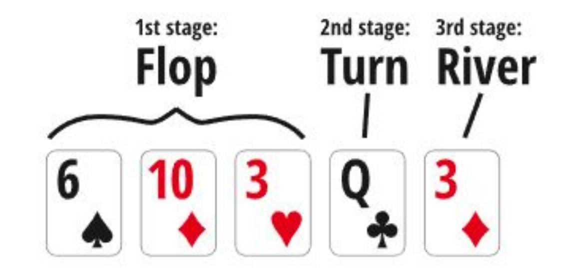
Libratus & Pluribus

Artificial Intelligence Seminar – 184.068 – 2022W

Simão Costa (e12202234)

Basics of Texas Hold'em Poker

- Minimum of 2 players
- Standard deck of cards (no Jokers Included)



- 2 ***hole*** cards for each player (private)
- 5 ***community*** cards are (eventually) dealt face up

- Rounds of betting take place before the flop is dealt and after each subsequent deal
- The ***pot*** is the accumulation of all the **money/chips** betted and it's given to the winner
- In **each** round, players have **3 betting options**:
    - call
    - raise
    - fold

Basics of Texas Hold'em Poker

In the end, players who have not folded form the **best five-card poker hand**

- using all the available cards **(hole + community)**

The winner is the one with **highest-ranking** five-card poker hand.

Basics of Texas Hold’em Poker

If the betting causes **all but one** player to **fold**

- that player wins the pot
- he doesn’t need to show his cards

Players don’t need to have the best hand to win the pot

- they can try to get others to **fold better hands**

Bluff **-** To make other players believe that he has a better hand than they might have

- by betting or raising

Poker is a psychological game!

Libratus: The Superhuman AI for No-Limit Poker

Libratus was developed by **Noam Brown and Tuomas Sandholm of Carnegie Mellon University** and it was designed to play the game of poker, requiring 100 CPUs to run.

First AI agent to beat professional players in Heads-up no-limit Texas hold ’em, defeating 4 professional poker players in a tournament (2017).
Libratus played **120000 hands** in **20 days**, **winning $1.8 million**.

Libratus plays **no-limit heads up Texas Hold'em poker**

- **imperfect-information**, **extensive-form**, **zero-sum** finite game

Imperfect-information game

A game in which one or more players **do not have complete information** about the state of the game.

Poker is an imperfect information game

- the cards that each player holds in their hand are hidden to the others

Extensive form game

Poker is a **sequential game -** only one player acts a time
Sequential games can be represented in what is called **extensive-form**, also known as a **game tree**

graph with **nodes that represent positions** in a game and **lines that represent potential actions**

The **poker game tree** is an abstract concept to better understand the best ways to develop strategies.

- allows them to capture more expected value than their opponents

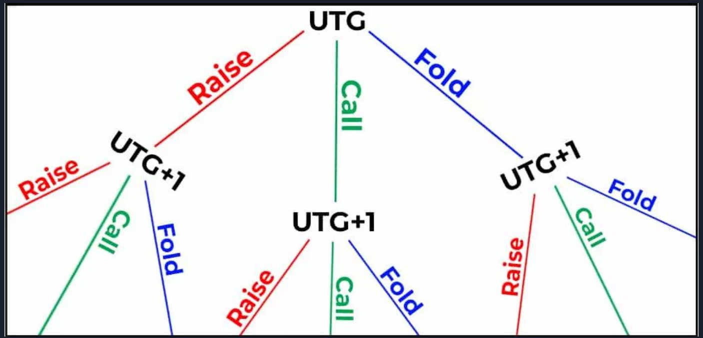
Game Tree

<u>Game Tree</u>
The tree starts from the **first preflop** decision

- A player (**UTG**) has 3 possible decisions: fold, call or raise.
- The next player (**UTG+1**) then makes his decision.

The deeper we go into the tree, the possibilities multiply exponentially!

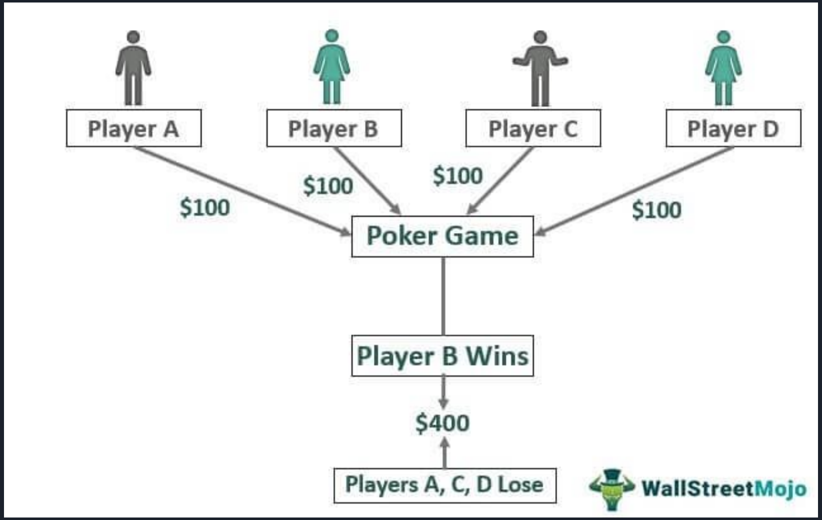
- When you add in the possibility of **multiple bet sizes**, then the tree becomes exponentially bigger once more.



Zero-sum game

One player's reward is the **negative** of their opponent(s)

- r1 + r2 + … + rN = 0



	Start	End	Profit
Player A	100	0	-100
Player B	100	400	+300
Player C	100	0	-100
Player D	100	0	-100
Total	400	400	0

Poker AI

- Rhode Island Hold'em: 10^9 decisions

  -> Solved with LP [Gilpin and Sandholm 2005]

- Limit Texas Hold'em: 10^{13} decisions

  -> Essentially solved with Counterfactual regret minimization+ [Zinkevich et al. 2007, Bowling et al. 2015, Tammelin et al. 2015]

  -> Required size compression (262TB -> 11TB)

- No-limit Texas Hold'em: 10^{161}

  -> Too big for the same approach

Libratus game plan

Uses **game theory** to make strategic decisions while playing poker.

branch of mathematics that studies decision-making in strategic situations, where the outcome of a decision is affected by the actions of other players

Libratus uses a technique called **Counterfactual regret minimization (CFR)** to determine its strategy in the game, producing an approximation of a **strategy profile** called **Nash Equilibrium**

Strategy profile

Set of strategies for all players involved in a game.

Given a **strategy profile** it's possible to emulate playing poker between players.

A single game play is a sequence of actions drawn from probability distributions given by players strategies. Once a game play is over, **players gain their utilities.**

Probabilistic framework → **expected** utilities

We can evaluate strategies and strategy profiles via expected utilities

Nash Equilibrium

Strategy profile such that no single player has incentive to deviate.

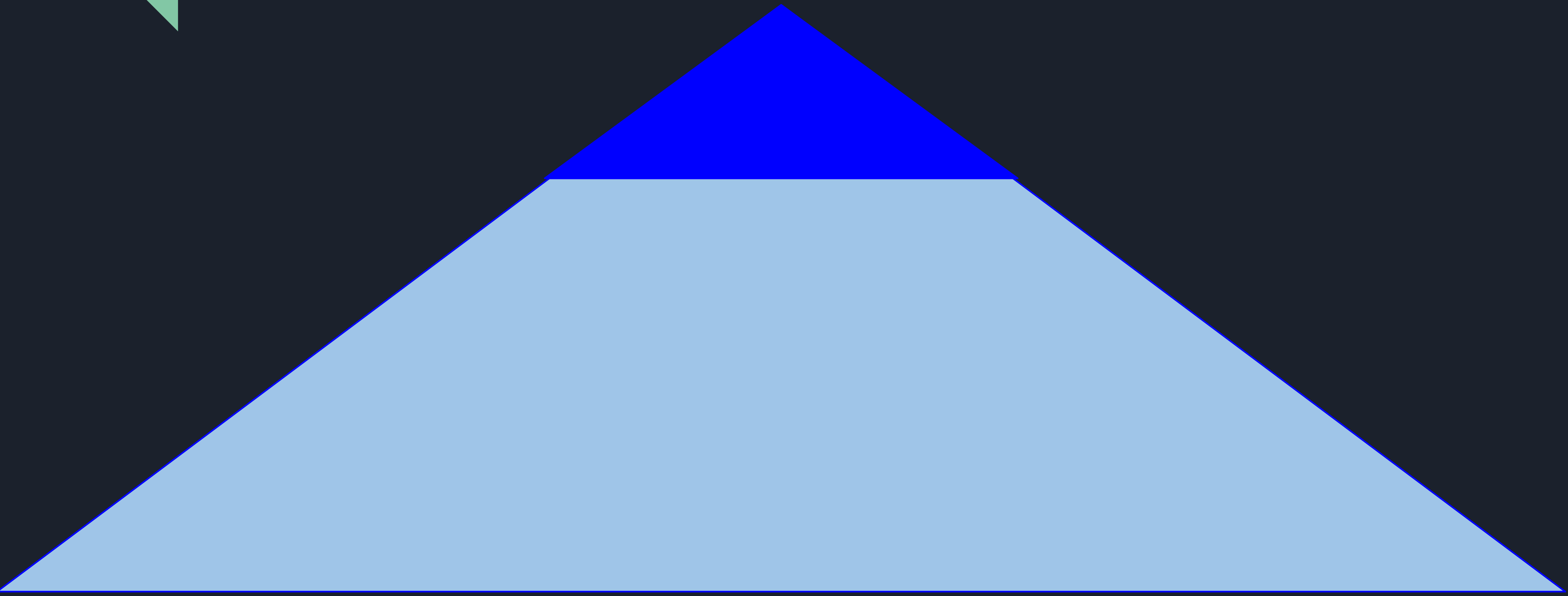
In 2 player zero-sum games, playing a Nash equilibrium ensures you will not lose in expectation.

A game reaches a state of **equilibrium** when no player in a game can improve their outcome by changing their strategy, assuming that all other players keep their strategies unchanged.

