CFR – Algorithm for Solving Games

(Zinkevich *et. al.* NIPS 2008)

Two Players

Imperfect Information

Sequential Decisions

Zero-Sum

Photo © lawrence on flickr

Games are Just Formal Problems

- Coast Guard Patrol Scheduling (Shieh *et. al.* AAAI 2012)
  - One player allocates security resources
  - Other player chooses where to attack
- Insulin & Blood Glucose (Chen & Bowling NIPS 2012)
  - One player chooses monitoring and treatment schedule
  - Other player chooses worst outcome from a sample

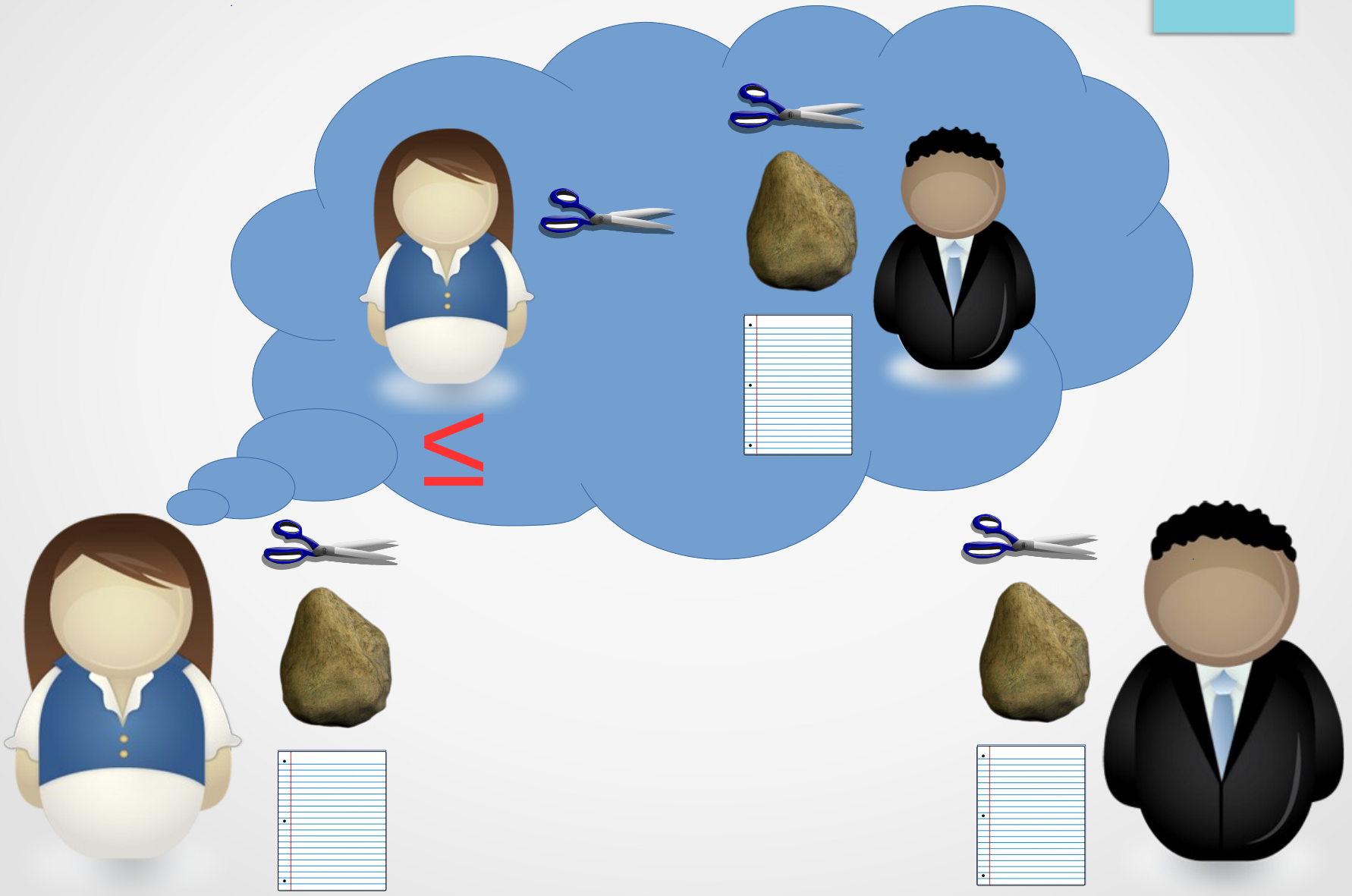
Poker: Heads-up Limit Texas Hold'em

- $3.16*10^{17}$ game states
- $1.38*10^{13}$ unique situations for a player
- 131TB for a strategy
- 1 CPU month to walk the whole game tree

