POLITECNICO
MILANO 1863

Regret-based Traces-Exploration Abstractions for Large Game Solving

Jacopo Pio Gargano

Advisor Prof. Nicola Gatti
Co-Advisor Prof. Barbara Caputo

Recreational Games

IBM

1996

MiniMax with alpha-beta pruning search

Google DeepMind

2016

Monte Carlo tree search

Deep neural networks

Reinforcement Learning

Real-world Strategic Scenarios

Security

Poaching

Sports

Military

Game Theory

Theoretical framework for strategic interaction

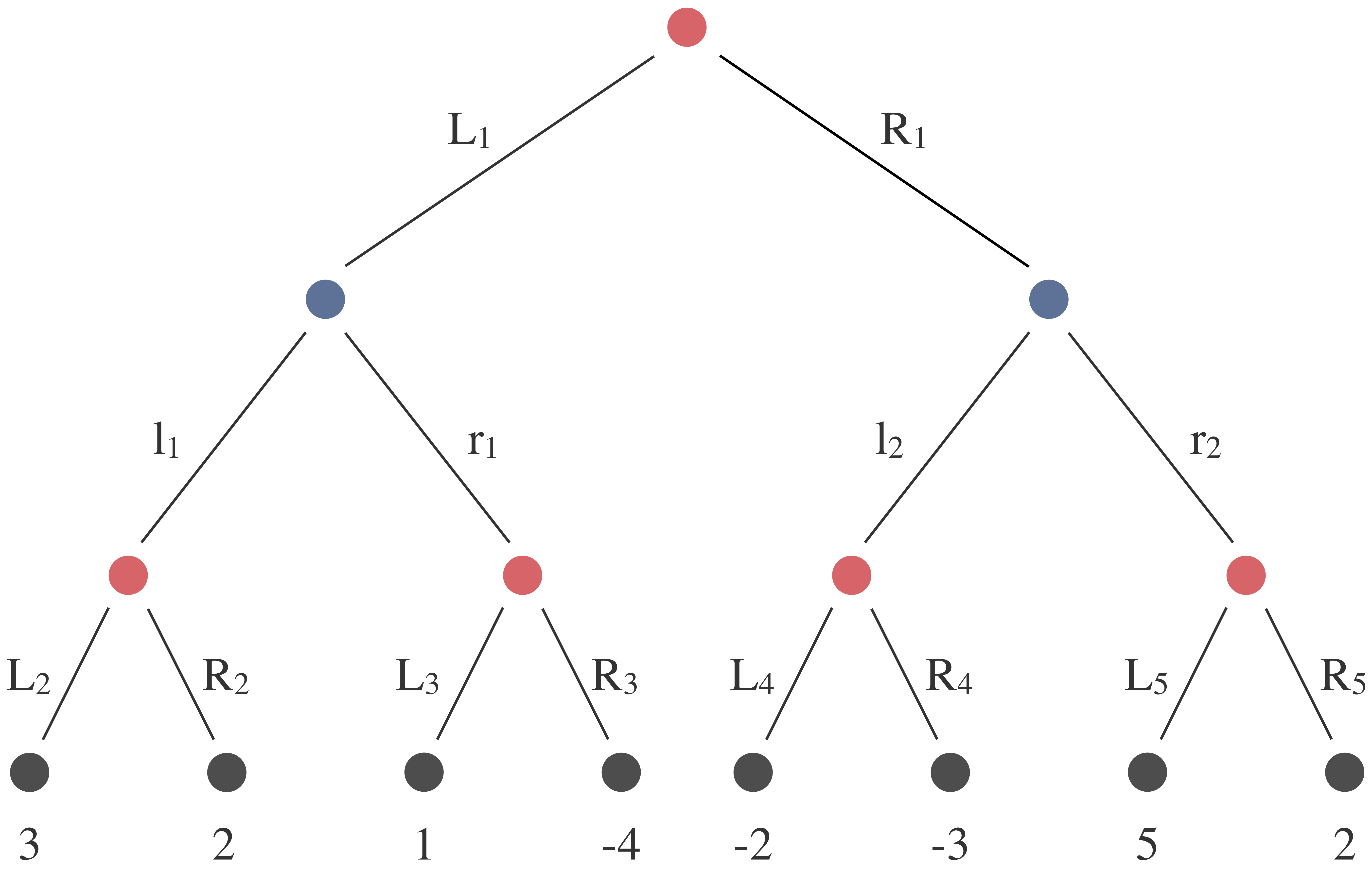
Mathematical models and algorithms (Algorithmic Game Theory)

