitment

Conditional benchmarks and observed opponent actions. A column-specific honest benchmark compares outcomes under the same opponent action, but still credits honest play before auditing. If

Player 1 observes p_2 before acting, bounding its gain over honest play under the same audit instead requires

$$\alpha(p_1, p_2)U^+(p_1, p_2) - \sum_{a_1} \mu_1(a_1)\alpha(a_1, p_2)U^+(a_1, p_2) \leq \gamma \qquad \forall p_1, p_2.$$

These linear inequalities control every opponent-contingent deviation, for any opponent distribution. They do not assert that the honest reference itself conditions on p_2. Such a model would require a specified conditional honest policy. The main opponent-averaged constraints need not imply these stronger inequalities.

Gain and honest retention. Honest payoff removed by an audit is $\sum_{p_1,p_2} \mu_1(p_1)\mu_2(p_2)(1 - \alpha(p_1, p_2))U^+(p_1, p_2)$. Since $0 \leq U^+ \leq 1$, this is at most the flag budget δ. A baseline tester bounds the best response against μ_2 by $W^+ + \epsilon$. Subtracting H_α gives same-audit gain at most $\epsilon + (W^+ - H_\alpha) \leq \epsilon + \delta$. This does not imply honest play is a best response, and the comparison requires accepted positive payoff to be the player's utility.

Proof of Theorem 5. For a fixed policy, the accepted payoff is linear in the deviating strategy. A pure maximizing pair therefore attains the evaluator's maximum, and monotonicity of the positive-part function gives worst loss $[\max_{p_1,p_2} \alpha(p_1, p_2)U^+(p_1, p_2) - W^+]_+$. Minimizing this continuous function over the nonempty compact feasible policy set gives the minimum external tolerance and an optimal commitment. Under the flag budget alone, this optimum is $\epsilon^*(\delta)$. Any maximizing pair is a best response to it, and all such responses give the same auditor loss. If the maximum does not exceed W^+, all responses have zero loss. The same argument applies after adding feasible affine constraints to the auditor's policy set, with the augmented LP value. It is not an equilibrium claim about the original two-player game. □

C SYNTHETIC EXPERIMENTS

Code and run parameters. The `experiments/` directory contains the data generators, optimization checks, figure and table renderers, and a README with the execution order and dependencies. The code is available at https://anonymous.4open.science/status/experiments-58D0. The game masks use seed 1000, reward magnitudes use 1001, reference-estimation samples use 1002, validation scripts use 1003–1005, and the joint LP sweep uses 2000–2001. No external data are used.

We use synthetic games to ask two questions: how much does payoff-aware auditing improve on uniform acceptance, and what changes when the audit must also protect the honest player's incentives and rewards? The first comparisons assume that the payoffs and reference strategies are known and both actions are observed. We then estimate the honest reference from samples to measure the resulting flag rates and accepted payoffs. The results describe performance in these controlled game simulations.

For each policy α, we report the honest payoff H_α, the best-response return $R_\alpha := \max_{p_1} v_\alpha(p_1)$, and the gain from deviation $G_\alpha := R_\alpha - H_\alpha$. We also report the external tolerance $\epsilon_\alpha := [\max_{p_1,p_2} \alpha(p_1,p_2)U^+(p_1,p_2) - W^+]_+$. Players choose before seeing the opponent's action. They receive U^+ on an accepted win and zero on a loss or a flag.

Game generation. We vary the number of monitored-player actions $n \in \{32, 128\}$ and the expected win rate $d \in \{0.10, 0.35, 0.70\}$, keeping 32 opponent actions. For each size and win rate, we generate 100 independent win/loss masks. For each row, we independently draw a win probability $\theta_i \sim \mathrm{Beta}(4d, 4(1-d))$. Conditional on this probability, its entries are independent draws $Z_{ij} \sim \mathrm{Bernoulli}(\theta_i)$. This makes some actions more successful than others within a game.

Each mask produces two payoff tables. In the binary version, $U_{ij} = 2Z_{ij} - 1$. In the varying-reward version, $U_{ij} = Z_{ij}H_{ij} - (1 - Z_{ij})$, where each winning reward H_{ij} is drawn independently from $\mathrm{Uniform}[0.1, 1]$. The two versions share their win/loss mask, giving 1,200 tables from 600 masks. The opponent reference μ_2 is uniform.