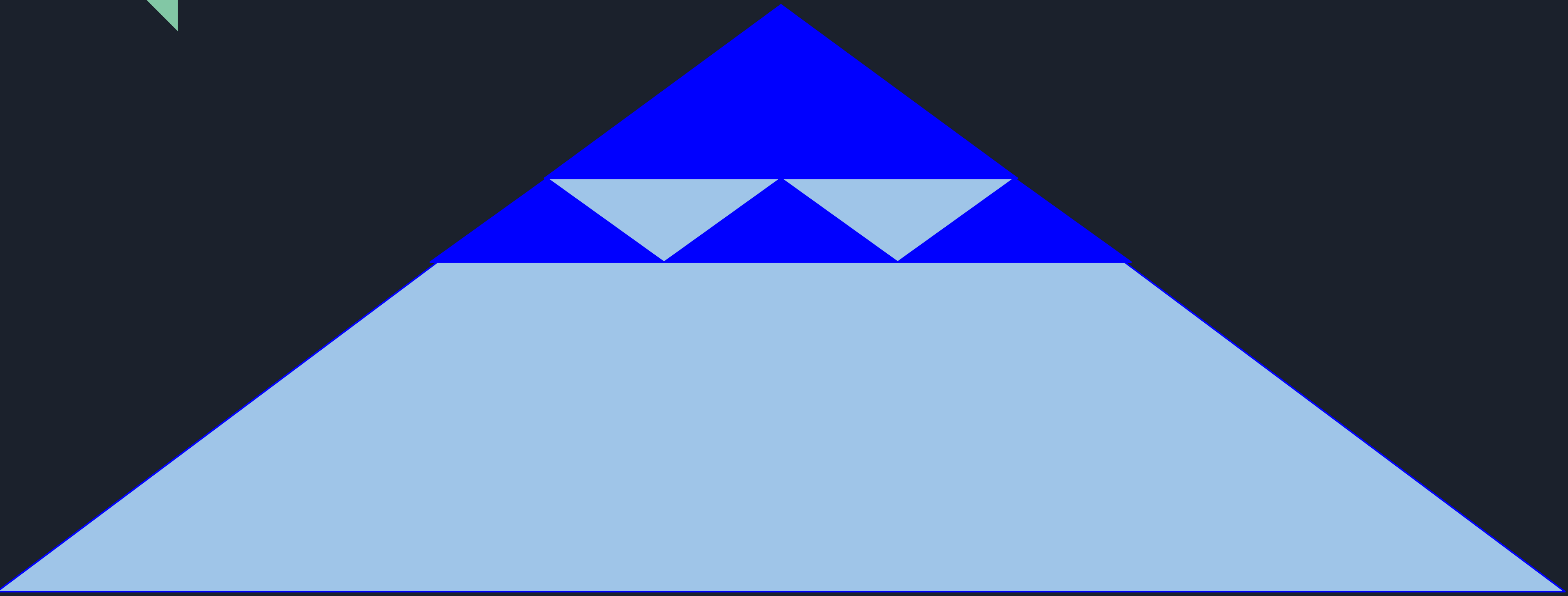
The goal is to find the Nash Equilibrium

Libratus

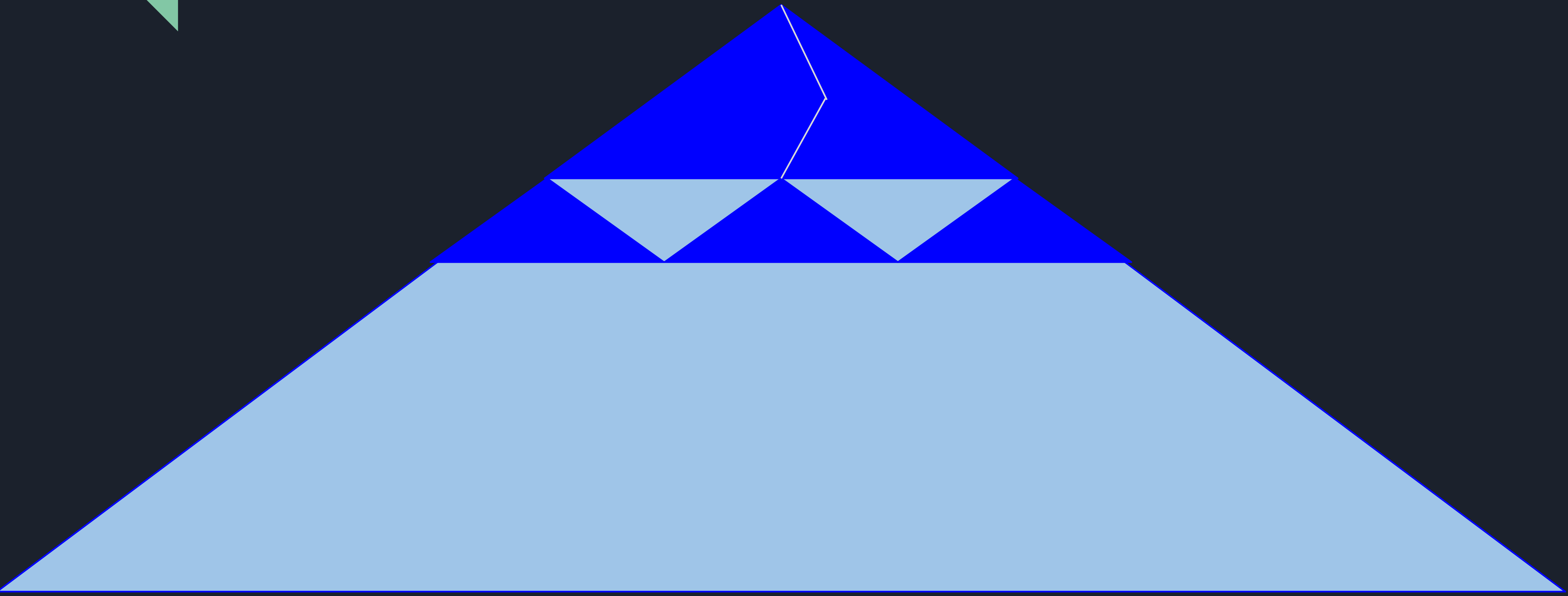
Libratus



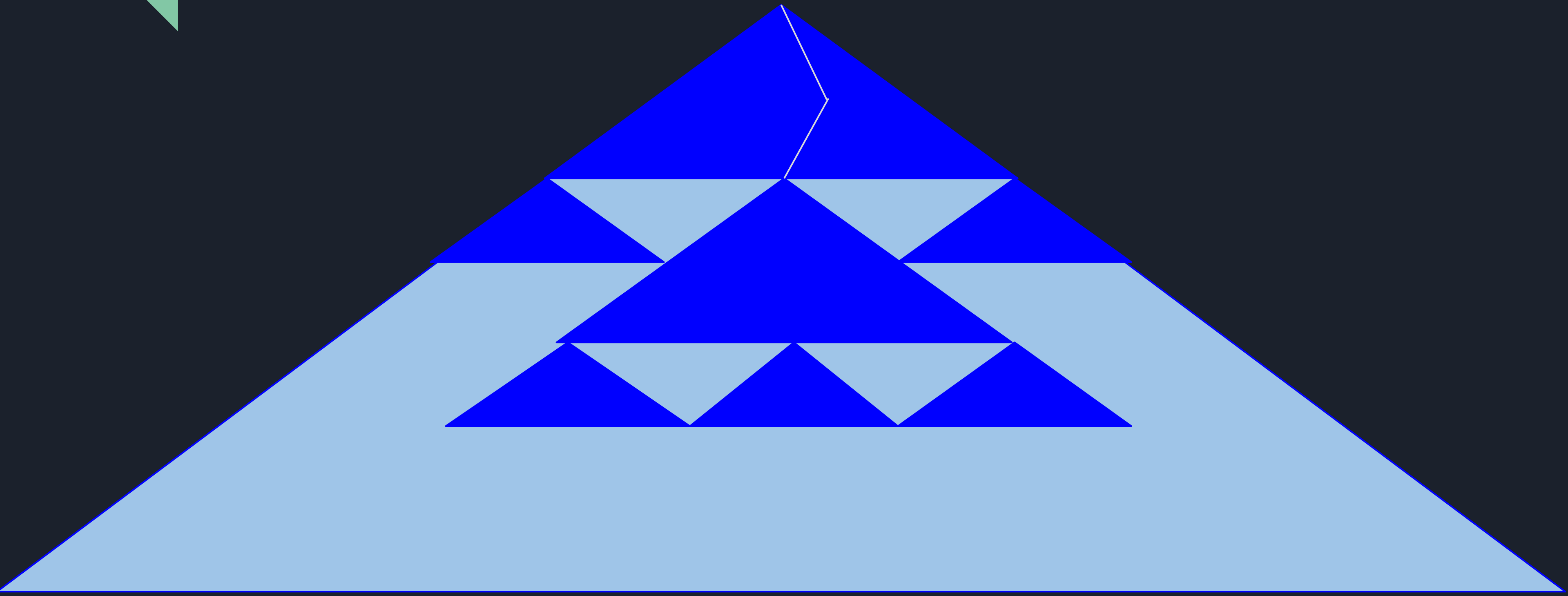
Libratus



Libratus



Libratus



Game abstraction

10^{161}

No-limit → **no restrictions on bets** → **number of possible actions is enormous**

It is costly and wasteful to construct a new betting strategy for a single-dollar difference in the bet.

