

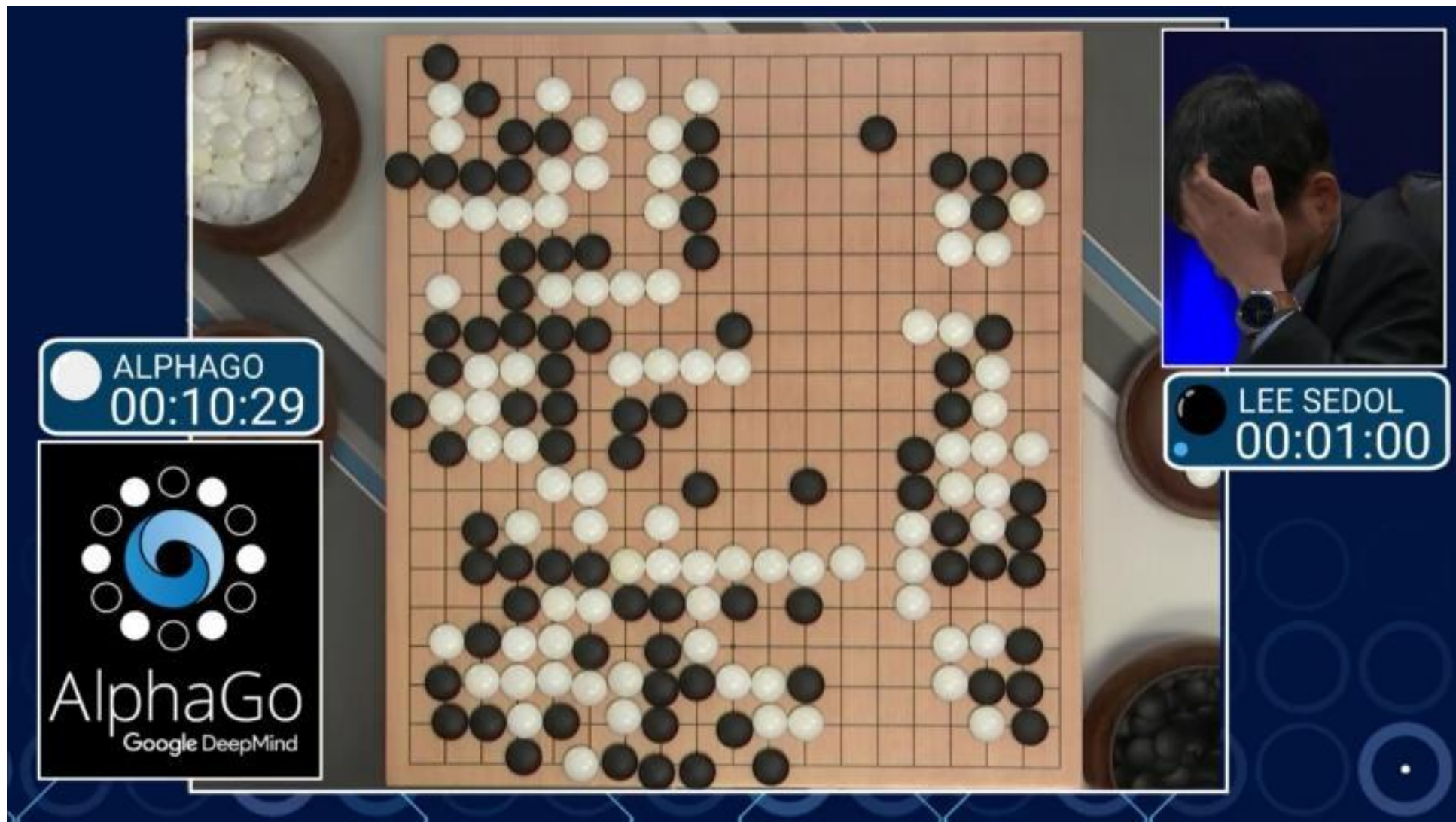
Announcement

- Homework 1: search, CSP, adversarial search
 - Available in Blackboard -> Homework
 - Due: Oct. 25, 11:59pm
- Programming Assignment 1B: adversarial search
 - Instructions at Blackboard -> “Programming Assignments”
 - Submission at AutoLab
 - Due: Nov. 1, 11:59pm

Adversarial Search

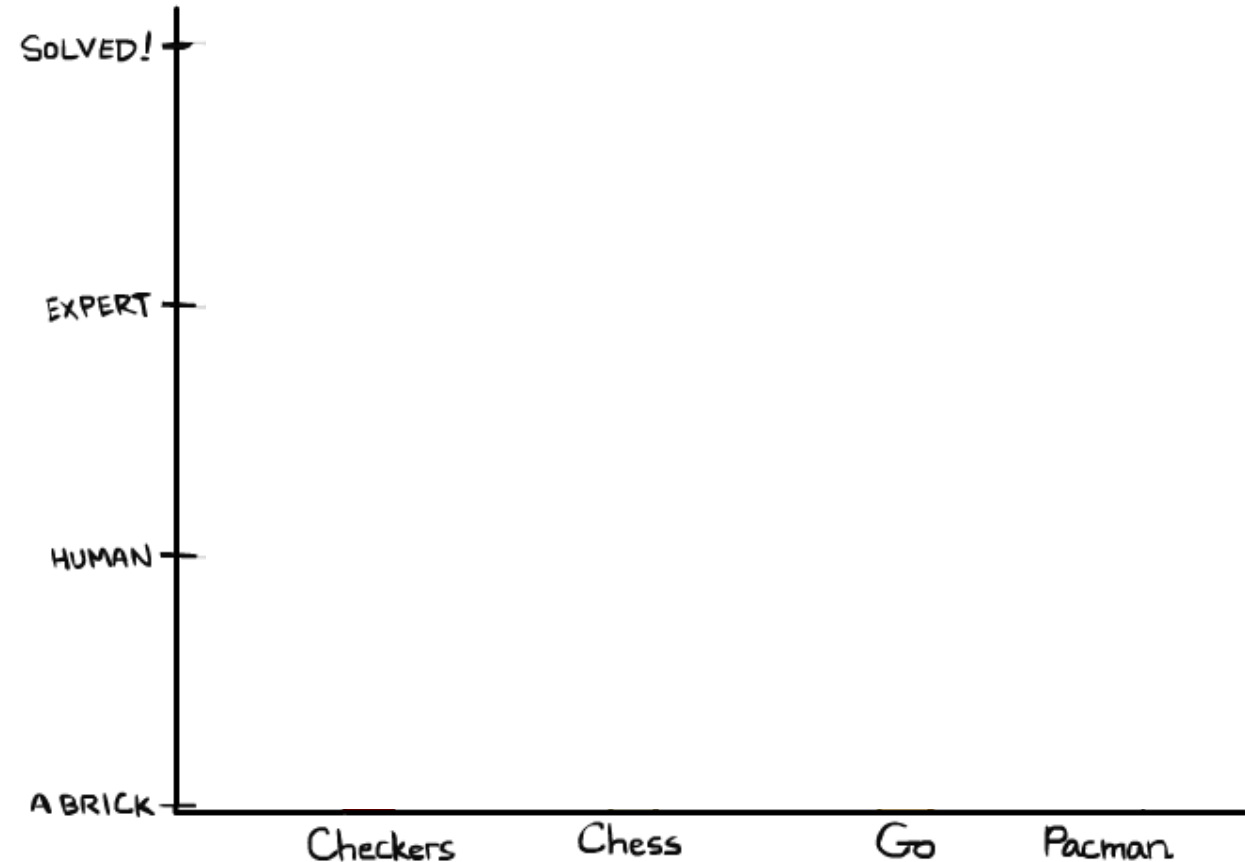


AlphaGo (2016)



Game Playing State-of-the-Art

- **Checkers:** 1950: First computer player. 1994: First computer champion: Chinook ended 40-year-reign of human champion Marion Tinsley using complete 8-piece endgame. 2007: Checkers solved!
- **Chess:** 1997: Deep Blue defeats human champion Gary Kasparov in a six-game match. Deep Blue examined 200M positions per second, used very sophisticated evaluation and undisclosed methods for extending some lines of search up to 40 ply. Current programs are even better, if less historic.
- **Go: 2016: AlphaGo defeats human champion! Uses Monte Carlo Tree Search, learned evaluation function.**
- **Pacman**



Mahjong

WORLD ARTIFICIAL INTELLIGENCE CONFERENCE

WAIC

2019世界人工智能大会
WORLD ARTIFICIAL INTELLIGENCE CONFERENCE

微软AI最新突破：超级麻将AI Suphx
探索隐藏的不确定性
Super Mahjong AI Suphx: exploring hidden information

隐藏的
不确定性
Hidden
information

下一个挑战：麻将！
Next challenge: Mahjong!

国际象棋
Chess
1997

国际象棋
Chess
2006

德州扑克
Poker
2017

麻将
Mahjong
2019

可观测的状态信息
Observable
States

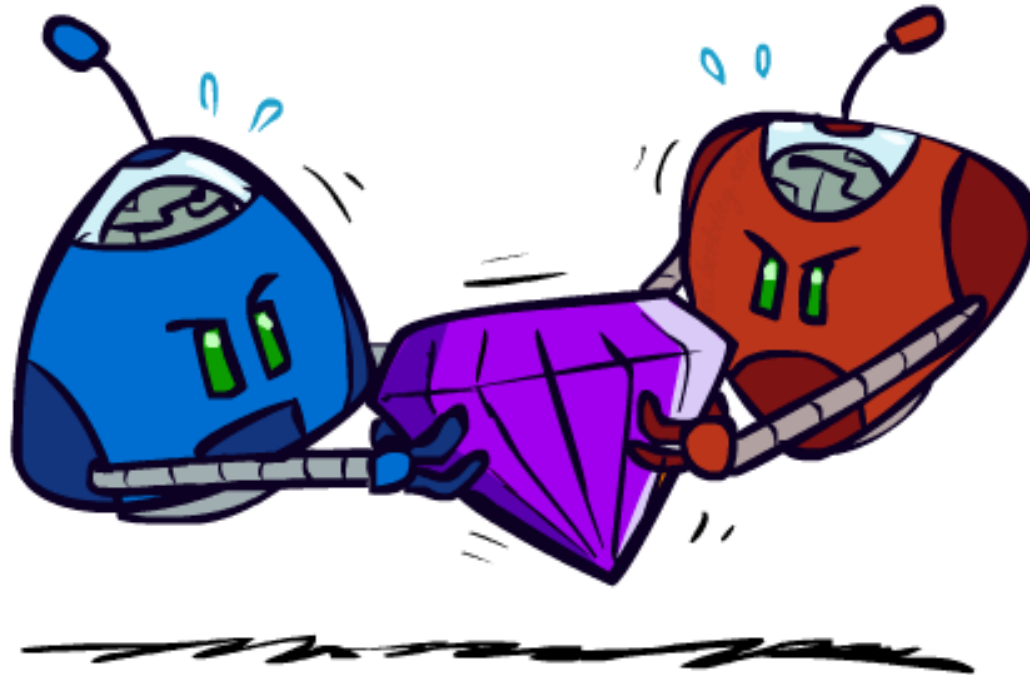
今天在上海我要向大家宣布，微软亚洲研究院创造出了历史上最强的麻将AI。
Today in Shanghai, I would like to announce to you that Microsoft Research Asia has created the strongest mahjong AI in history

2019世界人工智能大会
WORLD ARTIFICIAL INTELLIGENCE CONFERENCE

隐藏的不确定信息

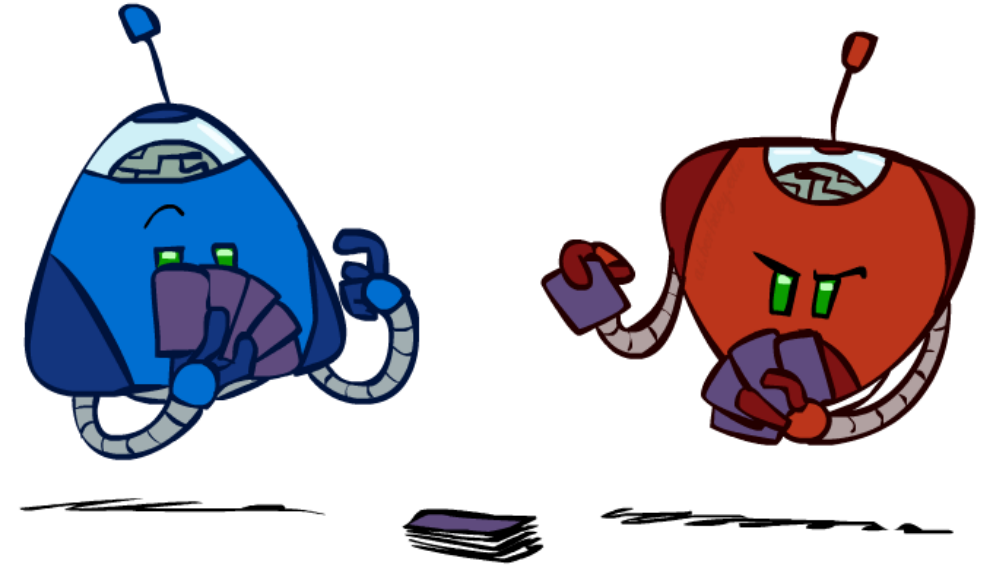


Adversarial Games



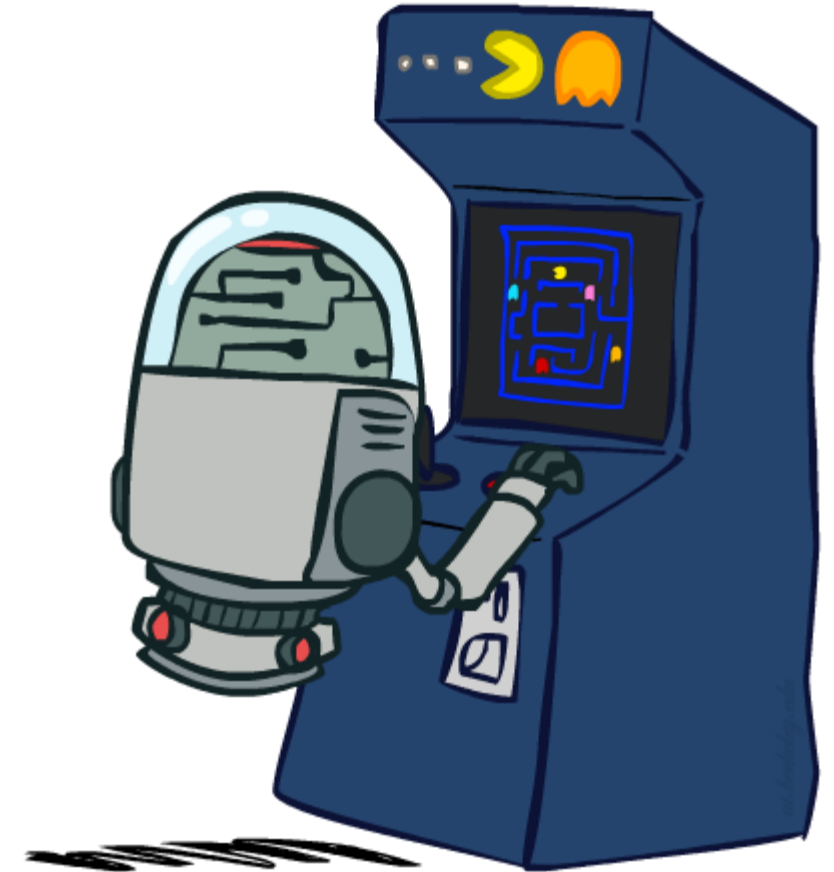
Types of Games

- Many different kinds of games!
- Differences:
 - Deterministic or stochastic?
 - One, two, or more players?
 - Zero sum?
 - Perfect information (can you see the state)?
- Want algorithms for calculating a **strategy (policy)** which recommends a move from each state

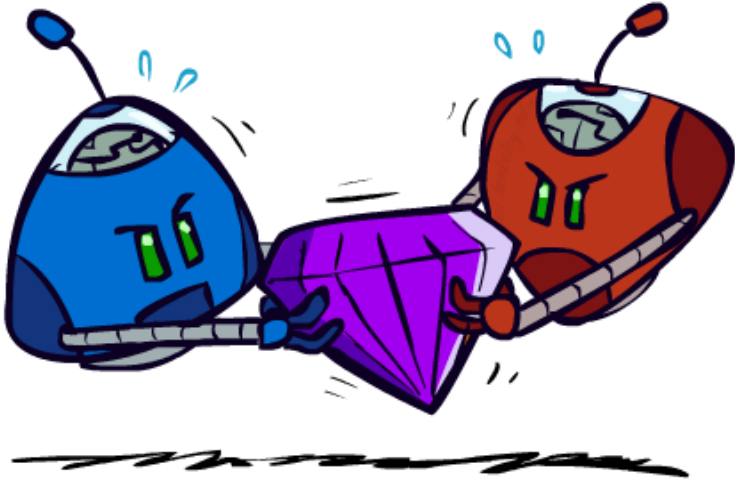


Deterministic Games

- Many possible formalizations, one is:
 - States: S (start at s_0)
 - Players: $P=\{1\dots N\}$ (usually take turns)
 - Actions: A (may depend on player / state)
 - Transition Function: $S \times A \rightarrow S$
 - Terminal Test: $S \rightarrow \{t, f\}$
 - Terminal Utilities: $S \times P \rightarrow R$
- Solution for a player is a **policy**: $S \rightarrow A$