To vary honest behavior, we rank the rows by win frequency, break ties by row index, and assign evenly spaced scores $s_i \in [-1, 1]$. We set $\mu_1(i) = \exp(\beta s_i)/\sum_\ell \exp(\beta s_\ell)$ with $\beta \in \{-4, 0, 4\}$. High-win-rate actions are respectively rare, equally likely, or common under these references. All three references have full support. The ranks are held fixed across the two reward versions, and the $\beta = \pm 4$ references have the same action probabilities, assigned in reverse rank order.

Payoff and response comparison. At each budget $\delta \in \{0, 0.01, 0.05, 0.10, 0.25, 0.50, 0.75, 1\}$, we compare uniform acceptance with the policy that minimizes the largest accepted pair payoff, which we call the *raw cap*. At that cap, we choose the largest permitted acceptance probability on every pair. This tie-break can affect honest payoff and deviation gain when the optimal external tolerance is zero. Algorithm 1 uses a different tie-break: it maximizes honest retention among all tolerance minimizers. The two policies achieve the same optimal external tolerance.

We also include a wins-only uniform rule. It accepts every nonwinning pair and applies one acceptance probability to all winning pairs, using as much of the flag budget as possible. For each policy, the player's response is the row with the highest $v_\alpha(p_1)$. We check the flag budget, budget monotonicity, and $0 \le G_\alpha \le \epsilon_\alpha + W^+ - H_\alpha$ on every instance. On 144 selected instances, an independent LP agrees with the sorting routine within 7×10^{-9}. Pointwise 95% Monte Carlo intervals are based on paired differences across the 100 masks in each condition.

Figure 3 shows the reduction in external tolerance relative to uniform auditing for 128 actions. Table 2 gives both action-set sizes at $\delta = 0.05$. In these simulations, the improvement is largest when high-win-rate actions are rare under honest play. With varying rewards, the policy also improves on uniform auditing when those actions are common.

With binary rewards, all wins have the same value, so the wins-only rule reaches the optimal raw cap. When winning rewards vary, their magnitudes matter: Table 3 shows the further reduction from tailoring acceptance to payoff size. The accompanying validation script compares the sorting routine with an independent raw-cap LP on 500 cases with correlated references, including zero-probability pairs, payoff ties, and budget endpoints.

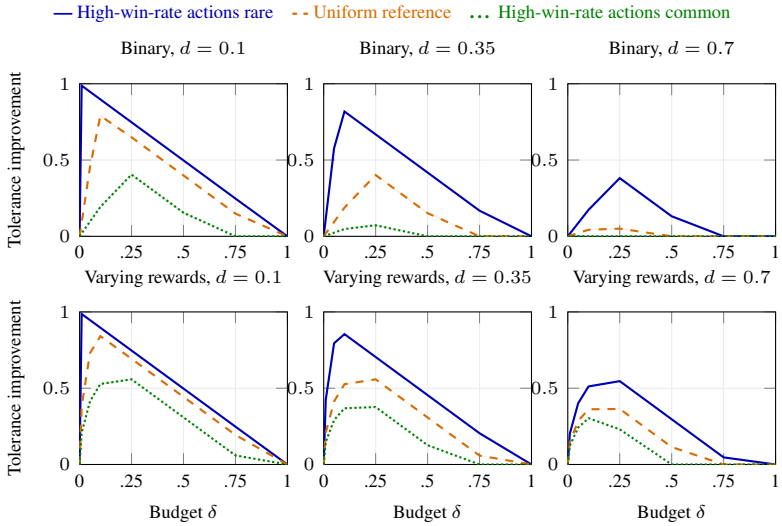

Figure 3: Improvement $\epsilon_{\text{unif}}(\delta) - \epsilon^*(\delta)$ from pair-dependent auditing with 128 player actions and 32 opponent actions. Each panel averages 100 independent games. Policies and reference conditions share the same games. Bands are pointwise approximate 95% Monte Carlo intervals for the mean paired improvement, clipped to $[0, 1]$ for display. Curves connect the eight evaluated budgets.

