DOMINATED ACTIONS SLOW DOWN CFR: COLD-START REGRET ATTENUATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Counterfactual regret minimization (CFR) is the leading framework for solving imperfect-information games and underlies superhuman poker AI. Its fastest variants, CFR$^+$, predictive CFR$^+$ (PCFR$^+$) and the discounted PDCFR$^+$, all start every information set from the uniform strategy, and that start carries a cost. At the first iteration the baseline against which regrets are measured is the unweighted mean of the counterfactual action values, so dominated actions pull it down and every non-dominated action receives the same additive offset in its first regret. The offset is uninformative, since pairwise differences are unchanged. Regret matching$^+$ keeps it in the cumulative regret until later updates outweigh it, so those iterations are spent separating good actions whose order the offset does not change. We propose *cold-start regret attenuation* (CSRA), which multiplies the first non-zero regret of each information set by a constant $c \in (0, 1]$ at the moment it appears and leaves every later update verbatim. The first strategy update is unchanged, the reduction costs one additive constant in the regret bound, and a single reduction preserves the convergence rate of any base algorithm whose analysis is sequence-agnostic. CSRA is the limiting case of PDCFR$^+$, and truncating PDCFR$^+$'s discount after its first iteration leaves its convergence within the variation across perturbed starts. Appending k dominated actions to every decision node of a controlled experiment multiplies the iterations PCFR$^+$ needs at $k = 32$ by 5.40, pushes CFR$^+$ past the budget, and changes those of CSRA by at most a factor of 1.27. On the standard benchmarks, evaluated with perturbation ensembles because the iteration at which a single run reaches a given exploitability is not reproducible on the games with large offsets, CSRA-PCFR$^+$ matches PDCFR$^+$ without a discount schedule. On the two games with the largest offsets, Battleship and Small Matrix, it reaches a given exploitability 2.43 to 9.38 times faster than PCFR$^+$ and up to 3.12 times faster than PDCFR$^+$, and it also accelerates PDCFR$^+$ and APCFR$^+$. A statistic computed from the first iteration alone, the common lower bound on the first regret, separates the games where the offset is large. The uniform start of CFR is not free, and CSRA removes its cost with a single multiplication at the first regret.

1 INTRODUCTION

Counterfactual regret minimization (CFR) (Zinkevich et al., 2007) decomposes the regret of an extensive-form game into one regret-minimization problem per information set, and it is the algorithmic core of every superhuman poker system (Bowling et al., 2015; Moravčík et al., 2017; Brown & Sandholm, 2018; 2019b). The fastest tabular variants replace regret matching (Hart & Mas-Colell, 2000) by regret matching$^+$ (CFR$^+$, Tammelin, 2014), add a prediction of the next regret (PCFR$^+$, Farina et al., 2021), or discount past regrets (DCFR, Brown & Sandholm, 2019a; PDCFR$^+$, Xu et al., 2024b). All of them initialise every information set at zero cumulative regret, that is, at the uniform strategy. This paper asks what that start costs.

At the first iteration the baseline against which regrets are measured is the unweighted mean of the counterfactual action values. When some actions are dominated (Ganzfried, 2025), a fold that gives up a won pot or a losing bet size in a no-limit tree, their low values pull the mean down, and every non-dominated action receives the same additive offset in its first regret (Lemma 1). The offset says nothing about which non-dominated actions belong to the equilibrium support, since their

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.

pairwise differences are unchanged, yet regret matching$^+$ retains it until later updates outweigh it. Proposition 1 bounds how long the strategy stays near-uniform over those actions in terms of the offset, so those iterations are spent separating good actions whose order the offset does not change.

Discounting reduces this cost only incidentally: DCFR and PDCFR$^+$ multiply the accumulated regret at iteration t by $t^\alpha/(t^\alpha+1)$ (Brown & Sandholm, 2019a; Xu et al., 2024b), which shrinks the first regret and every later one by a factor depending on α. We reduce the offset at its source. *Cold-start regret attenuation* (CSRA) multiplies the first non-zero regret of each information set by a constant $c \in (0, 1]$ at the moment it appears and leaves every subsequent update of the base algorithm verbatim. Because regret matching$^+$ is scale invariant from the zero accumulator, the first strategy update is unchanged, and the reduction adds only a constant independent of the horizon T to the regret bound, so the $O(1/\sqrt{T})$ rate is preserved (Theorem 1).

Experiments confirm the mechanism and delimit it. Appending k dominated actions to every decision node of Kuhn poker (Kuhn, 1950), Leduc poker (Southey et al., 2005) and Liar's dice (Neller & Hnath, 2011) multiplies the iterations PCFR$^+$ needs at $k = 32$ by 5.40 and pushes CFR$^+$ past the budget, while the iteration count of CSRA changes by at most a factor of 1.27; Goofspiel (Ross, 1971), whose equilibrium is nearly uniform over the good actions, is unaffected for every method (Section 5.2). A single run of PCFR$^+$ is not reproducible: two starts that differ only in a 10^{-9} perturbation of the first regret separate by a factor of 399.17 by $T = 5000$ (Farina et al., 2023). We therefore evaluate every method with perturbation ensembles and first-hit statistics (Section 5.1). Under that protocol CSRA-PCFR$^+$, PCFR$^+_{\gamma=5}$ and PDCFR$^+$ stay within a factor of 4.71 of one another on the games with two or three actions per information set. On the two games with the largest first-regret offsets, Battleship and Small Matrix (Xu et al., 2024a), CSRA-PCFR$^+$ reaches 10^{-6} 2.43 and 9.38 times faster than PCFR$^+$ respectively, matches or beats PDCFR$^+$, and CSRA also accelerates PDCFR$^+$ and APCFR$^+$ (Meng et al., 2026a); on poker-like games, where predictive methods lag CFR$^+$, CSRA is neutral (Section 5.3). PDCFR$^+$'s advantage over PCFR$^+$ on the small games comes mostly from its averaging exponent, and what its discount adds on top is a cold-start attenuation that has faded by the hundredth iteration (Section 5.4).

Contributions.

- We quantify the cost of the uniform start: a lower bound on the iterations regret matching$^+$ and its predictive variant spend separating the non-dominated actions, proportional to the offset that dominated actions create, an exact characterisation of the hitting time in a stationary two-action problem (Proposition 3), and a first-iteration statistic, the common lower bound on the first regret, which separates the games where the offset is large (Section 4.1).
- We introduce CSRA and prove that a single reduction adds one additive constant, independent of T, to the regret bound of any base algorithm whose bound holds for all regret sequences, covering CFR$^+$, PCFR$^+$, DCFR and PDCFR$^+$, so the $O(1/\sqrt{T})$ rate is preserved (Theorem 1); CSRA with $c = \frac{1}{2}$ is the limiting case of PDCFR$^+$ (Proposition 2).
- We isolate the mechanism in a controlled experiment that appends k dominated actions to every decision node, and evaluate all methods with perturbation ensembles and hitting-time statistics because a single run's first-hit iteration is not reproducible on the games with large offsets; under that protocol CSRA-PCFR$^+$ matches PDCFR$^+$ without a discount schedule on the standard benchmarks and is fastest on the two games with the largest offsets (Sections 5.1 and 5.3).

2 Related work

CFR variants. CFR$^+$ (Tammelin, 2014; Tammelin et al., 2015) introduced regret matching$^+$, linear averaging and alternating updates (Burch et al., 2019); DCFR (Brown & Sandholm, 2019a) discounts positive and negative cumulative regrets with separate exponents; PCFR$^+$ (Farina et al., 2021) derives predictive regret matching$^+$ from predictive Blackwell approachability (Abernethy et al., 2011) and connects it to optimistic mirror descent (Rakhlin & Sridharan, 2013; Farina et al., 2019b;a; Liu et al., 2022); PDCFR$^+$ (Xu et al., 2024b) combines prediction with discounting and is the strongest published tabular method on the non-poker benchmarks we use. APCFR$^+$ (Meng et al., 2026a) adapts the step size of the prediction, AutoCFR (Xu et al., 2022) and DDCFR (Xu et al., 2024a) learn discount schedules, and hyper-parameter schedules (Zhang et al., 2026) replace learned schedules by a hand-designed one. Parameter-free regret matching with $O(1/T)$ rate (Meng

et al., 2026b) and scale-invariant regret matching (Zhang et al., 2025) address the accumulated regret from the optimisation side. Discounting treats early regrets differently only through a schedule that also acts on every later regret (Section 4.3); CSRA isolates the first non-zero update and leaves the rest of the base algorithm unchanged, so it composes with all of them; we combine it with CFR$^+$, PCFR$^+$, PDCFR$^+$ and APCFR$^+$.

Initialisation of regret minimisers. Warm starting (Brown & Sandholm, 2016) constructs cumulative regrets that are consistent with a given strategy so that CFR can continue from it, and regret transfer (Brown & Sandholm, 2014) rescales regrets when the game's payoffs change. Both scale regrets and both rely on the same fact we use in Theorem 1: a bounded, one-off modification of the accumulator costs only an additive constant in the regret bound. Our contribution is to identify *which* part of the standard initialisation is harmful and to remove it without any auxiliary strategy.

Instability of regret matching$^+$. Farina et al. (2023) show that regret matching$^+$ can be unstable and that its alternating and predictive variants converge fast in practice without a matching theory; last-iterate results exist for smoothed variants (Cai et al., 2025; Meng et al., 2025). The dependence on the initial accumulator that we document in Section 5.1 is a practical consequence of this instability and motivates our ensemble-based evaluation.

3 PRELIMINARIES

Extensive-form games and CFR. We consider two-player zero-sum extensive-form games with perfect recall. Player i acts at information sets $I \in \mathcal{I}_i$ with action set $A(I)$; a strategy σ_i assigns a distribution $\sigma_i(I) \in \Delta(A(I))$ to every I. For a profile $\sigma = (\sigma_1, \sigma_2)$ the counterfactual value $v^\sigma(I, a)$ (Zinkevich et al., 2007) is the expected utility of player i after taking a at I, weighted by the probability that chance and the opponent reach I. The *instantaneous counterfactual regret* at iteration t is

$$r_t(I, a) = v^{\sigma_t}(I, a) - \textstyle\sum_{b \in A(I)} \sigma_t(I, b)\, v^{\sigma_t}(I, b), \qquad \text{so} \qquad \langle \sigma_t(I), r_t(I) \rangle = 0. \tag{1}$$

CFR (Zinkevich et al., 2007) runs an independent regret minimiser at every information set and returns the (weighted) average strategy $\bar{\sigma}_T$; its exploitability, half the sum of the two players' best-response gains $\frac{1}{2}\mathrm{NashConv}(\bar{\sigma}_T)$ (Johanson et al., 2013), is bounded by $\sum_I \max_a R_T(I, a)^+ / T$, where $R_T(I, a) = \sum_{t \le T} r_t(I, a)$ is the cumulative regret. All algorithms in this paper use alternating updates (Tammelin, 2014; Burch et al., 2019).