Zero-Sum Games



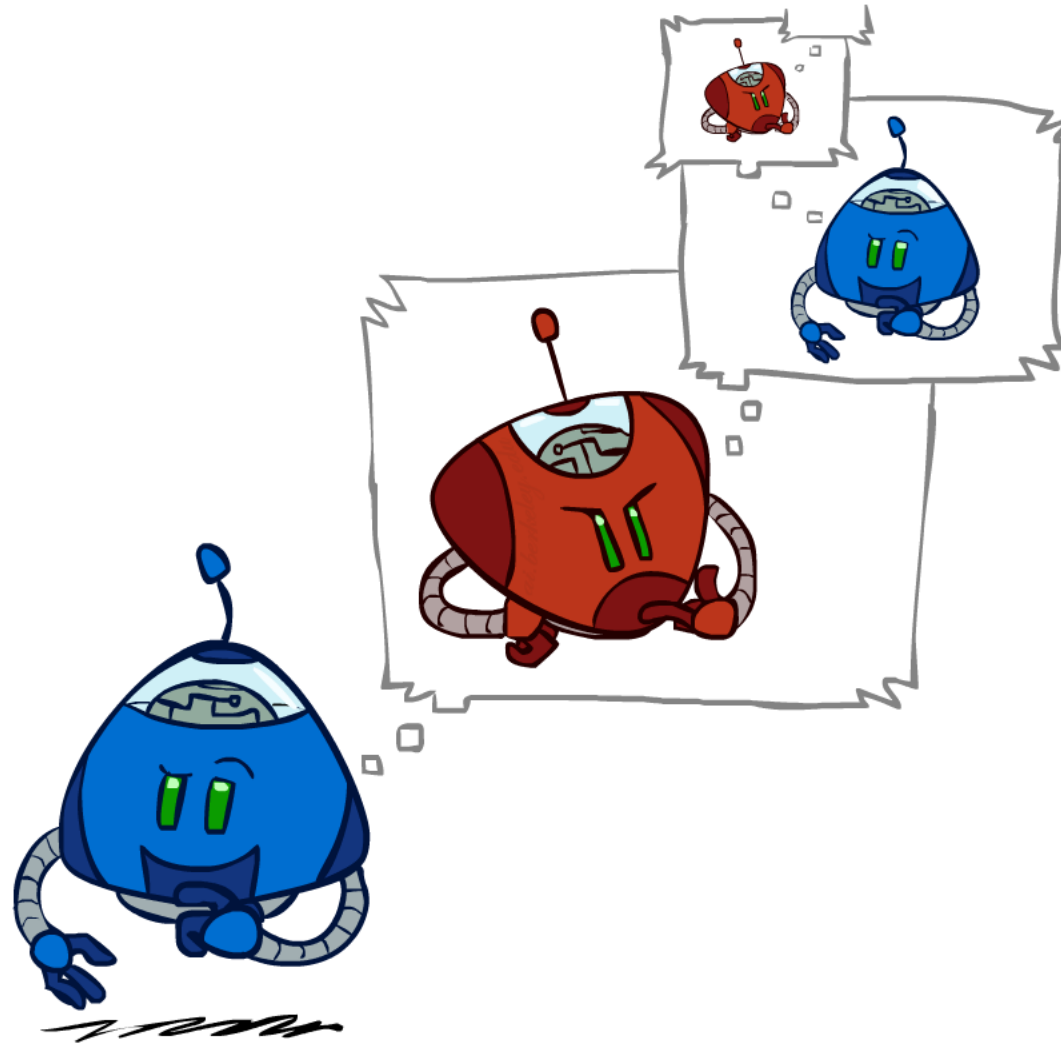
■ Zero-Sum Games

- Agents have opposite utilities (values on outcomes)
- Lets us think of a single value that one maximizes and the other minimizes
- Adversarial, pure competition

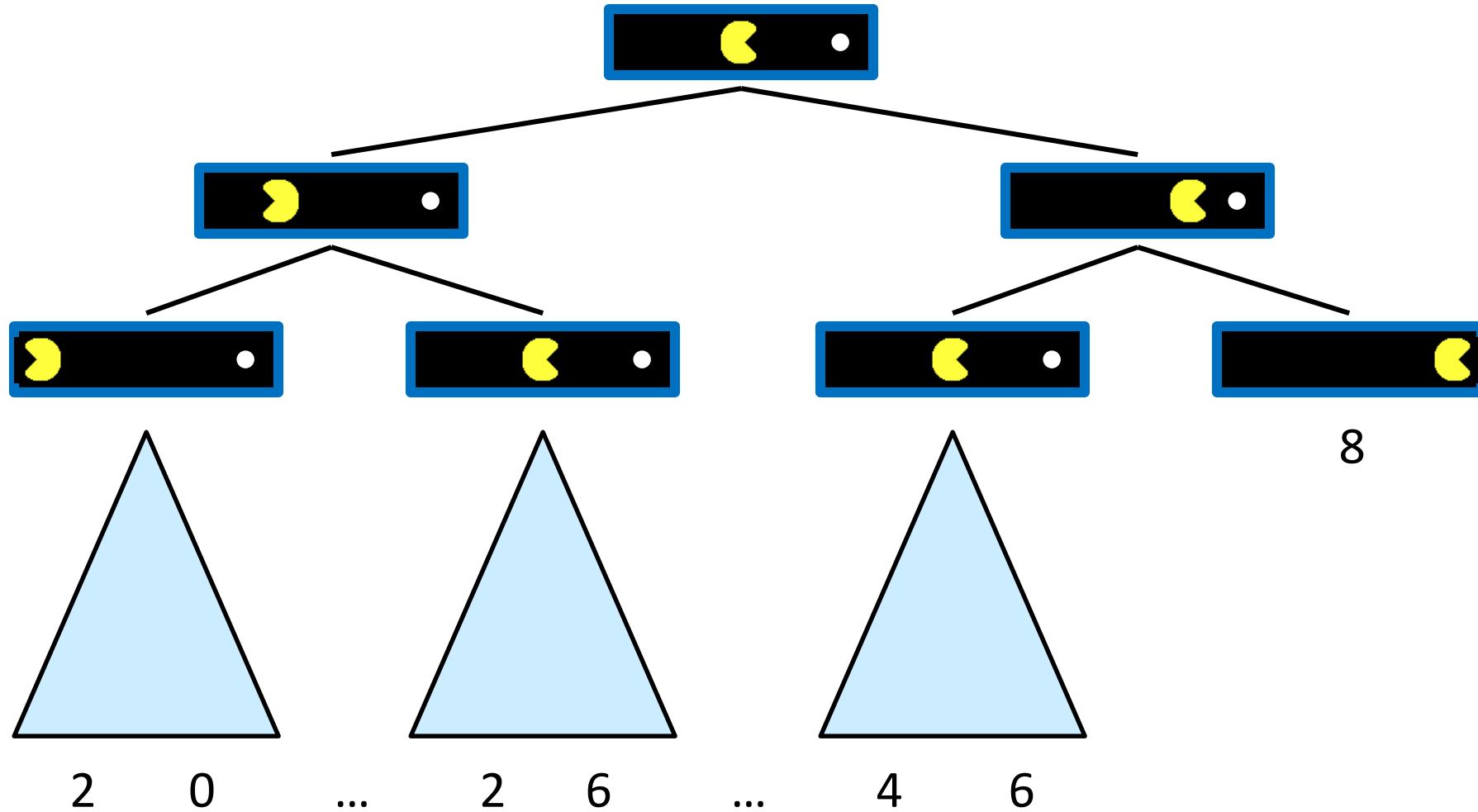
■ General Games

- Agents have independent utilities (values on outcomes)
- Cooperation, indifference, competition, and more are all possible
- More later on non-zero-sum games

Adversarial Search



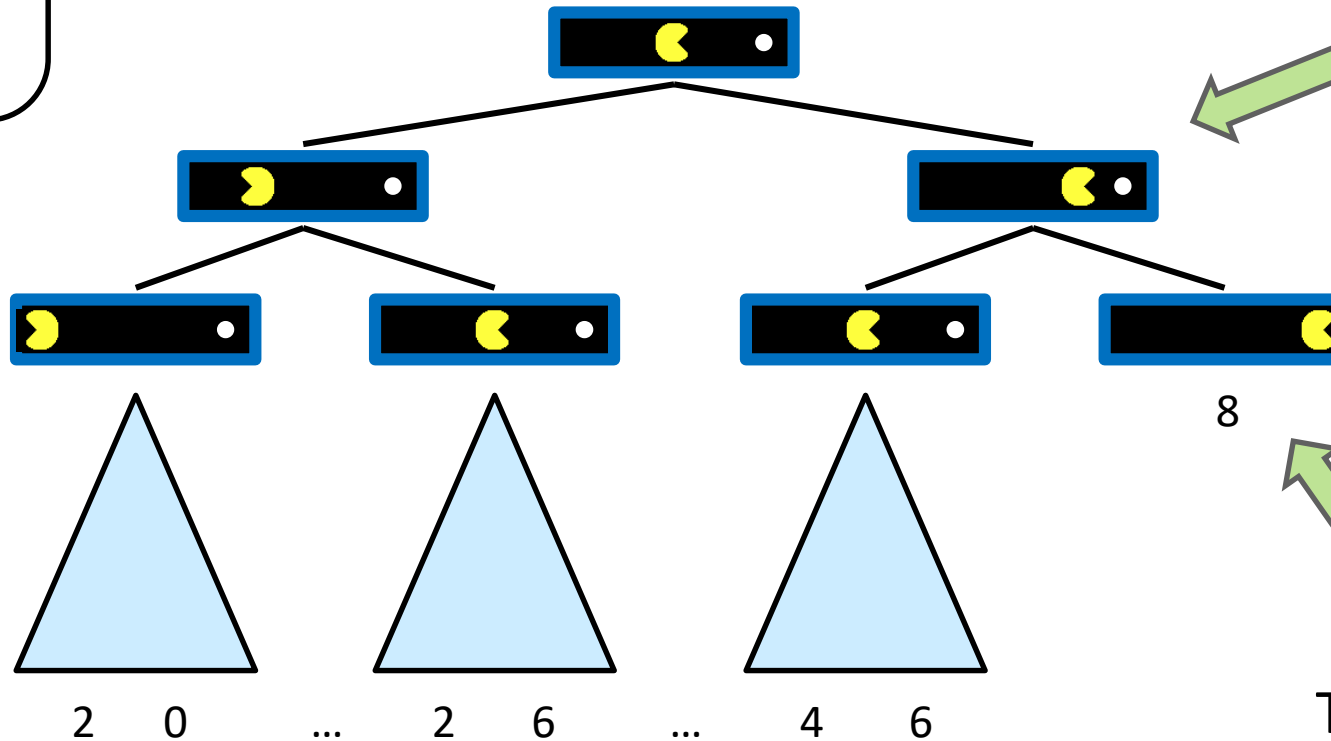
Single-Agent Trees



Value of a State

Value of a state:
The best achievable
outcome (utility)
from that state

*Policy: the agent should choose an action
leading to the state with the largest value*



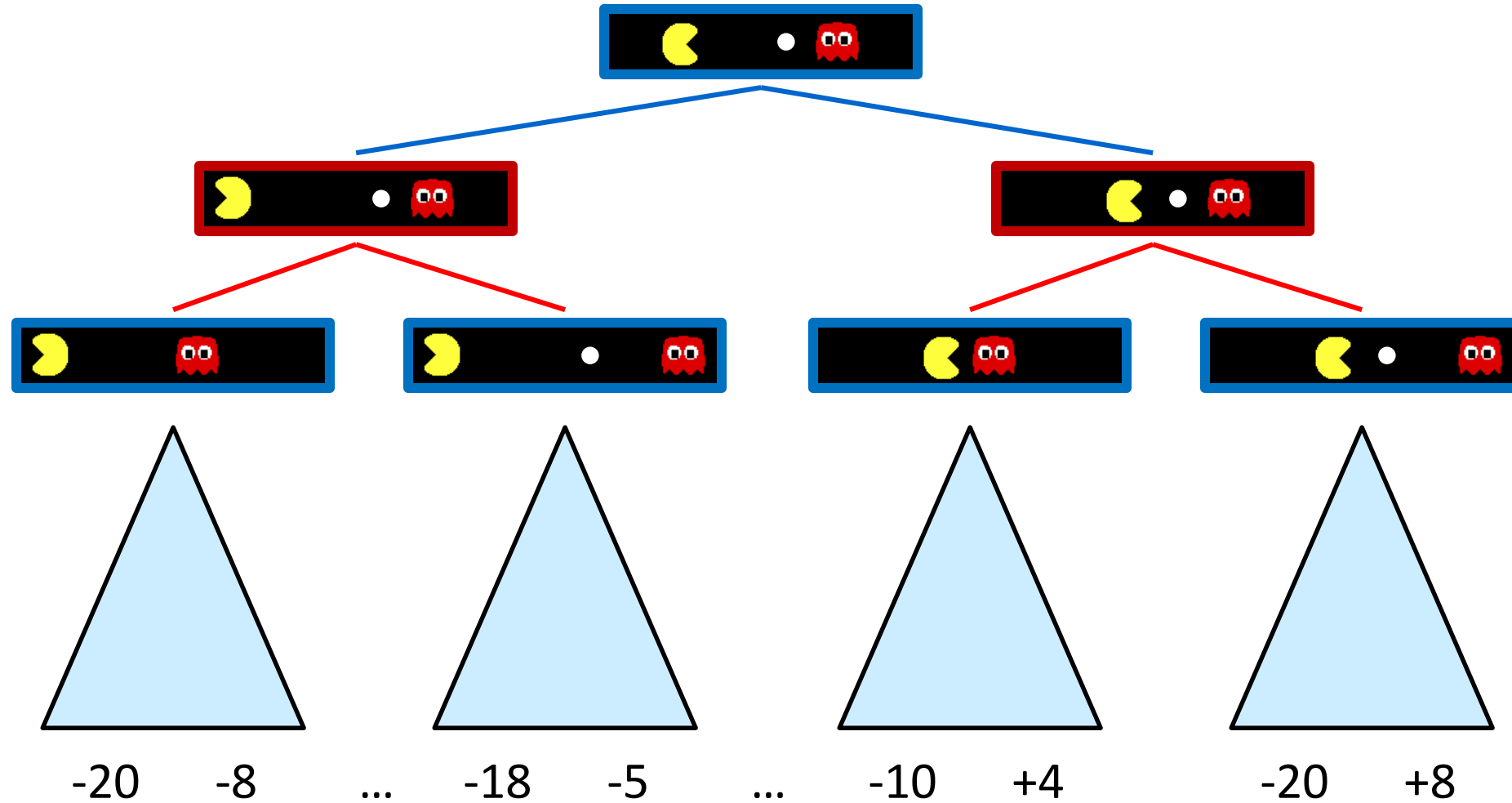
Non-Terminal States:

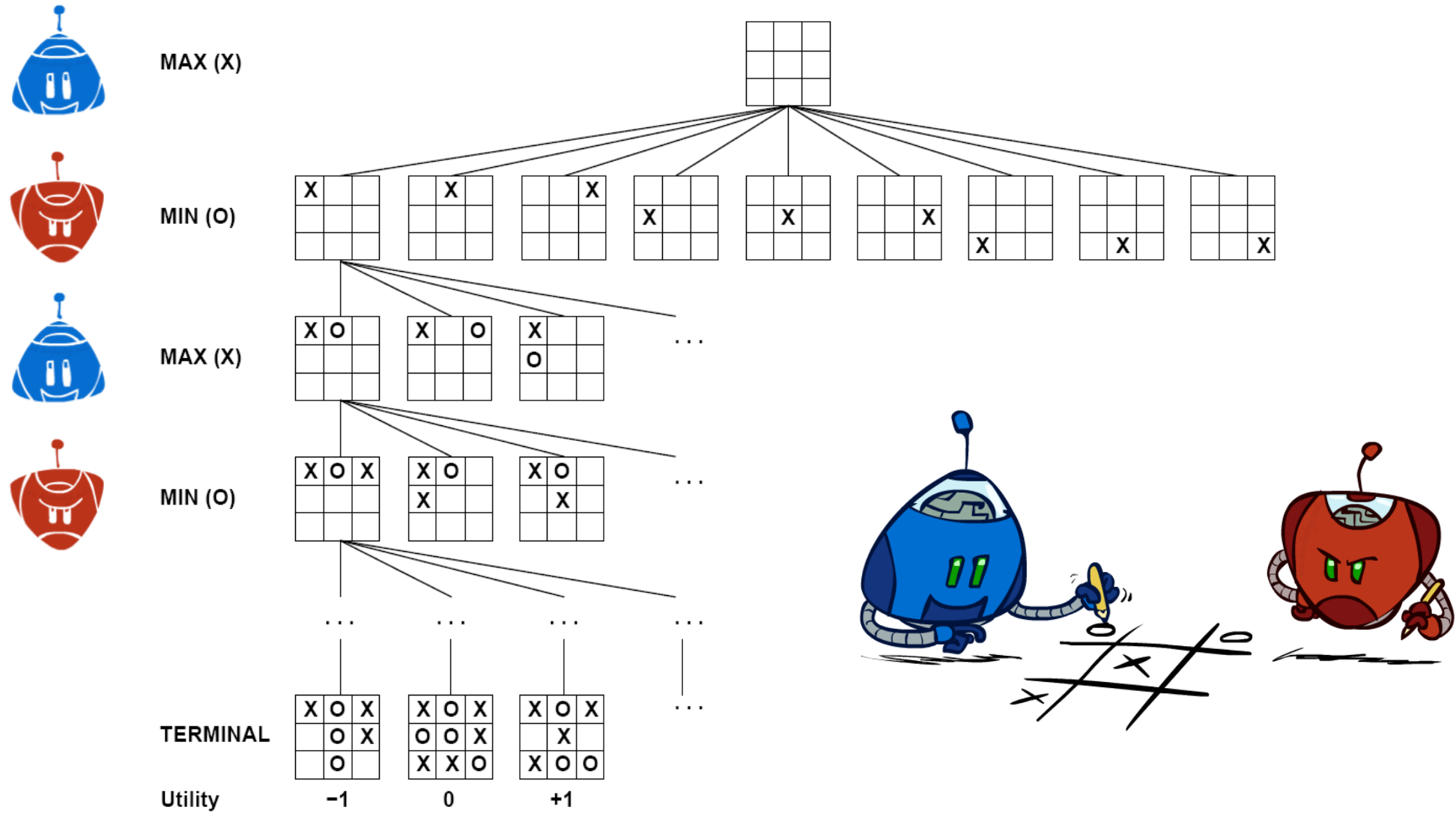
$$V(s) = \max_{s' \in \text{children}(s)} V(s')$$

Terminal States:

$$V(s) = \text{known}$$

Adversarial Game Trees





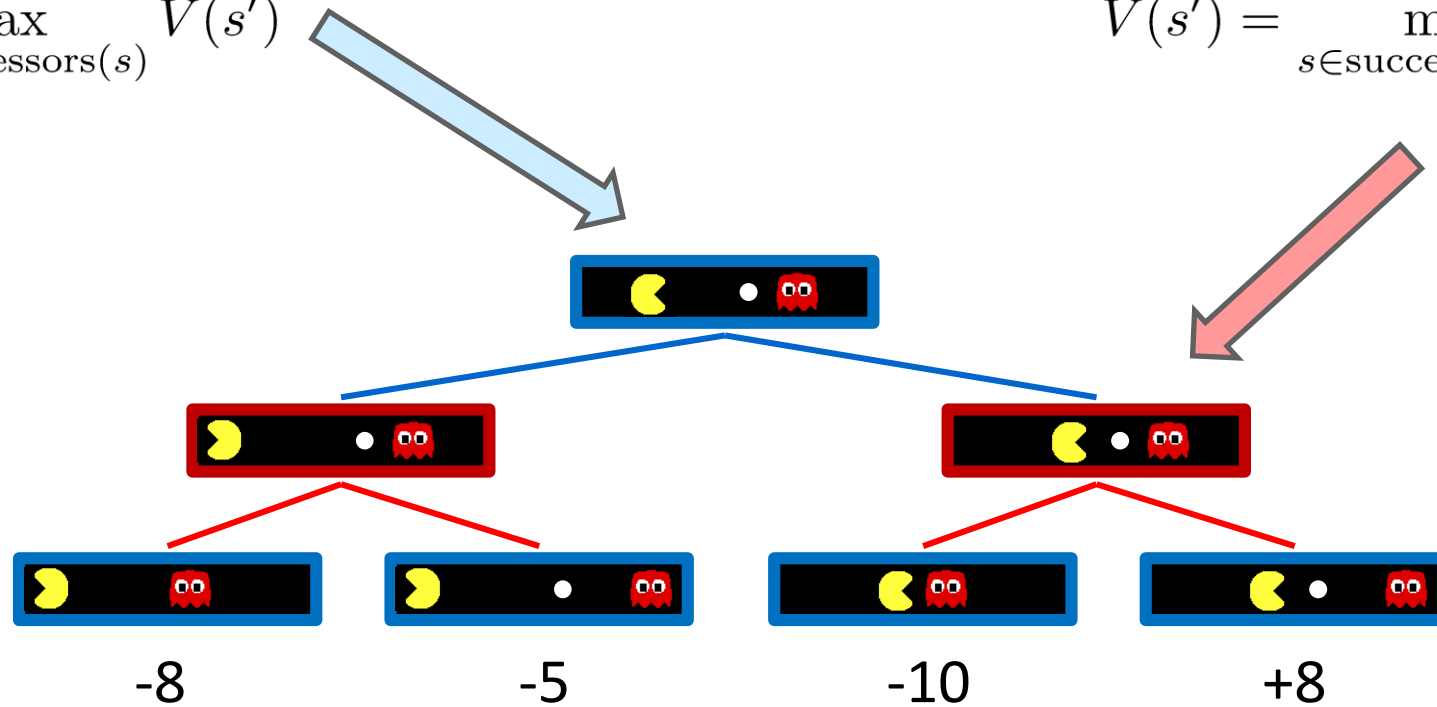
Minimax Values

States Under Agent's Control:

$$V(s) = \max_{s' \in \text{successors}(s)} V(s')$$

States Under Opponent's Control:

$$V(s') = \min_{s \in \text{successors}(s')} V(s)$$



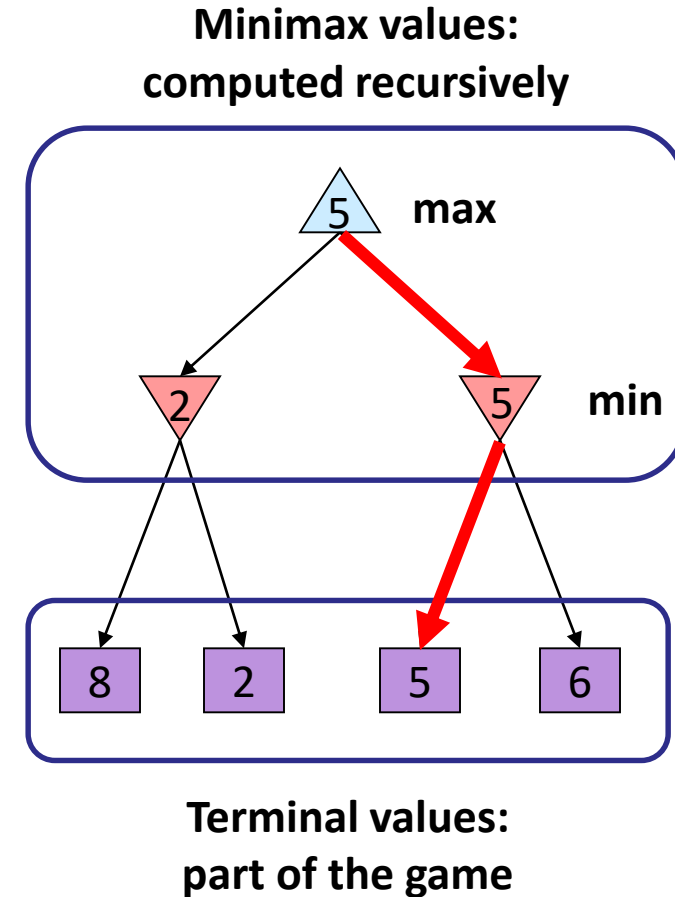
Policy: the agent should choose an action leading to the state with the largest value

Terminal States:

$$V(s) = \text{known}$$

Adversarial Search (Minimax)

- **Deterministic, zero-sum games:**
 - Tic-tac-toe, chess, checkers
 - Players alternate turns
 - One player maximizes result
 - The other minimizes result
- **Minimax search:**
 - A state-space search tree
 - Compute each node's **minimax value**: the best achievable utility against a rational (optimal) adversary



Minimax Implementation

```
def value(state):
```

if the state is a terminal state: return the state's utility

if the next agent is MAX: return max-value(state)

if the next agent is MIN: return min-value(state)

```
def max-value(state):
```

initialize $v = -\infty$

for each successor of state:

$v = \max(v, \text{value}(\text{successor}))$

return v

```
def min-value(state):
```

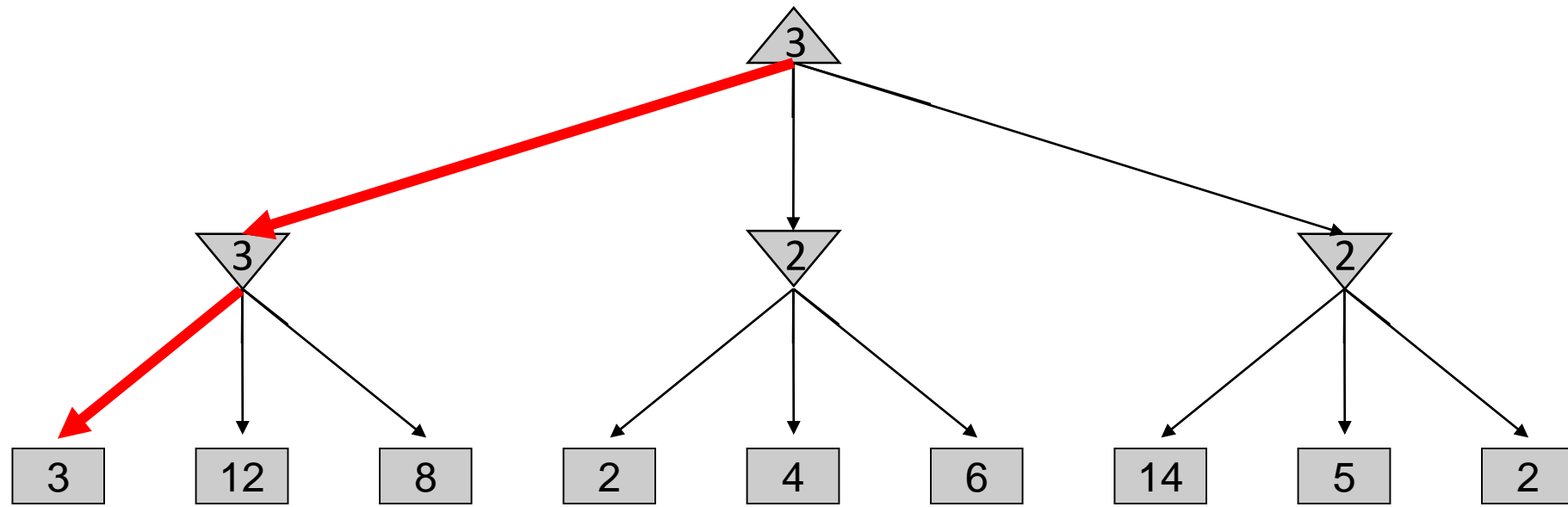
initialize $v = +\infty$

for each successor of state:

$v = \min(v, \text{value}(\text{successor}))$

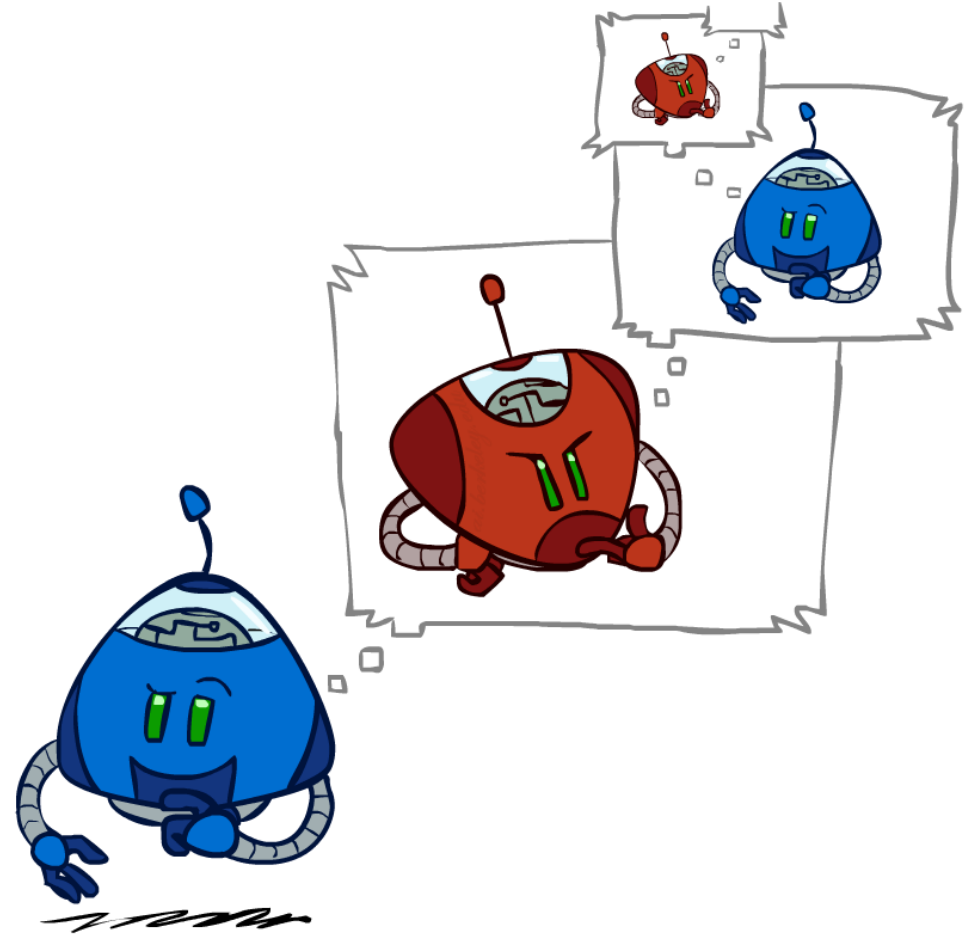
return v

Minimax Example



Minimax Efficiency

- How efficient is minimax?
 - Just like (exhaustive) DFS
 - Time: $O(b^m)$
 - Space: $O(bm)$
- Example: For chess, $b \approx 35$, $m \approx 100$
 - Exact solution is completely infeasible
 - But, do we need to explore the whole tree?



Resource Limits



Resource Limits

- Problem: In realistic games, cannot search to leaves!