Rewards	Actions	d	Rare	Uniform	Common
Binary	32	0.10	0.947 ± 0.000	0.476 ± 0.029	0.108 ± 0.008
Binary	32	0.35	0.579 ± 0.037	0.097 ± 0.004	0.026 ± 0.002
Binary	32	0.70	0.086 ± 0.004	0.022 ± 0.001	0.002 ± 0.000
Binary	128	0.10	0.947 ± 0.000	0.454 ± 0.013	0.097 ± 0.003
Binary	128	0.35	0.578 ± 0.021	0.094 ± 0.002	0.024 ± 0.001
Binary	128	0.70	0.087 ± 0.003	0.022 ± 0.000	0.002 ± 0.000
Varying	32	0.10	0.941 ± 0.002	0.727 ± 0.015	0.420 ± 0.012
Varying	32	0.35	0.784 ± 0.018	0.415 ± 0.006	0.295 ± 0.005
Varying	32	0.70	0.401 ± 0.008	0.287 ± 0.002	0.243 ± 0.003
Varying	128	0.10	0.946 ± 0.000	0.729 ± 0.008	0.414 ± 0.005
Varying	128	0.35	0.794 ± 0.010	0.413 ± 0.003	0.291 ± 0.002
Varying	128	0.70	0.401 ± 0.004	0.286 ± 0.001	0.242 ± 0.002

Table 2: Mean improvement over uniform auditing at $\delta = 0.05$, with approximate 95% Monte Carlo half-widths over 100 games. Rare and common refer to the reference mass assigned to high-win-rate actions. Rounding to three decimals can hide small positive half-widths.

Comparing payoff-aware objectives. Next we ask what is lost by optimizing only the external payoff cap. With 128 actions, $d = 0.35$, a uniform honest reference, and $\delta = 0.25$, we reuse the same 100 masks in both reward conditions. We compare five policies: the minimum raw cap, the minimum external tolerance with maximum honest payoff, the minimum same-audit gain, the minimum gain subject to $H_\alpha \geq (1 - \delta)W^+$, and uniform acceptance.

For gain minimization, we optimize γ in equation 3 without an external-ceiling constraint. Among gain minimizers we first maximize H_α, then minimize the worst accepted pair payoff. Each stage is an LP with 10^{-9} slack on preceding optima. These gain-minimizing policies accept pairs with $U^+ = 0$ surely. We verify the budget, payoff floors, objective ordering, and primal–dual agreement to 2×10^{-8}.

Table 4 shows why the objectives select different audits. For binary rewards, the raw-cap policy has zero external tolerance but leaves a mean same-audit gain of 0.161. Minimizing gain subject to the honest-payoff floor reduces mean gain from 0.432 under uniform auditing to 0.102, while retaining the same mean honest payoff, 0.261. The cost is a higher external tolerance: 0.652 rather than 0.402.

Actions	d	Uniform	Wins only	Optimal
32	0.10	0.886	0.413	0.159
32	0.35	0.758	0.661	0.343
32	0.70	0.566	0.544	0.279
128	0.10	0.892	0.440	0.163
128	0.35	0.758	0.664	0.345
128	0.70	0.564	0.543	0.279

Table 3: External tolerance at $\delta = 0.05$ for varying rewards and a uniform honest reference, averaged over 100 games. "Wins only" applies one acceptance probability to every positive-payoff pair and accepts other pairs surely. The optimal rule also uses payoff magnitudes. For binary rewards, wins-only acceptance is itself optimal.

Rewards	Objective	R_α	H_α	G_α	ϵ_α	Flags
Binary	Minimum raw cap	0.259	0.098	0.161	0.000	0.250
Binary	Ceiling, retain honest	0.322	0.122	0.200	0.000	0.226
Binary	Minimum same-audit gain	0.107	0.098	0.009	0.652	0.250
Binary	Gain, honest floor	0.363	0.261	0.102	0.652	0.087
Binary	Uniform	0.693	0.261	0.432	0.402	0.250
Varying	Minimum raw cap	0.102	0.039	0.063	0.000	0.250
Varying	Ceiling, retain honest	0.173	0.065	0.108	0.000	0.190
Varying	Minimum same-audit gain	0.027	0.026	0.001	0.590	0.250
Varying	Gain, honest floor	0.202	0.144	0.059	0.806	0.081
Varying	Uniform	0.399	0.144	0.255	0.558	0.250

Table 4: Paired objective comparison over 100 games per reward condition, with 128 player actions, $d = 0.35$, uniform references, and $\delta = 0.25$. All columns are means. The gain policies maximize H_α among gain minimizers and then minimize the raw pairwise cap. The honest-floor variant also requires $H_\alpha \geq (1 - \delta)W^+$. The ceiling/retain policy maximizes H_α among external-tolerance minimizers.

