Real-time Planning as Decision-making Under Uncertainty

Andrew Mitchell

Advisor: Wheeler Ruml

University of New Hampshire

What is Planning?

my work: improve real-time planning using ideas from decision-making!

What is Planning?

my work: improve real-time planning using ideas from decision-making!

planning is a way of finding a sequence of actions that accomplish some objective

What is Planning?

my work: improve real-time planning using ideas from decision-making!

planning is a way of finding a sequence of actions that accomplish some objective

one method of planning: heuristic search!

What is Planning?

my work: improve real-time planning using ideas from decision-making!

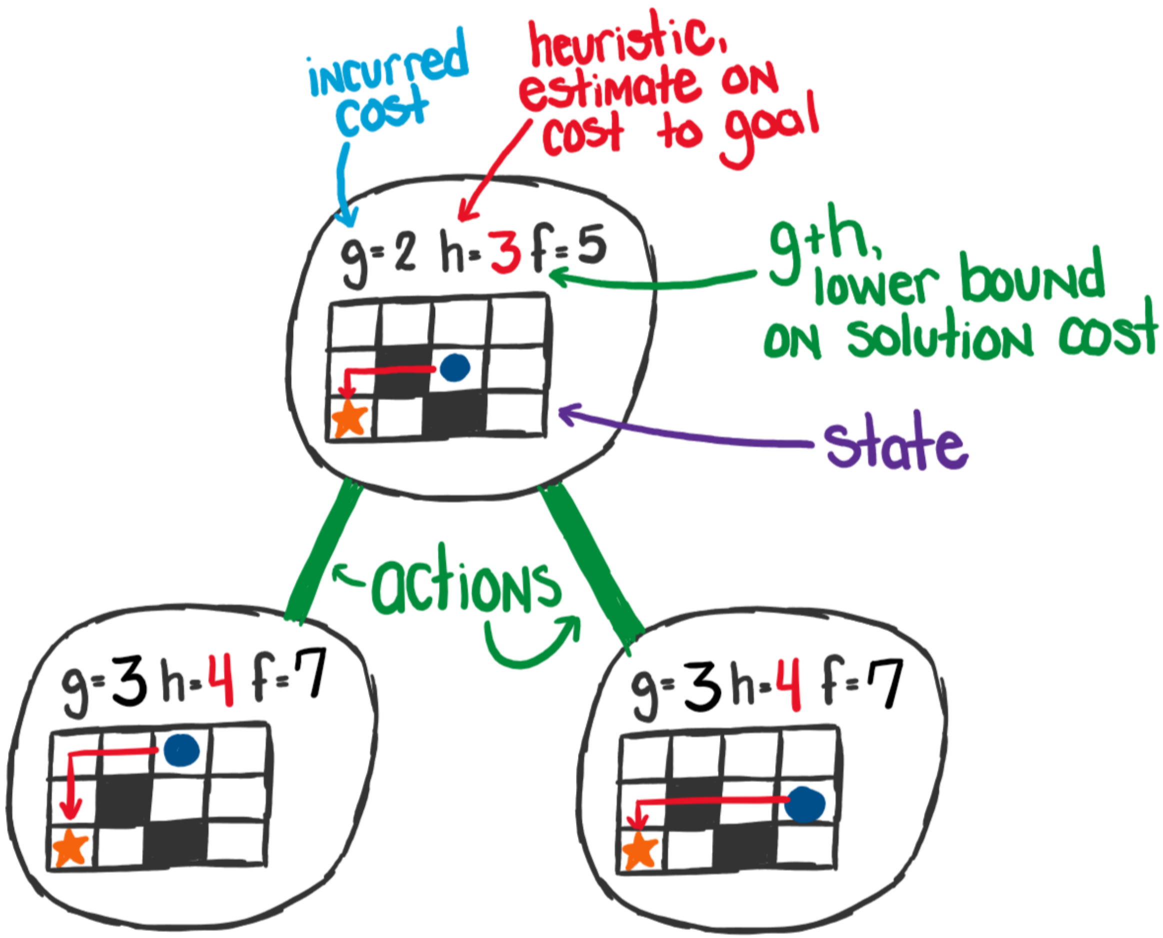
planning is a way of finding a sequence of actions that accomplish some objective

one method of planning: heuristic search!

heuristic search:

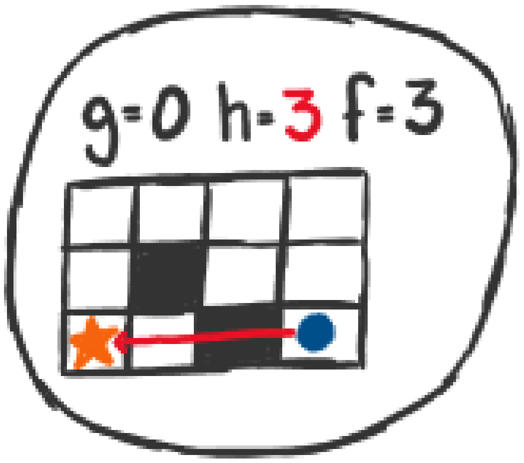
agent tasked with reaching a specific state

accomplished by searching graph of states + actions heuristic search associates costs with states, used to guide search