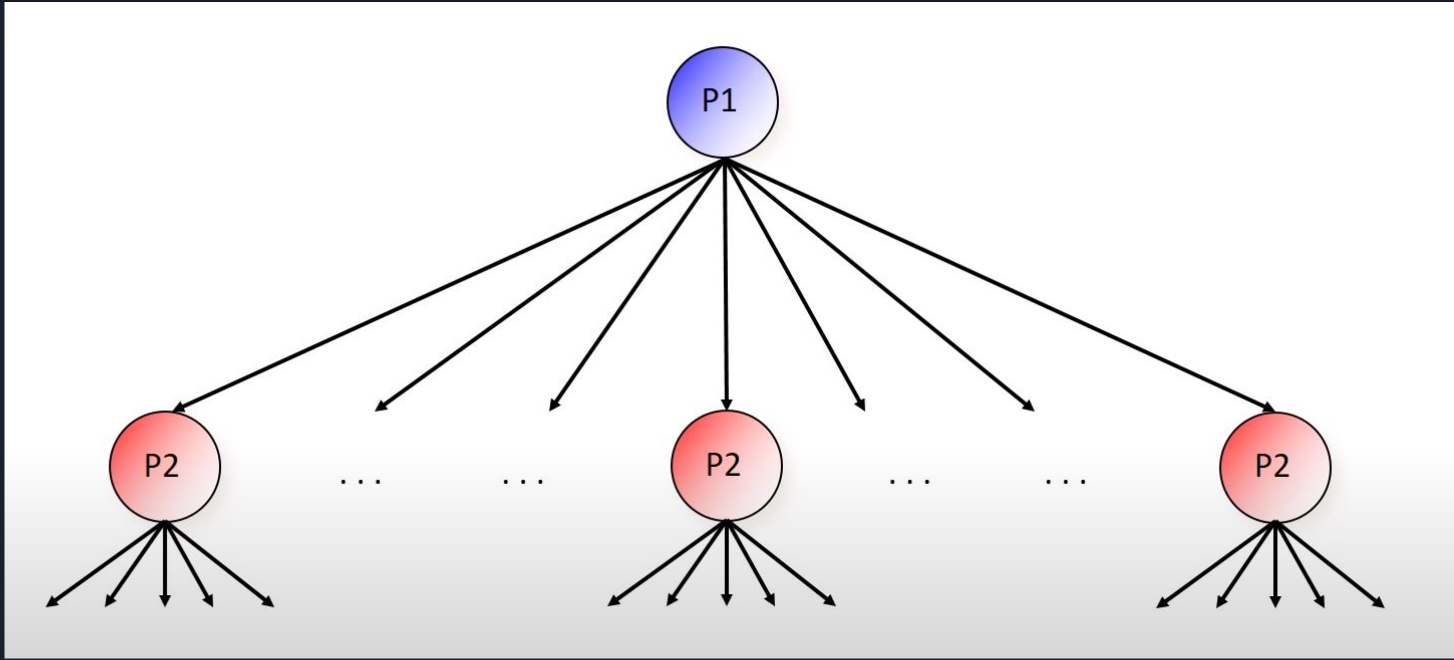
Libratus abstracts the game using a technique - **Blueprinting**

- by grouping similar bets (**action abstraction**) and card combinations (**card abstraction**)

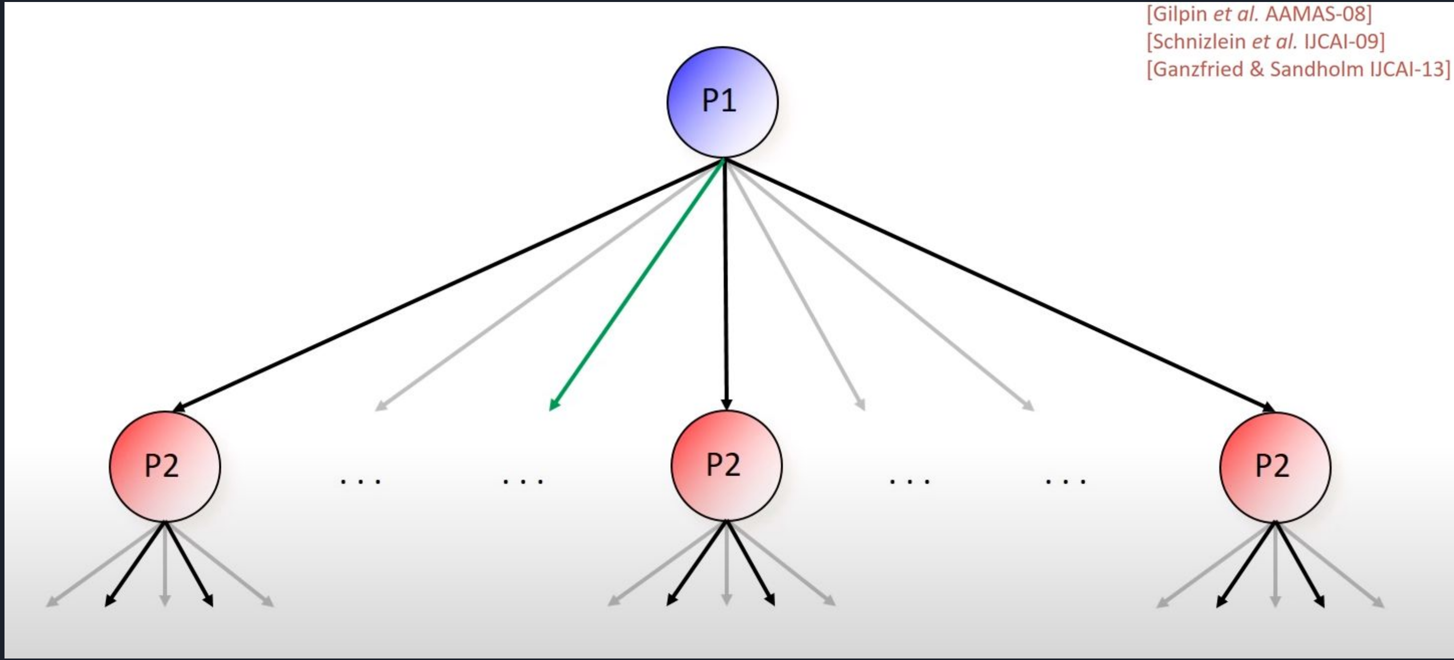
↓

This allows the AI to consider a **smaller number of possible actions**, making the decision-making process more **efficient**.