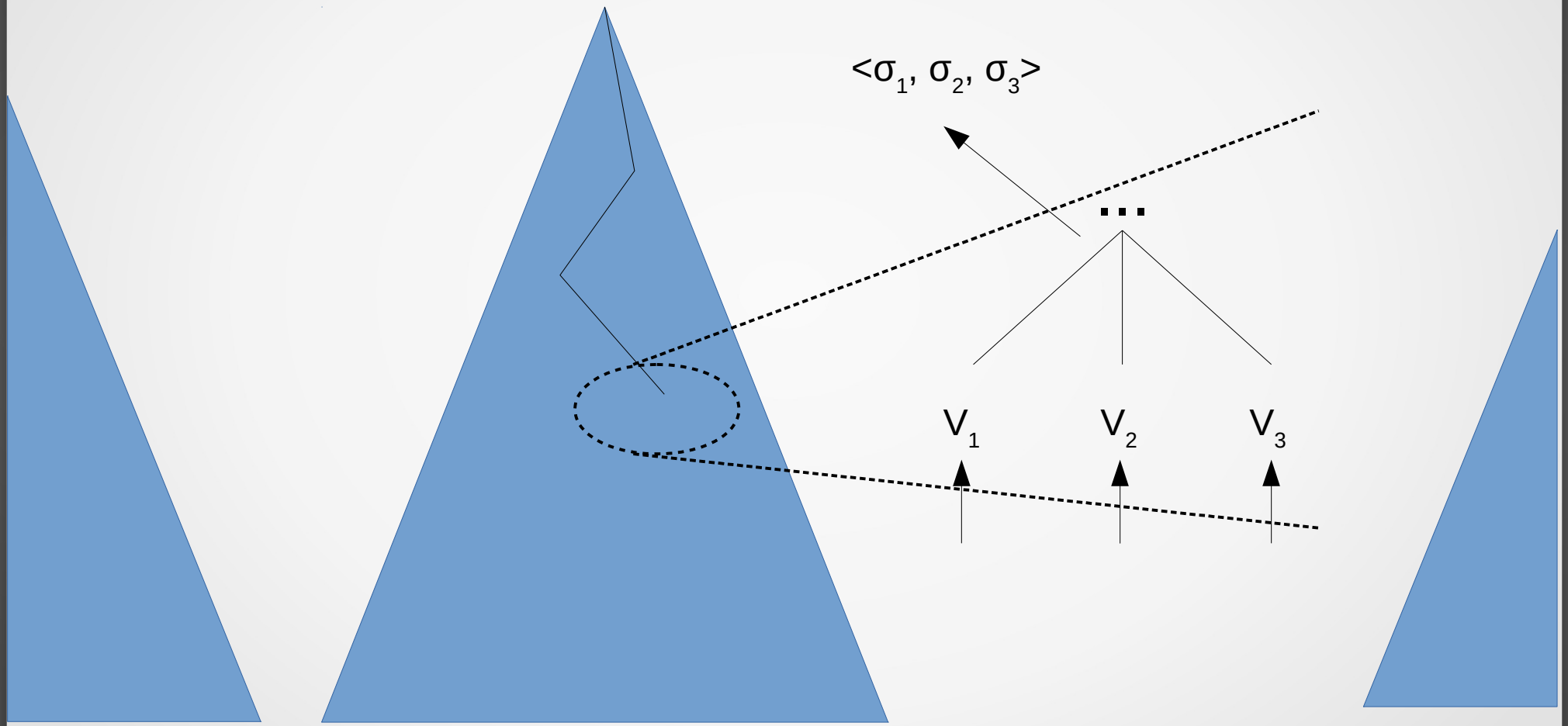
Game Solution?

