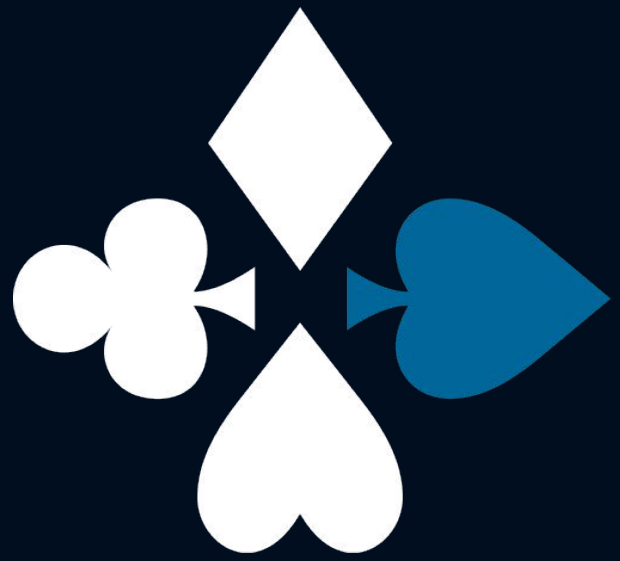


Pokerbots 2024

Lecture 6: Dr. Noam Brown

Sponsors

Jane Street

LOTS of Announcements

Weekly Tournament 2

($1000) WINNER: Pineapple
($500) BIGGEST UPSET: Grilled Cheese
($750) MOST IMPROVED: Poker Ace

Weekly Tournament 3

($1000) WINNER: Pineapple
($500) BIGGEST UPSET: Trader Joe
($750) MOST IMPROVED: ssad_people

Lightning Tournament Last Night

Strategy Proof Turngate	$400
Pineapple	$300
DKE Sophomores	$225
The Fish	$175
cxrt	$150

Today (Monday 1/29)

- Noam Brown Guest Lecture (now)
- Noam Brown Poker Social (immediately after)
- Scrimmage Challenges disabled at 11:59pm

Tomorrow (Tuesday 1/30)

Team Strategy Reports Due at 11:59pm

- Email to pokerbots@mit.edu
- Required to pass this class
- Expectations in Syllabus

Wednesday 1/31

- GTO Wizard Guest Lecture (1pm)
- GTO Wizard Poker Social (immediately after)
- Final bots must be uploaded by 11:59pm

Friday 2/2

Pokerbots Final Event - 4:30-7pm in Kresge

- 4:30-6:00 Sponsor Networking Session and Puzzle Contest in Kresge Lobby
- 6:00-7:00 Awards Presentation and Closing Ceremony in Kresge Auditorium

Today's Raffle: RSVP at pkr.bot/rsvp to win Sony Headphones

Saturday 2/3

GTO Wizard Poker Tournament- 5pm in BC Porter Room

- $3500 Cash Prize Pool
- Limited to 150 - RSVP to secure your spot @ pkr.bot/gto

Dr. Noam Brown

OpenAI

Libratus & Pluribus Creator

Building a Superhuman AI for No-Limit Poker

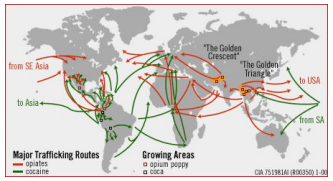
Noam Brown

OpenAI Reasoning

"And that's why there's never going to be a computer that will play World Class Poker. It's a people game."
-Doyle Brunson, *Super/System* 1979

"The analysis of a more realistic poker game than our very simple model should be quite an interesting affair."
-John Forbes Nash, 1951

Imperfect-Information Multi-Agent



Perfect-Information Multi-Agent

No-Limit Texas Hold'em Poker

- 

- Long-standing challenge problem in AI and game theory
- Massive in size (two-player has 10^{161} decision points)
- By far the most popular form of poker

2017 Brains vs AI

- Libratus (our 2017 AI) against four of the ***best*** heads-up no-limit Texas Hold'em poker pros



- 120,000 hands over 20 days in January 2017
- $200,000 divided among the pros based on performance
- Won with 99.98% statistical significance
- Each human lost individually to Libratus

2017 Brains vs AI

- 120,000 hands of poker against a team of pros trying to exploit the bot
- Trained from self-play; no human data
- No deep neural networks
- 3 million core hours to train (~$100,000) and 1,200 CPU cores to run (no GPUs)

2019 Pluribus Six-Player Poker AI

[Brown & Sandholm Science-19]

- Pluribus (our 2019 AI) against 15 top professionals in ***six-player*** no-limit Texas Hold'em

- 10,000 hands over 12 days in June 2019
  - Used variance-reduction techniques to decrease luck
  - One bot playing with five humans
- Won with >95% statistical significance
- Cost under $150 to train, runs on 28 CPU cores (no GPUs)

Who is the better poker player?

Option 1: Someone who, over a large enough sample size, wins head-to-head vs. any other player

Option 2: Someone who makes more money playing poker than anyone else

Who is the better poker player?

Minimax Equilibrium

Option 1: Someone who, over a large enough sample size, wins head-to-head vs. any other player

Population Best Response

Option 2: Someone who makes more money playing poker than anyone else

Minimax Equilibrium

Minimax Equilibrium: a set of strategies in which no player can improve by deviating

In two-player zero-sum games, playing a minimax equilibrium ensures you will not lose in expectation

Exploitability: How much we'd lose to a best response

	Round 1	Round 2	Round 3
Us			
Best Response			

Our Exploitability = 1

Minimax Equilibrium

Minimax Equilibrium: a set of strategies in which no player can improve by deviating

In two-player zero-sum games, playing a minimax equilibrium ensures you will not lose in expectation

Exploitability: How much we'd lose to a best response

	Round 1	Round 2	Round 3
Us			
Best Response			

Our Exploitability = 1

Minimax Equilibrium

Minimax Equilibrium: a set of strategies in which no player can improve by deviating

In two-player zero-sum games, playing a minimax equilibrium ensures you will not lose in expectation

Exploitability: How much we'd lose to a best response

Critical assumption: Our strategy is common knowledge, but the outcomes of random processes are **not** common knowledge

Minimax Equilibrium

Minimax Equilibrium: a set of strategies in which no player can improve by deviating

In two-player zero-sum games, playing a minimax equilibrium ensures you will not lose in expectation

Exploitability: How much we'd lose to a best response

	Round 1	Round 2	Round 3
Us			
Best Response			

Our Exploitability = 0

Minimax Equilibrium

“Poker is simple, as your opponents make mistakes, you profit.”

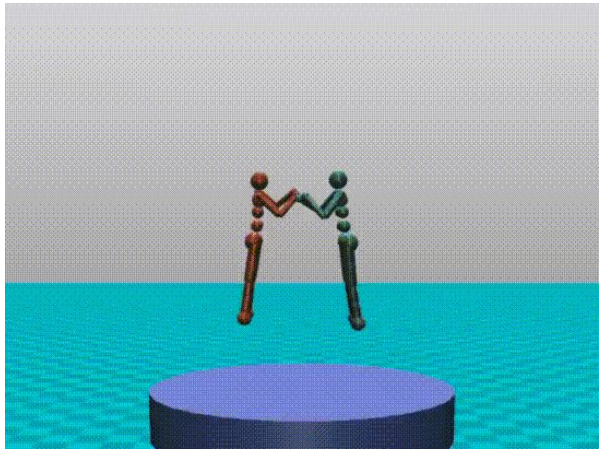
-Ryan Fee’s Poker Strategy Guide

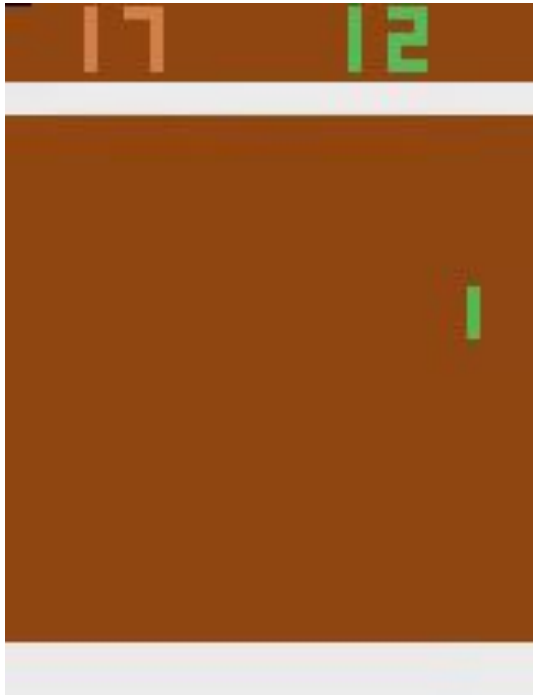
	Round 1	Round 2	Round 3
Us			
Best Response			

Our Exploitability = 0

Self-play in two-player zero-sum games

- In **self-play**, an agent gradually improves by playing against copies of itself
- Initial strategy can be completely random
- In balanced **two-player zero-sum** games, **sound self-play** provably converges to a **minimax equilibrium**
- Thus, given sufficient memory and compute, **any finite two-player zero-sum game can be "solved" via self-play**





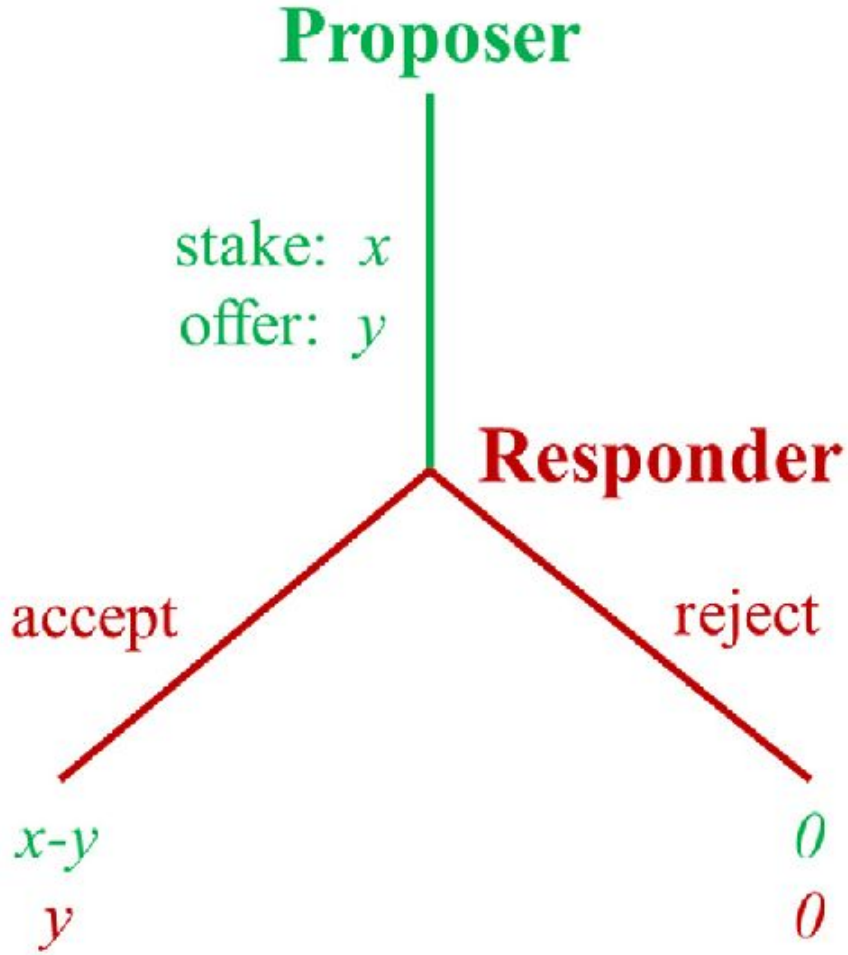
Self-play in two-player zero-sum games

- In **self-play**, an agent gradually improves by playing against copies of itself
- Initial strategy can be completely random
- In balanced **two-player zero-sum** games, **sound self-play** provably converges to a **minimax equilibrium**
- Thus, given sufficient memory and compute, **any finite two-player zero-sum game can be "solved" via self-play**

Question: What about non-two-player zero-sum games?

Ultimatum Game

- Alice is given $100
- First, Alice offers $0 - $100 to Bob
- Then, Bob must decide whether to **accept** or **reject**
  - If Bob **accepts**, then Alice and Bob keep their money
  - If Bob **rejects**, then Alice and Bob get nothing



Who is the better poker player?

~~Minimax Equilibrium~~

~~Option 1: Someone who, over a large enough sample size, wins head-to-head vs. any other player~~

Not meaningful in general games!

Population Best Response

Option 2: Someone who wins more often against the population of players than anyone else

Requires data on the population of players, i.e., human data

piKL- Human-regularized RL and planning

(Jacob et al. 2022)

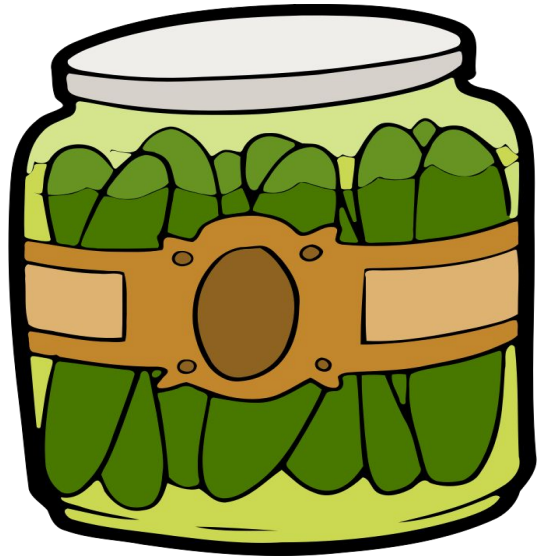
Idea: Given **anchor policy τ** from human imitation learning, when optimizing policy π, optimize the regularized utility:

$$u(\pi) = EV(\pi) - \lambda D_{KL}(\pi||\tau)$$

λ is the **anchor strength**:

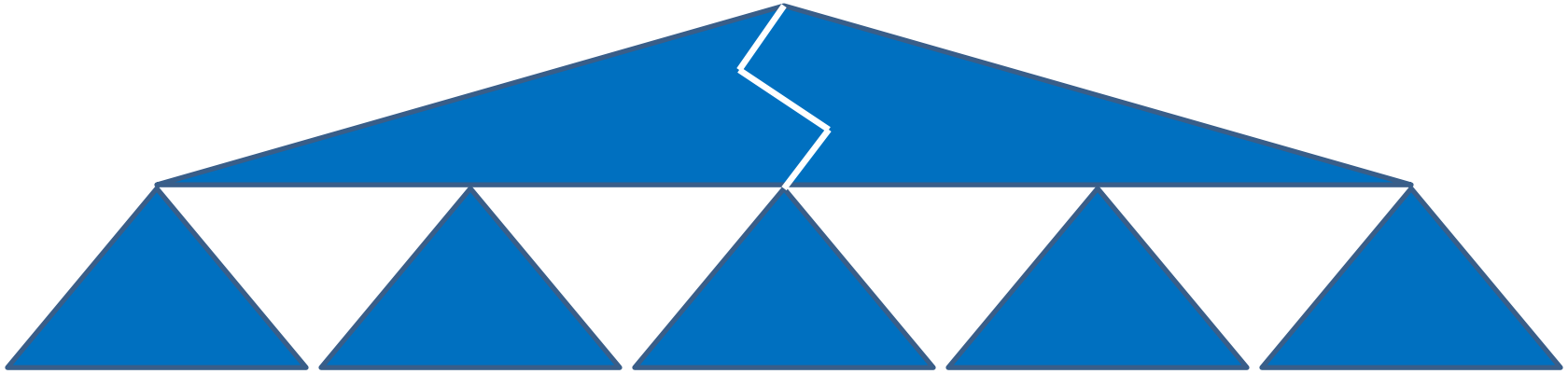
- **λ** = 0: self-play from scratch
- **λ** = infinity: human behavioral cloning
- Choosing **λ** in-between gains benefits of both.

Results: Significant policy improvement while maintaining high human compatibility.