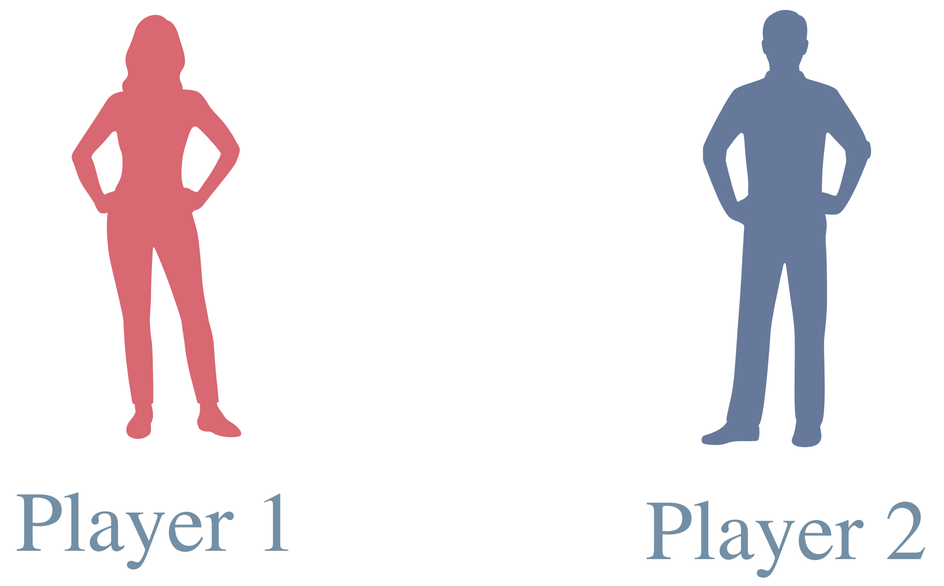
Conflict and cooperation

Intelligent rational decision-makers

Decisions influencing agents' welfare

Sequential Games Representation

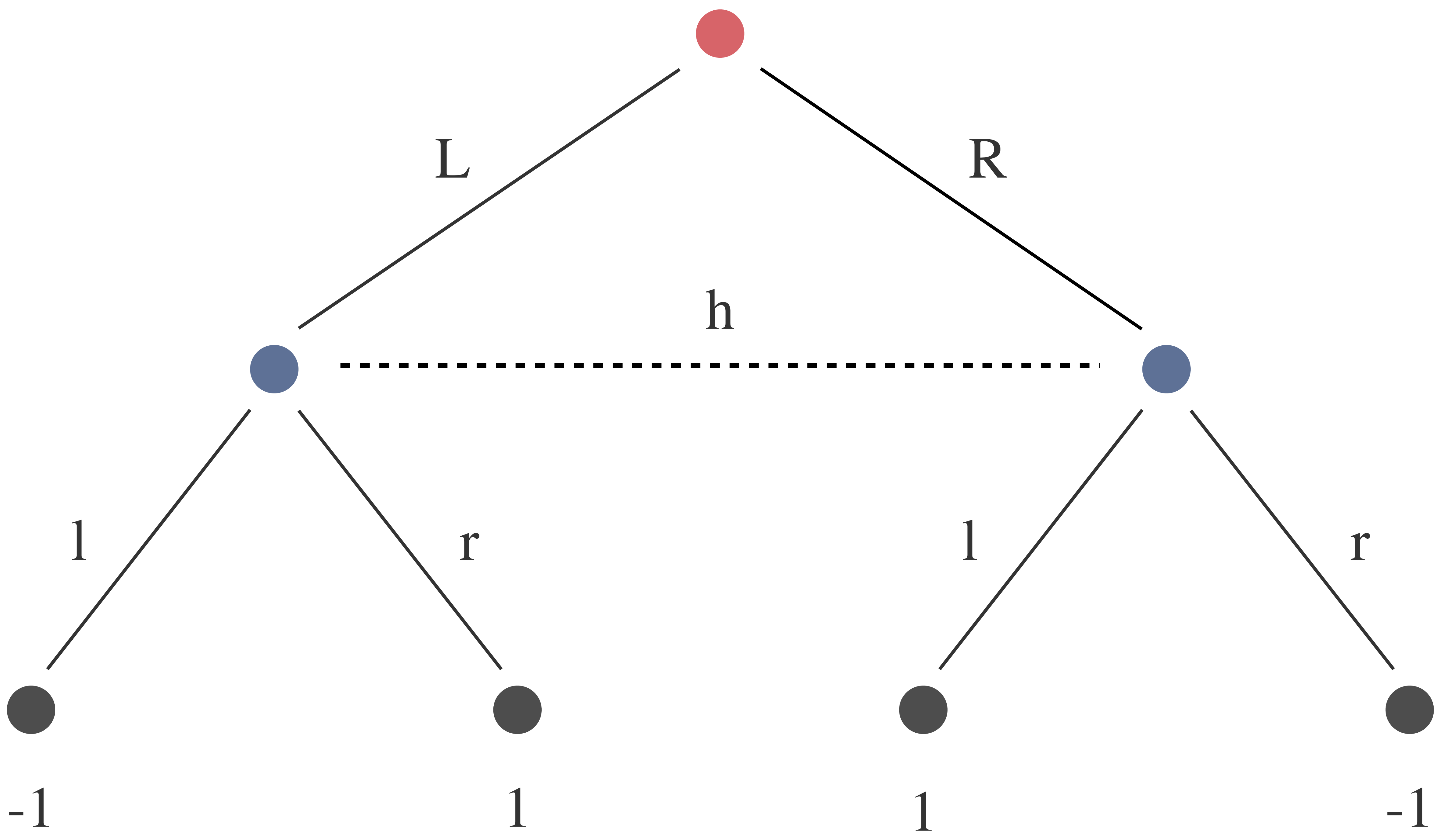




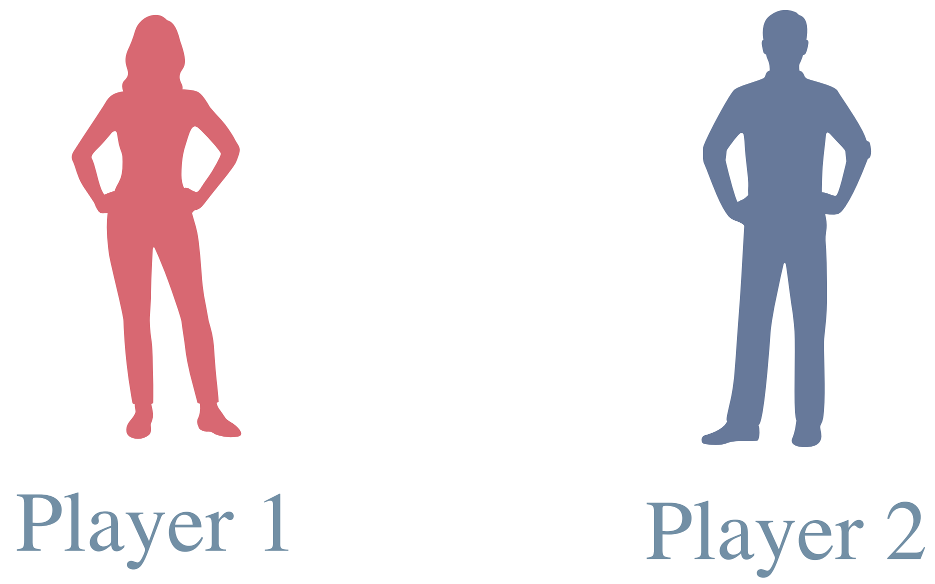
$$\sigma_i : H_i \to \Delta^{|A_{H_i}|}$$

Strategy

Sequential Games Representation



Imperfect-information extensive-form games



$$\sigma_i : H_i \rightarrow \Delta^{|A_{H_i}|}$$

Strategy

Nash Equilibrium

Definition

A Nash Equilibrium (NE) is a joint combination of strategies, stable w.r.t. unilateral deviations of a single player.

Theorem

Every n-player finite game has at least one NE profile in mixed strategies.

Approximations

An ε-approximate Nash Equilibrium (ε-NE) approximately satisfies the condition of NE.

Regret Minimization

Counterfactual Value

$$v_i^\sigma(h) = \sum_{z \in Z} \pi_{-i}^\sigma(h) \cdot \pi^\sigma(h, z) \cdot u_i(z)$$

Regret

$$r_i(h, a) = v_i^\sigma(h, a) - v_i^\sigma(h)$$

Regret Minimization

$$\text{if } \frac{R_i^T}{T} \leq \varepsilon \text{ then } 2\varepsilon\text{-NE}$$

Regret Matching (CFR)

$$\sigma_i^{T+1}(h, a) = \begin{cases} \frac{R_i^{T,+}(h,a)}{\sum_{a \in A_h} R_i^{T,+}(h,a)} & \text{if } \sum_{a \in A_h} R_i^{T,+}(h, a) > 0 \\ \frac{1}{|A_h|} & \text{otherwise} \end{cases}$$

Counterfactual Regret Minimization

Algorithm 1 CFR

Input the history *history*, the traverser player i, CFR iteration t, reach probabilities π_i, chance reach π_c

Output counterfactual value $v_i^\sigma(h)$

function CFR($history$, i, t, π_1, π_2, π_c):