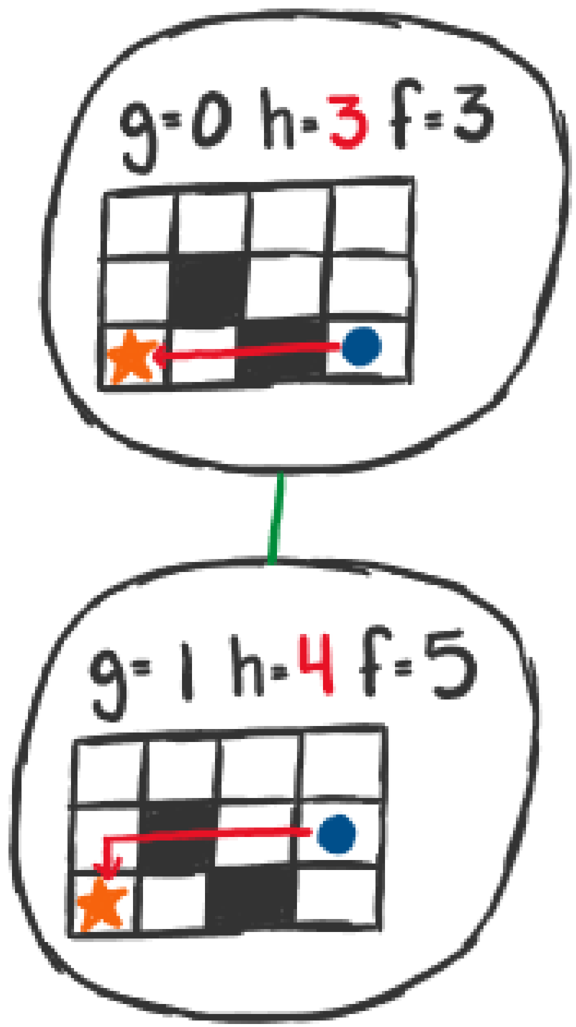
Heuristic Search Example: A*

want blue dot to be where star is



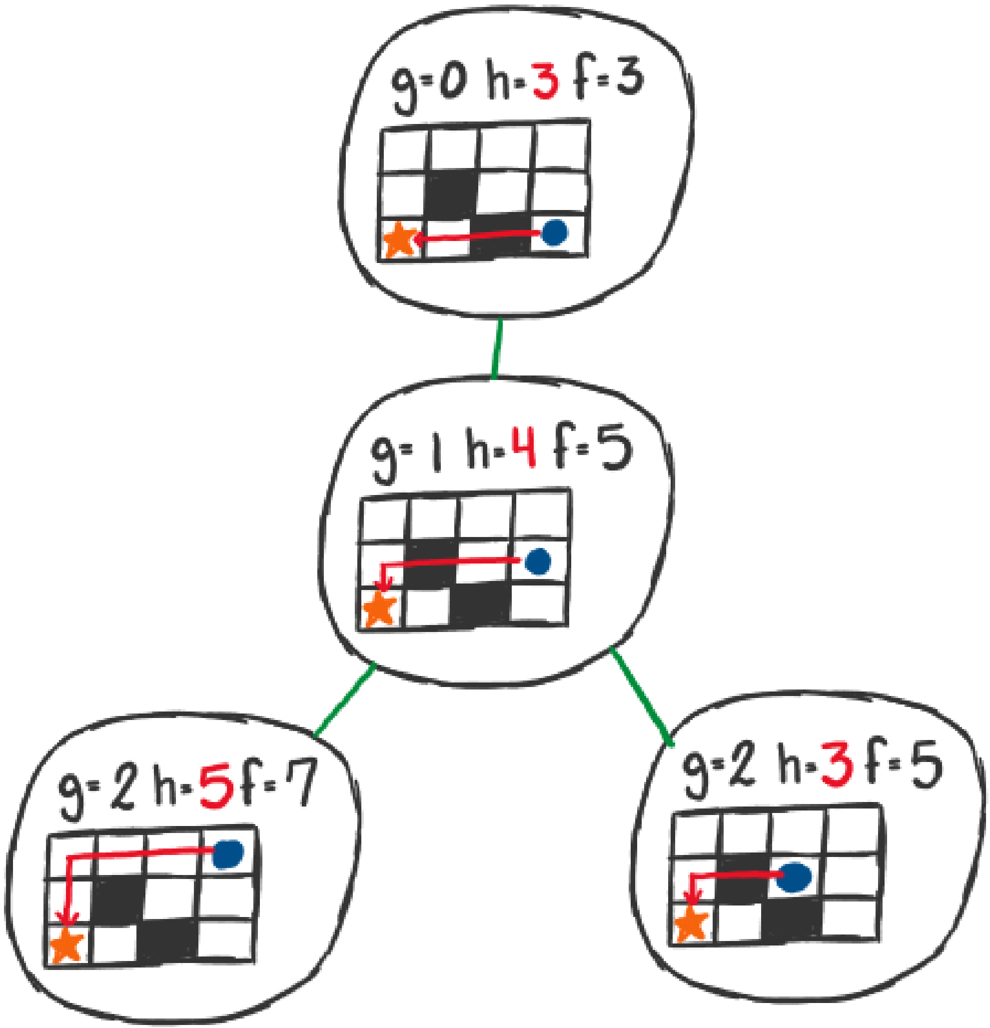
Heuristic Search Example: A*

want blue dot to be where star is



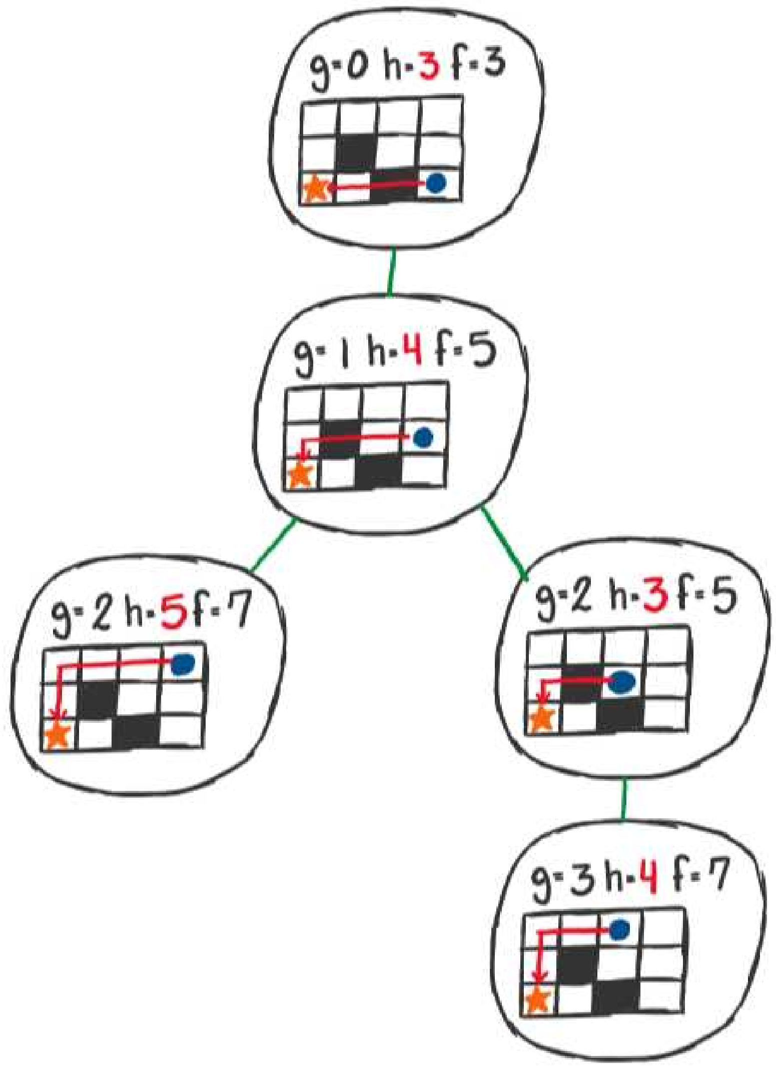
Heuristic Search Example: A*

want blue dot to be where star is



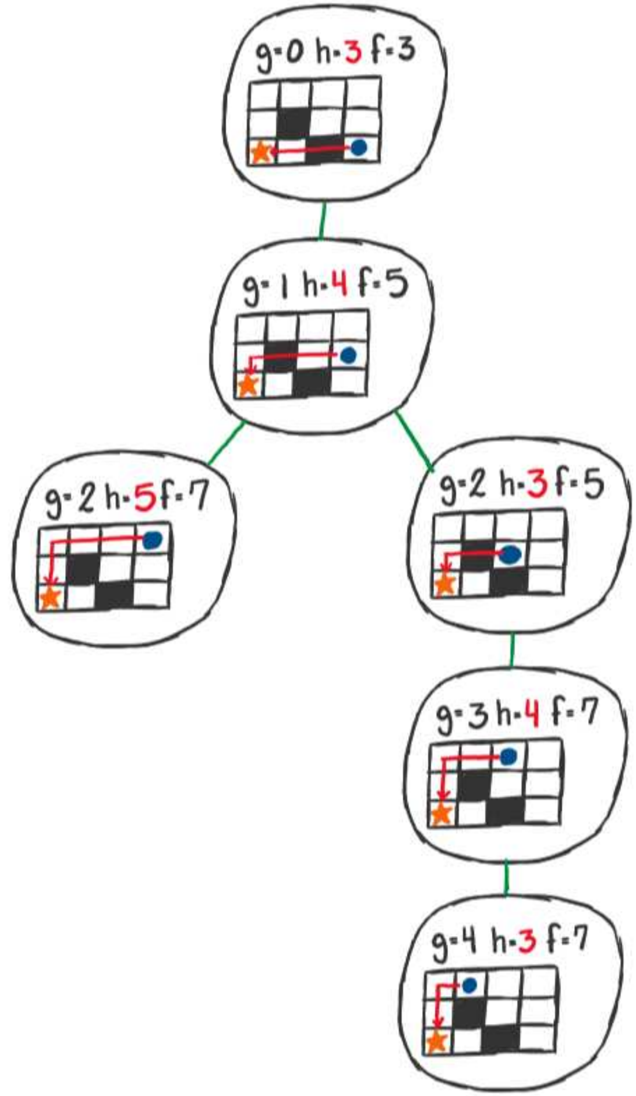
Heuristic Search Example: A*

want blue dot to be where star is

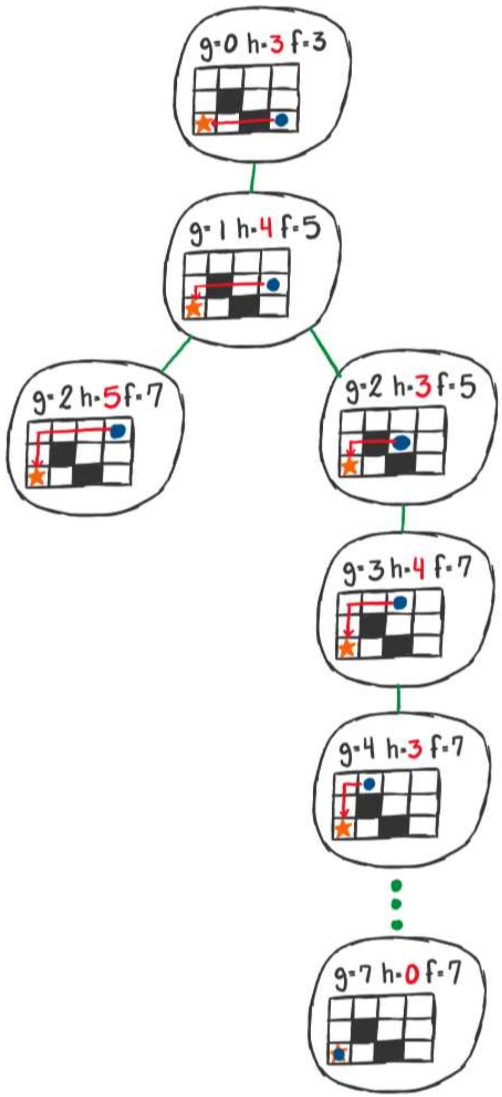


Heuristic Search Example: A*

want blue dot to be where star is



want blue dot to be where star is



Examples of Heuristic Search

A*: expands nodes with minimal f value
returns optimal path
optimal search can take too long!

Examples of Heuristic Search

A*: expands nodes with minimal f value
returns optimal path
optimal search can take too long!

alternatives to optimal search?

Examples of Heuristic Search

A*: expands nodes with minimal f value
returns optimal path
optimal search can take too long!