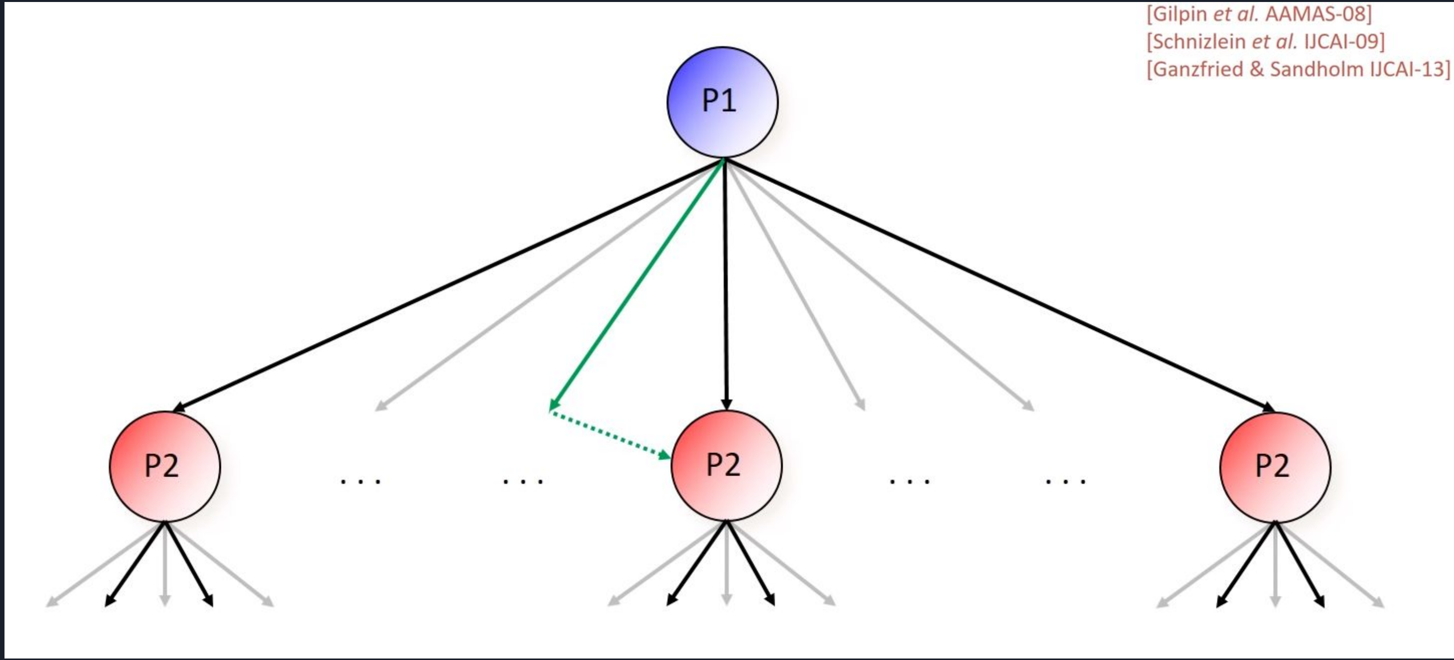
Action abstraction



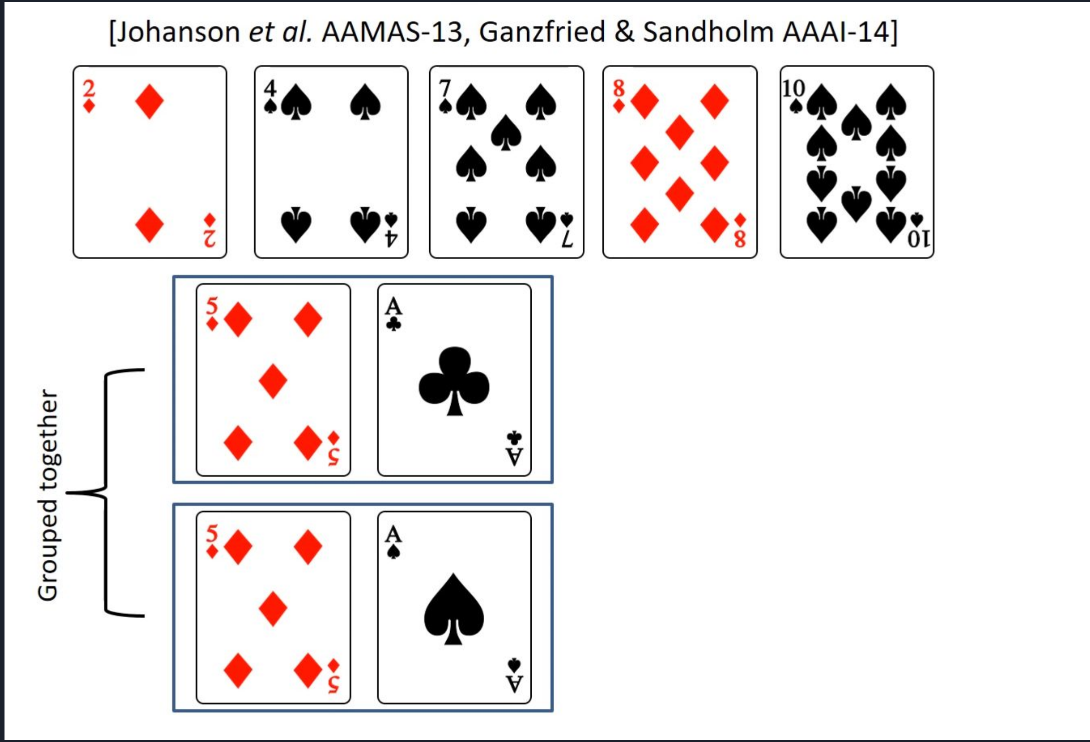
Action abstraction



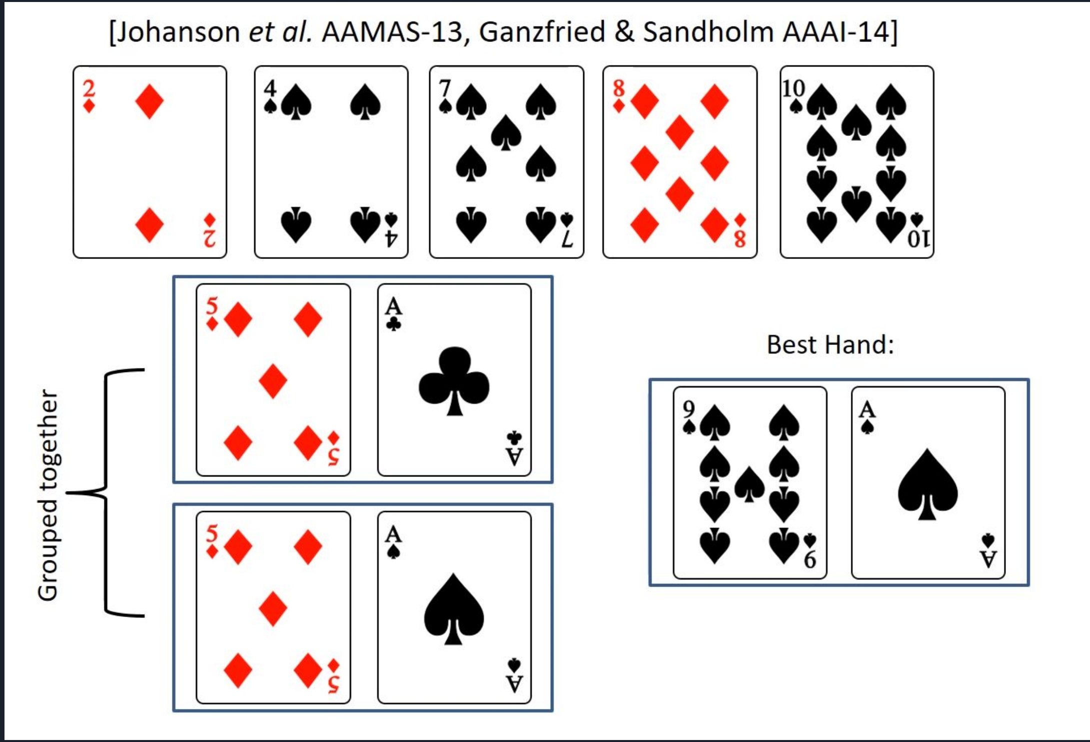
Action translation



Card abstraction



Card abstraction



Libratus abstraction

Libratus does not use any card abstraction on the **first (preflop)** and **second (flop)** betting rounds

The **last two betting rounds**, which are exponentially larger, are more coarsely abstracted.

Third round: **55 million** different hand possibilities are grouped into **2.5 million buckets**

Fourth round: **2.4 billion** different possibilities are grouped into **1.25 million buckets**

Counterfactual regret minimization

Regret

- To update the strategy profile for each player
- Keeps track of the **regrets** at **each decision point**
    - uses them to adjust the probability of each action in the strategy profile

Counterfactual regret - The **difference in expected utility** between

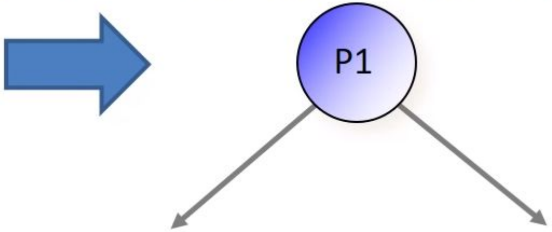
- the decision that was actually made
- the **best decision** that could have been made

given the information available at the time

Monte Carlo CFR

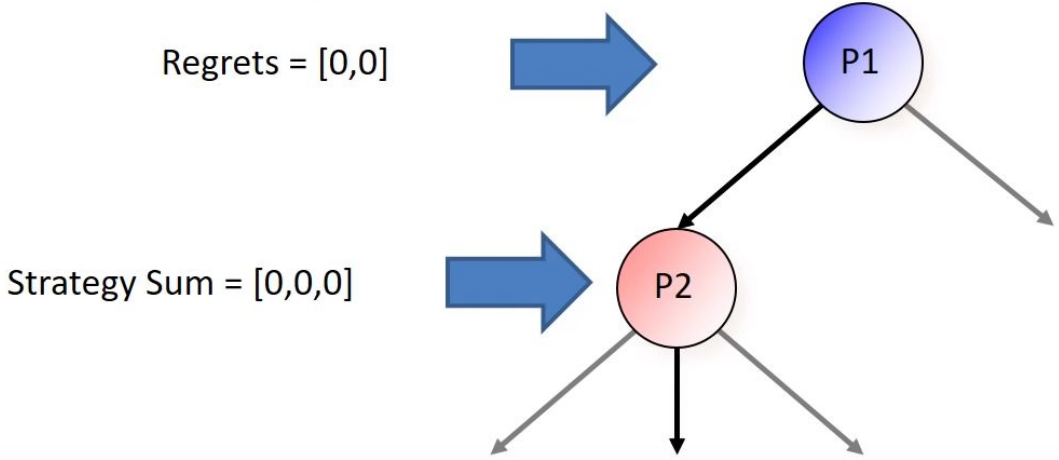
[Zinkevich *et al.* NIPS-07, Lanctot *et al.* NIPS-12]

Regrets = [0,0]