Joint optimization with an honest best response. To isolate the effect of adding incentive constraints, we generate 100 independent 8×8 games in which honest play is already a best response before auditing. Player 2 is uniform and $\mu_1 = (1/2, 1/2, 0, \ldots, 0)$. Its two supported actions have distinct payoff profiles but the same unaudited return $W^+ = 0.45$. Every other action returns at most 0.35.

To construct each supported row, we center eight independent $\mathrm{Uniform}[-1, 1]$ draws, scale the largest absolute value to a draw from $\mathrm{Uniform}[0.25, 0.42]$, and add 0.45. For each other row, we rescale eight $\mathrm{Uniform}[0.05, 0.8]$ draws to a mean from $\mathrm{Uniform}[0.12, 0.35]$. We solve equation 4 at $\delta = 0.10$, varying the permitted gain over $\gamma/W^+ \in \{0, 0.01, 0.02, 0.04, \infty\}$ and the honest-payoff floor over $h/W^+ \in \{0, 0.90, 1\}$. Among tolerance minimizers, we maximize H_α.

Table 5 gives the resulting external tolerance. With a floor of $0.90W^+$, requiring zero same-audit gain raises mean tolerance from 0.099 without a gain bound to 0.129. Raising the floor all the way to W^+ raises tolerance to 0.335 throughout this grid. Intermediate gain limits give intermediate tolerances.

The payoff-only policy with maximum honest reference retention violates both $\gamma = 0$ and $H_\alpha \geq 0.90W^+$ in all 100 games at $\delta = 0.10$. Table 6 compares it with the joint policy at three flag budgets. At $\delta = 0.10$, enforcing both requirements raises mean tolerance from 0.038 to 0.129, while eliminating same-audit gain and meeting the honest-payoff floor.

We repeat the comparison with all eight actions in the honest reference. Each row now has unaudited return 0.45, so the uniform reference remains a best response. At $\delta = 0.10$, $\gamma = 0$, and $h = 0.90W^+$, Table 7 shows the same trade-off: mean external tolerance increases from 0.037 to 0.179 when the incentive and retention requirements are imposed.

Reference alignment. We also ask how the honest reference changes the advantage of payoff-aware auditing. This comparison uses all 600 masks with both reward conditions, all three references, and budgets $0.05, 0.25, 0.50$, for 10,800 paired instances. Gain minimization uses the same gain and

Honest floor	$\gamma = 0$	$0.01W^+$	$0.02W^+$	$0.04W^+$	No gain bound
$0.00W^+$	0.063 ± 0.007	0.055 ± 0.007	0.049 ± 0.007	0.041 ± 0.006	0.038 ± 0.005
$0.90W^+$	0.129 ± 0.010	0.117 ± 0.010	0.109 ± 0.009	0.101 ± 0.009	0.099 ± 0.008
$1.00W^+$	0.335 ± 0.011	0.335 ± 0.011	0.335 ± 0.011	0.335 ± 0.011	0.335 ± 0.011

Table 5: Mean optimal external tolerance in the joint LP at $\delta = 0.10$, varying the same-audit gain limit γ and honest-payoff floor h. Each cell uses 100 games in which the prescribed mixed reference is an unaudited best response. Here $\pm$ is an approximate 95% Monte Carlo half-width. The policy maximizes honest payoff among tolerance minimizers.

δ	Payoff-only ϵ	Payoff-only G_α	Joint ϵ	Joint G_α	ϵ increase
0.05	0.124	0.011	0.204	0.000	0.080 ± 0.005
0.10	0.038	0.012	0.129	0.000	0.091 ± 0.006
0.25	0.000	0.012	0.007	0.000	0.007 ± 0.003

Table 6: Joint LP with exact best-response preservation ($\gamma = 0$) and the floor $h = (1 - \delta)W^+$, compared with Algorithm 1 using its maximum-retention tie-break. Each row averages 100 paired games. The last column gives an approximate 95% Monte Carlo half-width for the increase in external tolerance. All generated references are best responses before auditing.

retention objectives as above. Its final tie-break minimizes rowwise flagging, so the resulting external tolerances may differ. A separate validation script checks gain against independent LP solutions on 120 random instances, with tolerance 2×10^{-8}.