Regret matching$^+$ and its descendants. Fix one information set and write $r_t \in \mathbb{R}^n$, $n = |A(I)|$, and $\Delta \ge \|r_t\|_\infty$ for the utility range, and $\rho \le \Delta$ for a bound on $|r_t(I, a)|$ over the window of iterations a statement considers. Every method below keeps a non-negative accumulator Q_t and plays the normalised vector $\sigma_{t+1} \propto s_t$:

$$\text{RM}^+ \text{ (CFR}^+\text{)}: \quad Q_t = [Q_{t-1} + r_t]^+, \qquad s_t = Q_t; \tag{2}$$

$$\text{PRM}^+ \text{ (PCFR}^+\text{)}: \quad Q_t = [Q_{t-1} + r_t]^+, \qquad s_t = [Q_t + r_t]^+; \tag{3}$$

$$\text{DCFR}: \quad Q_t = Q_{t-1} \odot d_{t-1} + r_t, \qquad s_t = [Q_t]^+; \tag{4}$$

$$\text{PDCFR}^+: \quad Q_t = \left[Q_{t-1} \tfrac{(t-1)^\alpha}{(t-1)^\alpha + 1} + r_t\right]^+, \qquad s_t = \left[Q_t \tfrac{t^\alpha}{t^\alpha + 1} + r_t\right]^+, \tag{5}$$

with $Q_0 = 0$ and the uniform strategy at $t = 1$; in (4), $d_t(a) = t^\alpha/(t^\alpha + 1)$ if $Q_t(a) > 0$ and $t^\beta/(t^\beta + 1)$ otherwise. The average strategy weights iteration t by t (CFR$^+$), t^2 (PCFR$^+$) or t^γ (DCFR, PDCFR$^+$) (Tammelin, 2014; Brown & Sandholm, 2019a; Farina et al., 2021; Xu et al., 2024b). PRM$^+$ is a predictive (optimistic) method: the last regret serves as the prediction of the next one. We use the published hyper-parameters $(\alpha, \beta, \gamma) = (1.5, 0, 2)$ for DCFR (Brown & Sandholm, 2019a) and $(\alpha, \gamma) = (2.3, 5)$ for PDCFR$^+$ (Xu et al., 2024b), and also report PCFR$^+$ with $\gamma = 5$, since the averaging exponent turns out to matter (Section 5). APCFR$^+$ and SAPCFR$^+$ (Meng et al., 2026a) replace s_t in (3) by $[Q_t + r_t/(\alpha_t + 1)]^+$ with an adaptive or fixed α_t; their exact update is given in Appendix B.

4 COLD-START REGRET ATTENUATION

4.1 WHY THE FIRST REGRET IS SPECIAL

With the standard initialisation $Q_0 = 0$ every information set plays the uniform strategy at $t = 1$, so the baseline in (1) is the *unweighted mean* of the counterfactual action values. Split the actions into a set G of n_G "good" actions and a set B of k actions whose counterfactual value under the uniform profile is lower by a margin $g > 0$: $v(b) \le \min_{a\in G} v(a) - g$ for all $b \in B$ (strictly dominated actions are the canonical example).

Lemma 1 (uniform-baseline inflation)**.** *Under the uniform strategy, for every $a \in G$,*

$$r_1(a) \;\ge\; \big(v(a) - \bar v_G\big) + \frac{k\,g}{n_G + k}, \qquad \bar v_G = \tfrac{1}{n_G}\textstyle\sum_{a\in G} v(a),$$

while $r_1(a) - r_1(a') = v(a) - v(a')$ for all $a, a' \in G$.

The dominated actions therefore add a common offset $J \ge kg/(n_G + k)$ to the first regret of every good action. The offset carries no information about which good actions belong to the equilibrium support, since the differences among good actions are untouched, but RM$^+$ keeps it in Q until later regrets outweigh it. The next proposition quantifies the price. It covers regret matching$^+$ and predictive regret matching$^+$ and their discounted variants, and, through Lemma 1, any later iteration at which an information set receives its first non-zero regret, since its own strategy is still uniform there.

Proposition 1 (inertia of a common offset)**.** *Fix an information set and two actions a, b with $Q_\tau(a), Q_\tau(b) \ge J > 0$ and $Q_\tau(a) \le Q_\tau(b) + D$ ($D \ge 0$). Suppose $|r_t(a)|, |r_t(b)| \le \rho$ for $\tau < t \le \tau + s$, and let $P_s = \prod_{j=\tau}^{\tau+s-1} d_j \in (0, 1]$ be the product of the discount factors applied to Q between iterations τ and $\tau + s$ ($P_s = 1$ for RM$^+$ and PRM$^+$). If $s\rho < P_s J$, then for RM$^+$ and DCFR*

$$\frac{\sigma_{\tau+s+1}(a)}{\sigma_{\tau+s+1}(b)} \;\le\; \frac{P_s(J + D) + s\rho}{P_s J - s\rho},$$

and for PRM$^+$ and PDCFR$^+$ the same bound holds with s replaced by $s + 1$ and P_s by P_{s+1}, for every $s \ge 1$ with $(s + 1)\rho < P_{s+1}J$. Consequently the played ratio stays below $\kappa > 1$ as long as $(\kappa + 1)s\rho < P_s\big((\kappa - 1)J - D\big)$. Under CSRA$(c)$ the hypotheses hold with (cJ, cD), so the right-hand side of this condition is multiplied by c.

The bound is a lower bound on the number of iterations before regret matching$^+$ can favour one of the two actions by a factor κ. As long as the later regrets stay below ρ, at least $P_s((\kappa - 1)J - D)/((\kappa + 1)\rho)$ iterations pass with the two actions played almost equally. Discounting enters only through P_s, and CSRA(c) shortens the bound by the factor c at every information set. In a stationary problem this lower bound is tight up to a constant (Proposition 3, Appendix A), so the hitting time itself scales with c.

4.2 THE METHOD

CSRA rescales the regret of each information set at exactly one iteration. Let $\tau(I)$ be the first iteration at which $r_\tau(I) \ne 0$; before τ the accumulator is zero and the strategy uniform for every base algorithm above. CSRA feeds the base algorithm the rescaled vector

$$\tilde r_\tau(I) \;=\; c\,r_\tau(I), \qquad c \in (0, 1], \tag{6}$$

in place of $r_\tau(I)$, so that, because $Q_{\tau-1}(I) = 0$, the first accumulator becomes $Q_\tau(I) = c[r_\tau(I)]^+$, and leaves every other step of the base algorithm verbatim: the same accumulation rule, the same prediction, the same averaging. Two facts make this a clean intervention.

Lemma 2 (first-step invariance)**.** *Assume $Q_{\tau-1}(I) = 0$, as under the initialisation $Q_0 = 0$ and the perturbation protocol of Appendix C.4. The strategy played at iteration $\tau + 1$ is the same with and without CSRA, for RM$^+$, PRM$^+$ and their discounted variants.*

Proof. For RM$^+$, $Q_\tau/\|Q_\tau\|_1$ does not depend on $c > 0$. For PRM$^+$ the strategy vector is $[c[r_\tau]^+ + c\,r_\tau]^+ = 2c[r_\tau]^+$: on $\{r_\tau \ge 0\}$ the entry is $c\,r_\tau(a) + c\,r_\tau(a)$, and on $\{r_\tau < 0\}$ it is $c\cdot 0 + c\,r_\tau(a) < 0$

and is clipped; normalising removes the scalar. For PDCFR$^+$ the vector is $[d_\tau c[r_\tau]^+ + c\, r_\tau]^+ = c(1+d_\tau)[r_\tau]^+$ by the same argument, and DCFR-type variants scale Q_τ by a positive scalar before normalising. □

CSRA therefore changes the *magnitude* of the first regret and nothing about the first strategy update; by Proposition 1 the lower bound on the iterations lost to the offset shrinks by the factor c. The factor c is the only hyper-parameter. We fix $c = 0.1$ for every game and every base algorithm after a small grid search (Section 5.5); the method is insensitive to c over more than an order of magnitude ($0.02 \le c \le 0.2$ behave alike where the mechanism is active), while $c = 0$, which discards the first regret altogether, is harmful because r_τ also carries the ranking among the good actions.

4.3 Guarantees

Theorem 1 (a single reduction preserves the no-regret rate)**.** *Let ALG be RM$^+$, PRM$^+$, DCFR or PDCFR$^+$, let $w_t > 0$ be the weights under which ALG's guarantee is stated ($w_t \equiv 1$ for RM$^+$ and PRM$^+$), and let $B_T(\tilde r)$ be a bound on the weighted regret $\max_a \sum_{t\le T} w_t \tilde r_t(a)$ of ALG run on an input sequence $\tilde r$, valid for every sequence with $\langle \sigma_t, \tilde r_t\rangle = 0$ and $\|\tilde r_t\|_\infty \le \Delta$; B_T may depend on $\tilde r$ (as the predictive bound of Farina et al. (2021) does through $\sum_t \|\tilde r_t - \tilde r_{t-1}\|^2$) or not ($\Delta\sqrt{nT}$ for RM$^+$). Then the regret of CSRA(c)-ALG on its own trajectory satisfies*

$$\max_a \sum_{t\le T} w_t\, r_t(a) \;\le\; B_T(\tilde r) + (1-c)\, w_\tau \Delta,$$

where $\tilde r$ is the sequence CSRA fed to ALG ($\tilde r_t = r_t$ for $t \ne \tau$, $\tilde r_\tau = c\, r_\tau$). In particular CSRA-CFR$^+$ and CSRA-PCFR$^+$ inherit $B_T = \Delta\sqrt{nT}$ and the predictive bound of Farina et al. (2021), and the average strategy of CSRA-CFR$^+$ is an $O(|\mathcal{I}|\Delta\sqrt{n/T})$-Nash equilibrium (Nash Jr, 1950).

Proof. By definition (6), CSRA(c)-ALG is ALG run on the modified sequence $\tilde r_t = r_t$ for $t \ne \tau$ and $\tilde r_\tau = c\, r_\tau$; its accumulators, predictions and strategies are those of ALG on $\tilde r$. The modified sequence is admissible since $\langle \sigma_\tau, c\, r_\tau\rangle = 0$ and $\|c\, r_\tau\|_\infty \le \Delta$. Hence $\sum_t w_t \tilde r_t(a) \le B_T(\tilde r)$ for every a, and $\sum_t w_t r_t(a) = \sum_t w_t \tilde r_t(a) + (1-c) w_\tau r_\tau(a) \le B_T(\tilde r) + (1-c) w_\tau \Delta$. For $w \equiv 1$ the counterfactual regret decomposition of Zinkevich et al. (2007) bounds the exploitability of the average strategy by $\sum_I (B_T + \Delta)/T$; for the weighted variants the weighted decomposition of Brown & Sandholm (2019a) and Xu et al. (2024b) applies in the same way. □

CSRA costs one additive constant in the bound and nothing in the rate: since $B_T \ge \Delta\sqrt{n} \ge (1-c)\Delta$ for RM$^+$ and PRM$^+$, the total is at most $2B_T$ for every $T \ge 1$. The same proof shows that CSRA composes with any weighting or prediction scheme whose analysis is sequence-agnostic.