- Solution: Depth-limited search

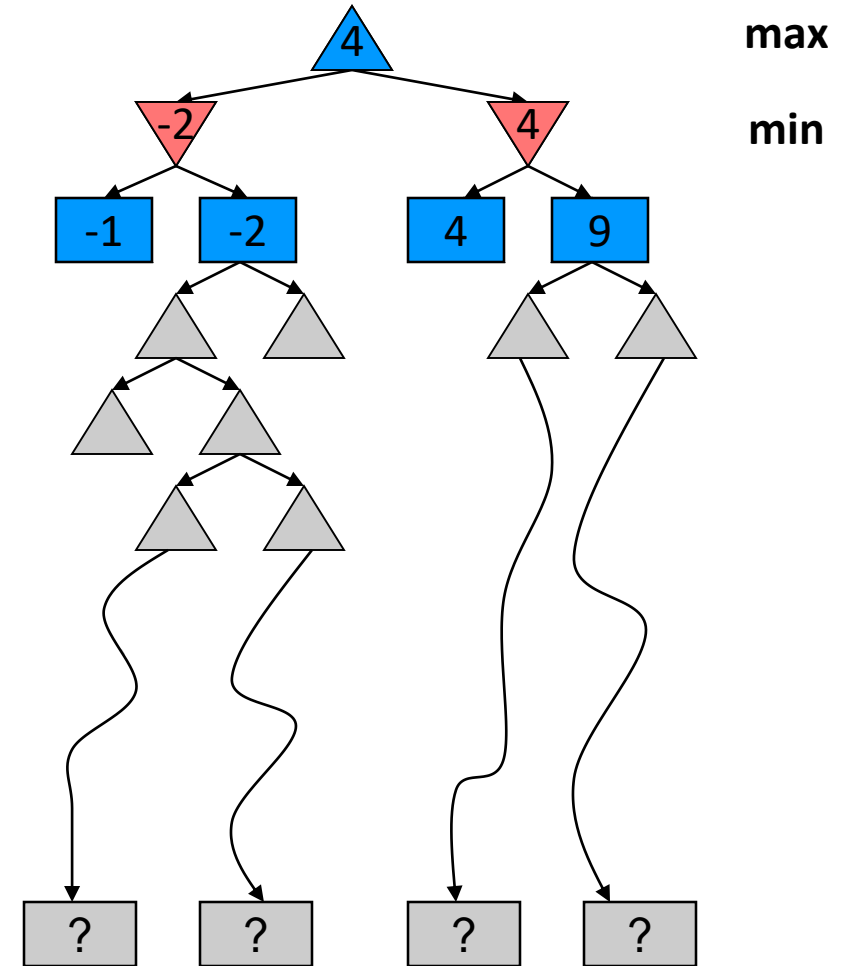
- Instead, search only to a limited depth in the tree
- Replace terminal utilities with an evaluation function for non-terminal positions

- Example:

- Suppose we have 100 seconds, can explore 10K nodes / sec
- So can check 1M nodes per move
- α - β reaches about depth 8 – decent chess program

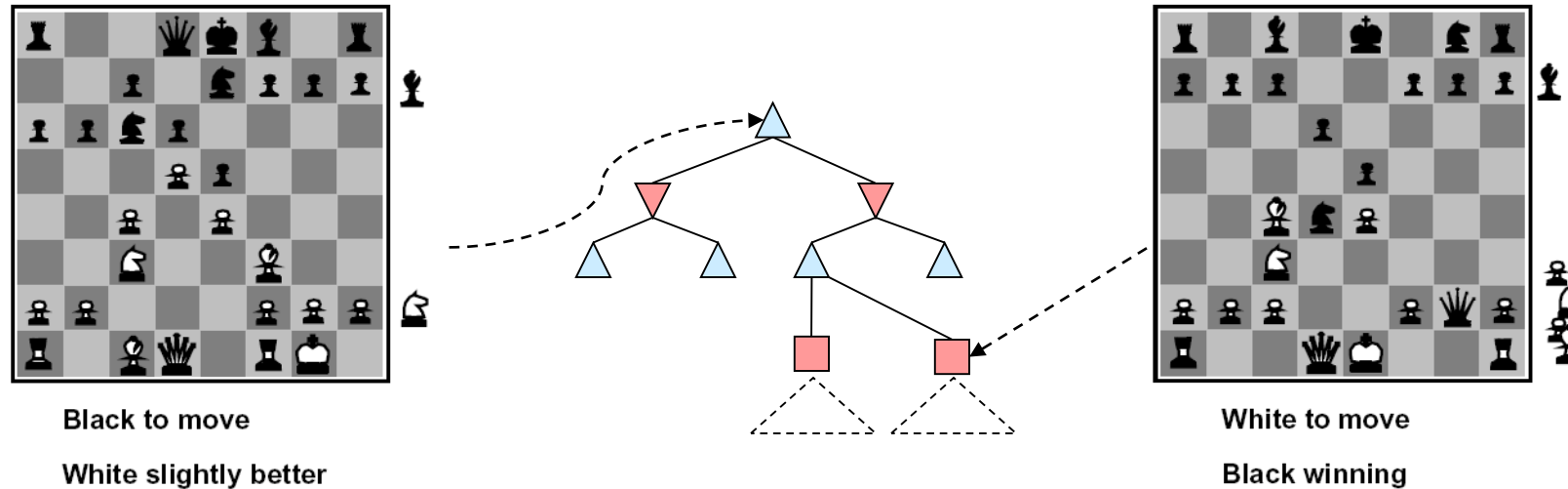
- Guarantee of optimal play is gone

- More depth makes a BIG difference



Evaluation Functions

- Evaluation functions score non-terminals in depth-limited search



- Ideal function: returns the actual minimax value of the position
- A simple solution in practice: weighted linear sum of features:

$$Eval(s) = w_1 f_1(s) + w_2 f_2(s) + \dots + w_n f_n(s)$$

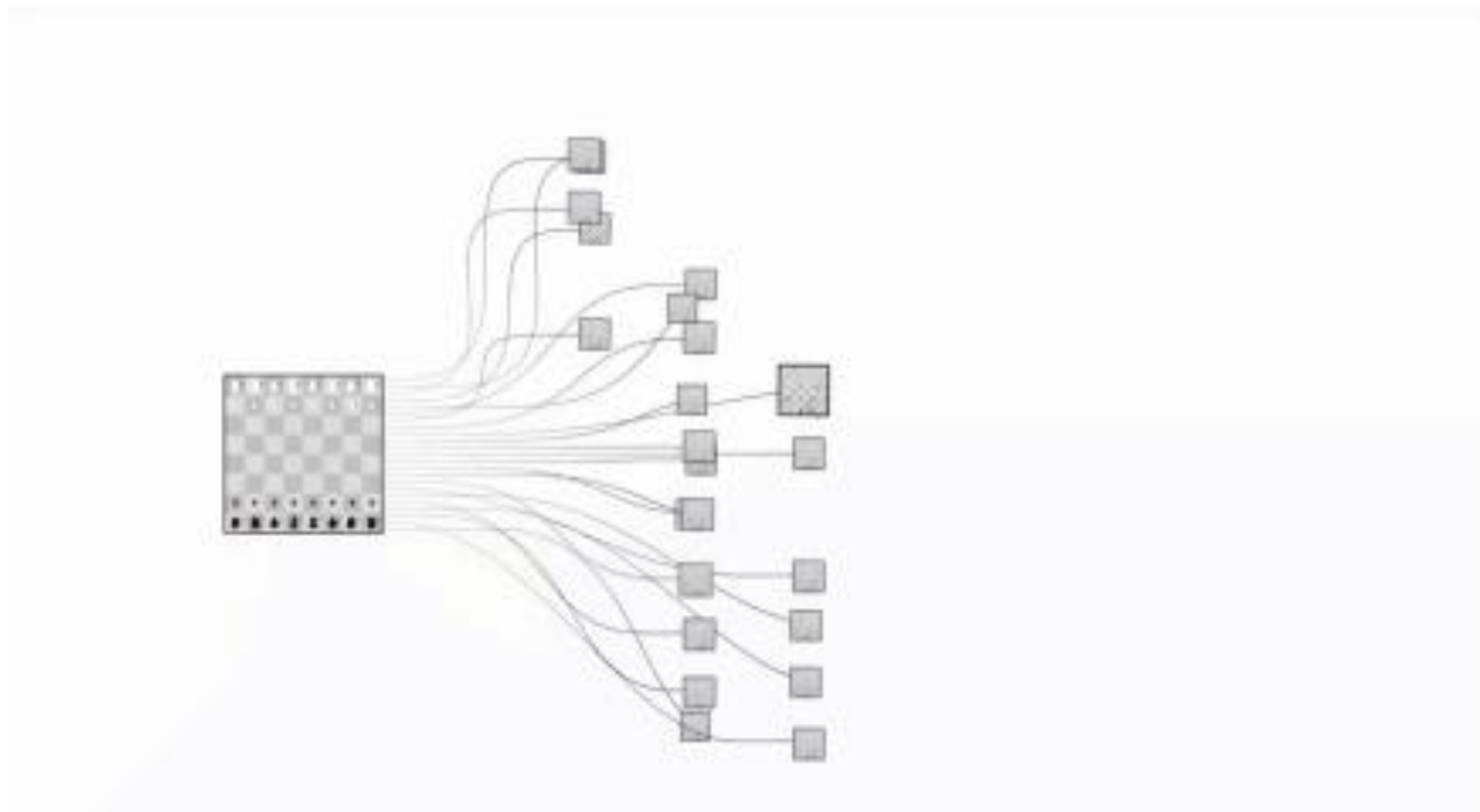
- e.g. $f_1(s) = (\text{num white queens} - \text{num black queens})$, etc.

Evaluation Functions

- More advanced solutions
 - Monte Carlo Tree Search
 - Randomly choose moves until the end of game
 - Repeat for many many times
 - Evaluate the state based on these simulations, e.g., the winning rate
 - Convolutional Neural Network (value network in AlphaGo)
 - Trained from records of game plays to predict a score of the state

Branching Factor

- Chess



Branching Factor

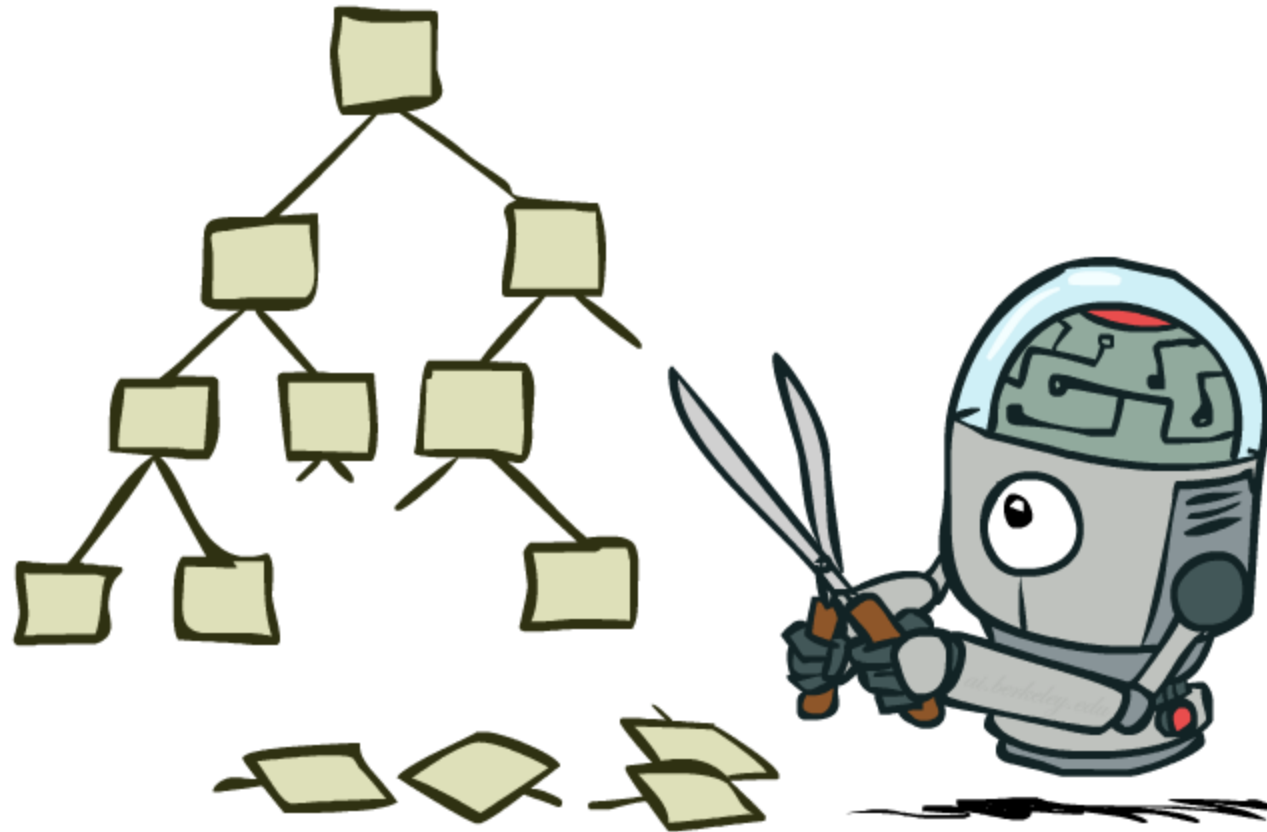
- Go



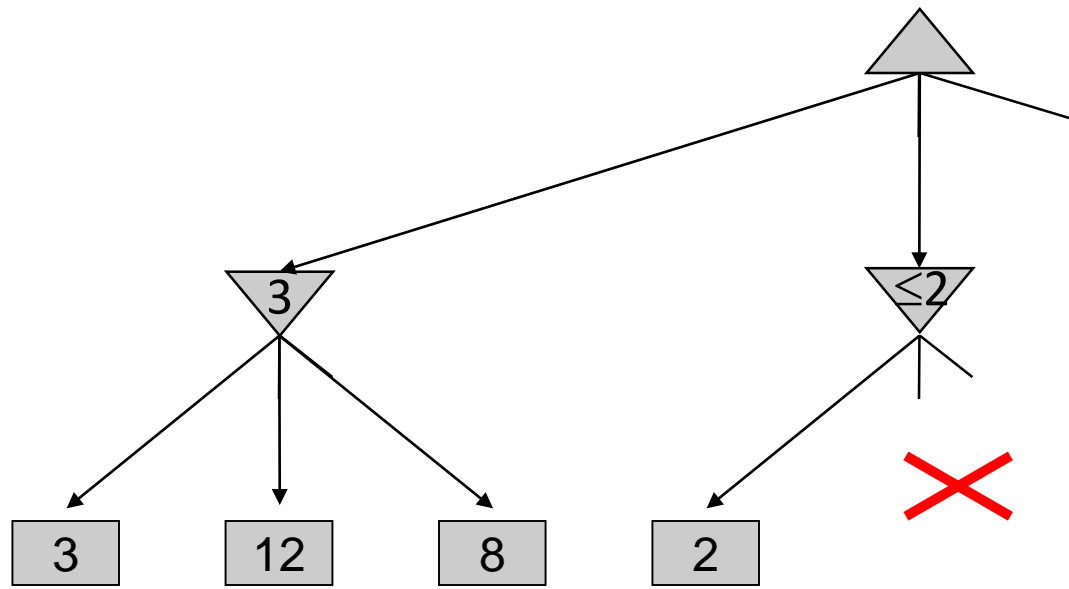
Branching Factor

- Go has a branching factor of up to 361
- Idea: limit the branching factor by considering only good moves
 - AlphaGo uses a Convolutional Neural Network (policy network)
 - Trained from records of game plays
 - Trained using reinforcement learning
 - AlphaGo Zero uses RL only

Game Tree Pruning

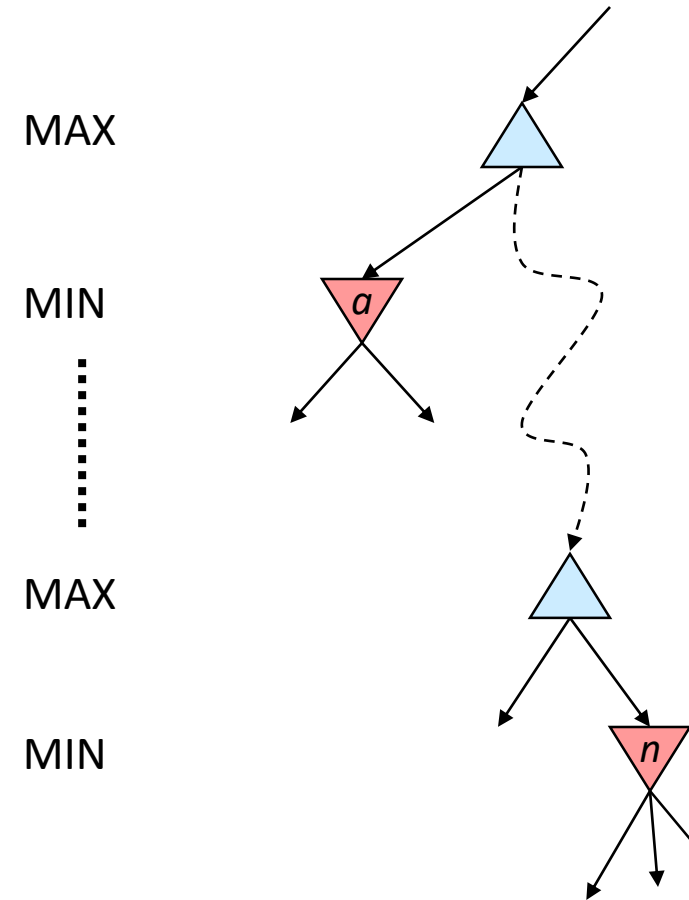


Minimax Pruning



Alpha-Beta Pruning

- General configuration (MIN version)
 - We're computing the MIN-VALUE at some node n
 - We're looping over n 's children, so n 's estimate is decreasing
 - Let a be the best value that MAX can get at any choice point along the current path from the root
 - If n becomes worse than a , then we can stop considering n 's other children
 - Reason: if n is eventually chosen, then the nodes along the path shall all have the value of n , but n is worse than a and hence the path shall not be chosen at the MAX
- MAX version is symmetric



