

Lecture 5: Search with Other Agents

Shuai Li

John Hopcroft Center, Shanghai Jiao Tong University

<https://shuaili8.github.io>

<https://shuaili8.github.io/Teaching/CS3317/index.html>

Part of slide credits: CMU AI & <http://ai.berkeley.edu>

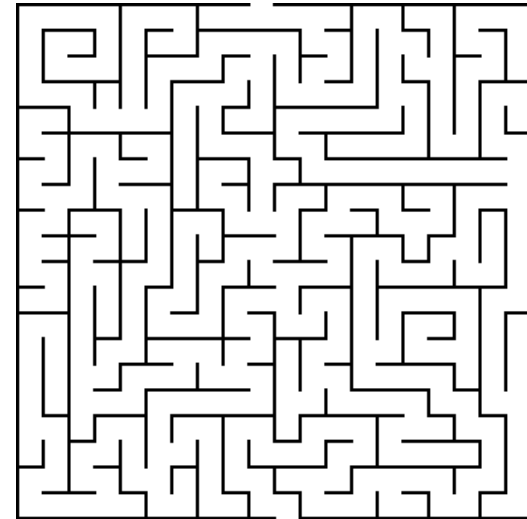
Warm Up

- How is Tic Tac Toe different from maze search?

X	O	X
	O	X
	O	

X	O	X
O	O	X
X	X	O

X	O	X
	X	
X	O	O



Warm Up 2

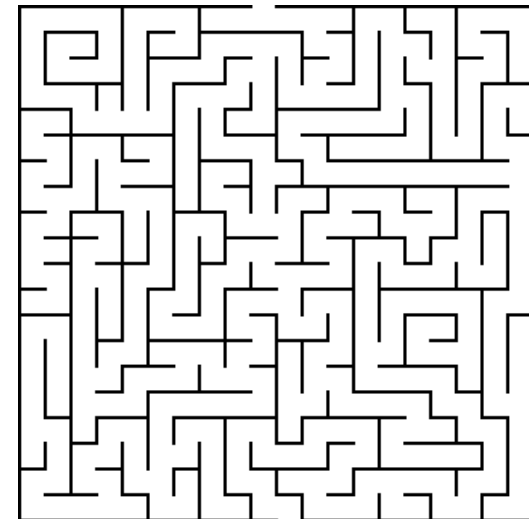
- How is Tic Tac Toe different from maze search?

X	O	X
	O	X
	O	

X	O	X
O	O	X
X	X	O

X	O	X
	X	
X	O	O

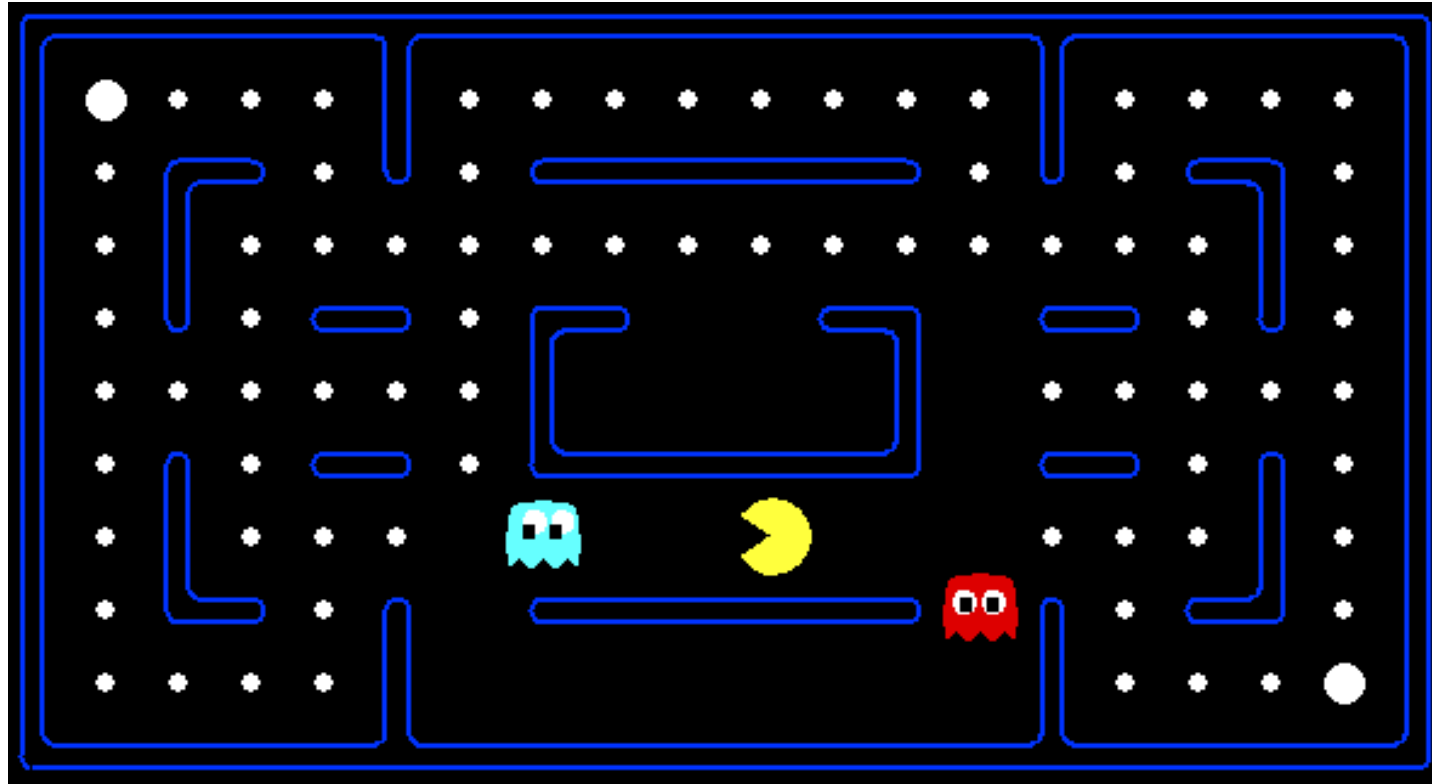
Multi-Agent, Adversarial, Zero Sum



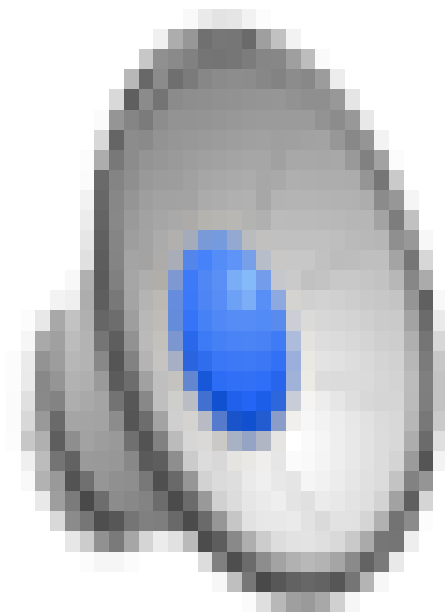
Single Agent

Game Type

Behavior from Computation



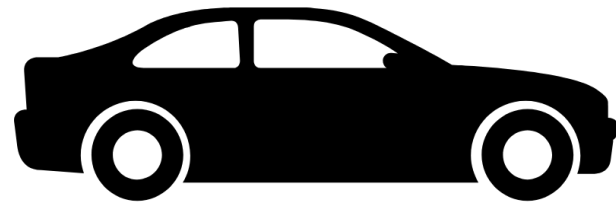
Video of Demo Mystery Pacman



Agents Getting Along with Other Agents

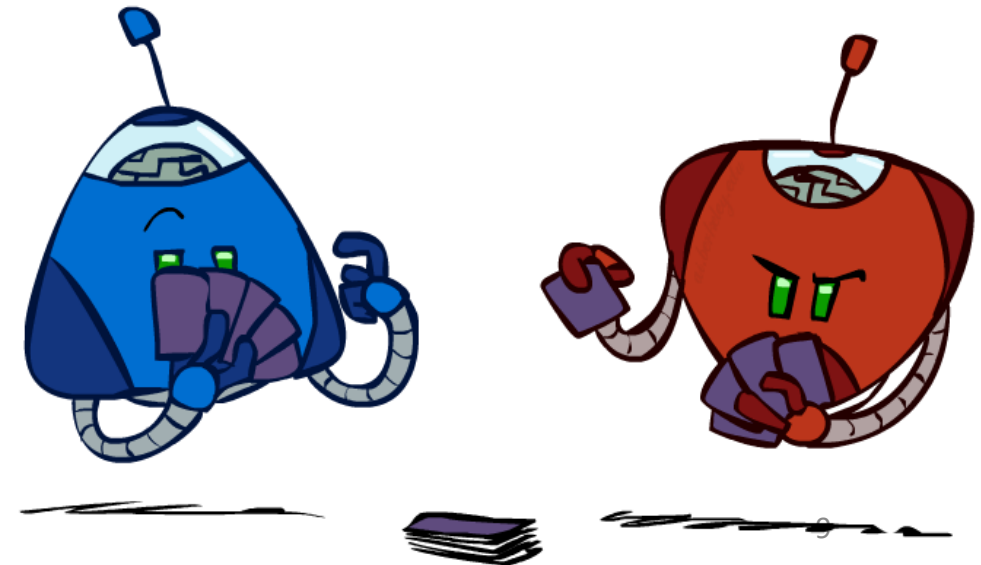
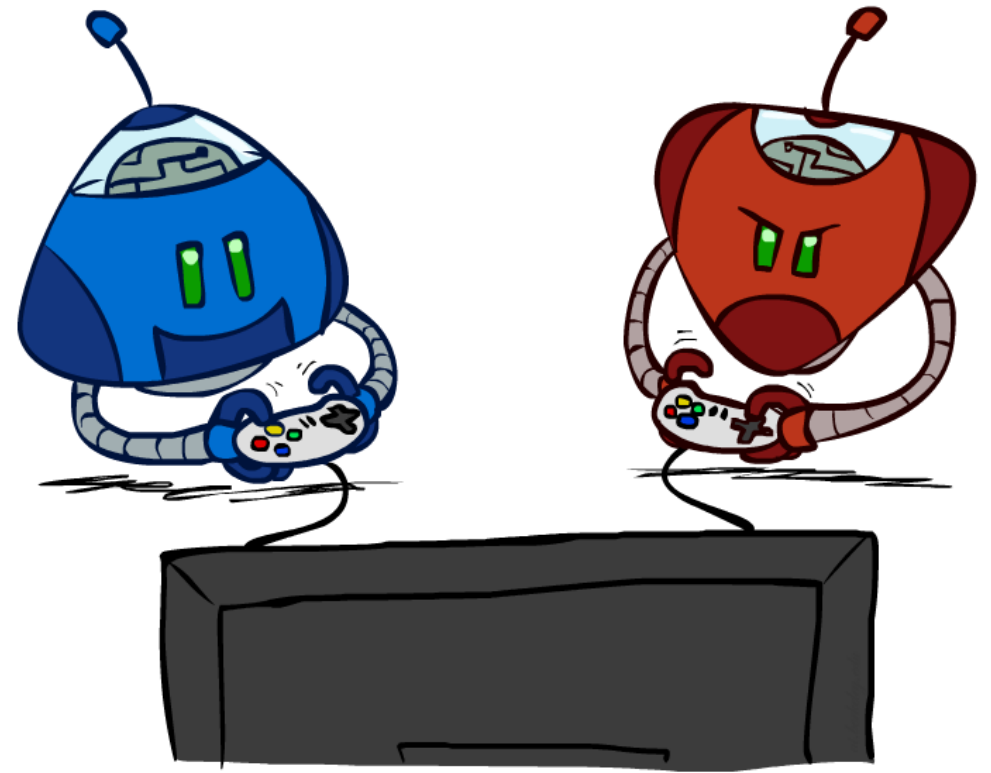


Agents Getting Along with Humans

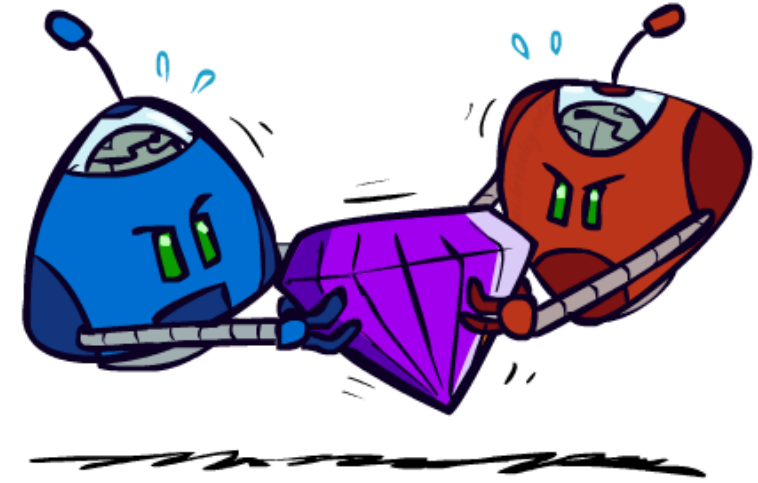


Types of Games

- Many different kinds of games!
- Axes:
 - Deterministic or stochastic?
 - One, two, or more players?
 - Zero sum?
 - Perfect information (can you see the state)?
- Want algorithms for calculating a **contingent plan** (a.k.a. **strategy** or **policy**) which recommends a move for every possible eventuality



Types of Games 2



- General Games

- Agents have **independent** utilities (values on outcomes)
- Cooperation, indifference, competition, shifting alliances, and more are all possible
 - We don't make AI to act in isolation, it should
 - a) work around people and b) help people
 - That means that every AI agent needs to solve a game

- 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

History of Game AI

1956 checkers

1992 backgammon

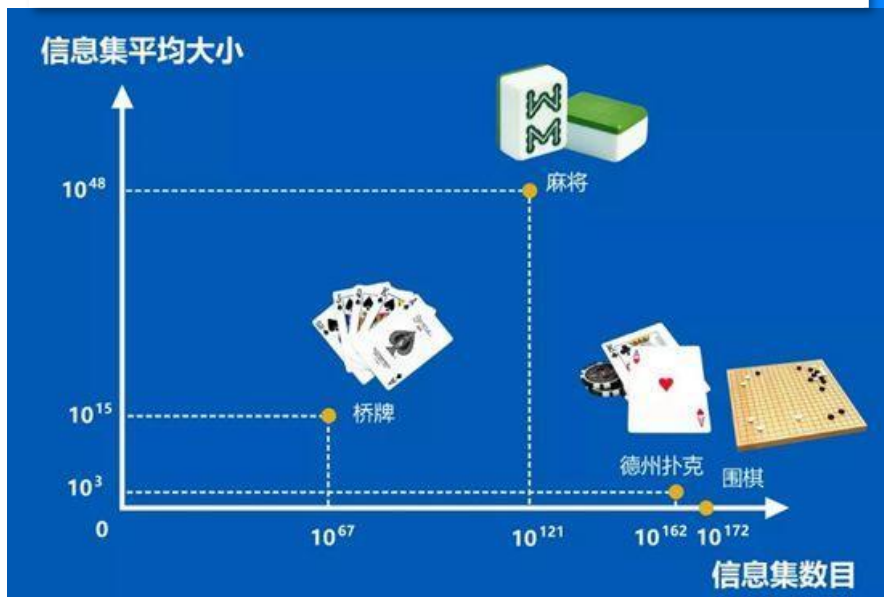
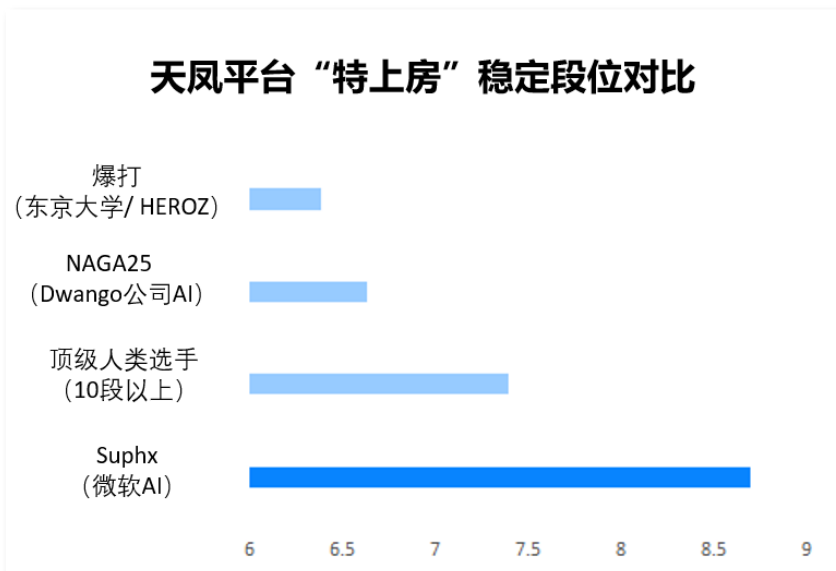
1994 checkers

1997 chess

2016 Go

2017 Texas hold'em

2019 Majiang



1956年

Arthur Samuel开发的AI程序首先应用于国际跳棋

1992年

Hans Berliner开发的AI程序在双陆棋中取得突破

1994年

Jonathan Schaeffer团队开发的AI程序战胜国际跳棋世界冠军

Chinook



许峰雄团队开发的AI程序击败国际象棋世界冠军

1997年

DeepMind团队开发的AI程序战胜围棋世界冠军

2016年

Deep Blue



2017年

卡耐基梅隆大学、Facebook AI以及阿尔伯特大学开发的AI程序在德州扑克取得突破

下一个里程碑在哪里?



AlphaGo



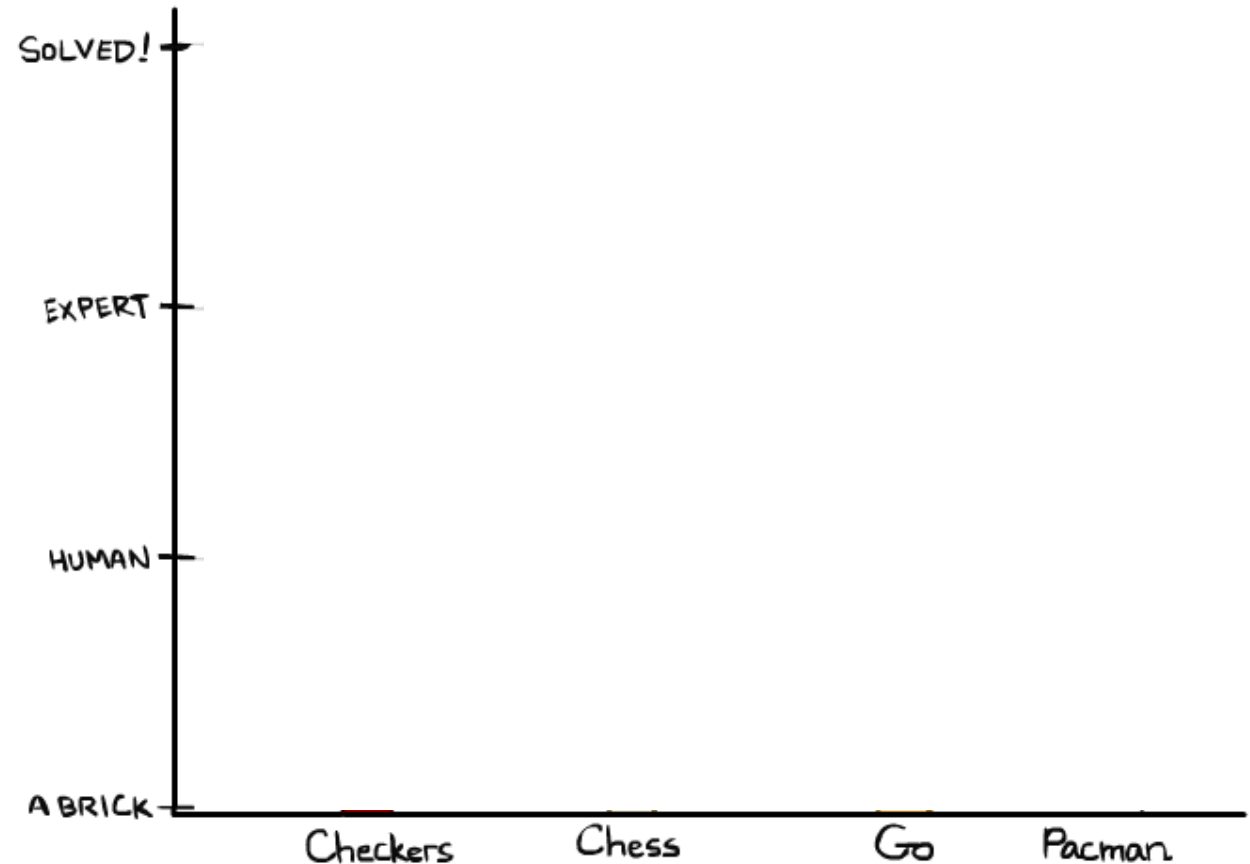
Libratus

Pluribus

DeepStack

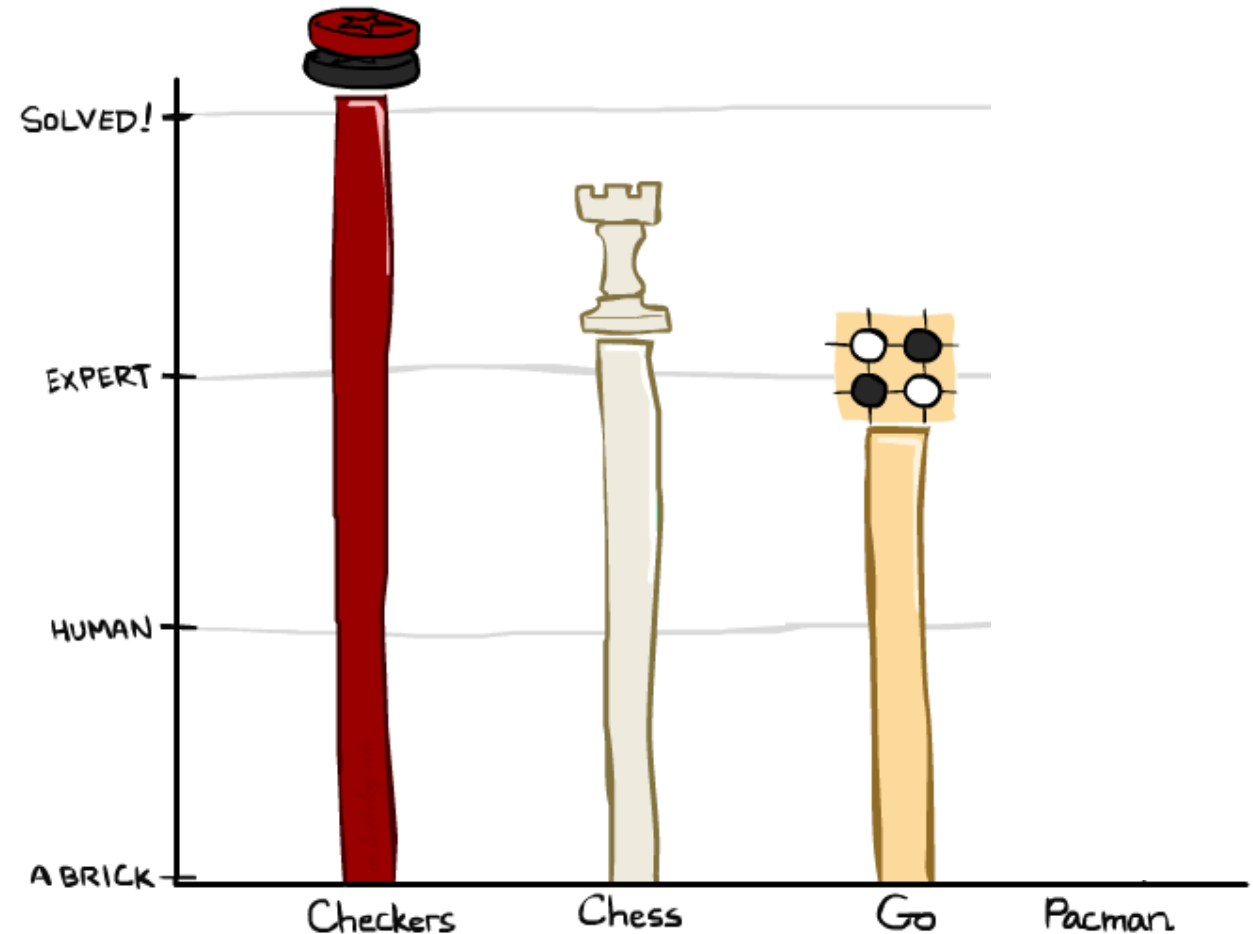
Zero-Sum Games

- **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: Alpha GO defeats human champion. Uses Monte Carlo Tree Search, learned evaluation function.