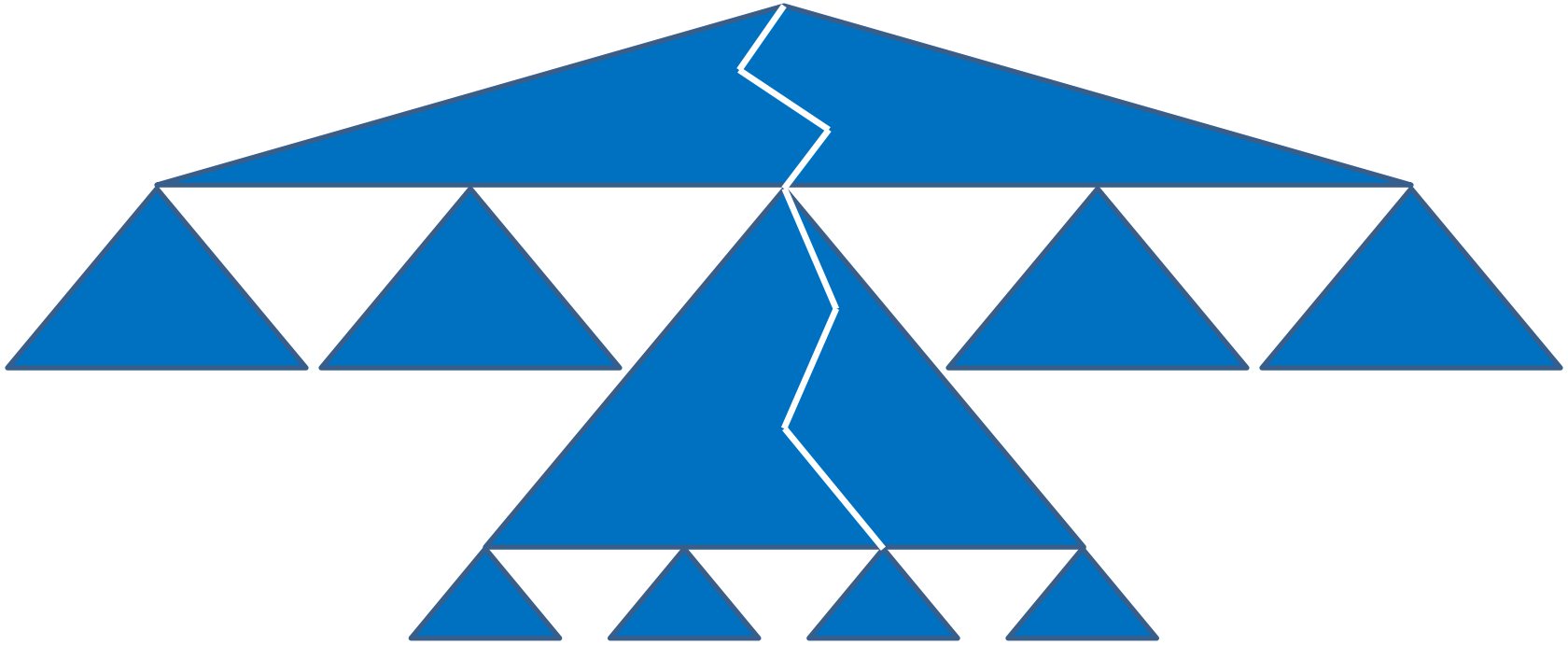


Libratus/Pluribus

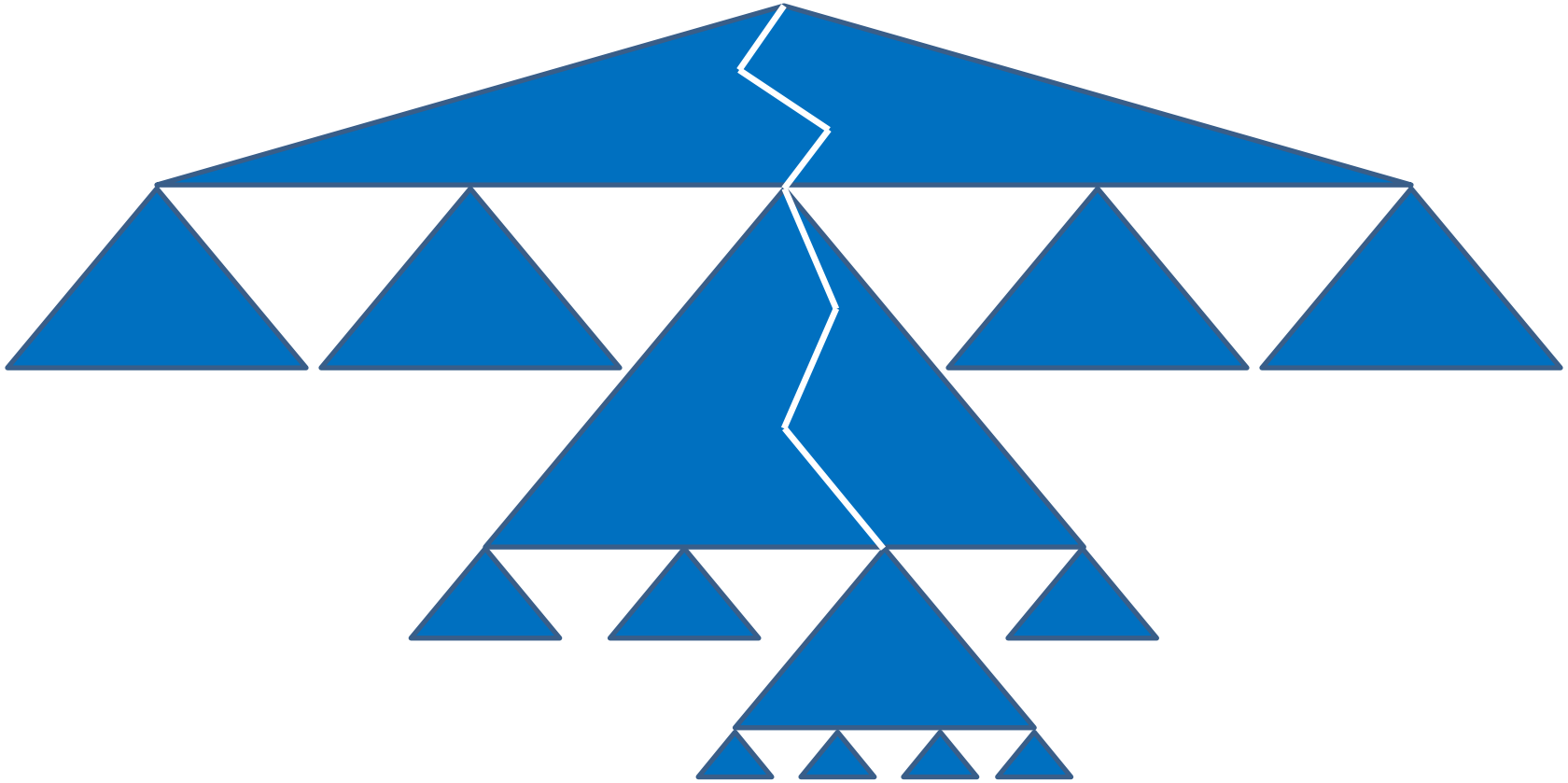
Libratus/Pluribus



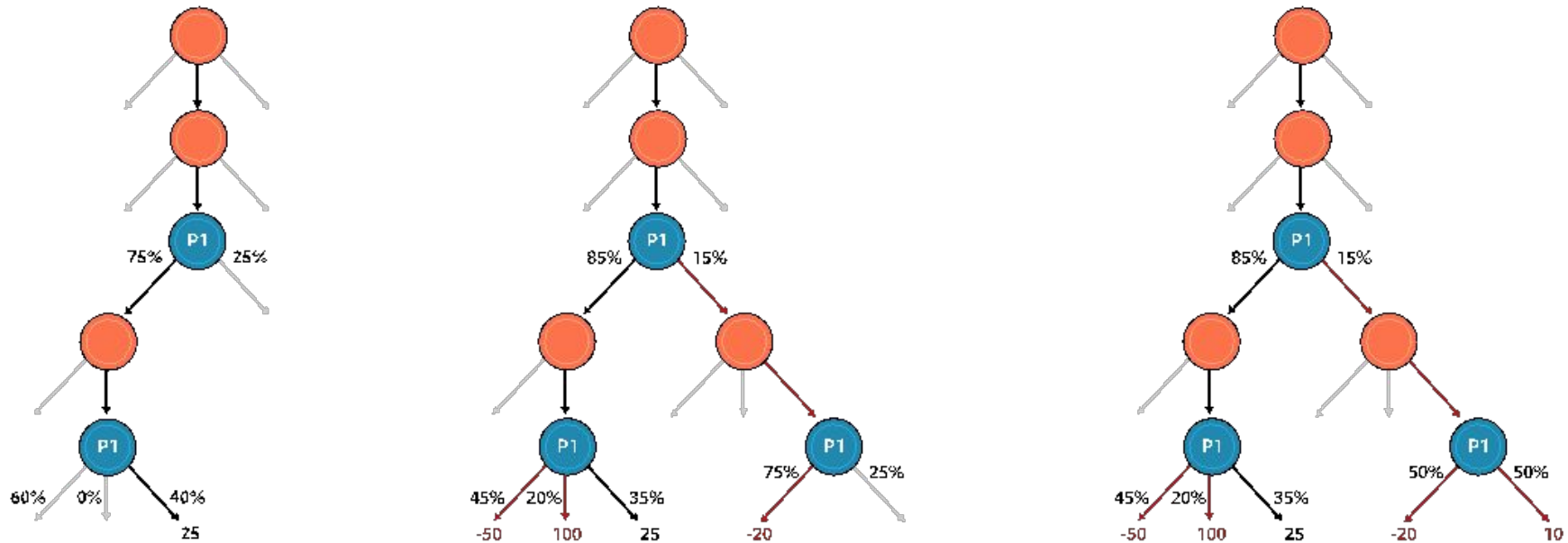
Libratus/Pluribus



Libratus/Pluribus



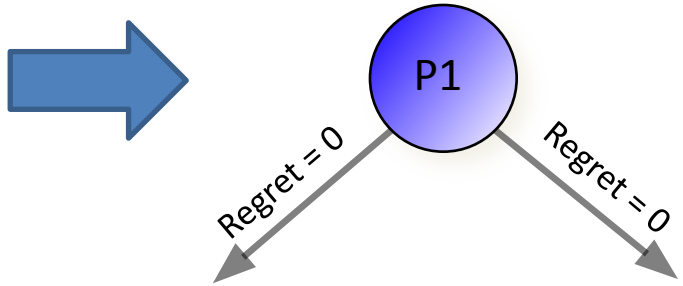
Computing Equilibria with Counterfactual Regret Minimization



Monte Carlo Counterfactual Regret Minimization (MCCFR)

[Zinkevich *et al.* NeurIPS-07, Lanctot *et al.* NeurIPS-09]

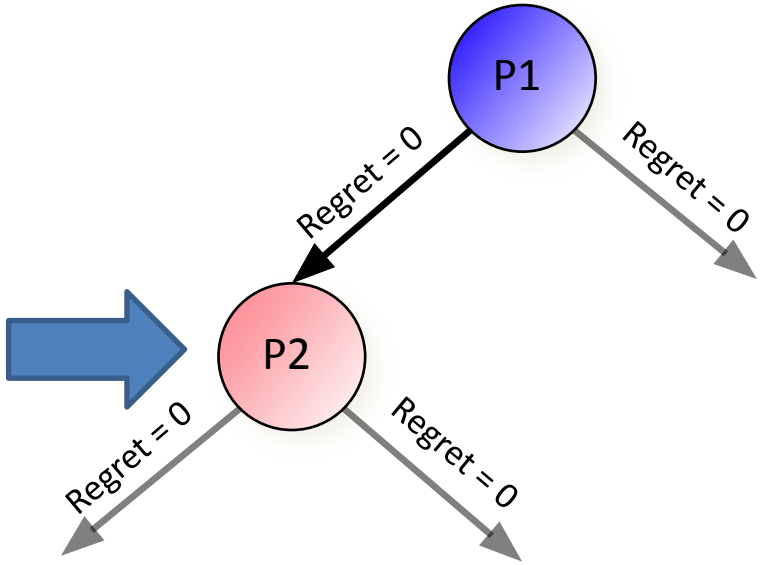
Pick action proportional to **positive regret**



Monte Carlo Counterfactual Regret Minimization (MCCFR)

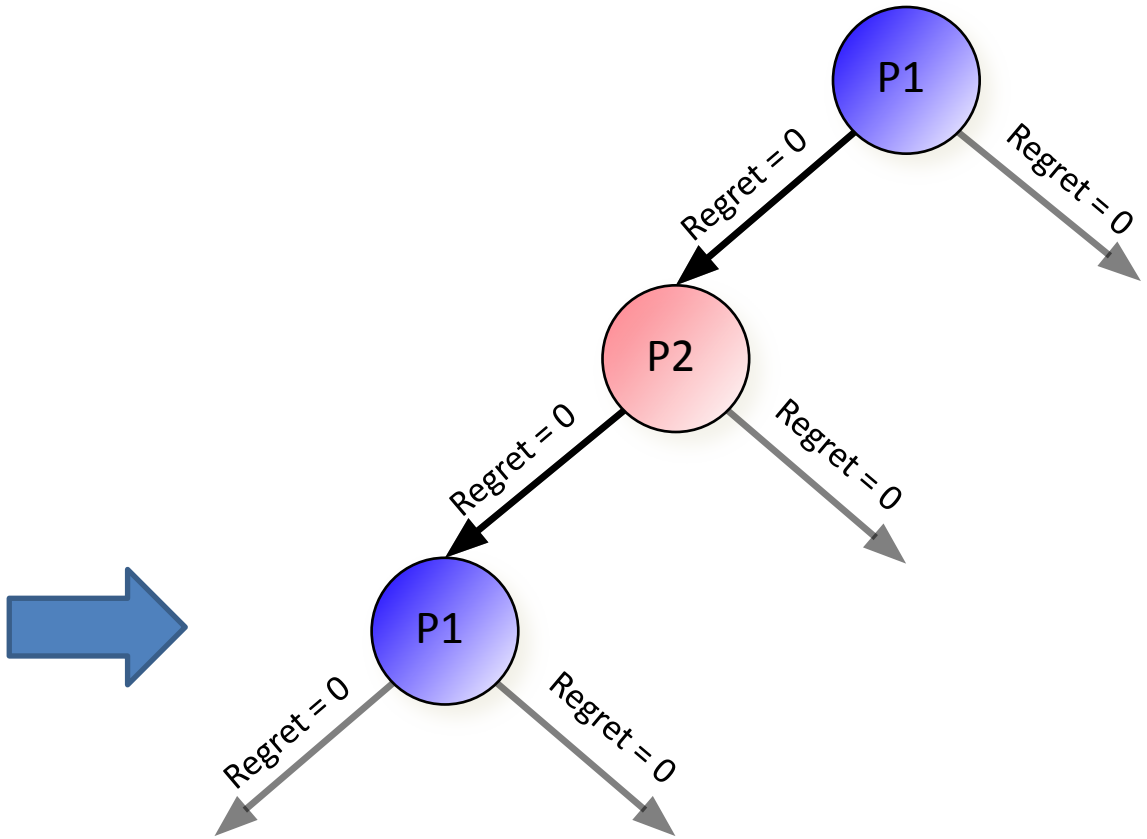
[Zinkevich *et al.* NeurIPS-07, Lanctot *et al.* NeurIPS-09]

Pick action proportional to **positive regret**



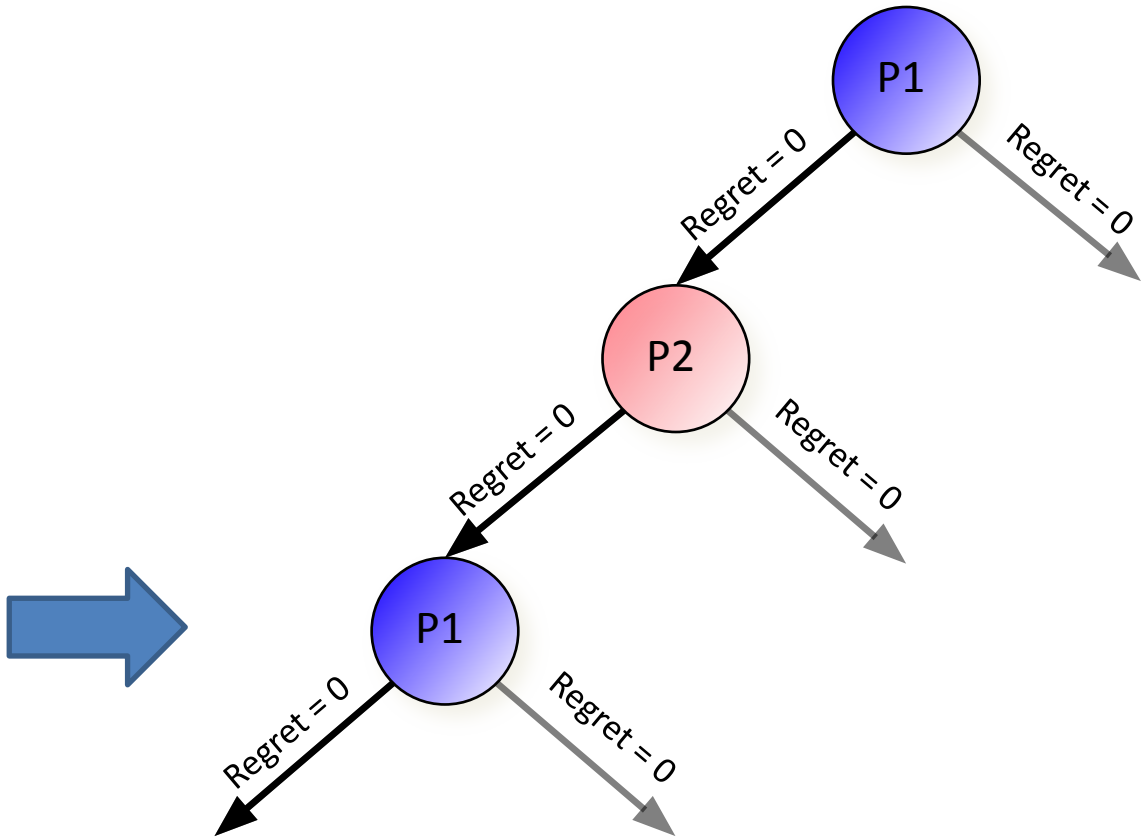
Monte Carlo Counterfactual Regret Minimization (MCCFR)

[Zinkevich *et al.* NeurIPS-07, Lanctot *et al.* NeurIPS-09]