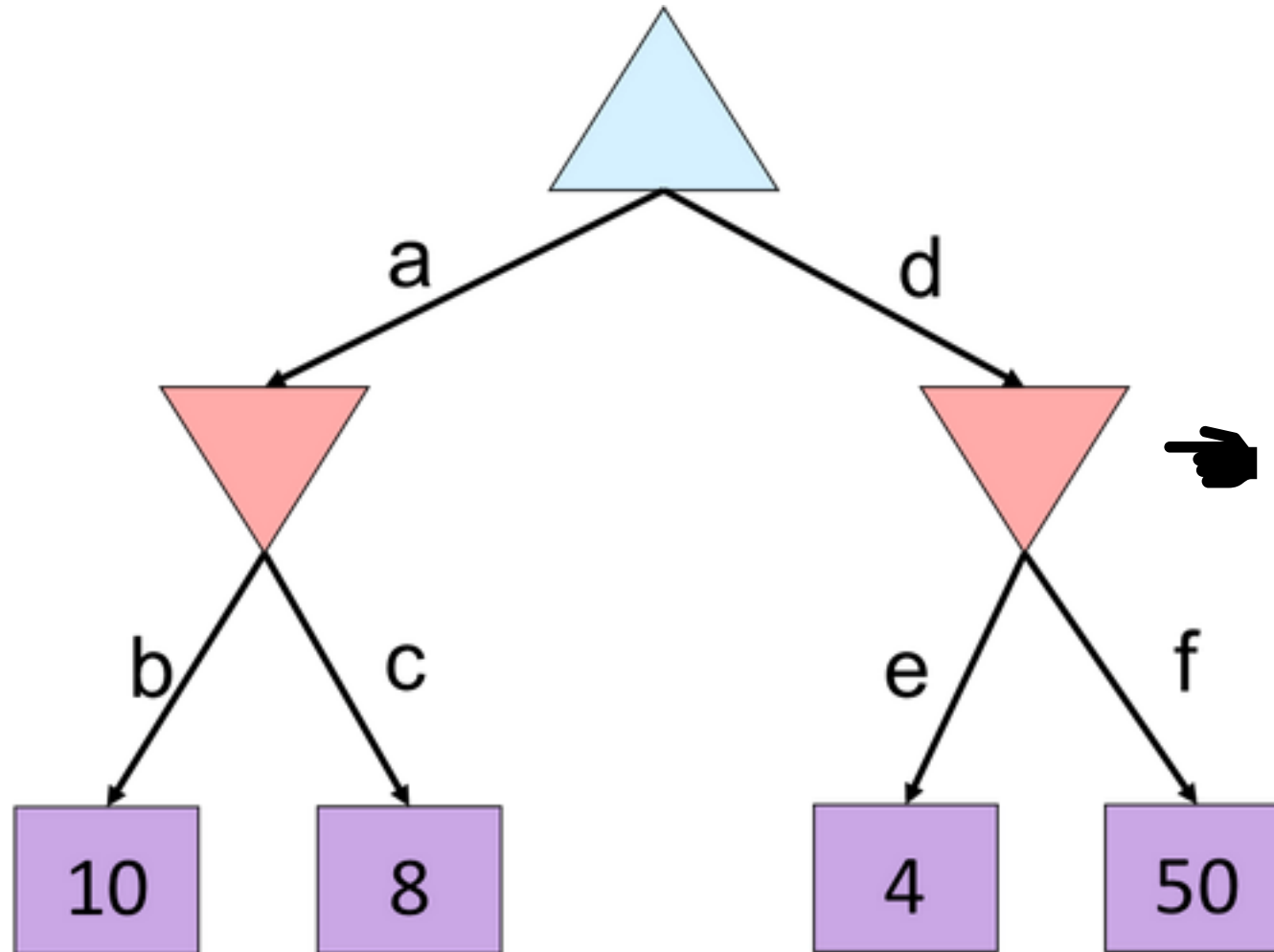
Alpha-Beta Implementation

α : MAX's best option on path to root
 β : MIN's best option on path to root

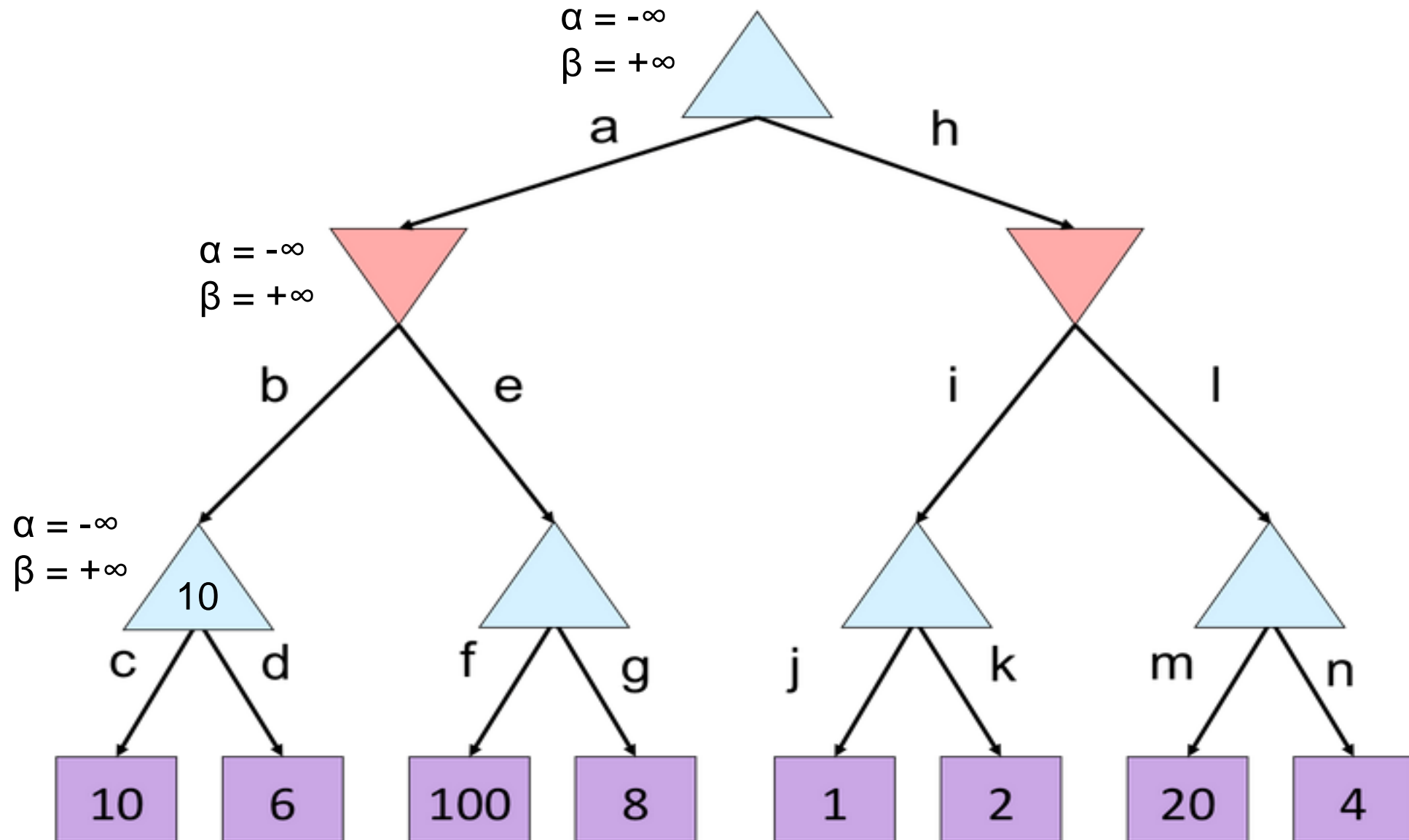
```
def max-value(state,  $\alpha$ ,  $\beta$ ):  
    initialize  $v = -\infty$   
    for each successor of state:  
         $v = \max(v, \text{value}(\text{successor}, \alpha, \beta))$   
        if  $v \geq \beta$  return  $v$   
         $\alpha = \max(\alpha, v)$   
    return  $v$ 
```

```
def min-value(state,  $\alpha$ ,  $\beta$ ):  
    initialize  $v = +\infty$   
    for each successor of state:  
         $v = \min(v, \text{value}(\text{successor}, \alpha, \beta))$   
        if  $v \leq \alpha$  return  $v$   
         $\beta = \min(\beta, v)$   
    return  $v$ 
```