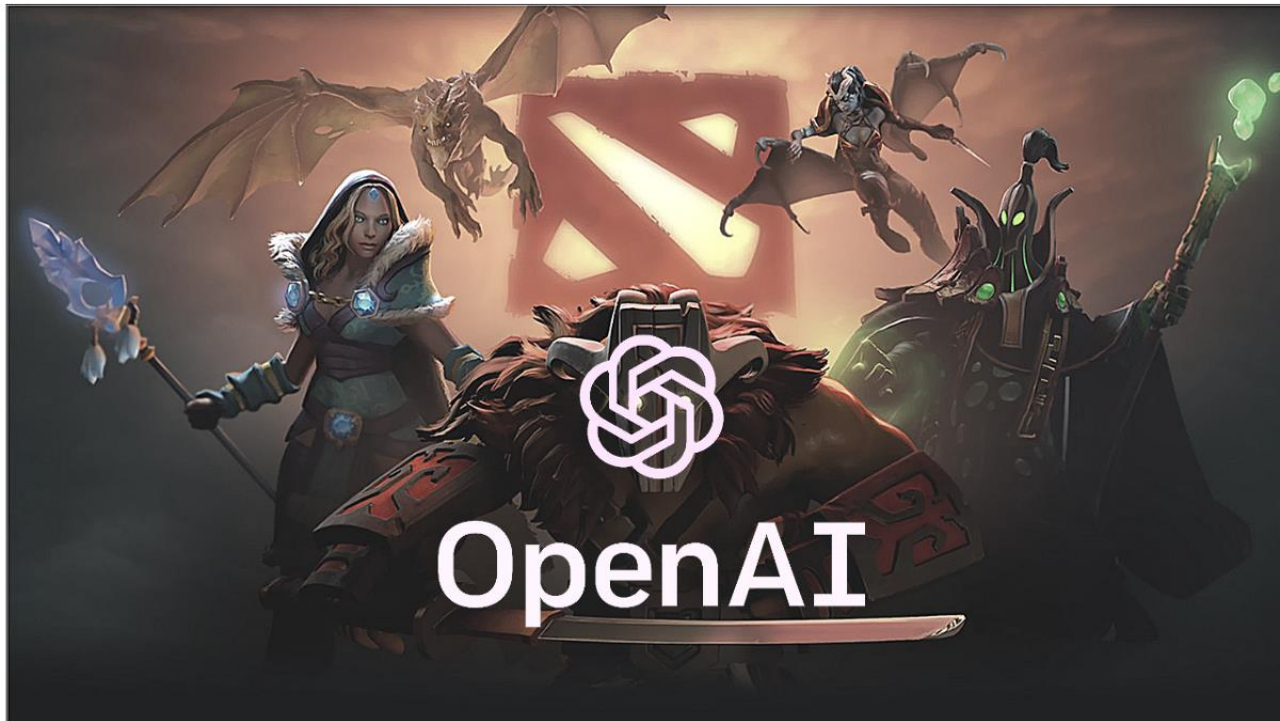


Zero-Sum Games 2

- **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: Alpha GO defeats human champion. Uses Monte Carlo Tree Search, learned evaluation function.
- **Pacman**



Game playing – state of the art

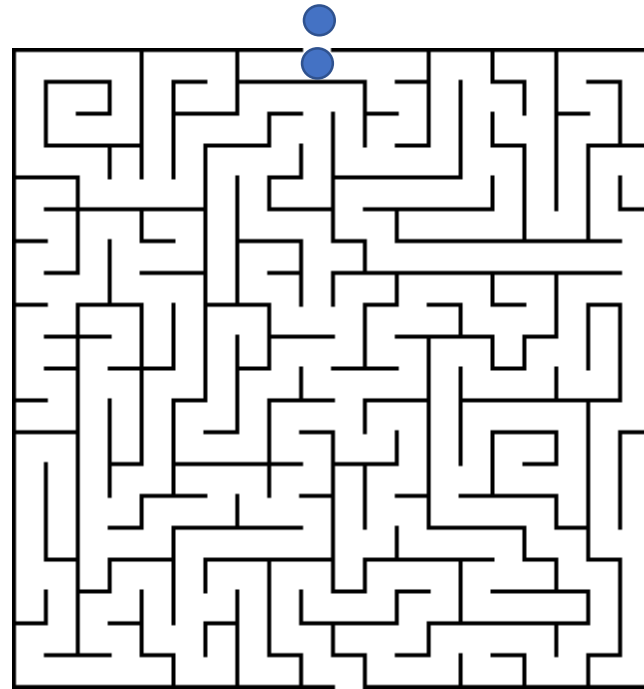


Types of Games 3



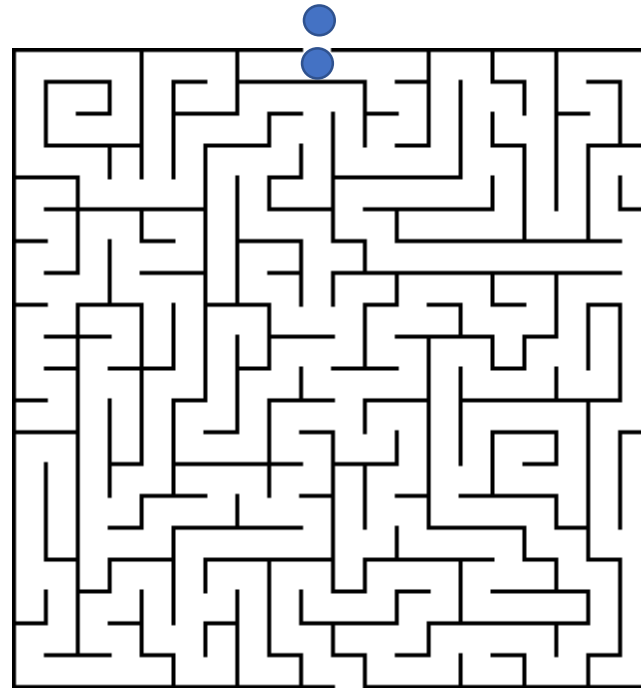
- Common payoff games
 - Discussion: Use a technique you've learned so far to solve one!

How could we model multi-agent collaborative problems?

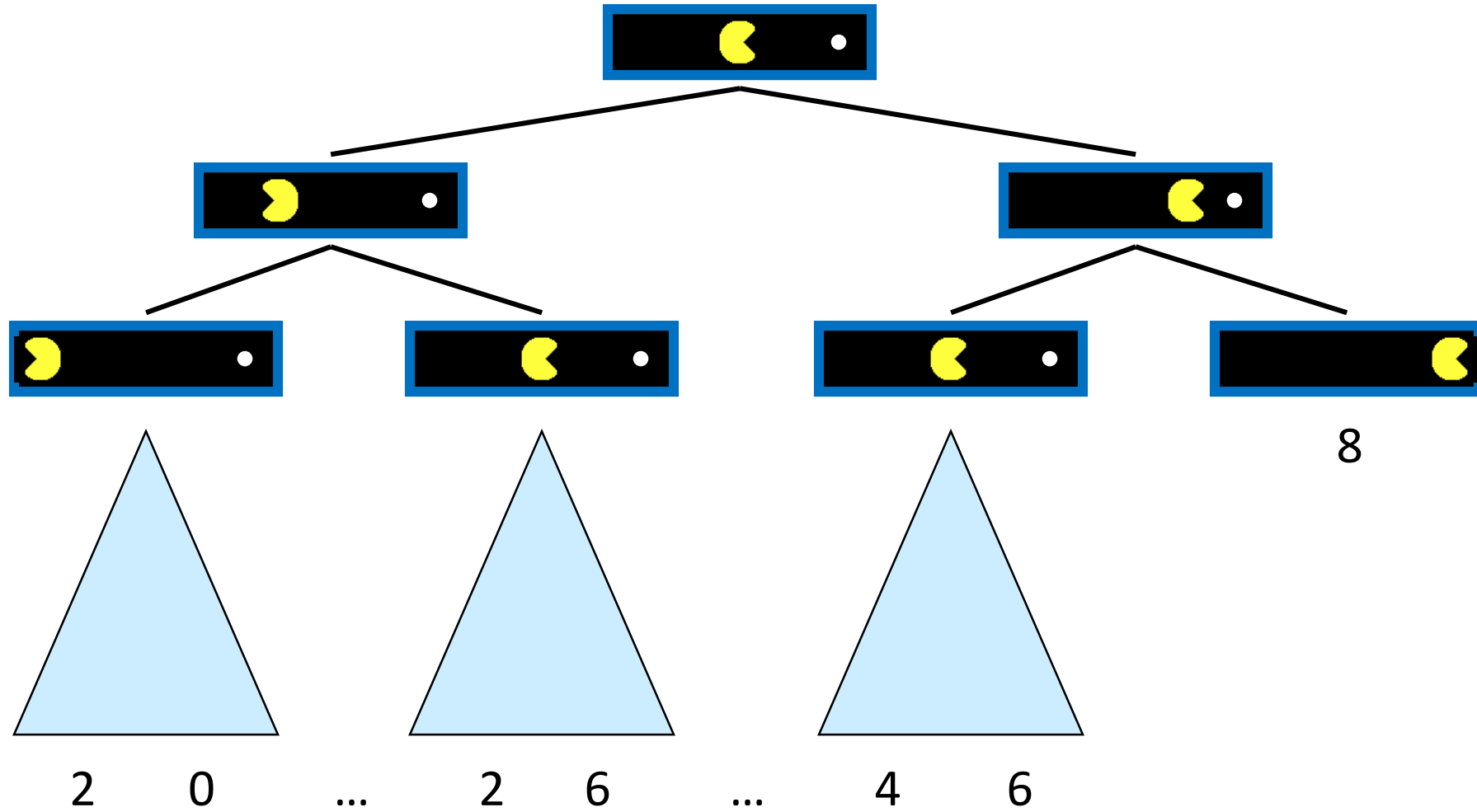


How could we model multi-agent collaborative problems? 2

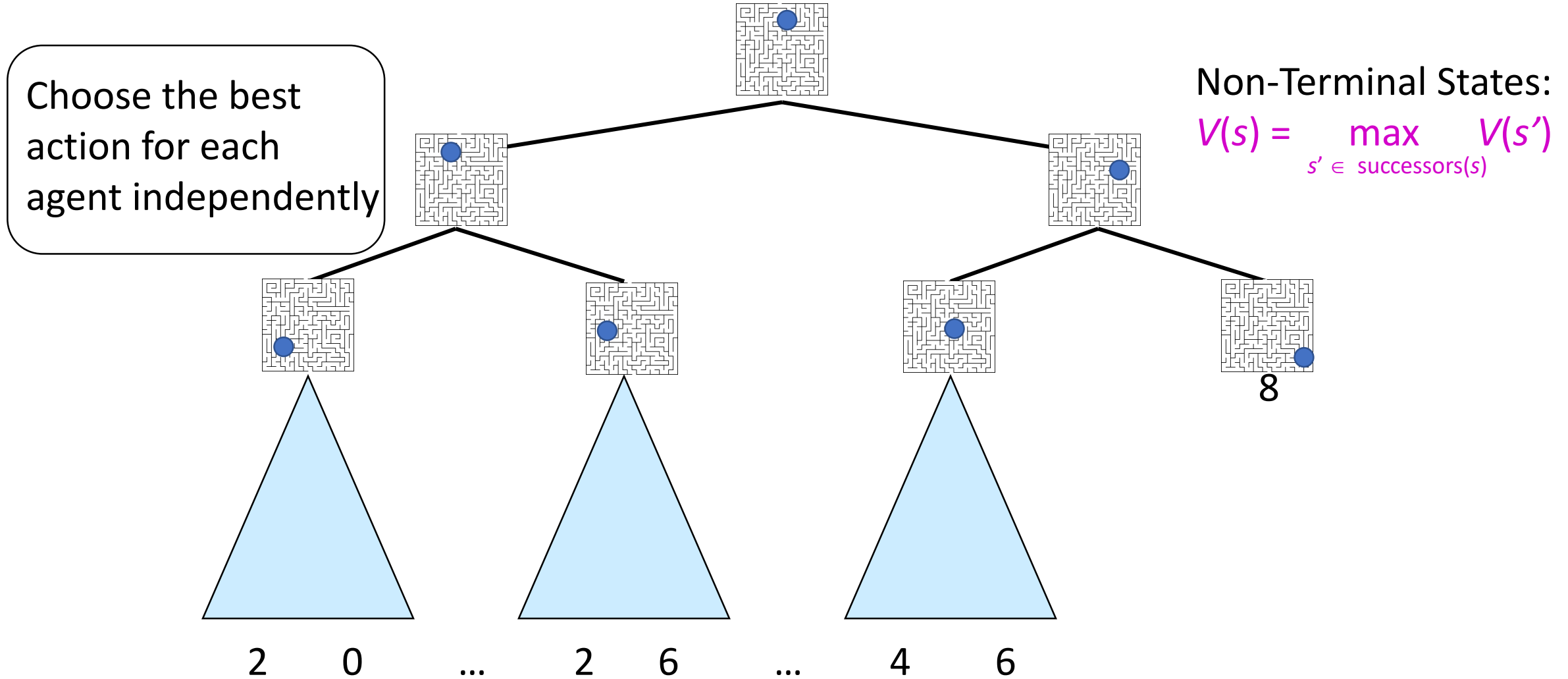
- Simplest idea: each agent plans their own actions separately from others



Single-Agent Trees



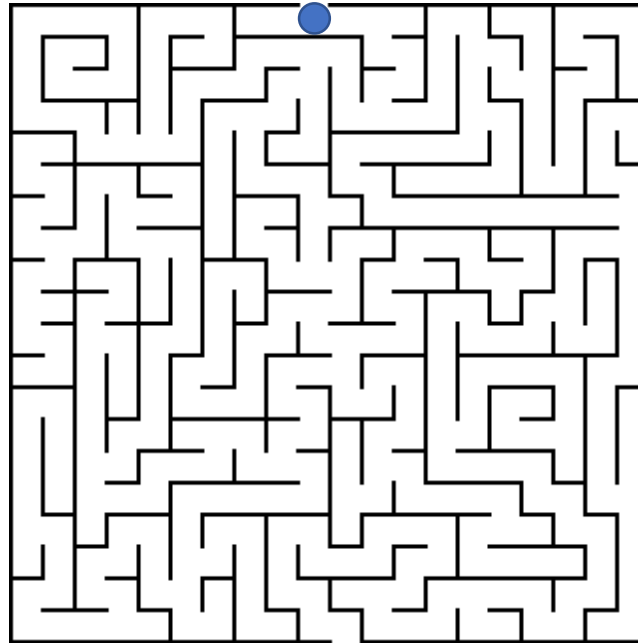
Many Single-Agent Trees



Idea 2: Joint State/Action Spaces

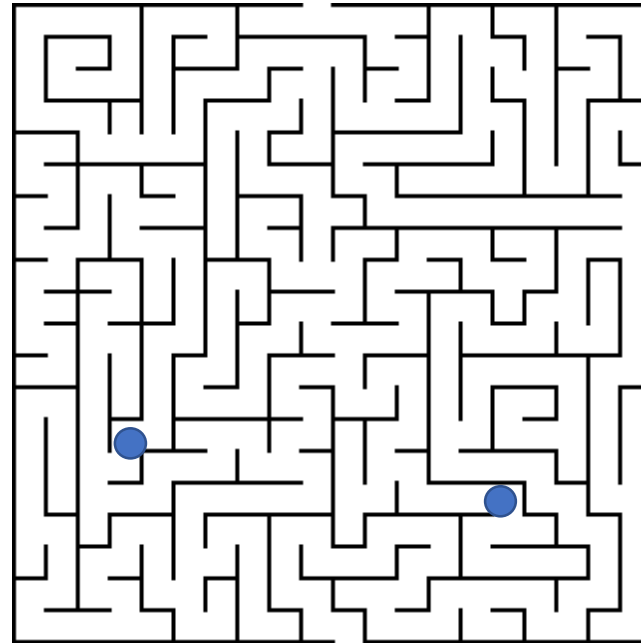
- Combine the states and actions of the N agents

$$s_0 = (s_0^A, s_0^B)$$



Idea 2: Joint State/Action Spaces 2

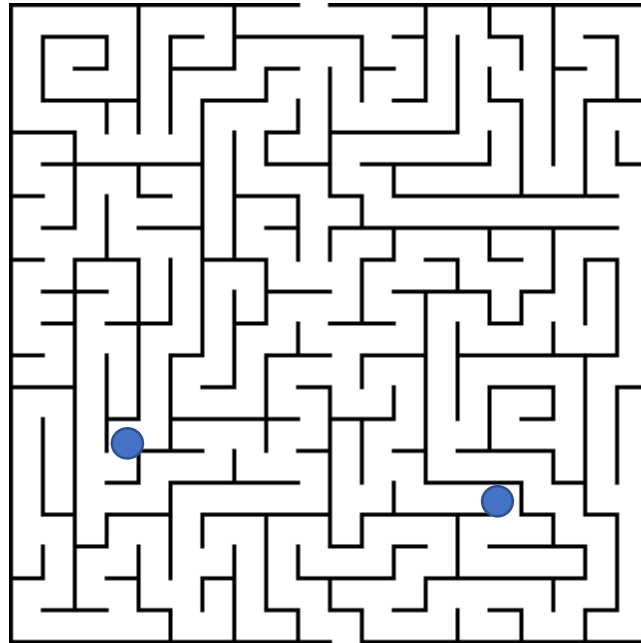
$$S_K = (S_K^A, S_K^B)$$



Idea 2: Joint State/Action Spaces 3

- Search looks through all combinations of all agents' states and actions
- Think of one brain controlling many agents

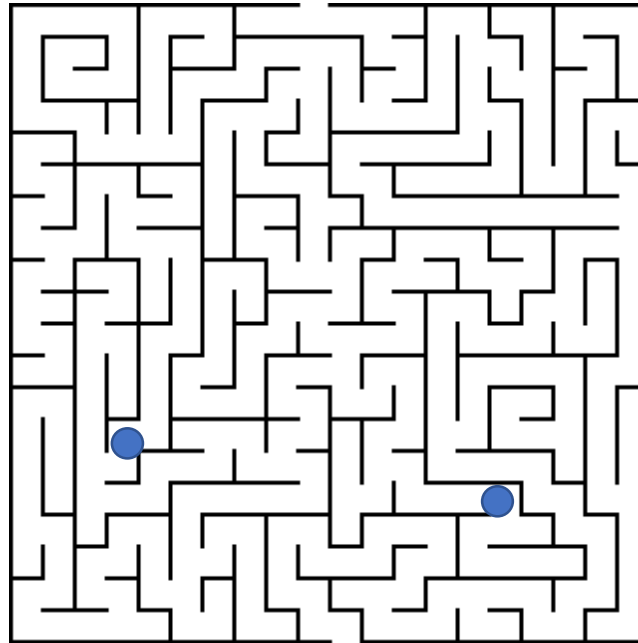
$$S_K = (S_K^A, S_K^B)$$



Idea 2: Joint State/Action Spaces 4

- Search looks through all combinations of all agents' states and actions
- Think of one brain controlling many agents

- What is the size of the state space?
- What is the size of the action space?
- What is the size of the search tree?



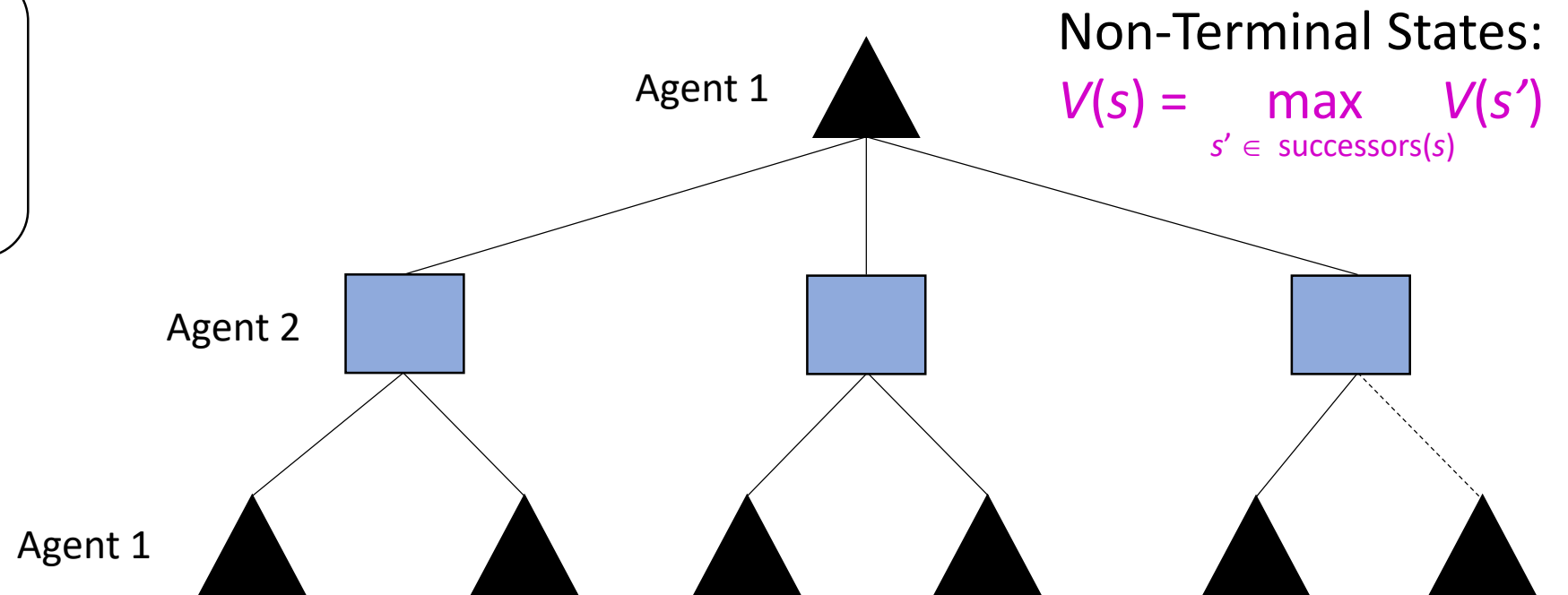
Idea 3: Centralized Decision Making

- Each agent proposes their actions and computer confirms the joint plan
 - Either accept or arrange for conflicts of the current submitted plans
- Example: Autonomous vehicles driving through intersections

Idea 4: Alternate Searching One Agent at a Time

- Search one agent's actions from a state, search the next agent's actions from those resulting states, etc...

Choose the best cascading combination of actions



Idea 4: Alternate Searching One Agent at a Time

- Search one agent's actions from a state, search the next agent's actions from those resulting states, etc...

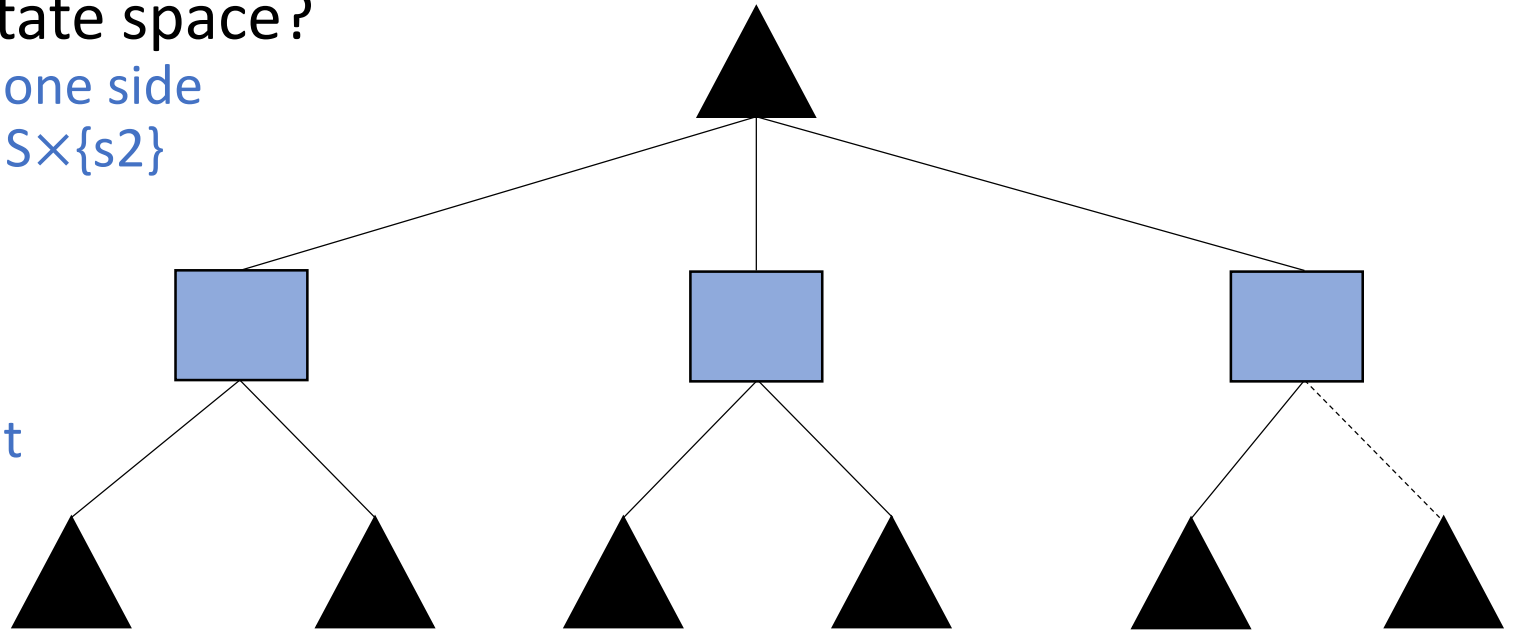
- What is the size of the state space?

- Try to expand (s_1, s_2) on one side
- Only consider space like $S \times \{s_2\}$

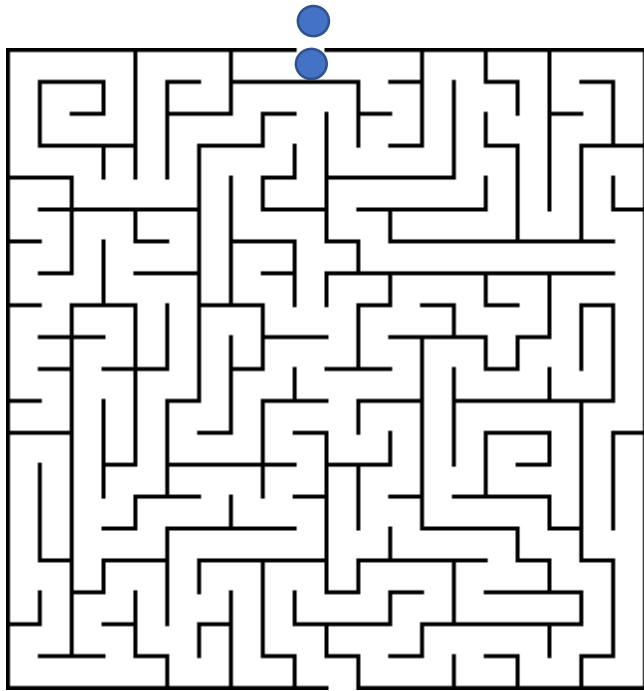
- What is the size of the action space?

- Take action for one agent

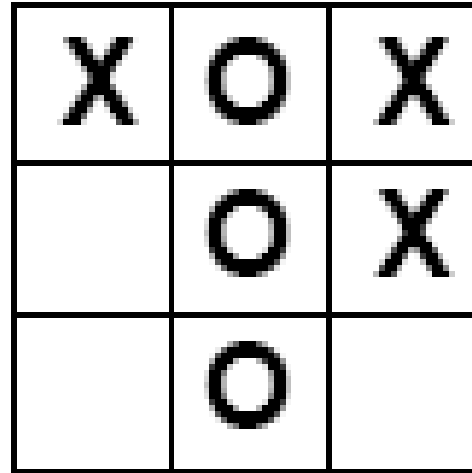
- What is the size of the search tree?