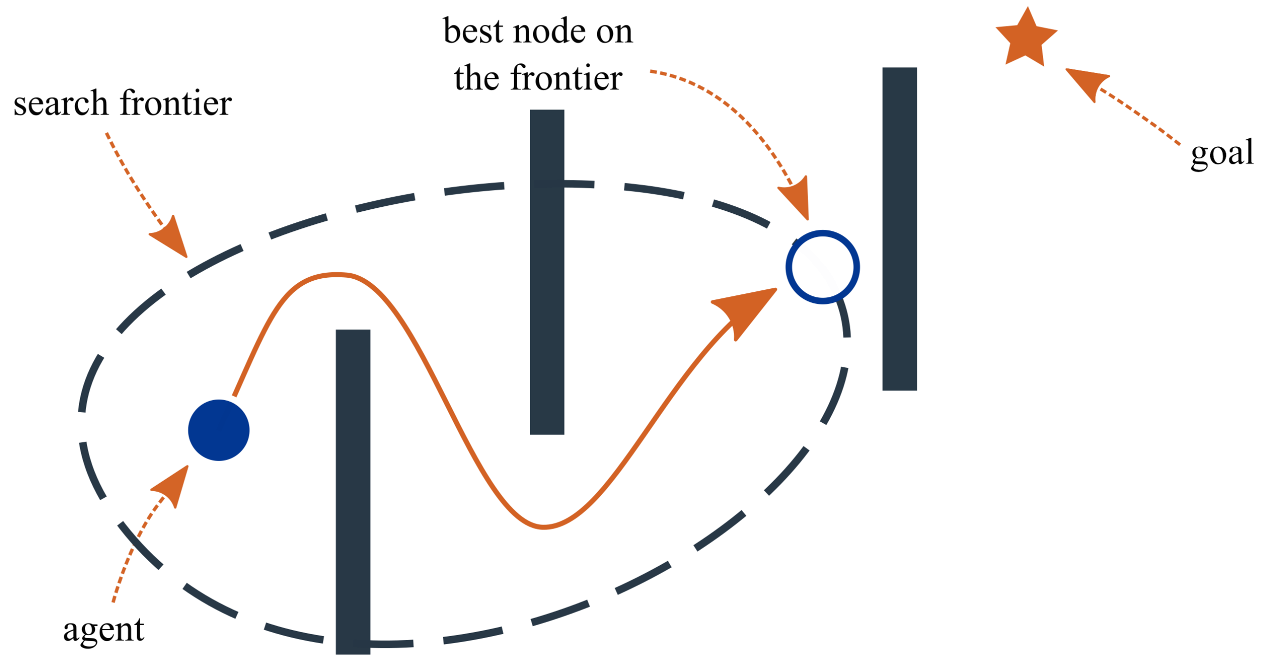
alternatives to optimal search?
real-time heuristic search!

Real-time Search Example

(Image credit: Bence Cserna)

agent performs search for a limited time

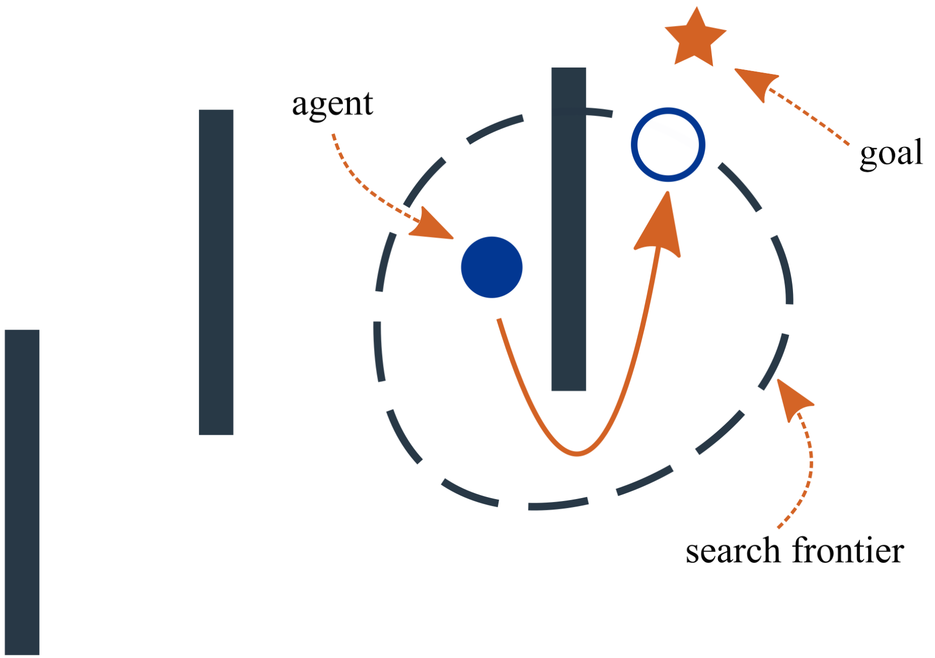
Real-time Search Example



(Image credit: Bence Cserna)

agent commits to a path to frontier and executes

Real-time Search Example



(Image credit: Bence Cserna)

agent continues interleaving search and path execution

How Real-time Planning Works

can be broken down into three phases...

1. Expansion Phase:

   expands nodes to explore the search space

How Real-time Planning Works

can be broken down into three phases...

1. Expansion Phase:
   expands nodes to explore the search space
2. Decision-making Phase:
   amasses information on search frontier (backup rules)
   uses information to select an action to execute

How Real-time Planning Works

can be broken down into three phases...

1. Expansion Phase:
   expands nodes to explore the search space
2. Decision-making Phase:
   amasses information on search frontier (backup rules)
   uses information to select an action to execute
3. Learning Phase:
   learns heuristic values

How Real-time Planning Works

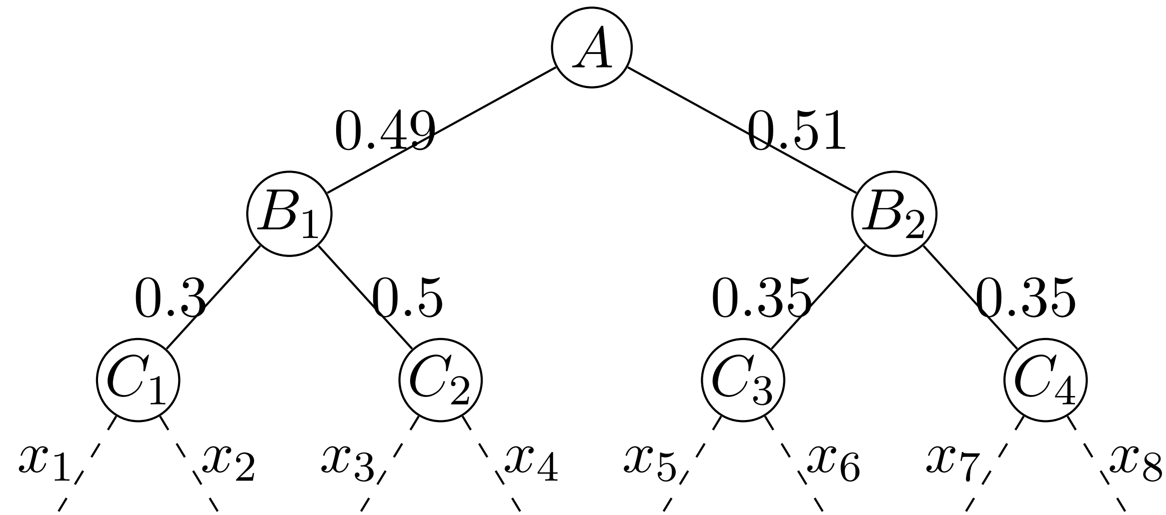
can be broken down into three phases...

1. Expansion Phase:
   expands nodes to explore the search space
2. Decision-making Phase:
   amasses information on search frontier (backup rules)
   uses information to select an action to execute
3. Learning Phase:
   learns heuristic values

Will focus on the first two stages!

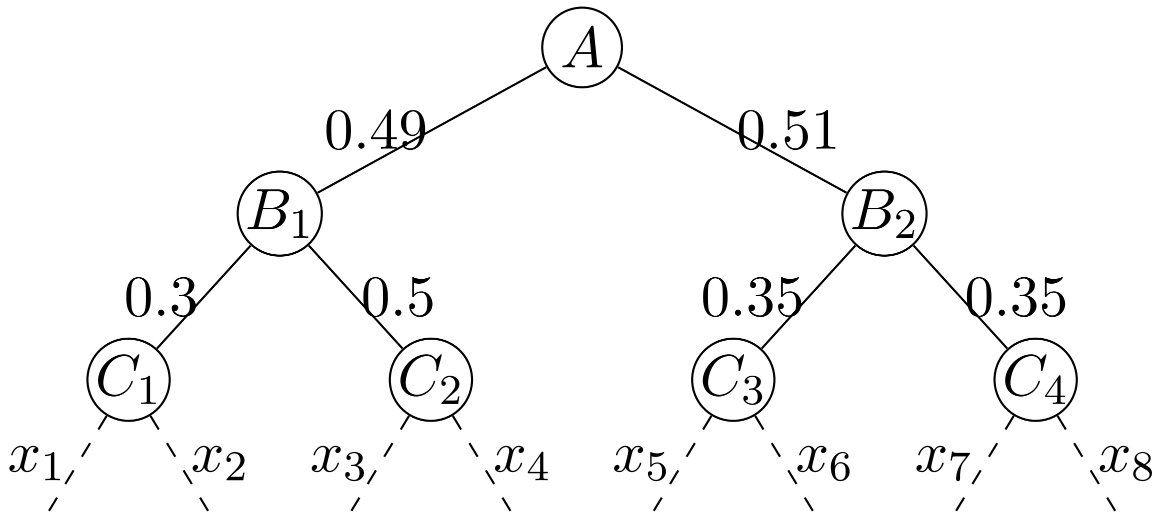
Overview

- Introduction
- Backup Rules
  - experiments
- Expansion Policies
  - experiments
- Conclusions



Should an agent at A move to B_1 or B_2?
(x_i are unknown but independent and identically distributed)