Proposition 2 (CSRA is the $\alpha \to \infty$ limit of PDCFR$^+$)**.** *Let $d_t = \frac{t^\alpha}{t^\alpha+1}$; then $d_1 = \frac{1}{2}$ for every α and $d_t \to 1$ for $t \ge 2$ as $\alpha \to \infty$. In* (5) *the factor d_1 multiplies Q_1 when Q_2 and σ_2 are formed, so the limit of PDCFR$^+(\alpha,\gamma)$ is PCFR$^+$ with t^γ averaging in which the accumulator of the first iteration is halved once, that is, CSRA$(\frac{1}{2})$ applied to every information set whose first non-zero regret appears at $t = 1$; information sets first reached later receive no reduction in the limit, so CSRA is slightly more general. Likewise DCFR(α,γ) with a clipped accumulator converges to CSRA$(\frac{1}{2})$ applied to CFR$^+$.*

Proposition 2 places discounting and CSRA on one axis, and PDCFR$^+$ at its published $\alpha = 2.3$ already sits near that end of it (Section 5.4). Proofs of Lemma 1 and Proposition 1 are in Appendix A.

5 Experiments

5.1 Protocol

Games. We use the benchmark games of Xu et al. (2024b) and Meng et al. (2026a): Kuhn poker, Leduc poker, Liar's dice with 4/5/6-sided dice, imperfect-information Goofspiel with 4/5/6 cards, and Battleship (3×2 grid, one ship of length 2, three shots), all in their OpenSpiel implementations (Lanctot et al., 2019). We add three no-limit poker games in which *every* bet size is legal (two

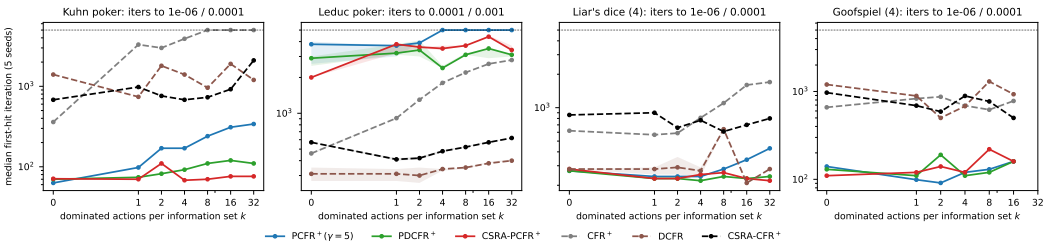

Figure 1: Iterations to reach a fixed exploitability (median over five perturbed starts, shaded: range) as a function of the number k of dominated actions appended to every decision node. Solid: predictive methods, threshold 10^{-6} (10^{-4} on Leduc); dashed: non-predictive methods, threshold 10^{-4} (10^{-3} on Leduc). The dotted line is the iteration budget. CSRA uses $c = 0.1$.

betting rounds, stacks of 6, 10 and 12 chips; the last with a six-rank deck), because unrestricted betting trees are the natural source of many dominated actions. We also include the Small Matrix game of Xu et al. (2024a), a rock–paper–scissors variant with two catastrophic trap actions (tree sizes in Table 3).

Algorithms. CFR$^+$, Linear CFR (Brown & Sandholm, 2019a), DCFR$(1.5, 0, 2)$, PCFR$^+$ with quadratic (γ=2) and quintic (γ=5) averaging, PDCFR$^+(2.3, 5)$, SAPCFR$^+$ and APCFR$^+$ with the settings of their authors, and CSRA applied to CFR$^+$, PCFR$^+(\gamma{=}5)$, PDCFR$^+$ and APCFR$^+$ with $c = 0.1$ throughout. All methods use alternating updates and share one implementation of the tree traversal, so they differ only in the local update (Appendix B). Exploitability is $\frac{1}{2}\mathrm{NashConv}$ of the average strategy, floored at 10^{-12} as in Xu et al. (2024b).

Chaos and ensembles. Table 5 in Appendix C.4 shows why a single run per configuration is not enough. Perturbing the first non-zero regret of every information set by a factor $1 + 10^{-9}$, far below any quantity that enters the algorithm (Lemma 3), moves the exploitability of PCFR$^+$ at $T = 100$ by at most a factor of 1.39 on any game, but the same starts separate as they run: by $T = 5000$ the spread over the starts of one method reaches 399.17× on Battleship and 205.48× on Goofspiel (5), against at most 8.51× on the other games. We therefore run every configuration from the exact uniform start and from four perturbed starts, the methods of the headline comparison from nine on Battleship and Small Matrix, and report the *median first-hit iteration*, the first logged iteration at which exploitability falls below a threshold, together with the range over the runs.

5.2 Dominated actions slow CFR$^+$ and PCFR$^+$; CSRA removes the dependence

To isolate the mechanism of Section 4.1 we append k dominated actions to every decision node of a game: each is a terminal move in which the acting player loses twice the game's largest payoff. The equilibria of the original game are unchanged, and after the first update every regret-matching$^+$ method assigns the new actions probability zero, so k acts on the first iteration only: it adds the offset of Lemma 1 to the acting player's first regrets, and it makes the uniform opponent blunder with probability $k/(|A| + k)$ while those regrets are computed. The inflation share of Appendix C.6 grows from 0.40 to 0.58 on Leduc poker and from 0.65 to 0.80 on Kuhn poker as k goes from 2 to 8. Figure 1 reports the median first-hit iteration as a function of $k \in \{0, 1, 2, 4, 8, 16, 32\}$ on Kuhn poker, Leduc poker and Liar's dice (4), and $k \le 16$ on Goofspiel (4), for $T \le 5000$.

The picture is the one the theory predicts. On Kuhn poker the iterations PCFR$^+$ needs grow from 63 at $k = 0$ to 340 at $k = 32$, and CFR$^+$ leaves the budget at $k = 8$; on Leduc poker PCFR$^+$ leaves it from $k = 4$ on; on Liar's dice PCFR$^+$ goes from 270 to 430. With CSRA the same curves are flat: CSRA-PCFR$^+$ stays below 110 iterations on Kuhn poker and below 280 on Liar's dice over all seven values of k. CSRA-CFR$^+$ stays bounded as well, at most 2100 iterations on Kuhn poker, where CFR$^+$ leaves the budget at $k = 8$, and at most 620 on Leduc poker, where CFR$^+$ reaches 2800. Discounting mitigates the dependence: PDCFR$^+$ goes from 70 to 120 on Kuhn poker and is flat on Liar's dice, and DCFR, whose discount also acts on negative regrets, varies by less than a factor of two on Leduc poker and less than four on Kuhn poker and Liar's dice. Goofspiel is the exception: no method's iteration count varies by more than a factor of four over $k \le 16$, because

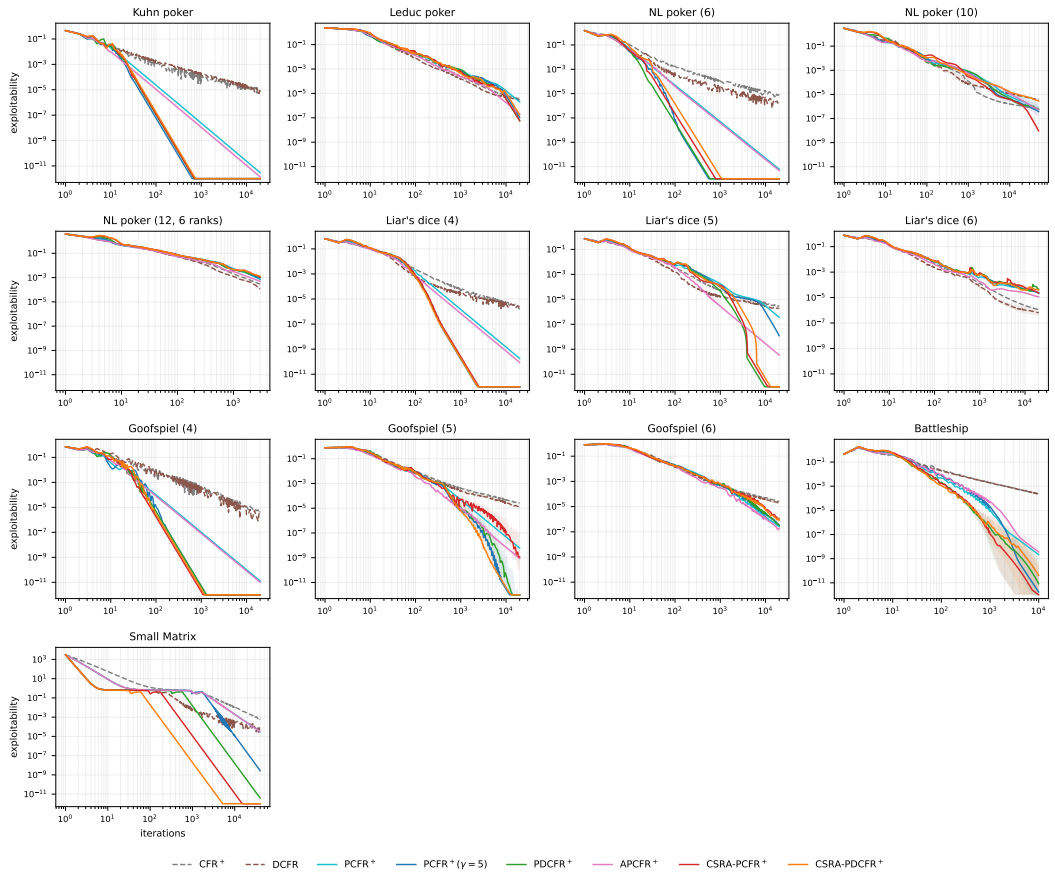

Figure 2: Exploitability of the average strategy on the benchmark games (median over the perturbed starts, shaded: min–max). Dashed: non-predictive methods.

Table 1: Median first-hit iteration to exploitability $\leq 10^{-6}$ (ten starts on Battleship and Small Matrix for the headline methods, five elsewhere). "$> T\,(j/n)$": the threshold was not reached within the common budget T (Table 4) by $n - j$ of the n starts, so the starts pin no median. Bold: fastest identified median.