~~Mini-max~~

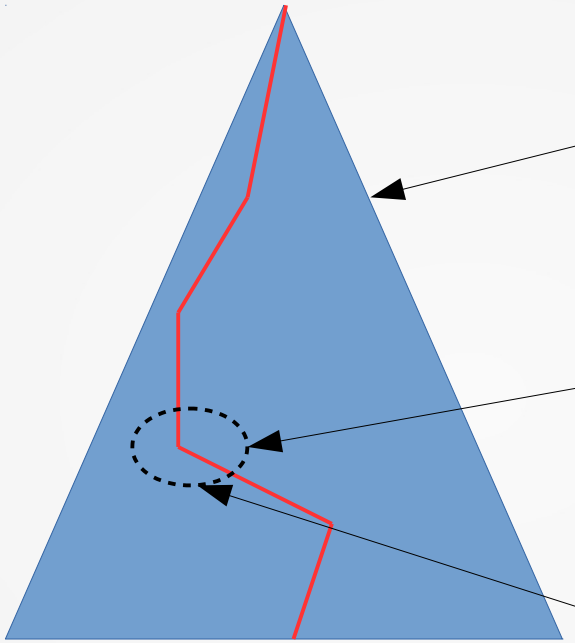
Game Solution: Nash Equilibrium



CFR is similar to UCT self-play



CFR is not UCT



Updates and stores whole tree

Regret-matching instead of UCB

Counterfactual values

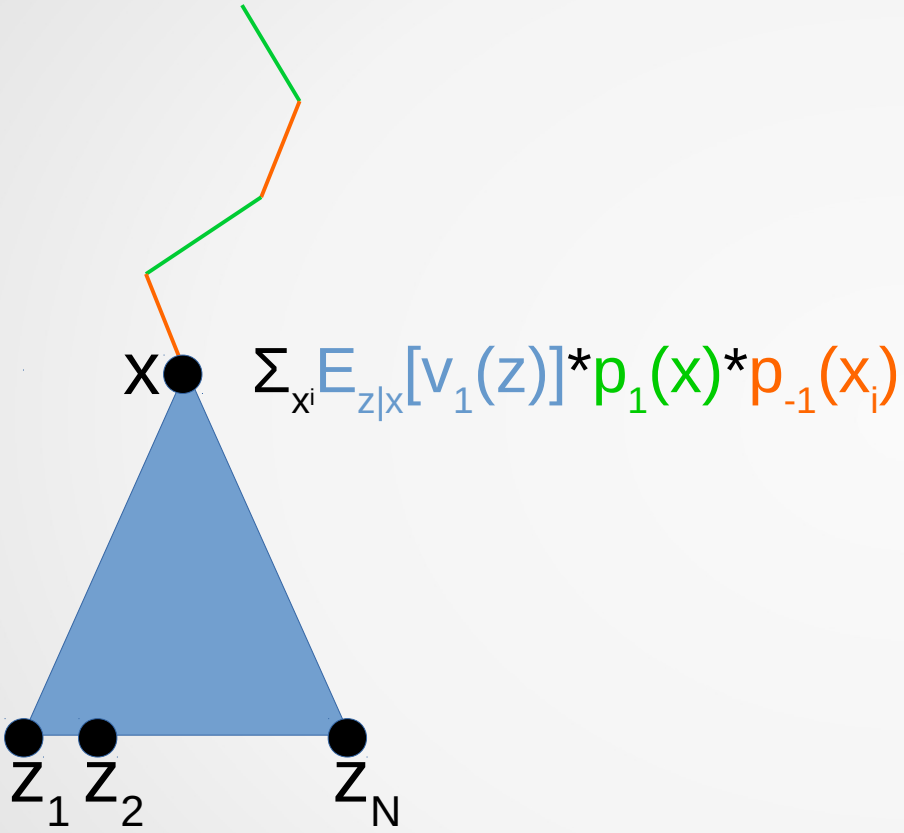
time T

...