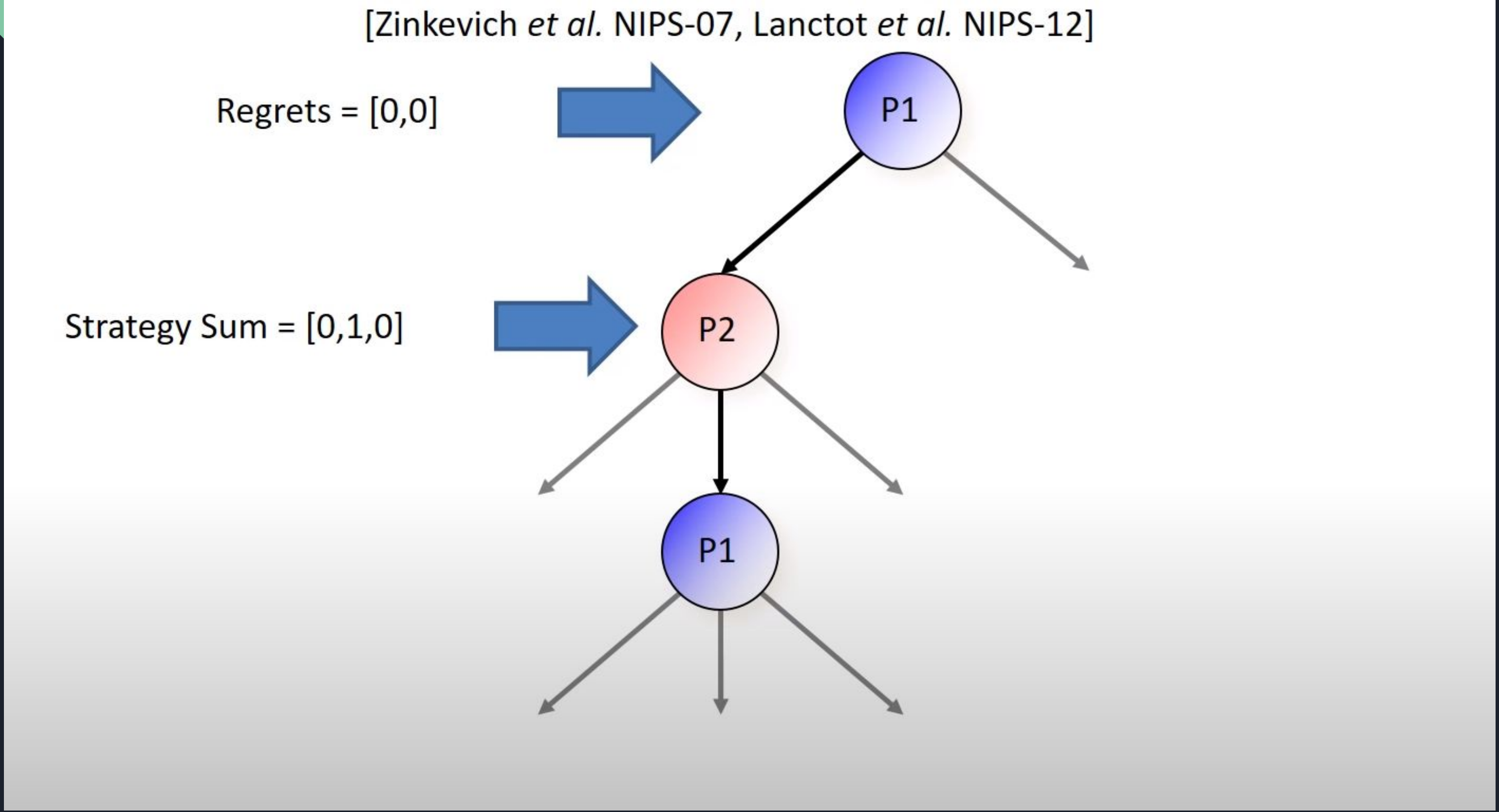


Monte Carlo CFR

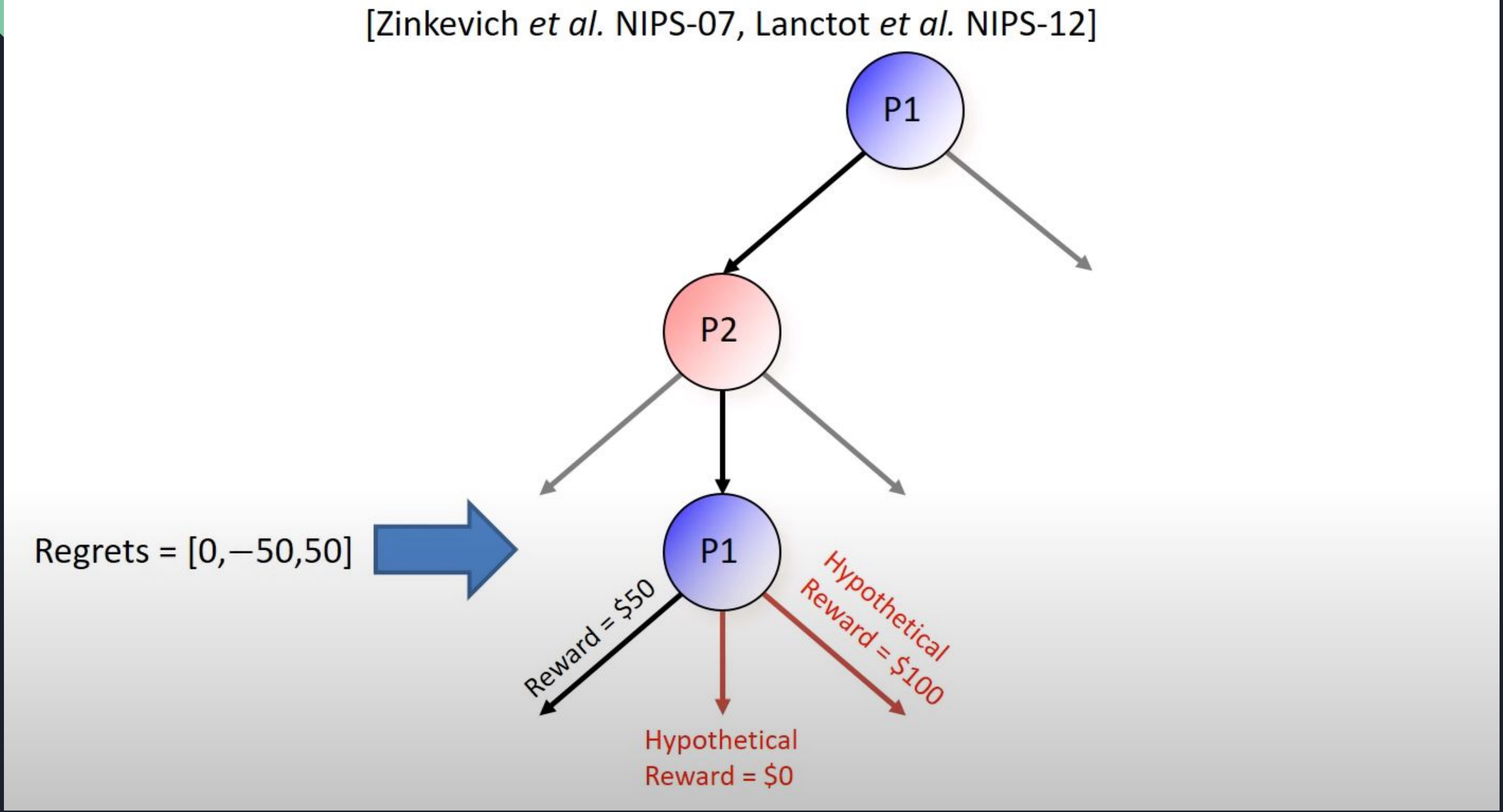
[Zinkevich *et al.* NIPS-07, Lanctot *et al.* NIPS-12]



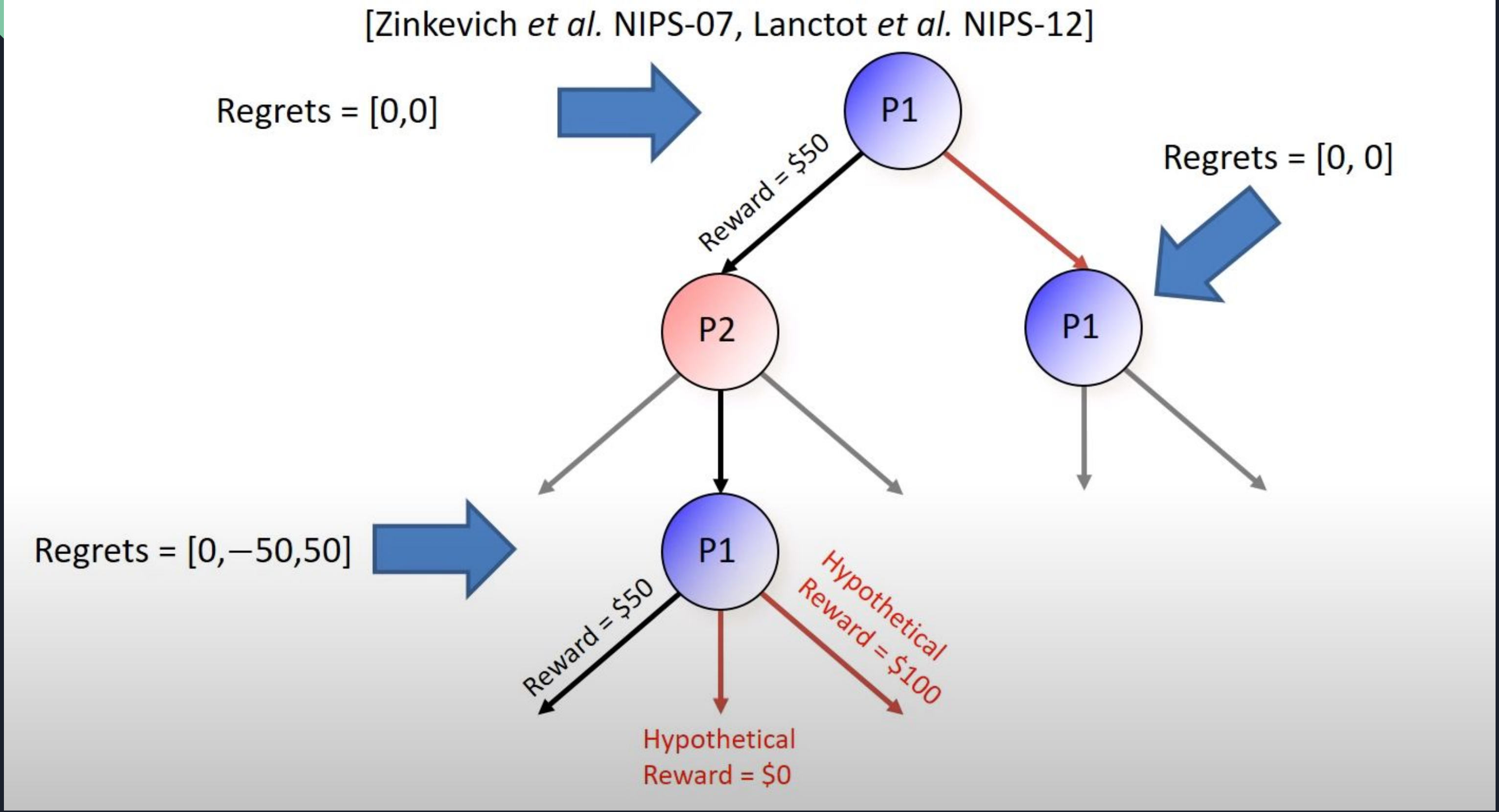
Monte Carlo CFR



Monte Carlo CFR

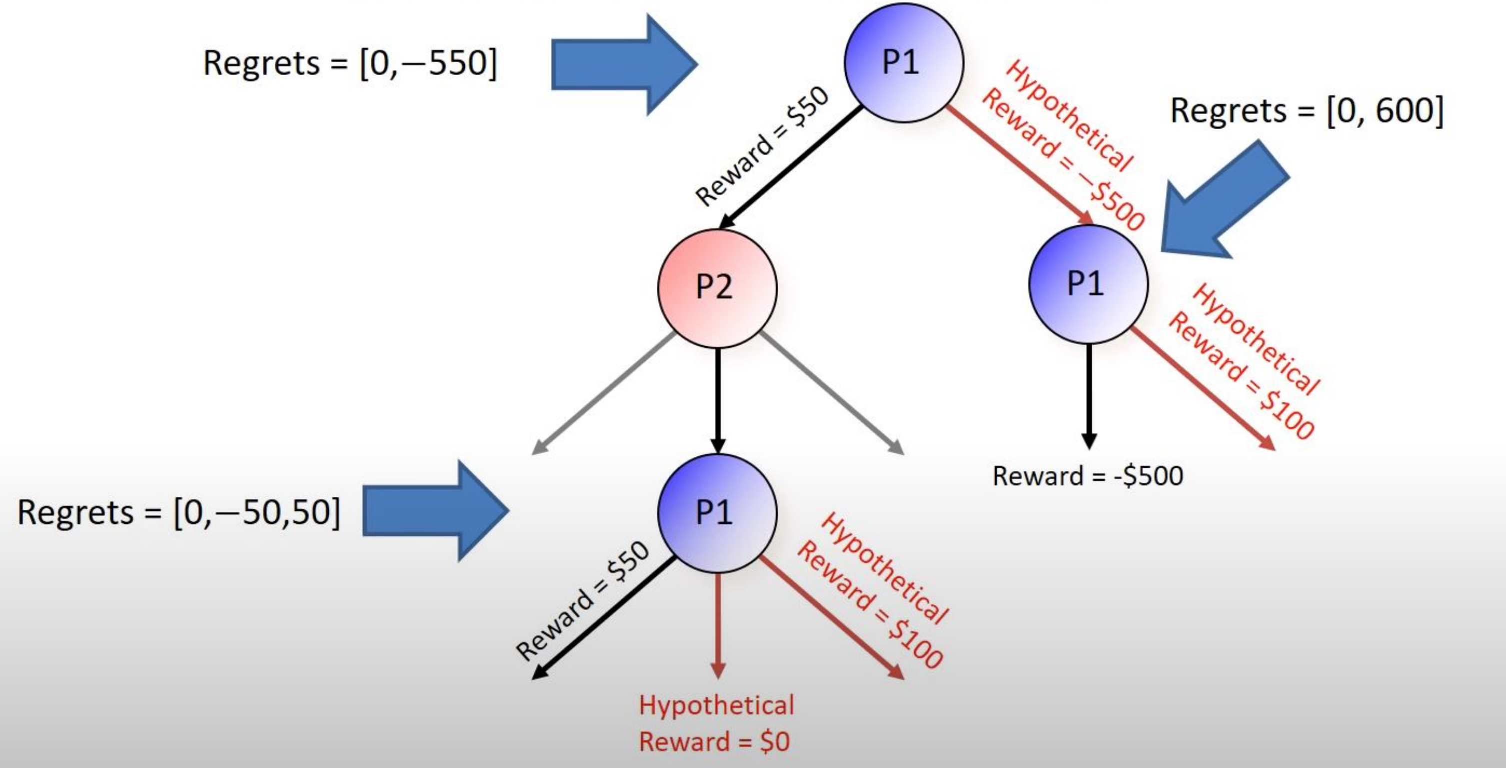


Monte Carlo CFR



Monte Carlo CFR

[Zinkevich *et al.* NIPS-07, Lanctot *et al.* NIPS-12]