Table 8 reports the reduction in same-audit gain at budget 0.25, while retaining at least the honest payoff of uniform auditing. In these simulations, the reduction is largest when the honest reference assigns little mass to high-win-rate actions. Each interval uses the 100 independent masks within its condition.

Estimated-reference stress test. The theory above uses known references. To study one source of estimation error, we sample $N \in \{100, 1000, 10000, 100000\}$ independent Player 1 actions from μ_1 for the first 30 masks in each condition and use their frequencies to estimate μ_1. We use flag budgets 0.05 and 0.25. The payoff table and opponent reference remain known. We first compute the audit from the estimated benchmark and flag budget, then evaluate it under the true reference.

The total-variation error is $\|\widehat{\mu}_1 - \mu_1\|_{\mathrm{TV}} := \frac{1}{2} \sum_{p_1} |\widehat{\mu}_1(p_1) - \mu_1(p_1)|$.

Across $8{,}640$ policy evaluations, $1{,}852$ exceeded the true flag budget. The largest excess was approximately 0.026. Table 9 reports the observed estimation errors, which decrease as the sample size grows. These results measure sensitivity to reference estimation in the generated games.

Honest support	Payoff-only ϵ	Payoff-only G_α	Joint ϵ	Joint G_α	ϵ increase
Two actions	0.038	0.012	0.129	0.000	0.091 ± 0.006
All eight actions	0.037	0.029	0.179	0.000	0.142 ± 0.006

Table 7: Support sensitivity at $\delta = 0.10$, $\gamma = 0$, and $h = 0.90W^+$. In both cohorts the honest mixed reference is an unaudited best response. Each entry averages 100 games. The final column gives an approximate 95% paired Monte Carlo half-width.

Rewards	Actions	d	Rare	Uniform	Common
Binary	32	0.10	0.339 ± 0.019	0.184 ± 0.013	0.097 ± 0.008
Binary	32	0.35	0.510 ± 0.014	0.262 ± 0.011	0.107 ± 0.007
Binary	32	0.70	0.425 ± 0.008	0.179 ± 0.005	0.041 ± 0.003
Binary	128	0.10	0.466 ± 0.017	0.305 ± 0.014	0.205 ± 0.012
Binary	128	0.35	0.580 ± 0.008	0.330 ± 0.007	0.170 ± 0.007
Binary	128	0.70	0.430 ± 0.004	0.180 ± 0.002	0.041 ± 0.002
Varying	32	0.10	0.186 ± 0.011	0.101 ± 0.007	0.054 ± 0.004
Varying	32	0.35	0.284 ± 0.009	0.146 ± 0.007	0.063 ± 0.004
Varying	32	0.70	0.261 ± 0.007	0.128 ± 0.005	0.054 ± 0.004
Varying	128	0.10	0.264 ± 0.011	0.174 ± 0.009	0.119 ± 0.008
Varying	128	0.35	0.334 ± 0.006	0.197 ± 0.006	0.109 ± 0.005
Varying	128	0.70	0.288 ± 0.004	0.151 ± 0.003	0.076 ± 0.003

Table 8: Reduction in same-audit gain relative to uniform auditing at $\delta = 0.25$, while retaining at least its honest payoff. Each entry is a mean paired reduction and approximate 95% Monte Carlo half-width over 100 games. Rare/common indicate alignment of honest reference mass with high-win-rate actions. The full grid also includes budgets 0.05 and 0.50.

Actions	Samples	Mean TV	Max. flag error	Max. tolerance excess
32	100	0.1681	0.0543	0.0461
32	1,000	0.0541	0.0140	0.0217
32	10,000	0.0172	0.0032	0.0043
32	100,000	0.0054	0.0014	0.0020
128	100	0.3433	0.0405	0.0578
128	1,000	0.1122	0.0114	0.0151
128	10,000	0.0355	0.0032	0.0044
128	100,000	0.0112	0.0013	0.0021

Table 9: Reference-estimation diagnostics. TV is the error in the estimated honest distribution. Flag error compares true and empirical flag probabilities. Tolerance excess is true external tolerance minus its empirical value, clipped at zero. Means and maxima pool the tested densities, reference alignments, rewards, and budgets. All values are empirical summaries of the simulated games.