Offline

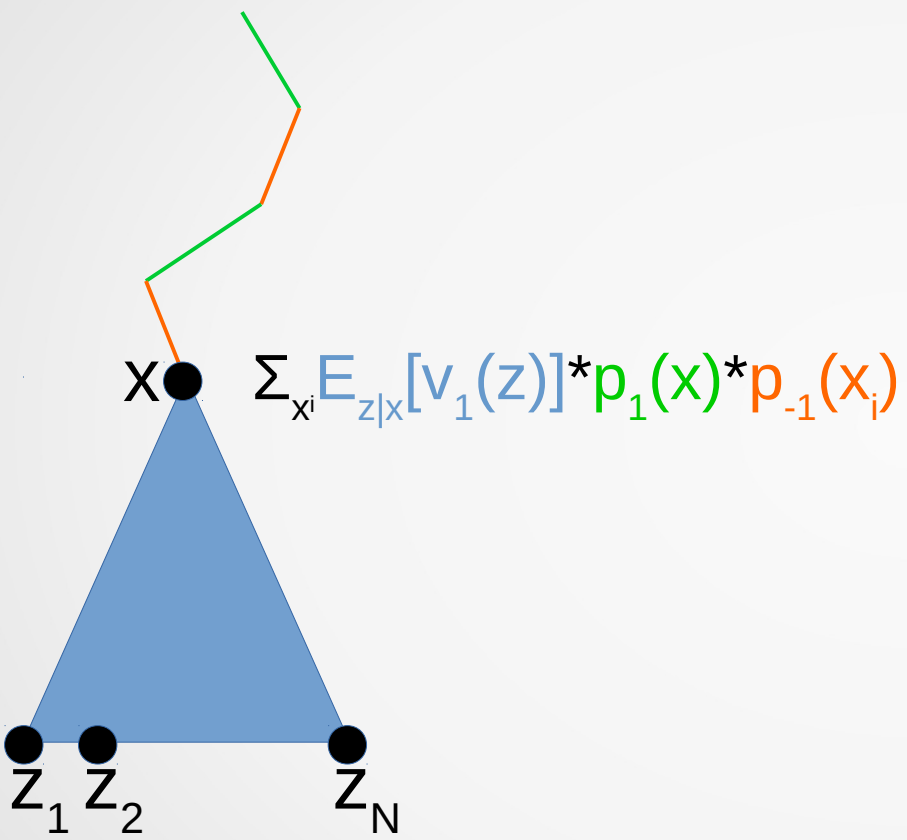
Uses average strategy

Counterfactual Value?

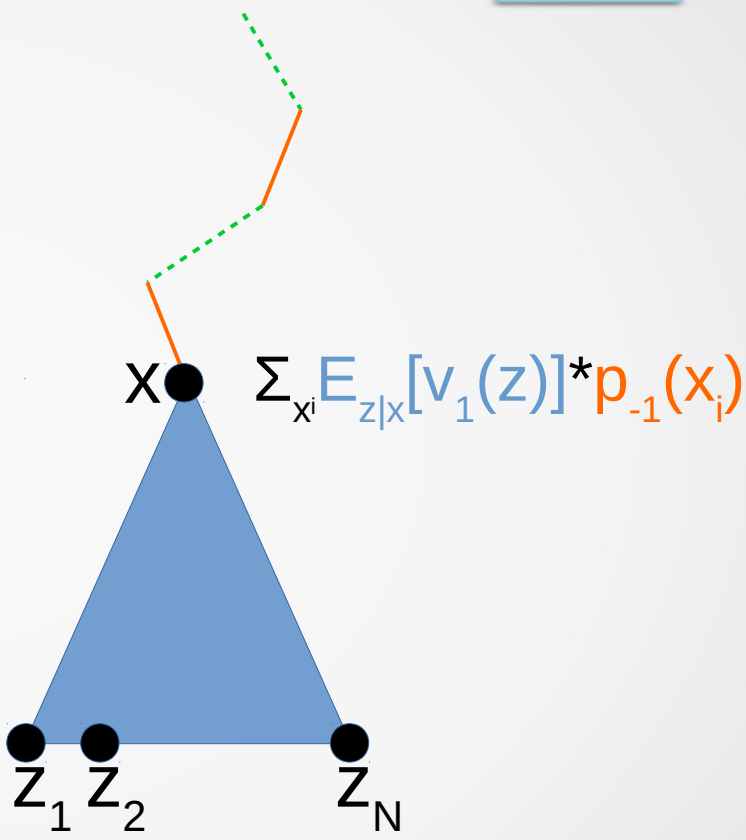


Expected value

Counterfactual Value



Expected value



Counterfactual value
“What if p1 plays to x?”

Not so strange: perfect information games do this too

Regret Minimisation

Online action selection algorithm

- Repeatedly make decisions
- Full information
- Adversarial environment

Guarantee average value → optimal value

Hedge, Regret-matching

Regret: how well could I have done?