Unsafe Subgame solving

- An **estimate of the opponent's strategy** is made

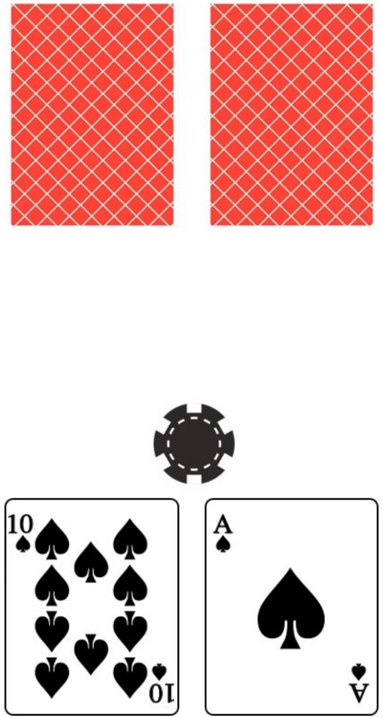
- This gives a **belief** distribution over states

- Update beliefs via **Bayes Rule -** mathematical formula for determining conditional probability. Conditional probability is the likelihood of an outcome occurring, based on a previous outcome having occurred in similar circumstances

May result in very high exploitability!

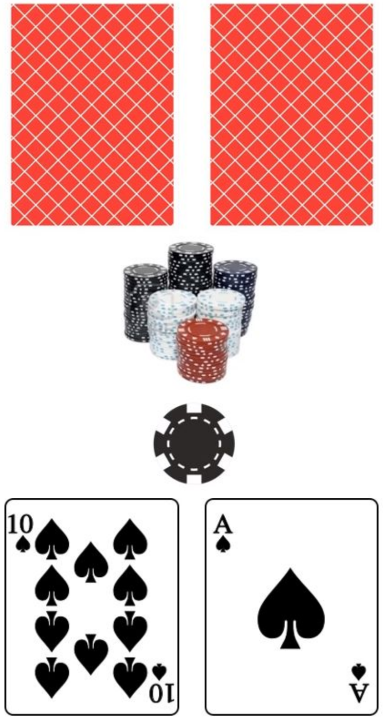
Unsafe Subgame solving

[Ganzfried & Sandholm AAMAS 2015]



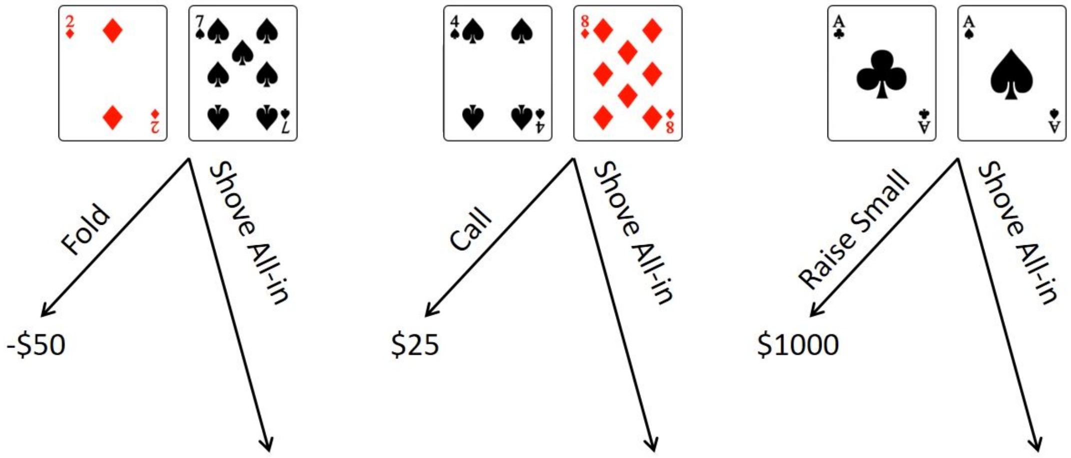
Unsafe Subgame solving

[Ganzfried & Sandholm AAMAS 2015]



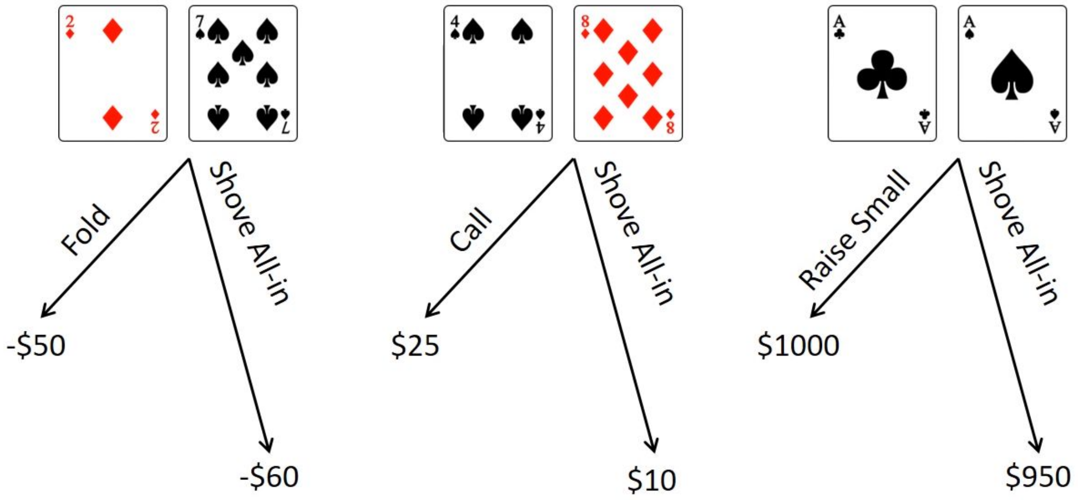
Safe Subgame solving

[Burch et al AAAI-14, Moravcik et al AAAI-16, Brown & Sandholm NIPS-17]