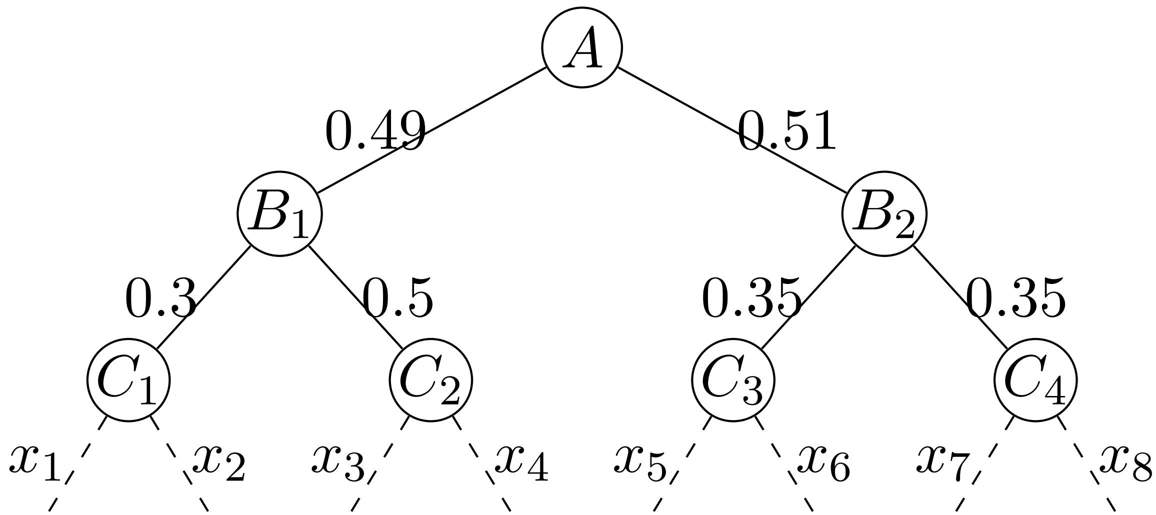
Online Decision-making



Should an agent at A move to B_1 or B_2?
(x_i are unknown but independent and identically distributed)

$f = g + h = g + 0$ is lower bound on optimal plan cost

Online Decision-making

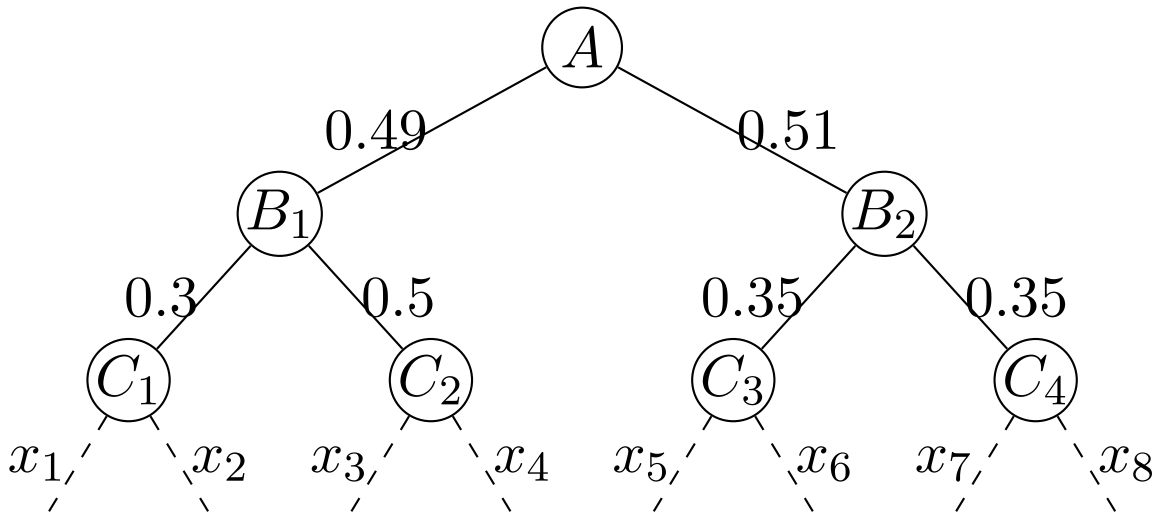


Should an agent at A move to B_1 or B_2?
(x_i are unknown but independent and identically distributed)

$f = g + h = g + 0$ is lower bound on optimal plan cost

f is not the answer: need statistical perspective

Online Decision-making



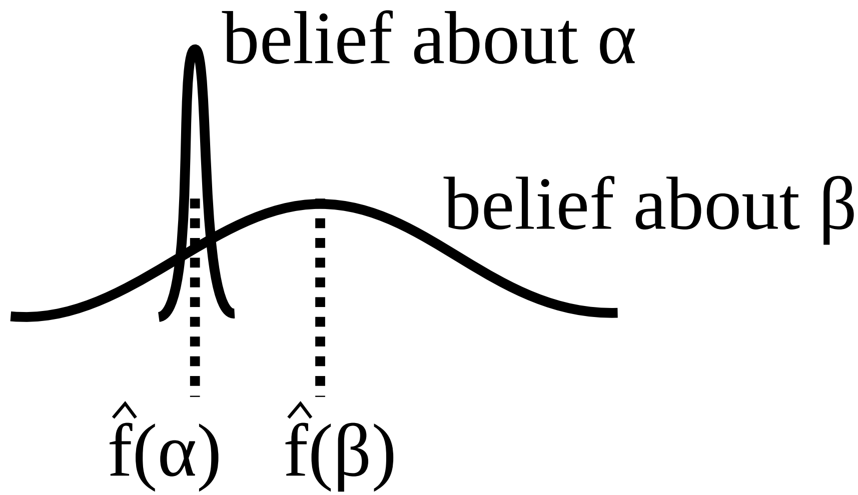
Should an agent at A move to B_1 or B_2?
(x_i are unknown but independent and identically distributed)

$f = g + h = g + 0$ is lower bound on optimal plan cost

f is not the answer: need statistical perspective

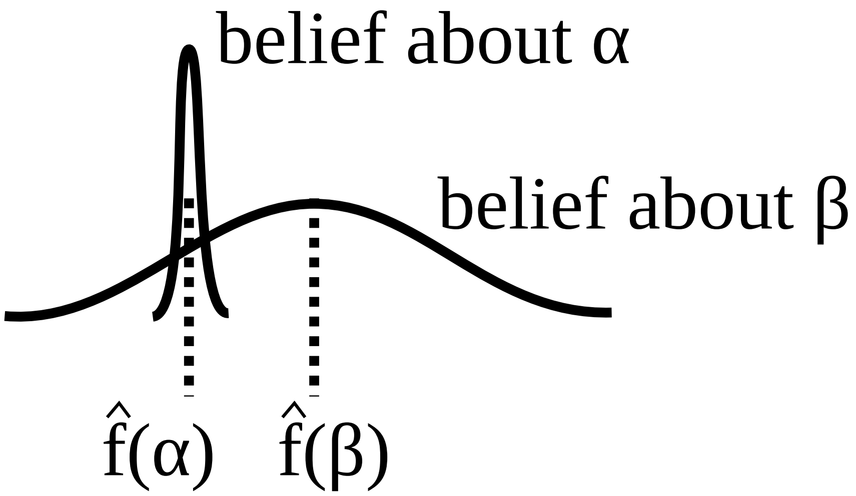
decision theory gives us principle of rationality:
should minimize expected value!

Which Nodes to Expand?



($\hat{f}$ is expected value)
($\hat{f} = f + \epsilon * d$)

Should an agent expand nodes under α or β?



($\hat{f}$ is expected value)
($\hat{f} = f + \epsilon * d$)

Should an agent expand nodes under α or β?

$\hat{f}$ is not the answer: what to do?

The Two Central Questions in Real-time Planning

1. Which action to select?

   minimum $\hat{f}$ by principle of rationality

The Two Central Questions in Real-time Planning

1. Which action to select?
   minimum $\hat{f}$ by principle of rationality
2. How to backup from frontier?
   minimin optimal for deterministic (A*)
   Bellman optimal for stochastic (VI)
   what about online?

The Two Central Questions in Real-time Planning

1. Which action to select?
   minimum $\hat{f}$ by principle of rationality
2. How to backup from frontier?
   minimin optimal for deterministic (A*)
   Bellman optimal for stochastic (VI)
   what about online?
3. Which nodes to expand?
   minimum f optimal for optimal search (A*)
   what about online?

The Two Central Questions in Real-time Planning

1. Which action to select?
   minimum $\hat{f}$ by principle of rationality
2. How to backup from frontier?
   minimin optimal for deterministic (A*)
   Bellman optimal for stochastic (VI)
   what about online?
3. Which nodes to expand?
   minimum f optimal for optimal search (A*)
   what about online?

this work: a practical investigation of the two questions from the perspective of decision-making under uncertainty.

