Optimal Game Solving

Counterfactual Regret Minimization

Jacopo Pio Gargano
gargj@amazon.com

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

Agenda

Why there *should* always be a purpose

How some theoretical stuff on game theory

What algorithm + code + demo

Open Challenges this was easy! what's next?

Game Theory

Theoretical framework for strategic interaction

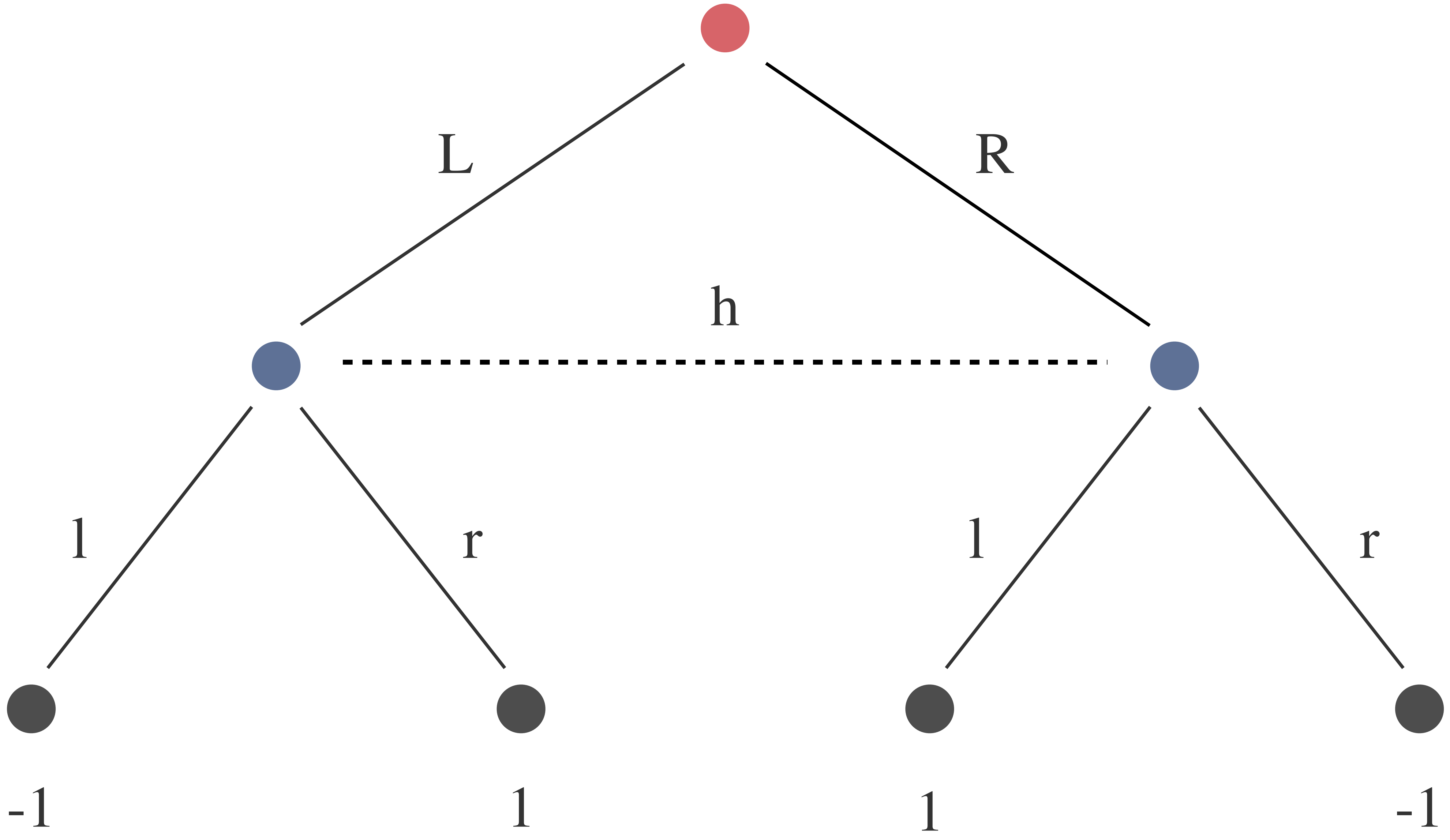
Mathematical models and algorithms to find strategies (Algorithmic Game Theory)