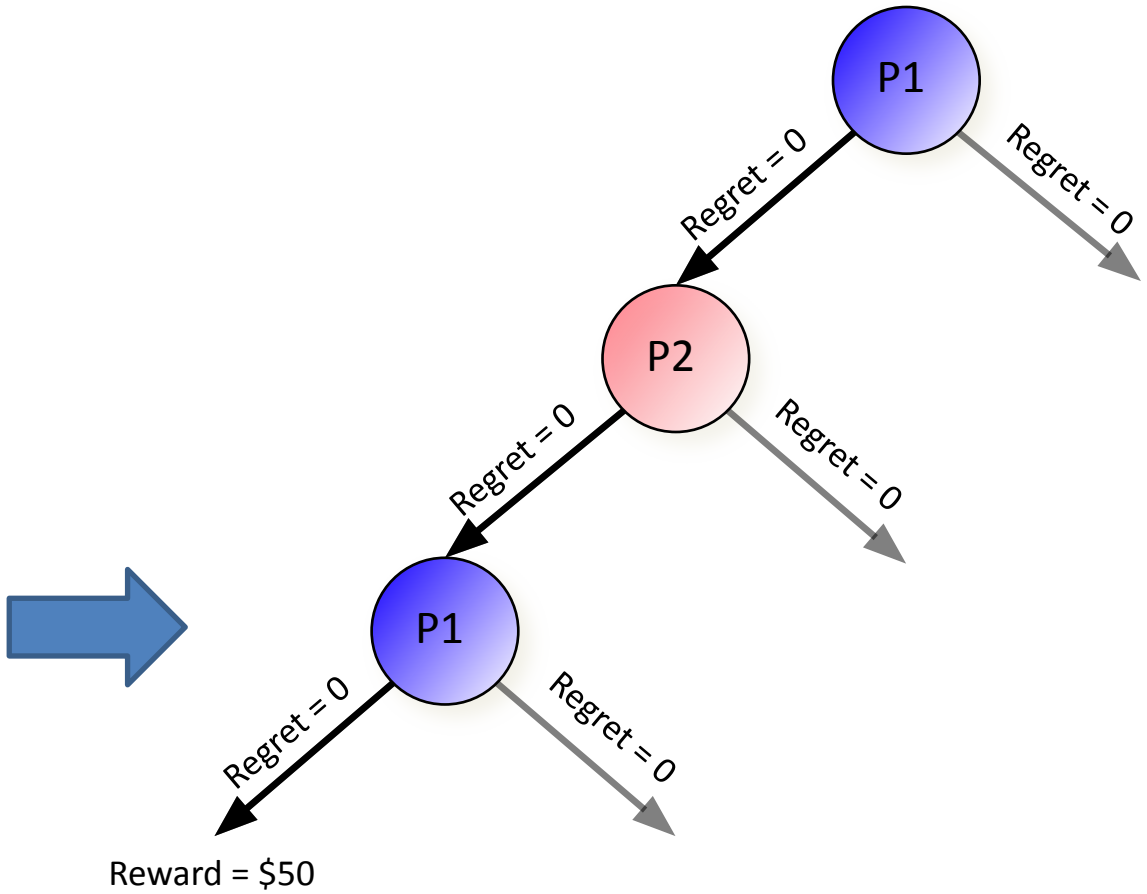
Monte Carlo Counterfactual Regret Minimization (MCCFR)

[Zinkevich *et al.* NeurIPS-07, Lanctot *et al.* NeurIPS-09]



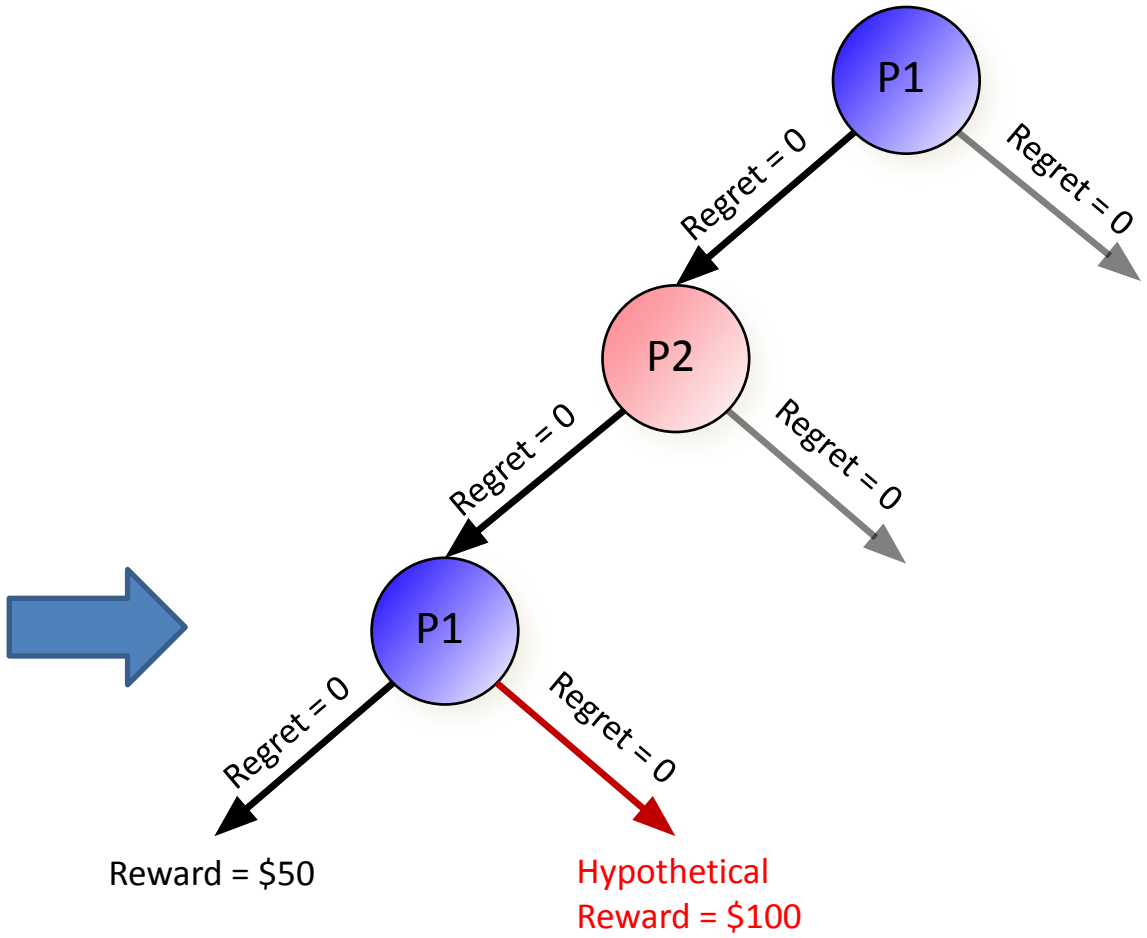
Monte Carlo Counterfactual Regret Minimization (MCCFR)

[Zinkevich *et al.* NeurIPS-07, Lanctot *et al.* NeurIPS-09]



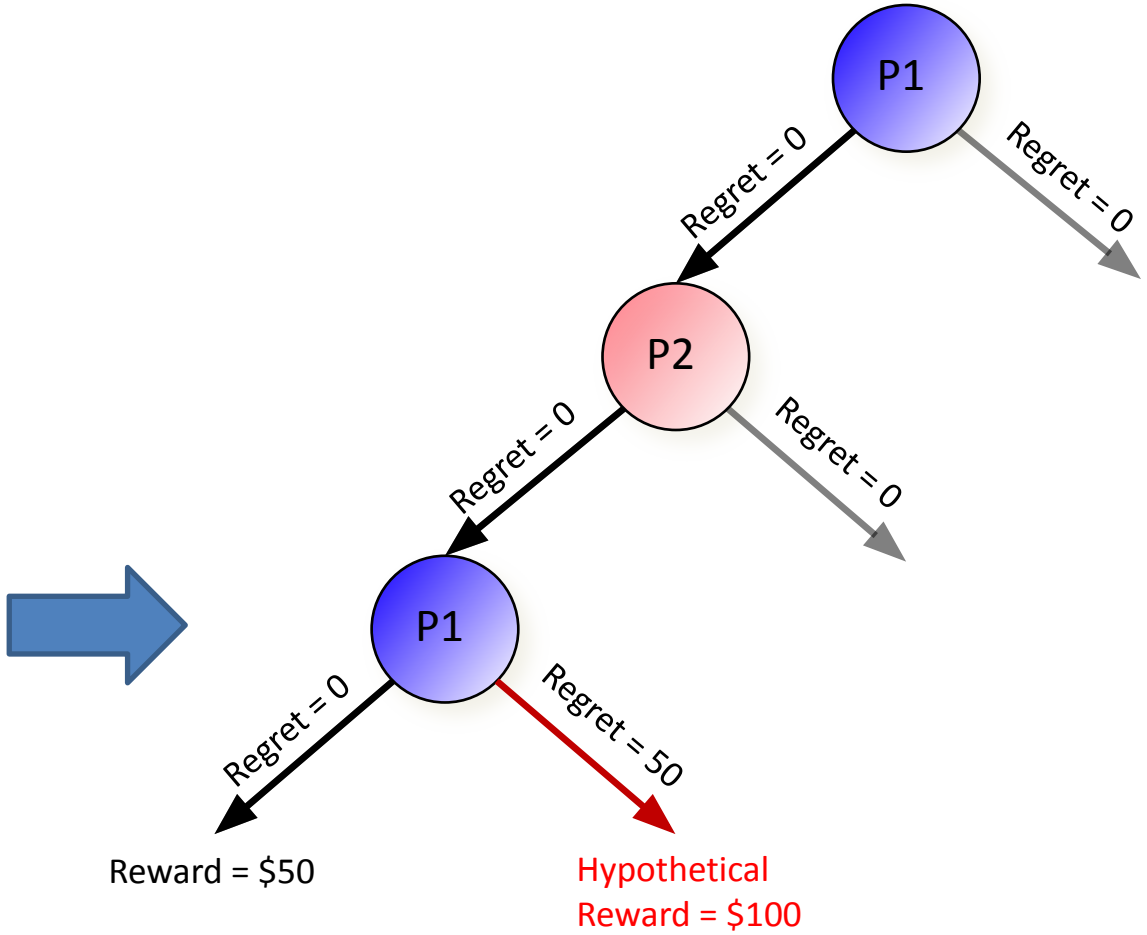
Monte Carlo Counterfactual Regret Minimization (MCCFR)

[Zinkevich *et al.* NeurIPS-07, Lanctot *et al.* NeurIPS-09]



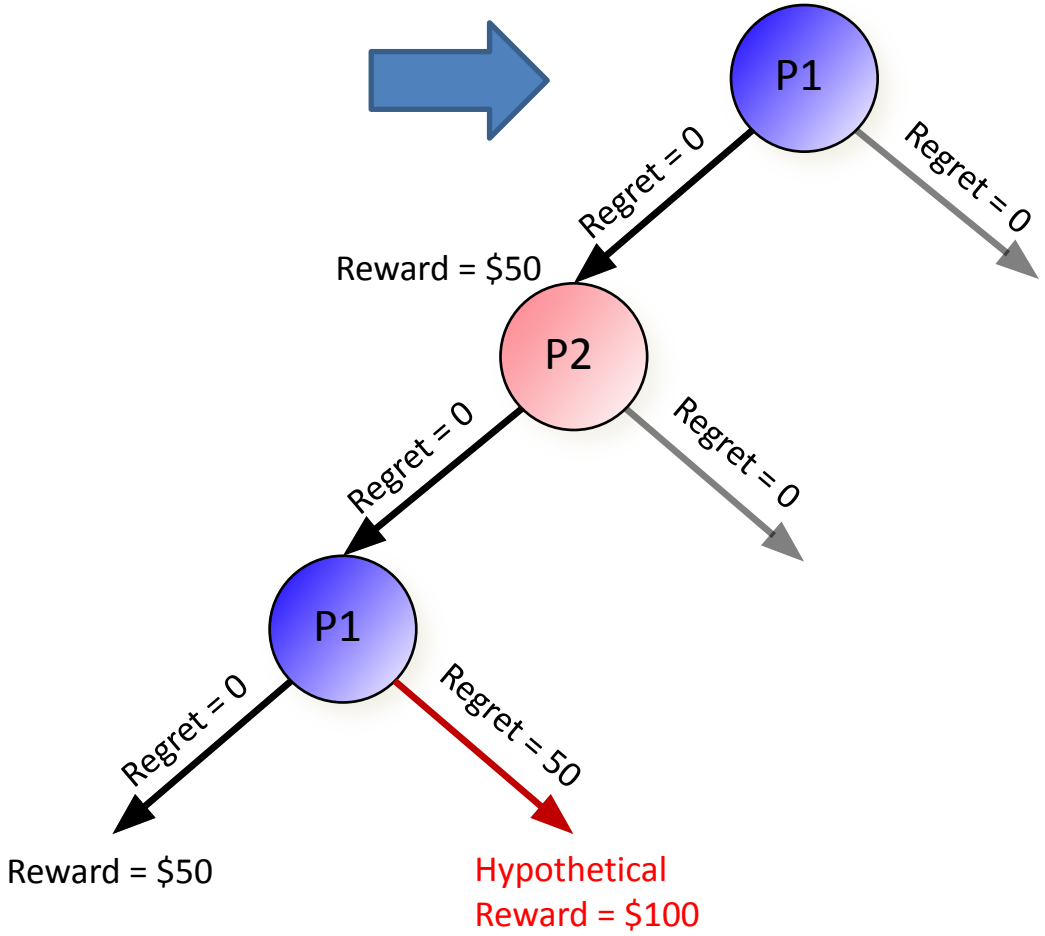
Monte Carlo Counterfactual Regret Minimization (MCCFR)

[Zinkevich *et al.* NeurIPS-07, Lanctot *et al.* NeurIPS-09]



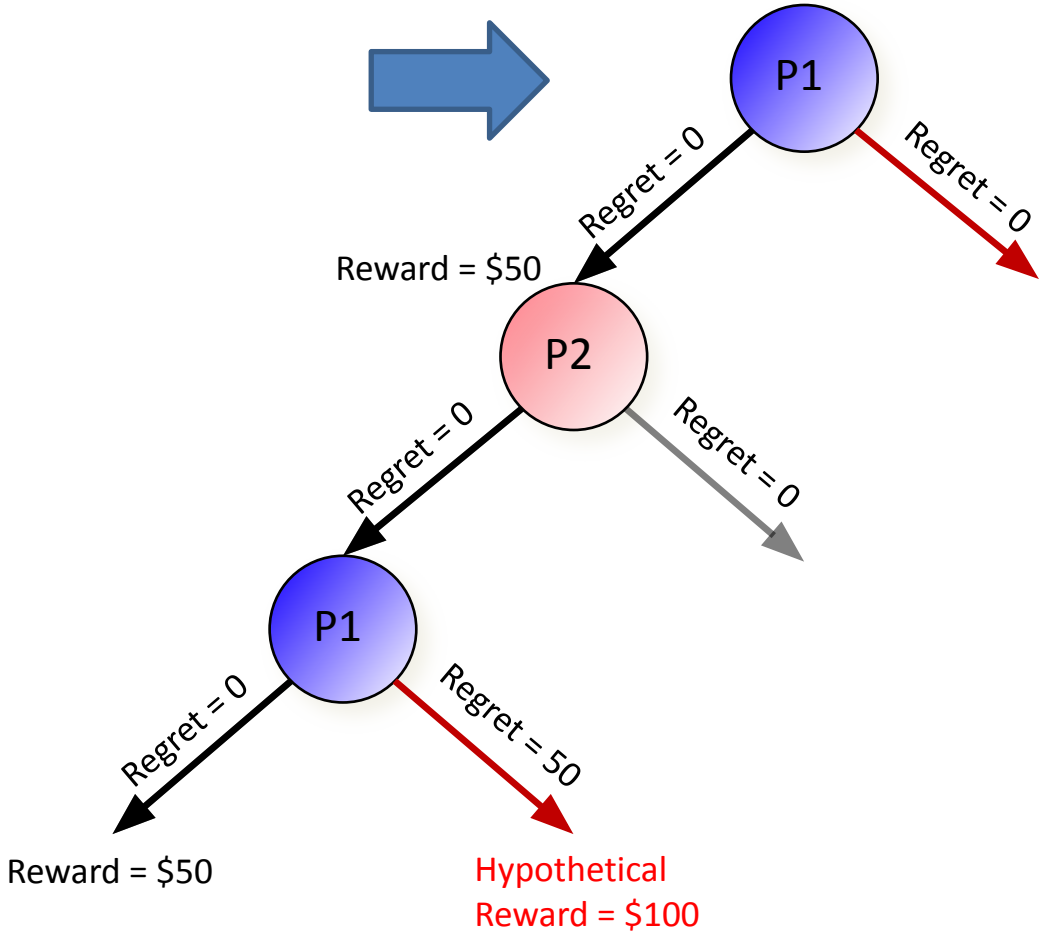
Monte Carlo Counterfactual Regret Minimization (MCCFR)

[Zinkevich *et al.* NeurIPS-07, Lanctot *et al.* NeurIPS-09]



