# Faster Game Solving via Hyperparameter Schedules

Naifeng Zhang[1], Stephen McAleer[2], Tuomas Sandholm[1345]

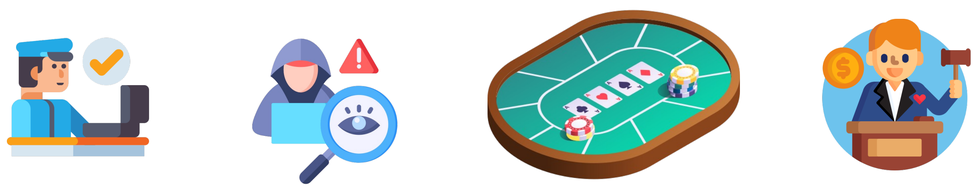
[1]Carnegie Mellon University, [2]Anthropic, [3]Strategy Robot, Inc., [4]Strategic Machine, Inc., [5]Optimized Markets, Inc.

Carnegie Mellon University

## I. MOTIVATION & BACKGROUND

### 1. Many real-world settings are games

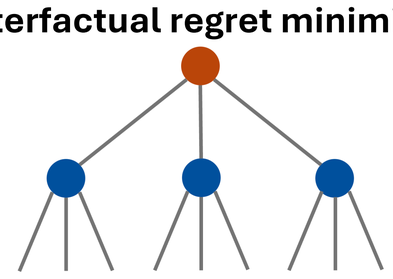
Specifically, **imperfect-information games (IIGs)**



…to deceive and to understand deception

### 2. Our focus: solving two-player zero-sum IIGs

via **Counterfactual regret minimization (CFR)** algorithms



"How much better I could have done?"

Iteratively reducing regret to guide the average strategy toward a Nash equilibrium

### 3. Evolution of CFR variants

**CFR** (Zinkevich et al. 2007) —discounting average strategy→ **CFR+** (Tammelin 2014) —$(\alpha, \beta, \gamma)$: discounting average strategy & cumulative regrets→ **Discounted CFR (DCFR)** (Brown and Sandholm 2019) —dynamic discounting schemes via RL→ **Dynamic DCFR (DDCFR)** (Xu et al. 2024)

$\alpha = 1.5, \beta = 0, \gamma = 2$

$\alpha = f(t), \beta = g(t), \gamma = h(t)$

CFR+ → **Predictive CFR+ (PCFR+)** (Farina, Kroer, and Sandholm 2021) → **Dynamic PCFR+ (DPCFR+)**

$\gamma$: discounting average strategy & predictive regret matching

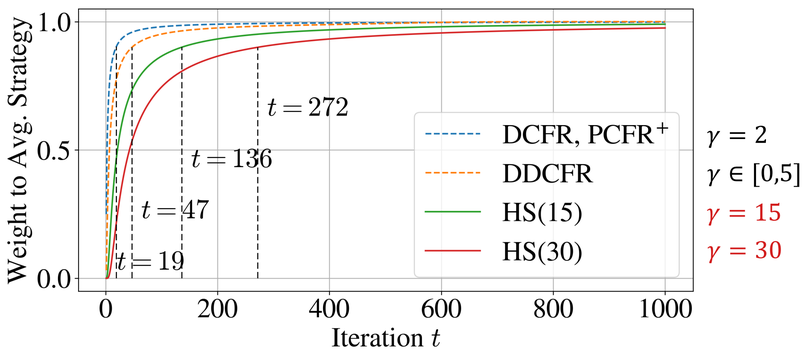
**Training:** 24 hours on 200 CPU cores
**Inference:** game-specific, multiprocessing required