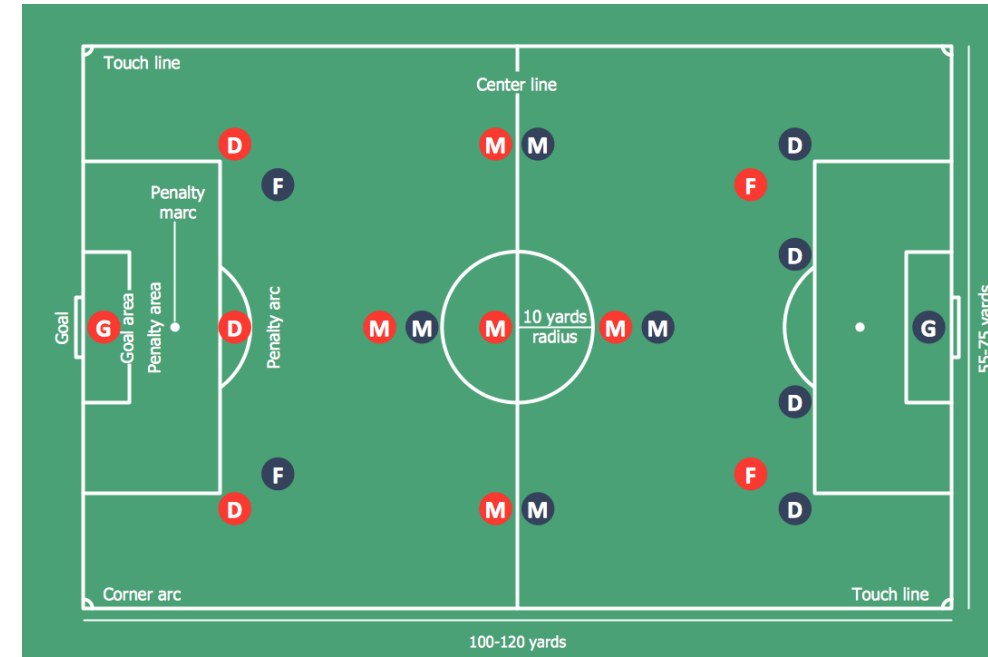
Multi-Agent Applications



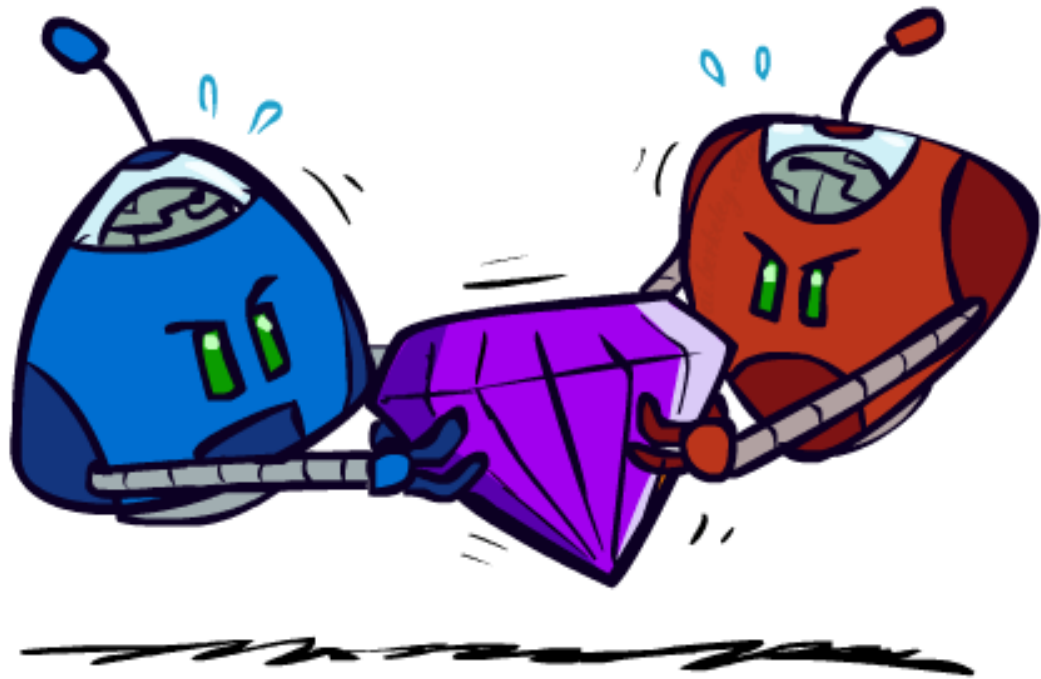
Collaborative Maze Solving



Adversarial

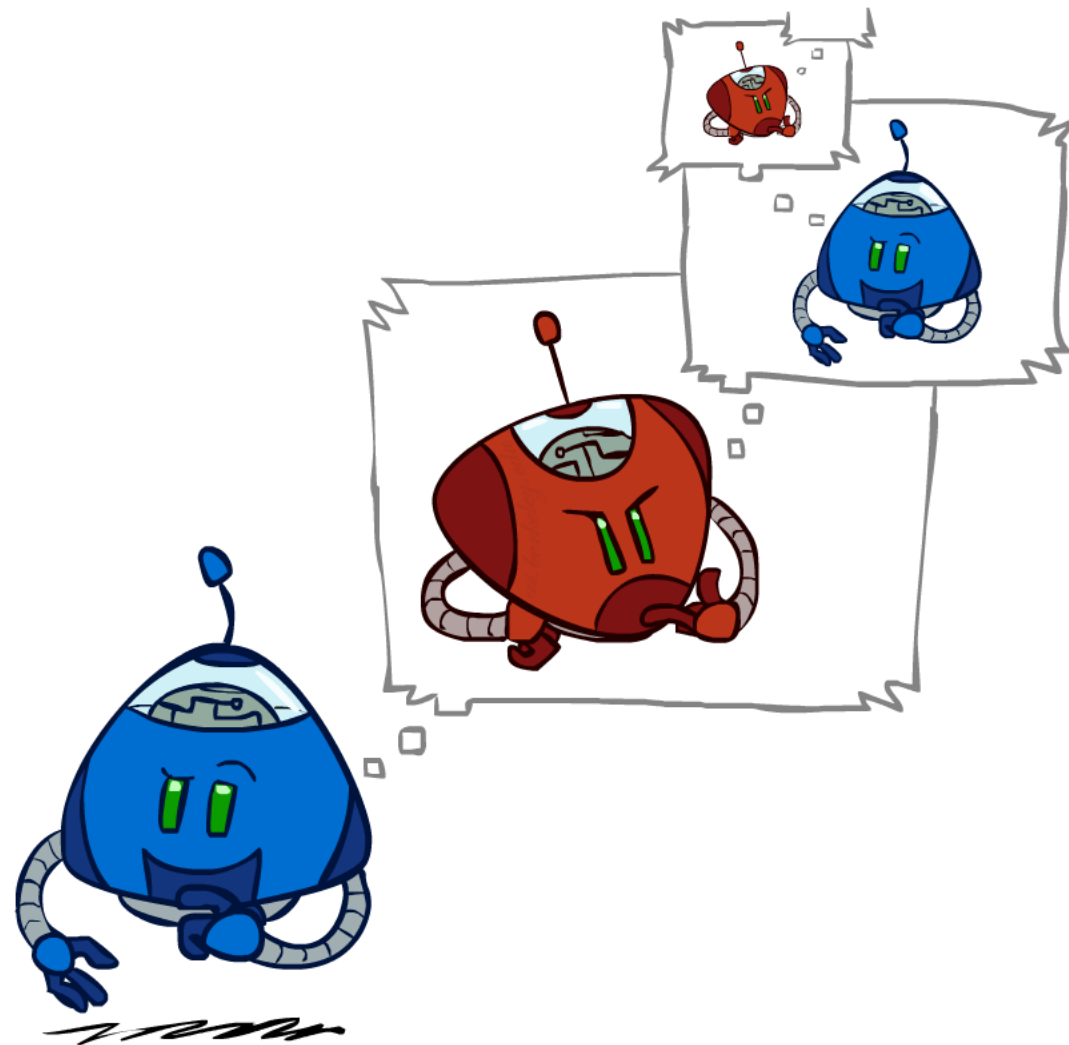


Team: Collaborative
Competition: Adversarial



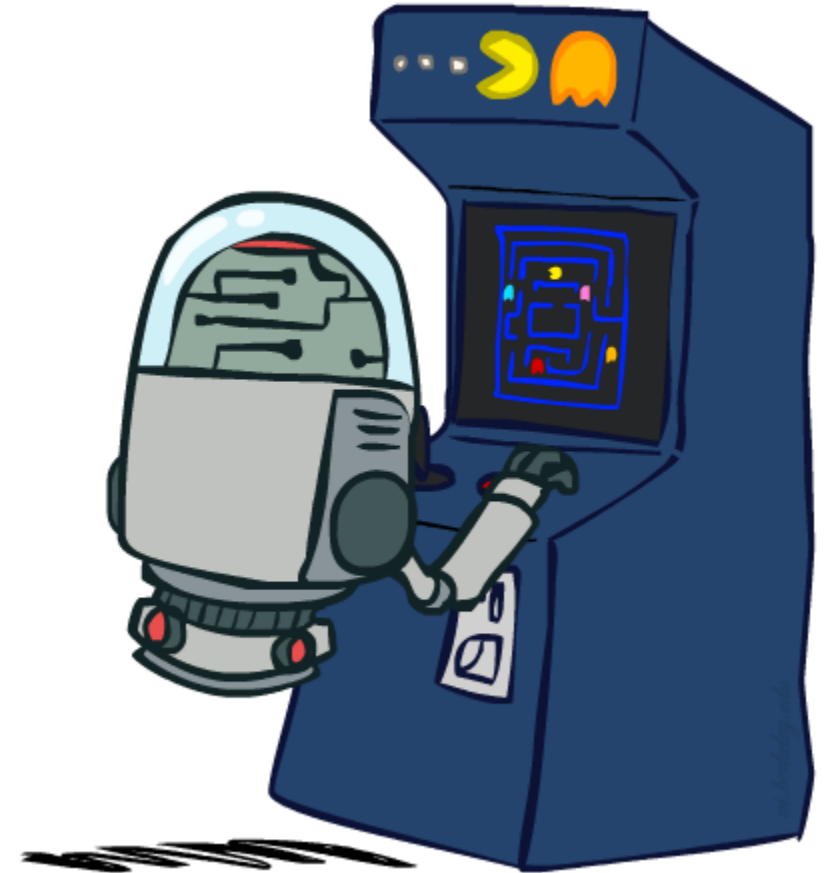
Adversarial Search

Cost \rightarrow Utility!



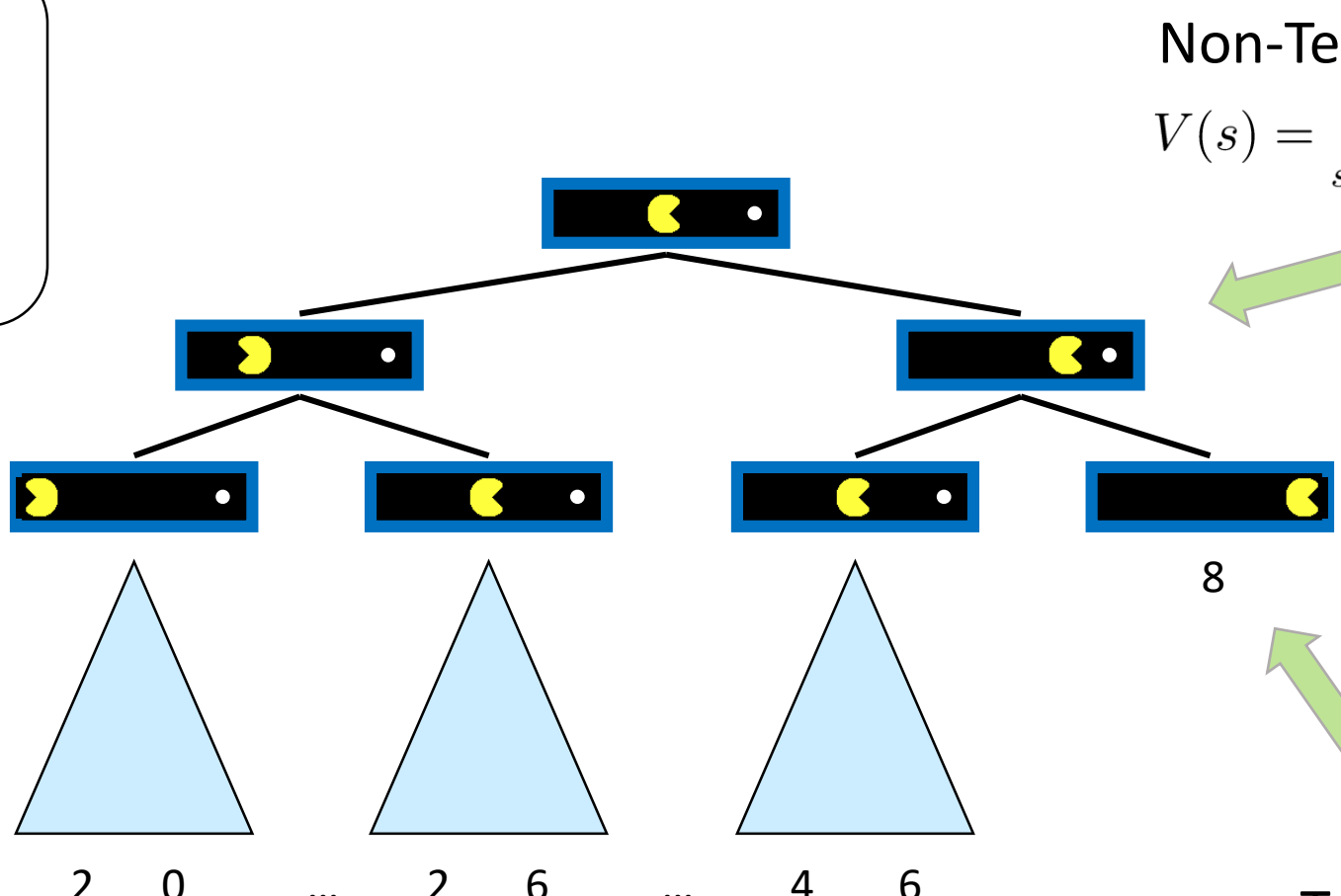
“Standard” Games

- Standard games are **deterministic**, observable, two-player, turn-taking, zero-sum
- Game formulation:
 - 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**: $S \rightarrow A$



Single-Agent Trees: Value of a State

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



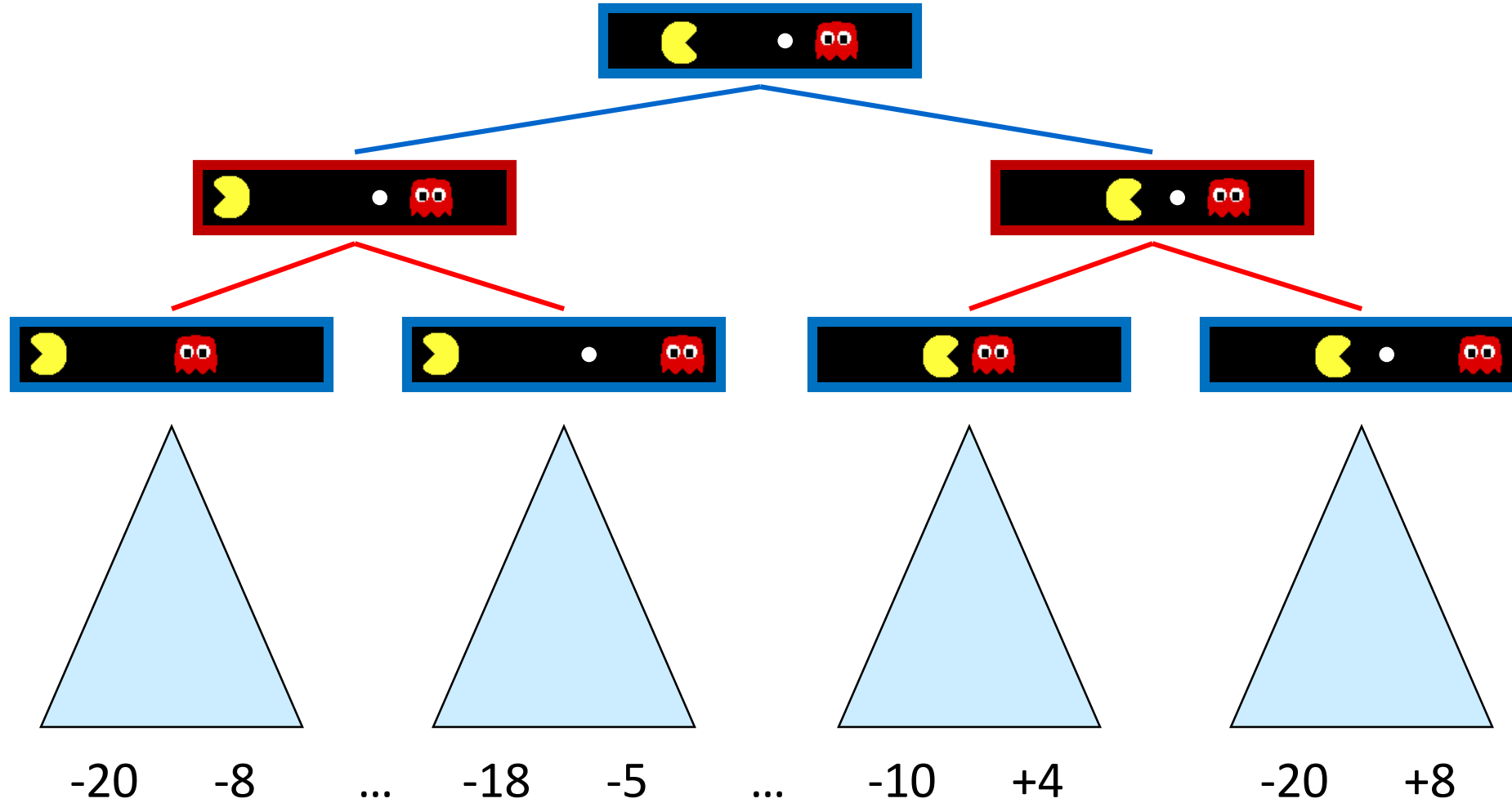
Non-Terminal States:

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

Terminal States:

$$V(s) = \text{known } s_0$$

Adversarial Game Trees



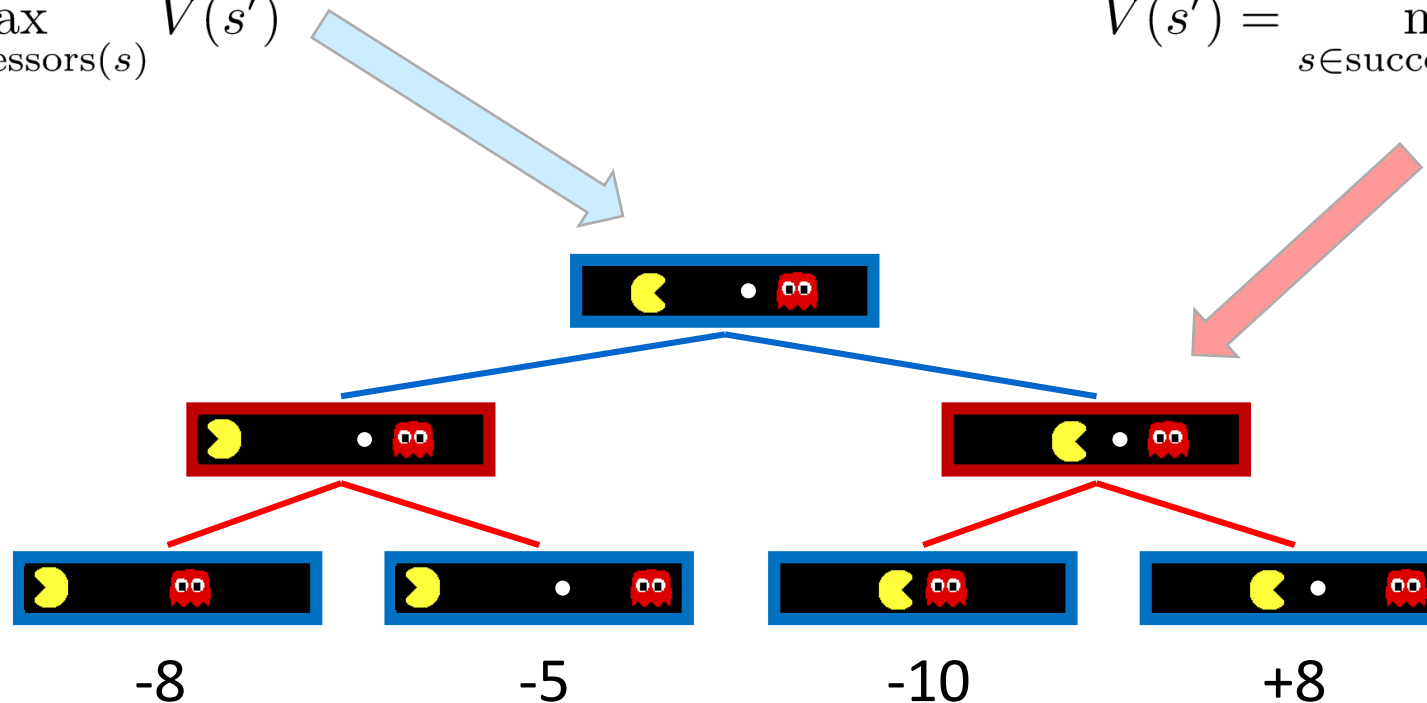
Adversarial Game Trees: Minimax Values

States Under Agent's Control:

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

States Under Opponent's Control:

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



Terminal States:

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

Example: Tic-Tac-Toe Game Tree

Instead of taking the max utility at every level, alternate max and min

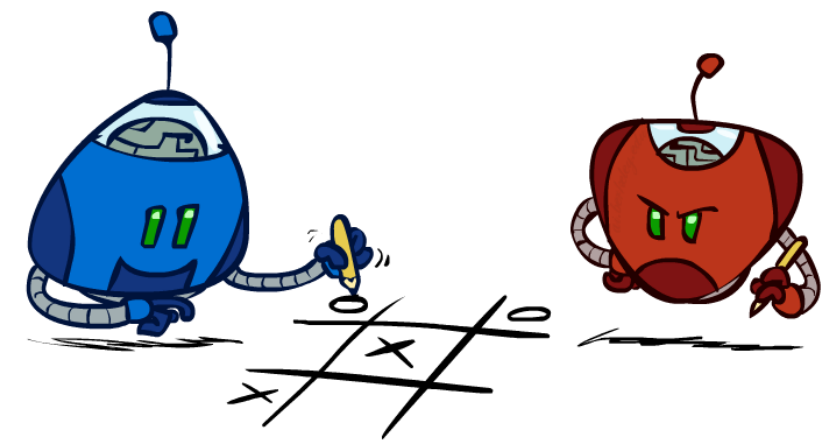
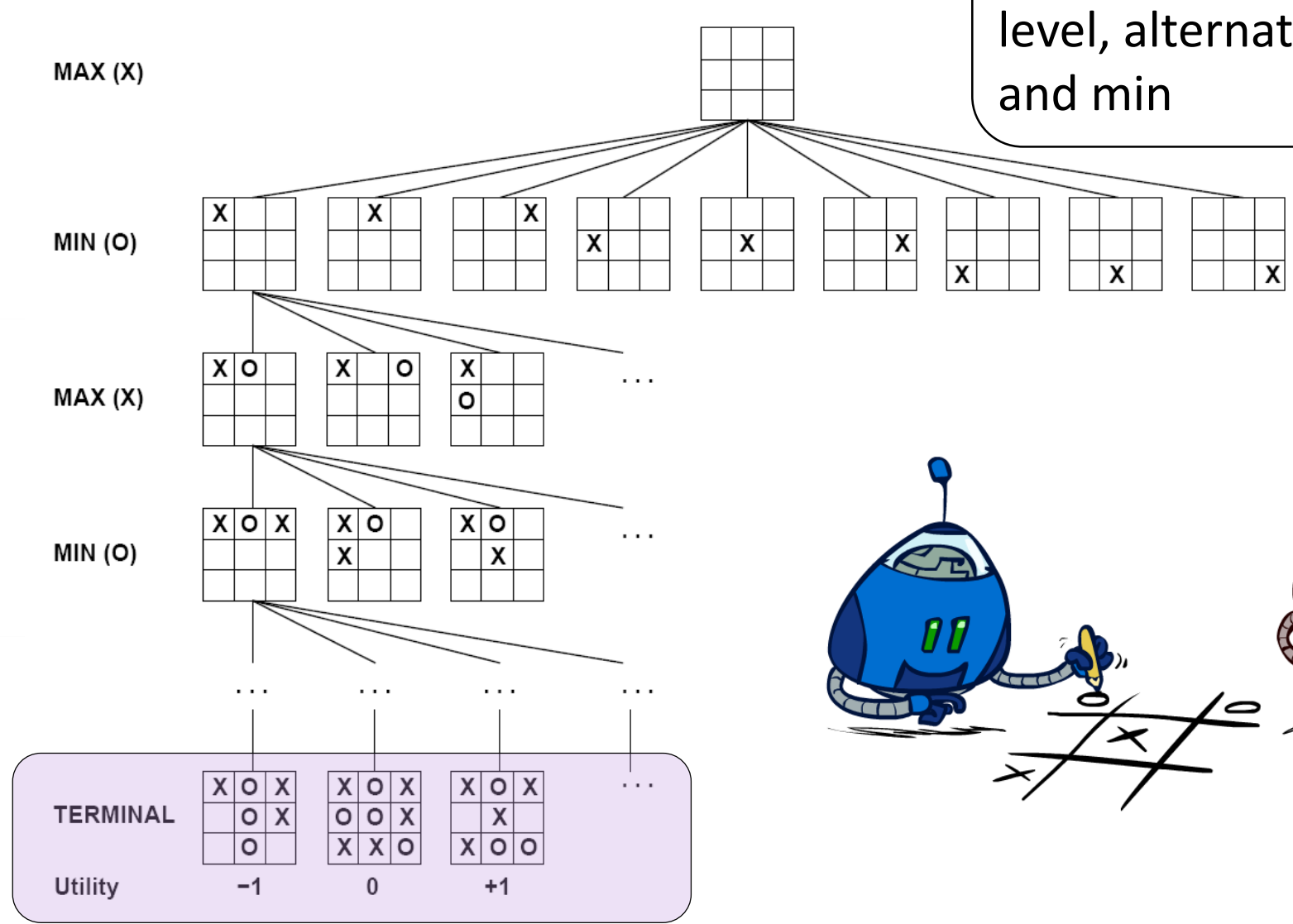
- States



- Actions

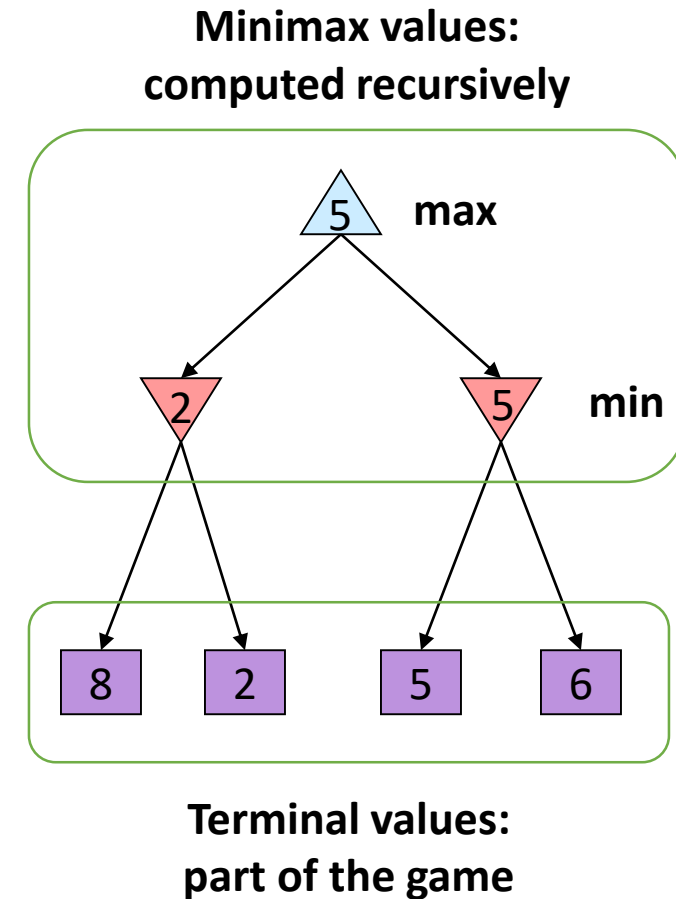


- Values



Minimax Search

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



Minimax Implementation

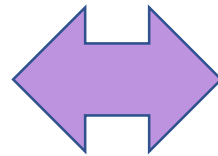
```
def max-value(state):
```

```
    initialize v =  $-\infty$ 
```

```
    for each successor of state:
```

```
        v = max(v, min-value(successor))
```

```
    return v
```



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

```
    initialize v =  $+\infty$ 
```

```
    for each successor of state:
```

```
        v = min(v, max-value(successor))
```

```
    return v
```