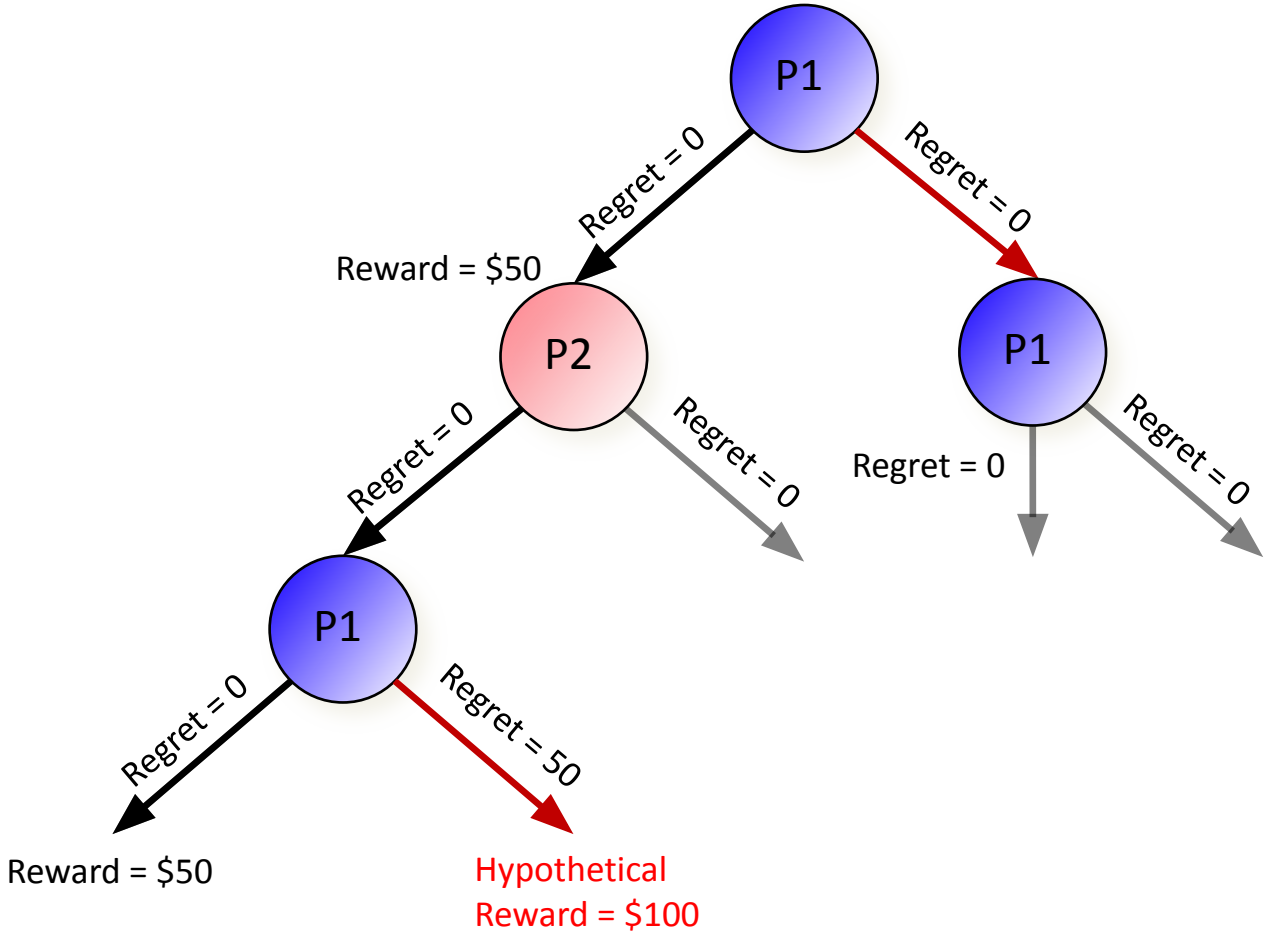
Monte Carlo Counterfactual Regret Minimization (MCCFR)

[Zinkevich *et al.* NeurIPS-07, Lanctot *et al.* NeurIPS-09]



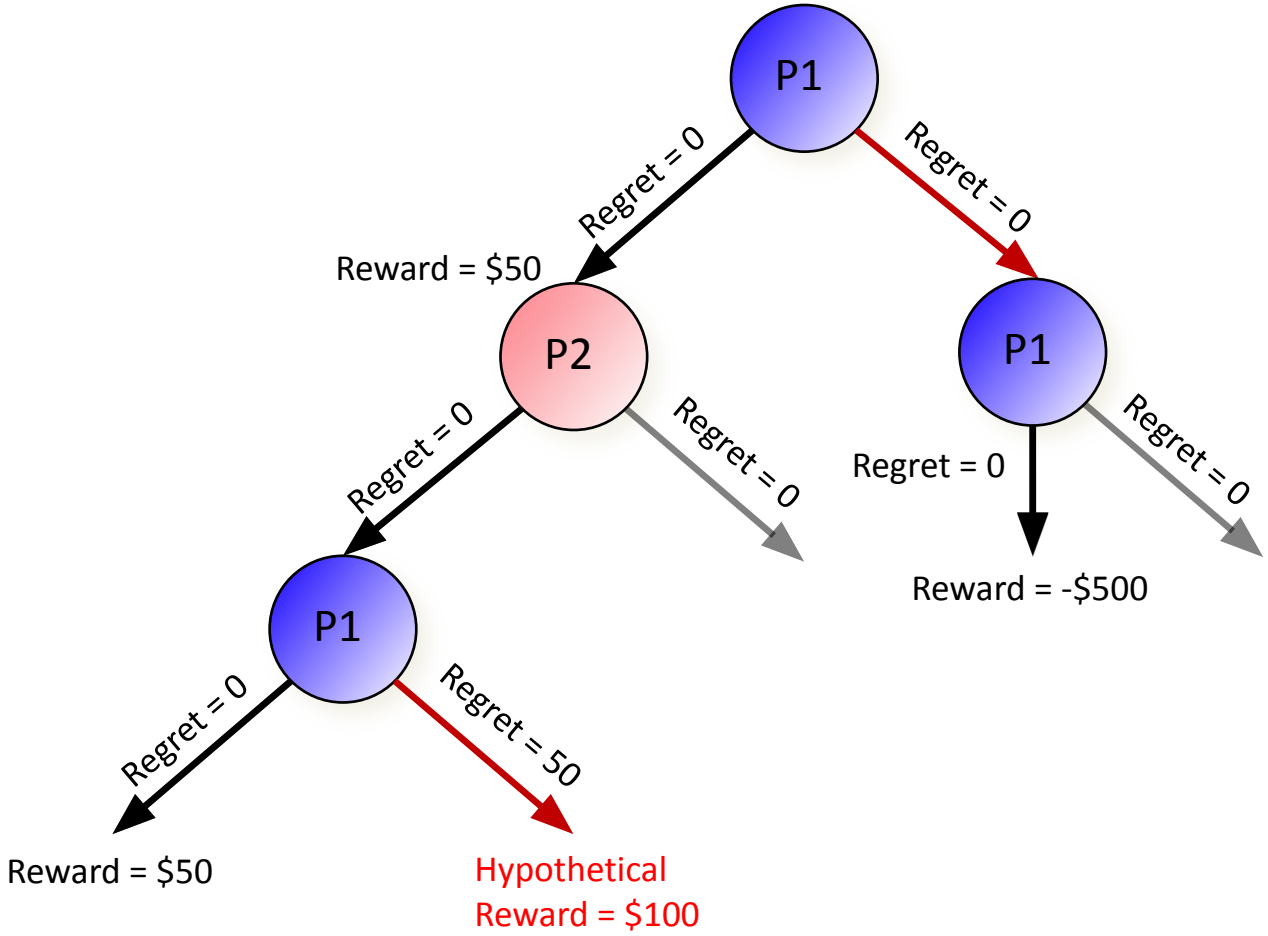
Monte Carlo Counterfactual Regret Minimization (MCCFR)

[Zinkevich *et al.* NeurIPS-07, Lanctot *et al.* NeurIPS-09]



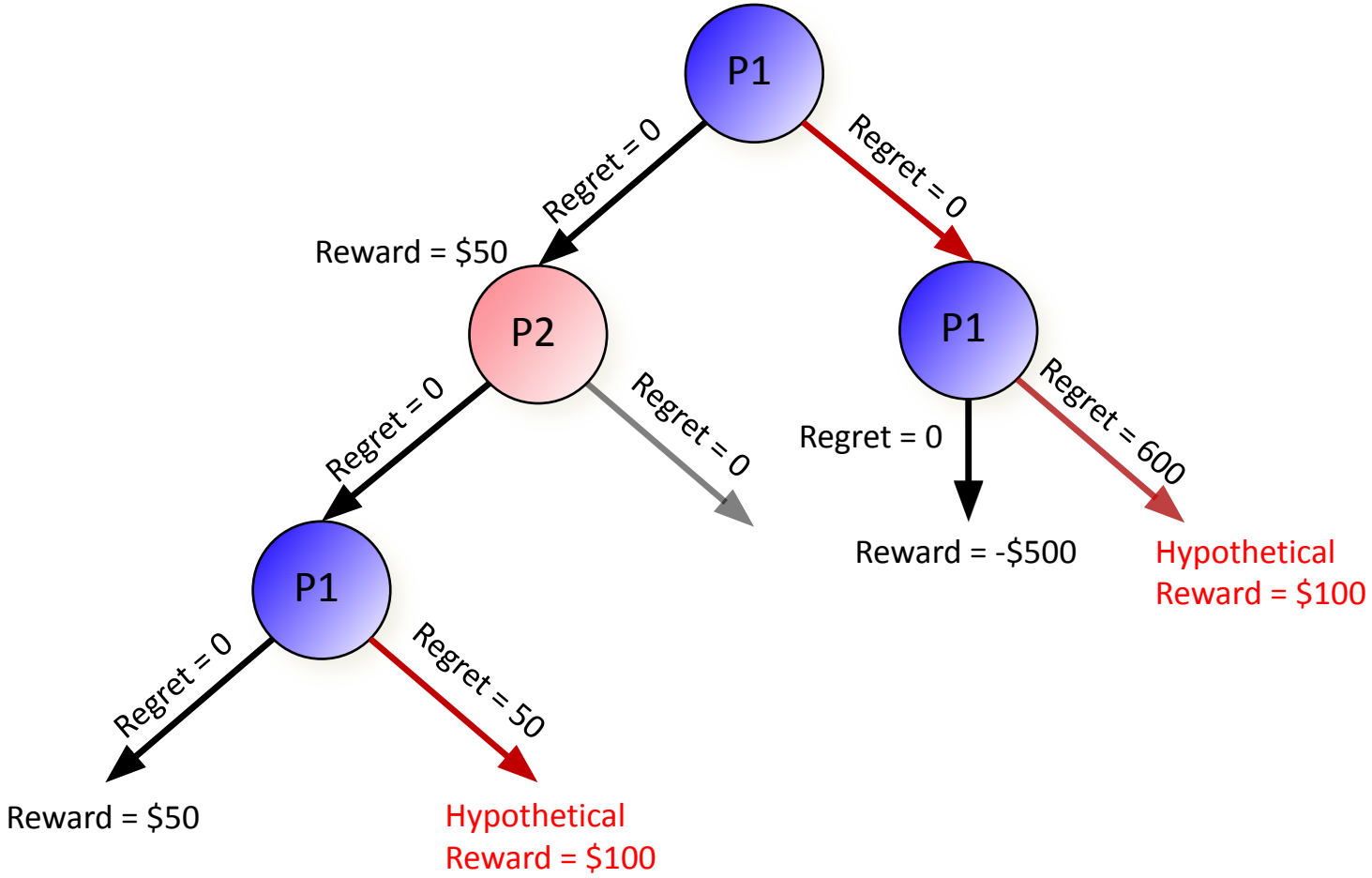
Monte Carlo Counterfactual Regret Minimization (MCCFR)

[Zinkevich *et al.* NeurIPS-07, Lanctot *et al.* NeurIPS-09]



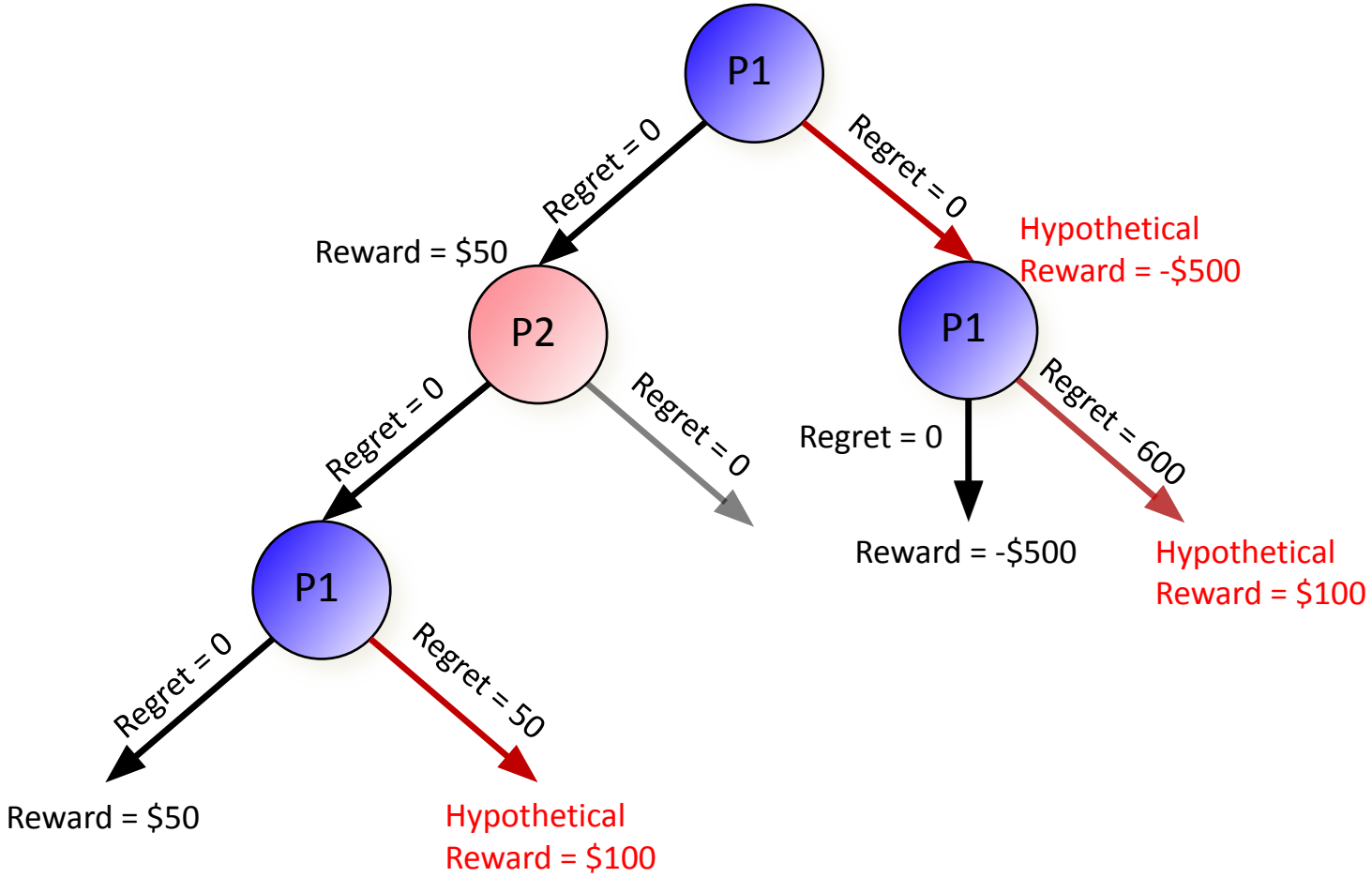
Monte Carlo Counterfactual Regret Minimization (MCCFR)

[Zinkevich *et al.* NeurIPS-07, Lanctot *et al.* NeurIPS-09]