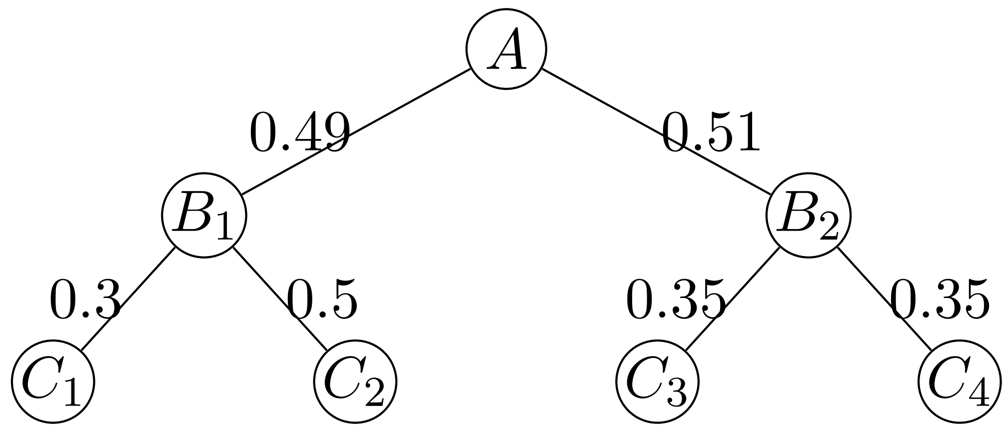
Backup Rules

Real-time Planning as Decision-making Under Uncertainty

heuristics are inaccurate/uncertain estimates on cost to goal
- due to unexplored state space
- true heuristic could be much higher!

view heuristics as distribution over potential values
- distributions centered about expected value

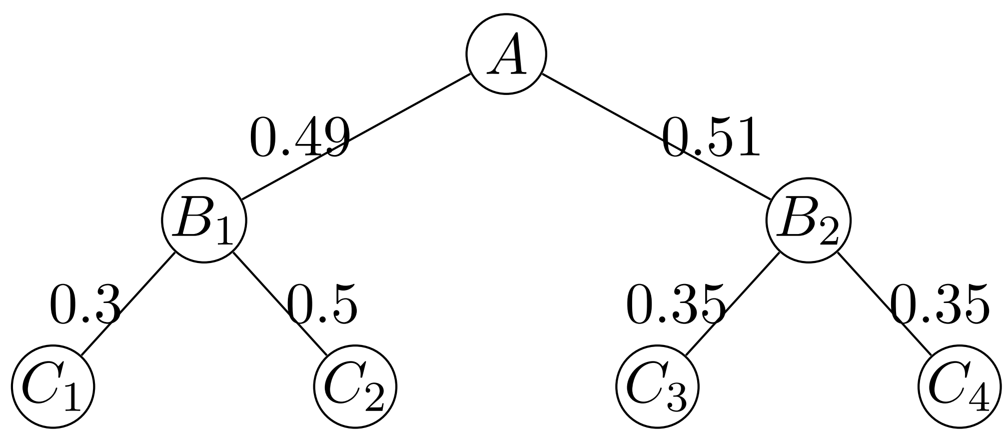
assume most accurate heuristics on the frontier



heuristics are inaccurate/uncertain estimates on cost to goal
- due to unexplored state space
- true heuristic could be much higher!

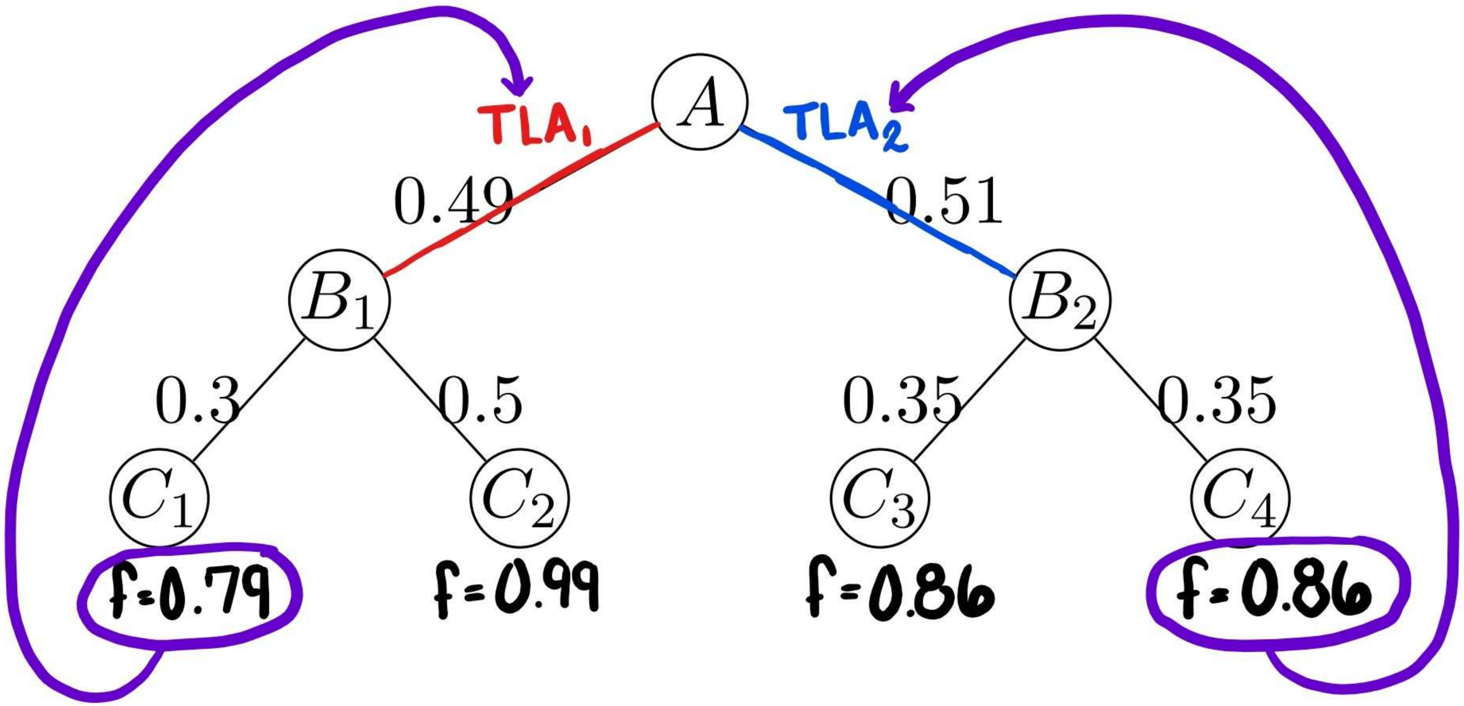
view heuristics as distribution over potential values
- distributions centered about expected value

assume most accurate heuristics on the frontier



how to form beliefs?
how to gather information from the search frontier?

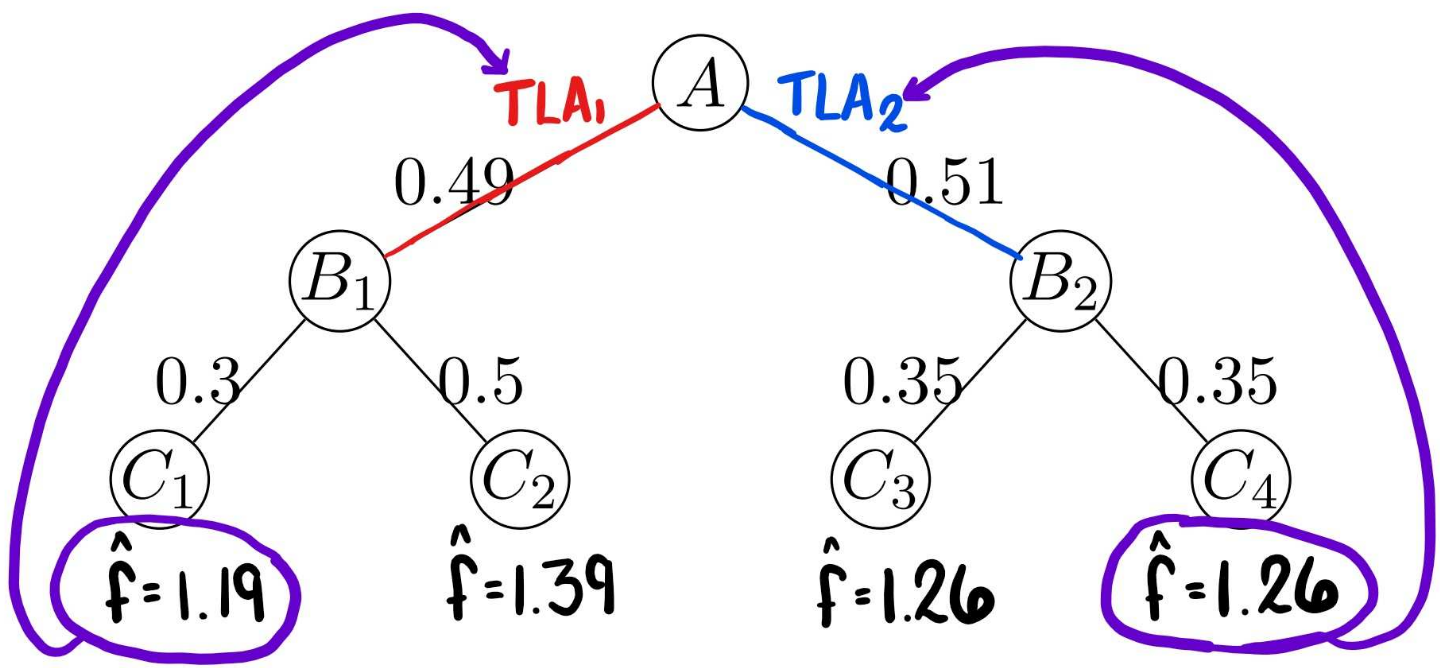
Backup Rules (1/5): Minimin



Minimin:

parent $\leftarrow$ minimum f among successors

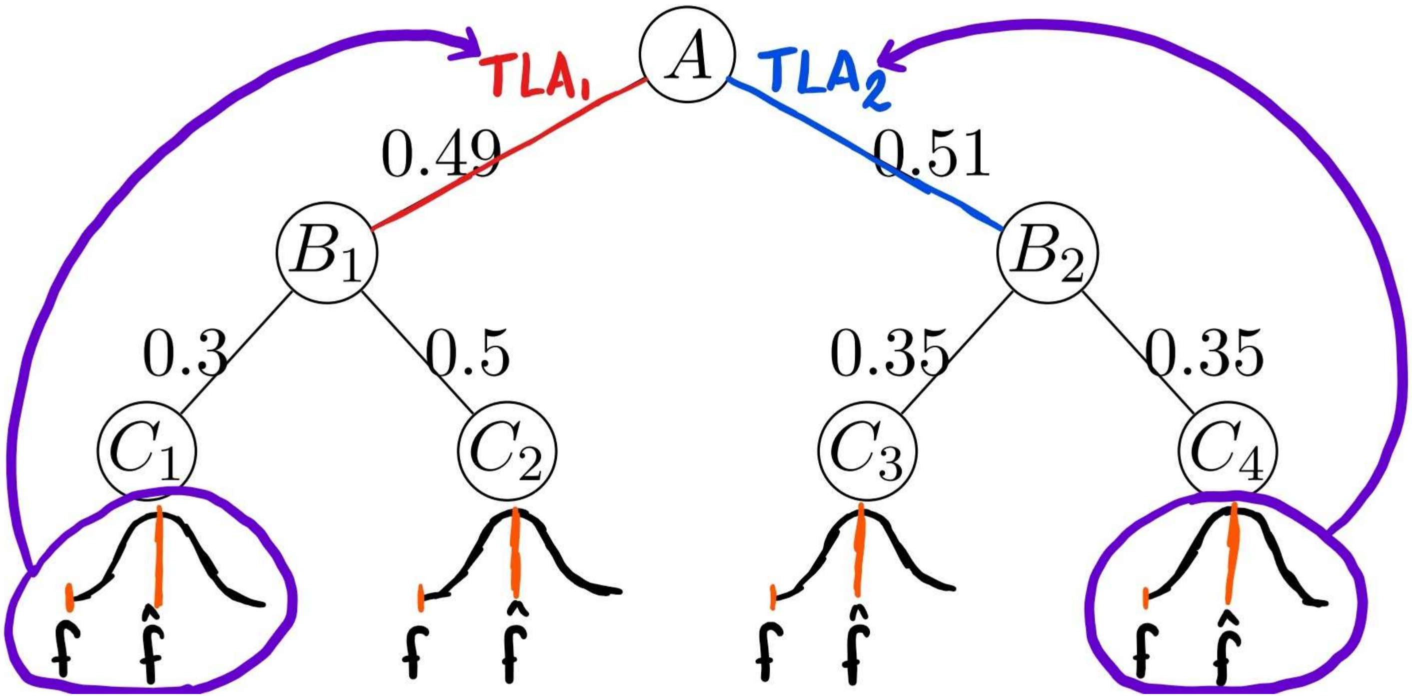
lower bound: not suitable for rational action selection



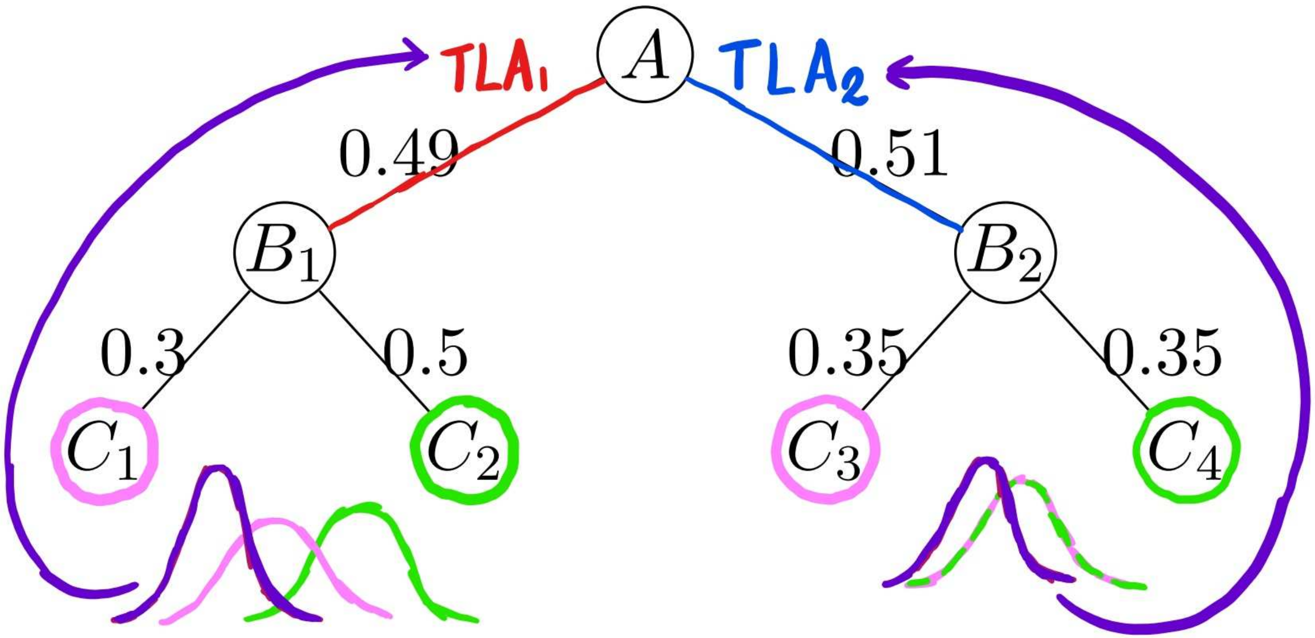
Bellman:

parent $\leftarrow$ minimum $\hat{f}$ among successors

only conveys a scalar value

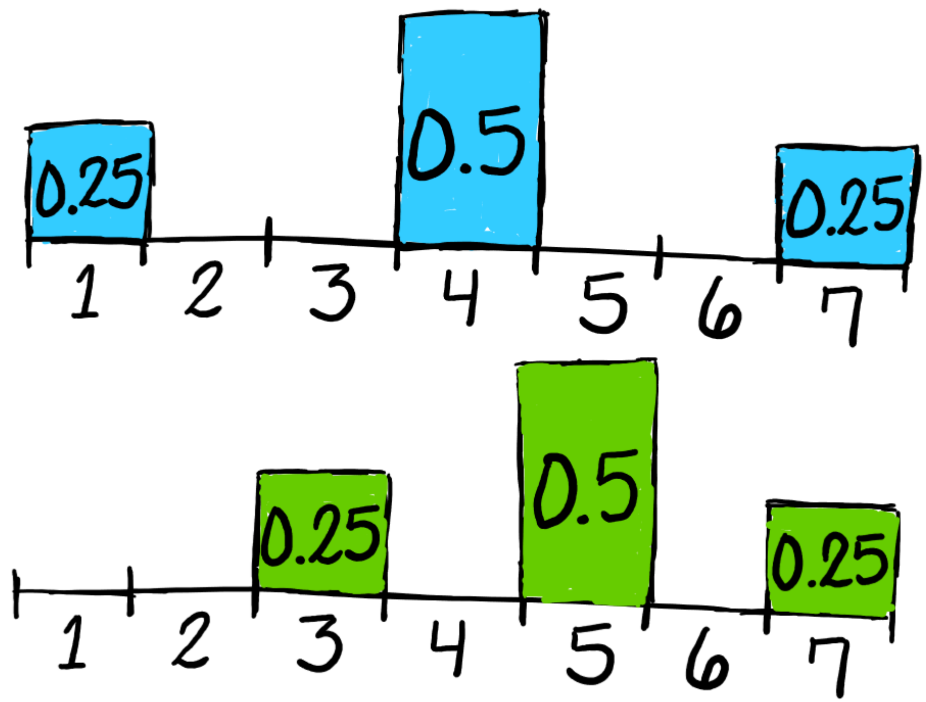


Nancy (new!):
parent $\leftarrow$ belief with minimum $\hat{f}$ among successors
conveys an entire belief distribution



Cserna (new!):
parent $\leftarrow$ distribution of minimum value of successors
assumes that true values of successors will be revealed