***Q: SIMPLE, EFFECTIVE & TRAINING-FREE METHOD?***

## II. APPROACH

### 1. Hyperparameter Schedules (HSs)

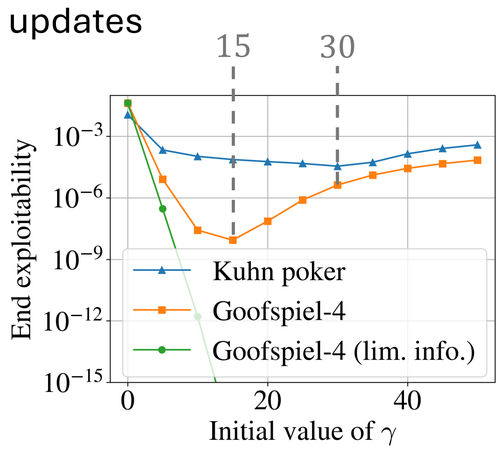
Discounting schemes that control how hyperparameter changes




Prior CFR variants' discounting schemes are **not sufficiently aggressive**

### 2. Identifying effective HSs

that **aggressively downweight** unrefined strategies from early updates




### 3. HS-powered algorithms

On iteration $t+1$, **DCFR** multiplies

- positive cumulative regrets by $\frac{t^\alpha}{t^\alpha+1}$
- negative cumulative regrets by $\frac{t^\beta}{t^\beta+1}$
- contributions to the average strategy by $\left(\frac{t}{t+1}\right)^\gamma$

**Proposed HSs:**

$$\text{HS}_\alpha : \alpha = 1 + \frac{3}{n}t, \qquad \text{HS}_\beta : \beta = -1 - \frac{2}{n}t,$$

$$\text{HS}_{\gamma 30} : \gamma_{30} = 30 - \frac{5}{n}t, \quad \text{HS}_{\gamma 15} : \gamma_{15} = 15 - \frac{5}{n}t.$$

$n$ is the total number of iterations

### 4. HS-powered algorithms, cont.

**HS-powered DCFR (HS-DCFR)**
$(HS_\alpha, HS_\beta, HS_\gamma)$

**HS-powered PCFR+ (HS-PCFR+)**
$HS_\gamma$

$$\frac{t^\alpha}{t^\alpha+1} \to \frac{t^{HS_\alpha}}{t^{HS_\alpha}+1}$$

$$\frac{t^\beta}{t^\beta+1} \to \frac{t^{HS_\beta}}{t^{HS_\beta}+1}$$

$$\left(\frac{t}{t+1}\right)^\gamma \to \left(\frac{t}{t+1}\right)^{HS_\gamma}$$

### 5. Theoretical guarantees

**Provable convergence** to a Nash equilibrium in two-player zero-sum games

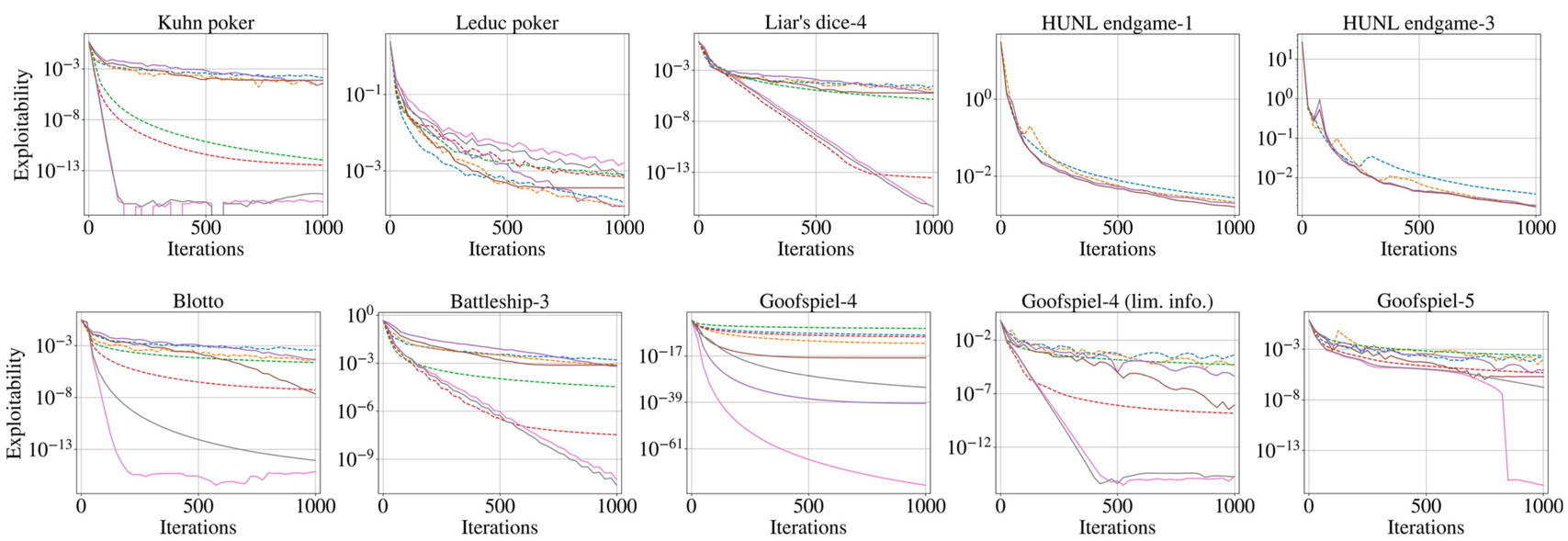
**Theorem 3.1.** *Suppose $T$ iterations of HS-DCFR, with simultaneous updates, are played in a two-player zero-sum game, and $U$ is the upper bound of $\gamma$ across all iterations. If $\alpha \in [1,5]$, $\beta \in [-5,0]$, and $\gamma \in [0,U]$, the weighted average strategy profile is a $(U+1)\Delta|\mathcal{I}|\left(\frac{8}{3}\sqrt{|\mathcal{A}|} + \frac{2}{\sqrt{T}}\right)/\sqrt{T}$-Nash equilibrium.*