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

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

Minimax Implementation (Dispatch)

```
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 $v = +\infty$

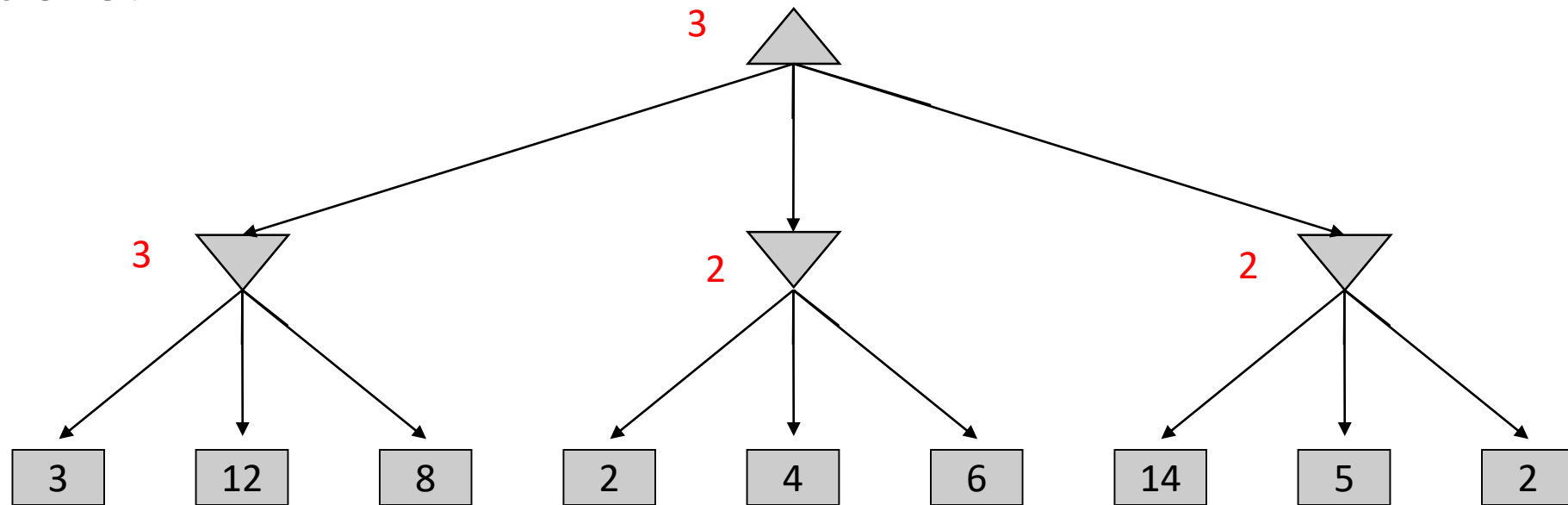
for each successor of state:

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

return v

Example

- Actions?

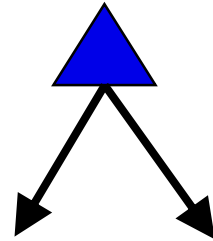


Pseudocode for Single Agent

```
def max_value(state):  
    if state.is_leaf:  
        return state.value  
    # TODO Also handle depth limit  
  
    best_value = -10000000  
  
    for action in state.actions:  
        next_state = state.result(action)  
  
        next_value = max_value(next_state)  
  
        if next_value > best_value:  
            best_value = next_value  
  
    return best_value
```

Pseudocode for Minimax Search

```
def max_value(state):  
    if state.is_leaf:  
        return state.value  
    # TODO Also handle depth limit  
  
    best_value = -10000000  
  
    for action in state.actions:  
        next_state = state.result(action)  
        next_value = min_value(next_state)  
  
        if next_value > best_value:  
            best_value = next_value  
  
    return best_value  
  
def min_value(state):
```

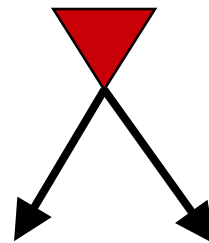


$$V(s) = \max_a V(s'),$$

where $s' = \text{result}(s, a)$

$$\hat{a} = \operatorname{argmax}_a V(s'),$$

where $s' = \text{result}(s, a)$



Pseudocode for Generic Game Tree

```
function minimax_decision( state )  
    return argmaxa in state.actions value( state.result(a) )
```

```
function value( state )  
    if state.is_leaf  
        return state.value  
  
    if state.player is MAX  
        return maxa in state.actions value( state.result(a) )  
  
    if state.player is MIN  
        return mina in state.actions value( state.result(a) )
```


Quiz

- Minimax search belongs to which class?

A) BFS

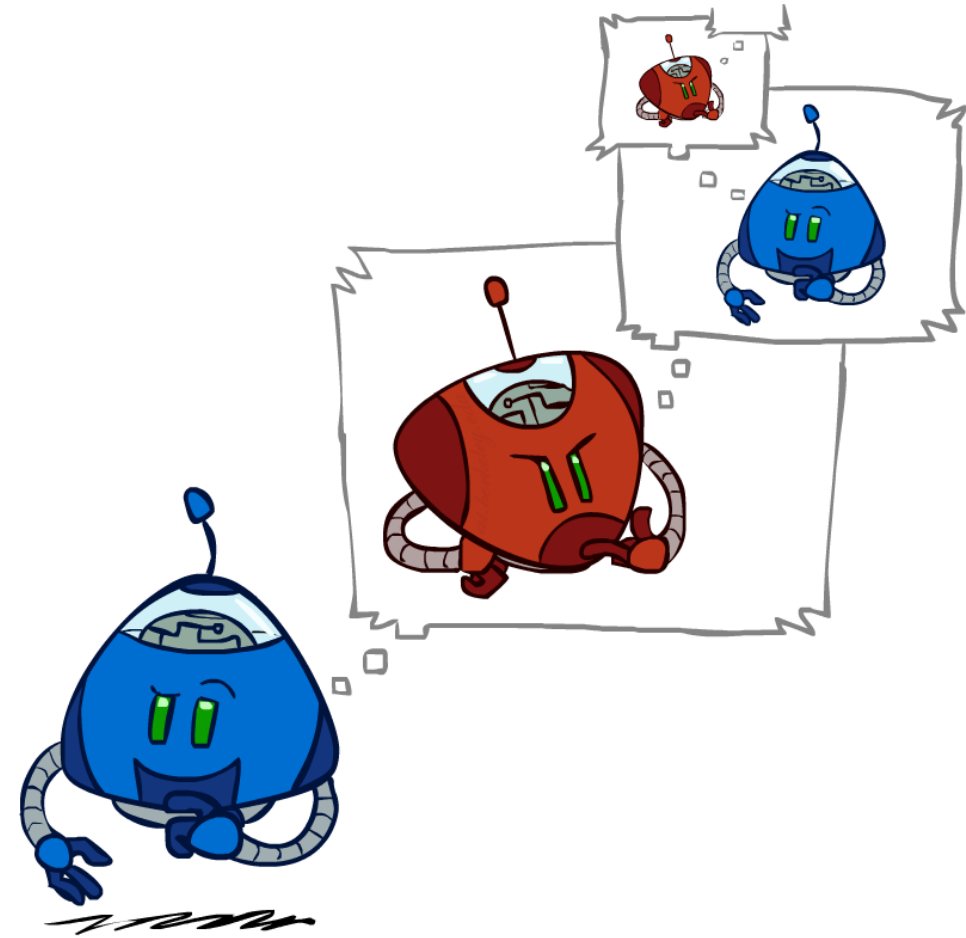
B) DFS

C) UCS

D) A*

Minimax Efficiency

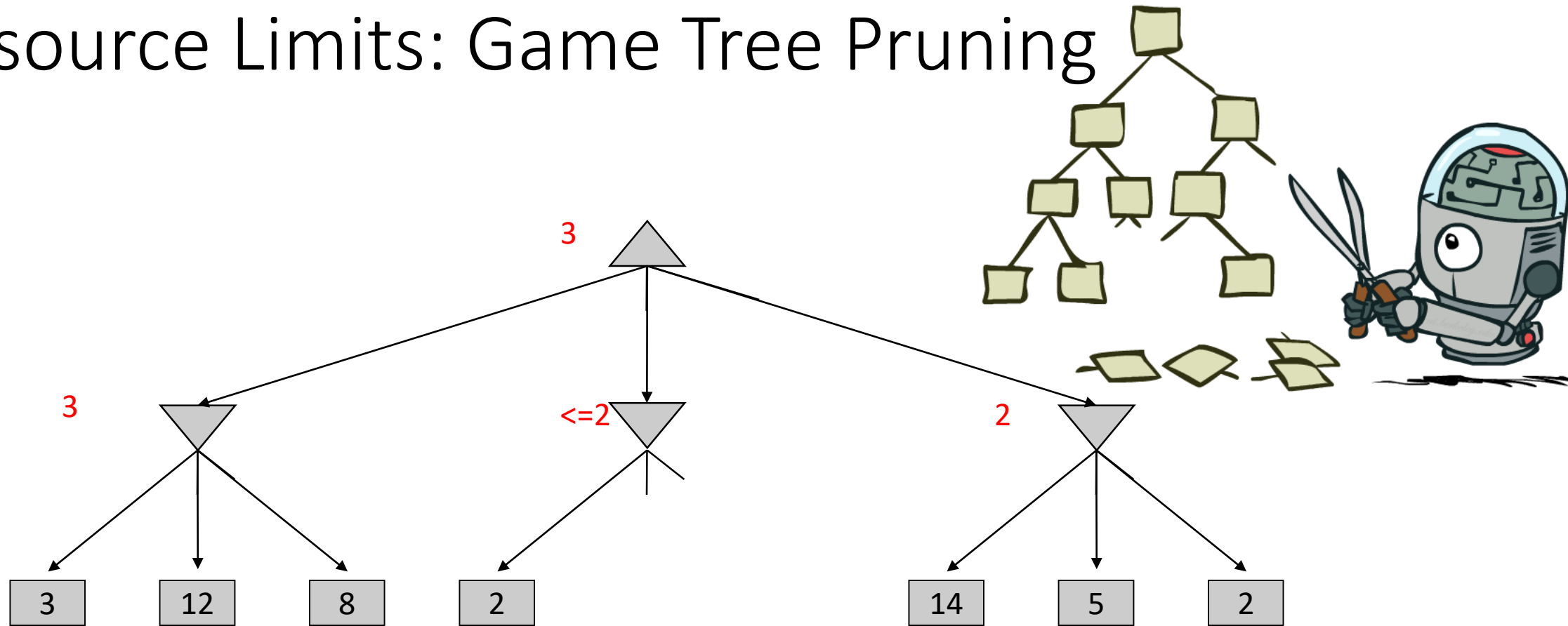
- How efficient is minimax?
 - Just like (exhaustive) DFS
 - Time: $O(b^m)$
 - Space: $O(bm)$
- Example: For chess, $b \approx 35$, $m \approx 100$
 - Exact solution is completely infeasible
 - But, do we need to explore the whole tree?
 - Humans can't do this either, so how do we play chess?
 - **Bounded rationality** – Herbert Simon



Resource Limits



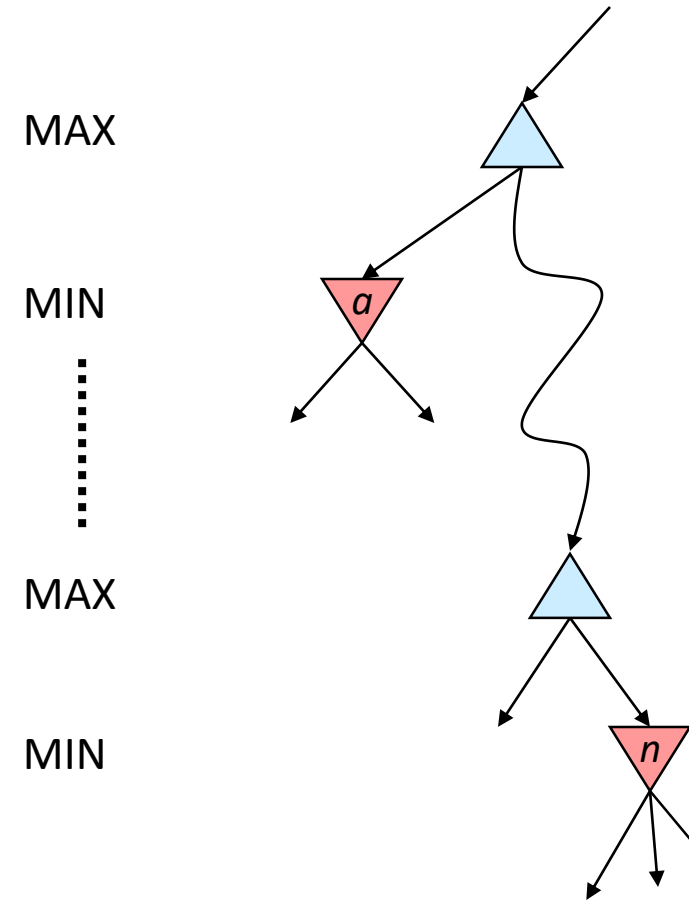
Resource Limits: Game Tree Pruning



The order of generation matters: more pruning is possible if good moves come first

Game Tree 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
 - n 's estimate of the childrens' min is dropping
 - Who cares about n 's value? MAX
 - Let a be the best value that MAX can get at any choice point along the current path from the root
 - If n becomes smaller than a , MAX will avoid it, so we can stop considering n 's other children (it's already bad enough that it won't be played)
- MAX version is symmetric



Alpha-Beta Implementation

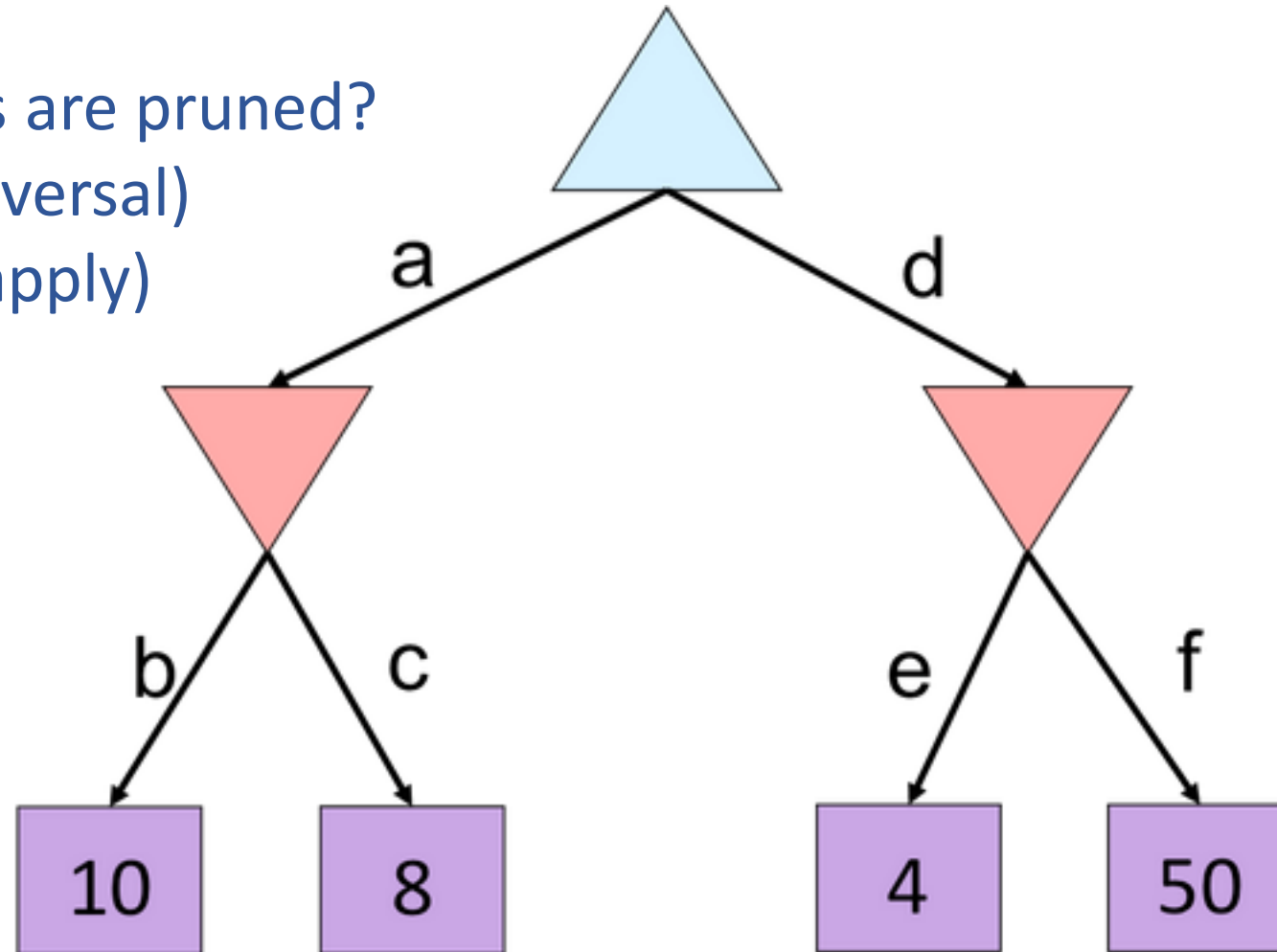
α : MAX's best option on path to root
 β : MIN's best option on path to root

```
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  $v$ 
```

Quiz a

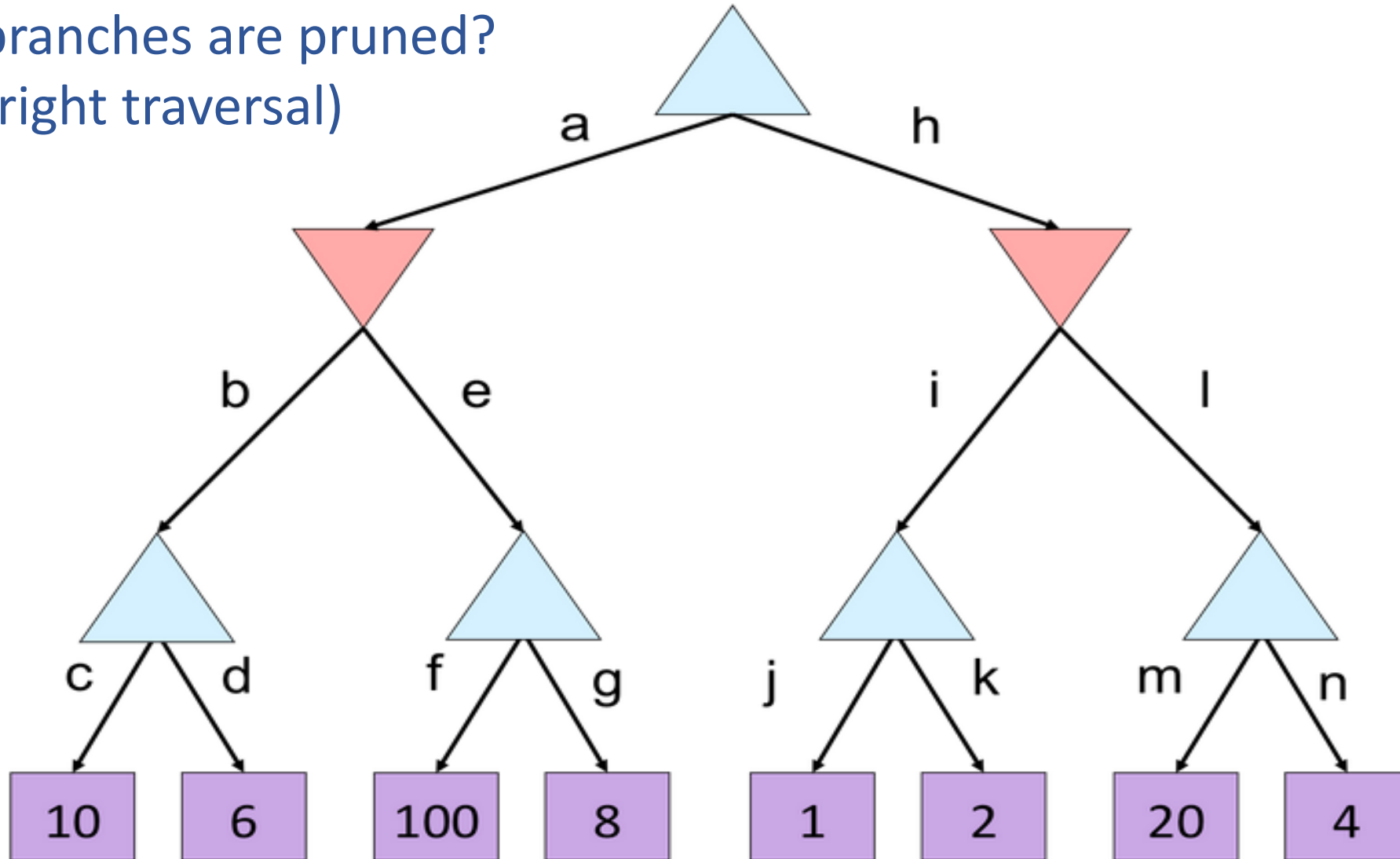
Which branches are pruned?
(Left to right traversal)
(Select all that apply)



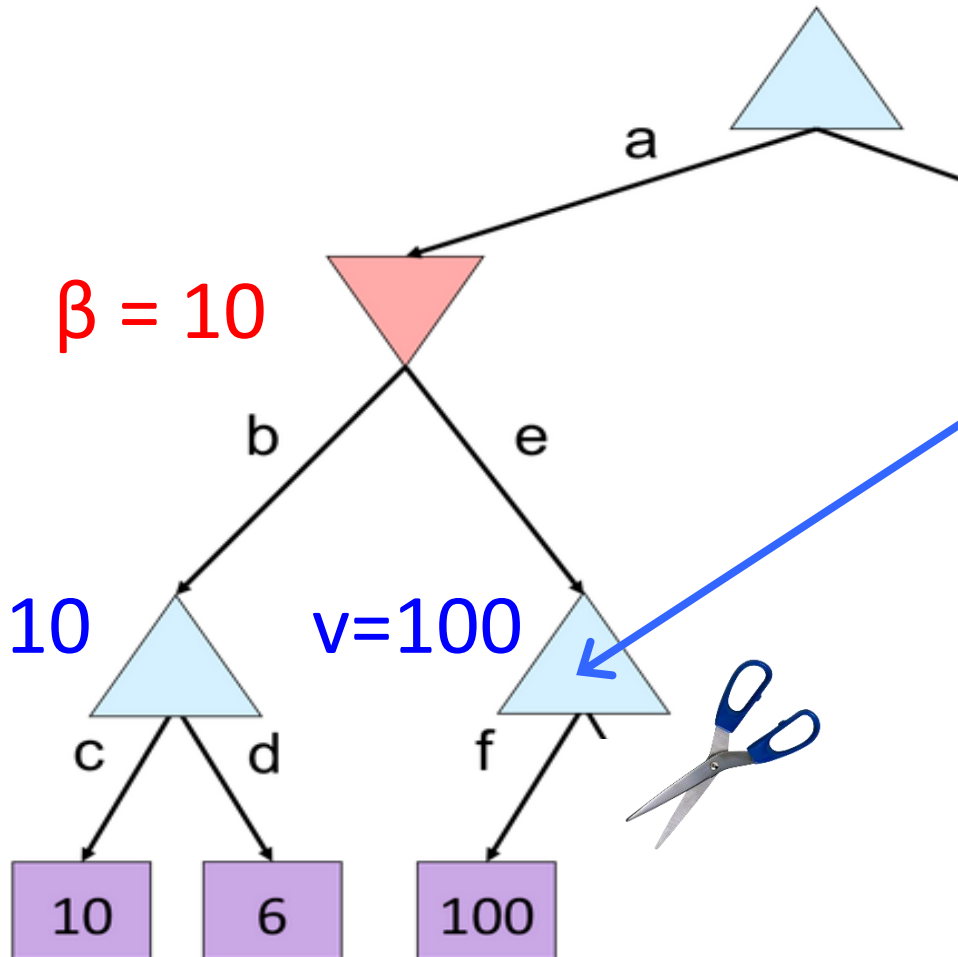
Quiz b

Which branches are pruned?
(Left to right traversal)

- A) e, l
- B) g, l
- C) g, k, l
- D) g, n



Quiz b - 2



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

def max-value(state, α , β):

initialize $v = -\infty$

for each successor of state:

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

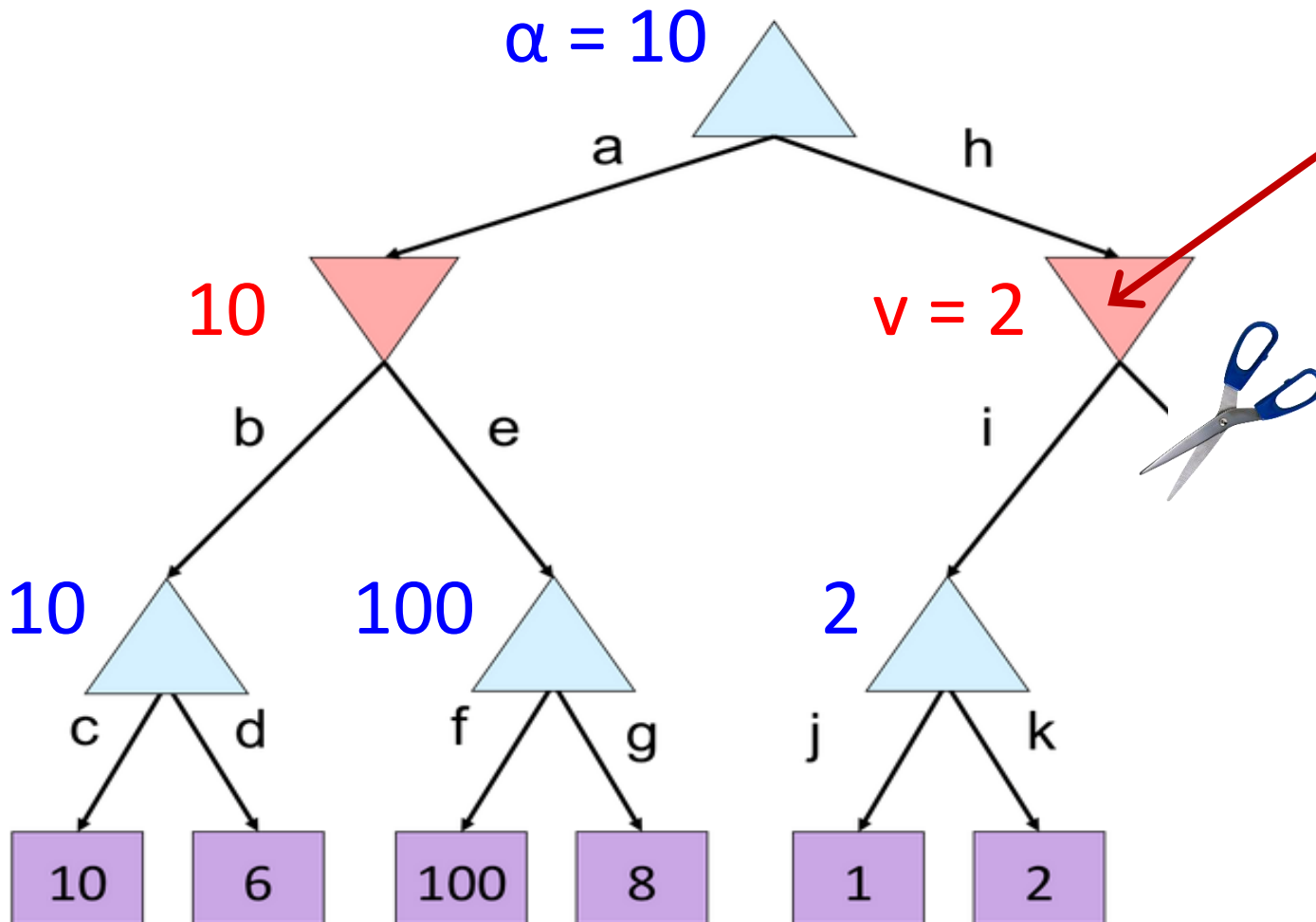
if $v \geq \beta$

return v

$\alpha = \max(\alpha, v)$

return v

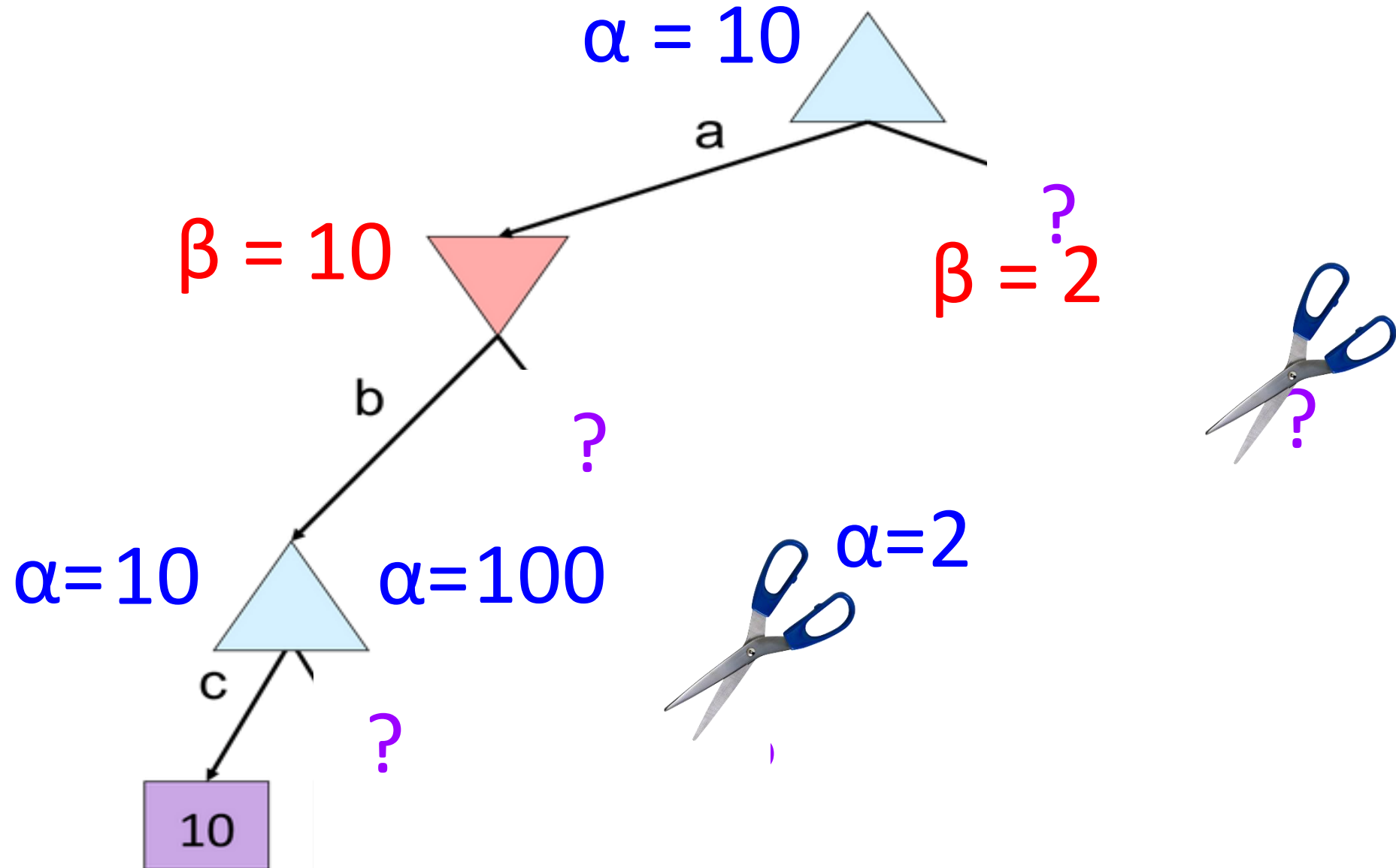
Quiz b - 3



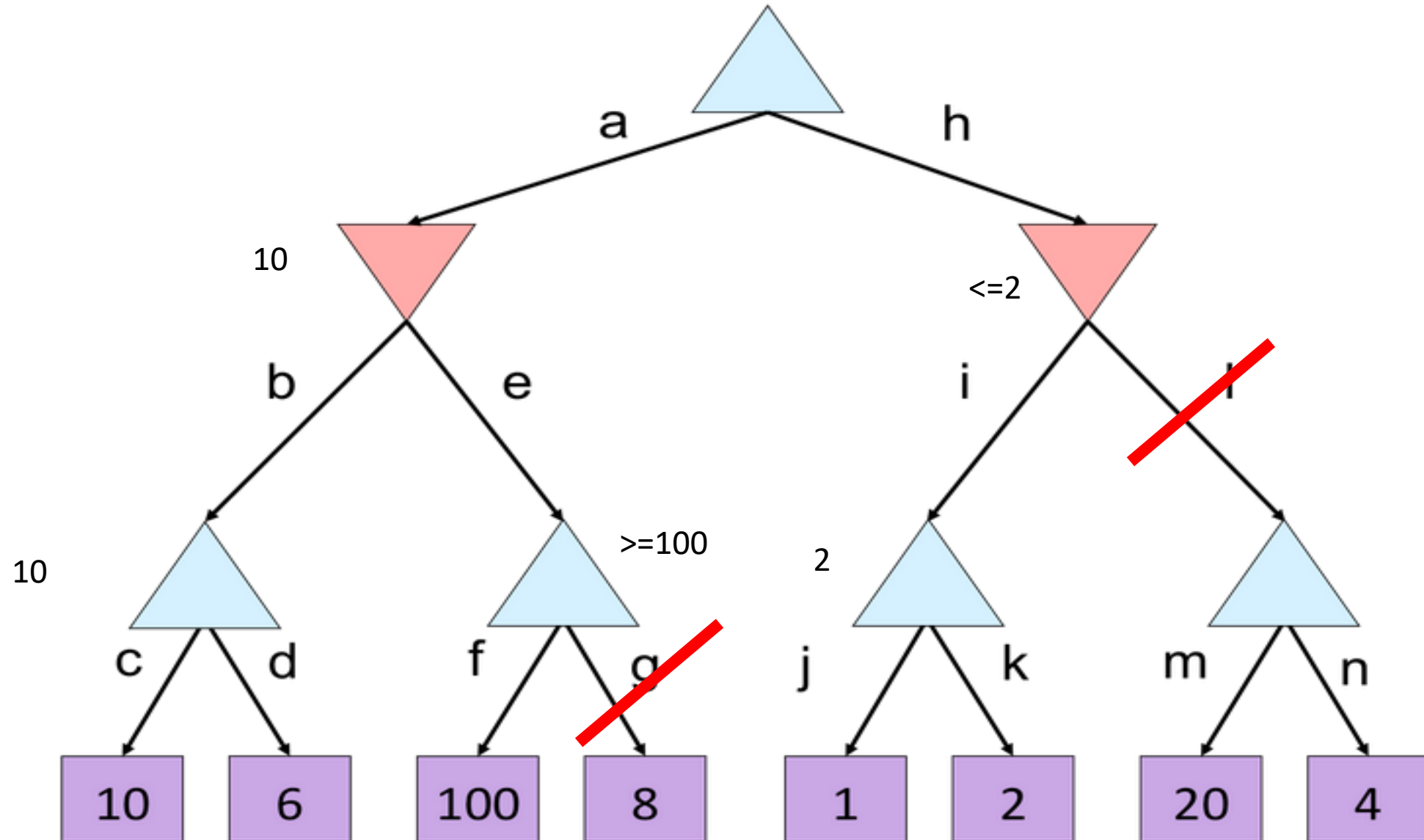
α : MAX's best option on path to root
 β : MIN's best option on path to root