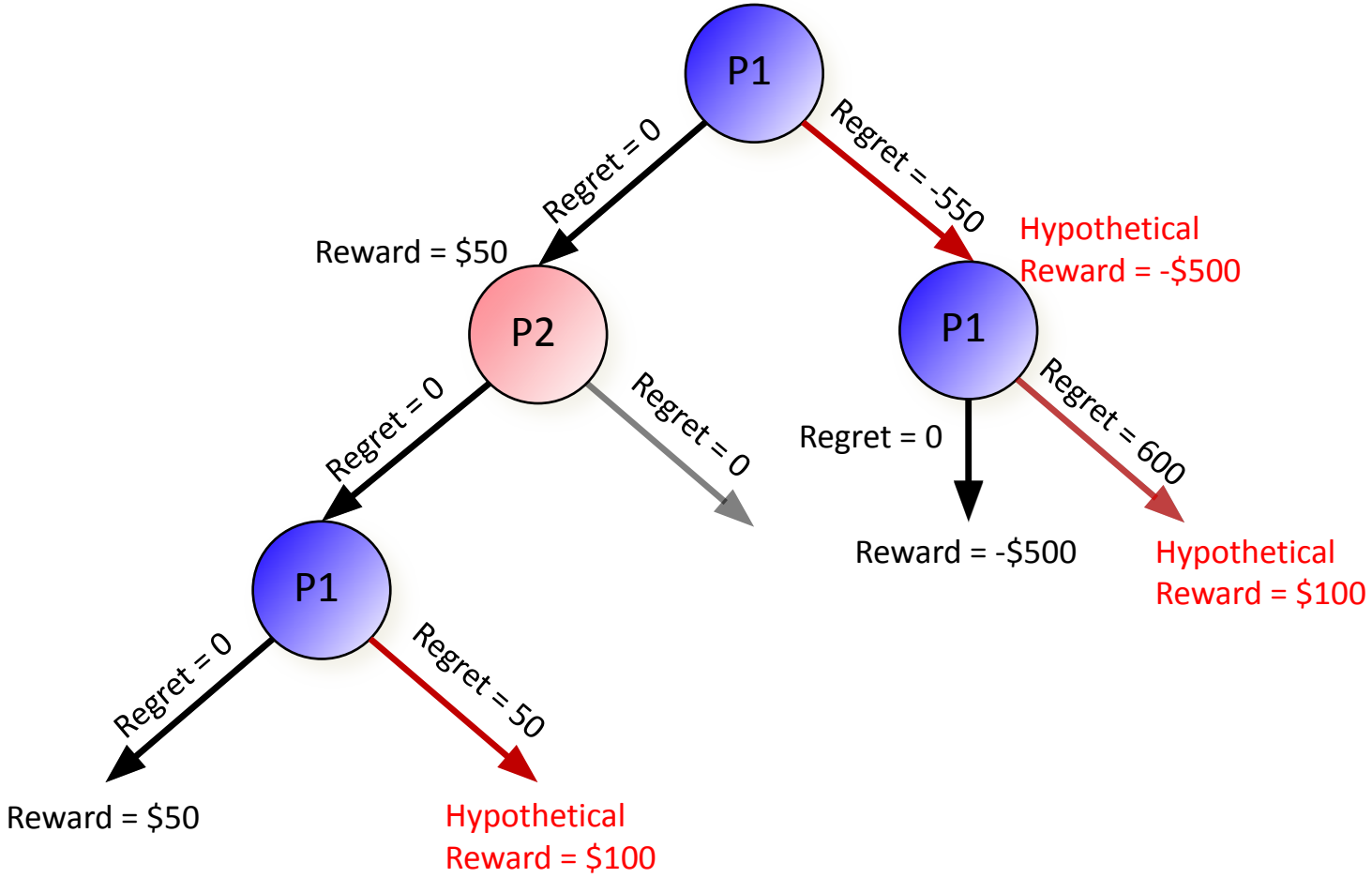
Monte Carlo Counterfactual Regret Minimization (MCCFR)

[Zinkevich *et al.* NeurIPS-07, Lanctot *et al.* NeurIPS-09]



Monte Carlo Counterfactual Regret Minimization (MCCFR)

[Zinkevich *et al.* NeurIPS-07, Lanctot *et al.* NeurIPS-09]



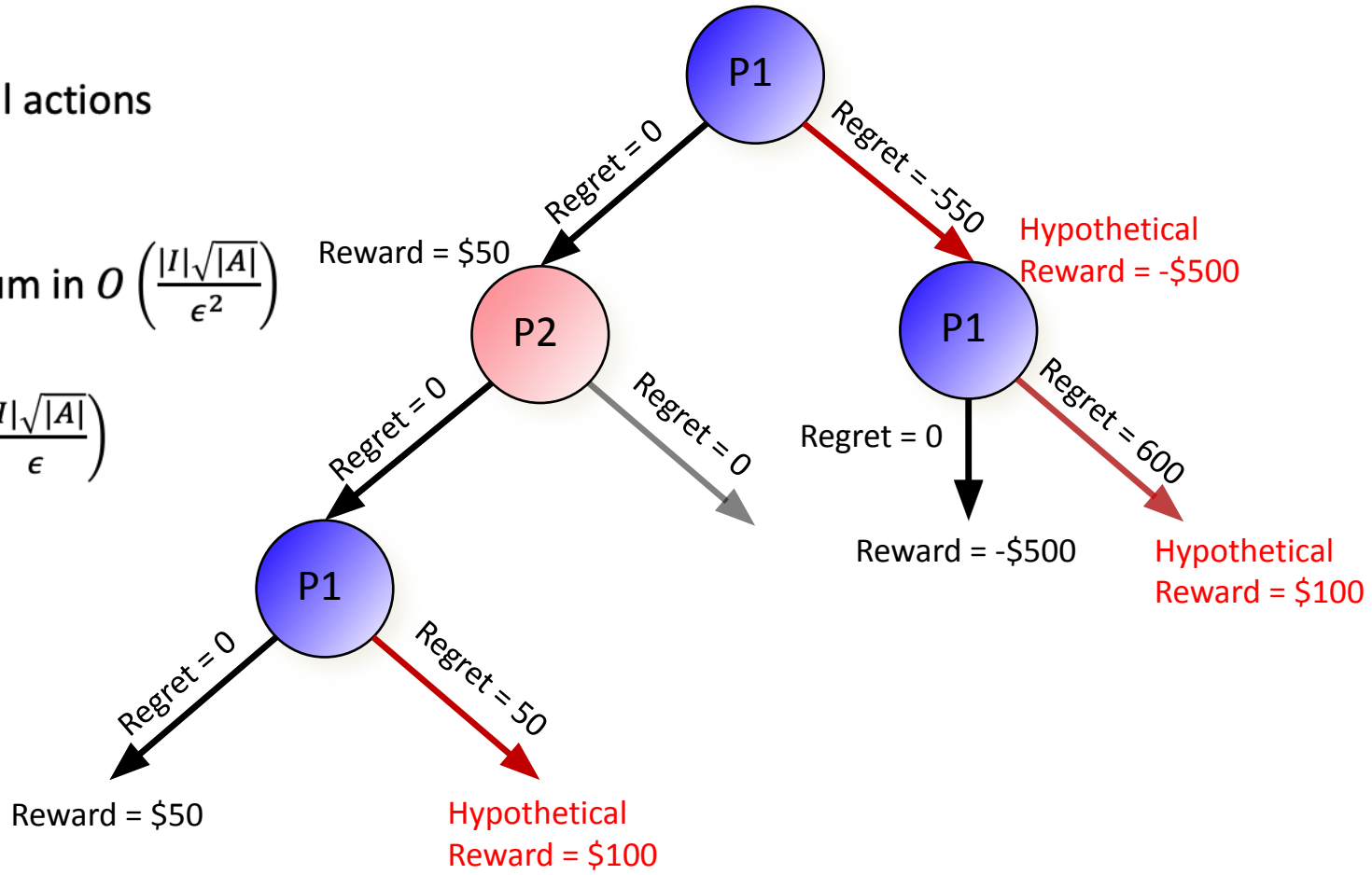
Counterfactual Regret Minimization (CFR)

[Zinkevich *et al.* NeurIPS-07]

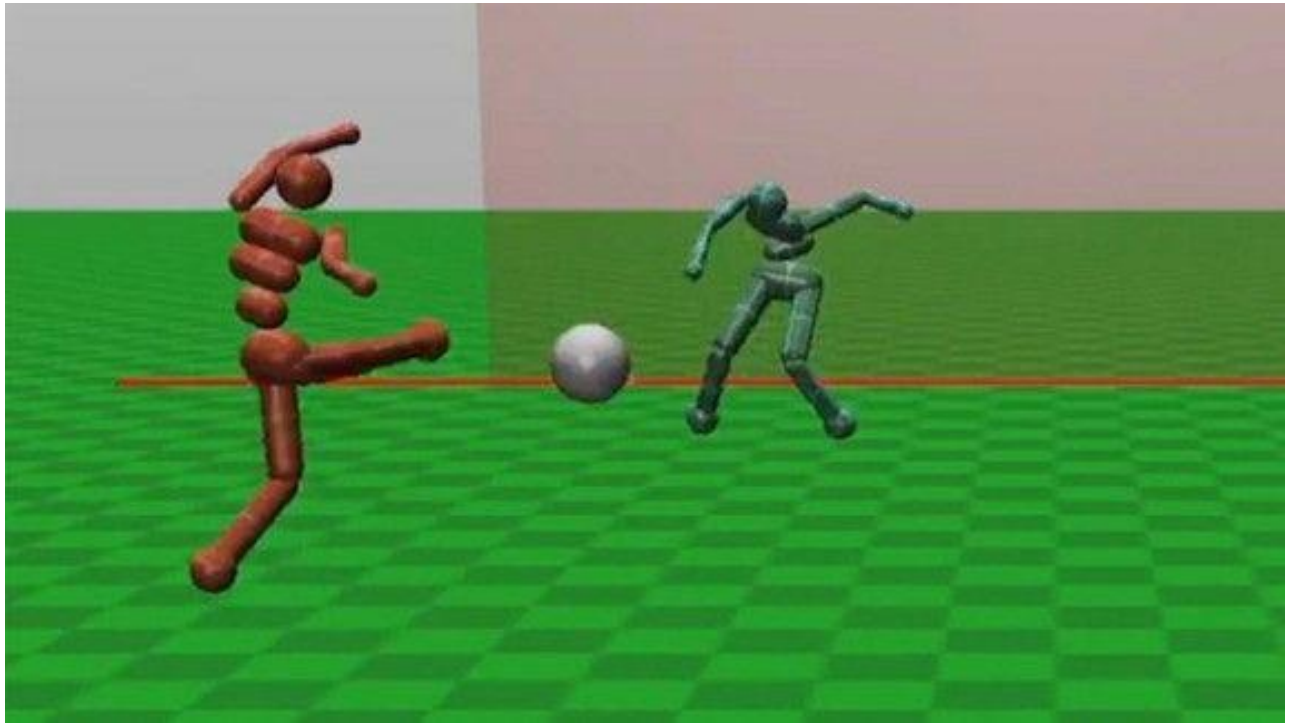
Similar, but takes the EV over all actions rather than sampling

Average converges to equilibrium in $O\left(\frac{|I|\sqrt{|A|}}{\epsilon^2}\right)$

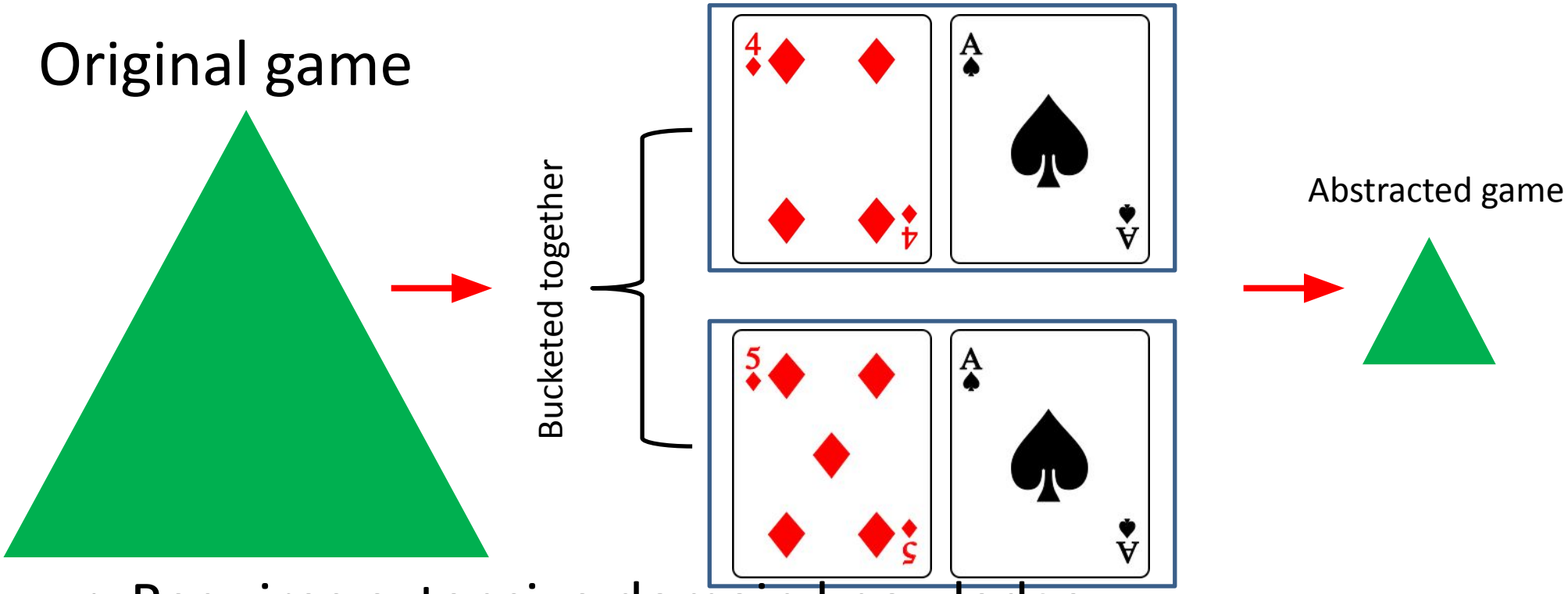
But in practice converge in $O\left(\frac{|I|\sqrt{|A|}}{\epsilon}\right)$



Extending CFR to Large Games



Prior Approach: Abstraction in Games

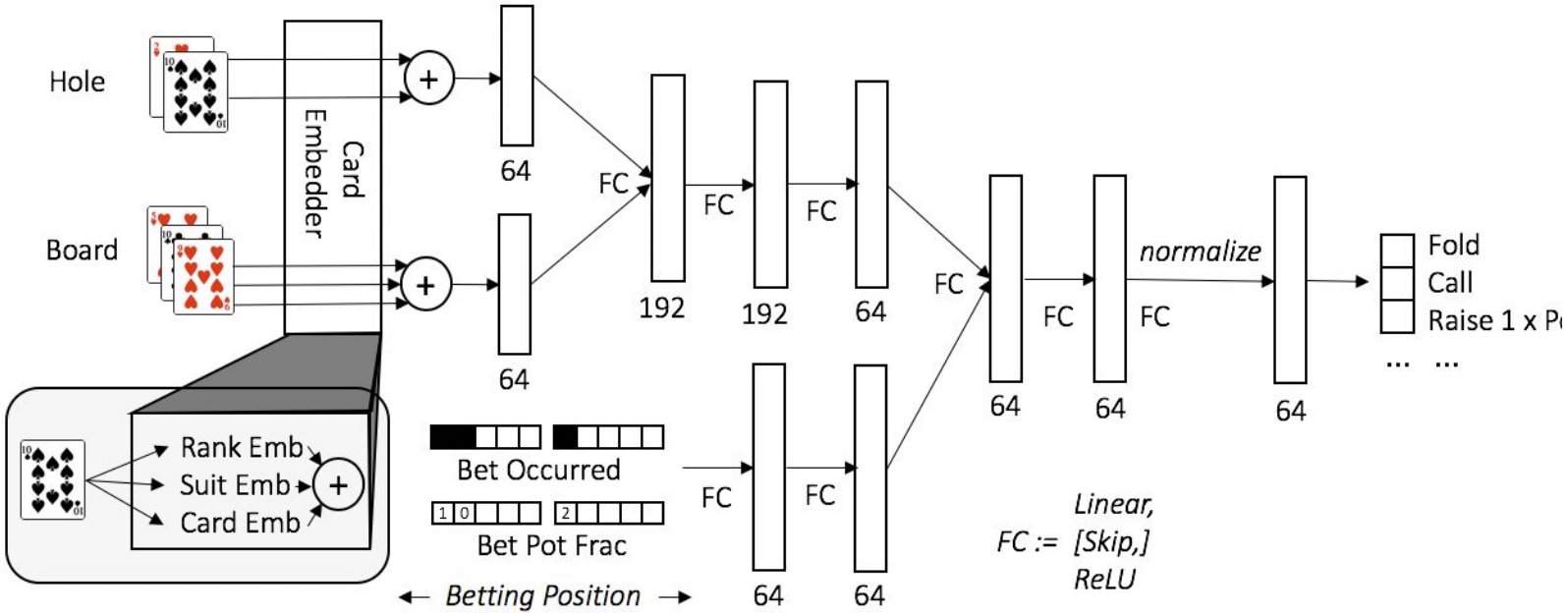


- Requires extensive domain knowledge
  - Several papers written on how to do abstraction just in poker [Johanson et al. AAMAS-13, Ganzfried & Sandholm AAAI-14]
  - Difficult to extend to other games

Deep CFR / DREAM

[Brown et al. ICML-19; Steinberger, Lerer, Brown arXiv-20]