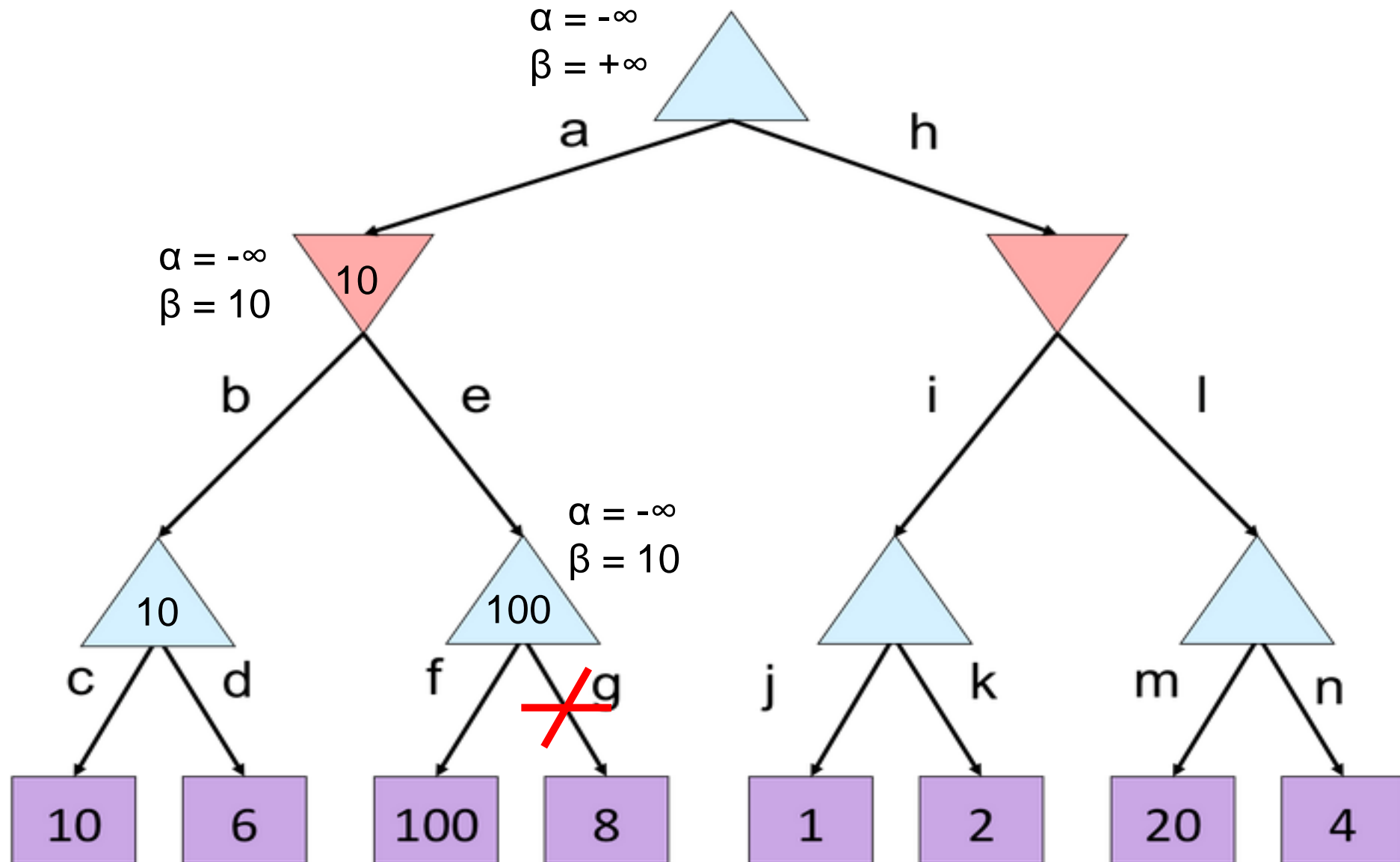
Alpha-Beta Example



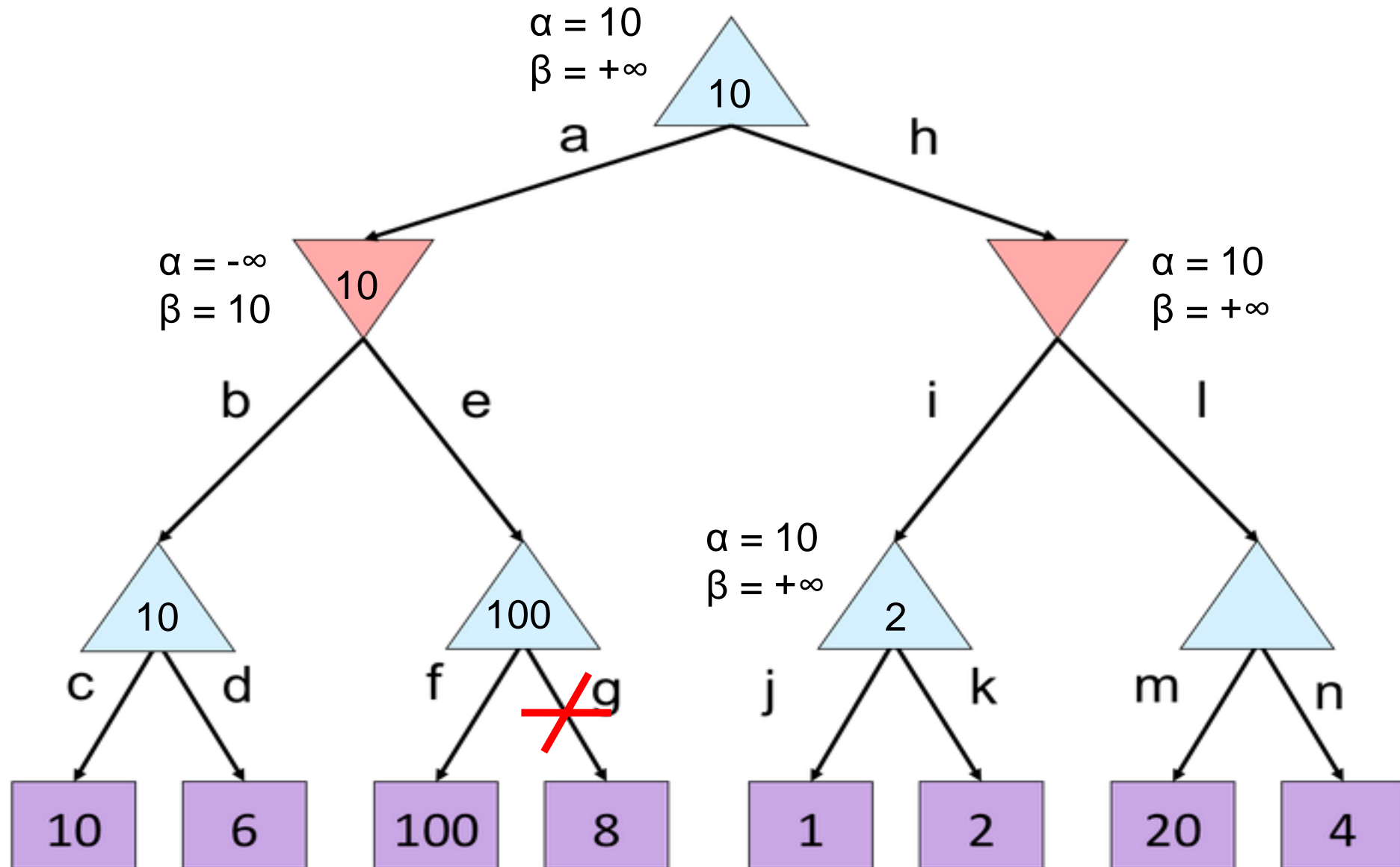
Alpha-Beta Example 2



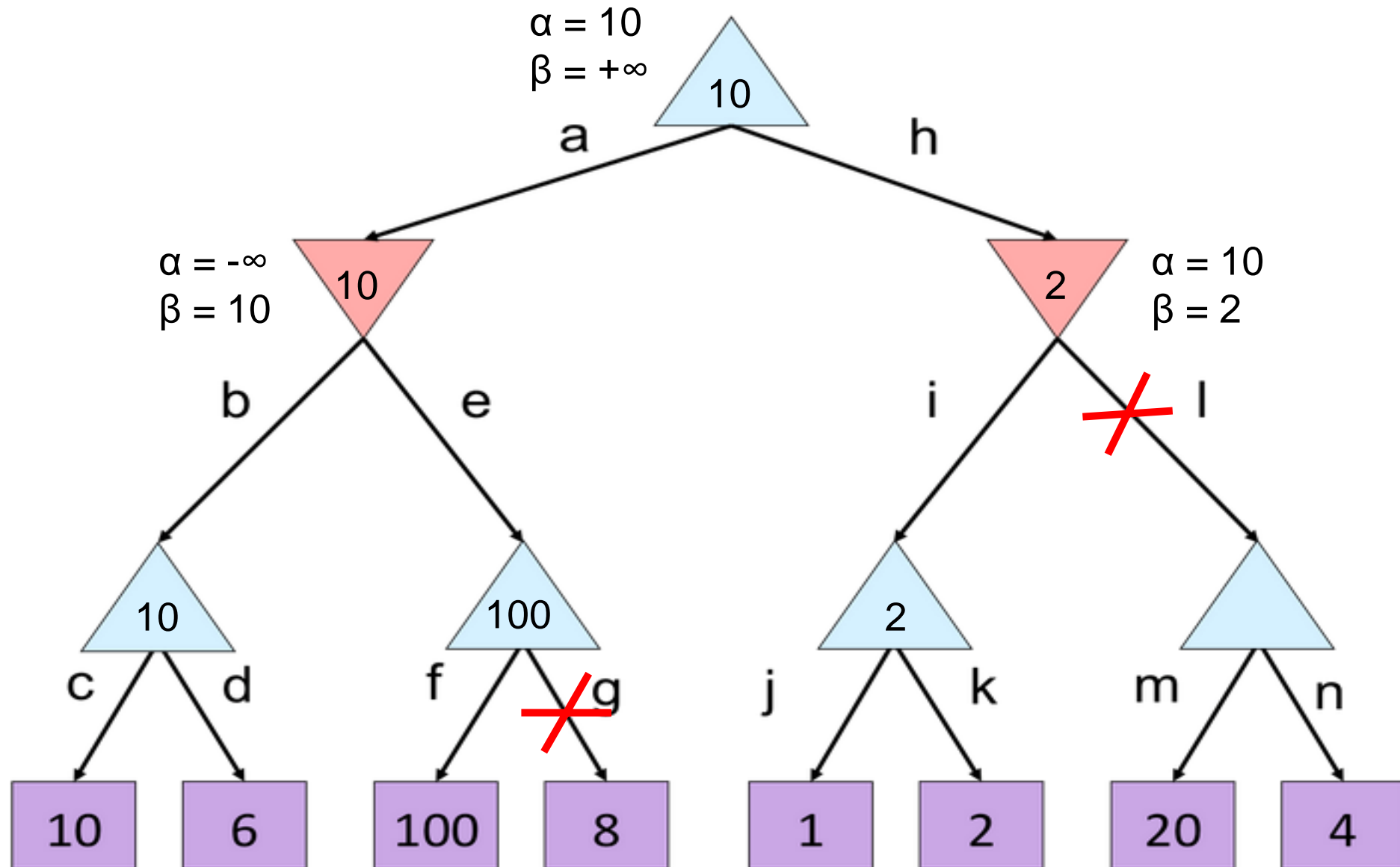
Alpha-Beta Example 2



Alpha-Beta Example 2

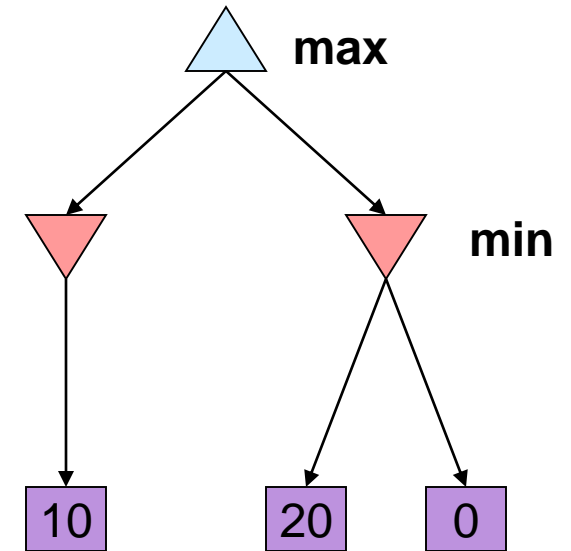


Alpha-Beta Example 2

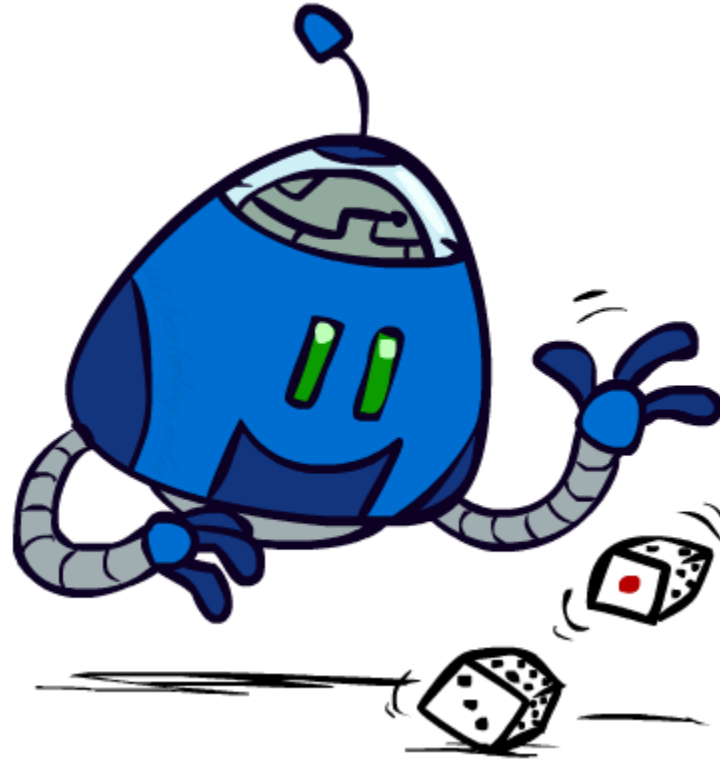


Alpha-Beta Pruning Properties

- Good child ordering improves effectiveness of pruning
- With “perfect ordering”:
 - Time complexity drops to $O(b^{m/2})$
 - Doubles solvable depth!

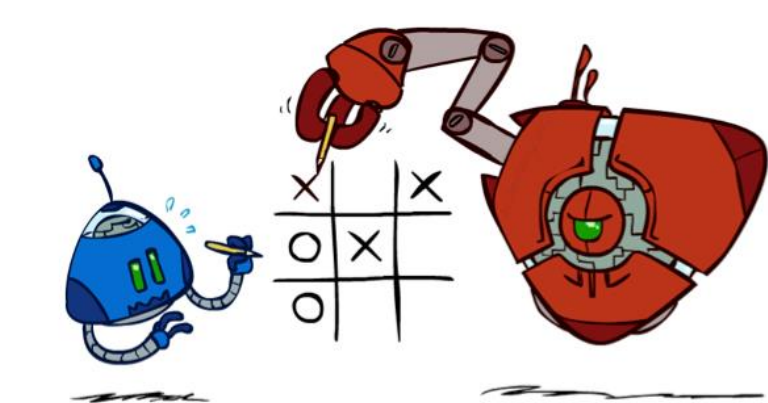
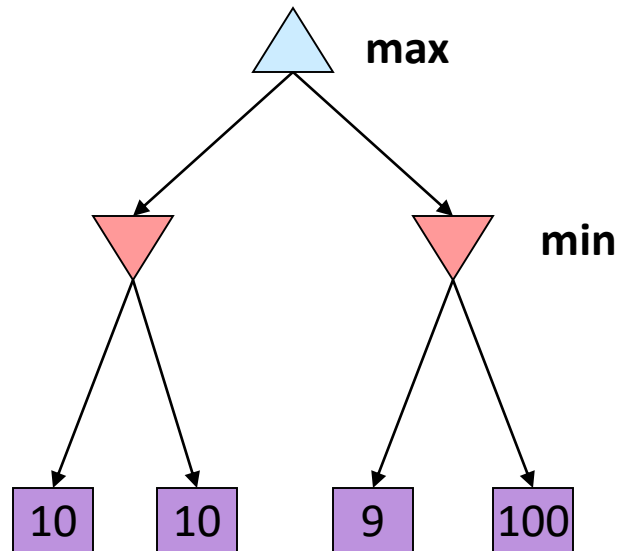


Uncertain Outcomes



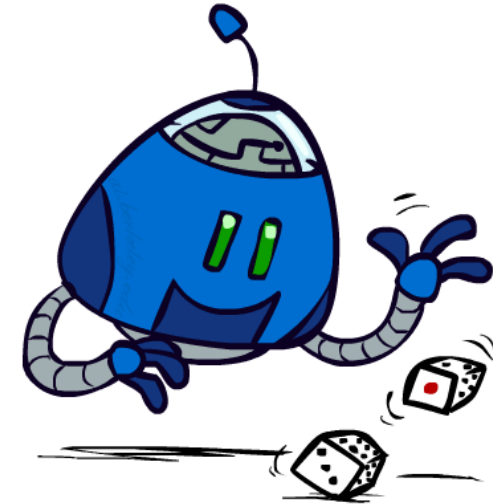
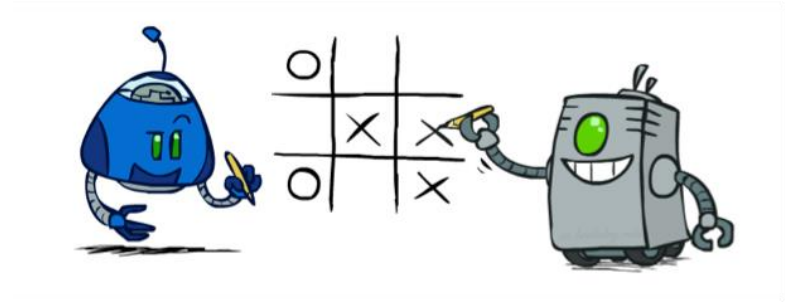
Worst-Case vs. Average Case

- The hidden assumption behind minimax
 - Your opponent is rational and smart



Worst-Case vs. Average Case

- What if...
 - Unpredictable opponents
 - E.g., the ghosts respond randomly
 - Explicit randomness
 - E.g., rolling dice
 - Actions can fail
 - E.g., when moving a robot, wheels might slip



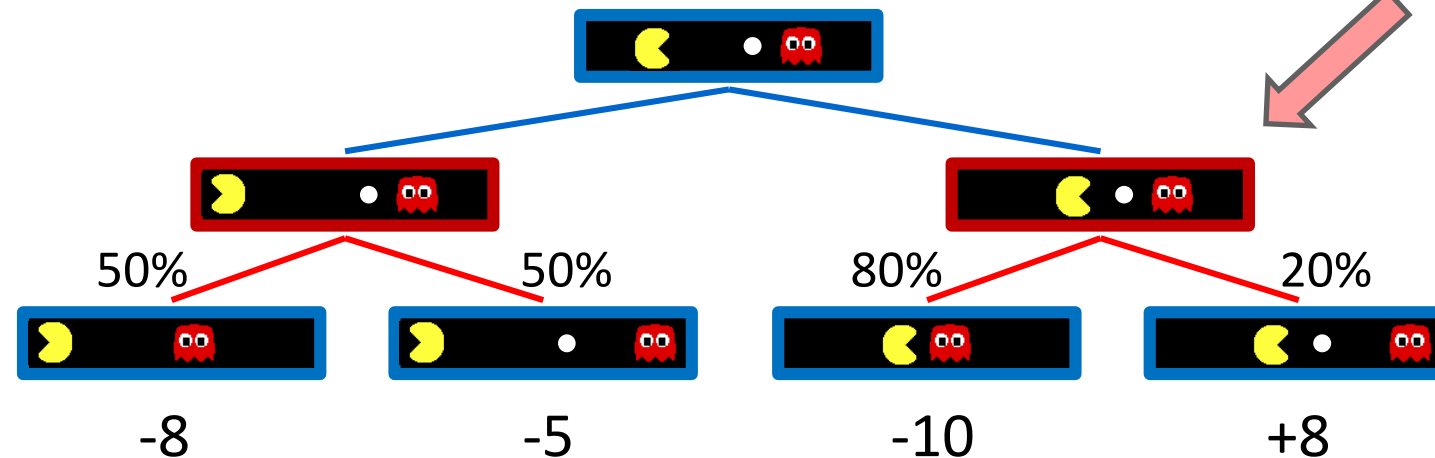
State Values

States Under Agent's Control:

$$V(s) = \max_{s' \in \text{successors}(s)} V(s')$$

States Under Opponent's Control:

$$V(s') = \sum_{s \in \text{successors}(s')} P(s) \times V(s)$$



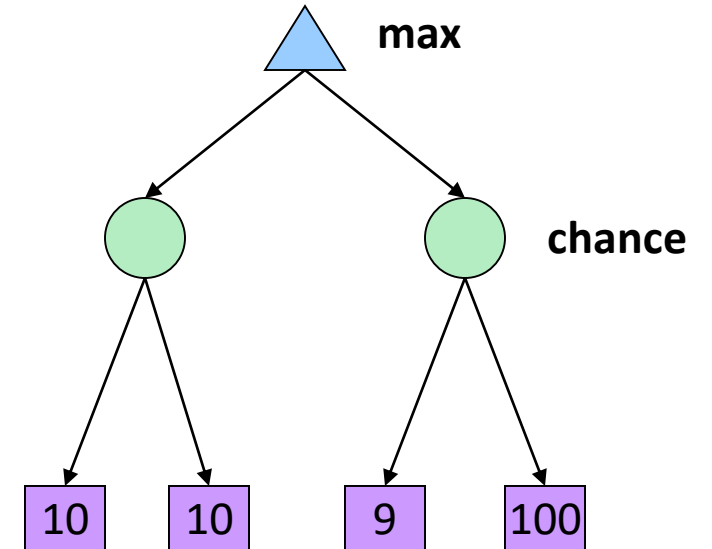
Value of a state: The best achievable **expected** utility from that state

Terminal States:

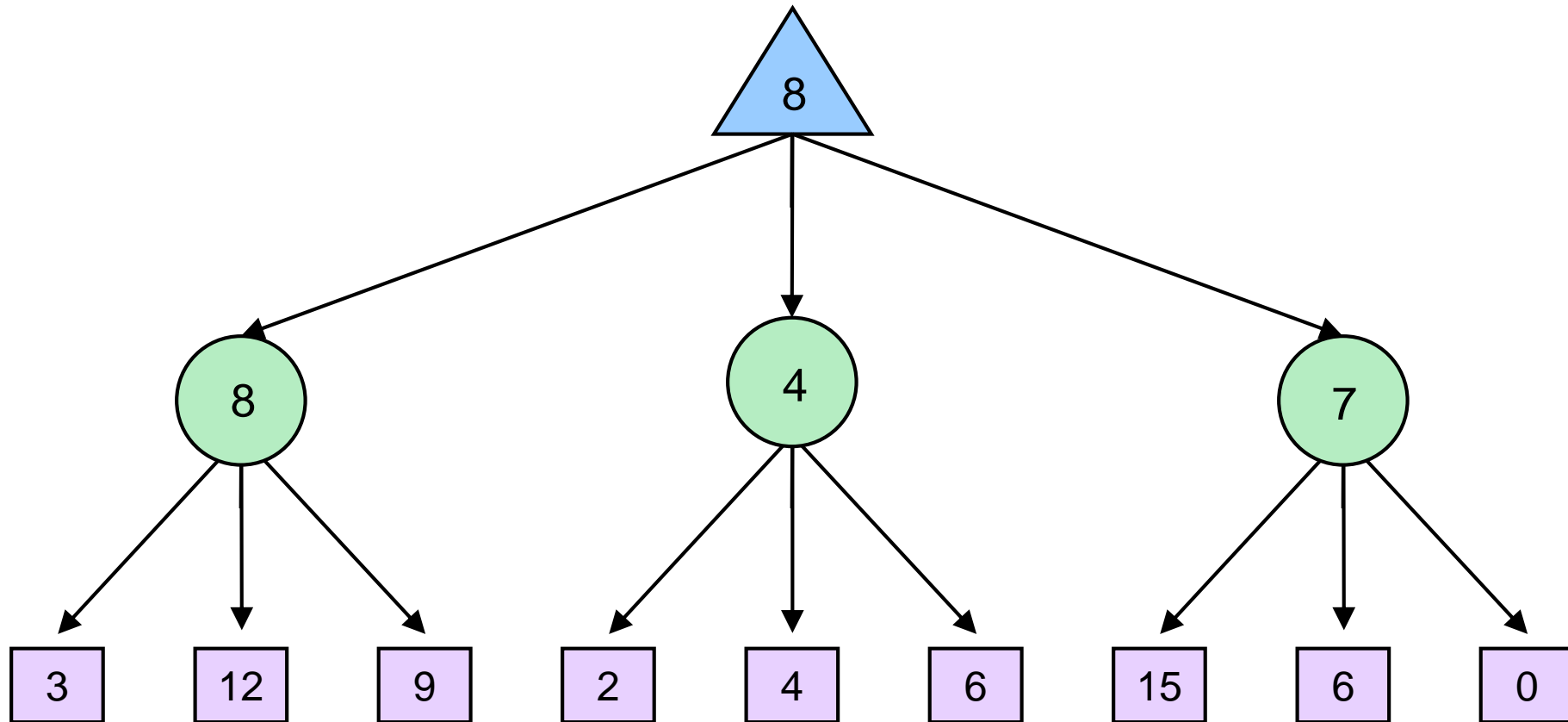
$$V(s) = \text{known}$$

Expectimax Search

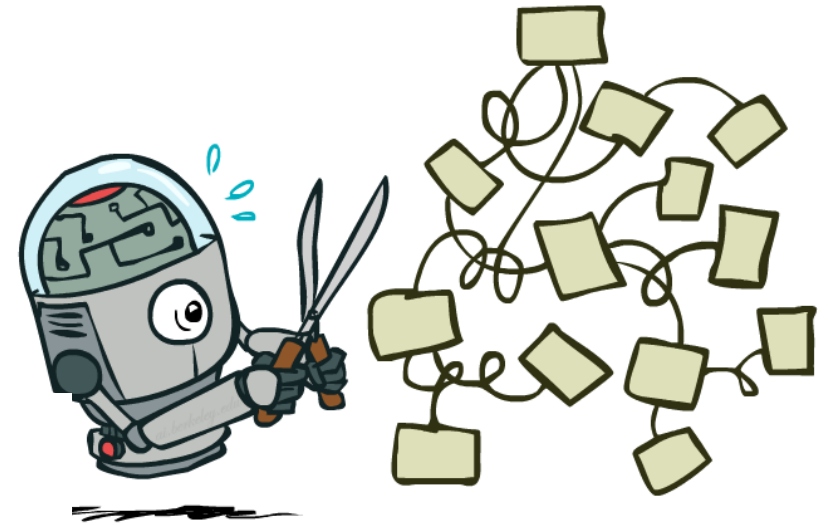
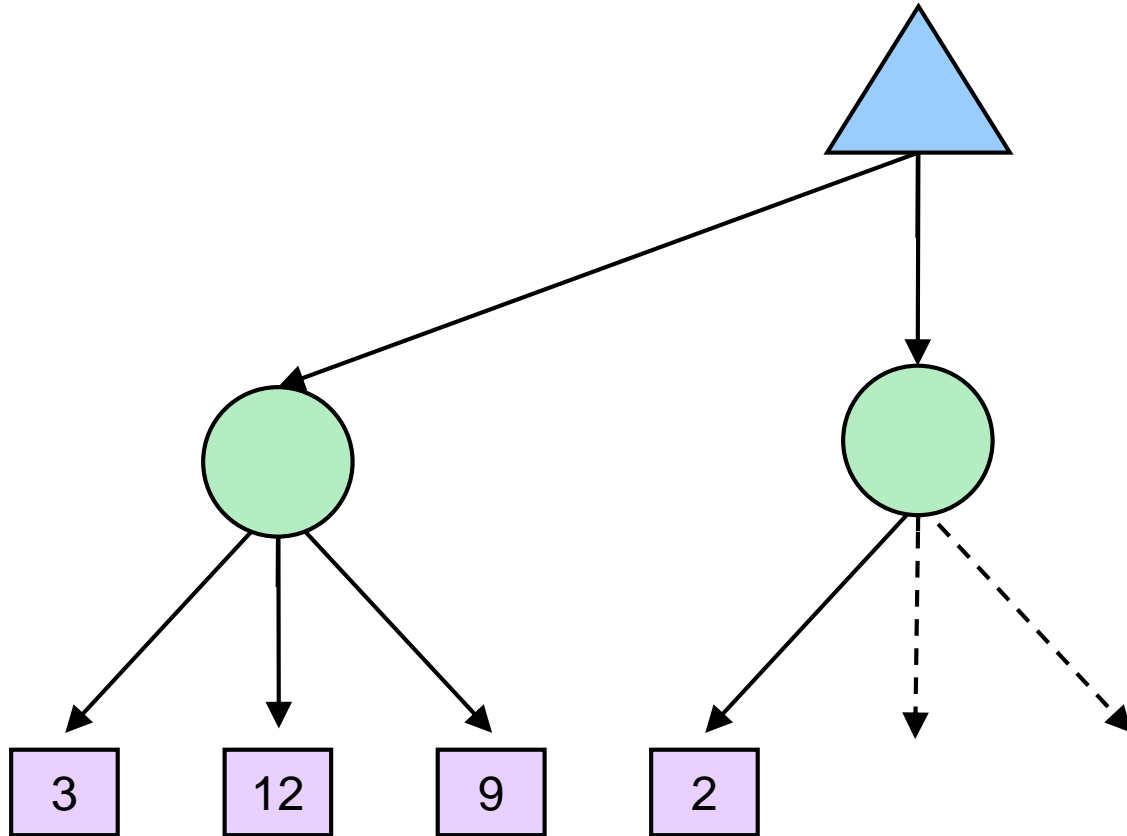
- **Expectimax search:** compute the average score under optimal play
 - Max nodes as in minimax search
 - Chance nodes are like min nodes but the outcome is uncertain
 - Calculate their **expected utilities**, i.e. taking weighted average (expectation) of children
- Later, we'll learn how to formalize the underlying uncertain-result problems as **Markov Decision Processes**



Expectimax Example



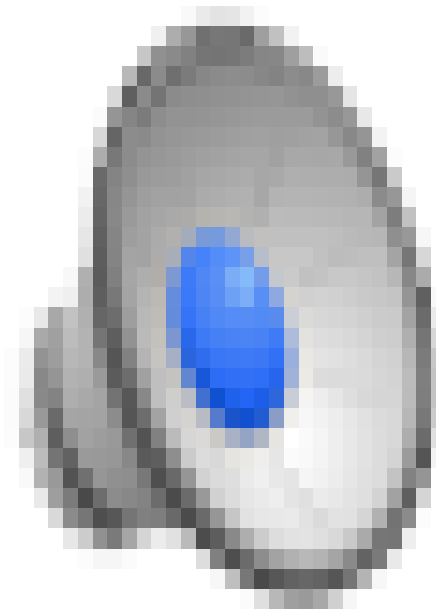
Expectimax Pruning?



Video of Demo Minimax vs Expectimax (Min)

The game:

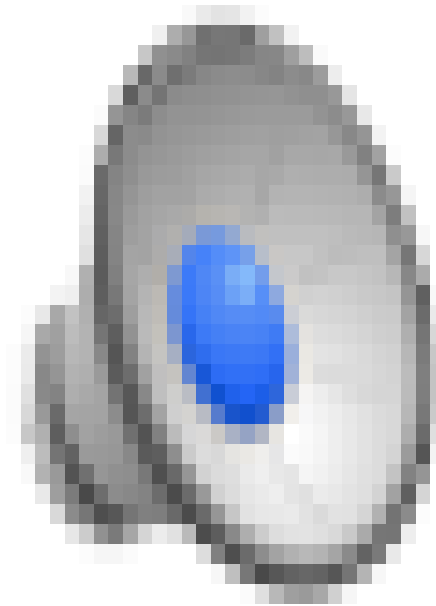
- +10 for eating each dot
- +500 for eating all the dots
- 500 for being eaten
- 1 for each move



Video of Demo Minimax vs Expectimax (Exp 1)

The game:

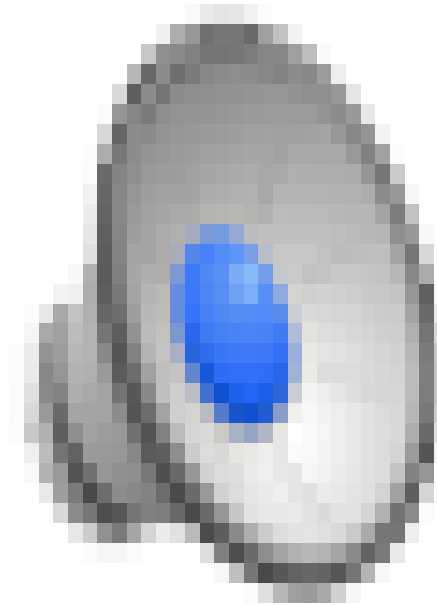
- +10 for eating each dot
- +500 for eating all the dots
- 500 for being eaten
- 1 for each move



Video of Demo Minimax vs Expectimax (Exp 2)

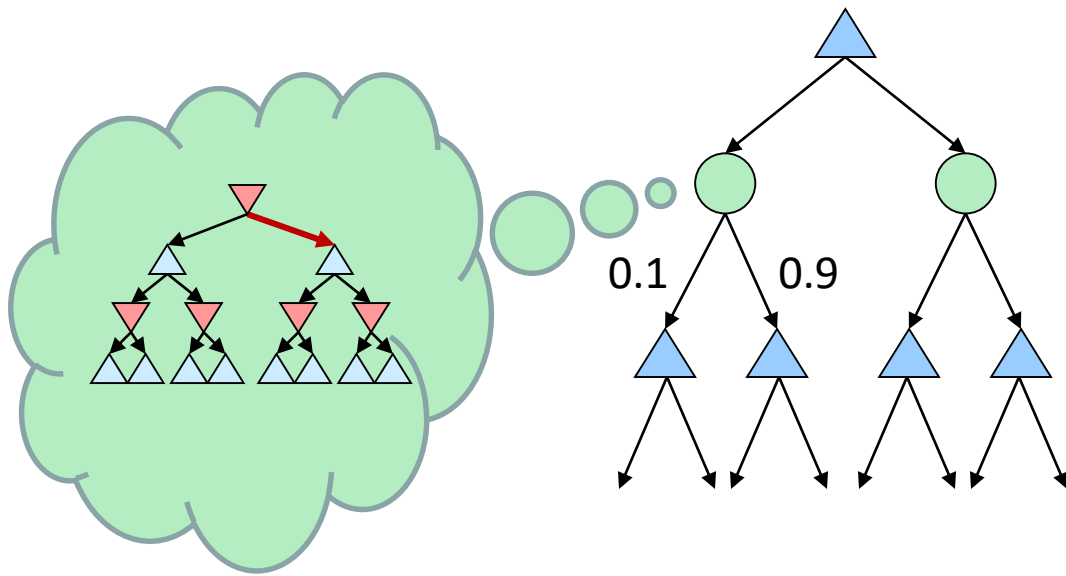
The game:

- +10 for eating each dot
- +500 for eating all the dots
- 500 for being eaten
- 1 for each move



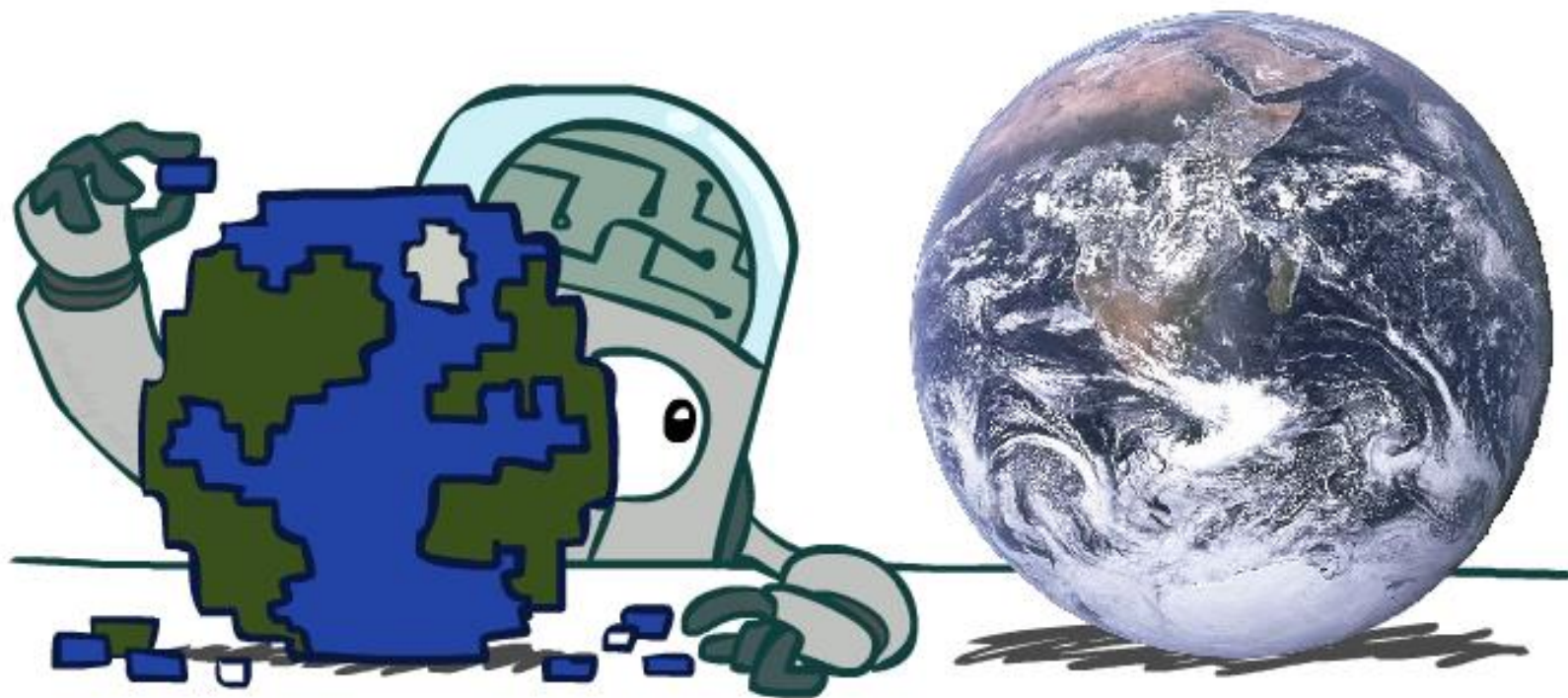
Quiz: Informed Probabilities

- Let's say you know that your opponent is actually running a depth 2 minimax, using the result 80% of the time, and moving randomly otherwise
- Question: What tree search should you use?



- Answer: Expectimax!
 - To figure out EACH chance node's probabilities, you have to run a simulation of your opponent
 - This kind of thing gets very slow very quickly
 - Even worse if you have to simulate your opponent simulating you...