Method	Kuhn poker	Leduc poker	NLP (6)	NLP (10)	NLP (12, 6 ranks)	LD (4)	LD (5)	LD (6)	GS (4)	GS (5)	GS (6)	Battleship	Small Matrix
CFR$^+$	>20k (0/5)	>20k (0/5)	>20k (0/5)	**23000**	>3k (0/5)	>20k (0/5)	>20k (0/5)	>20k (0/5)	>20k (0/5)	>20k (0/5)	>20k (0/5)	>10k (0/5)	>40k (0/5)
DCFR	>20k (0/5)	>20k (0/5)	14000	>50k (0/5)	>3k (0/5)	>20k (0/5)	>20k (0/5)	**11000**	18000	>20k (0/5)	>20k (0/5)	>10k (0/5)	>40k (0/5)
PCFR$^+$	**290**	>20k (0/5)	370	35000	>3k (0/5)	1200	15000	>20k (0/5)	470	3700	12000	1600	>40k (0/5)
PCFR$^+_{\gamma=5}$	**63**	14000	77	29000	>3k (0/5)	270	9900	>20k (0/5)	140	1900	14000	1750	15000
PDCFR$^+$	70	13000	**59**	39000	>3k (0/5)	270	2100	>20k (0/5)	130	1900	15000	760	5000
SAPCFR$^+$	**220**	12000	360	43000	>3k (0/5)	780	>20k (1/5)	>20k (0/5)	510	2700	9300	4600	>40k (0/5)
APCFR$^+$	**230**	13000	340	35000	>3k (0/5)	890	1500	>20k (0/5)	440	2200	9200	2500	>40k (0/10)
CSRA-CFR$^+$	>20k (0/5)	>20k (0/5)	**9800**	22000	>3k (0/5)	>20k (0/5)	>20k (0/5)	17000	11000	>20k (0/5)	>20k (0/5)	>10k (0/5)	>40k (0/5)
CSRA-PCFR$^+$	**71**	13000	79	22000	>3k (0/5)	280	2500	>20k (0/5)	110	3800	>20k (2/5)	720	1600
CSRA-PDCFR$^+$	**75**	15000	110	>50k (0/5)	>3k (0/5)	260	3200	>20k (0/5)	120	1600	18000	745	530
CSRA-APCFR$^+$	**220**	18000	440	42000	>3k (0/5)	1000	11000	>20k (0/5)	580	3600	13000	1150	13000

the good actions are tied by symmetry at the first iteration, so the offset costs nothing (Proposition 1 with $\kappa \approx 1$).

5.3 Standard benchmarks

Figure 2 and Table 1 summarise the benchmark runs, each game over its own iteration budget T_g (Table 4). The table reports medians and ranges over the starts, ten on Battleship and Small Matrix for the headline methods and five elsewhere, and we quote medians below. Three regimes appear, and a diagnostic computed from the first iteration alone (Appendix C.6) separates them: a common offset can only delay the separation of good actions at information sets whose first regret has at least

Table 2: Truncating the discount of PDCFR$^+$: median first-hit iteration to 10^{-6} [min, max] over the perturbed starts; notation as in Table 1. The truncation variants are the $T = 5000$ ablation runs of Appendix C.8; the reference rows are repeated at T_g.

Variant	Leduc poker	Liar's dice (4)	Liar's dice (5)	Goofspiel (4)	Goofspiel (5)	Battleship
PCFR$^+_{\gamma=5}$ (no discount)	14000 [14000, 15000]	270 [270, 270]	9900 [9900, 9900]	140 [140, 140]	1900 [1400, 2900]	1750 [1600, 1800]
discount only at $t \leq 1$	>5k (0/5) [–, >5k]	280 [280, 280]	>5k (0/5) [–, >5k]	130 [130, 130]	1200 [1200, 1400]	840 [770, 1400]
discount only at $t \leq 3$	>5k (0/5) [–, >5k]	240 [240, 240]	4700 [4700, 4700]	130 [130, 130]	2400 [2000, 2700]	790 [660, 840]
discount only at $t \leq 10$	>5k (0/5) [–, >5k]	280 [280, 280]	1500 [1500, 1500]	140 [140, 140]	2000 [1800, 3300]	790 [620, 920]
PDCFR$^+$ (full discount)	13000 [13000, 13000]	270 [270, 270]	2100 [2100, 2100]	130 [130, 130]	1900 [1700, 2400]	760 [540, 1200]
CSRA-PCFR$^+$	13000 [13000, 13000]	280 [280, 280]	2500 [2500, 2500]	110 [110, 110]	3800 [2700, 5200]	720 [590, 1600]

two positive entries, and the fraction of such sets is 0.50 in Small Matrix and 0.64 in Battleship, 0.15–0.32 in Liar's dice and Goofspiel, and 0.00–0.25 in the poker games. In Small Matrix and Battleship the positive entries of those sets are equal, so the whole positive mass is offset; in the poker games and Liar's dice the offset (the smallest positive entry) is 0.19–0.40 of the positive mass (Table 6).

Games with large offsets. On Small Matrix, where half of the information sets have two good actions and two catastrophic ones, CSRA-PCFR$^+$ reaches 10^{-6} in 1600 iterations against 15000 for PCFR$^+_{\gamma=5}$ and 5000 for PDCFR$^+$, and its median lies below the fastest of the ten PCFR$^+_{\gamma=5}$ runs. CSRA-PDCFR$^+$ needs 530 and CSRA-APCFR$^+$ 13000, while APCFR$^+$ does not reach the threshold at all; on this game the first-hit iteration of CSRA-PCFR$^+$ increases monotonically with c from $c = 0.1$ to $c = 1$ (Table 8). On Battleship CSRA-PCFR$^+$ reaches 10^{-6} in 720 iterations, again below the fastest of the ten PCFR$^+_{\gamma=5}$ runs, against 1750 and 760 for PCFR$^+_{\gamma=5}$ and PDCFR$^+$; at 10^{-9} the medians are 2600, 3750 and 4200. CSRA also speeds up APCFR$^+$ (1150 against 2500) and PDCFR$^+$ (745 against 760) on this game. CFR$^+$, Linear CFR and DCFR do not reach 10^{-6} within the budget with or without CSRA (Table 7). The two other games with many actions behave like the small games: on Goofspiel (6) CSRA-PCFR$^+$, PCFR$^+_{\gamma=5}$ and PDCFR$^+$ reach 10^{-5} in 8900, 5800 and 5700 iterations, the three medians within a factor of 1.6, and on Liar's dice (6), where the offset share is 0.28, they reach 10^{-4} in 6300, 4400 and 5800 iterations while CFR$^+$ (1200) and DCFR (920) are faster than every predictive method on this game.

Games with two or three actions per information set. Where all three of CSRA-PCFR$^+$, PCFR$^+_{\gamma=5}$ and PDCFR$^+$ reach 10^{-6} inside the common budget, the three medians stay within a factor of 4.71 of one another, and the widest of those games is Liar's dice (5) (9900 against 2100 and 2500 iterations). The largest median regression of CSRA-PCFR$^+$ against PCFR$^+_{\gamma=5}$ is 100% on Goofspiel (5) (3800 against 1900, the latter ranging from 1400 to 2900 across its starts), and every other regression is at most 13%. CSRA-PDCFR$^+$ recovers that loss (1600). This is the regime Lemma 1 predicts when k is small.

Poker-like games. On Leduc poker and the 10-chip no-limit poker game the predictive methods reach 10^{-6} at medians beyond the nominal budget, and CFR$^+$ and DCFR are competitive or better, in line with Xu et al. (2024b) and Meng et al. (2026a). On the 12-chip game, where no method reaches 10^{-6} even then, CSRA is slower: at $T = 3000$ the geometric-mean exploitability of CSRA-PCFR$^+$ is 1.1e-03 [9.8e-04, 1.2e-03] against 9.5e-04 [8.3e-04, 1.1e-03] for PCFR$^+_{\gamma=5}$ and 1.1e-03 [1.0e-03, 1.2e-03] for PDCFR$^+$, and that of CSRA-CFR$^+$ is 3.6e-04 [3.1e-04, 5.1e-04] against 2.9e-04 [2.8e-04, 3.4e-04] for CFR$^+$ (DCFR: 1.3e-04 [1.0e-04, 1.8e-04]), factors of 1.16 and 1.24 with ranges that touch. The first-update diagnostic is 0.11–0.25 on these games with an offset share below 0.31: there is little offset to remove, and CSRA does not act on whatever slows the predictive methods there.

5.4 Where the gains of discounting come from

Two observations locate the effect of PDCFR$^+$'s discount. First, it is not an averaging effect. Raising the averaging exponent of PCFR$^+$ from $\gamma = 2$ to $\gamma = 5$ already accounts for the gap to PDCFR$^+$ on the games with two or three actions (Table 1, e.g. Liar's dice (4), 1200 against 270 and 270). Xu et al. (2024b) report that PDCFR$^+$ surpasses PCFR$^+$ with $\gamma = 5$ averaging by five to seven

orders of magnitude on Liar's dice (5) and Goofspiel (5); we reproduce the first at $T = 5000$, where the median exploitability of PCFR$^+_{\gamma=5}$ is 9.3×10^{-6} against 4.5×10^{-11}, but not the second, where the two medians are 3.2×10^{-9} and 1.0×10^{-8} and the $\gamma = 5$ method is the lower of the two. Second, truncating the discount, that is, applying it only for the first K iterations and running PCFR$^+$ verbatim afterwards, changes little on the games where the truncated runs still reach the threshold (Table 2): on Battleship a truncation after three iterations matches the full discount (790 against 760, from 1750 for PCFR$^+_{\gamma=5}$ with no discount). Proposition 2 says why truncation is cheap: PDCFR$^+$'s discount acts on the first accumulator with a cumulative factor of ≈ 0.33 that has faded by the hundredth iteration, which is what constant-c CSRA with $c = \frac{1}{3}$ reproduces.

5.5 Ablations

Where the offset is large every $c \in [0.02, 0.2]$ gives essentially the same first-hit iteration and $c = 1$ recovers the base method (Table 8). On the other games of that table $c \in [0.1, 0.5]$ stays within a factor of three, except on Liar's dice (5), where the medians span a factor of eight, and on Goofspiel (6), where the $c = 0.1$ entry is not identified; $c = 0.05$ is the slow end on the smallest no-limit poker game (110 against 53–79). Neither reducing the first $K > 1$ iterations in place of the first regret nor the per-information-set factors is consistently better (Appendix C.8, which also reports the $k = 8$ ablation).