**Theorem 3.2.** *Suppose $T$ iterations of HS-PCFR$^+$, with simultaneous updates, are played in a two-player zero-sum game, and $U$ is the upper bound of $\gamma$ across all iterations. If $\gamma \in [0,U]$, the weighted average strategy profile is a $(U+1)|\mathcal{I}|O(1)/\sqrt{T}$-Nash equilibrium.*

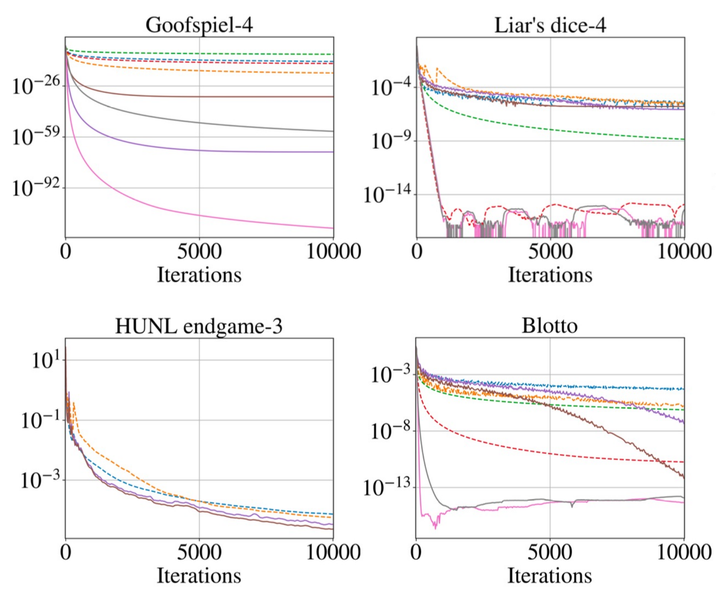
***< 15 LINES OF CODE CHANGES!***

## III. RESULTS

### 1. State of the art on ten diverse games




### 2. Results with extended iterations




DCFR DDCFR PCFR$^+$ DPCFR$^+$ HS-DCFR(30) HS-DCFR(15) HS-PCFR$^+$(30) HS-PCFR$^+$(15)

***ORDERS-OF-MAGNITUDE IMPROVEMENT OVER PRIOR SOTA***

Reach us at naifengz@cmu.edu, mcaleer.stephen@gmail.com, sandholm@cs.cmu.edu