  $h \leftarrow$ get information set associated to *history*

  if h is terminal **then**

    return $u_i(h)$

  else if h is chance **then**

    $H' \leftarrow \theta(h, RA)$ ▷ RA being the random action

    for all $h' \in H'$ **do**

      $history' \leftarrow history + info(h')$ ▷ *info()* returns the public info

      $v_i^\sigma(h) \leftarrow v_i^\sigma(h) + \mathrm{CFR}(history', i, t, \pi_1, \pi_2, \frac{\pi_c}{|H'|})$

    return $mean(v_i^\sigma(h))$

  else

    $v_i^\sigma(h) \leftarrow 0$

    $v_i^\sigma\big(\theta(h,a)\big) \leftarrow 0$ for all $a \in A_h$

    for all $a \in A_h$ **do**

      if $\rho(h) = 1$ **then**

        $v_i^\sigma\big(\theta(h,a)\big) \leftarrow \mathrm{CFR}(history + a, i, t, \sigma^t(h,a) \cdot \pi_1, \pi_2, \pi_c)$

      else

        $v_i^\sigma\big(\theta(h,a)\big) \leftarrow \mathrm{CFR}(history + a, i, t, \pi_1, \sigma^t(h,a) \cdot \pi_2, \pi_c)$

      $v_i^\sigma(h) \leftarrow v_i^\sigma(h) + \sigma^t(h,a) \cdot v_i^\sigma\big(\theta(h,a)\big)$

    if $\rho(h) = i$ **then**

      for all $a \in A_h$ **do**

        $r_i(h,a) \leftarrow r_i(h,a) + \pi_c \cdot \pi_{-i} \cdot \big(v_i^\sigma(\theta(h,a)) - v_i^\sigma(h)\big)$

        $s_i(h,a) \leftarrow s_i(h,a) + \pi_i \cdot \sigma^t(h,a)$

      $\sigma^{t+1}(h) \leftarrow$ regret-matching values

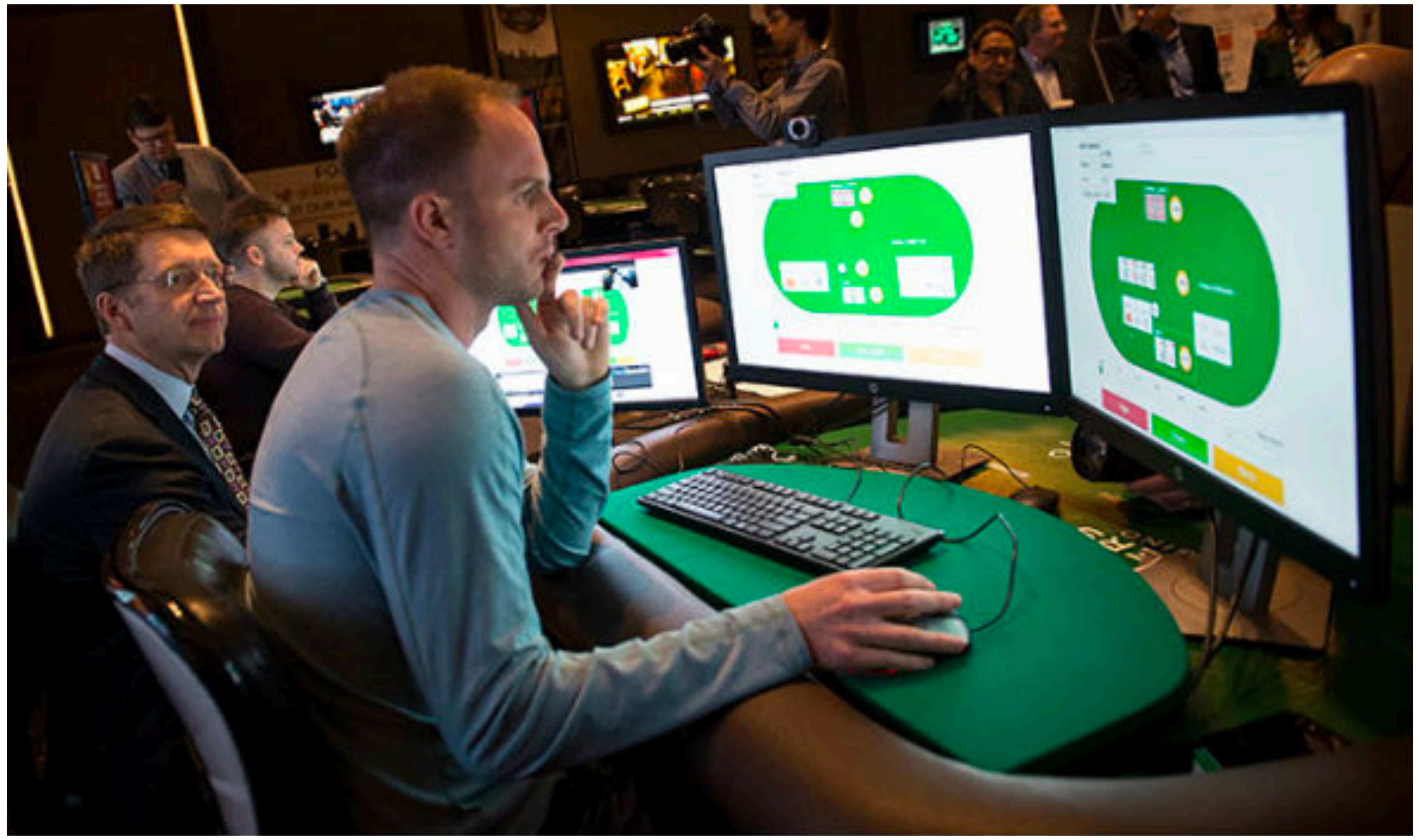
  return $v_i^\sigma(h)$

Abstractions

Abstractions are a smaller version of the game capturing the most essential properties of the real domain, reducing complexity.

The solution of the abstracted game provides a useful approximation of the optimal strategy.

- Information abstractions
- Action abstractions
- Simulation-based abstractions



Libratus, 2017

The Problem

Complexity

Real-world games and strategic scenarios are too large to be represented and analyzed.

Domain-independence

No clear domain-independent abstraction approach was presented to solve large games with many actions.

Legacy

Most works are based on classical abstraction techniques not leveraging new powerful learning approaches.