6 Conclusion

The uniform start of CFR carries a cost that has gone unnoticed: dominated actions inflate the first regret of every non-dominated action by a common offset, and regret matching$^+$ then spends a number of iterations proportional to that offset before it can separate the non-dominated actions. Cold-start regret attenuation removes the cost with a single multiplication at the moment the first regret appears and adds one additive constant to the regret bound of whatever base algorithm it is attached to. In a controlled experiment the iteration count of CSRA-CFR$^+$ and CSRA-PCFR$^+$ changes by at most a factor of 1.27 as the number of dominated actions grows to 32, while PCFR$^+$ needs 5.40 times as many iterations and CFR$^+$ is pushed past the budget. On standard benchmarks CSRA-PCFR$^+$ matches PDCFR$^+$ without a discount schedule and is 2.43 to 9.38 times faster than PCFR$^+$ on the two games with the largest first-regret offsets, the games a first-iteration statistic identifies in advance; the analysis of PDCFR$^+$'s discount shows that it is cold-start regret attenuation in disguise. The mechanism also delimits the method: CSRA has nothing to remove at information sets with few actions or whose equilibrium is close to uniform over the good actions, and a principled per-information-set choice of c is open. Since the first-hit iteration of a single run of predictive regret matching$^+$ is not reproducible on the games with large offsets, comparisons of CFR variants should use perturbation ensembles.

AI use statement

We used generative AI tools (Claude, Anthropic) to draft and edit the manuscript and to write parts of the experiment code. All AI-assisted text was reviewed and edited by the authors, and the code was tested against OpenSpiel's reference implementations. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

Reproducibility statement

All algorithms share one vectorised tree traversal whose CFR and CFR$^+$ outputs match OpenSpiel's reference implementations to floating-point precision. Every number is reproducible from the anonymous repository in the supplementary material, which holds the run drivers, the job lists and the compiled solver core.

References

Jacob Abernethy, Peter L Bartlett, and Elad Hazan. Blackwell approachability and no-regret learning are equivalent. In *Proceedings of the 24th Annual Conference on Learning Theory*, pp. 27–46.

JMLR Workshop and Conference Proceedings, 2011.

Michael Bowling, Neil Burch, Michael Johanson, and Oskari Tammelin. Heads-up limit hold'em poker is solved. *Science*, 347(6218):145–149, 2015.

Noam Brown and Tuomas Sandholm. Regret transfer and parameter optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28, 2014.

Noam Brown and Tuomas Sandholm. Strategy-based warm starting for regret minimization in games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016.

Noam Brown and Tuomas Sandholm. Superhuman ai for heads-up no-limit poker: Libratus beats top professionals. *Science*, 359(6374):418–424, 2018.

Noam Brown and Tuomas Sandholm. Solving imperfect-information games via discounted regret minimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 1829–1836, 2019a.

Noam Brown and Tuomas Sandholm. Superhuman ai for multiplayer poker. *Science*, 365(6456): 885–890, 2019b.

Neil Burch, Matej Moravcik, and Martin Schmid. Revisiting cfr+ and alternating updates. *Journal of Artificial Intelligence Research*, 64:429–443, 2019.

Yang Cai, Gabriele Farina, Julien Grand-Clément, Christian Kroer, Chung-Wei Lee, Haipeng Luo, and Weiqiang Zheng. Last-iterate convergence properties of regret-matching algorithms in games. In *The Thirteenth International Conference on Learning Representations*, 2025.

Gabriele Farina, Christian Kroer, Noam Brown, and Tuomas Sandholm. Stable-predictive optimistic counterfactual regret minimization. In *International conference on machine learning*, pp. 1853–1862. PMLR, 2019a.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Optimistic regret minimization for extensive-form games via dilated distance-generating functions. *Advances in neural information processing systems*, 32, 2019b.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Faster game solving via predictive blackwell approachability: Connecting regret matching and mirror descent. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 5363–5371, 2021.

Gabriele Farina, Julien Grand-Clément, Christian Kroer, Chung-Wei Lee, and Haipeng Luo. Regret Matching+: (In)Stability and Fast Convergence in Games. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=nYgs0qZJ97.

Sam Ganzfried. Dominated actions in imperfect-information games. *CoRR*, abs/2504.09716, 2025.

Sergiu Hart and Andreu Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5):1127–1150, 2000.

Michael Johanson, Neil Burch, Richard Valenzano, and Michael Bowling. Evaluating state-space abstractions in extensive-form games. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, pp. 271–278, 2013.

Harold W. Kuhn. A simplified two-person poker. In Harold W. Kuhn and Albert W. Tucker (eds.), *Contributions to the Theory of Games, Volume I*, volume 24 of *Annals of Mathematics Studies*, pp. 97–103. Princeton University Press, 1950.

Marc Lanctot, Edward Lockhart, Jean-Baptiste Lespiau, Vinicius Zambaldi, Satyaki Upadhyay, Julien Pérolat, Sriram Srinivasan, Finbarr Timbers, Karl Tuyls, Shayegan Omidshafiei, Daniel Hennes, Dustin Morrill, Paul Muller, Timo Ewalds, Ryan Faulkner, János Kramár, Bart De Vylder, Brennan Saeta, James Bradbury, David Ding, Sebastian Borgeaud, Matthew Lai, Julian Schrittwieser, Thomas Anthony, Edward Hughes, Ivo Danihelka, and Jonah Ryan-Davis. Openspiel: A framework for reinforcement learning in games. *arXiv preprint arXiv:1908.09453*, 2019.

Weiming Liu, Huacong Jiang, Bin Li, and Houqiang Li. Equivalence analysis between counterfactual regret minimization and online mirror descent. In *International Conference on Machine Learning*, pp. 13717–13745. PMLR, 2022.

Linjian Meng, Youzhi Zhang, Zhenxing Ge, Tianyu Ding, Shangdong Yang, Zheng Xu, Wenbin Li, and Yang Gao. Last-iterate convergence of smooth regret matching$^+$ variants in learning nash equilibria. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=v771lscnlS.

Linjian Meng, Tianpei Yang, Youzhi Zhang, Zhenxing Ge, and Yang Gao. Faster game solving via asymmetry of step sizes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 17161–17169, 2026a.

Linjian Meng, Youzhi Zhang, Shangdong Yang, Wenbin Li, Tianyu Ding, and Yang Gao. A faster parameter-free regret matching algorithm. In *The Fourteenth International Conference on Learning Representations*, 2026b. URL https://openreview.net/forum?id=JLllvi7dsg.

Matej Moravčík, Martin Schmid, Neil Burch, Viliam Lisý, Dustin Morrill, Nolan Bard, Trevor Davis, Kevin Waugh, Michael Johanson, and Michael Bowling. Deepstack: Expert-level artificial intelligence in heads-up no-limit poker. *Science*, 356(6337):508–513, 2017.

John F Nash Jr. Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, 36(1):48–49, 1950.

Todd W Neller and Steven Hnath. Approximating optimal dudo play with fixed-strategy iteration counterfactual regret minimization. In *Advances in Computer Games*, pp. 170–183. Springer, 2011.

Sasha Rakhlin and Karthik Sridharan. Optimization, learning, and games with predictable sequences. *Advances in Neural Information Processing Systems*, 26, 2013.

Sheldon M Ross. Goofspiel—the game of pure strategy. *Journal of Applied Probability*, 8(3): 621–625, 1971.

Finnegan Southey, Michael Bowling, Bryce Larson, Carmelo Piccione, Neil Burch, Darse Billings, and Chris Rayner. Bayes' bluff: opponent modelling in poker. In *Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence*, pp. 550–558, 2005.

Oskari Tammelin. Solving large imperfect information games using cfr+. *arXiv preprint arXiv:1407.5042*, 2014.

Oskari Tammelin, Neil Burch, Michael Johanson, and Michael Bowling. Solving heads-up limit texas hold'em. In *IJCAI*, volume 15, pp. 645–652, 2015.

Hang Xu, Kai Li, Haobo Fu, Qiang Fu, and Junliang Xing. Autocfr: Learning to design counterfactual regret minimization algorithms. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 5244–5251, 2022.

Hang Xu, Kai Li, Haobo Fu, Qiang Fu, Junliang Xing, and Jian Cheng. Dynamic discounted counterfactual regret minimization. In *The Twelfth International Conference on Learning Representations*, 2024a.

Hang Xu, Kai Li, Bingyun Liu, Haobo Fu, Qiang Fu, Junliang Xing, and Jian Cheng. Minimizing weighted counterfactual regret with optimistic online mirror descent. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pp. 5272–5280, 2024b.

Brian Hu Zhang, Ioannis Anagnostides, and Tuomas Sandholm. Scale-invariant regret matching and online learning with optimal convergence: Bridging theory and practice in zero-sum games. *arXiv preprint arXiv:2510.04407*, 2025.

Naifeng Zhang, Stephen Marcus McAleer, and Tuomas Sandholm. Faster game solving via hyperparameter schedules. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 17319–17326, 2026.

Martin Zinkevich, Michael Johanson, Michael Bowling, and Carmelo Piccione. Regret minimization in games with incomplete information. *Advances in neural information processing systems*, 20, 2007.

A PROOFS

Proof of Lemma 1. Let $\bar{v} = \frac{1}{n_G+k}\big(\sum_{a\in G} v(a) + \sum_{b\in B} v(b)\big)$ be the uniform baseline. Using $v(b) \le \min_G v - g \le \bar{v}_G - g$ for $b \in B$,

$$\bar{v} \;\le\; \frac{n_G\bar{v}_G + k(\bar{v}_G - g)}{n_G + k} \;=\; \bar{v}_G - \frac{kg}{n_G+k}.$$

Hence $r_1(a) = v(a) - \bar{v} \ge v(a) - \bar{v}_G + kg/(n_G+k)$. The difference claim holds because $\bar{v}$ is common to all actions. □

Proof of Proposition 1. Write $q_a = Q_\tau(a)$, $q_b = Q_\tau(b)$. One update of a discounted RM$^+$ maps $Q_{t-1}(x)$ to $[d_{t-1}Q_{t-1}(x) + r_t(x)]^+$, so $Q_t(x) \le d_{t-1}Q_{t-1}(x) + \rho$ (if the argument of $[\cdot]^+$ is non-negative this is $|r_t| \le \rho$; otherwise $Q_t(x) = 0 \le d_{t-1}Q_{t-1}(x) + \rho$) and $Q_t(x) \ge d_{t-1}Q_{t-1}(x) + r_t(x) \ge d_{t-1}Q_{t-1}(x) - \rho$. Unrolling from τ to $\tau + s$ and using that every partial product of the d_j is at most one,

$$Q_{\tau+s}(a) \le P_s q_a + s\rho \le P_s(q_b + D) + s\rho, \qquad Q_{\tau+s}(b) \ge P_s q_b - s\rho.$$