Conflict and cooperation

Intelligent rational decision-makers

Decisions influencing agents' welfare

Sequential Games Representation



Player 1 Player 2

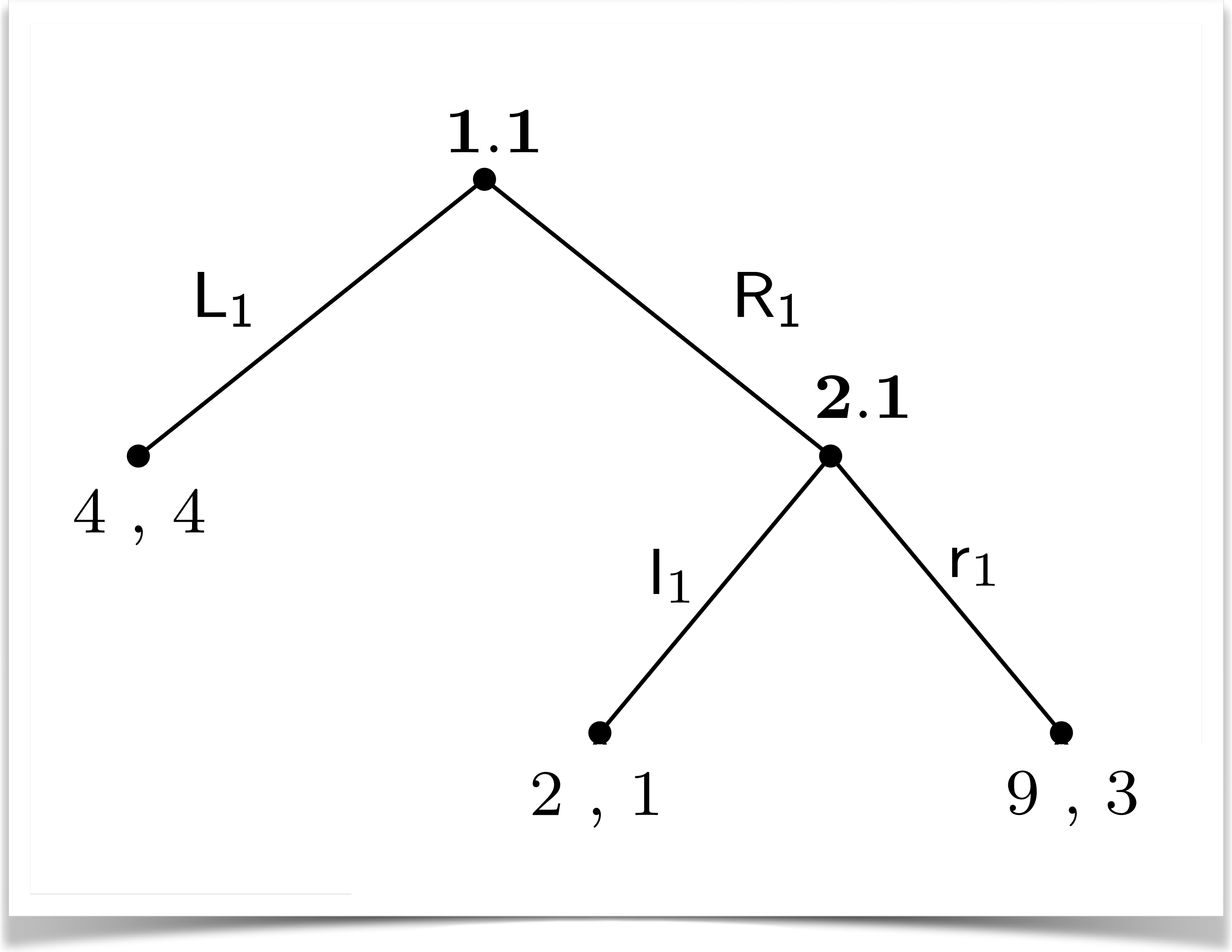
$$\sigma_i : H_i \rightarrow \Delta^{|A_{H_i}|}$$