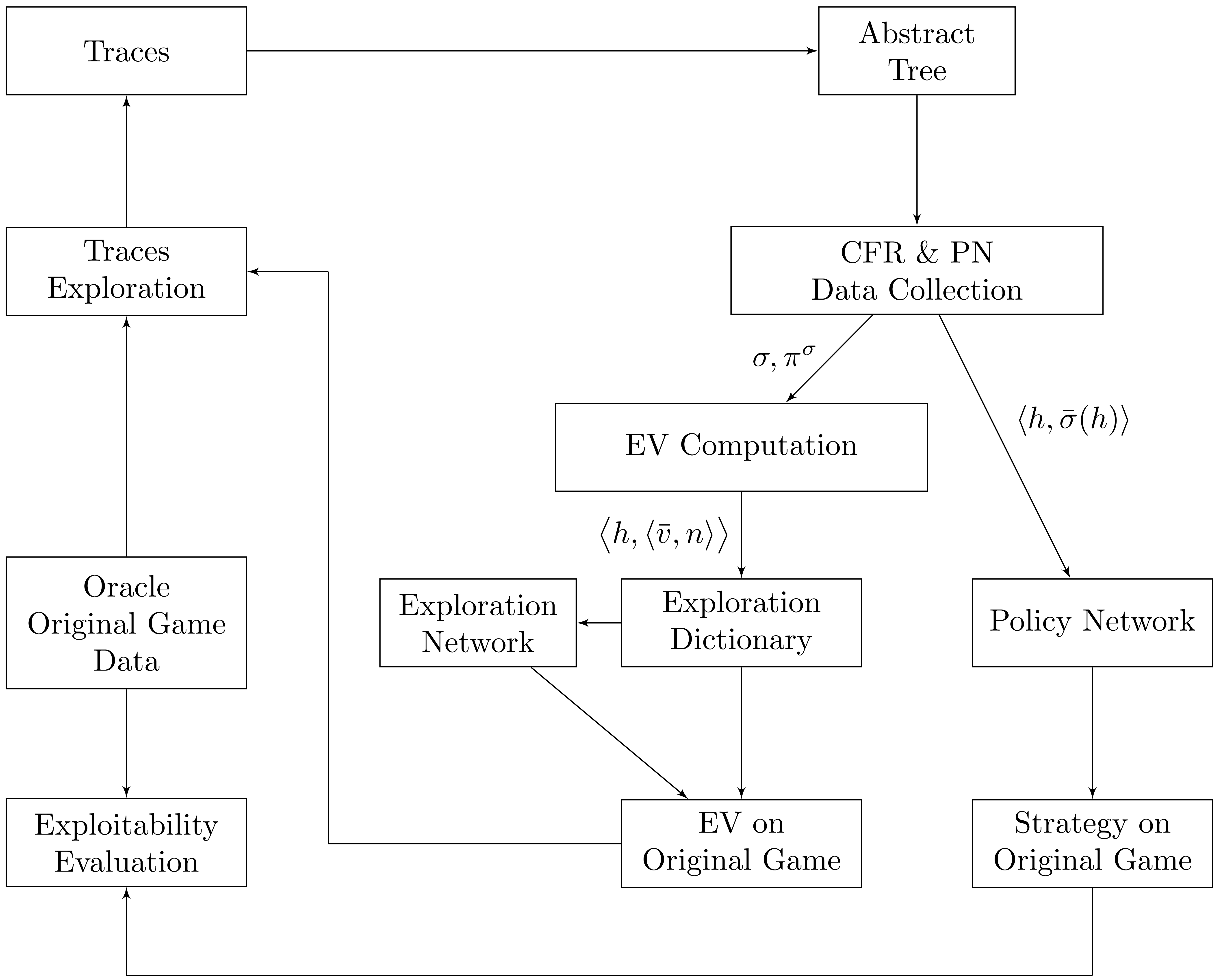
ReTrE - Regret-based Traces-Exploration Counterfactual Regret Minimization

Regret-based Traces-Exploration Counterfactual Regret Minimization (ReTrE) is a domain-independent model-free abstraction framework, able to find approximate mixed strategy Nash Equilibria in any extensive-form game in a simulation-based fashion.

ReTrE leverages deep neural networks and confidence-based exploration techniques to approximate the behavior of CFR in the full game.

ReTrE is a scalable pre-play iterative algorithm, focusing on the most exploitable parts of the game to obtain competitive suboptimal strategies.

RETRE - Overview

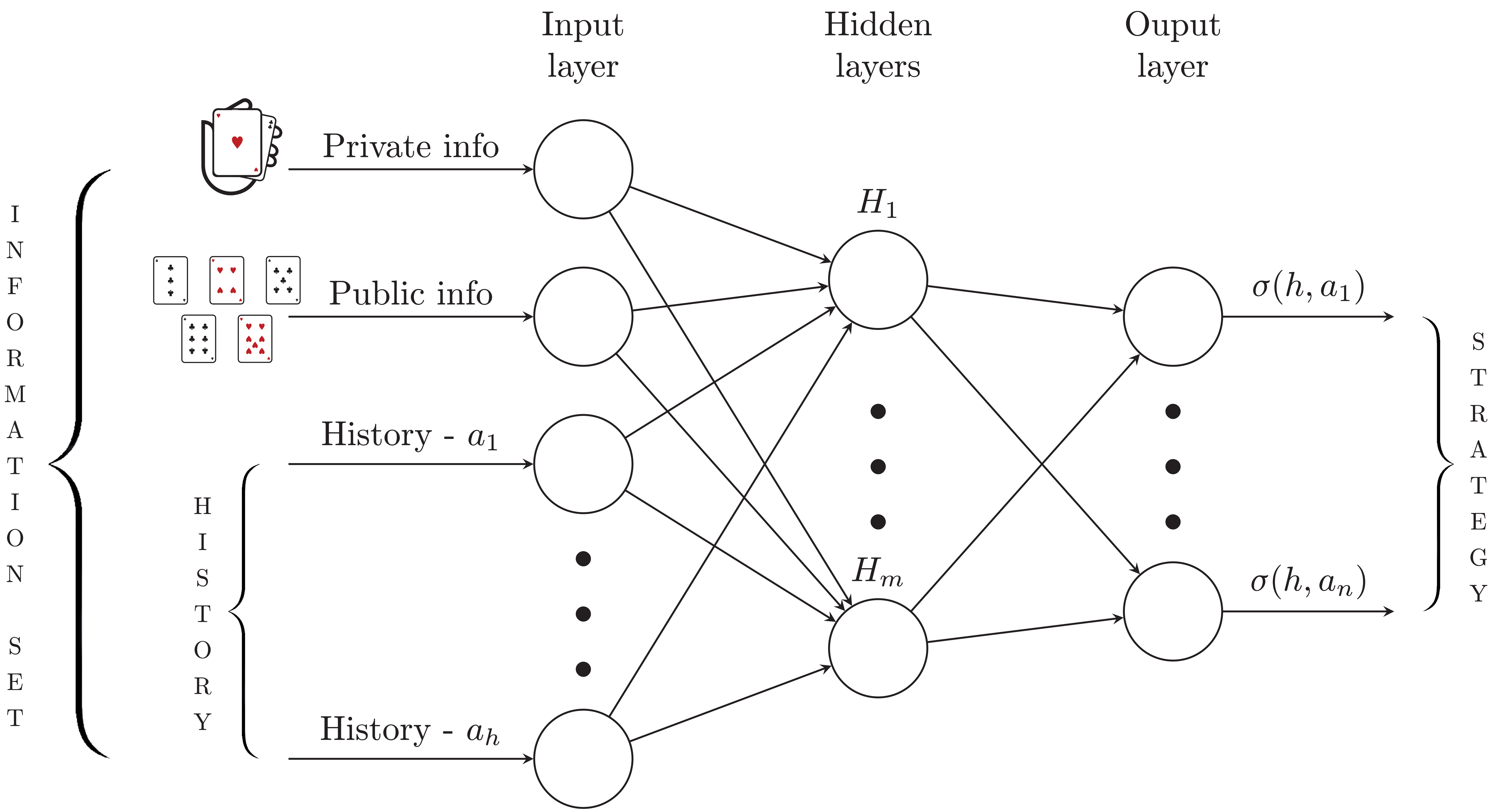


ReTrE - The Algorithm

Algorithm 1 ReTrE

function ReTrE(k, N, T, RT):
    Initialize PN, ED, EN for each player $i \in N$
    for ReTrE iteration $rt = 1$ to RT **do**
        $traces \leftarrow$ GetTracesUCB(ED, EN, rt, k)
        for CFR iteration $t = 1$ to T **do**
            for all $i \in N$ **do**
                CFR($[\,], i, t, rt, 1, 1, 1$)
        ComputeEV($rt, \sigma, \pi^{\sigma}, ED, EN$)
        Train EN through ED
    Train PN