```
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  $v$ 
```

Quiz b - 4

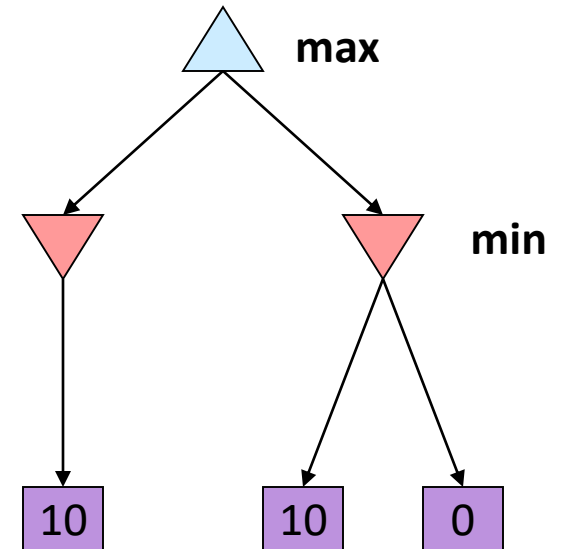


Quiz b - 5



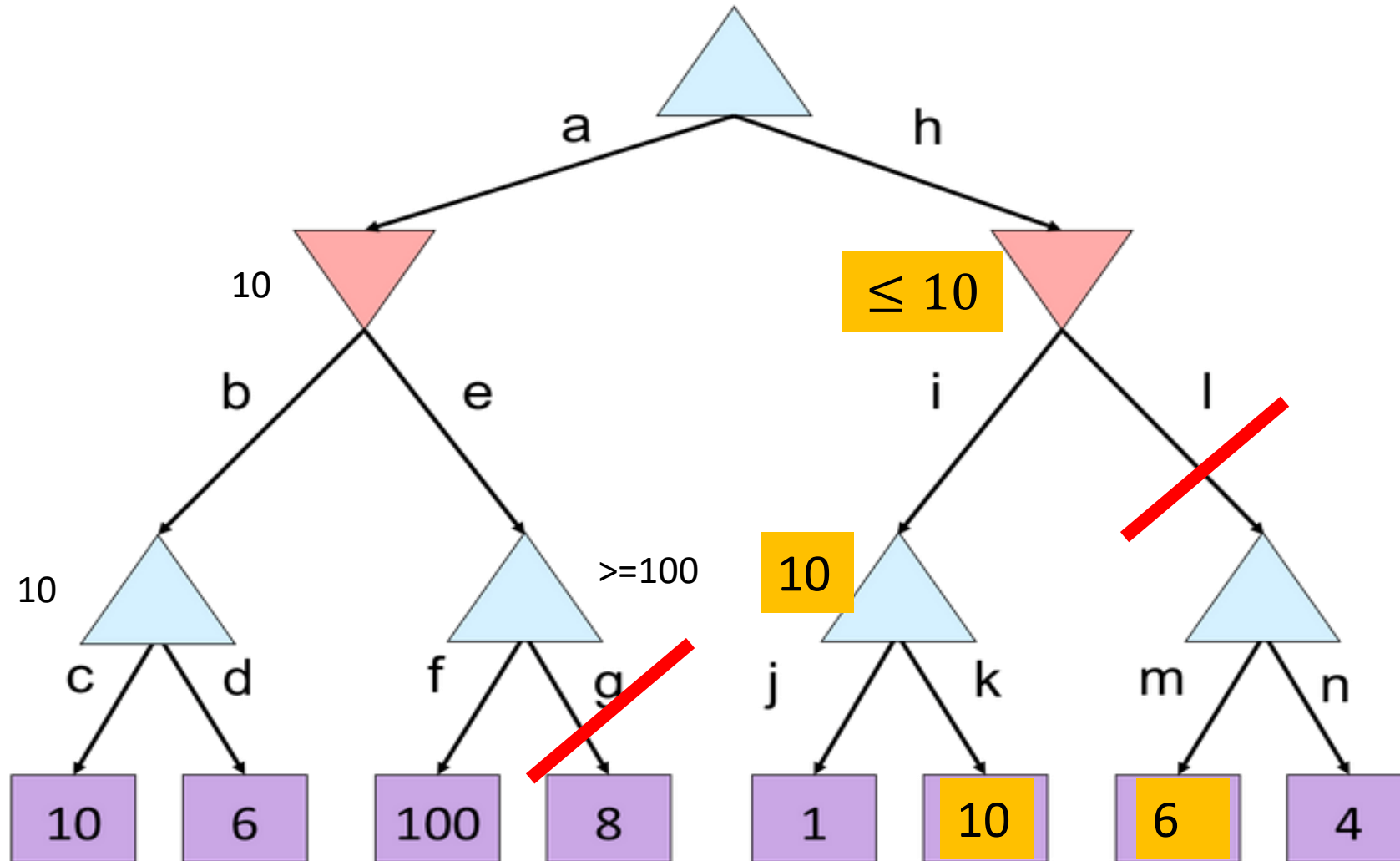
Alpha-Beta Pruning Properties

- This pruning has **no effect** on minimax value computed for the root!
- Values of intermediate nodes might be wrong
 - Important: children of the root may have the wrong value
 - **So the most naïve version won't let you do action selection**
- Good child ordering improves effectiveness of pruning
- With “perfect ordering”:
 - Time complexity drops to $O(b^{m/2})$
 - Doubles solvable depth!
 - Chess: 1M nodes/move => depth=8, respectable
 - Full search of complicated games, is still hopeless...



- This is a simple example of **metareasoning** (computing about what to compute)

Alpha-Beta Pruning: Action Selection Failure



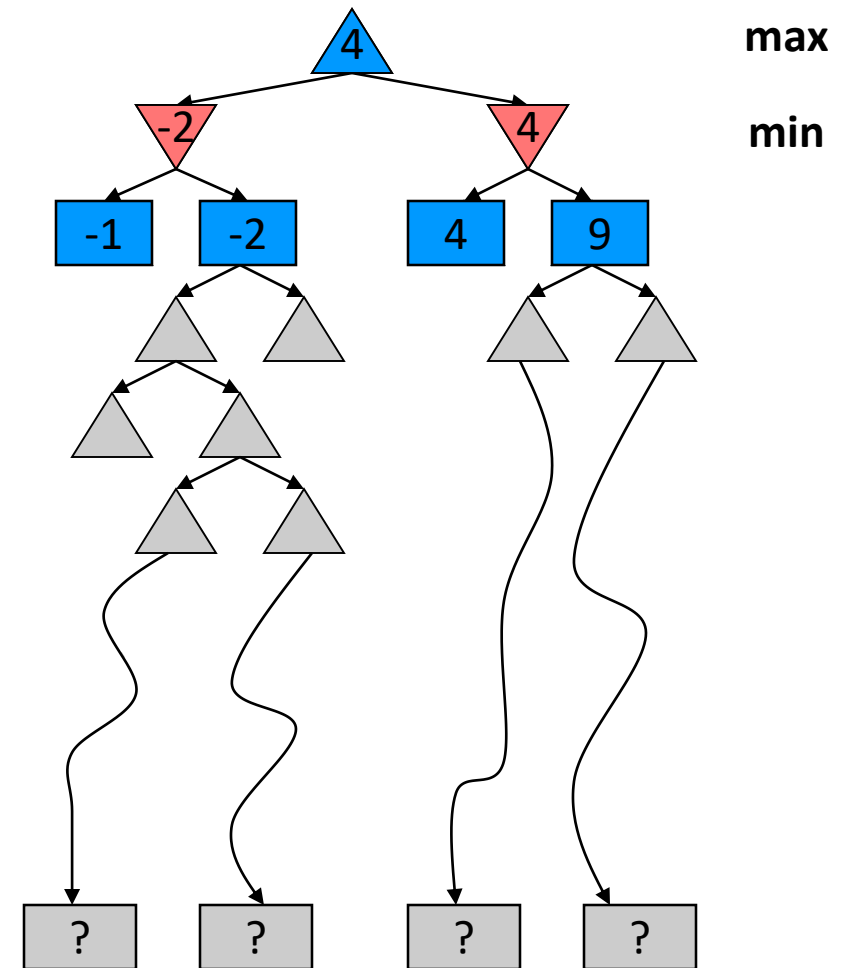
Resource Limits II

Bounded lookahead



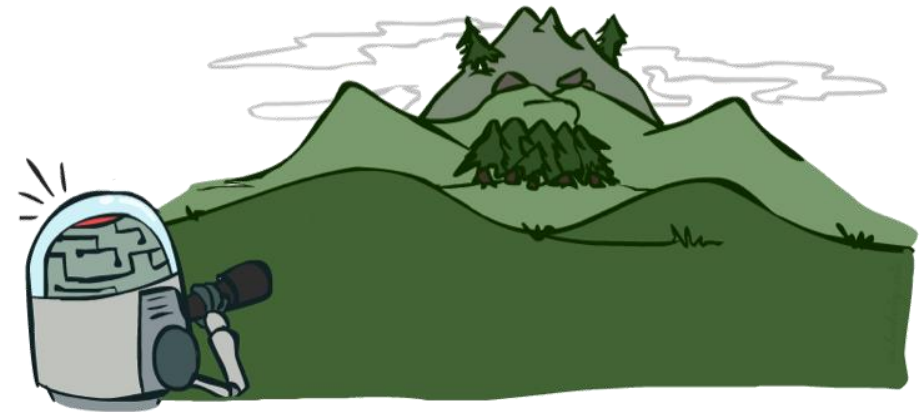
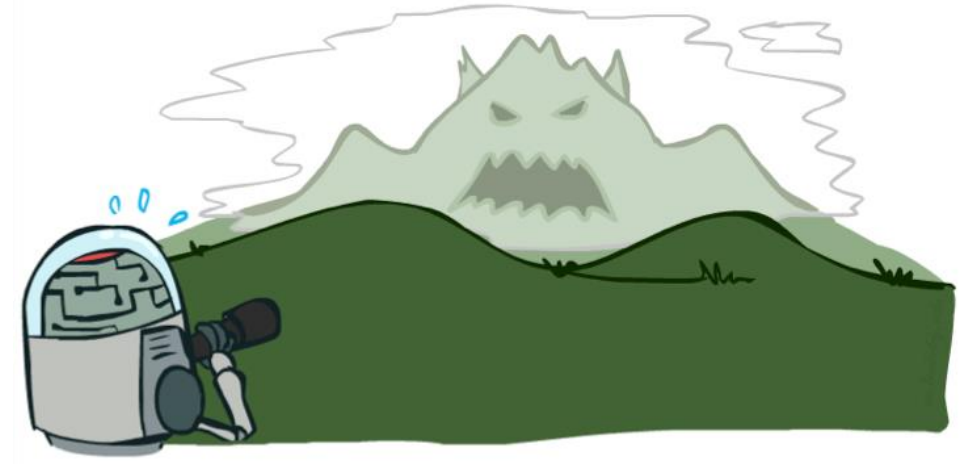
Depth-limited search

- 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
 - For chess, $b \approx 35$ so reaches about depth 4 – not so good
 - α - β reaches about depth 8 – decent chess program
- **Guarantee of optimal play is gone**
- **More plies makes a BIG difference**
- Use iterative deepening for an anytime algorithm

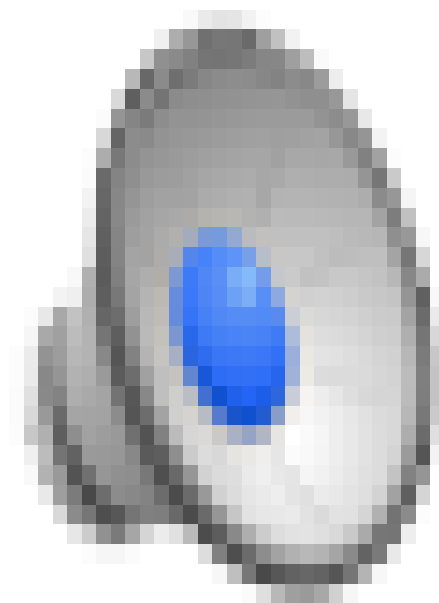


Depth Matters

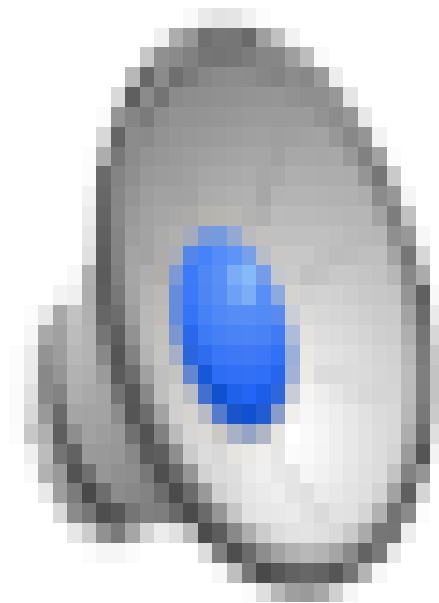
- Evaluation functions are always imperfect
- The deeper in the tree the evaluation function is buried, the less the quality of the evaluation function matters
- An important example of the **tradeoff** between complexity of features and complexity of computation



Video of Demo Limited Depth (2)



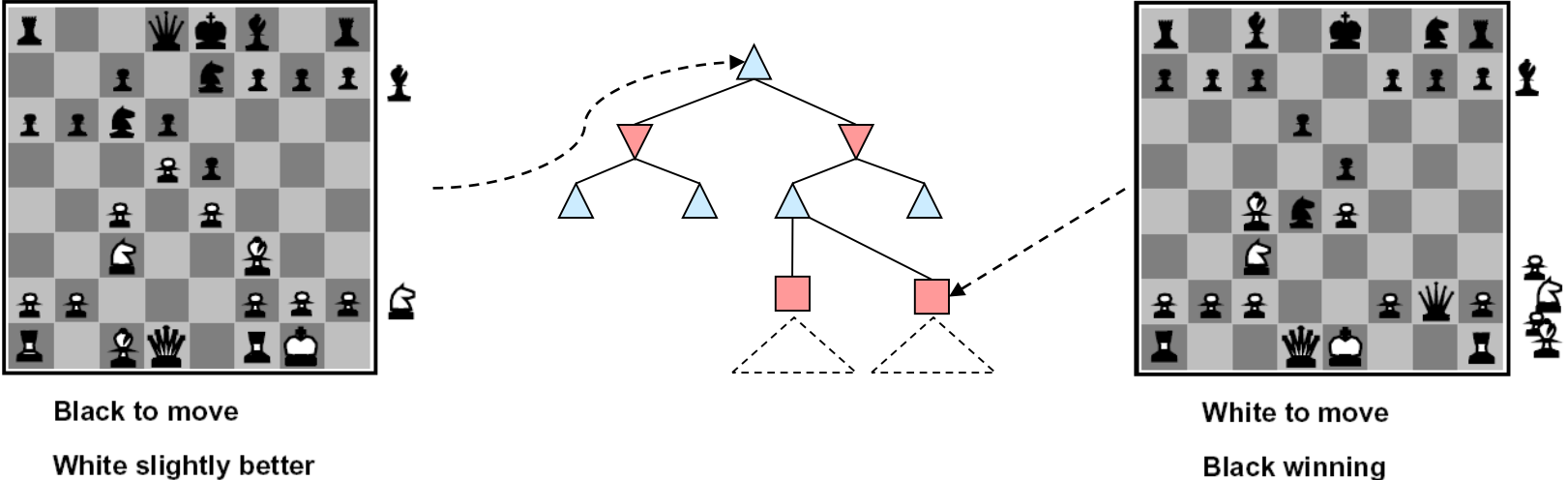
Video of Demo Limited Depth (10)





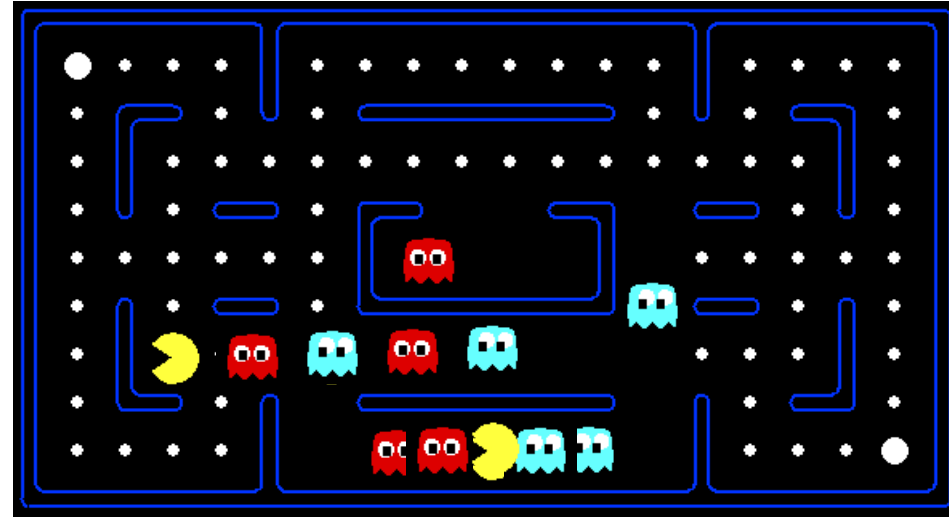
Evaluation Functions

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



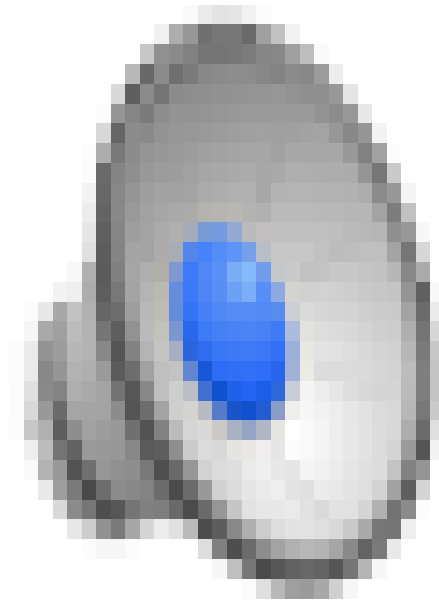
- Ideal function: returns the actual minimax value of the position
- In practice: typically weighted linear sum of features:
$$\text{EVAL}(s) = w_1 f_1(s) + w_2 f_2(s) + \dots + w_n f_n(s)$$
 - e.g. $w_1 = 9$, $f_1(s) = (\text{num white queens} - \text{num black queens})$, etc.

Evaluation for Pacman

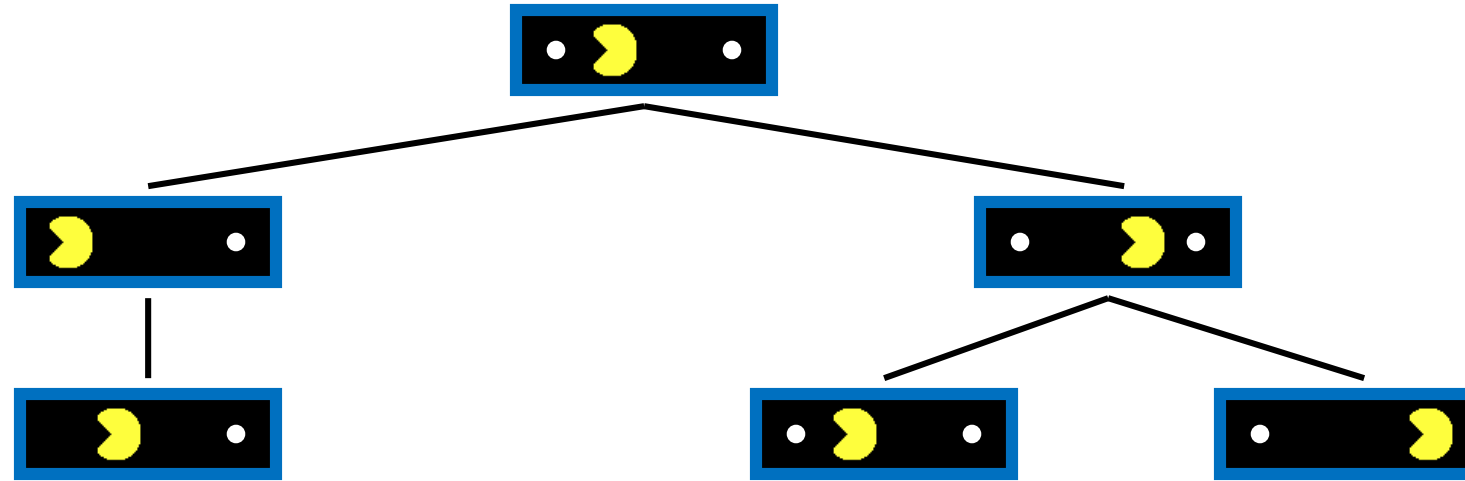


[Demo: thrashing $d=2$, thrashing $d=2$ (fixed evaluation function), smart ghosts coordinate (L6D6,7,8,10)]

Video of Demo Thrashing (d=2) (Failure)



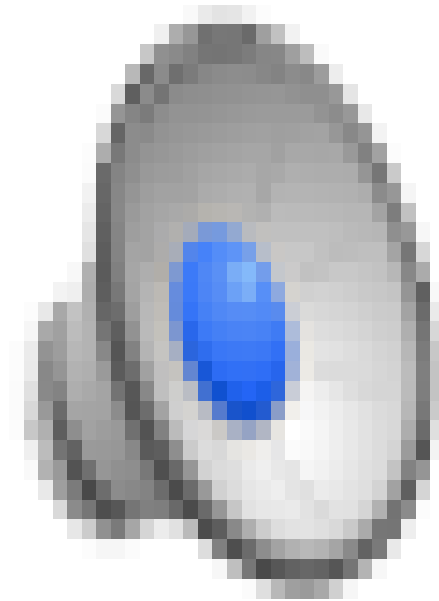
Why Pacman Starves



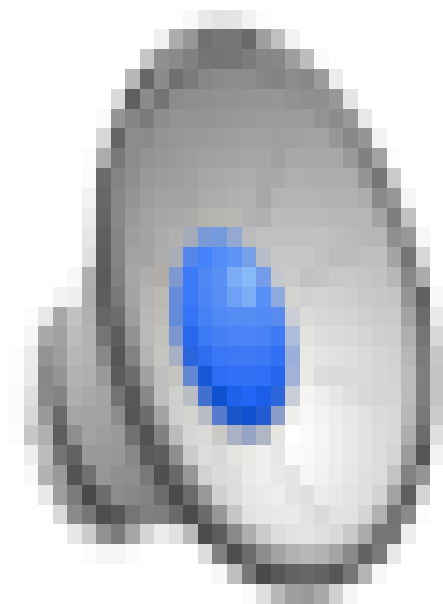
- A danger of replanning agents!

- He knows his score will go up by eating the dot now (west, east)
- He knows his score will go up just as much by eating the dot later (east, west)
- There are no point-scoring opportunities after eating the dot (within the horizon, two here)
- Therefore, waiting seems just as good as eating: he may go east, then back west in the next round of replanning!

Video of Demo Thrashing -- Fixed (d=2)

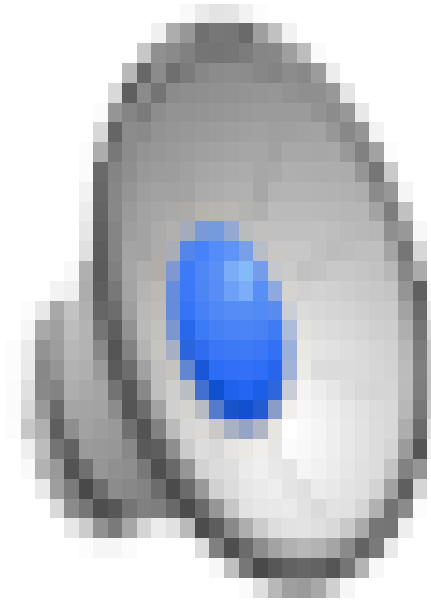


Video of Demo **Smart** Ghosts (Coordination)



Video of Demo Smart Ghosts (Coordination) – Zoomed In

- Cooperation behaviors naturally arise based on the minimax strategies and common evaluation functions

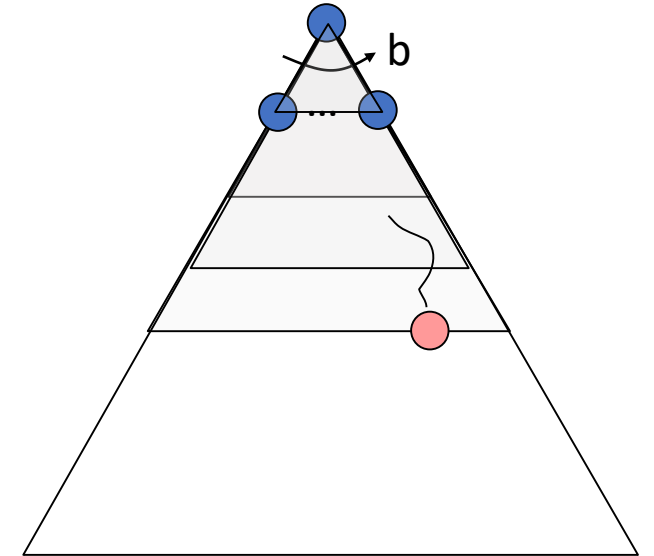


Iterative Deepening

Iterative deepening uses DFS as a subroutine:

1. Do a DFS which only searches for paths of length 1 or less. (DFS gives up on any path of length 2)
2. If “1” failed, do a DFS which only searches paths of length 2 or less.
3. If “2” failed, do a DFS which only searches paths of length 3 or less.

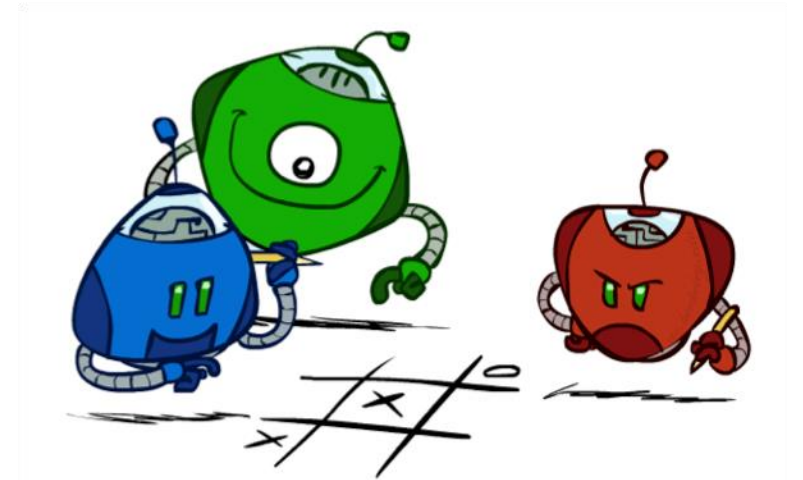
....and so on.



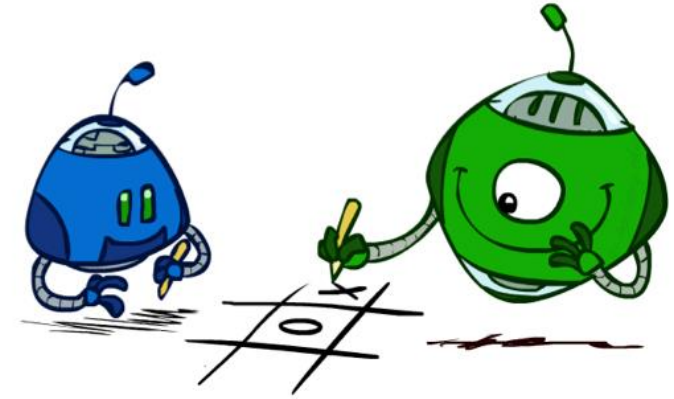
Why do we want to do this for multiplayer games?

Note: wrongness of eval functions matters less and less the deeper the search goes!

Generalized minimax



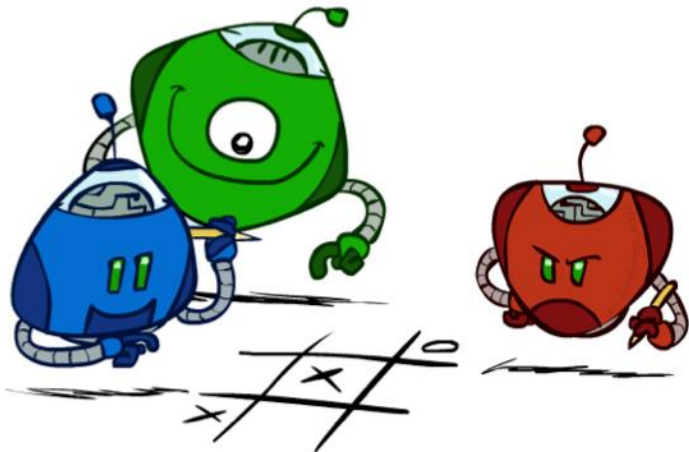
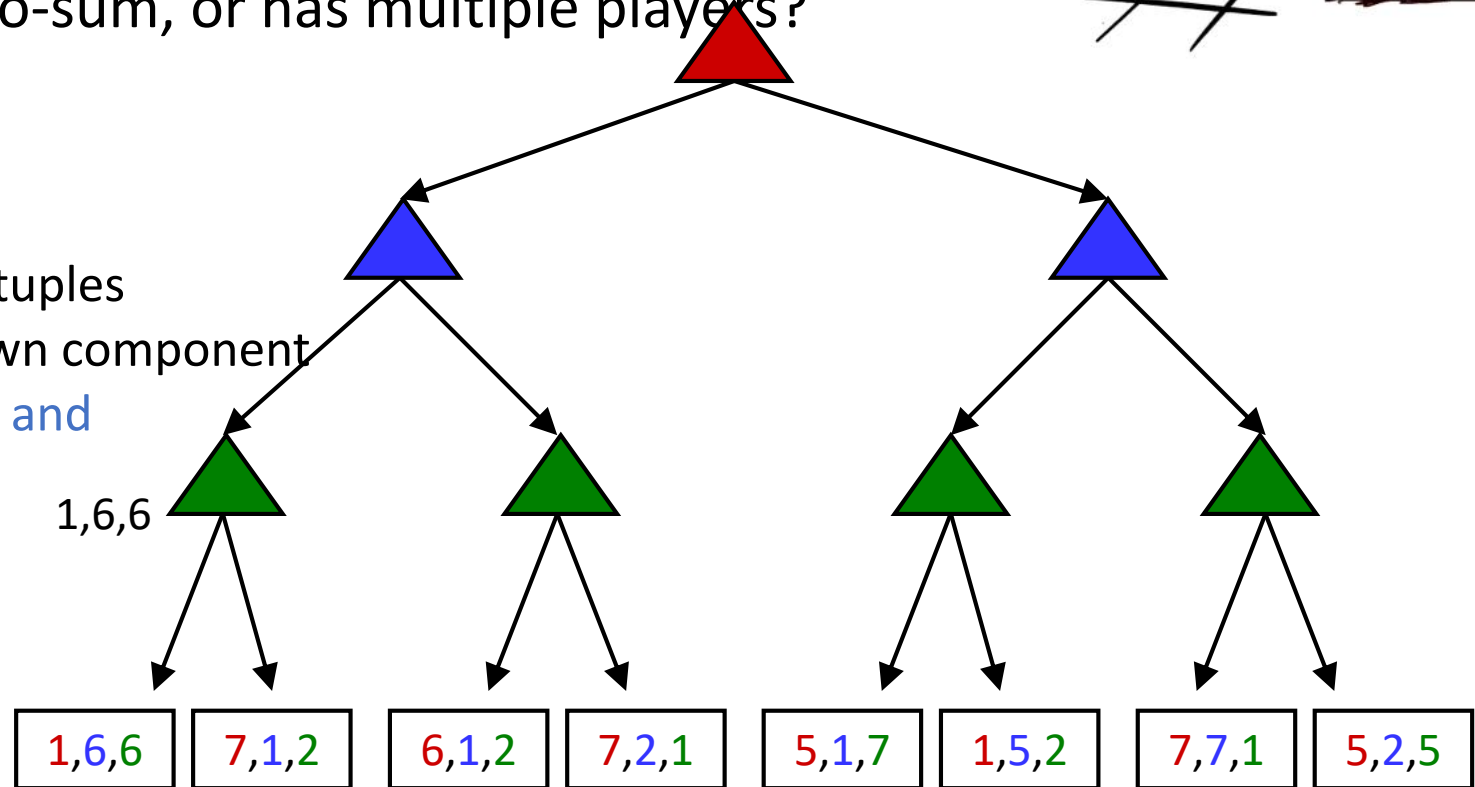
Multi-Agent Utilities



- What if the game is not zero-sum, or has multiple players?

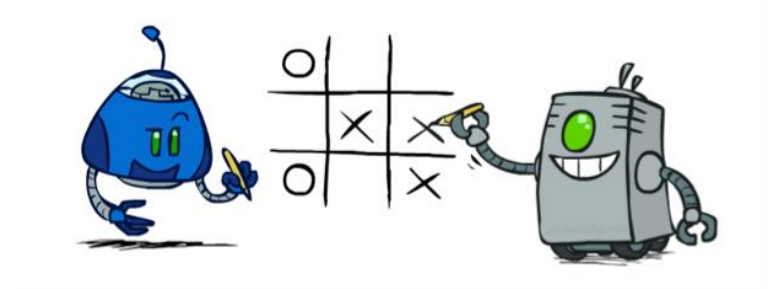
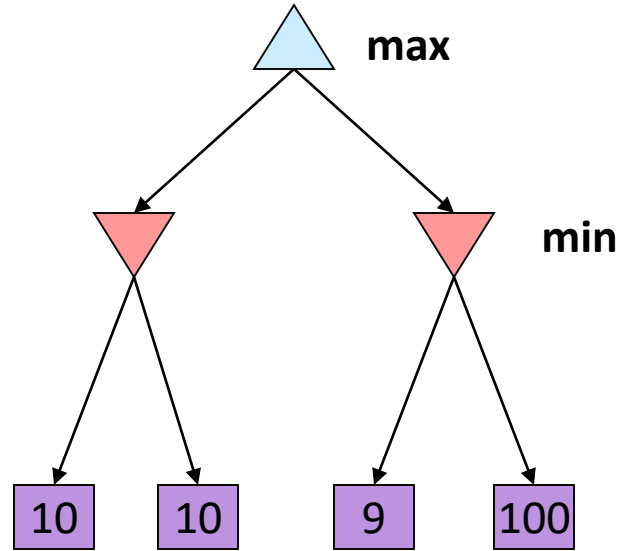
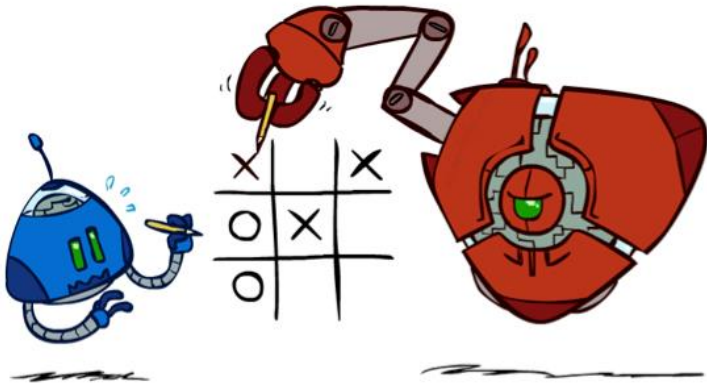
- Generalization of minimax:

- Terminals have utility tuples
- Node values are also utility tuples
- Each player maximizes its own component
- Can give rise to cooperation and competition dynamically...



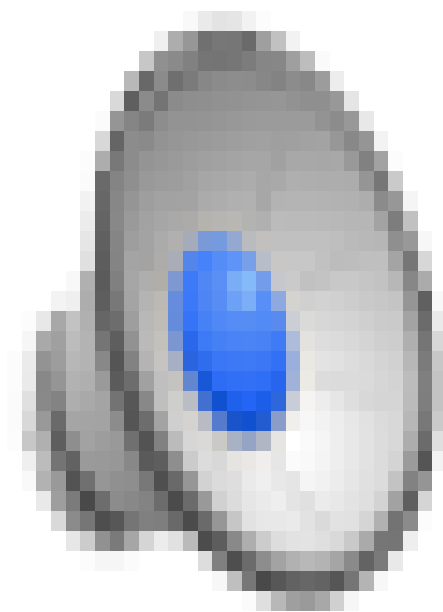
Modeling Assumptions

What if your opponent isn't playing optimally?

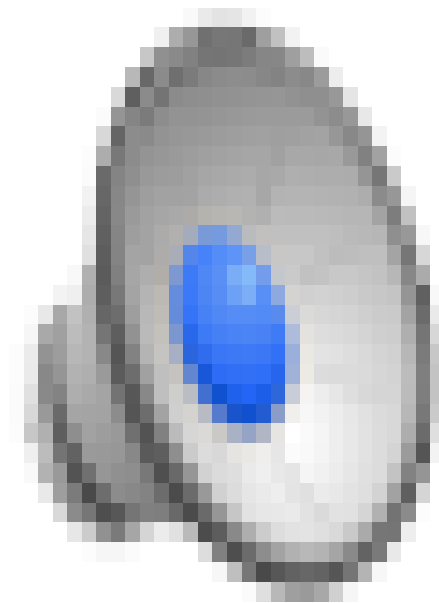


Optimal against a perfect player. Otherwise?

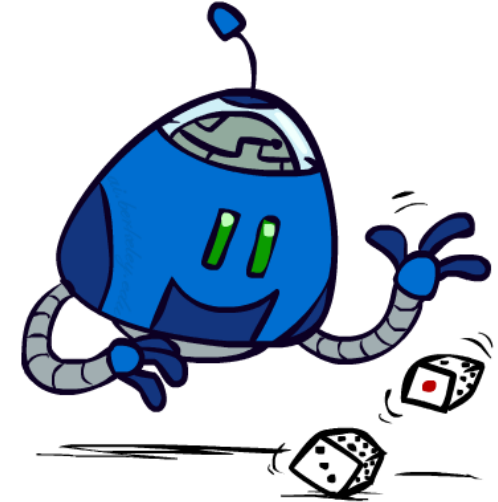
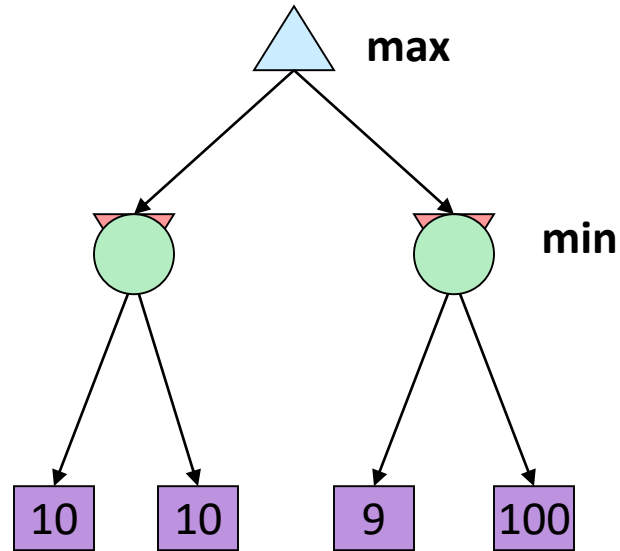
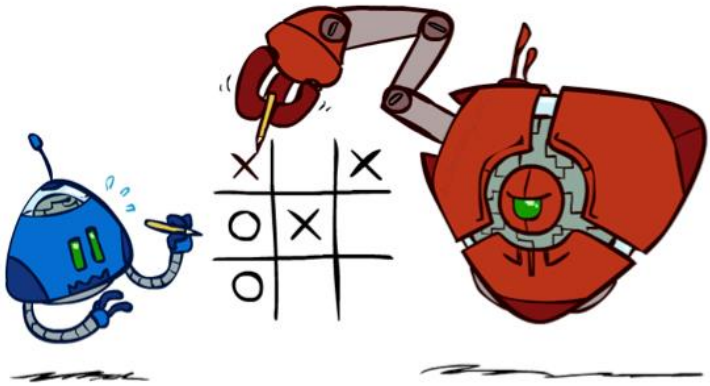
Video of Demo Min vs. Exp (Min)



Video of Demo Min vs. Exp (Exp)



Worst-Case vs. Average Case



Idea: Uncertain outcomes controlled by chance, not an adversary!

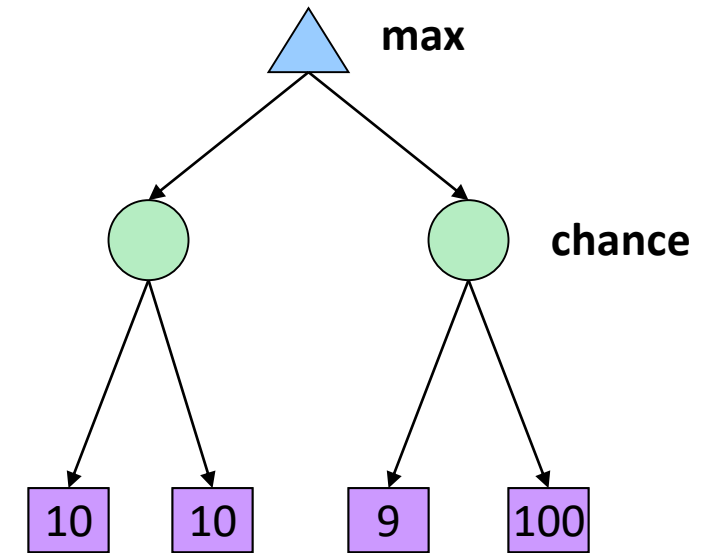
Why not minimax?

- Worst case reasoning is too conservative
- Need average case reasoning

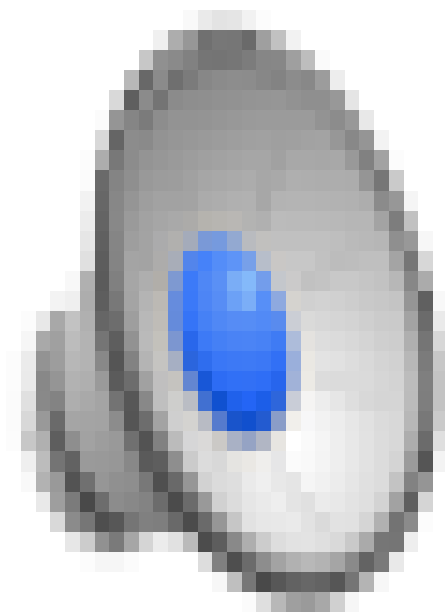


Expectimax Search

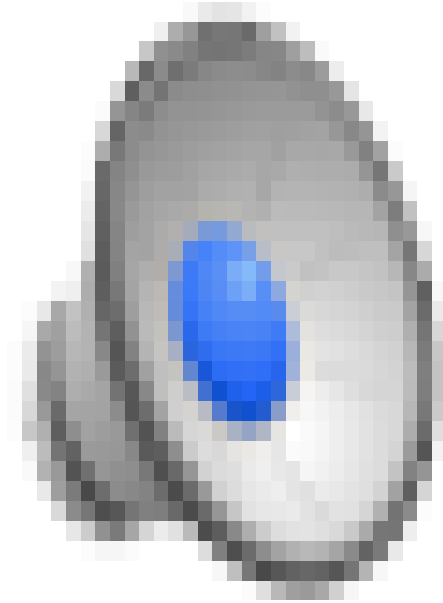
- Why wouldn't we know what the result of an action will be?
 - Explicit randomness: rolling dice
 - Unpredictable opponents: the ghosts respond randomly
 - Unpredictable humans: humans are not perfect
 - Actions can fail: when moving a robot, wheels might slip
- Values should now reflect average-case (expectimax) outcomes, not worst-case (minimax) outcomes
- **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. take weighted average (expectation) of children
- Later, we'll learn how to formalize the underlying uncertain-result problems as **Markov Decision Processes**



Video of Demo Minimax vs Expectimax (Min)



Video of Demo Minimax vs Expectimax (Exp)



Expectimax Pseudocode