… H_T H HT H_T → $

… H H H H → $ R_H=+

… T T T T → $ R_T=-

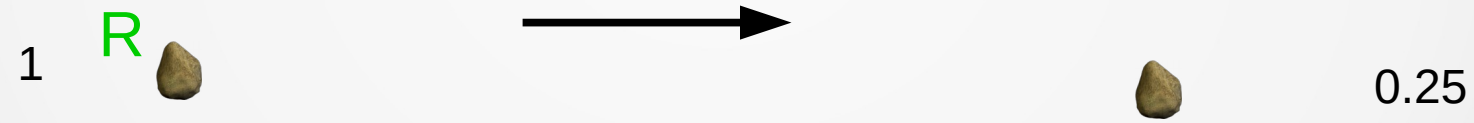
Regret-matching: current strategy

$$R_a^{T,+} = \max(R_a^T, 0) \qquad \sigma_a^{T+1} = R_a^{T,+} / (\Sigma_b R_b^{T,+})$$

3 R_{scissors}

1 R_{rock}

-2 R_{paper}



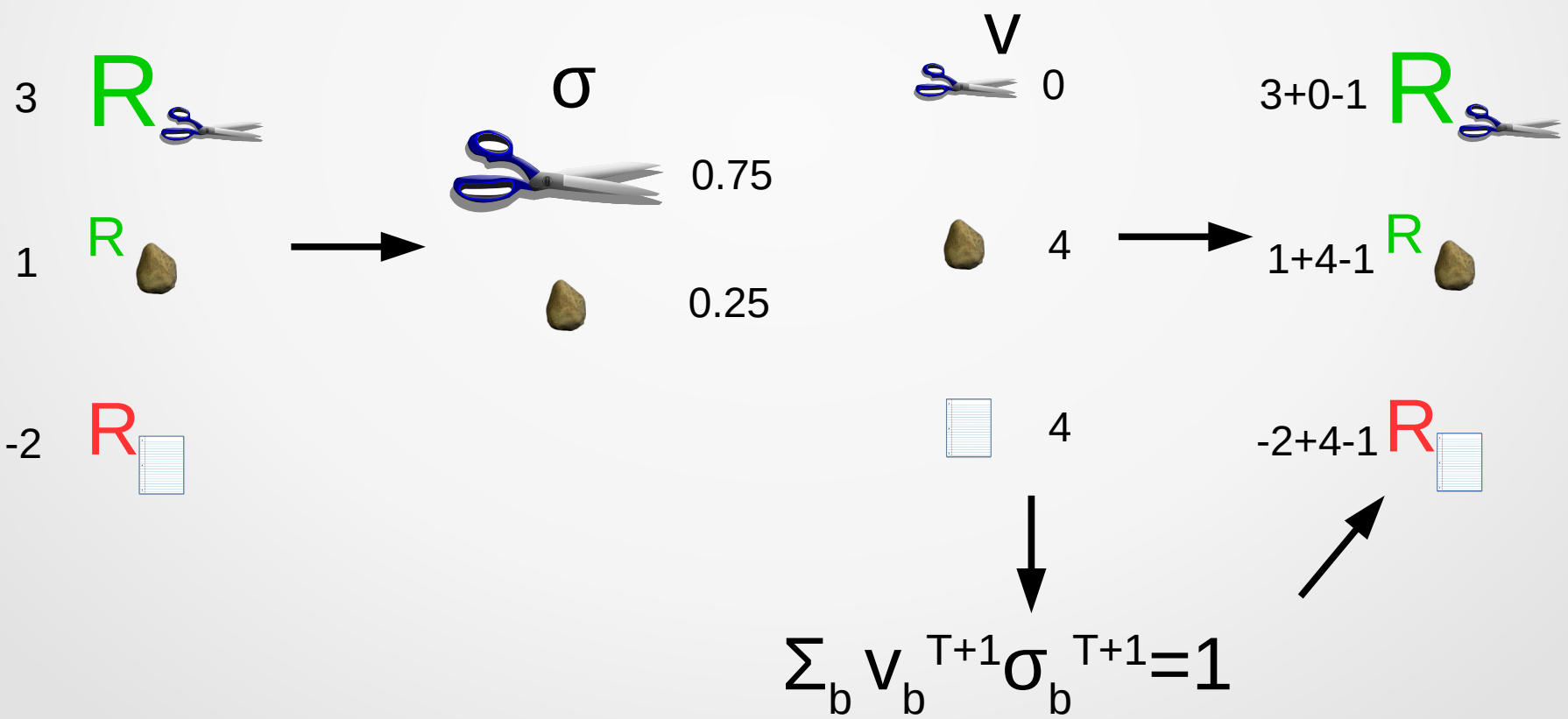
0.75

0.25

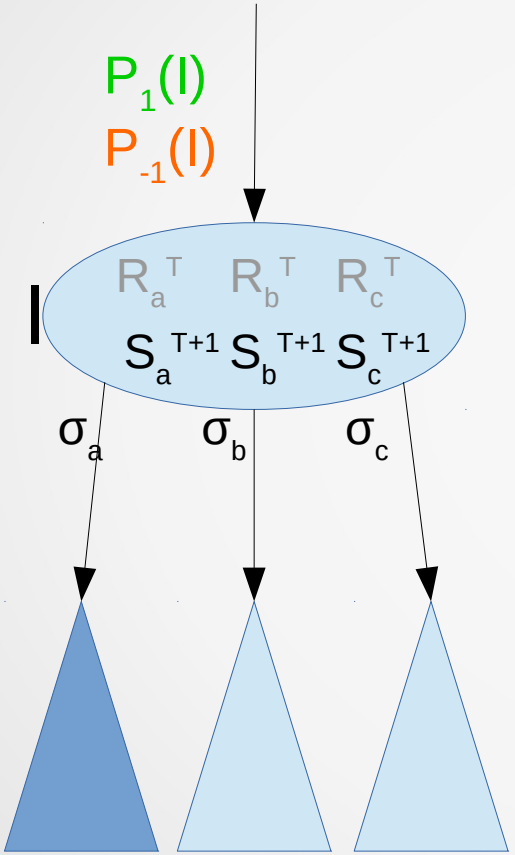
$$\max_a R_a^T \leq \sqrt{|A|T}$$