ReTrE - Policy Network



ReTrE - Exploration

Exploration Value (EV)

$$\hat{v}_i(h,a) = \frac{\sum_{t=1}^{T}\left(t \cdot \pi_{-i}^{\sigma,t}(h) \cdot \left(v_i^{\sigma,t}(\theta(h,a)) - v_i^{\sigma,t}(h)\right)\right)}{\sum_{t=1}^{T}\left(t \cdot \pi_{-i}^{\sigma,t}(h)\right)}$$

Exploration Dictionary (ED)

$\langle h, \langle \bar{v}, n\rangle\rangle$, where $\bar{v}$ is the average EV

n are the visits throughout ReTrE

Exploration Parameter (EP)

k, maximum number of children when exploring

Upper Confidence Bound

$$ucb = \bar{v} + \sqrt{\frac{2\log(t)}{n}}$$

Exploration Network (EN)

estimate $\bar{v}$ through $EN : H \rightarrow \mathbb{R}$

$$n = \min_{h' \in ED} n_{h'}$$

ReTrE - Exploration

Algorithm 1 ReTrE - EV Computation and Collection

function ComputeEV(rt, π^σ, ED, EN):
  for all information sets $h \in H$ **do**
    $i \leftarrow \rho(h)$
    for all $a \in A_h$ **do**
      $h' \leftarrow \theta(h, a)$
      $\hat{v}_i \leftarrow$ compute EV
      if $rt = 1$ **then**
        $\bar{v} \leftarrow \hat{v}_i$
        $n \leftarrow 1$
      else
        if $h' \in ED_i$ **then**
          $\bar{v}, n \leftarrow$ get $\langle \bar{v}, n \rangle$ of h' from ED_i
        else
          $\bar{v} \leftarrow$ predict $\bar{v}(h')$ through EN_i
          $n \leftarrow$ estimate n
        $\bar{v} \leftarrow \frac{\bar{v} \cdot n + \hat{v}_i}{n+1}$
        $n \leftarrow n + 1$
      Store $\langle h', \langle \bar{v}, n \rangle \rangle$

ReTrE - Exploitability Evaluation

Best Response (BR)

$$\text{if } u_i(BR(\sigma_{-i}), \sigma_{-i}) = \max_{\sigma_i' \in \Sigma_i} u_i(\sigma_i', \sigma_{-i})$$

Exploitability

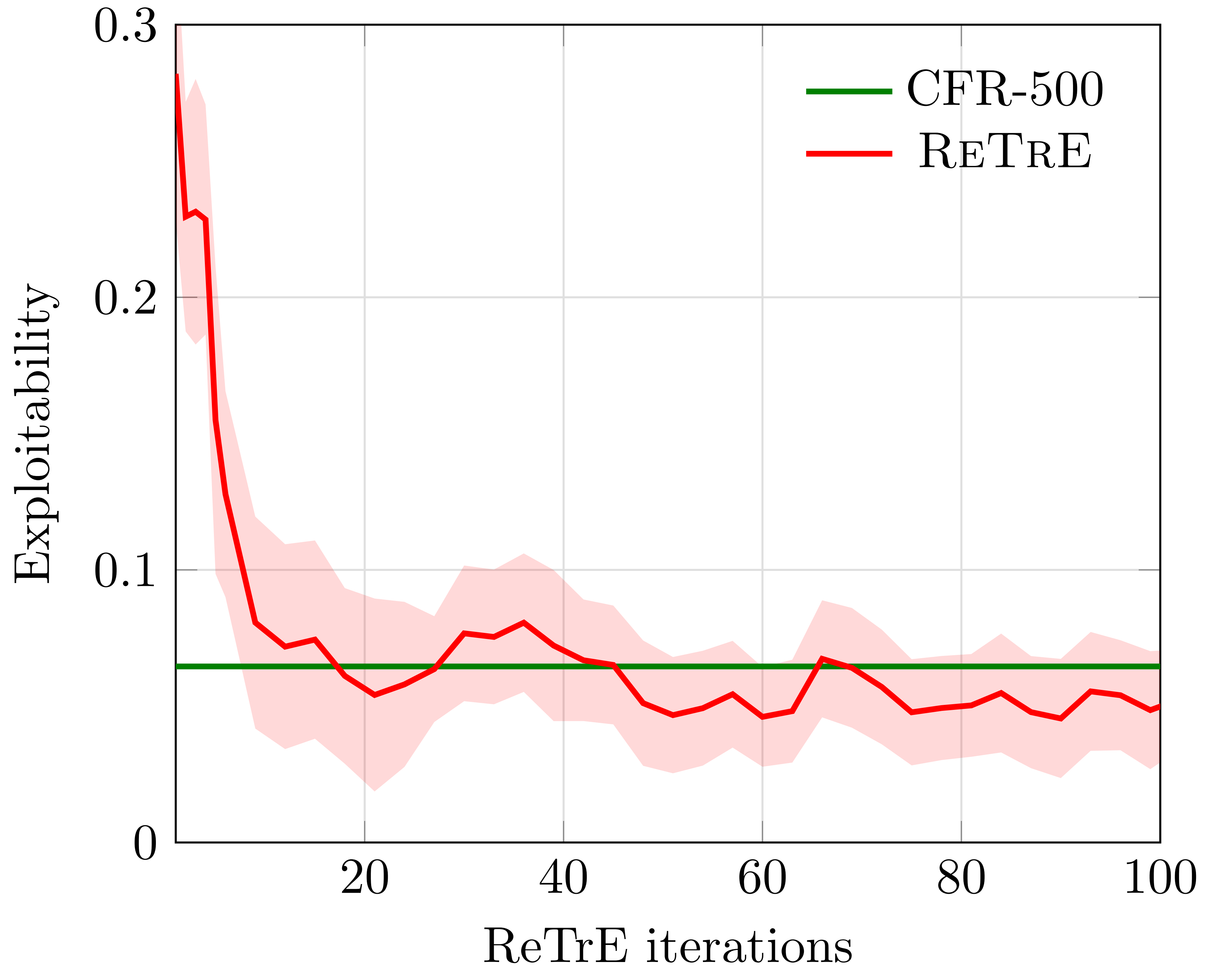
$$e(\sigma_i) = u_i\big(\sigma_i^*, BR(\sigma_i^*)\big) - u_i\big(\sigma_i, BR(\sigma_i)\big)$$

NashConv

$$\textsc{NashConv}(\sigma) = \sum_{i \in N} \max_{\sigma_i' \in \Sigma_i} u_i(\sigma_i', \sigma_{-i})$$