RM$^+$ and DCFR play $\sigma_{\tau+s+1} \propto Q_{\tau+s}$, hence

$$\frac{\sigma_{\tau+s+1}(a)}{\sigma_{\tau+s+1}(b)} \le f(q_b), \qquad f(q) = \frac{P_s(q+D) + s\rho}{P_s q - s\rho}.$$

The function f is decreasing on $q > s\rho/P_s$, since $f'(q) = -P_s(P_sD + 2s\rho)/(P_sq - s\rho)^2 < 0$; with $q_b \ge J$ this gives $f(q_b) \le f(J)$, which is the claim. For PRM$^+$ and PDCFR$^+$ the strategy at $\tau+s+1$ is proportional to $[d_{\tau+s}Q_{\tau+s} + r_{\tau+s}]^+$ with the same scalar $d_{\tau+s} \le 1$ for both actions; for $s \ge 1$ the hypothesis bounds $|r_{\tau+s}| \le \rho$, so the numerator of a is at most $d_{\tau+s}(P_sq_a + s\rho) + \rho \le P_{s+1}q_a + (s+1)\rho$ and that of b at least $P_{s+1}q_b - (s+1)\rho$, and the argument repeats with $(s+1, P_{s+1})$. The condition for the ratio to reach κ follows by solving $f(J) \ge \kappa$ for s. Under CSRA(c), Q_τ is multiplied by c coordinate-wise, so (q_a, q_b) satisfy the hypotheses with (cJ, cD). □

Proposition 3 (the bound is tight in a stationary problem)**.** *Let RM$^+$ act at one information set with actions a, b and a set B of further actions, and let the counterfactual values be fixed: $v(a) = v(b) + \delta$ with $\delta > 0$ and $v(d) < v(b)$ for $d \in B$. Suppose that at iteration τ, $Q_\tau(d) = 0$ for all $d \in B$ and $Q_\tau(a), Q_\tau(b) \in [J, J+D]$. Let s_κ be the number of further iterations until $Q(a) \ge \kappa\, Q(b)$ (with $Q(b) = 0$ counting as an infinite ratio). Then*

$$\frac{(\kappa-1)J - D}{(\kappa+1)\delta} \;\le\; s_\kappa \;\le\; \left\lceil \frac{(\kappa-1)J + \kappa D}{\delta} \right\rceil.$$

Under CSRA(c) applied at τ the same bounds hold with (cJ, cD). In particular, for $D \le J/2$ and fixed κ, $s_\kappa = \Theta(J/\delta)$ without CSRA and $\Theta(cJ/\delta)$ with it, with the same constants.

Proof of Proposition 3. While $Q_t(b) > 0$ the strategy is supported on $\{a, b\}$ (the actions in B have zero accumulator and, as shown next, keep it), so $\langle \sigma_t, v\rangle \in [v(b), v(a)]$ and

$$r_t(a) = \sigma_t(b)\,\delta \in [0, \delta], \qquad r_t(b) = -\sigma_t(a)\,\delta \in [-\delta, 0], \qquad r_t(d) = v(d) - \langle \sigma_t, v\rangle < 0.$$

Hence $Q_t(d)$ stays at zero, $Q_t(b) \le Q_\tau(b) \le J + D$ is non-increasing, and $|r_t(a)|, |r_t(b)| \le \delta$, so Proposition 1 applies with $\rho = \delta$, $P_s = 1$, and gives the lower bound. For the upper bound note that, until $Q(b)$ is clipped to zero (at which point the ratio is infinite and s_κ has been reached), each update adds $r_t(a) - r_t(b) = \delta$ to $Q_t(a) - Q_t(b)$, so $Q_{\tau+s}(a) - Q_{\tau+s}(b) = Q_\tau(a) - Q_\tau(b) + s\delta \ge s\delta - D$. The ratio is at least κ as soon as $Q(a) - Q(b) \ge (\kappa-1)Q(b)$, for which $s\delta - D \ge (\kappa-1)(J+D)$ suffices because $Q(b) \le J + D$; this is the stated upper bound. CSRA(c) multiplies Q_τ by c and leaves δ unchanged, so both bounds hold with (cJ, cD). For $D \le J/2$ the lower bound is at least $(\kappa-1)J/(2(\kappa+1)\delta)$ and the upper bound at most $((\kappa-1) + \kappa/2)J/\delta + 1$. □

Proof of Proposition 2. For fixed $t \ge 2$, $d_t = t^\alpha/(t^\alpha + 1) \to 1$ as $\alpha \to \infty$, while $d_1 = \frac{1}{2}$ for every α. With $Q_0 = 0$, PDCFR$^+$ has $Q_1 = [r_1]^+$, $s_1 = [d_1Q_1 + r_1]^+ = [\frac{1}{2}[r_1]^+ + r_1]^+ = \frac{3}{2}[r_1]^+$ and $Q_2 = [d_1Q_1 + r_2]^+ = [\frac{1}{2}[r_1]^+ + r_2]^+$. CSRA$(\frac{1}{2})$-PCFR$^+$ is fed $\frac{1}{2}r_1$ at $t = 1$, so $Q_1 = \frac{1}{2}[r_1]^+$,

$s_1 = [Q_1 + \frac{1}{2}r_1]^+ = [\frac{1}{2}[r_1]^+ + \frac{1}{2}r_1]^+ = [r_1]^+$ (on $\{r_1 \geq 0\}$ the entry is r_1, on $\{r_1 < 0\}$ it is negative and clipped), which normalises to the same σ_2, and $Q_2 = [Q_1 + r_2]^+ = [\frac{1}{2}[r_1]^+ + r_2]^+$. For $t \geq 2$ the PDCFR$^+$ maps $Q_{t+1} = [d_t Q_t + r_{t+1}]^+$ and $s_t = [d_t Q_t + r_t]^+$ converge pointwise to the PCFR$^+$ maps $[Q_t + r_{t+1}]^+$ and $[Q_t + r_t]^+$, so the update rule of PDCFR$^+$ converges to that of CSRA($\frac{1}{2}$)-PCFR$^+$ with the same t^γ averaging. If an information set's first non-zero regret appears at $\tau \geq 2$ the limit applies the factor $d_\tau \to 1$, i.e. no reduction. The iterates over a fixed horizon T converge as well whenever the strategy numerators along the limit trajectory are non-zero at every information set, since each update is then continuous in (Q, d); at an all-zero numerator the normalisation is discontinuous and the two trajectories may differ. □

B Algorithms

All methods share one tree traversal per player and iteration (alternating updates), differ only in the local update, and use $Q_0 = 0$. With t the player's own iteration counter:

Method	accumulator Q_t	strategy $\sigma_{t+1} \propto$ / average weight
CFR$^+$	$[Q_{t-1} + r_t]^+$	Q_t / t
Linear CFR	$Q_{t-1} + t\, r_t$	$[Q_t]^+$ / t
DCFR(α, β, γ)	$Q_{t-1} \odot d_{t-1} + r_t$	$[Q_t]^+$ / t^γ
PCFR$^+(\gamma)$	$[Q_{t-1} + r_t]^+$	$[Q_t + r_t]^+$ / t^γ
PDCFR$^+(\alpha, \gamma)$	$[Q_{t-1}\frac{(t-1)^\alpha}{(t-1)^\alpha+1} + r_t]^+$	$[Q_t\frac{t^\alpha}{t^\alpha+1} + r_t]^+$ / t^γ
SAPCFR$^+(\alpha, \gamma)$	$[Q_{t-1} + r_t]^+$	$[Q_t + r_t/(\alpha+1)]^+$ / t^γ
APCFR$^+(\gamma, \bar{\alpha})$	$[Q_{t-1} + r_t]^+$	$[Q_t + r_t/(\alpha_t+1)]^+$ / t^γ

APCFR$^+$ uses the per-information-set adaptive $\alpha_t = \min\{\sqrt{S_t^{\mathrm{gap}}/S_t^{\Delta Q}}, \bar{\alpha}\}$ with $S_t^{\mathrm{gap}} = \sum_{s\leq t} \|r_{s-1} - r_s\|_2$ and $S_t^{\Delta Q} = \sum_{s\leq t} \|Q_s - Q_{s-1}\|_2$, exactly as in the authors' released code (Meng et al., 2026a), with $\gamma = 2$, $\bar{\alpha} = 5$; SAPCFR$^+$ uses $\alpha = 2$, $\gamma = 2$. CSRA rescales the vector fed to the update at $\tau(I)$, so both the accumulator and, for predictive variants, the prediction term see $c\, r_\tau$, for whichever base is used.

C Experimental details

C.1 Games

Table 3: Benchmark games: tree size, number of information sets (both players), number of (information set, action) pairs, and actions per information set.

Game	nodes	infosets	pairs	mean $\lvert A\rvert$	max $\lvert A\rvert$
Kuhn poker	58	12	24	2.00	2
Leduc poker	9,457	936	2,184	2.33	3
NL poker (stack 6)	10,027	912	2,118	2.32	5
NL poker (stack 10)	106,507	9,360	24,582	2.63	9
NL poker (stack 12, 6 ranks)	3,060,829	107,040	289,932	2.71	11
Liar's dice (4)	8,181	1,024	2,040	1.99	8
Liar's dice (5)	51,181	5,120	10,230	2.00	10
Liar's dice (6)	294,883	24,576	49,140	2.00	12
Goofspiel (4)	1,077	162	348	2.15	4
Goofspiel (5)	26,931	2,124	4,566	2.15	5
Goofspiel (6)	969,523	34,482	74,076	2.15	6
Battleship	732,607	81,027	327,068	4.04	7
Small Matrix	21	2	8	4.00	5

The no-limit poker games are OpenSpiel instances with two players, two betting rounds, blinds 1/2, one hole card, one board card in the second round, a two-suit deck with three ranks (six ranks for the

12-chip game), stacks of 6/10/12 chips and no betting abstraction, so every raise size up to the stack is a legal action. Battleship uses a 3×2 board, one ship of length 2 worth 2, three shots and no shot repeated on the same cell; Goofspiel is played with imperfect information and a descending prize order; Liar's dice uses one die per player with 4/5/6 sides. Small Matrix is the game of Appendix J of Xu et al. (2024a): rock–paper–scissors in which any outcome involving scissors is worth 2 where the others are worth 1, plus two row-player actions that lose 10,000 and 20,000 unconditionally; we play it as a turn-based game with the column player moving second without observing the row.

C.2 Dominated-action augmentation

Given a game tree, we append k additional actions to every decision node. Each additional action leads to a terminal node in which the acting player receives $-L$ and the opponent $+L$, with L twice the largest absolute payoff of the game, so the action is strictly dominated at every information set and the set of Nash equilibria of the original game is preserved (the new actions receive zero probability in every equilibrium). The augmentation is applied to the flattened tree, so the traversal cost grows only by the number of new terminal nodes.