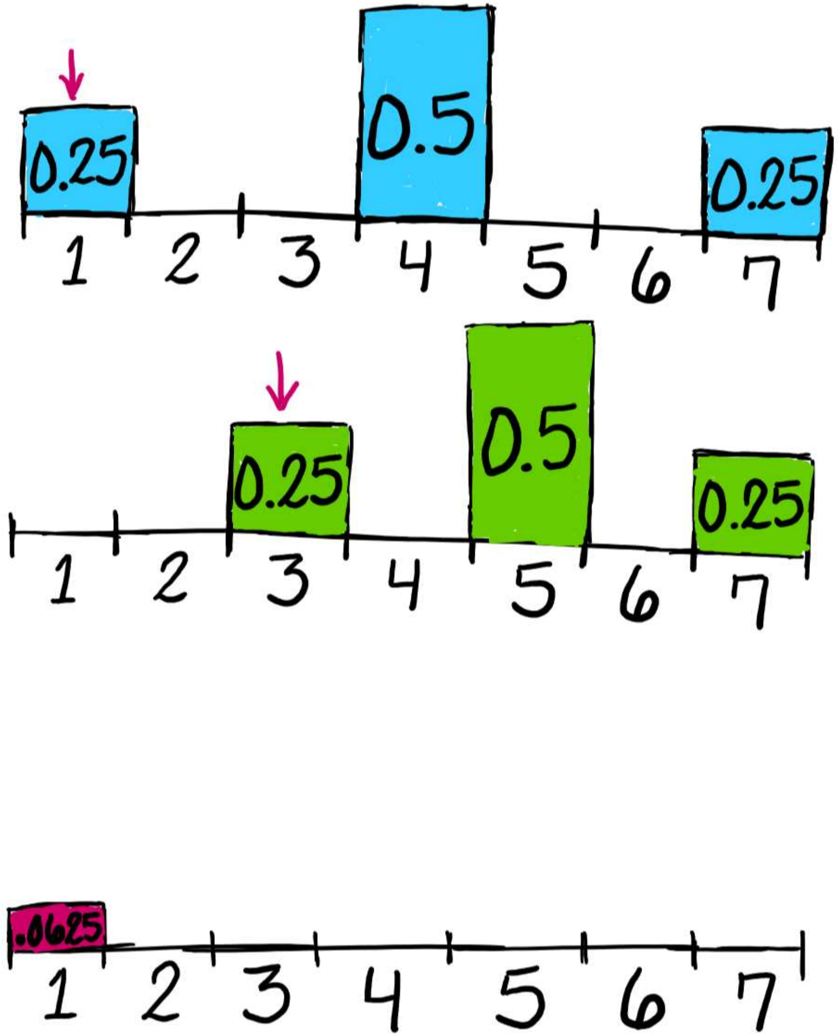
two beliefs on the frontier



how to obtain belief at TLA?