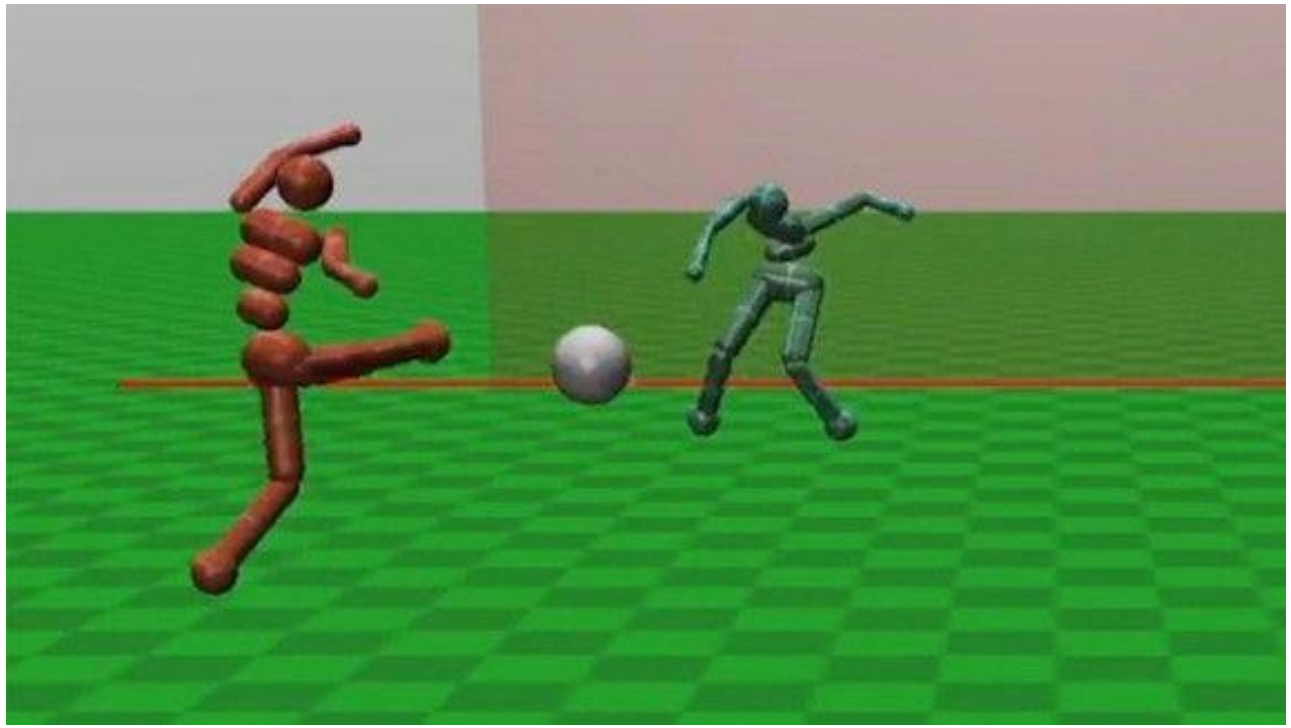
- Replaces abstraction with neural network approximation of regrets
- Deep CFR / DREAM require far less domain knowledge



Deep CFR / DREAM

[Brown et al. ICML-19; Steinberger, Lerer, Brown arXiv-20]

- Replaces abstraction with neural network approximation of regrets
- Deep CFR / DREAM require far less domain knowledge



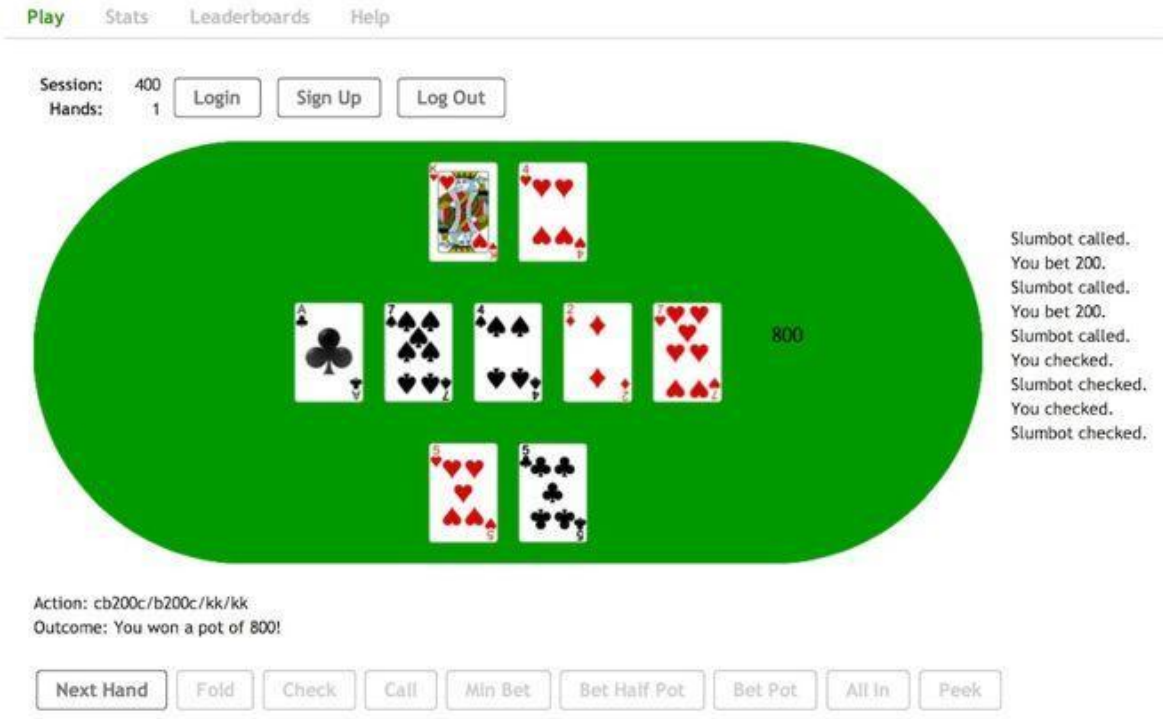
Searching for a better strategy in real time

Image Credit: UC Berkeley CS-188 Lecture 6

Annual Computer Poker Competition

- Each year, research labs would make poker bots and play them against each other.

- It turned into a competition of scaling models:
  - **2012:** 5,000 buckets
  - **2013:** 30,000 buckets
  - **2014:** 90,000 buckets
  - **2015:** 600,000 buckets

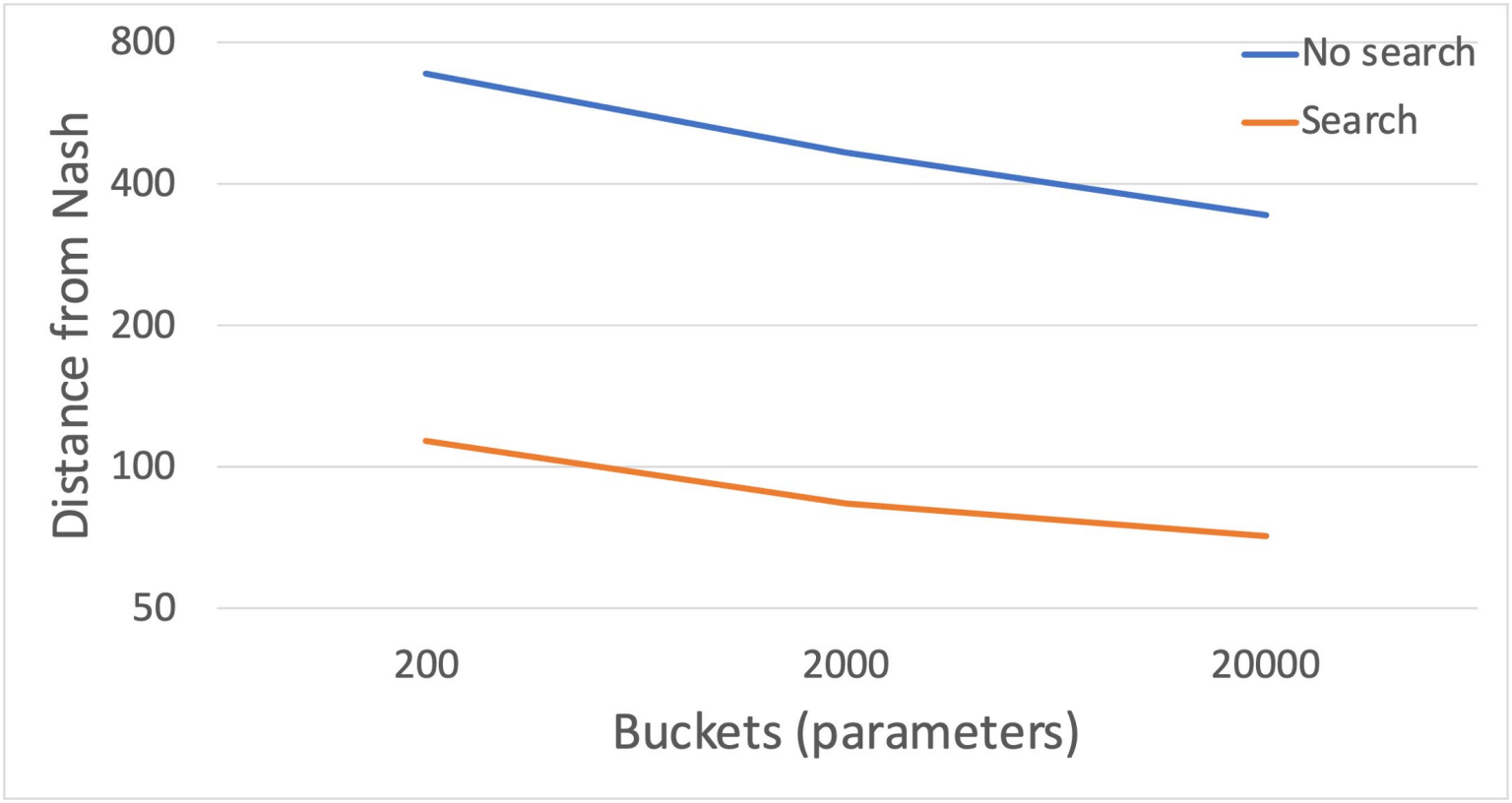


2015 Brains vs. AI Poker Competition

- In 2015 we (CMU) challenged 4 top poker pros to an 80,000-hand poker competition
- $120,000 in prize money to incentivize them
- Our bot (Claudico) lost by 9.1 bb/100

The importance of search in poker

Scaling laws in a medium-sized poker game [1]



[1] “Safe and Nested Subgame Solving in Imperfect-Information Games.” Brown & Sandholm. NeurIPS 2017 Best Paper.

2017 Brains vs AI Two-Player Poker AI

[Brown & Sandholm Science-17]

- Libratus against 4 top poker pros



- 120,000 hands of poker
- $200,000 in prize money
- **Won by 15 bb/100** (Claudico had lost by 9 bb/100)
  - P-value ≈ 0.0002
- Each human lost individually to Libratus

Why is search in imperfect-information games hard?

Because “states” as traditionally defined don’t have well-defined values

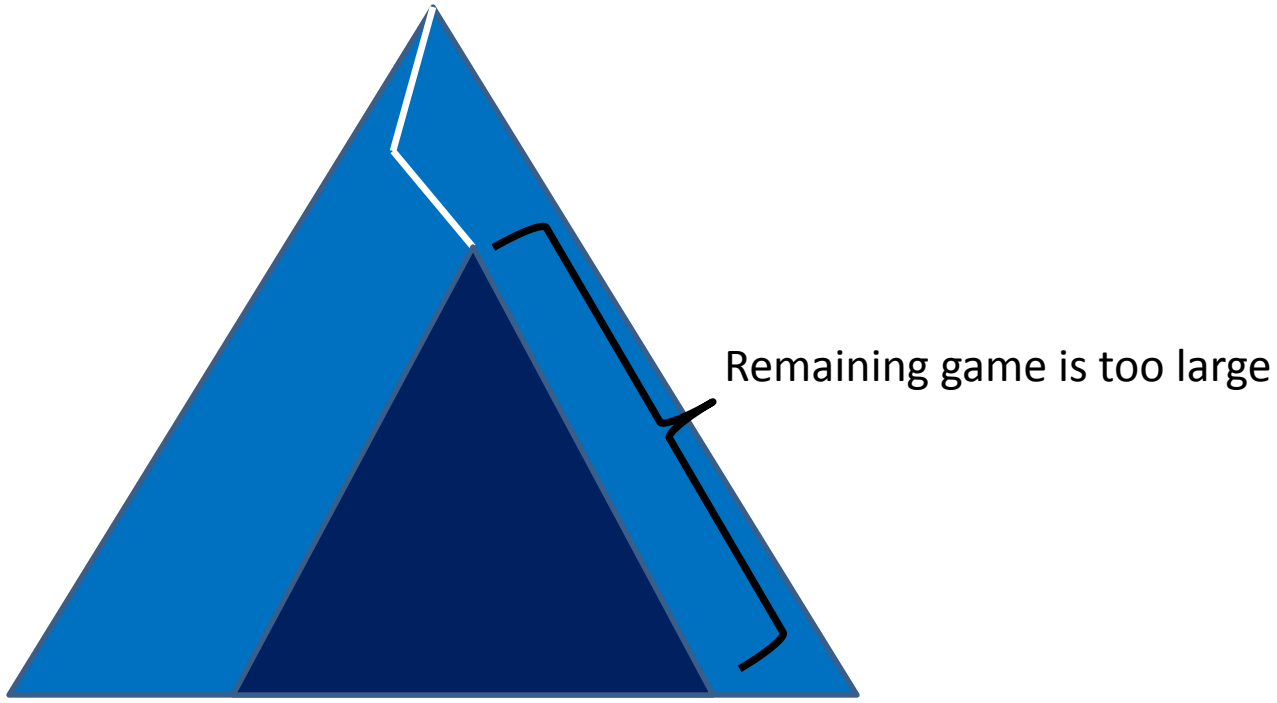
Search in Perfect-Information Games

- In perfect-information games, the **value of a state** is the **unique** value of both players playing optimally from that point forward

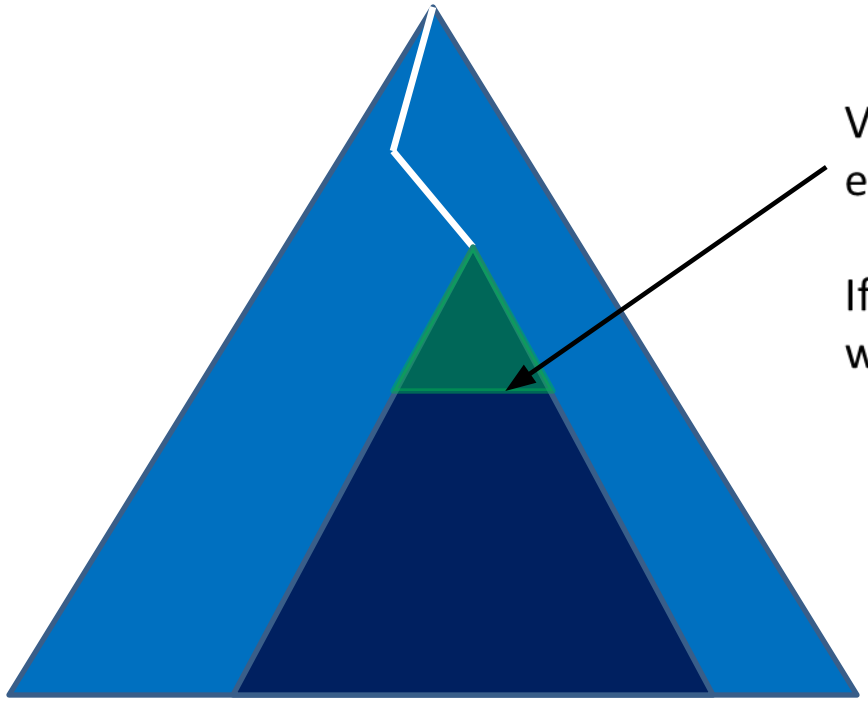
- A **value network** takes a state as input and outputs an estimate of the state value