```
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 EXP: return exp-value(state)
```

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

```
    initialize v =  $-\infty$ 
```

```
    for each successor of state:
```

```
        v = max(v, value(successor))
```

```
    return v
```

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

```
    initialize v = 0
```

```
    for each successor of state:
```

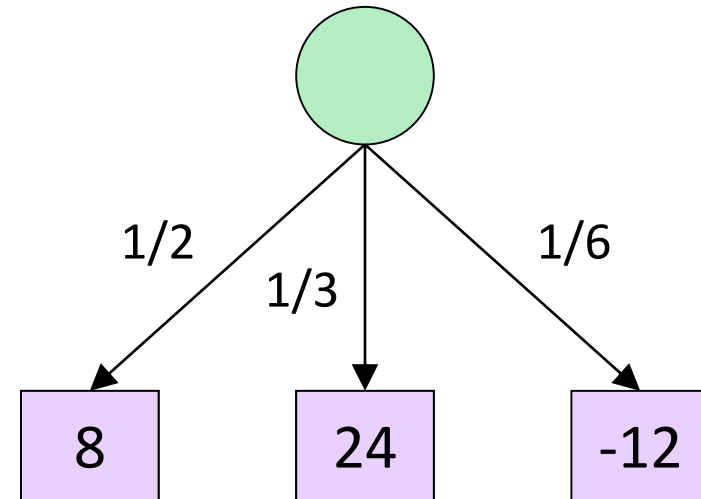
```
        p = probability(successor)
```