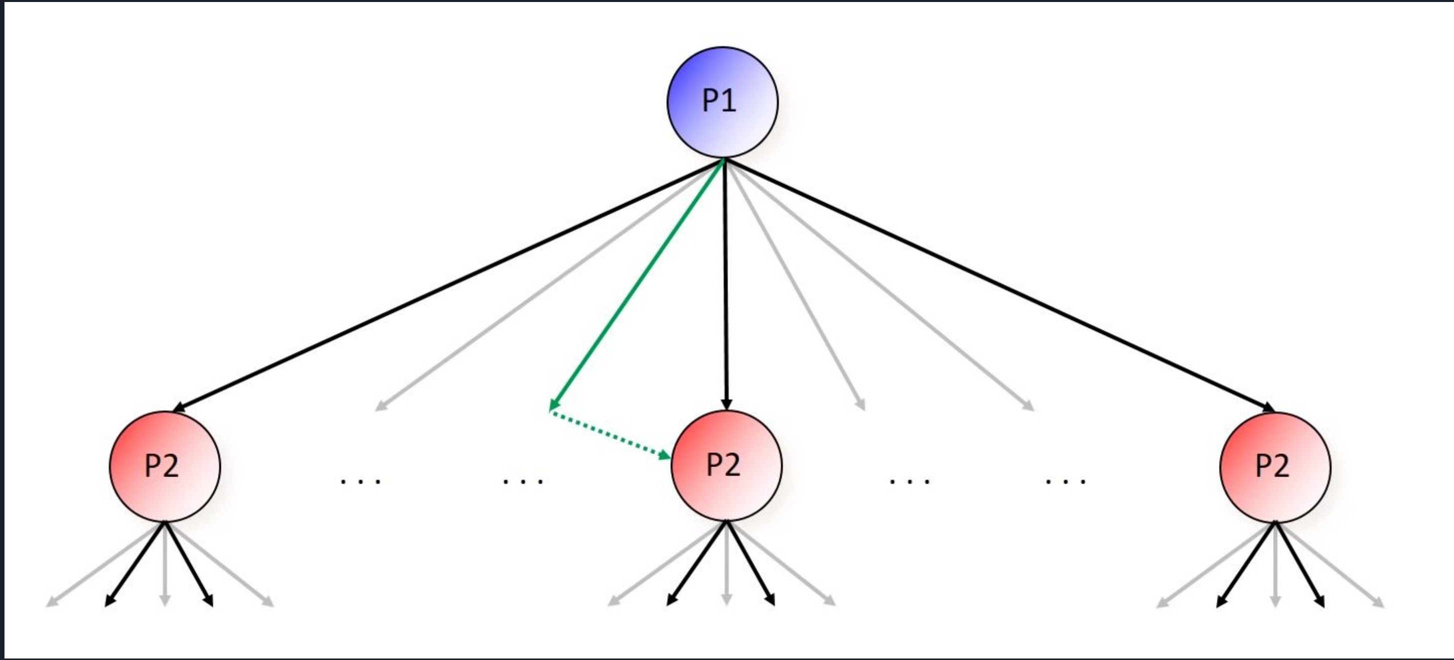


Safe Subgame solving

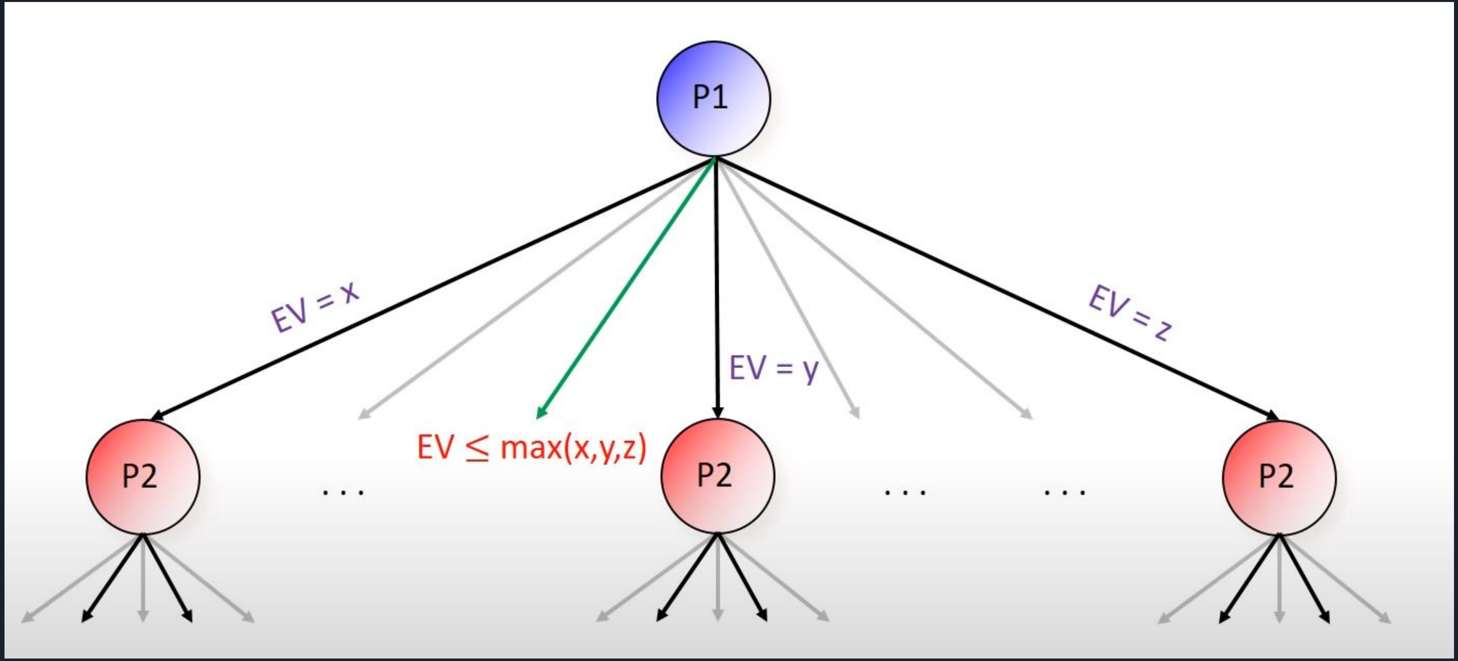
[Burch et al AAAI-14, Moravcik et al AAAI-16, Brown & Sandholm NIPS-17]