$$f_{white}(\quad) = 1$$

Search in Perfect-Information Games



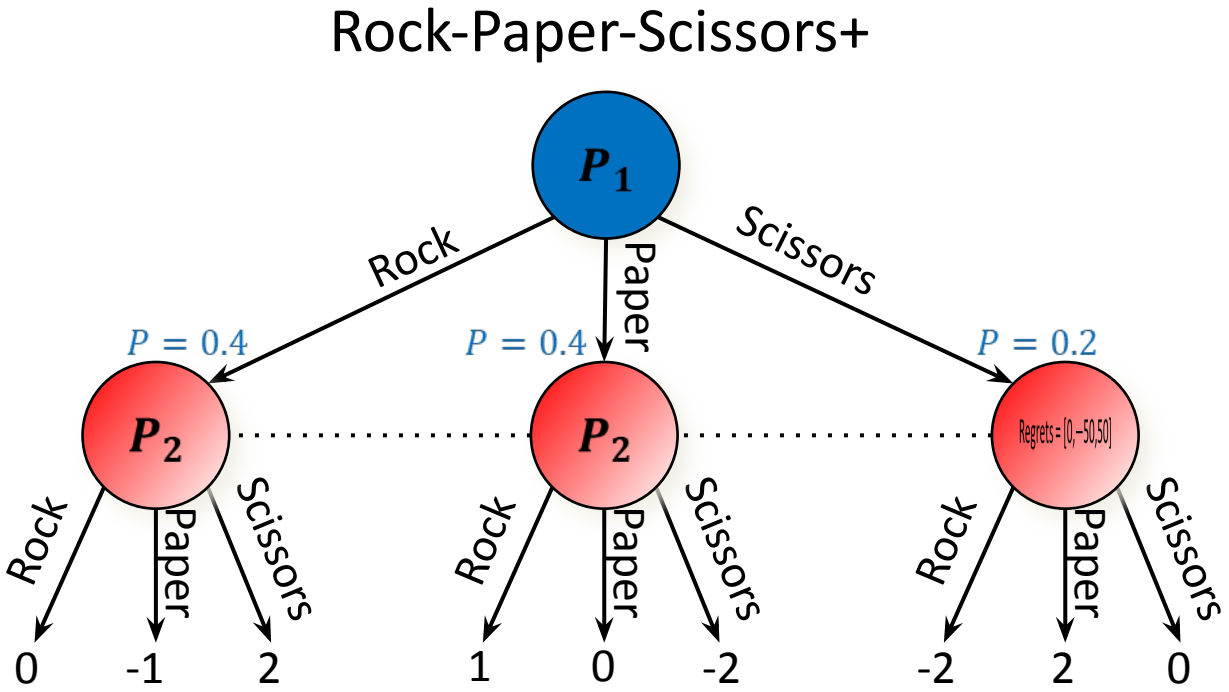
Search in Perfect-Information Games



Value substituted at leaf node s is estimate of equilibrium value $\hat{v}(s)$

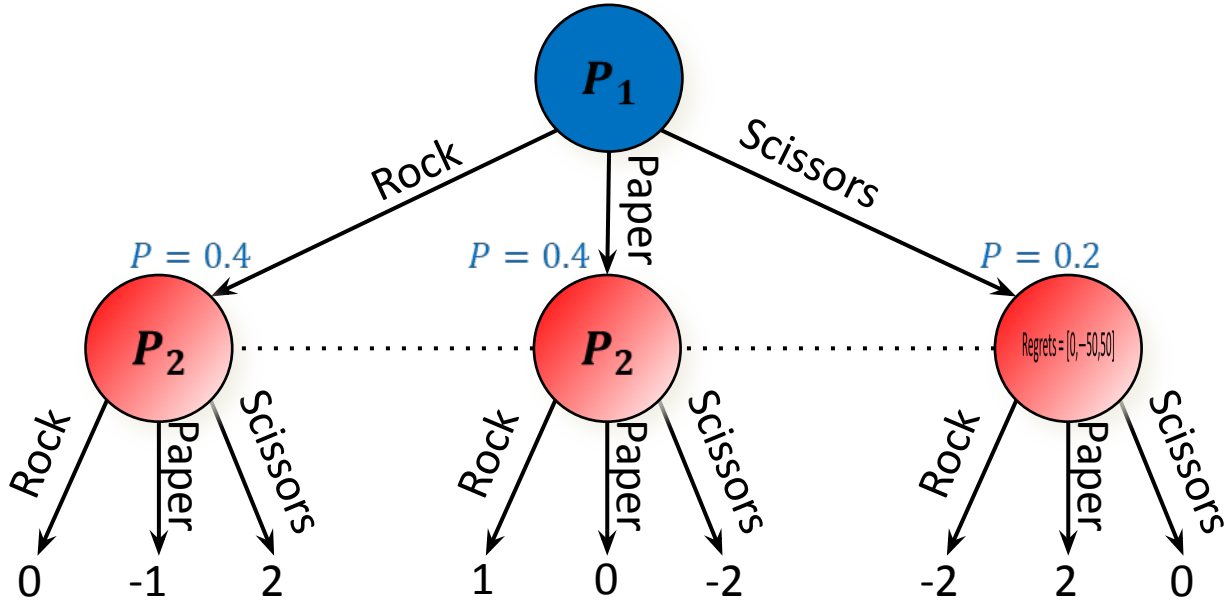
If estimate is perfect, depth-limited search will find optimal strategy