Regret-matching: new regrets

$$\sigma_a^{T+1} = R_a^{T,+} / (\Sigma_b R_b^{T,+}) \qquad R_a^{T+1} = R_a^T + v_a - \Sigma_b v_b^{T+1} \sigma_b^{T+1}$$

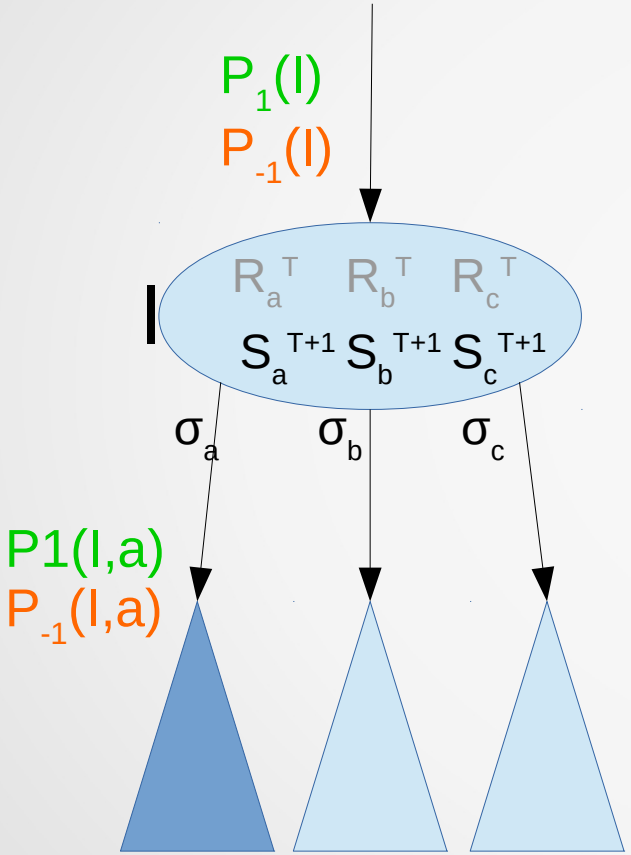


CFR update



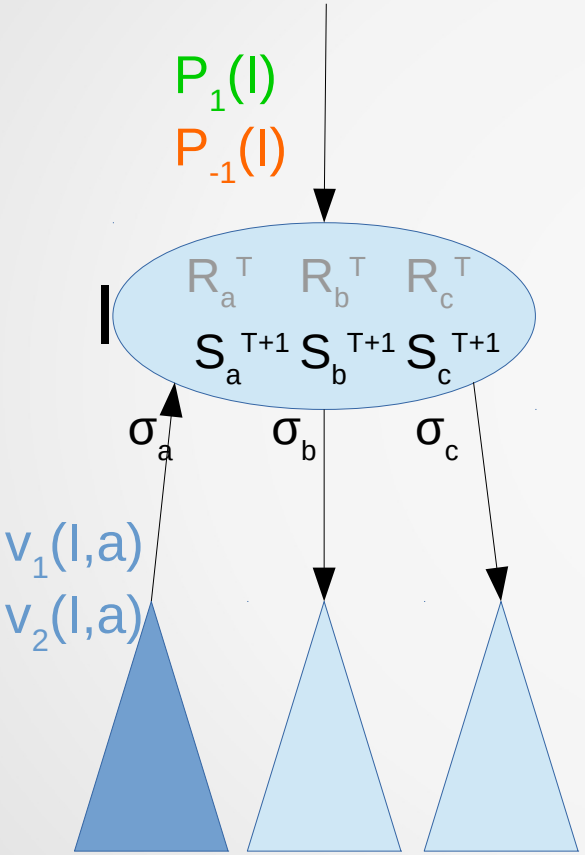
- Get current policy
- Update average
- Update child subtrees
- Get child values
- Update regrets
- Pass back value