```
        v += p * value(successor)
```

```
    return v
```

Expectimax Pseudocode 2

```
def exp-value(state):  
    initialize v = 0  
    for each successor of state:  
        p = probability(successor)  
        v += p * value(successor)  
    return v
```

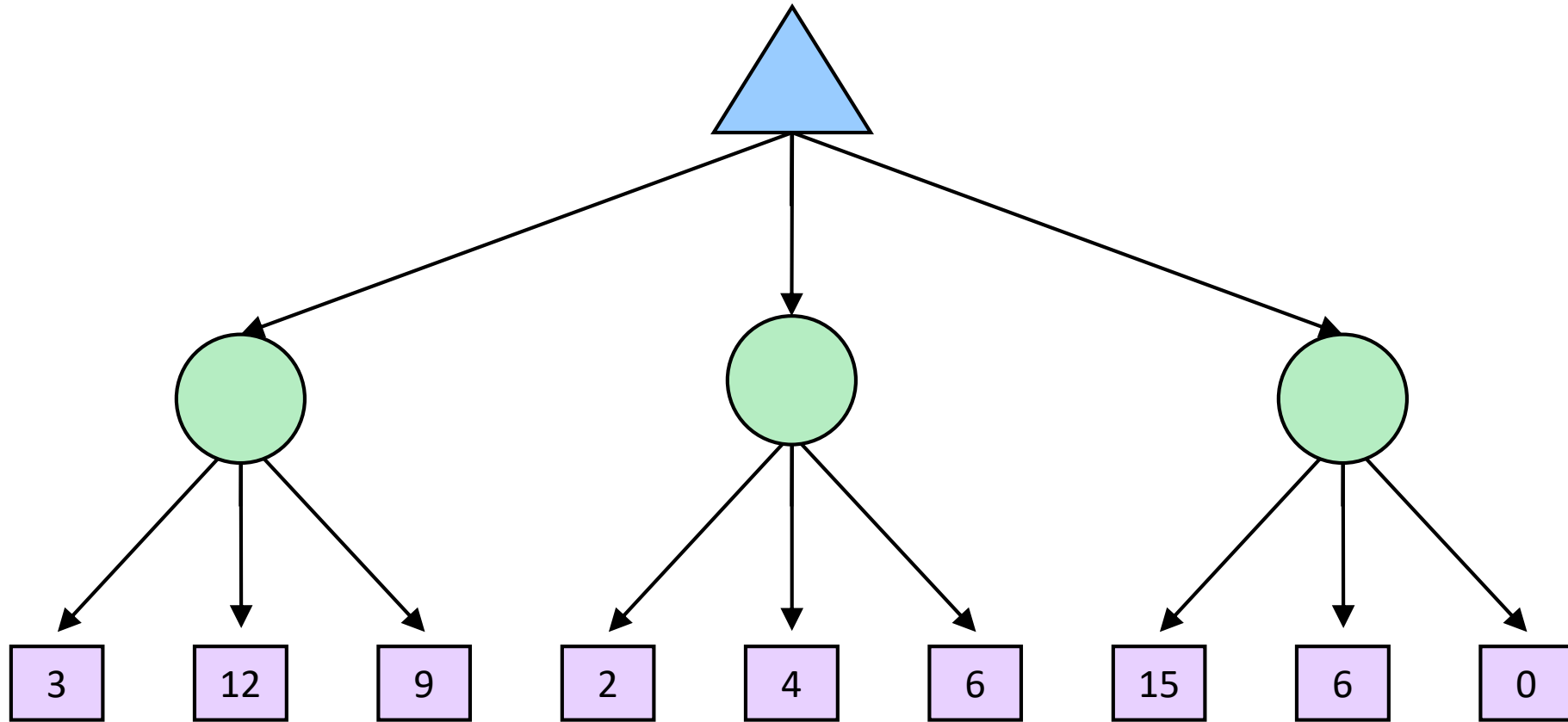


$$v = (1/2) (8) + (1/3) (24) + (1/6) (-12) = 10$$

Expectimax Pseudocode 3

- function **value**(state)
 - if state.is_leaf
 - return state.value
 - if state.player is **MAX**
 - return **max**_{a in state.actions} **value**(state.result(a))
 - if state.player is **MIN**
 - return **min**_{a in state.actions} **value**(state.result(a))
 - if state.player is **CHANCE**
 - return **sum**_{s in state.next_states} **P**(s) * **value**(s)

Example



Quiz

Expectimax tree search:

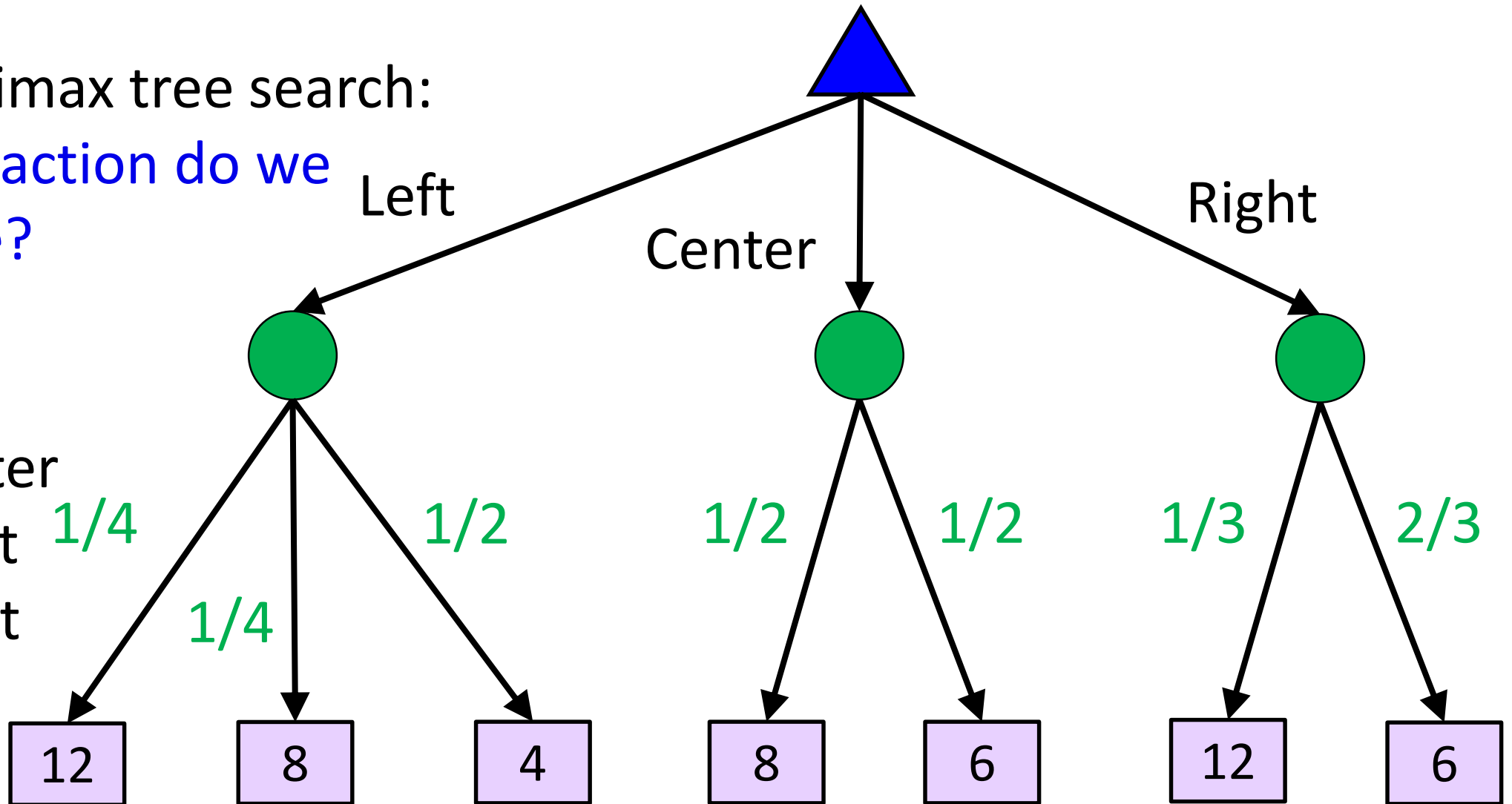
Which action do we choose?

A: Left

B: Center

C: Right

D: Eight



Quiz 2

Expectimax tree search:

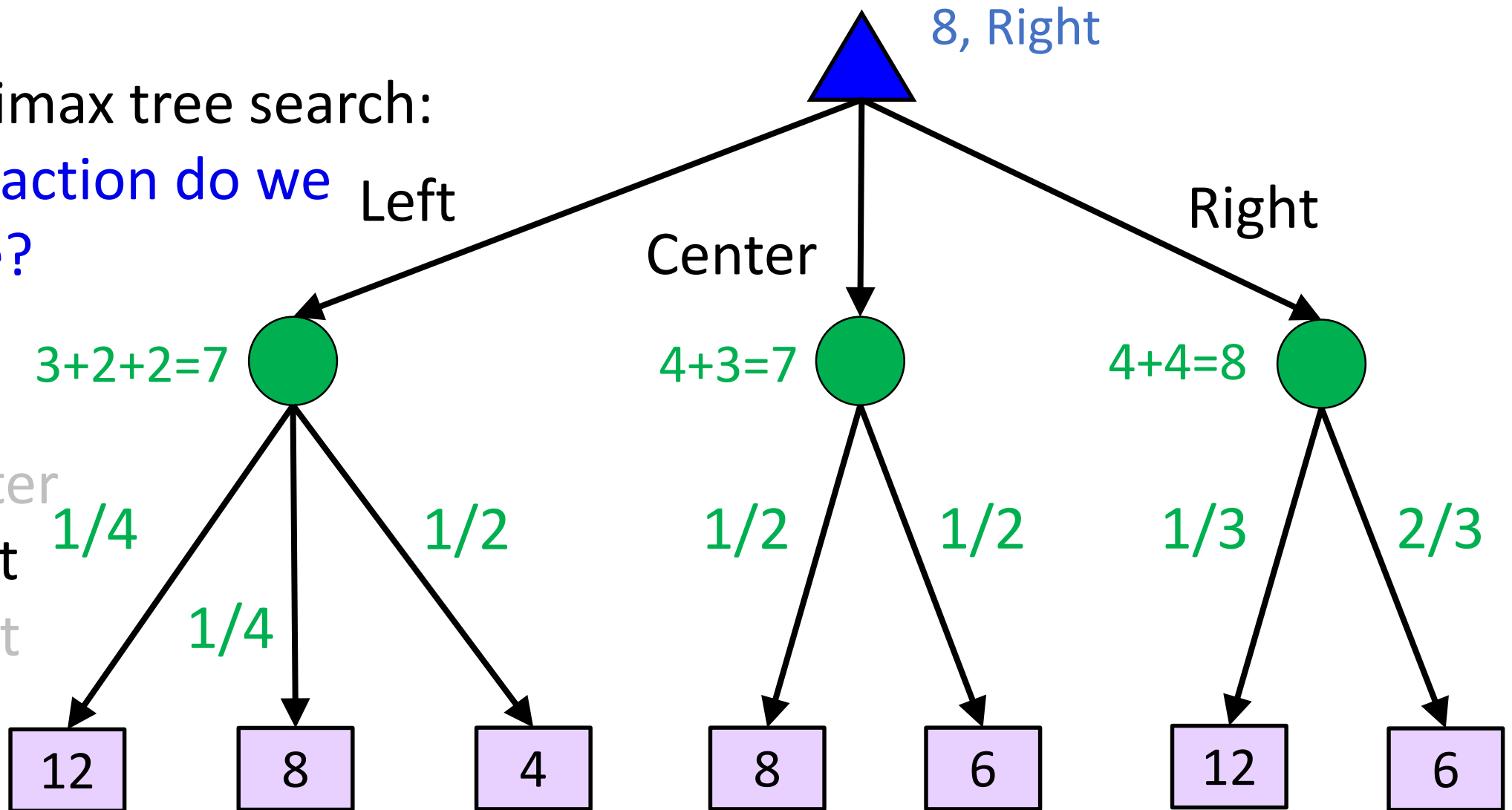
Which action do we choose?

A: Left

B: Center

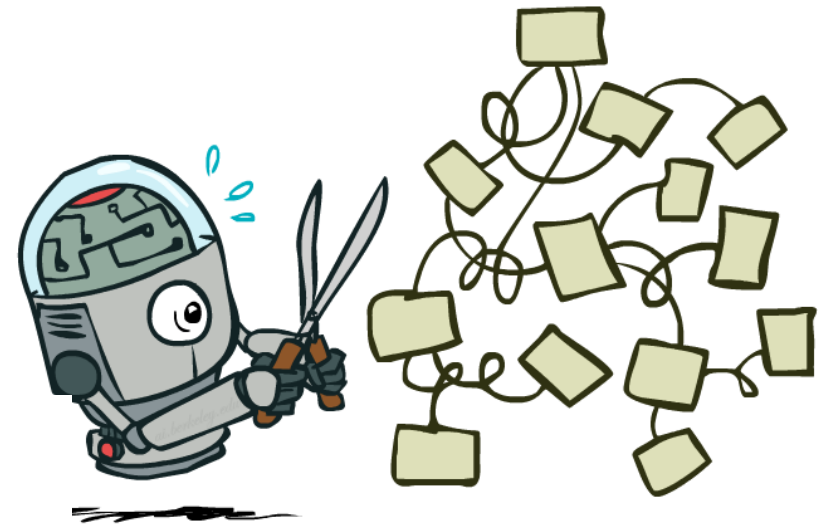
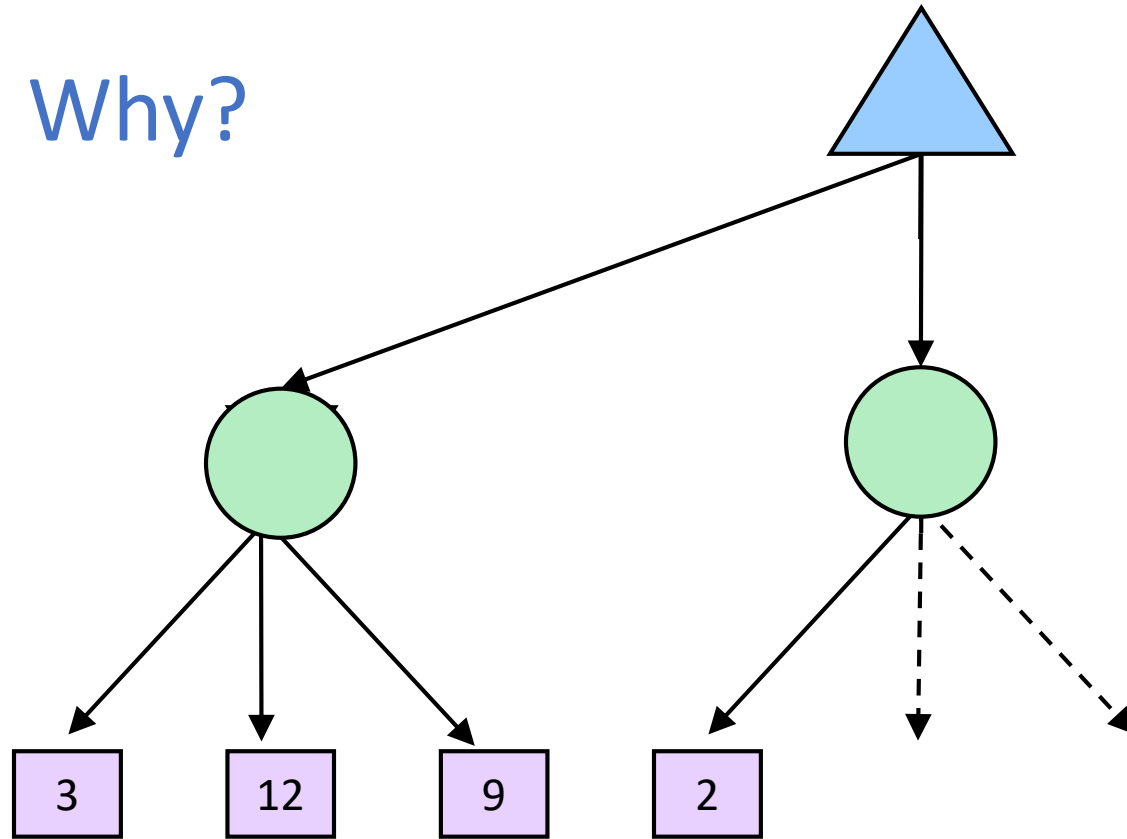
C: Right

D: Eight

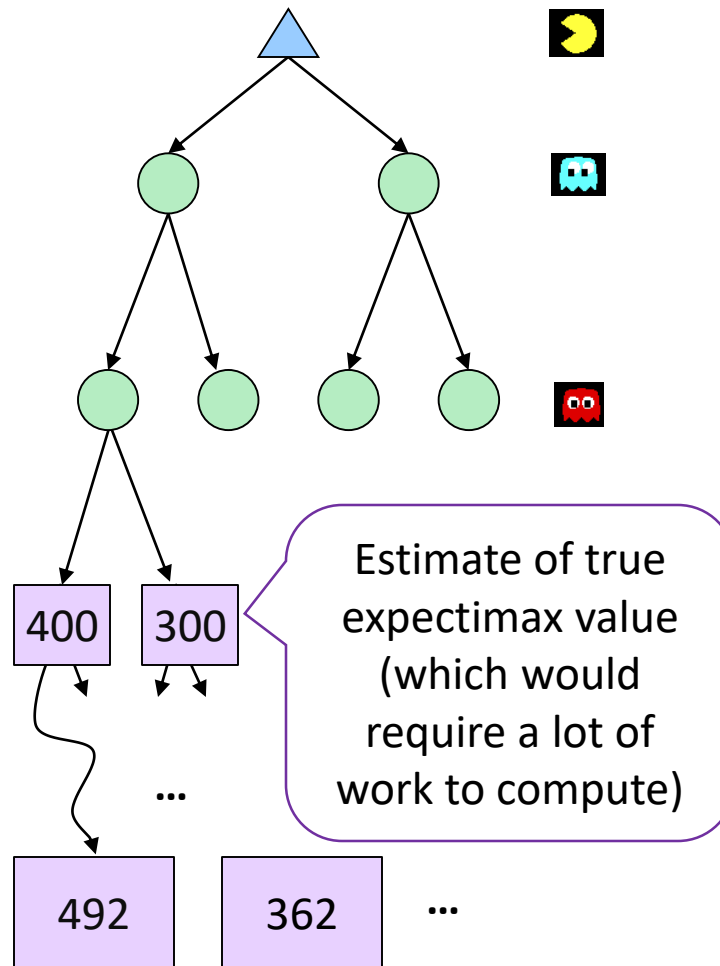


Expectimax Pruning?

No! Why?

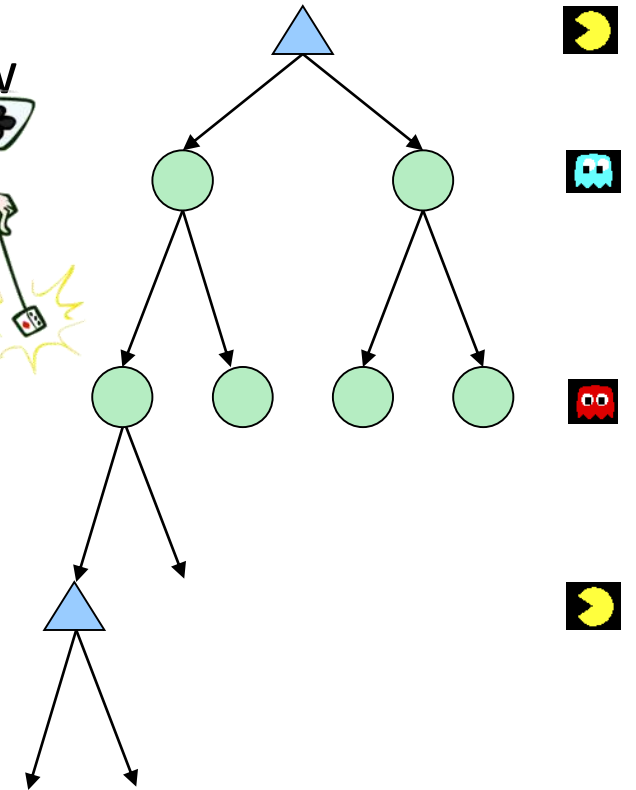


Expectimax: Depth-Limited



We know how to compute with probabilities. But where do probabilities arise?

- In expectimax search, we have a probabilistic model of how the opponent (or environment) will behave in an action
 - Model could be a simple uniform distribution (roll a die)
 - Model could be sophisticated and require a great deal of computation
 - We have a chance node for any outcome out of our control: opponent or environment
 - The model might say that adversarial actions are likely!
- For now, assume each chance node magically comes along with probabilities that specify the distribution over its outcomes



Having a probabilistic belief about another agent's action does not mean that the agent is flipping any coins!

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...
- ... except for minimax and maximax, which have the nice property that it all collapses into one game tree

This is basically how you would model a human, except for their utility: their utility might be the same as yours (i.e. you try to help them, but they are depth 2 and noisy), or they might have a slightly different utility (like another person navigating in the office)

Dangerous Pessimism/Optimism

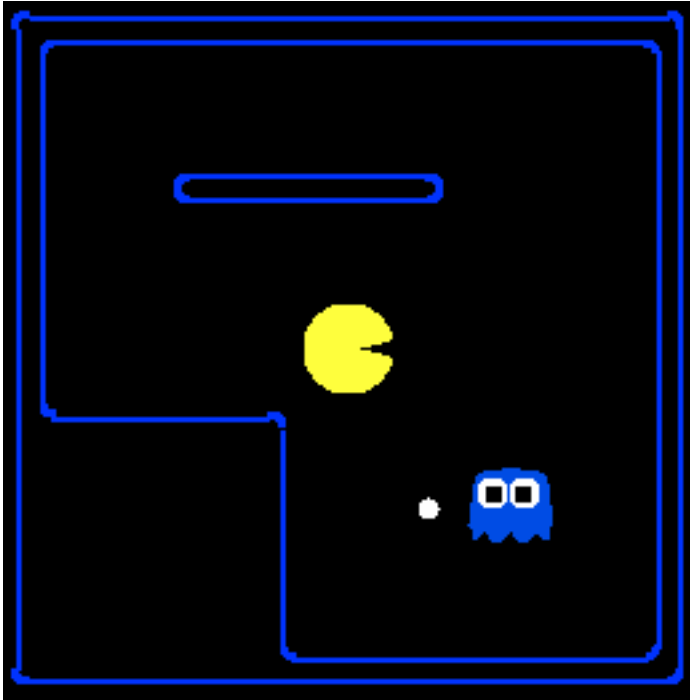
Assuming the worst case when it's not likely



Assuming chance when the world is adversarial



Assumptions vs. Reality

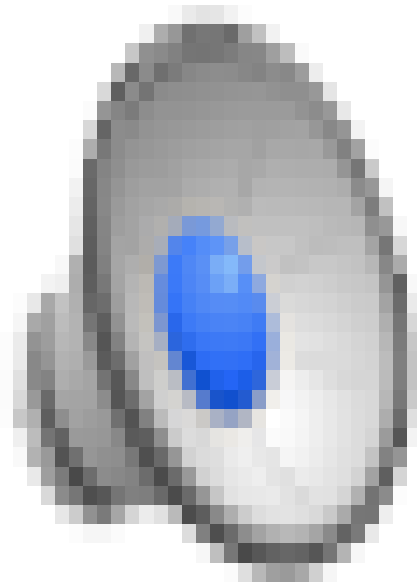


	Adversarial Ghost	Random Ghost
Minimax Pacman		
Expectimax Pacman		

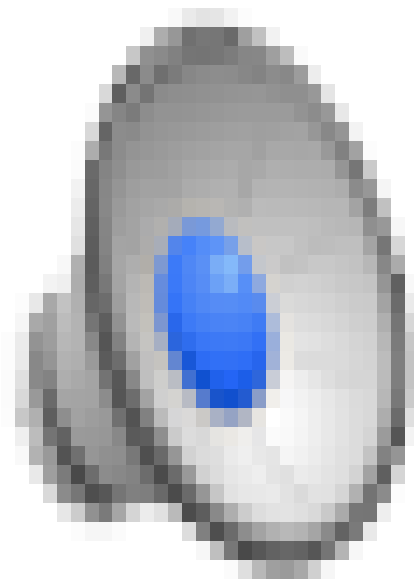
Results from playing 5 games

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

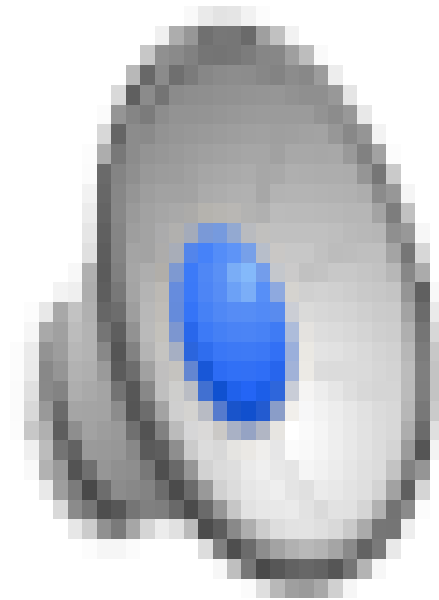
Video of Demo World Assumptions Random Ghost – Expectimax Pacman



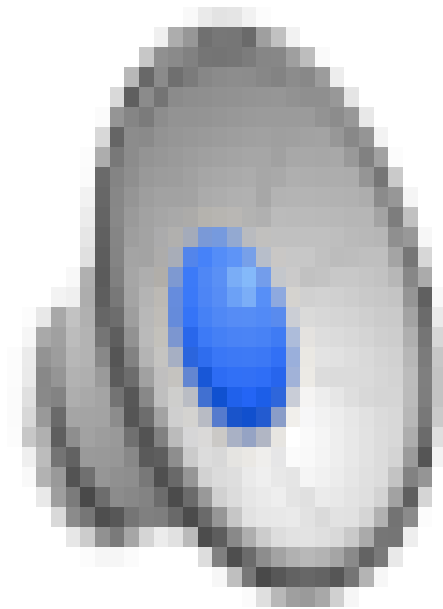
Video of Demo World Assumptions Adversarial Ghost – Minimax Pacman



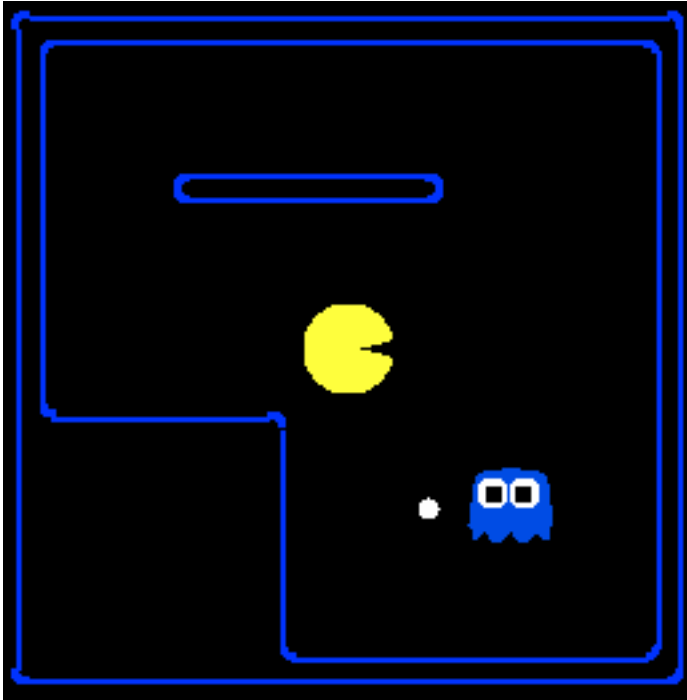
Video of Demo World Assumptions Adversarial Ghost – Expectimax Pacman



Video of Demo World Assumptions Random Ghost – Minimax Pacman



Assumptions vs. Reality



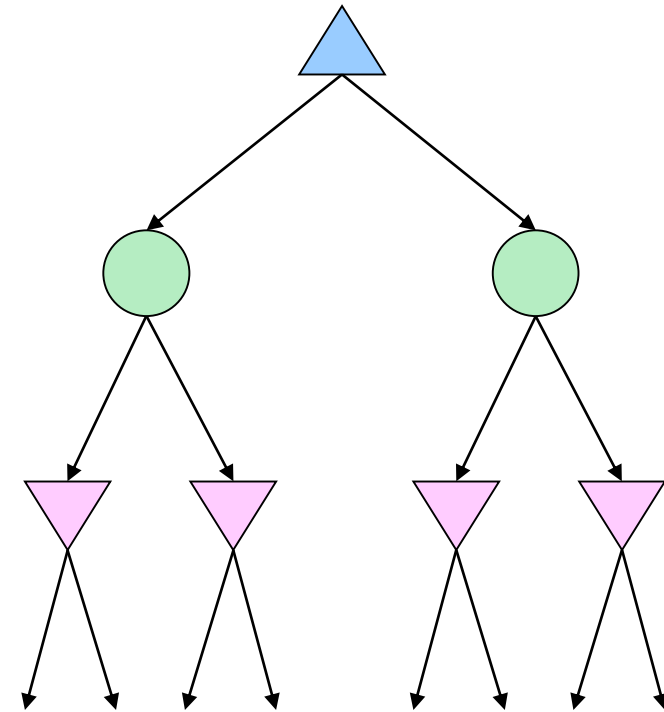
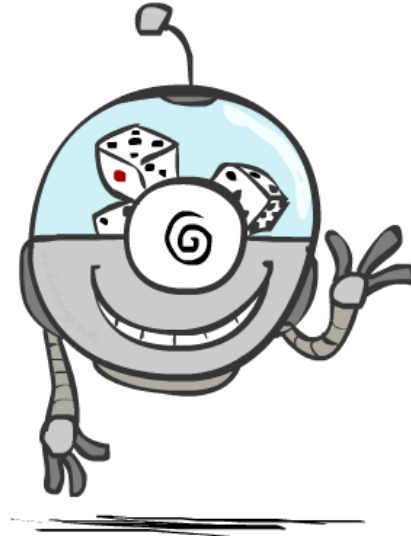
	Adversarial Ghost	Random Ghost
Minimax Pacman	Won 5/5 Avg. Score: 483	Won 5/5 Avg. Score: 493
Expectimax Pacman	Won 1/5 Avg. Score: -303	Won 5/5 Avg. Score: 503

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

Mixed Layer Types

- E.g. 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

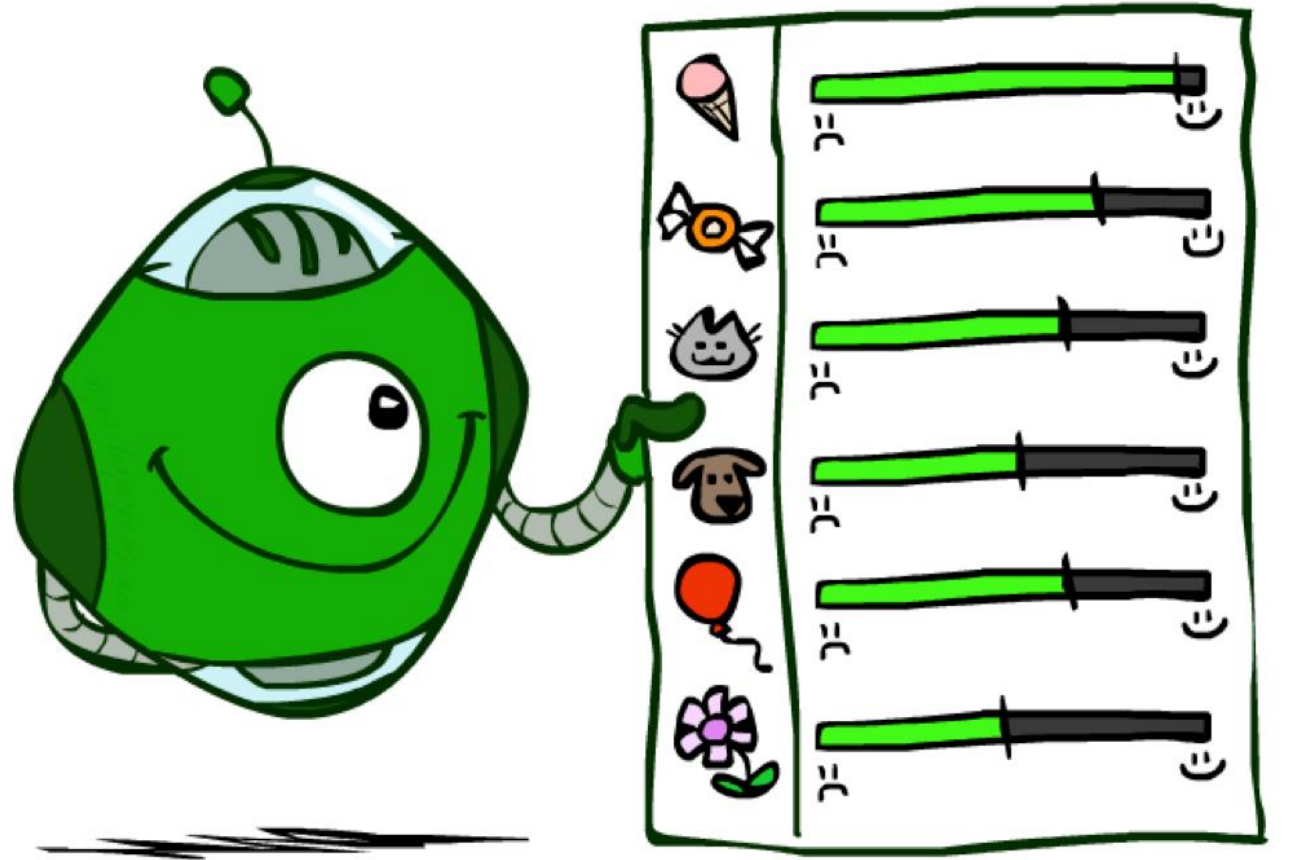


Example: Backgammon

- Dice rolls increase b : 21 possible rolls with 2 dice
 - Backgammon ≈ 20 legal moves
 - Depth 2 = $20 \times (21 \times 20)^3 = 1.2 \times 10^9$
- As depth increases, probability of reaching a given search node shrinks
 - So usefulness of search is diminished
 - So limiting depth is less damaging
 - But pruning is trickier...
- Historic AI: TDGammon uses depth-2 search + very good evaluation function + reinforcement learning:
[world-champion level play](#)
- 1st AI world champion in any game!



Utilities



We know how to compute with utilities given.
But how to model utilities in real applications?
And what properties of utilities are required?

Utilities

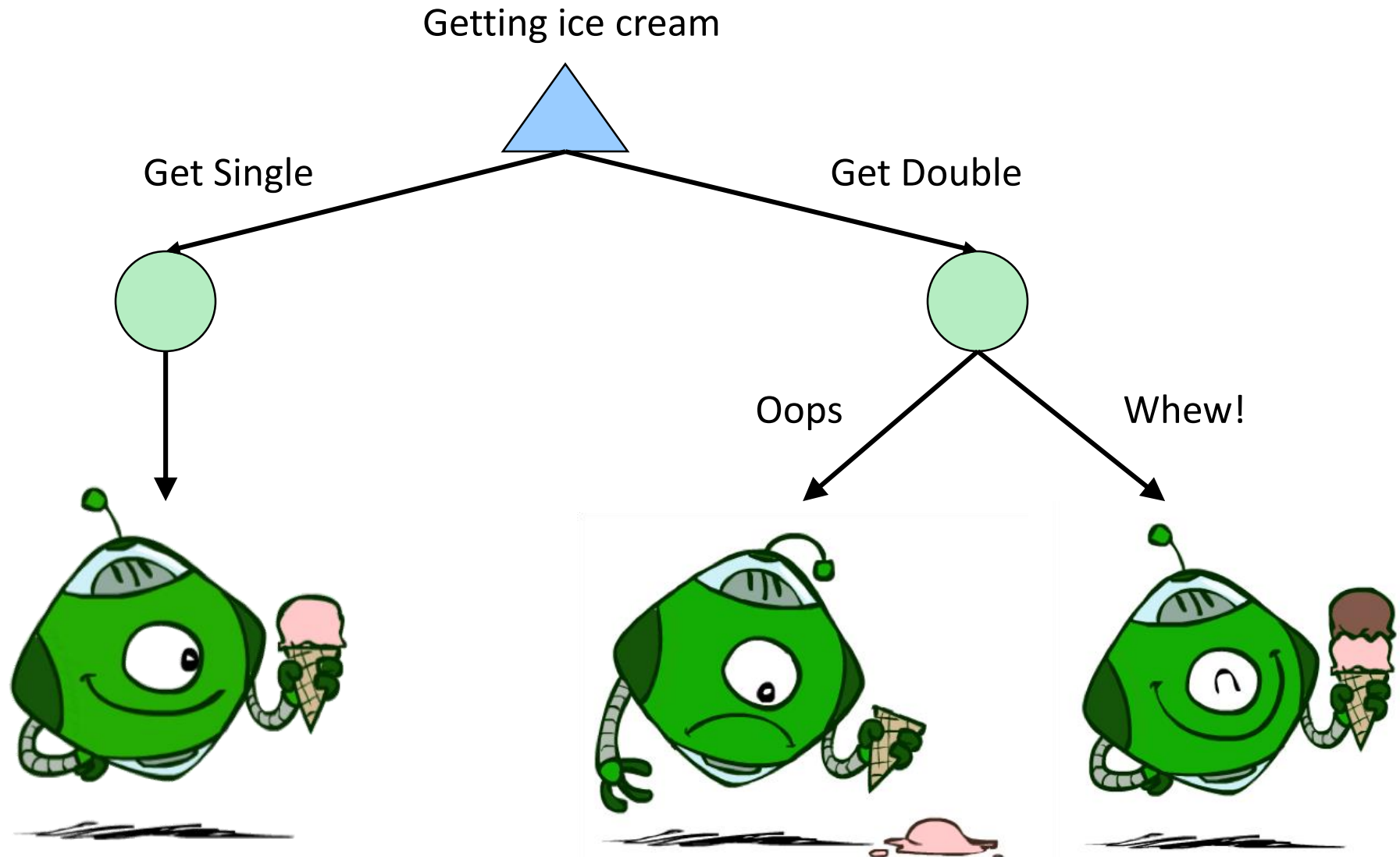
- Utilities are functions from outcomes (states of the world) to real numbers that describe an agent's preferences

- Where do utilities come from?
 - In a game, may be simple (+1/-1)
 - Utilities summarize the agent's goals
 - **Theorem**: any "rational" preferences can be summarized as a utility function

- We hard-wire utilities and let behaviors emerge
 - Why don't we let agents pick utilities?
 - Why don't we prescribe behaviors?



Utilities with Uncertain Outcomes

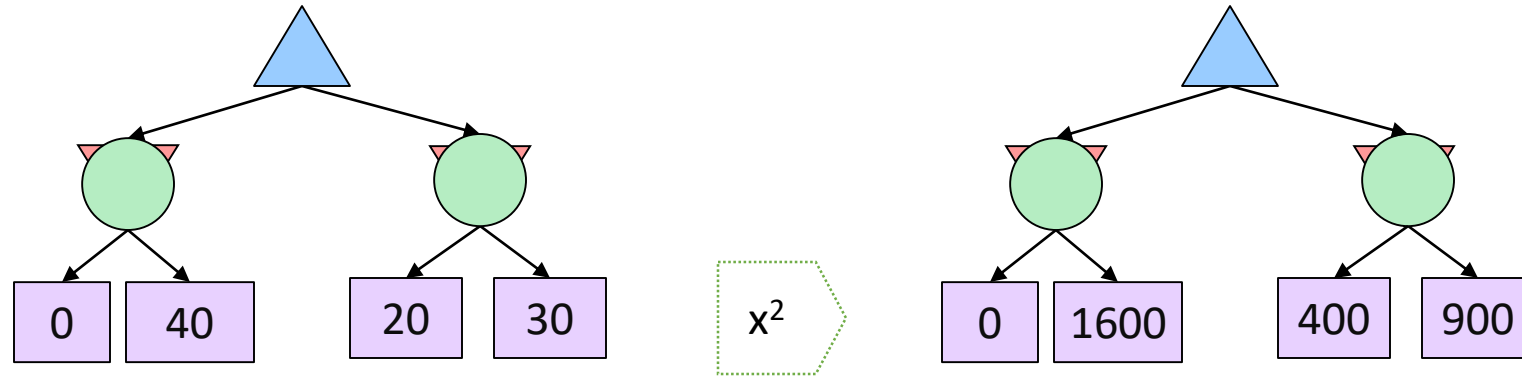


Maximum Expected Utility

- Why should we average utilities? Why not minimax?
- Principle of maximum expected utility:
 - A rational agent should choose the action that **maximizes its expected utility, given its knowledge**
- Questions:
 - Where do utilities come from?
 - How do we know such utilities even exist?
 - How do we know that averaging even makes sense?
 - What if our behavior (preferences) can't be described by utilities?



What Utilities to Use?



- For worst-case minimax reasoning, terminal function scale doesn't matter