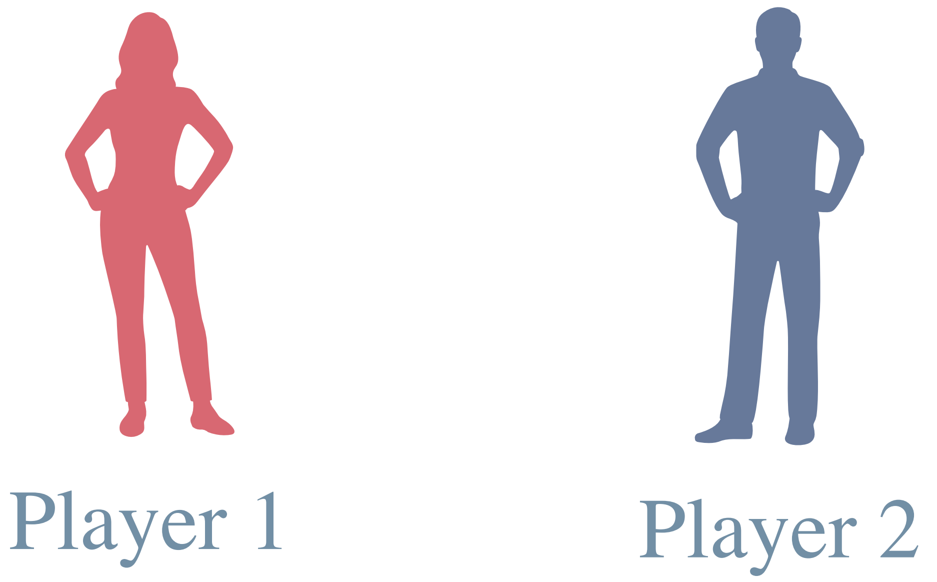
Strategy

Nash Equilibrium

Definition

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

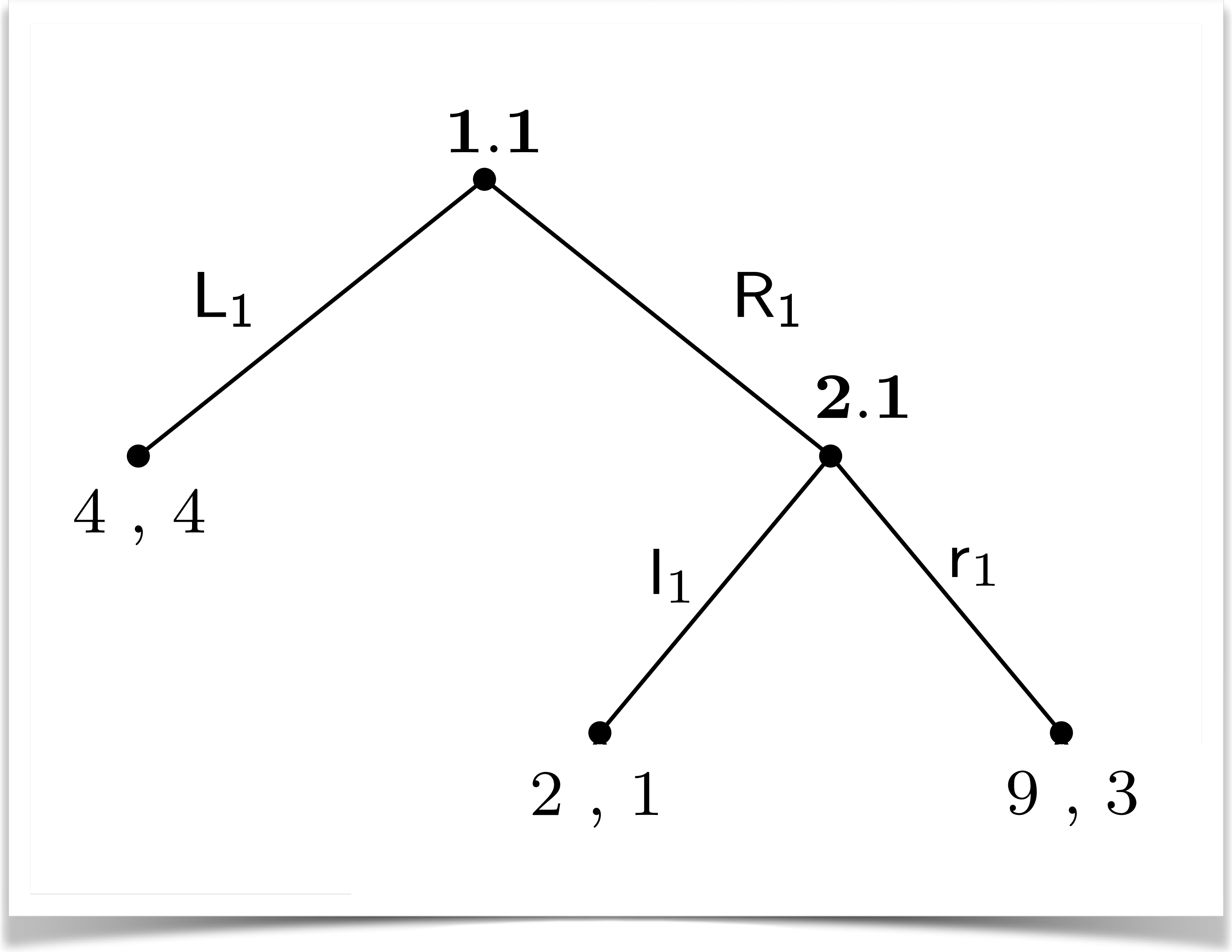


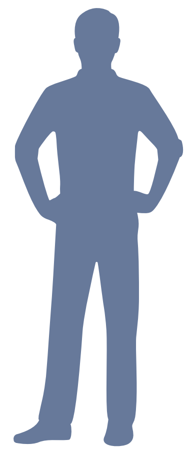


Nash Equilibrium

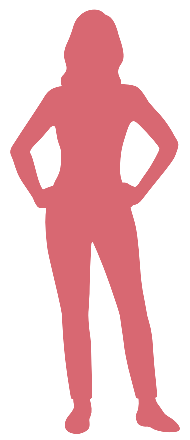
Definition

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





Player 2



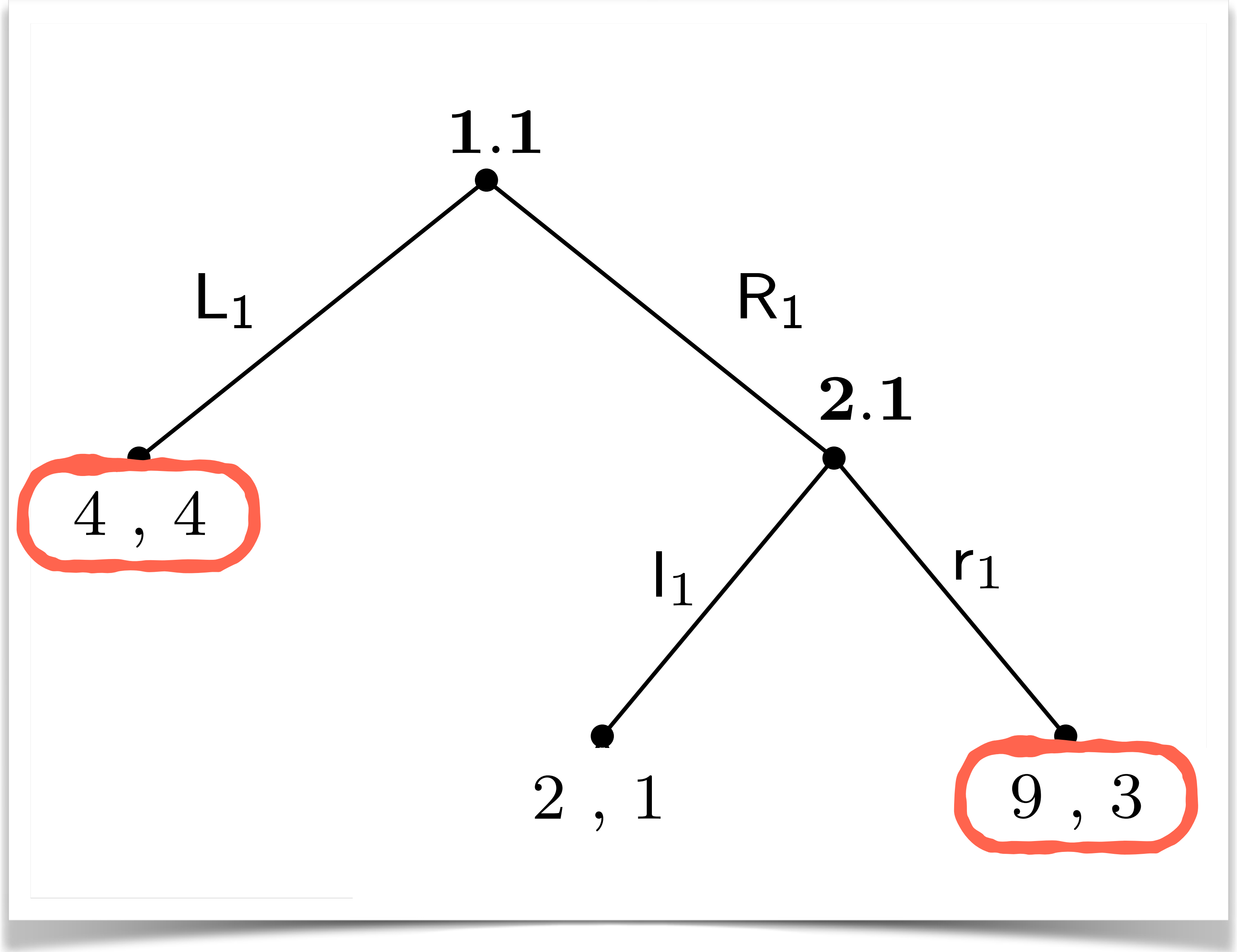
Player 1

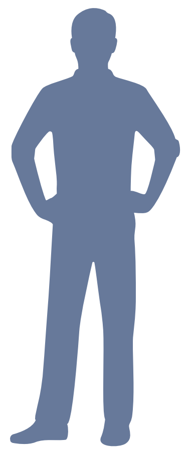
	l_1	r_1
L_1	4 , 4	4 , 4
R_1	2 , 1	9 , 3

Nash Equilibrium

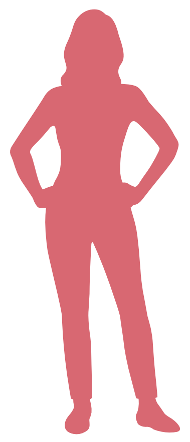
Definition

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





Player 2



Player 1

	l_1	r_1
L_1	4 , 4	4 , 4
R_1	2 , 1	9 , 3

Nash Equilibrium

Definition

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

Why?

If an agent sticks to the Equilibrium strategy, their opponent won't benefit from deviating from the Equilibrium strategy itself

Therefore, if one plays the Equilibrium strategy, the other will play the Equilibrium strategy in the worst case


https://www.youtube.com/watch?v=jwG_qR6XmDQ&feature=emb_logo

"So it really was a decision I had to make for myself and the framework I found, which made the decision incredibly easy, was a regret minimization framework.
So I wanted to project myself forward to age 80 and say 'okay now I'm looking back on my life, I want to have minimized the number of regrets I have', and I knew that when I was going to be 80 I was not going to regret having tried this."