CFR update



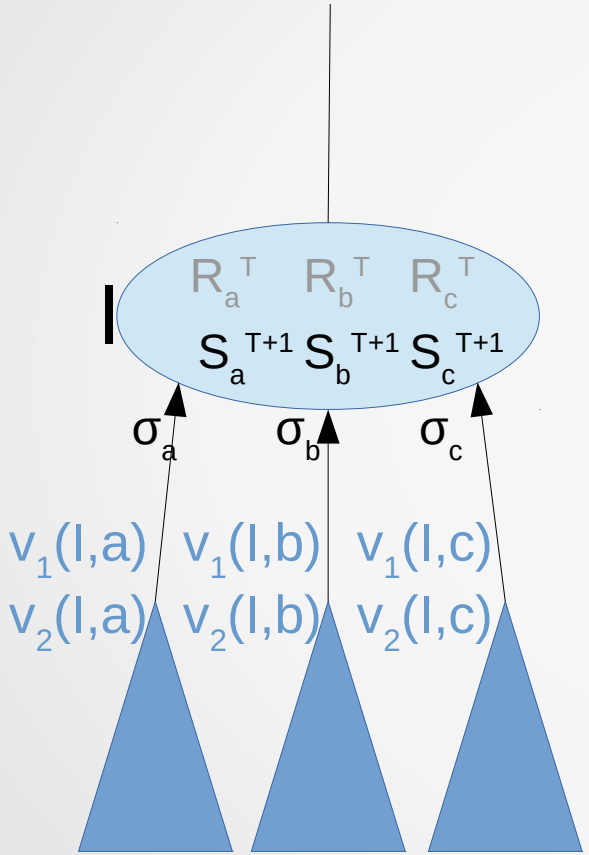
- Get current policy
- Update average
- Update child subtrees
- Get child values
- Update regrets
- Pass back value

CFR update



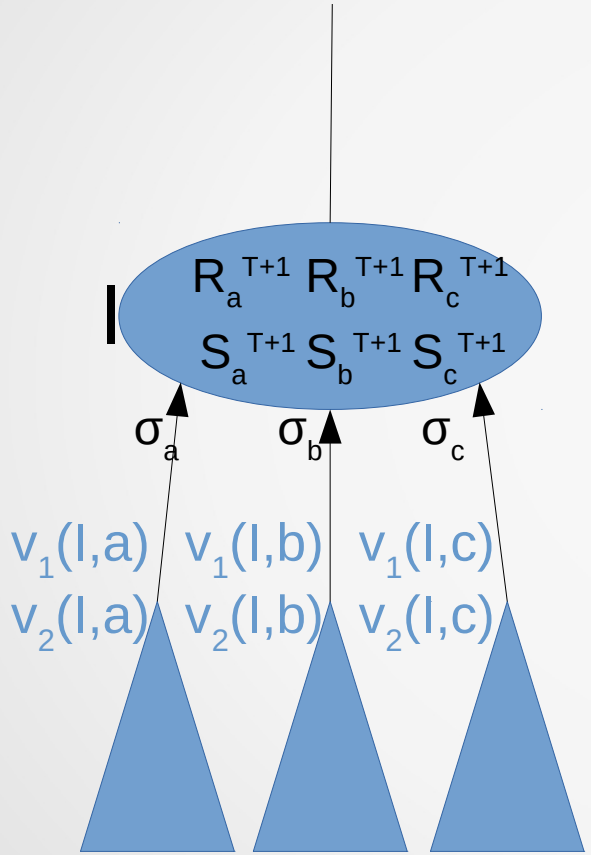
- Get current policy
- Update average
- Update child subtrees
- Get child values
- Update regrets
- Pass back value