RETRE - Exploitability Evaluation

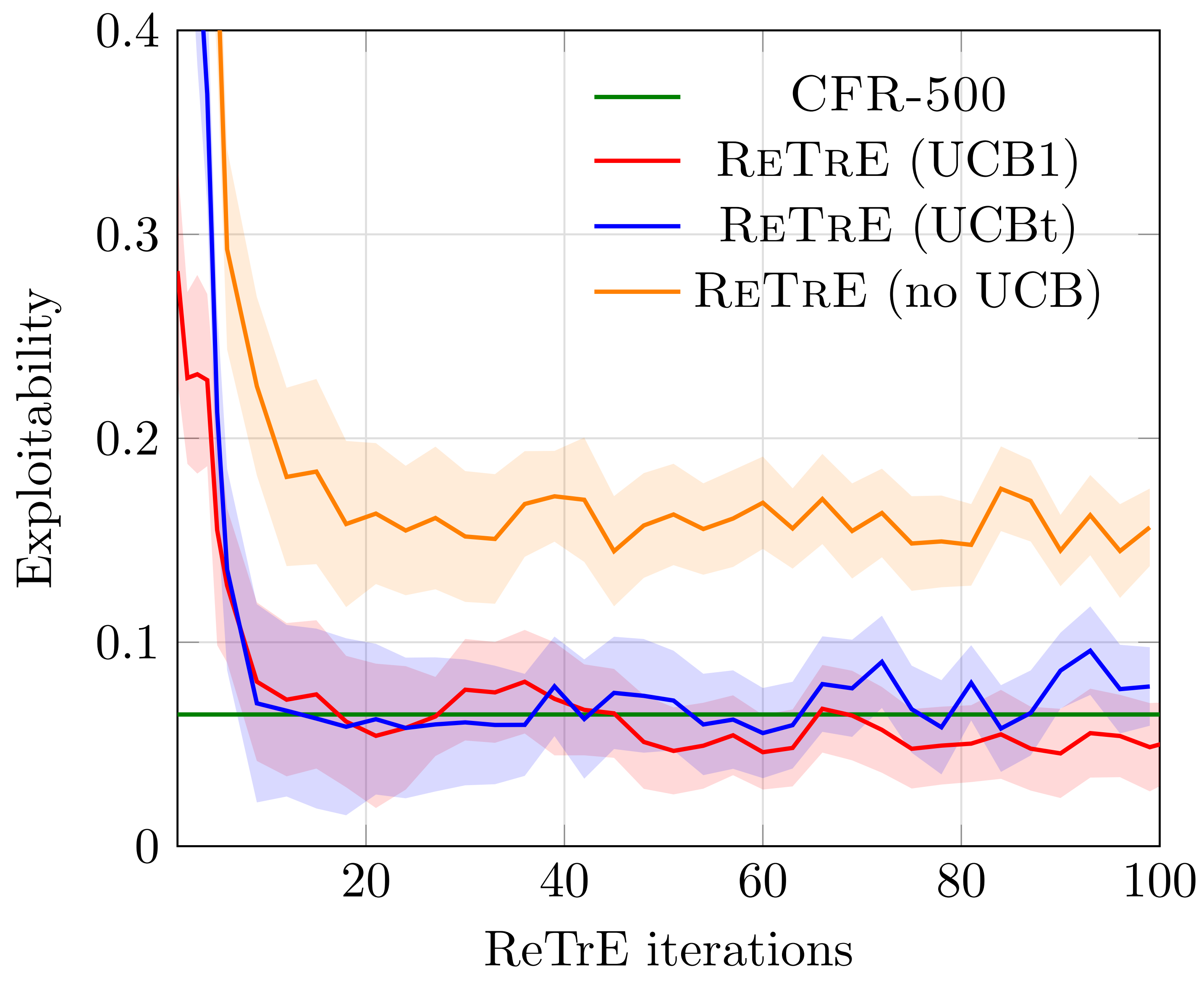
RETRE shows lower exploitability, after an exploration phase, compared to CFR.