C.3 Run manifest

Section 5.2: 4 games $\times$ $k \in \{0, 1, 2, 4, 8, 16\}$ (32 on three games) $\times$ 5 starts $\times$ 10 configurations (CFR$^+$, DCFR, PCFR$^+_{\gamma=5}$, PDCFR$^+$ and CSRA with $c \in \{0.1, 0.3\}$ on CFR$^+$, PCFR$^+_{\gamma=5}$, PDCFR$^+$), $T = 5000$. Section 5.3: 13 games $\times$ 5 starts $\times$ 18 configurations (the eight baselines, CSRA-PCFR$^+_{\gamma=5}$ with $c \in \{0.1, 0.2, 0.3, 0.5\}$, CSRA on CFR$^+$, PDCFR$^+$ and APCFR$^+$ with $c \in \{0.1, 0.3\}$); the six headline methods (PCFR$^+_{\gamma=5}$, PDCFR$^+$, APCFR$^+$ and their CSRA versions with $c = 0.1$) use 10 starts on Battleship and Small Matrix; the 12-chip no-limit poker game uses 5 starts and 8 configurations because of its size. Ablations (Appendix C.8): $T = 5000$, 5 starts. One iteration of the vectorised engine takes 0.2–1.3 s on the largest games and under 5 ms on Kuhn poker on a single CPU core; the 3,315 logged runs together took 198 core-hours of wall-clock time. CSRA adds no measurable per-iteration cost: the ratio of per-iteration time of CSRA-PCFR$^+_{\gamma=5}$ to PCFR$^+_{\gamma=5}$ over paired runs is 0.89–1.12 on Liar's dice (5) and 1.01–1.13 on Battleship, against a 1.08 and 1.18 spread between repeated runs of the base method alone; after an information set has started, the reduction is inactive there, so the only recurring work is a scan over the (I, a) vector per player per iteration. Every run of a game uses the same budget T_g (Table 4), so no start is censored earlier than its siblings; a cell whose median first-hit iteration is not identified within T_g is reported as $> T_g$ together with the number of starts that reached the threshold. Applying the same rule to the results of the earlier protocol, in which the budget varied across starts, rejects 105 of its cells; the sweep re-runs those cells at T_g.

Table 4: Iteration budget T_g per game. Every run of a game uses the same budget, so no start is censored earlier than its siblings. T_g is the initial budget, doubled where a headline method had not reached 10^{-6} within it (†), and fixed by hand where doubling would change no reading (‡).

Game	T_g
Kuhn poker	20,000
Leduc poker	20,000
NL poker (6)	20,000
NL poker (10)	50,000 ‡
NL poker (12, 6 ranks)	3,000 ‡
Liar's dice (4)	20,000
Liar's dice (5)	20,000
Liar's dice (6)	20,000 †
Goofspiel (4)	20,000
Goofspiel (5)	20,000
Goofspiel (6)	20,000 †
Battleship	10,000
Small Matrix	40,000 †

C.4 PERTURBATION ENSEMBLES

Seed 0 is the exact uniform start ($Q_0 = 0$). A perturbed start $k \geq 1$ keeps $Q_0 = 0$ and multiplies the first non-zero regret of every information set coordinate-wise by $1 + \varepsilon\,\xi_{I,a}$, with $\varepsilon = 10^{-9}$ and $\xi_{I,a} \sim U(0,1)$ drawn once from a generator seeded with k; every later update is unchanged. (Perturbing the accumulator instead, $Q_0 = \varepsilon\xi$, does not give a small perturbation: regret matching normalises Q_0, plays $\xi/\|\xi\|_1$ at $t = 1$, and the scale ε cancels.)

Lemma 3 (size of the perturbation)**.** *Let $r = r_\tau(I) \neq 0$ and $r^\varepsilon = r \odot (1 + \varepsilon\xi)$ with $\xi \in [0,1]^n$. For RM$^+$, PRM$^+$, their discounted variants and their CSRA(c) versions, feeding r^ε in place of r at τ gives $Q^\varepsilon_\tau = Q_\tau \odot (1 + \varepsilon\xi)$ and $\|\sigma^\varepsilon_{\tau+1}(I) - \sigma_{\tau+1}(I)\|_1 \leq 2\varepsilon$.*

Proof. Multiplying by $1 + \varepsilon\xi_a \in [1, 1 + \varepsilon]$ preserves the sign of every coordinate, so $[r^\varepsilon]^+ = [r]^+ \odot (1 + \varepsilon\xi)$, which is the accumulator claim (the factor c of CSRA and the zero discount of $Q_{\tau-1} = 0$ commute with it). For each of the algorithms the strategy numerator at $\tau + 1$ is a positive scalar times a vector $p = [\,\lambda_1[r]^+ + \lambda_2 r\,]^+$ with $\lambda_1, \lambda_2 \geq 0$ (Lemma 2), and the perturbed numerator is $p \odot (1 + \varepsilon\xi) =: p^\varepsilon$ by the same sign argument; thus $\|p^\varepsilon - p\|_1 \leq \varepsilon\|p\|_1$. For non-zero $x, y \geq 0$,

$$\Big\|\frac{x}{\|x\|_1} - \frac{y}{\|y\|_1}\Big\|_1 \leq \frac{\|x - y\|_1}{\|x\|_1} + \|y\|_1\Big|\frac{1}{\|x\|_1} - \frac{1}{\|y\|_1}\Big| \leq \frac{2\|x - y\|_1}{\|x\|_1},$$

and with $x = p$, $y = p^\varepsilon$ this is at most 2ε. □

Table 5 lists the spread of exploitability across the starts at fixed iterations for PCFR$^+_{\gamma=5}$. Up to a thousand iterations no game separates by more than a factor of 2.87; the separation then concentrates in Battleship and Goofspiel (5), and at $T = 5000$ every other game stays within 8.51×.

Table 5: Spread of exploitability across the perturbed starts (min–max, PCFR$^+_{\gamma=5}$; ten on Battleship and Small Matrix, five elsewhere) at fixed iterations.

Game	T=100	T=1000	T=5000
Kuhn poker	5.9e-08–5.9e-08 (1×)	1.0e-12–1.0e-12 (1×)	1.0e-12–1.0e-12 (1×)
Leduc poker	1.8e-02–1.8e-02 (1×)	6.3e-04–6.5e-04 (1×)	2.8e-05–7.2e-05 (3×)
NL poker (6)	7.3e-08–7.3e-08 (1×)	1.0e-12–1.0e-12 (1×)	1.0e-12–1.0e-12 (1×)
NL poker (10)	7.5e-03–8.0e-03 (1×)	4.7e-04–7.0e-04 (1×)	3.9e-06–3.4e-05 (9×)
NL poker (12, 6 ranks)	6.6e-02–6.9e-02 (1×)	3.7e-03–4.8e-03 (1×)	n/a
Liar's dice (4)	5.1e-04–5.1e-04 (1×)	1.6e-10–1.6e-10 (1×)	1.0e-12–1.0e-12 (1×)
Liar's dice (5)	6.1e-03–6.1e-03 (1×)	9.4e-05–9.4e-05 (1×)	9.3e-06–9.3e-06 (1×)
Liar's dice (6)	6.9e-03–6.9e-03 (1×)	7.0e-04–7.0e-04 (1×)	7.8e-05–7.9e-05 (1×)
Goofspiel (4)	1.1e-05–1.1e-05 (1×)	4.8e-12–4.8e-12 (1×)	1.0e-12–1.0e-12 (1×)
Goofspiel (5)	7.3e-03–1.0e-02 (1×)	5.5e-06–1.6e-05 (3×)	1.1e-09–2.3e-07 (205×)
Goofspiel (6)	1.7e-02–2.0e-02 (1×)	4.7e-04–5.6e-04 (1×)	1.6e-05–2.4e-05 (2×)
Battleship	6.8e-03–7.7e-03 (1×)	1.6e-05–2.3e-05 (1×)	1.0e-11–4.1e-09 (399×)
Small Matrix	6.6e-01–6.6e-01 (1×)	3.1e-01–3.1e-01 (1×)	3.3e-04–3.3e-04 (1×)

C.5 PRE-REGISTERED PREDICTIONS FOR THE HEADLINE EXPERIMENTS

We recorded four predictions before the runs of Tables 1 and 5 were submitted, on the games, thresholds and budgets fixed in advance. Each prediction is stated as it was written, followed by the reading.

- **P1, Battleship.** The median first-hit iteration of CSRA-PCFR$^+$ to 10^{-6} is at most $0.6\times$ that of PCFR$^+_{\gamma=5}$. Measured over ten starts that all reach the threshold: 720 against 1750, a ratio of 0.41. Holds.
- **P2, Small Matrix.** At most $0.3\times$. Measured: 1600 against 15000, a ratio of 0.11. Holds.
- **P3, mechanism.** The median of PCFR$^+_{\gamma=5}$ at $k = 32$ dominated actions is at least $3\times$ its median at $k = 0$ on at least two of Kuhn poker, Leduc poker and Liar's dice (4), and

the median of CSRA-PCFR$^+$ at $k = 32$ is at most $1.5\times$ its median at $k = 0$. Measured on Kuhn poker and Liar's dice (4): $5.40\times$ and $1.59\times$ for PCFR$^+_{\gamma=5}$, $1.07\times$ and $0.79\times$ for CSRA-PCFR$^+$. On Leduc poker no start of either method reaches the threshold within the budget, so the ratio cannot be formed, and the requirement of two games is met on one. Holds for CSRA-PCFR$^+$, fails for PCFR$^+_{\gamma=5}$.

- **P4, spread across starts.** At $T = 1000$ iterations the spread of exploitability across starts, measured as max/min, is at least 10 on Battleship or Small Matrix. Measured: 1.49 and 1.00. Fails; the spreads we observe are those of Table 5.

C.6 Inflation diagnostics

For every information set that receives a non-zero regret at its first update, split the actions into $G^+ = \{a : r_\tau(a) > 0\}$ and its complement. Since the actions outside G^+ have counterfactual value at most the uniform baseline and those in G^+ strictly above it, Lemma 1 applies to this partition, and its offset equals the mean of the positive regrets, $\bar{v}_{G^+} - \bar{v} = \frac{1}{|G^+|}\sum_{a\in G^+} r_\tau(a)$. Under regret matching$^+$ the actions outside G^+ receive zero accumulator and zero probability at $\tau + 1$, so the inertia of Proposition 1 concerns the actions of G^+ alone; there $J = \min_{a\in G^+} r_\tau(a)$ is the largest common lower bound on Q_τ (a lower bound on the offset above) and $S = \max_{G^+} r_\tau - J$ is the spread D. We record, for the information sets with $|G^+| \geq 2$, the *inflation share* $J/(J + S)$; it is the ratio $J/(J + D)$ that enters the proposition, and is 1 when the positive regrets are equal. Table 6 reports the fraction of information sets with at least two positive first regrets and the median inflation share, with and without appended dominated actions.