CFR update



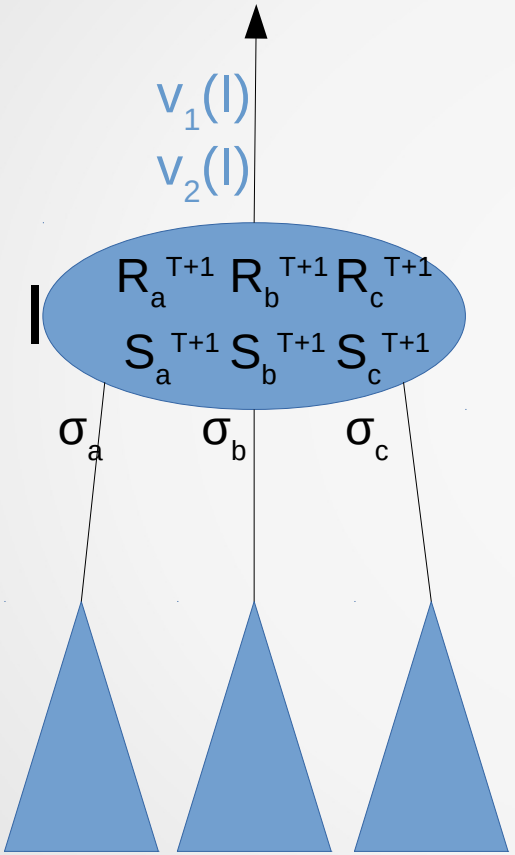
- Get current policy
- Update average
- Update child subtrees
- Get child values
- Update regrets
- Pass back value

CFR update



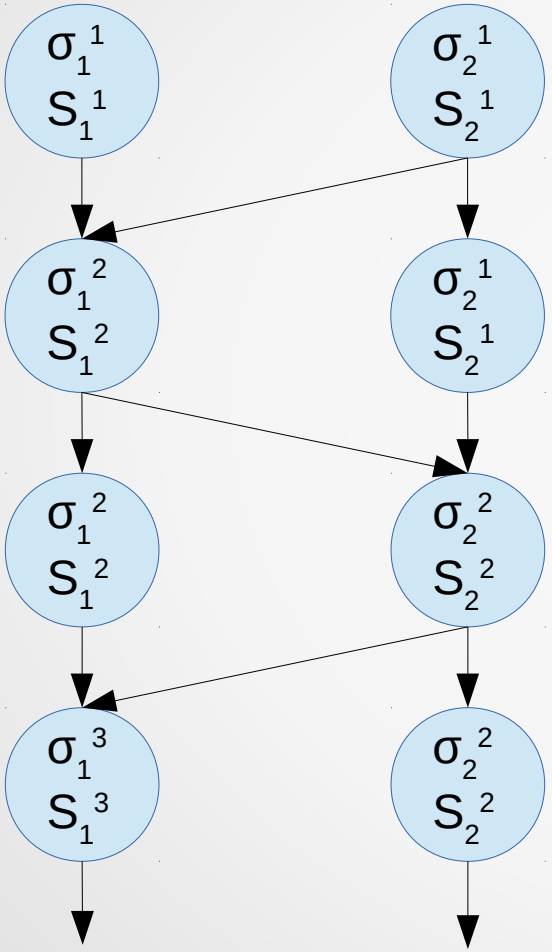
- Get current policy
- Update average
- Update child subtrees
- Get child values
- Update regrets
- Pass back value

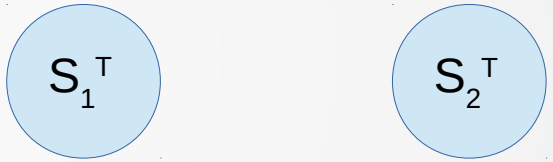
CFR update



- Get current policy
- Update average
- Update child subtrees
- Get child values
- Update regrets
- Pass back value

CFR





ε-Nash equilibrium

$\varepsilon \leq \Sigma_p L M_p \sqrt{|A|/T}$

(or $2L|I|\sqrt{|A|/T}$)

What else?

- Efficient implementation tricks
- Even more like MCTS: MCCFR (Lanctot *et. al.* NIPS 2009)
- One sided: CFR-BR (Johanson *et. al.* AAAI 2012)
- Save space: CFR-D (Burch *et. al.* AAAI 2014)
- Save time: CFR+ (Tammelin *et. al.* IJCAI 2015)
- ...