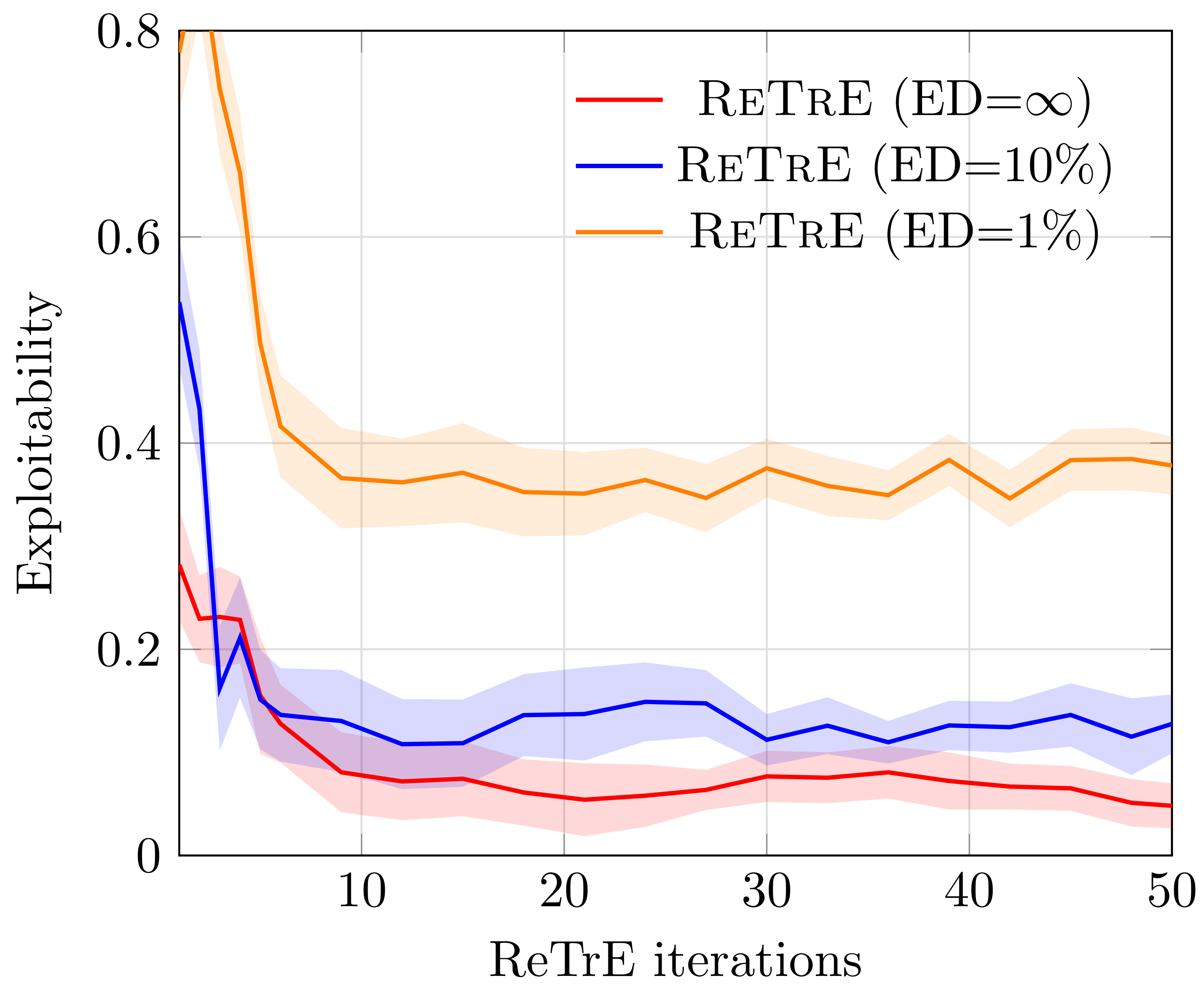
CFR-500 and RETRE

ReTrE - Exploitability Evaluation

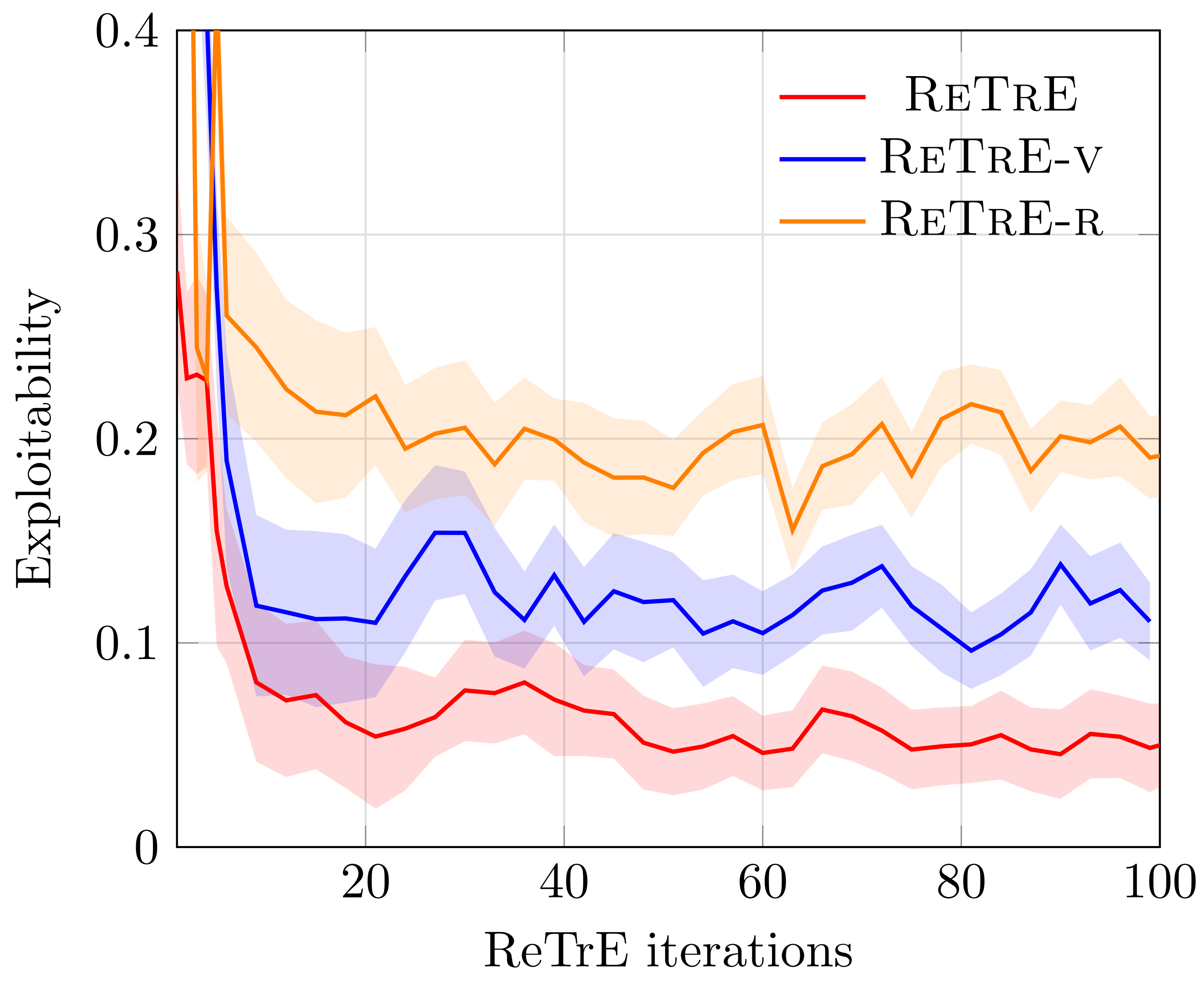
We compare different configurations to examine ReTrE's behavior.