Modeling Assumptions



The Dangers of Optimism and Pessimism

Dangerous Optimism

Assuming chance when the world is adversarial

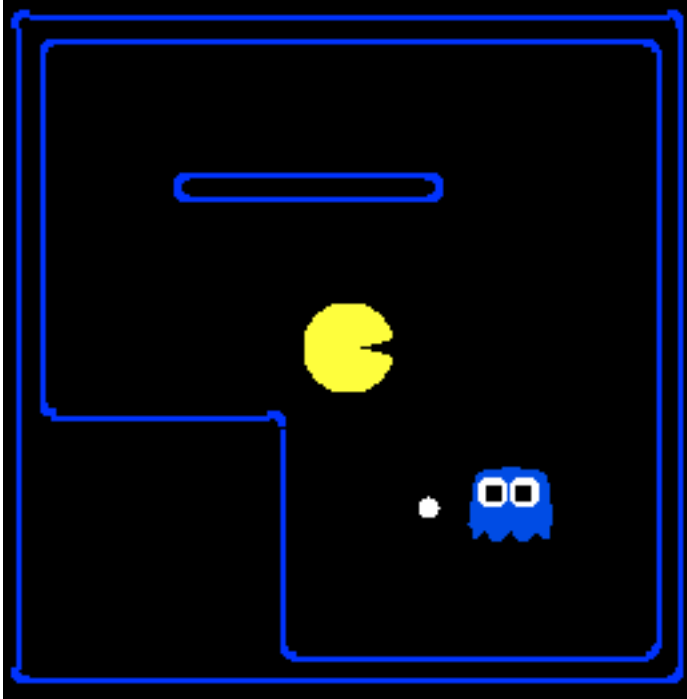


Dangerous Pessimism

Assuming the worst case when it's not likely



Assumptions vs. Reality

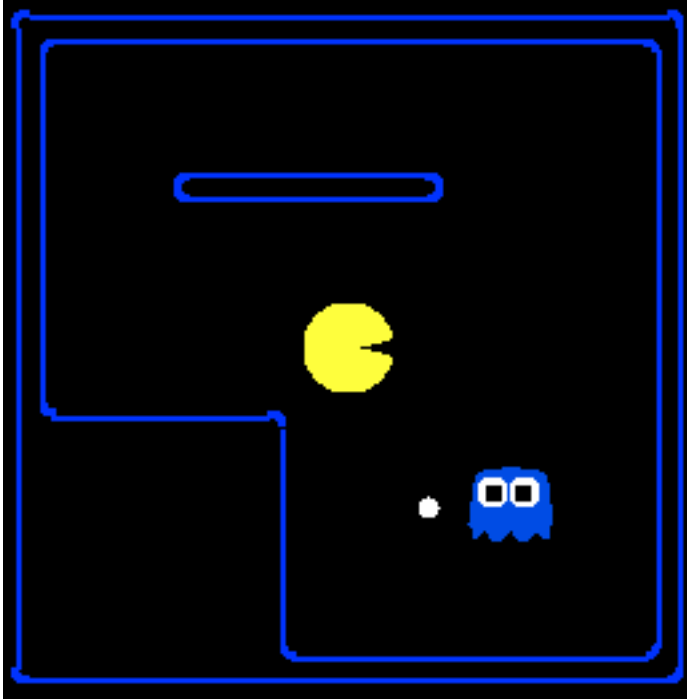


	Adversarial Ghost	Random Ghost
Minimax Pacman		
Expectimax Pacman		

Results from playing 5 games

Pacman used depth 4 search with an eval function that avoids trouble
Ghost used depth 2 search with an eval function that seeks Pacman

Assumptions vs. Reality



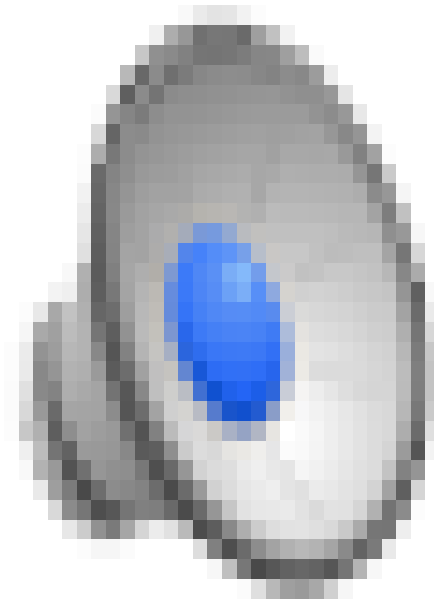
	Adversarial Ghost	Random Ghost
Minimax Pacman	Won 5/5 Avg. Score: 483	Won 5/5 Avg. Score: 453
Expectimax Pacman	Won 1/5 Avg. Score: -303	Won 5/5 Avg. Score: 503

Results from playing 5 games

Pacman used depth 4 search with an eval function that avoids trouble
Ghost used depth 2 search with an eval function that seeks Pacman

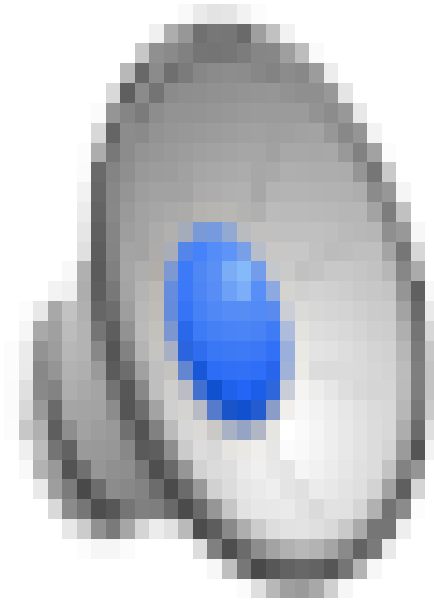
Video of Demo World Assumptions

Random Ghost – Expectimax Pacman



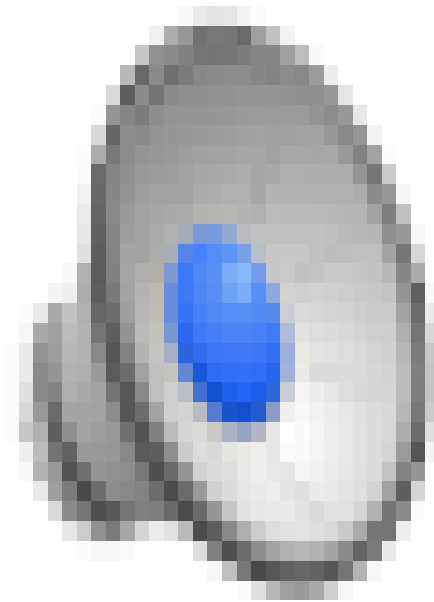
Video of Demo World Assumptions

Adversarial Ghost – Minimax Pacman



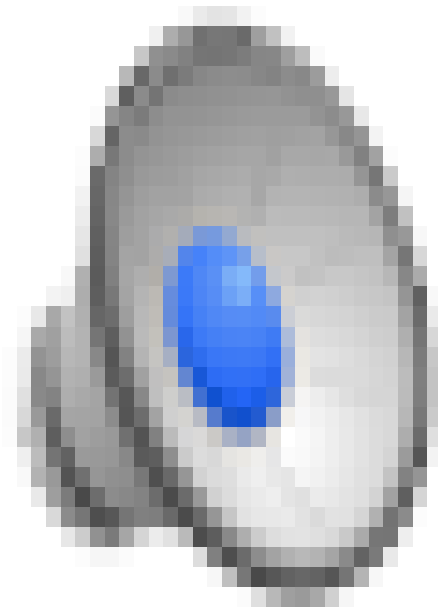
Video of Demo World Assumptions

Adversarial Ghost – Expectimax Pacman

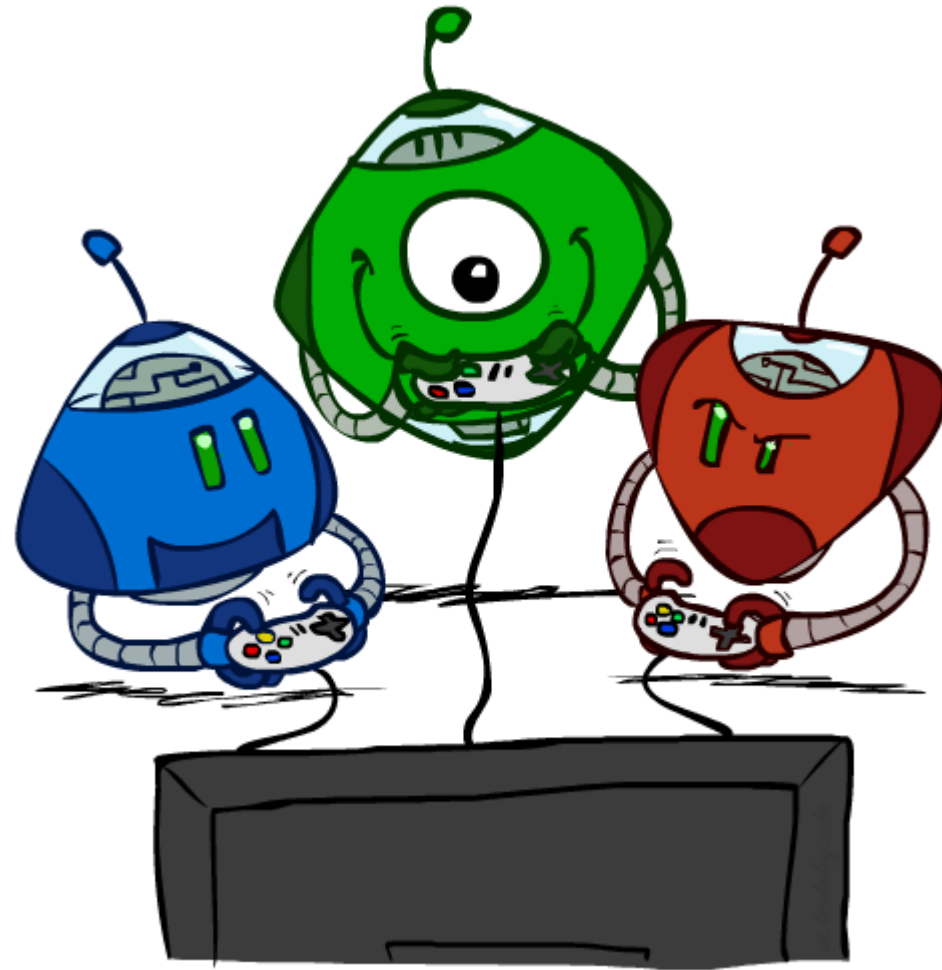


Video of Demo World Assumptions

Random Ghost – Minimax Pacman

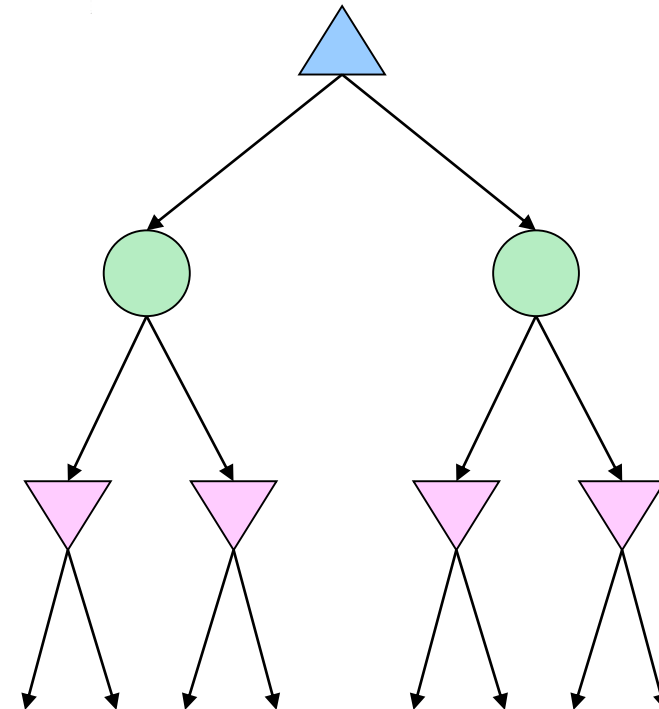
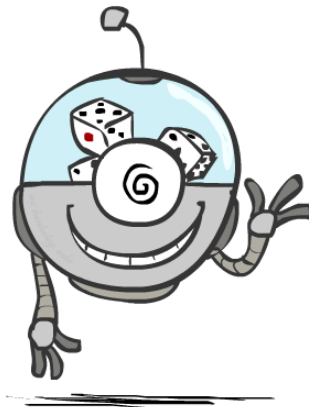


Other Game Types



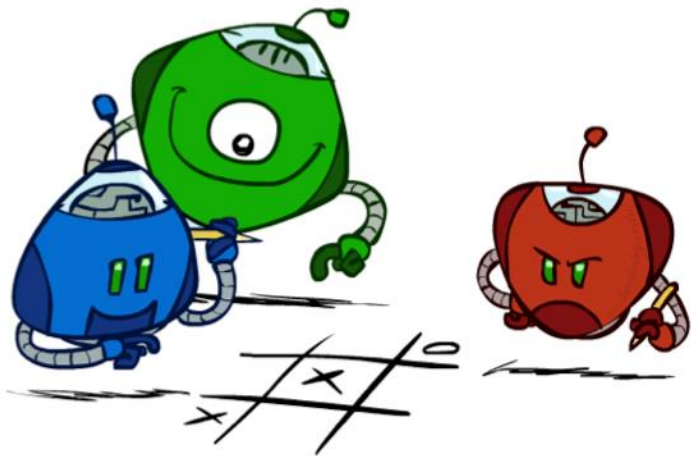
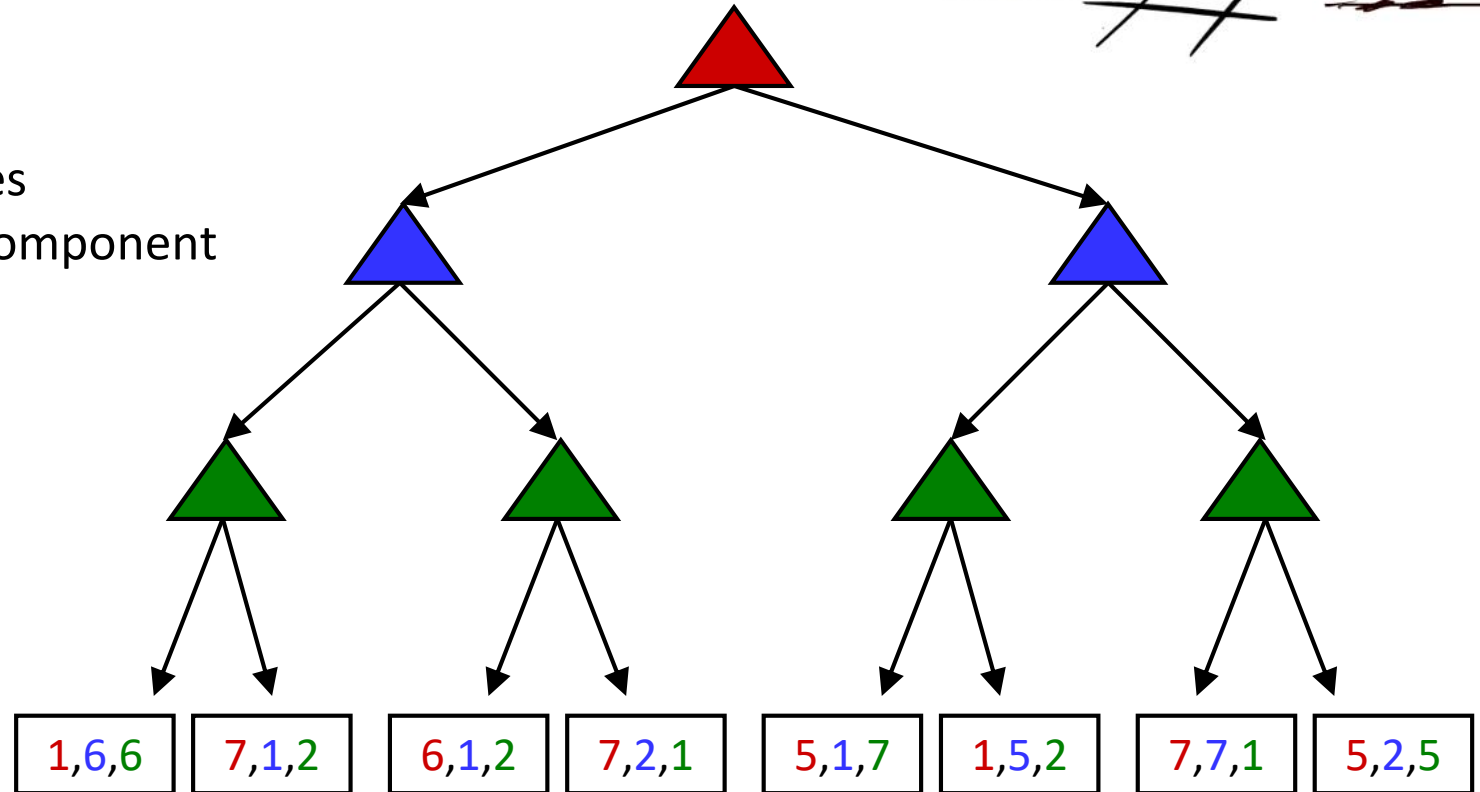
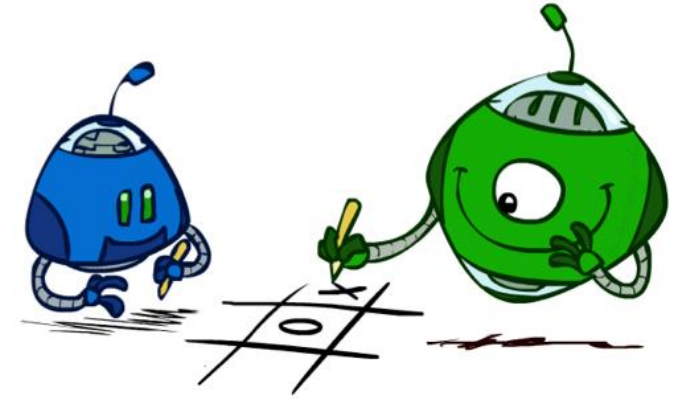
Mixed Layer Types

- Backgammon
- Expectiminimax
 - Environment is an extra “random agent” player that moves after each min/max agent
 - Each node computes the appropriate combination of its children



Multi-Agent Utilities

- What if the game is not zero-sum, or has multiple players?
- Generalization of minimax:
 - Terminals have utility tuples
 - Node values are also utility tuples
 - Each player maximizes its own component
 - Can give rise to cooperation and competition dynamically...



Summary

- Adversarial Games
- Adversarial Search
 - Minimax
- Resource Limits
 - Depth-limited search, limiting branching factor
- Game Tree Pruning (alpha-beta pruning)
- Uncertain Outcomes
 - Expectimax
- Other Game Types