J. Bezos

Regret Minimization

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

Regret Minimization

Regret

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

Regret Minimization

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

Regret Matching (CFR)

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

Regret Minimization

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

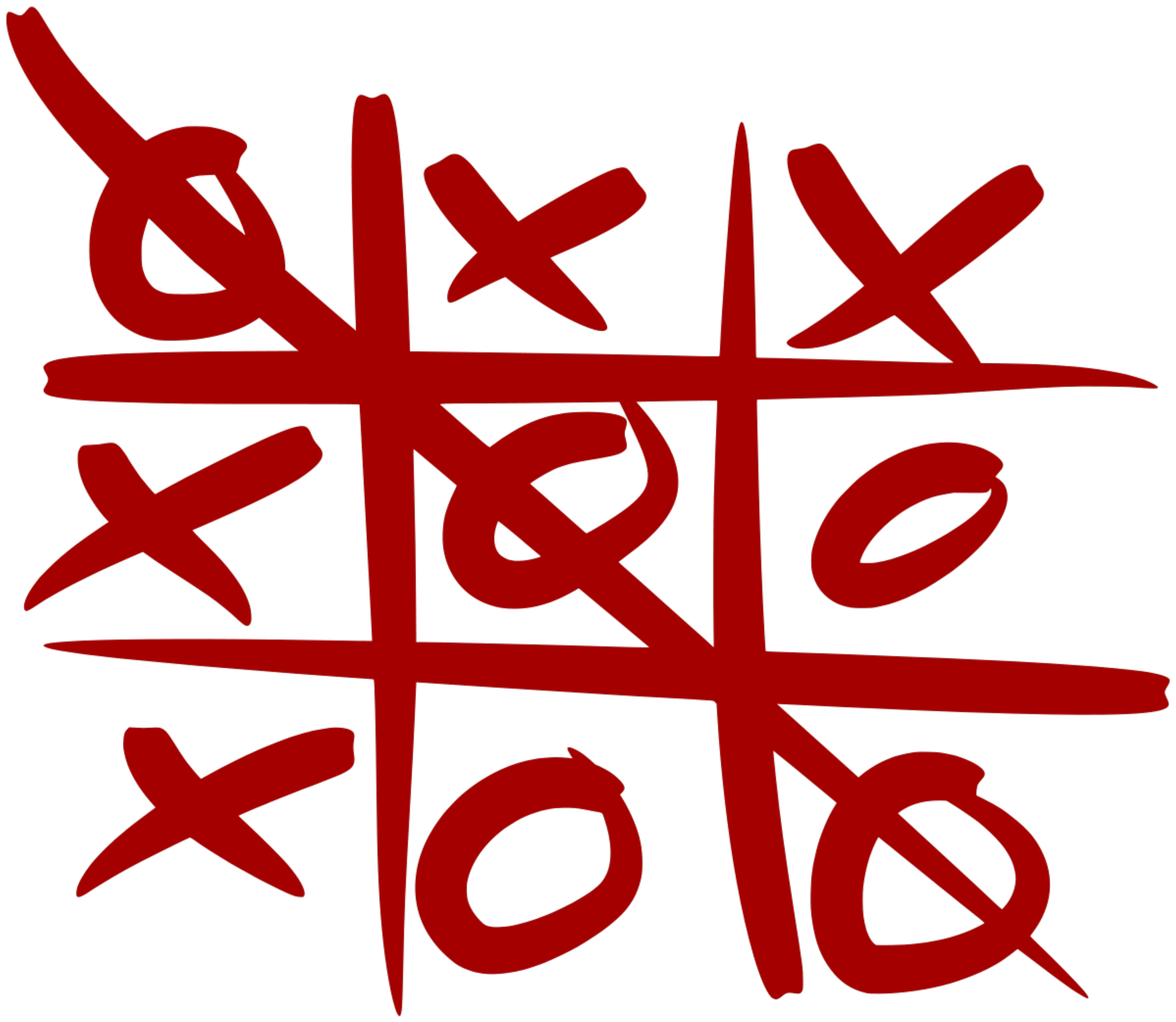
Algorithm 1 CFR

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 $v_i^\sigma(h)$

Demo



Open Challenges

What about large games?

Domain-independence

Model-freedom

Strategy refinement

References

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

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

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

Neller, T., Lanctot, M. (2013). *An introduction to counterfactual regret minimization*.

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

Thank you

Jacopo Pio Gargano