Different UCBs



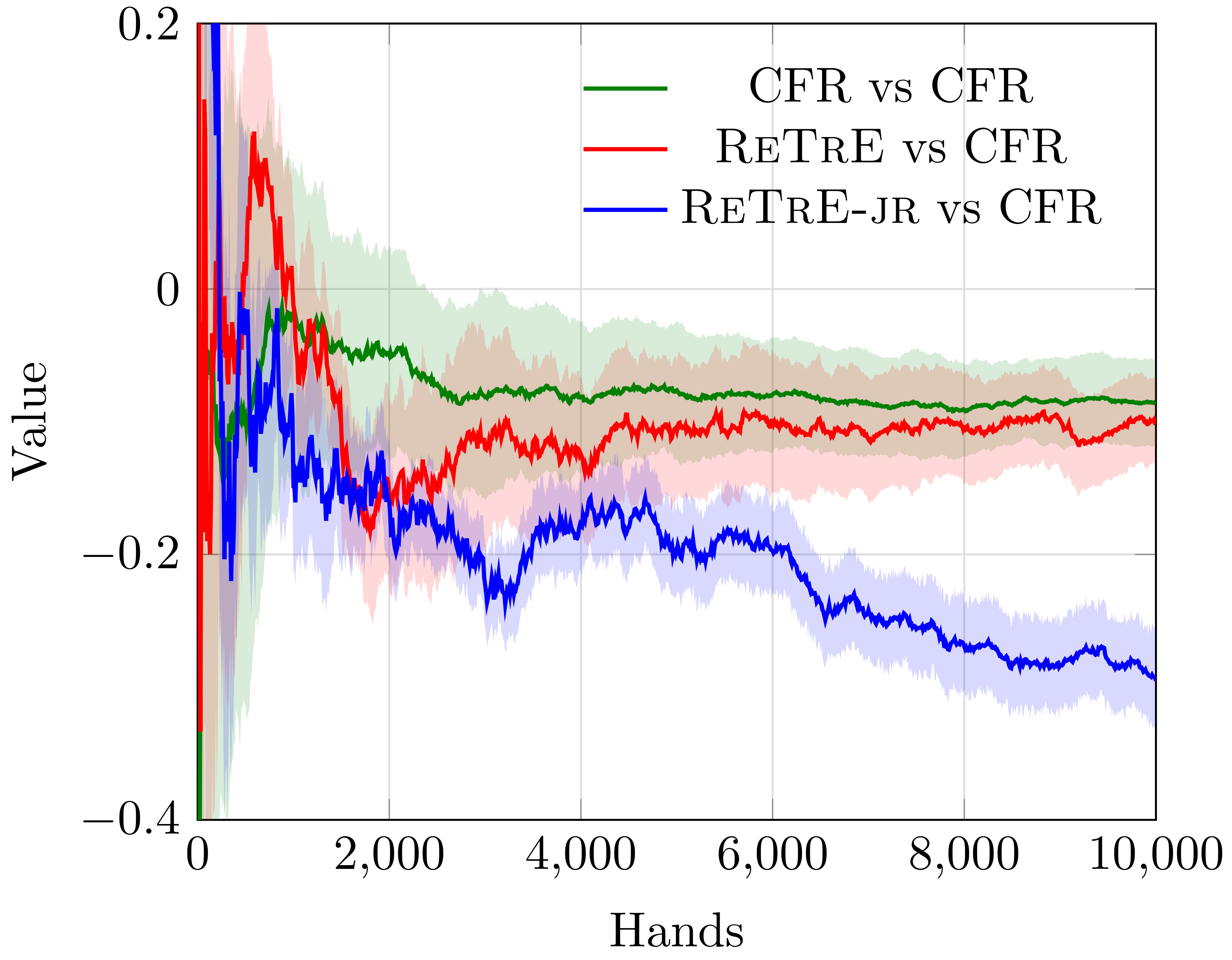
Different ED sizes



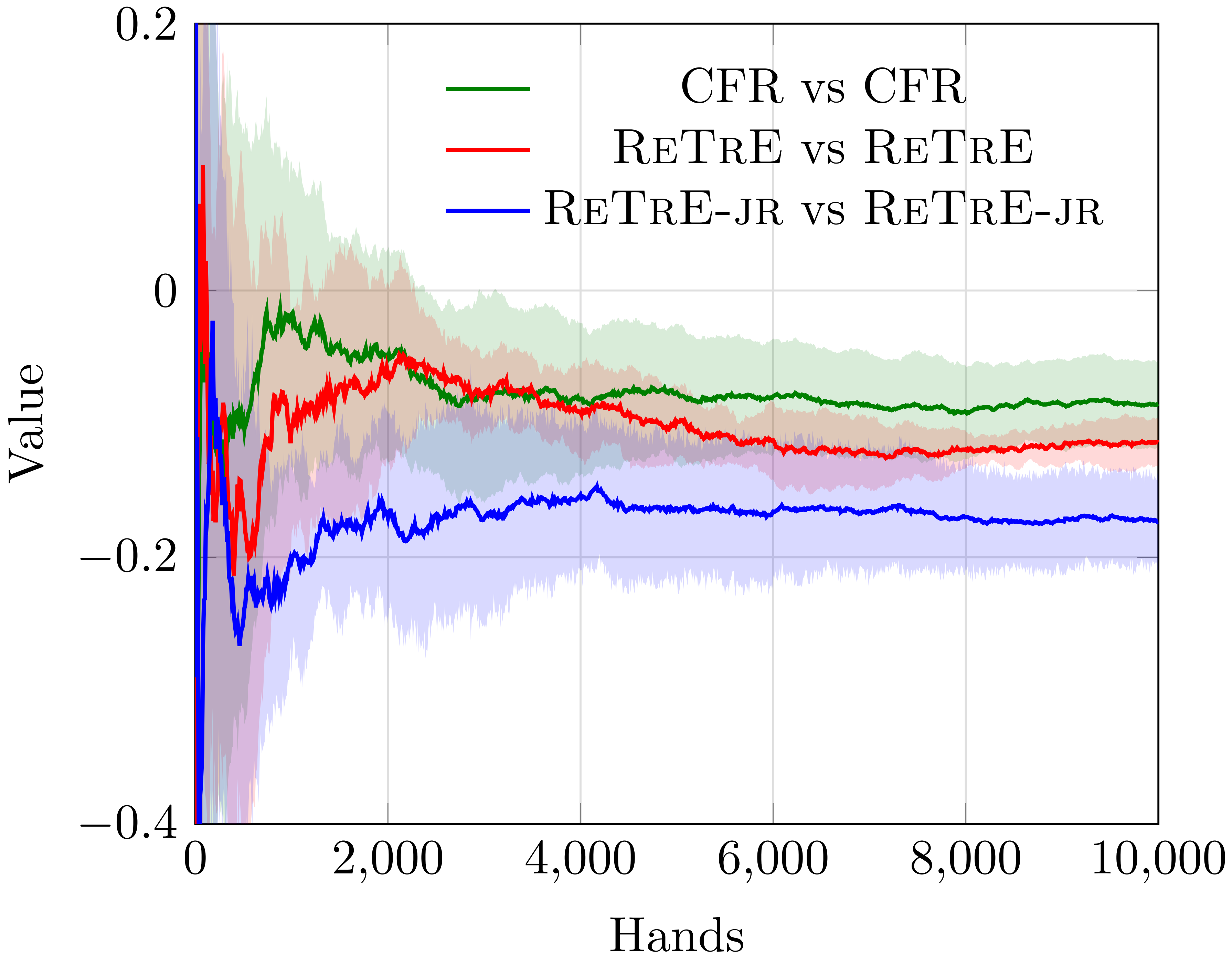
Different EVs

ReTrE - Head-to-head Simulations

We run head-to-head games to evaluate actual performance during play.



AI vs CFR



AI vs AI

Conclusions

Focused on the challenge of analyzing large and infinite extensive-form games, capturing the essence of real world-scenarios, to find ε-NEs.

Presented ReTrE, a scalable domain-independent model-free abstraction framework, able to solve large extensive-form games in a simulation-based fashion.

Leveraged neural networks and confidence-based exploration to approximate the behavior of optimal regret minimization algorithm CFR in the full game.

Evaluated performance on games small enough to be analyzed by CFR, measuring exploitability, and observed low distance from equilibrium for ReTrE.

Future work

Evaluate performance on large games.

Comparison with abstraction algorithms.

Integration of CFR variants.

Optimize information embedding capturing potential.

Leverage other upper confidence bound methods.

Integrate strategy refinement techniques.

References

Areyan Viqueira, E., Cousins, C., Greenwald, A. (2020, May). *Improved Algorithms for Learning Equilibria in Simulation-Based Games*.

Auer, P., Cesa-Bianchi, N., Fischer, P. (2002). *Finite-time analysis of the multiarmed bandit problem*.

Brown, N., Lerer, A., Gross, S., Sandholm, T. (2019, May). Deep counterfactual regret minimization. In *International conference on machine learning* (pp. 793-802).

Brown, N., Sandholm, T. (2018). *Superhuman AI for heads-up no-limit poker: Libratus beats top professionals*.

Maschler, M., Solan, E., Zamir, S. (2013). *Game Theory*.

References

Nash, J. (1950). *Equilibrium points in n-person games*.

Nash, J. (1951). *Non-cooperative games*.

Russell, S., Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*.

Sandholm, T. (2010). *The state of solving large incomplete-information games, and application to poker*.

Waugh, K. (2009). *Abstraction in large extensive games*.

Zinkevich, M., Johanson, M., Bowling, M., Piccione, C. (2008). *Regret minimization in games with incomplete information*.

POLITECNICO
MILANO 1863

Thank you

JACOPO PIO GARGANO

ADVISOR PROF. NICOLA GATTI
CO-ADVISOR PROF. BARBARA CAPUTO