Depth-Limited Search

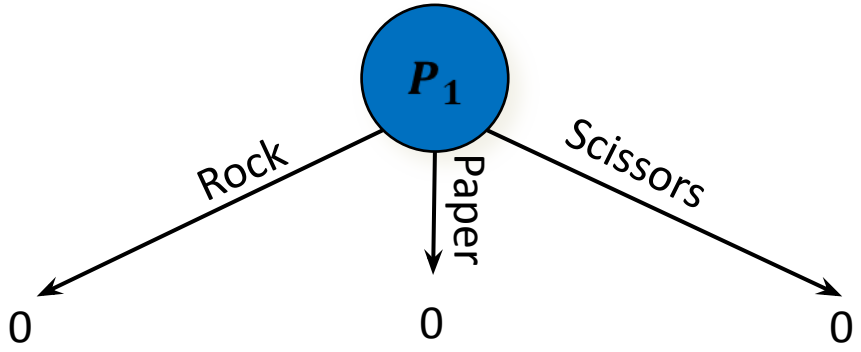


Depth-Limited Search

Rock-Paper-Scissors+

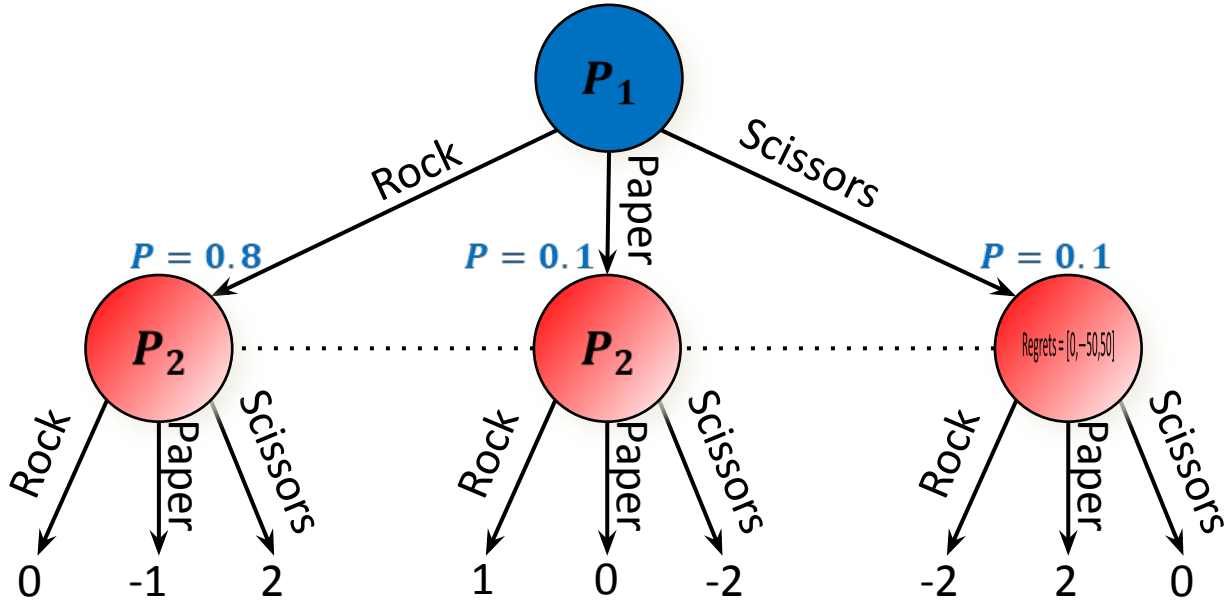


Depth-Limited Rock-Paper-Scissors+

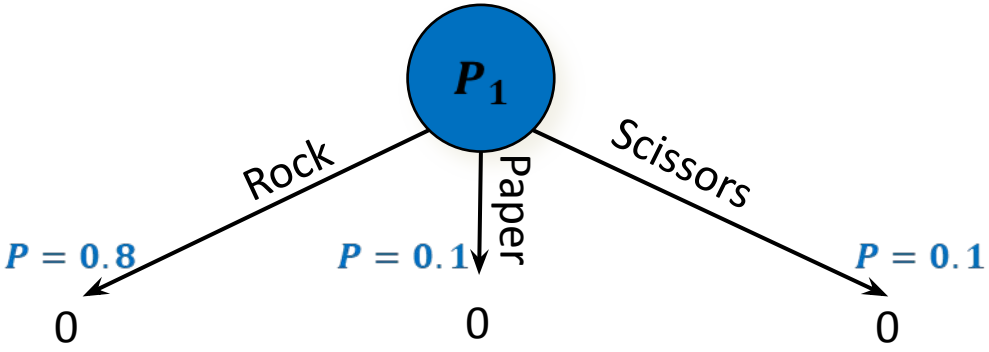


Depth-Limited Search

Rock-Paper-Scissors+

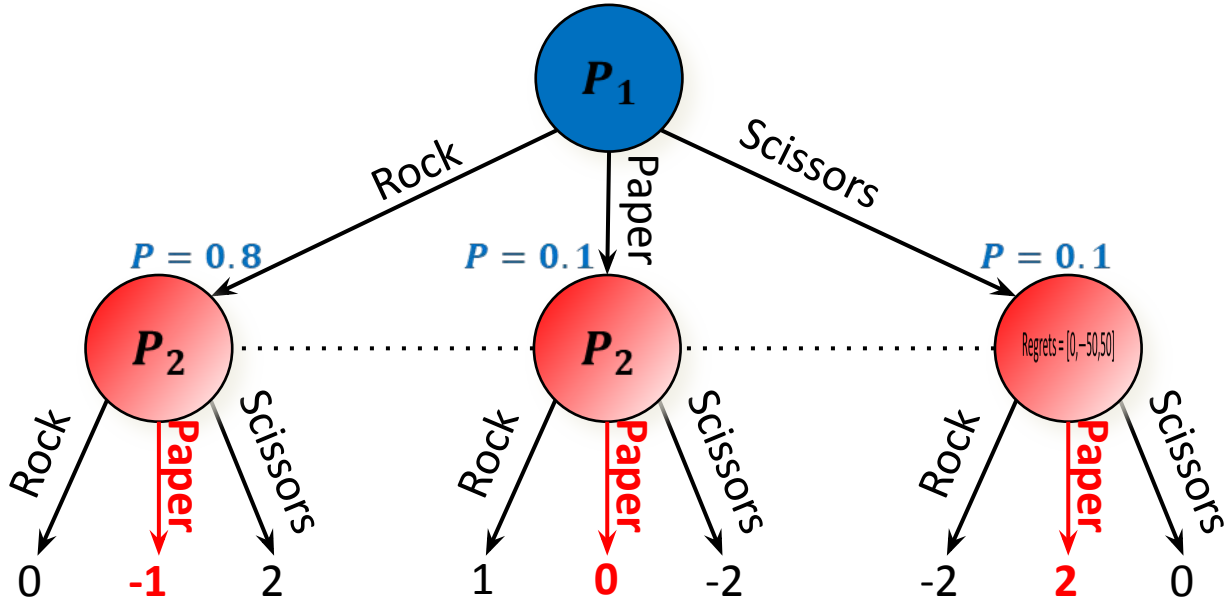


Depth-Limited Rock-Paper-Scissors+

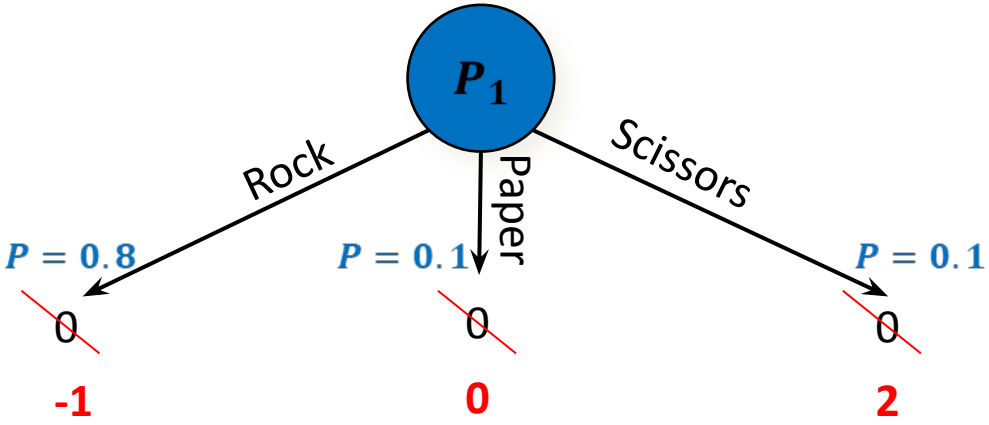


Depth-Limited Search

Rock-Paper-Scissors+

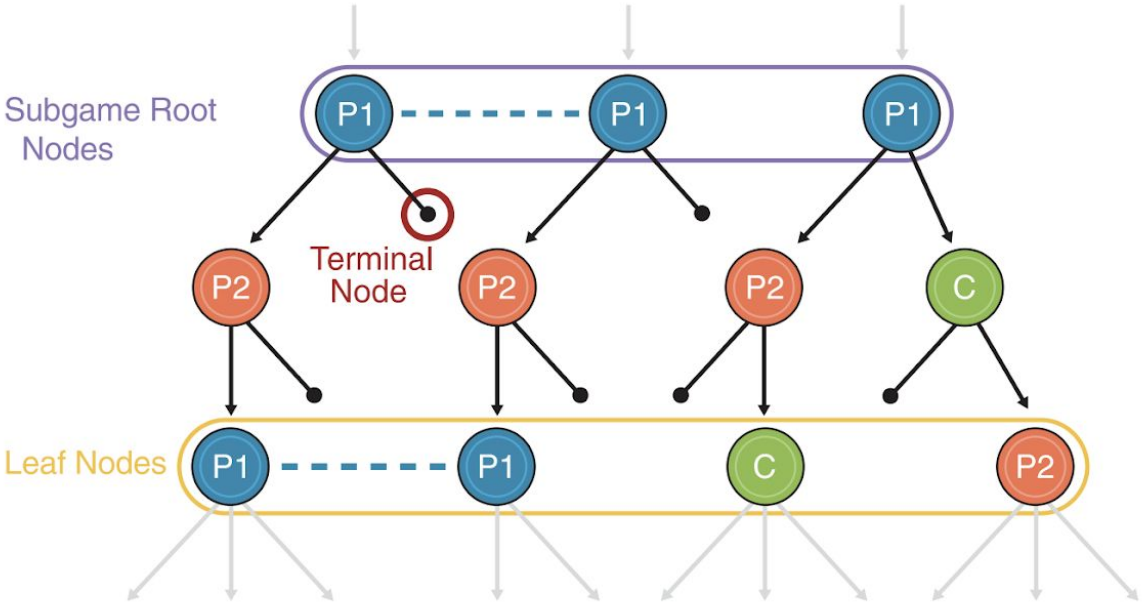


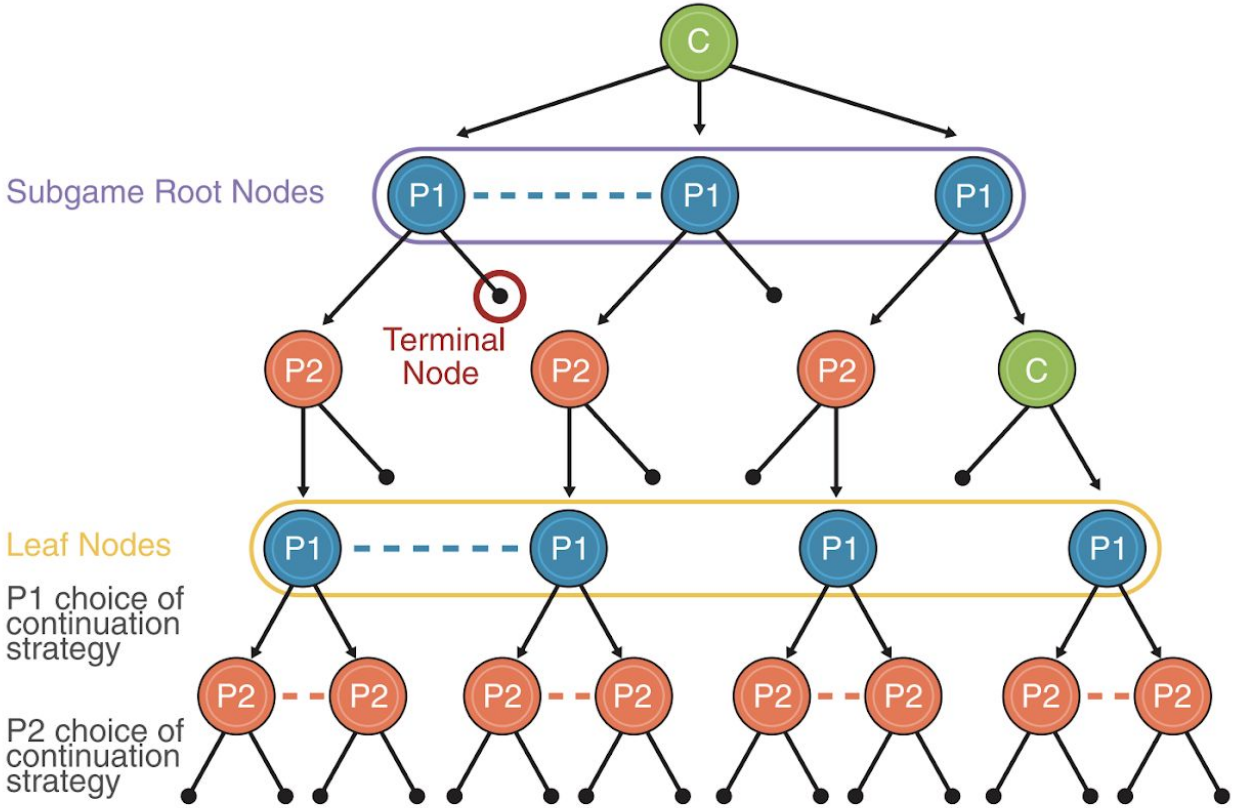
Depth-Limited Rock-Paper-Scissors+



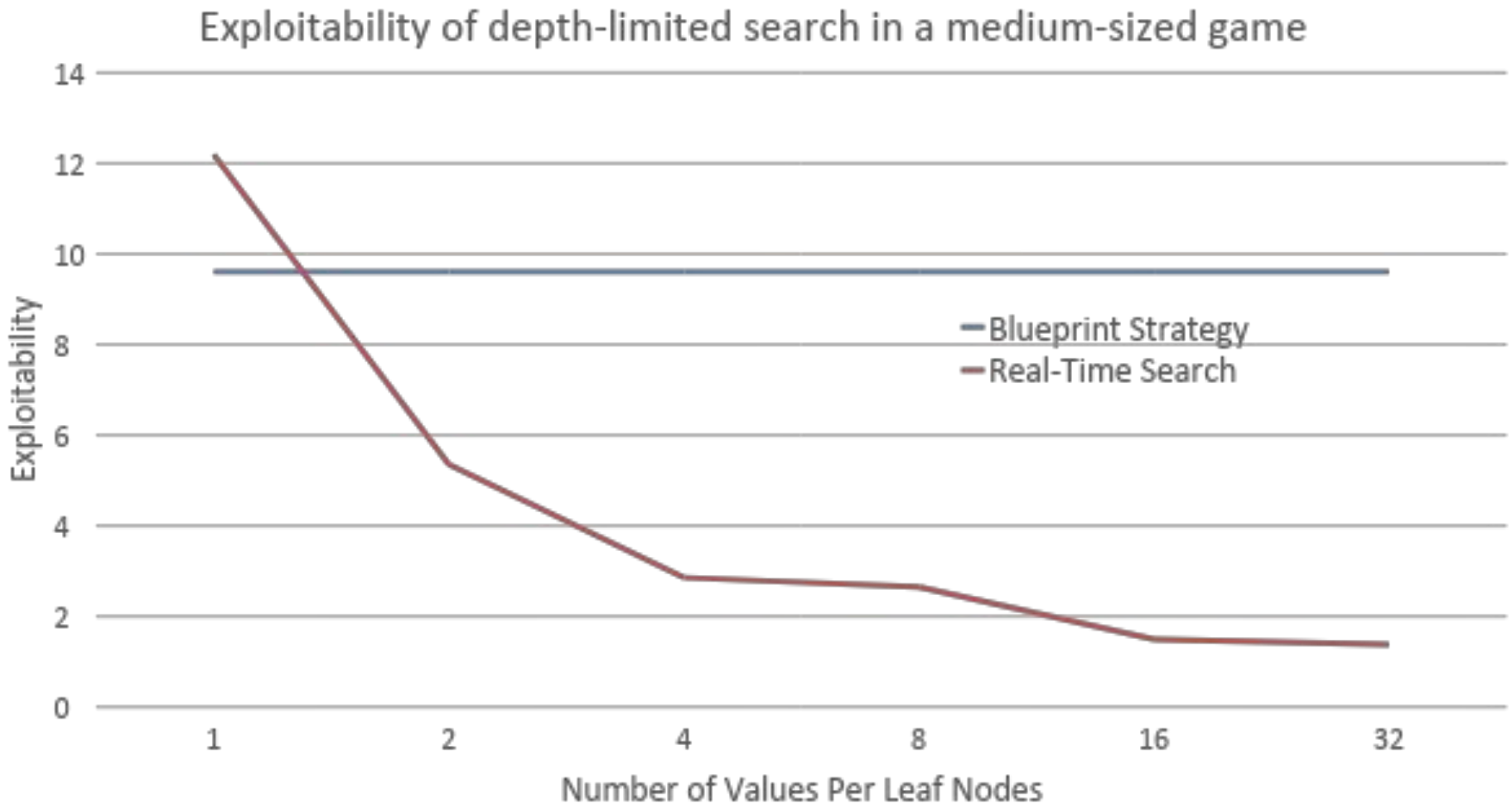
Depth-Limited Search in Pluribus

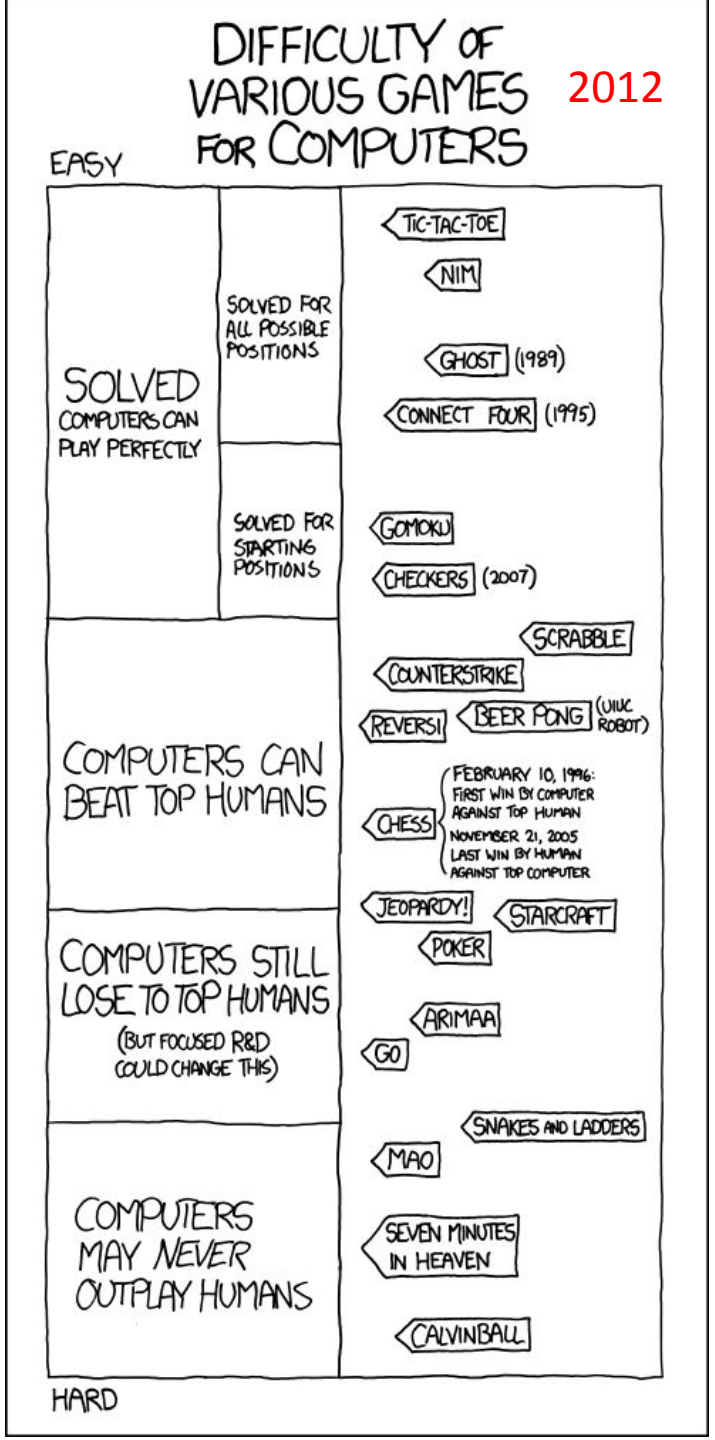
[Brown, Sandholm, Amos NeurIPS-18; Brown & Sandholm Science-19]

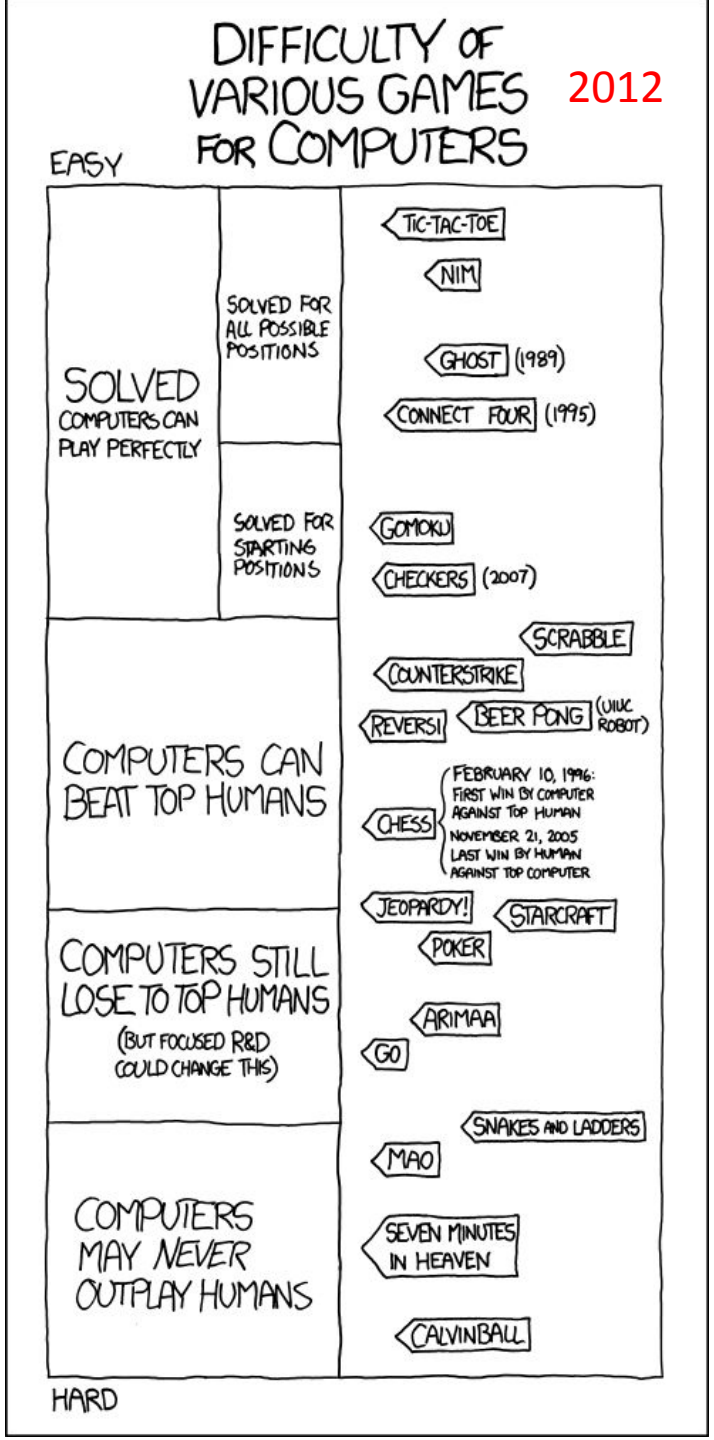


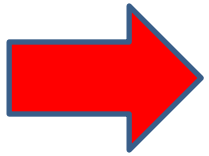


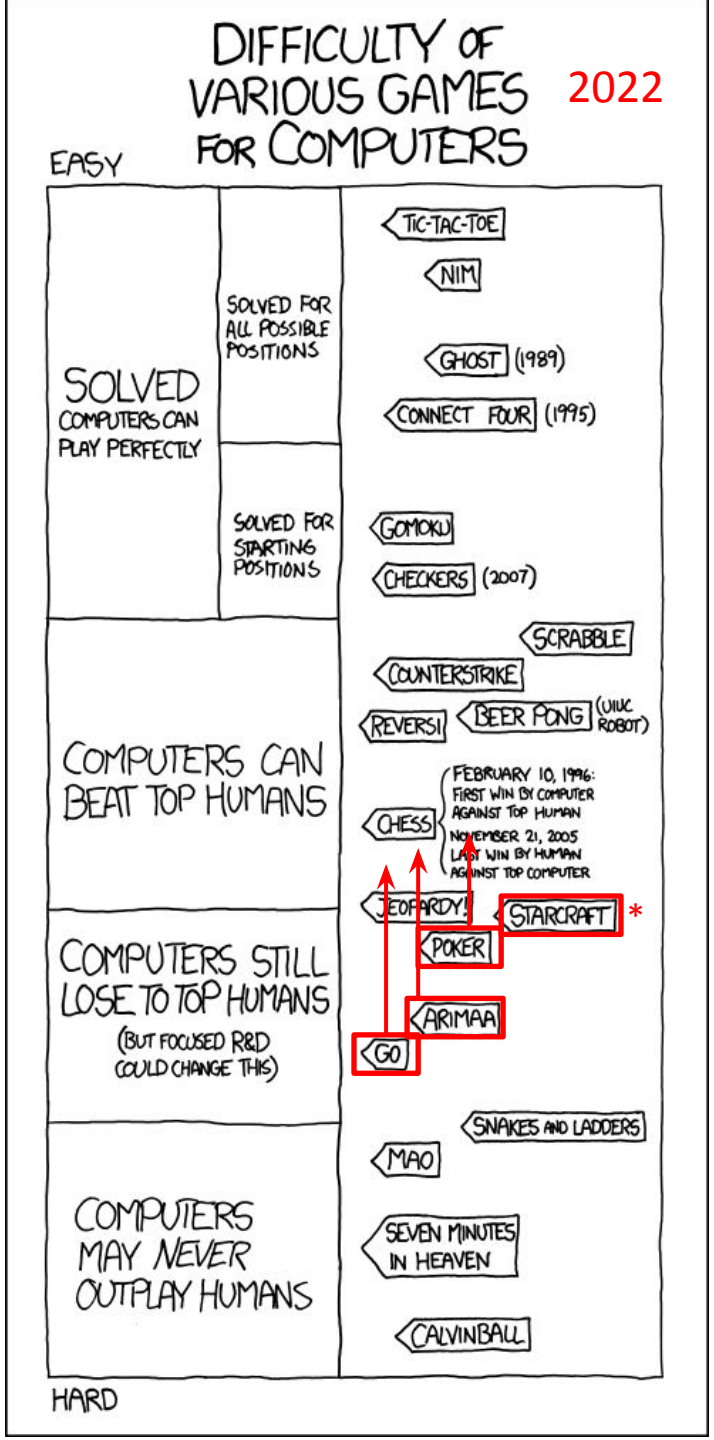
Exploitability Measurements




DIFFICULTY OF VARIOUS GAMES FOR COMPUTERS
2012
EASY
SOLVED
COMPUTERS CAN PLAY PERFECTLY
SOLVED FOR ALL POSSIBLE POSITIONS
SOLVED FOR STARTING POSITIONS
COMPUTERS CAN BEAT TOP HUMANS
COMPUTERS STILL LOSE TO TOP HUMANS
(BUT FOCUSED R&D COULD CHANGE THIS)
COMPUTERS MAY NEVER OUTPLAY HUMANS
HARD
TIC-TAC-TOE
NIM
GHOST (1989)
CONNECT FOUR (1995)
GOMOKU
CHECKERS (2007)
SCRABBLE
COUNTERSTRIKE
REVERSI
BEER PONG (UIUC ROBOT)
CHESS
FEBRUARY 10, 1996: FIRST WIN BY COMPUTER AGAINST TOP HUMAN
NOVEMBER 21, 2005 LAST WIN BY HUMAN AGAINST TOP COMPUTER
JEOPARDY!
STARCRAFT
POKER
ARIMAA
GO
SNAKES AND LADDERS
MAO
SEVEN MINUTES IN HEAVEN
CALVINBALL


DIFFICULTY OF VARIOUS GAMES FOR COMPUTERS
2012
EASY
SOLVED
COMPUTERS CAN PLAY PERFECTLY
SOLVED FOR ALL POSSIBLE POSITIONS
SOLVED FOR STARTING POSITIONS
COMPUTERS CAN BEAT TOP HUMANS
COMPUTERS STILL LOSE TO TOP HUMANS
(BUT FOCUSED R&D COULD CHANGE THIS)
COMPUTERS MAY NEVER OUTPLAY HUMANS
HARD
TIC-TAC-TOE
NIM
GHOST (1989)
CONNECT FOUR (1995)
GOMOKU
CHECKERS (2007)
SCRABBLE
COUNTERSTRIKE
REVERSI
BEER PONG (UIUC ROBOT)
CHESS
FEBRUARY 10, 1996: FIRST WIN BY COMPUTER AGAINST TOP HUMAN
NOVEMBER 21, 2005 LAST WIN BY HUMAN AGAINST TOP COMPUTER
JEOPARDY!
STARCRAFT
POKER
ARIMAA
GO
SNAKES AND LADDERS
MAO
SEVEN MINUTES IN HEAVEN
CALVINBALL




DIFFICULTY OF VARIOUS GAMES FOR COMPUTERS
2022
EASY
SOLVED
COMPUTERS CAN PLAY PERFECTLY
SOLVED FOR ALL POSSIBLE POSITIONS
SOLVED FOR STARTING POSITIONS
COMPUTERS CAN BEAT TOP HUMANS
COMPUTERS STILL LOSE TO TOP HUMANS
(BUT FOCUSED R&D COULD CHANGE THIS)
COMPUTERS MAY NEVER OUTPLAY HUMANS
HARD
TIC-TAC-TOE
NIM
GHOST (1989)
CONNECT FOUR (1995)
GOMOKU
CHECKERS (2007)
SCRABBLE
COUNTERSTRIKE
REVERSI
BEER PONG (UIUC ROBOT)
CHESS
FEBRUARY 10, 1996: FIRST WIN BY COMPUTER AGAINST TOP HUMAN
NOVEMBER 21, 2005 LAST WIN BY HUMAN AGAINST TOP COMPUTER
JEOPARDY!
STARCRAFT
*
POKER
ARIMAA
GO
SNAKES AND LADDERS
MAO
SEVEN MINUTES IN HEAVEN
CALVINBALL

The Bitter Lesson by Richard Sutton

"The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective… The two methods that seem to scale arbitrarily in this way are ***search*** and ***learning***."

Thank You!

Noam Brown
noam.brown@gmail.com
@polynoamial
www.noambrown.com