 - We just want better states to have higher evaluations (get the ordering right)
 - We call this **insensitivity to monotonic transformations**
- For average-case expectimax reasoning, we need *magnitudes* to be meaningful

Preferences

- An agent must have preferences among:

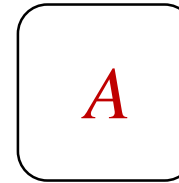
- Prizes: A , B , etc.
- Lotteries: situations with uncertain prizes

$$L = [p, A; (1 - p), B]$$

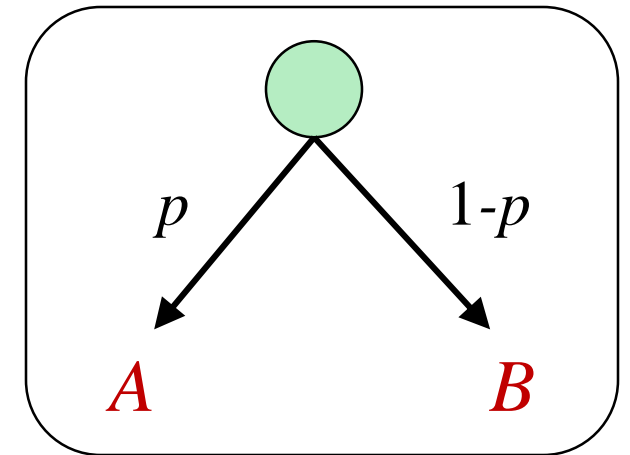
- Notation:

- Preference: $A \succ B$
- Indifference: $A \sim B$

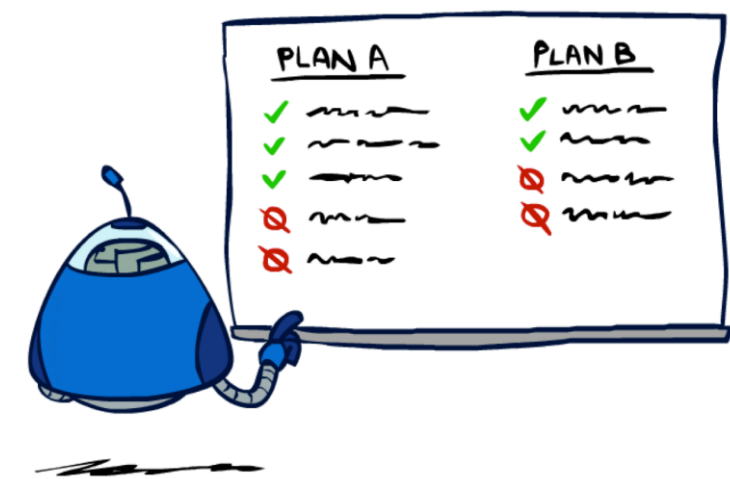
A Prize



A Lottery



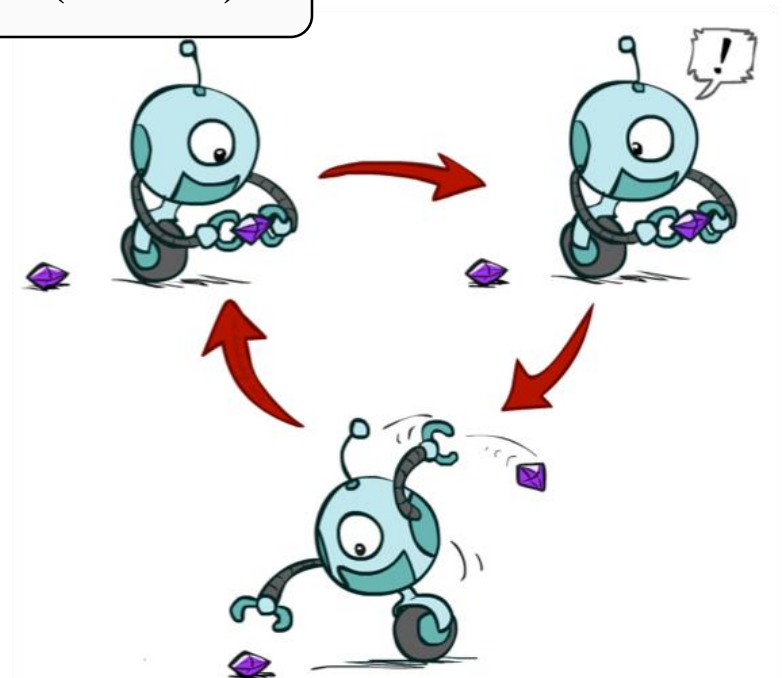
Rational Preferences



- We want some constraints on preferences before we call them rational, such as:

$$\text{Axiom of Transitivity: } (A \succ B) \wedge (B \succ C) \Rightarrow (A \succ C)$$

- For example: an agent with **intransitive preferences** can be induced to give away all of its money
 - If $B \succ C$, then an agent with C would pay (say) 1 cent to get B
 - If $A \succ B$, then an agent with B would pay (say) 1 cent to get A
 - If $C \succ A$, then an agent with A would pay (say) 1 cent to get C



Rational Preferences 2

- The **Axioms** of Rationality

Orderability

$$(A \succ B) \vee (B \succ A) \vee (A \sim B)$$

Transitivity

$$(A \succ B) \wedge (B \succ C) \Rightarrow (A \succ C)$$

Continuity

$$A \succ B \succ C \Rightarrow \exists p [p, A; 1 - p, C] \sim B$$

Substitutability

$$A \sim B \Rightarrow [p, A; 1 - p, C] \sim [p, B; 1 - p, C]$$

Monotonicity

$$A \succ B \Rightarrow$$

$$(p \geq q \Leftrightarrow [p, A; 1 - p, B] \succeq [q, A; 1 - q, B])$$



- **Theorem:** Rational preferences imply behavior describable as maximization of expected utility

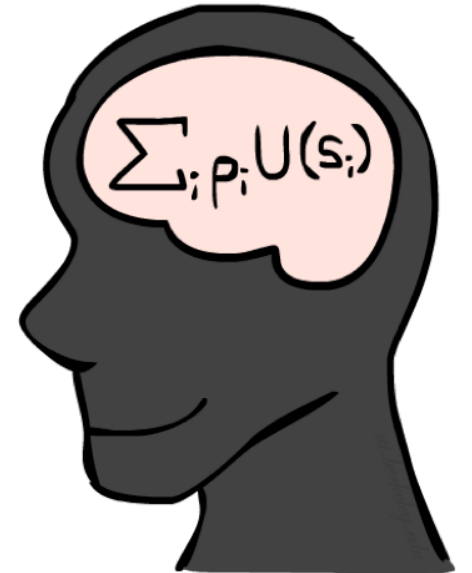
MEU Principle

- **Theorem** [Ramsey, 1931; von Neumann & Morgenstern, 1944]
 - Given any preferences satisfying these constraints, there exists a real-valued function U such that:

$$U(A) \geq U(B) \Leftrightarrow A \succeq B$$

$$U([p_1, S_1; \dots ; p_n, S_n]) = \sum_i p_i U(S_i)$$

- i.e. values assigned by U preserve preferences of both prizes and lotteries!
- **Maximum expected utility (MEU) principle:**
 - Choose the action that maximizes expected utility
 - Note: an agent can be entirely rational (consistent with MEU) without ever representing or manipulating utilities and probabilities
 - E.g., a lookup table for perfect tic-tac-toe, a reflex vacuum cleaner

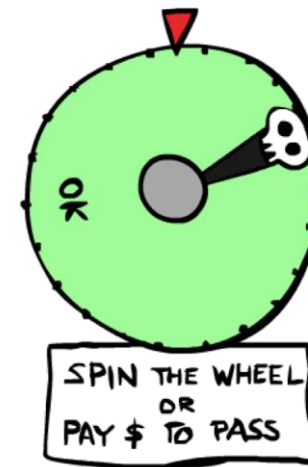


Human Utility Scales

- **Normalized utilities:** $u_+ = 1.0$, $u_- = 0.0$
- **Micromorts:** one-millionth chance of death, useful for paying to reduce product risks, etc.
- **QALYs:** quality-adjusted life years, useful for medical decisions involving substantial risk
- Note: behavior is invariant under positive linear transformation

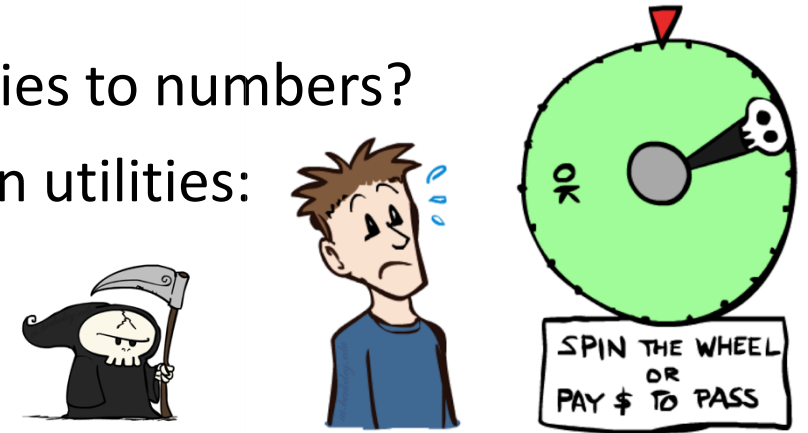
$$U'(x) = k_1 U(x) + k_2 \quad \text{where } k_1 > 0$$

- With deterministic prizes only (no lottery choices), only **ordinal utility** can be determined, i.e., total order on prizes



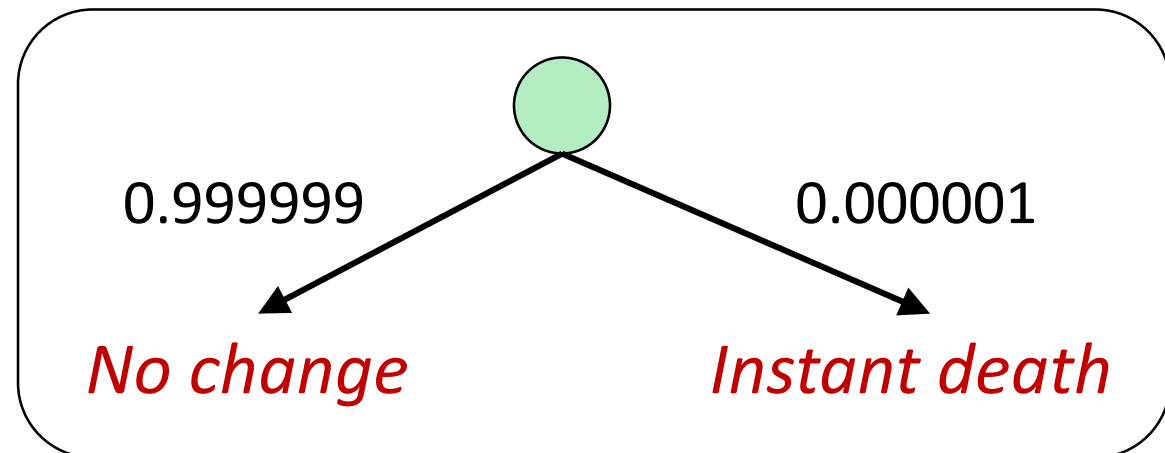
Human Utilities

- Once we have total order on utilities, how to map utilities to numbers?
- Standard approach to assessment (elicitation) of human utilities:
 - Compare a prize A to a **standard lottery** L_p between
 - “best possible prize” u_+ with probability p
 - “worst possible catastrophe” u_- with probability $1-p$
 - Adjust lottery probability p until indifference: $A \sim L_p$
 - Resulting p is a utility in $[0,1]$



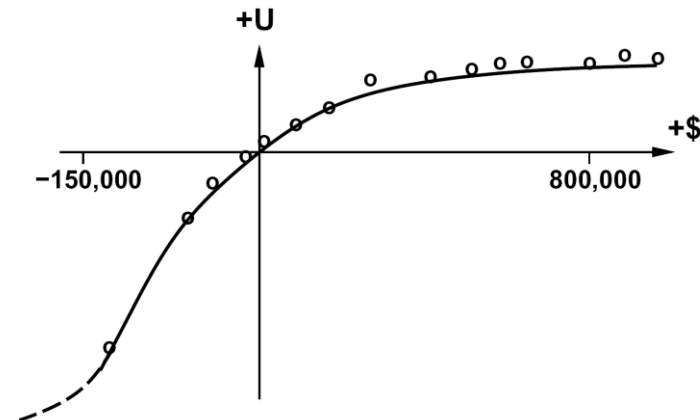
Pay \$30

~



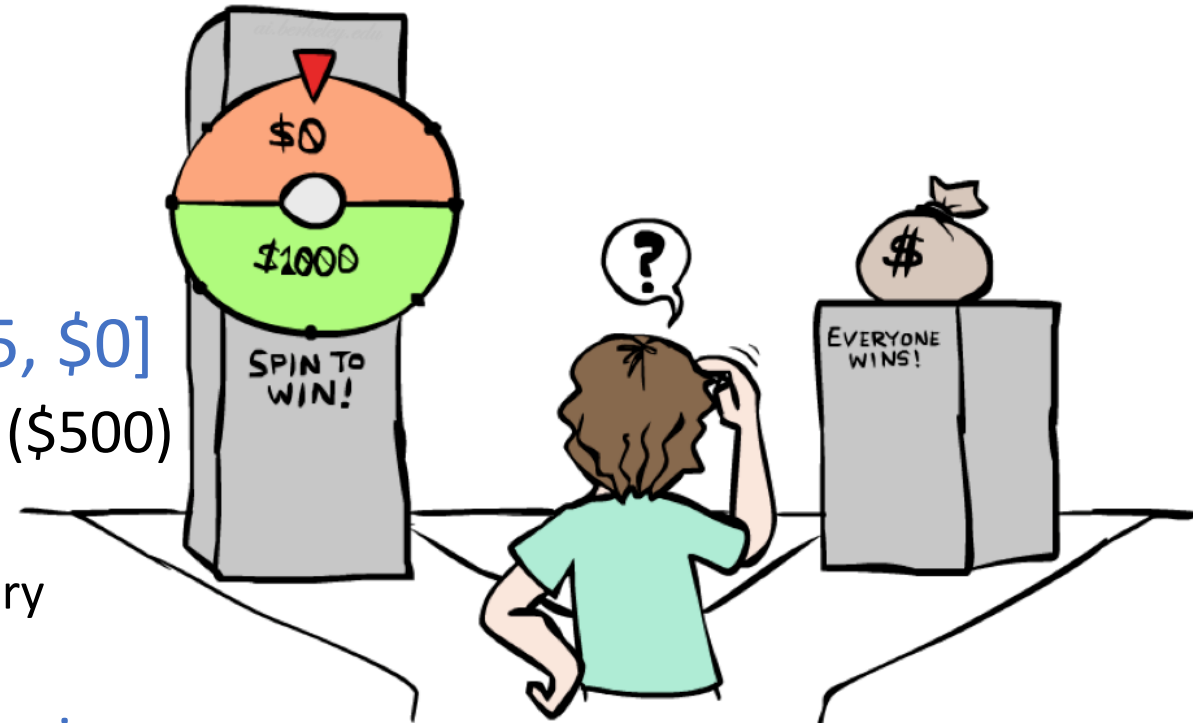
Human Utilities: Money

- Money does not behave as a utility function, but we can talk about **the utility of having money** (or being in debt)
- Given a lottery $L = [p, \$X; (1-p), \$Y]$
 - The **expected monetary value** $EMV(L)$ is $p*X + (1-p)*Y$
 - $U(L) = p*U(\$X) + (1-p)*U(\$Y)$
 - Typically, $U(L) < U(EMV(L))$
 - In this sense, people are **risk-averse**
 - When deep in debt, people are **risk-prone**
 - When $X, Y < 0$, $U(L) > U(EMV(L))$



Example: Insurance

- Consider the lottery $[0.5, \$1000; 0.5, \$0]$
 - What is its **expected monetary value**? (\$500)
 - What is its **certainty equivalent**?
 - Monetary value acceptable in lieu of lottery
 - \$400 for most people
 - Difference of \$100 is the **insurance premium**
 - There's an insurance industry because people will pay to reduce their risk
 - If everyone were risk-neutral, no insurance needed!
 - It's win-win: you'd rather have the \$400 and the insurance company would rather have the lottery (their utility curve is flat and they have many lotteries)



Example: Human Rationality?

- Famous example of Allais (1953)

- A: [0.8, \$4k; 0.2, \$0] ←
- B: [1.0, \$3k; 0.0, \$0]
- C: [0.2, \$4k; 0.8, \$0]
- D: [0.25, \$3k; 0.75, \$0]

- Most people prefer $B > A$, $C > D$

- But if $U(\$0) = 0$, then

- $B > A \Rightarrow U(\$3k) > 0.8 U(\$4k)$
- $C > D \Rightarrow 0.8 U(\$4k) > U(\$3k)$



Summary

- Types of games
- Adversarial Game Trees: Minimax search
- Resource Limits I: Alpha-Beta Pruning
- Resource Limits II: Depth-limited search
 - Evaluation functions
 - Iterative deepening
- Expectimax Search
 - Pruning ✘/Depth-limited ✓
 - Adversarial/Random environments vs. Minimax/Expectimax agents
 - Probabilities/Simulations
- Utilities
 - Rational preferences/MEU principle/Human utilities/Money

Shuai Li

<https://shuaili8.github.io>

Questions?