Table 6: First-update diagnostics. "≥ 2 pos.": fraction of information sets whose first regret has at least two positive entries; "share": median inflation share among them.

	original		$k = 2$		$k = 8$	
Game	≥ 2 pos.	share	≥ 2 pos.	share	≥ 2 pos.	share
Kuhn poker	0.00	–	1.00	0.65	1.00	0.80
Leduc poker	0.17	0.40	1.00	0.42	1.00	0.58
NL poker (6)	0.11	0.19	1.00	0.54	1.00	0.70
NL poker (10)	0.25	0.31	1.00	0.32	1.00	0.47
Liar's dice (4)	0.32	0.30	0.42	0.36	0.50	0.41
Liar's dice (5)	0.27	0.36	0.41	0.42	0.50	0.47
Liar's dice (6)	0.24	0.28	0.38	0.48	0.50	0.49
Goofspiel (4)	0.16	1.00	1.00	1.00	1.00	1.00
Goofspiel (5)	0.15	0.54	1.00	1.00	1.00	1.00
Battleship	0.64	1.00	1.00	0.60	1.00	0.79
Small Matrix	0.50	1.00	–	–	–	–

C.7 Benchmark results at a coarser threshold

C.8 Ablations

Table 8 reports the sensitivity to c on the benchmark games and Table 9 the ablations with $k = 8$ dominated actions per decision node, where the mechanism is active. On Kuhn poker the first-hit iteration decreases monotonically from $c = 1$ to $c = 0.1$ and stays near its minimum down to $c = 0.02$; on Liar's dice it is flat for every $c \leq 0.5$; on Leduc poker $c \in [0.1, 0.3]$ is best. Reducing the first K iterations in place of the first regret is worse for $K = 20$ on all three games; $K = 5$ is worse on Kuhn poker and Liar's dice and better on Leduc poker. The per-information-set factors $c_I = \|[r_\tau - \langle\sigma_\lambda, r_\tau\rangle\mathbf{1}]^+\|_1/\|[r_\tau]^+\|_1$ of Section 4.2, with $\lambda = 1$ and σ_λ the strategy induced by r_τ, match the constant on Kuhn poker (where $c_I \approx 0.06$ on average) but not on Liar's dice, and $\lambda = 0.5$ is a mild reduction ($c_I \approx 0.5$). Re-baselining the first regret against the induced strategy (subtracting $\langle\sigma_\lambda, r_\tau\rangle$ in place of scaling) is not better than scaling. Truncating PDCFR$^+$'s discount after 1–10 iterations changes its first-hit iteration by less than the seed range.

Table 7: Repeats Table 1 at threshold 10^{-4}, at which the non-predictive methods are resolved. Median first-hit iteration to exploitability $\leq 10^{-4}$ (ten starts on Battleship and Small Matrix for the headline methods, five elsewhere). "$> T\,(j/n)$": the threshold was not reached within the common budget T (Table 4) by $n - j$ of the n starts, so the starts pin no median. Bold: fastest identified median.

Method	Kuhn poker	Leduc poker	NLP (6)	NLP (10)	NLP (12, 6 ranks)	LD (4)	LD (5)	LD (6)	GS (4)	GS (5)	GS (6)	Battleship	Small Matrix
CFR$^+$	**360**	1800	750	1100	>3k (0/5)	620	660	1200	660	4400	4300	>10k (0/5)	>40k (0/5)
DCFR	1400	1400	**250**	780	>3k (0/5)	280	260	920	1200	2300	3300	>10k (0/5)	8700
PCFR$^+$	**62**	4100	79	3200	>3k (0/5)	250	1300	2100	110	790	1800	510	28000
PCFR$^+_{\gamma=5}$	**28**	3800	45	2400	>3k (0/5)	150	990	4400	65	550	2200	610	6400
PDCFR$^+$	32	2900	**26**	2600	>3k (0/5)	150	790	5800	59	610	2600	255	2400
SAPCFR$^+$	**47**	1800	90	3200	>3k (0/5)	170	500	1200	110	510	1600	890	28000
APCFR$^+$	**48**	1900	76	2700	>3k (0/5)	200	340	1400	93	410	1500	700	28000
CSRA-CFR$^+$	680	2000	**390**	1500	>3k (0/5)	860	860	1500	970	3400	4000	>10k (0/5)	23000
CSRA-PCFR$^+$	**31**	2000	37	1900	>3k (0/5)	150	1200	6300	50	570	2400	280	700
CSRA-PDCFR$^+$	**35**	3000	49	2600	>3k (0/5)	140	1100	2000	61	590	1900	260	250
CSRA-APCFR$^+$	**47**	1900	94	2800	>3k (0/5)	220	1100	1300	130	370	1300	335	2800

Table 8: Sensitivity to c: median first-hit iteration to 10^{-6} of CSRA-PCFR$^+_{\gamma=5}$ over the perturbed starts. "n/a": the configuration was not run. Not listed, because no value of c reaches the threshold within its budget and the columns would not be comparable: NL poker (12, 6 ranks), Liar's dice (6).

Game	c=1 (PCFR$^+$)	c=0.05	c=0.1	c=0.2	c=0.3	c=0.5
Kuhn poker	63	70	71	69	62	64
Leduc poker	14000	13000	13000	14000	12000	12000
NL poker (6)	77	110	79	56	53	54
NL poker (10)	29000	30000	22000	50000	39000	40000
Liar's dice (4)	270	270	280	270	270	280
Liar's dice (5)	9900	7500	2500	20000	7200	9600
Goofspiel (4)	140	130	110	110	140	140
Goofspiel (5)	1900	1500	3800	2800	1900	3100
Goofspiel (6)	14000	n/a	>20k (2/5)	17000	16000	13000
Battleship	1750	660	720	710	680	840
Small Matrix	15000	760	1600	3100	4500	7500

C.9 Pre-registered additional games

Before running them we selected two games outside the benchmark set of Xu et al. (2024b) that exercise the mechanism of Section 4.1, computed the first-update diagnostics of Appendix C.6 from their trees alone, and recorded a prediction for each: a game whose first regrets have many positive entries with a large inflation share should show CSRA-PCFR$^+$ reaching the threshold in at most half the iterations of PCFR$^+_{\gamma=5}$ and in a seed range disjoint from it, while a game with a small share should show overlapping ranges. The registered candidates were a 3×3 Battleship with three shots, Liar's dice with two three-sided dice, and Goofspiel (7); the first (38.4M nodes) and the third (47.5M nodes) exceed the 5M-node budget we set, so the Battleship was run in its registered fallback form with two shots (832k nodes) and the Goofspiel candidate was dropped, Goofspiel (4–6) already covering that family. Table 10 reports the diagnostics, the predictions and the measured first-hit iterations. On the 3×3 Battleship CSRA-PCFR$^+$ reaches the threshold in a median less than half that of PCFR$^+_{\gamma=5}$ and of PDCFR$^+$, as predicted for a game whose first regrets have an inflation share of 1.00, but one PCFR$^+_{\gamma=5}$ start reaches the threshold at 170 iterations, so the registered disjointness of the two seed ranges does not hold. On Liar's dice with two three-sided dice, whose inflation share is 0.35, the three medians have overlapping ranges, as predicted; here CFR$^+$ and DCFR do not reach the threshold within the budget.

Table 9: Ablations with $k = 8$ dominated actions per decision node: median first-hit iteration [min, max] over five perturbed starts, thresholds 10^{-6} (Kuhn, Liar's dice) and 10^{-4} (Leduc), budget 5000 for every row. A start that had not reached the threshold within the shared budget is reported as ">5k", the number of starts that reached it being in Table 1's notation.

Variant	Kuhn poker	Leduc poker	Liar's dice (4)
PCFR$^+_{\gamma=5}$ (no reduction)	240 [240, 240]	>5k (0/5) [–, >5k]	280 [280, 280]
CSRA, $c = 0.02$	72 [72, 72]	3000 [3000, 5000]	230 [230, 230]
CSRA, $c = 0.05$	67 [67, 67]	4400 [4300, 4400]	230 [230, 230]
CSRA, $c = 0.1$	70 [70, 70]	3700 [3700, 3700]	260 [260, 260]
CSRA, $c = 0.2$	83 [83, 83]	3000 [2800, 3000]	250 [250, 250]
CSRA, $c = 0.3$	110 [110, 110]	3400 [2900, 3600]	230 [230, 230]
CSRA, $c = 0.5$	200 [200, 200]	3500 [3500, 3500]	230 [230, 230]
CSRA, $c = 1$	240 [240, 240]	>5k (0/5) [–, >5k]	280 [280, 280]
reduce first 5 iterations, $c = 0.1$	78 [78, 78]	3000 [3000, 3100]	330 [330, 330]
reduce first 20 iterations, $c = 0.1$	150 [150, 150]	3000 [3000, 3100]	320 [320, 320]
per-infoset factor, $\lambda = 0.5$	200 [200, 200]	3500 [3500, 3500]	230 [230, 230]
per-infoset factor, $\lambda = 1$	67 [67, 67]	3600 [3500, 3700]	280 [280, 280]
re-baselined regret, $\lambda = 0.5$	150 [150, 150]	>5k (0/5) [–, >5k]	210 [210, 210]
re-baselined regret, $\lambda = 1$	82 [82, 82]	>5k (0/5) [–, >5k]	280 [270, 280]
PDCFR$^+$	110 [110, 110]	3100 [3100, 3200]	240 [240, 240]
PDCFR$^+$, discount only at $t \leq 1$	200 [200, 200]	3500 [3500, 3500]	230 [230, 230]
PDCFR$^+$, discount only at $t \leq 3$	170 [170, 170]	3500 [2900, 3500]	230 [230, 230]
PDCFR$^+$, discount only at $t \leq 10$	110 [110, 110]	3200 [2800, 3200]	250 [250, 250]

Table 10: Pre-registered additional games: first-update diagnostics (as in Table 6) and median first-hit iteration to 10^{-6} [min, max] over five perturbed starts, budget 10000 iterations (>10k: as in Table 1). The predictions were written down before the runs (Appendix C.9).

	diagnostics		prediction	PCFR$^+_{\gamma=5}$		PDCFR$^+$		CSRA-PCFR$^+$	
Game	≥ 2 pos.	share		med	[min, max]	med	[min, max]	med	[min, max]
Battleship (3×3, 2 shots)	1.00	1.00	CSRA $\leq$ PCFR$^+$/2	440	[170, 750]	900	[540, 990]	510	[290, 910]
Liar's dice (2 dice, 3 sides)	0.31	0.35	overlap	660	[660, 1400]	1400	[890, 1400]	680	[670, 860]