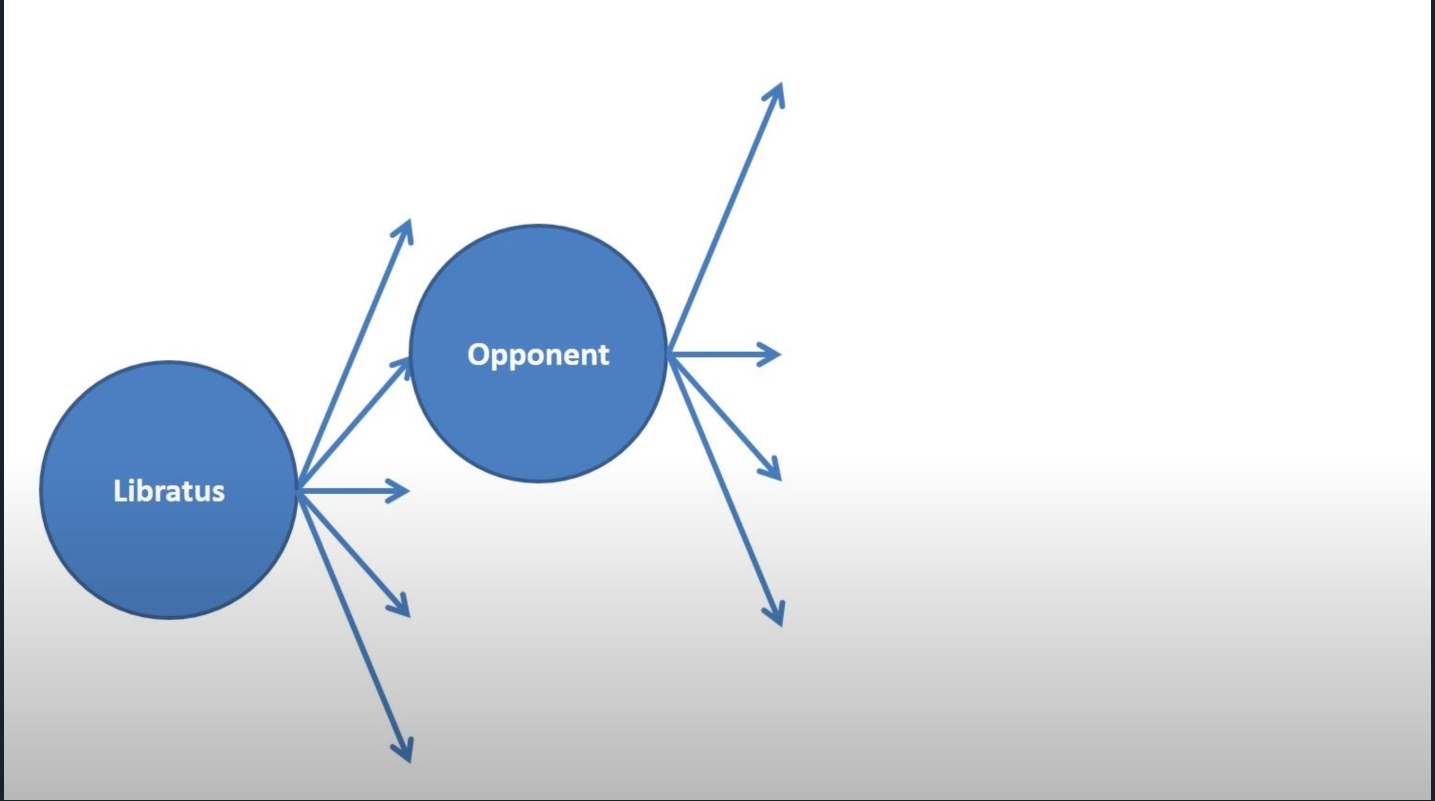
Action translation



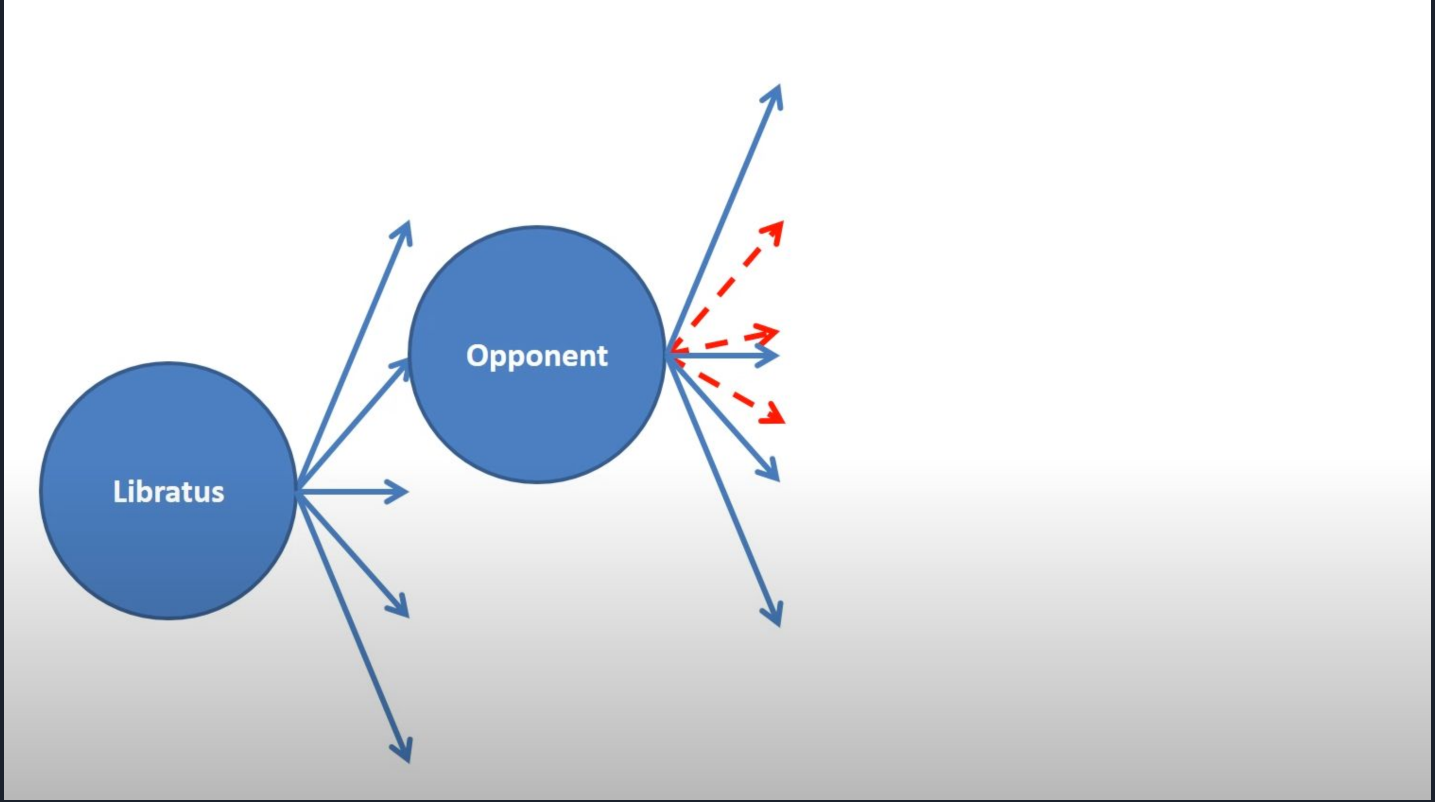
Subgame Solving – Off-tree actions



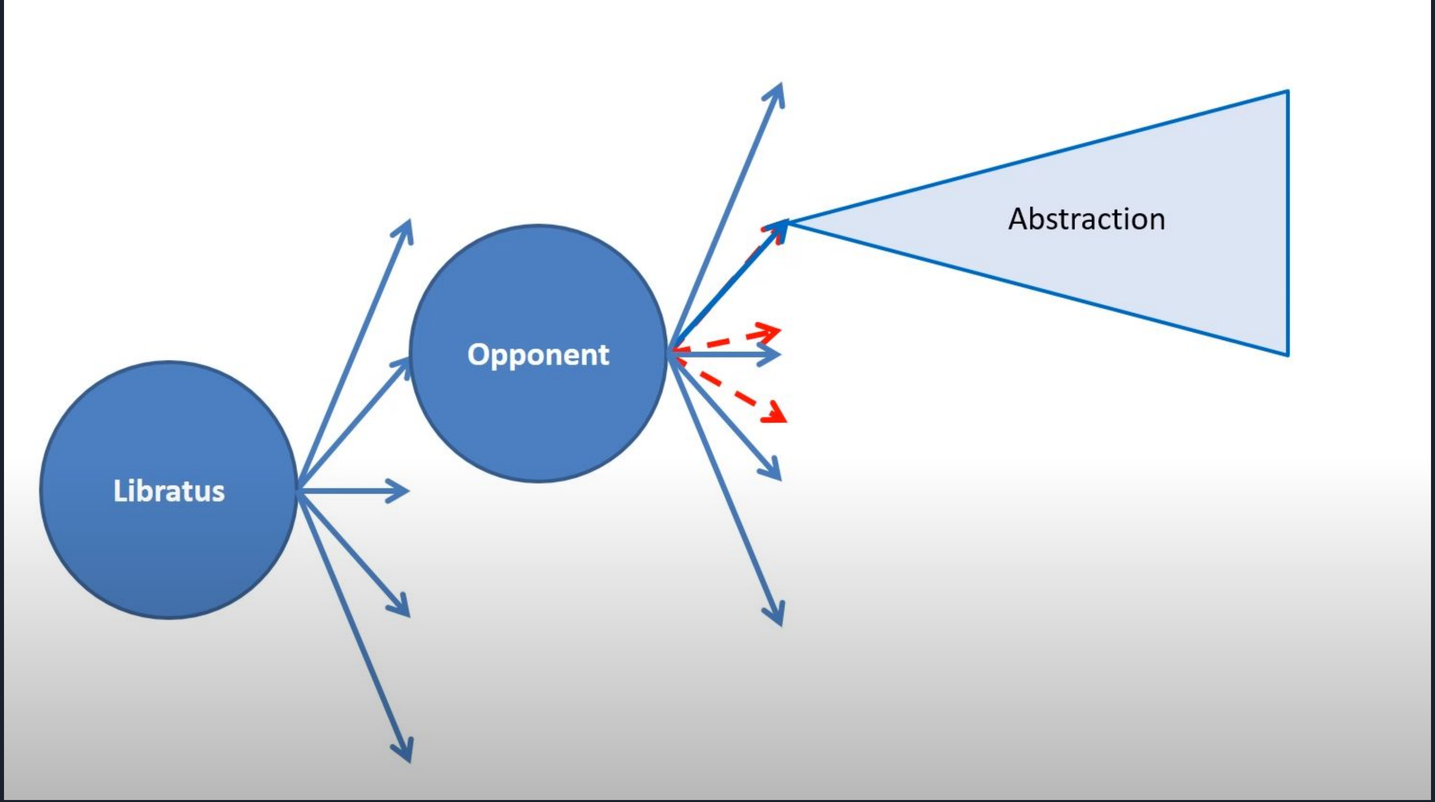
Self-Improvement



Self-Improvement



Self-Improvement



Pluribus

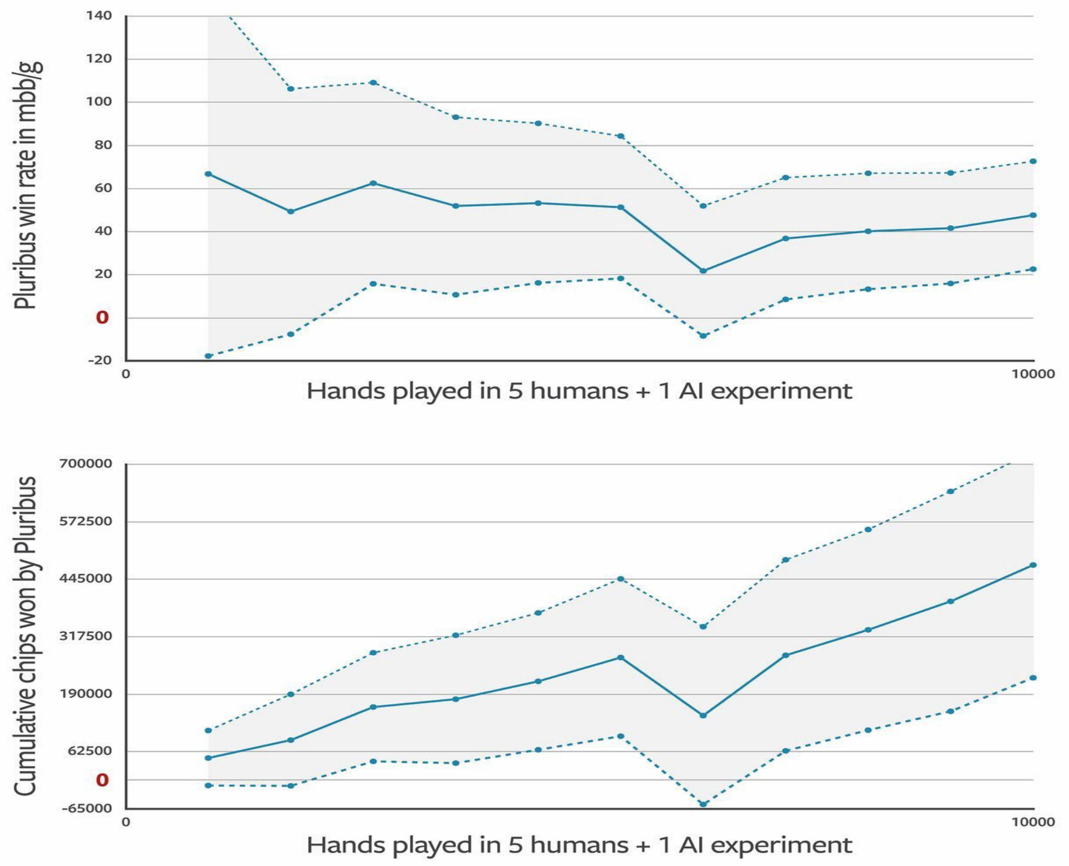
- Designed to play in a **multiplayer** environment (**3+ players**)
- Improvement from **Libratus**
- Team of AI researchers from Facebook in collaboration with Carnegie Mellon University

Was able to defeat **12** professional poker players (2019) in a **six-player** game, playing **10000 hands and winning on average $480** from its human competitors for **every 100 hands-on.**

Pluribus has to deal with a **higher level of uncertainty**, and has to take into account the different strategies and behaviors of multiple opponents at the same time.

Pluribus has the use of additional techniques like **fictitious play** and **abstract reasoning** which allows the AI to reason about the strategies of its opponents and the state of the actions **on the long run**. This allows the AI to make more accurate predictions about the actions of its opponents and make more strategic decisions.

Pluribus



[Brown and Sandholm 2019]

Observations

After talking to Poker professionals, they stated their main **strengths**:

- Many different bet sizes
- Huge overbets
- Near-perfect balance

The only **“weakness”** found was that there were no opponent exploitation

Both Pluribus and Libratus have the potential to be applied in other fields such as **finance, cybersecurity, and military strategy**.

Pluribus, due to its ability to play in a multi-player game, could have more potential applications in decision making in group dynamics.

References

1. [Bard et al., 2014] Nolan Bard, Michael Johanson, and Michael Bowling. Asymmetric abstractions for adversarial settings. In International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), pages 501–508, 2014.

2. [Bowling et al., 2015] Michael Bowling, Neil Burch, Michael Johanson, and Oskari Tammelin. Heads-up limit hold'em poker is solved. Science, 347(6218):145–149, January 2015.

3. [Brown and Sandholm, 2014] Noam Brown and Tuomas Sandholm. Regret transfer and parameter optimization. In Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, pages 594–601. AAAI Press, 2014.

4. [Brown and Sandholm, 2015] Noam Brown and Tuomas Sandholm. Regret-based pruning in extensive-form games. In Advances in Neural Information Processing Systems, pages 1972–1980, 2015.

5. [Brown and Sandholm, 2016a] Noam Brown and Tuomas Sandholm. Baby Tartanian8: Winning agent from the 2016 annual computer poker competition. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16), pages 4238–4239, 2016.

6. [Brown and Sandholm, 2016b] Noam Brown and Tuomas Sandholm. Strategy-based warm starting for regret minimization in games. In AAAI Conference on Artificial Intelligence (AAAI), 2016.

7. 49. N. Brown, T. Sandholm, Proceedings of the AAAI-17 Workshop on Computer Poker and Imperfect Information Games (2017).

8. [Noam Brown, 2017] Superhuman AI for heads-up no-limit poker: Libratus beats top professionals.

9. A. Rubinstein, Inapproximability of Nash Equilibrium. *SIAM J. Comput.* 47, 917–959 (2018).

10. OpenAI, OpenAI Five, https://blog.openai.com/openai-five/ (2018).

Thank you for your time!

Artificial Intelligence Seminar – 184.068 – 2022W

Simão Costa (e12202234)