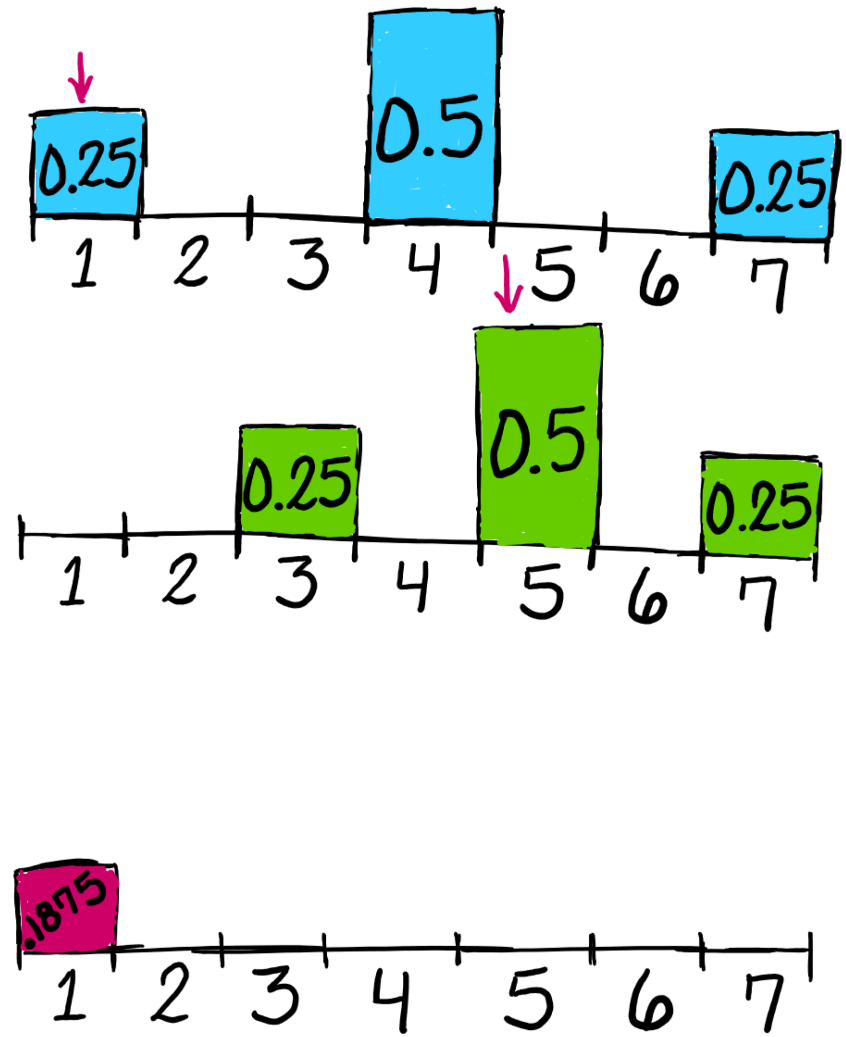
Backup Rules (4/5): Cserna Example

multiply probabilities and add to minimum value



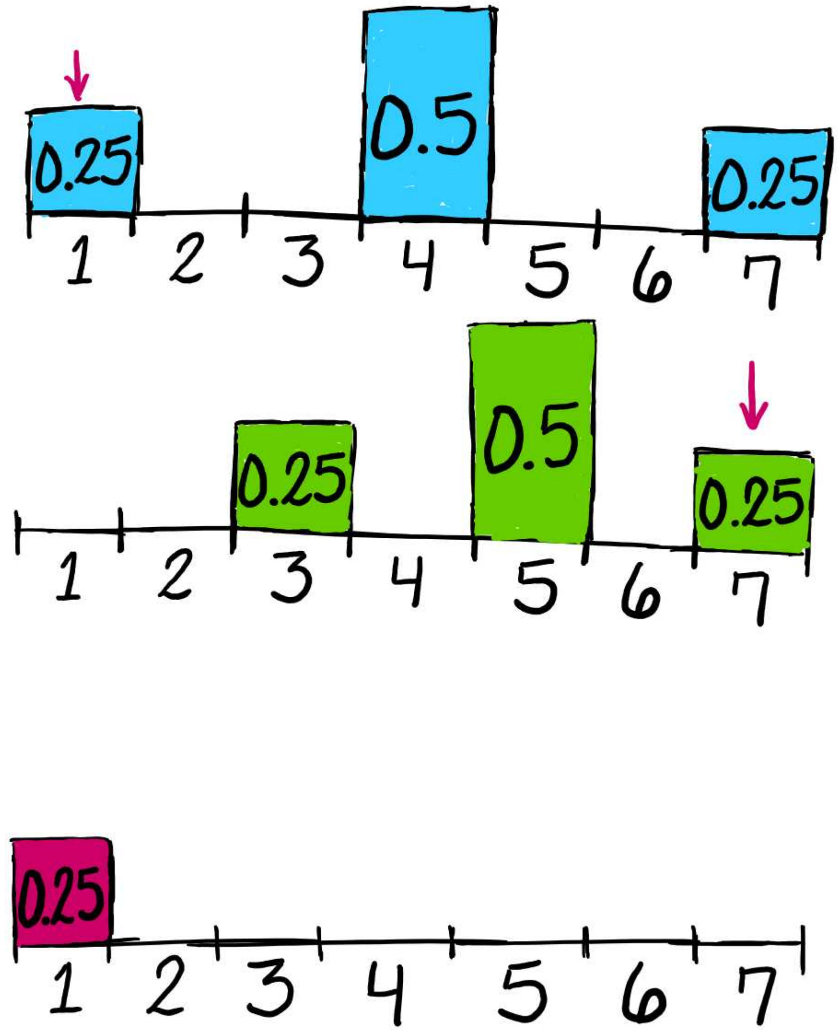
Backup Rules (4/5): Cserna Example

multiply probabilities and add to minimum value

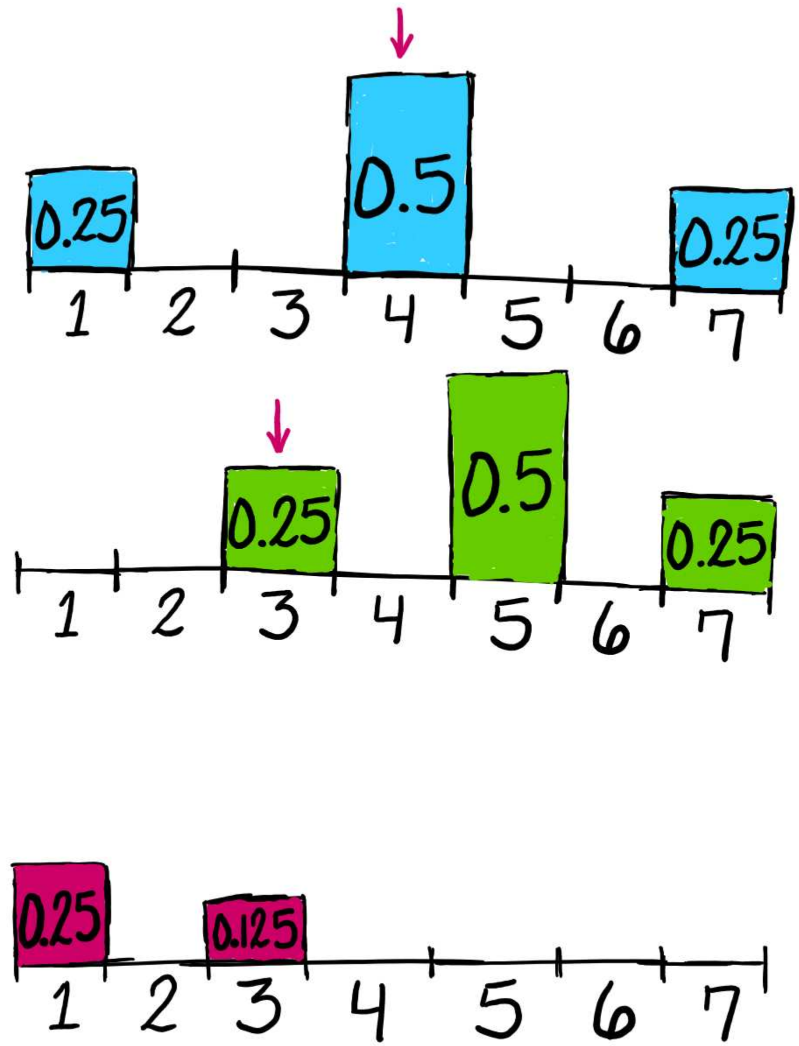


Backup Rules (4/5): Cserna Example

multiply probabilities and add to minimum value

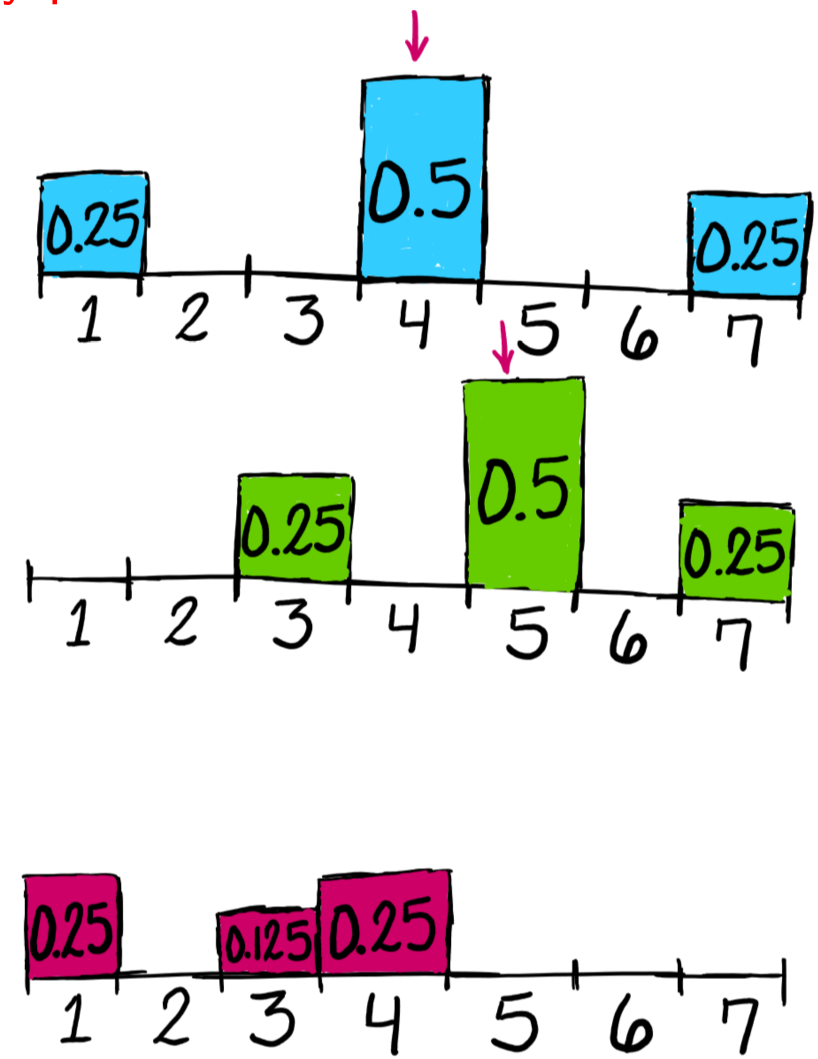


multiply probabilities and add to minimum value



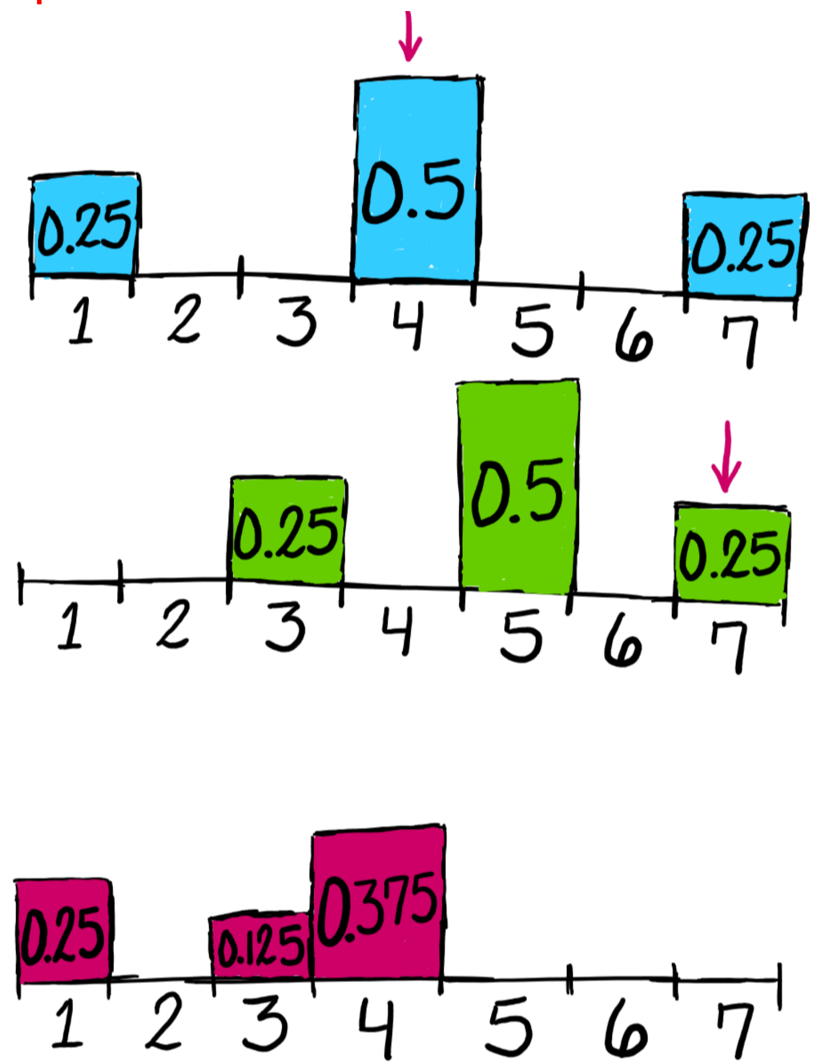
Backup Rules (4/5): Cserna Example

multiply probabilities and add to minimum value



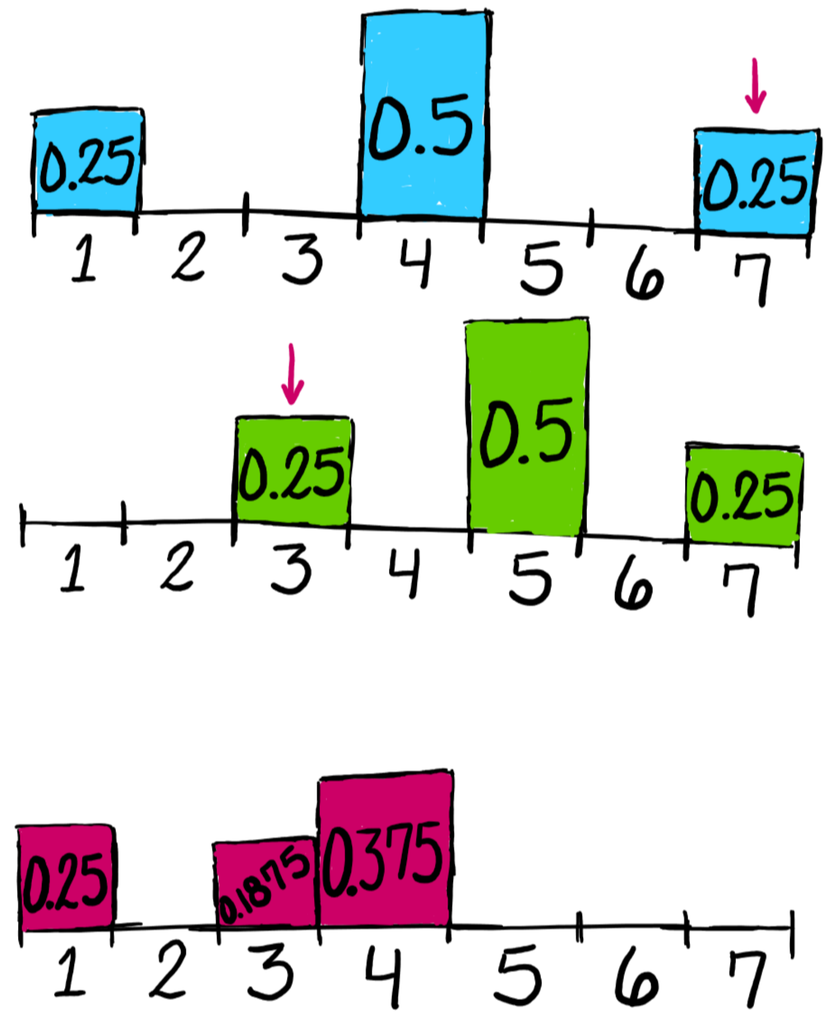
Backup Rules (4/5): Cserna Example

multiply probabilities and add to minimum value



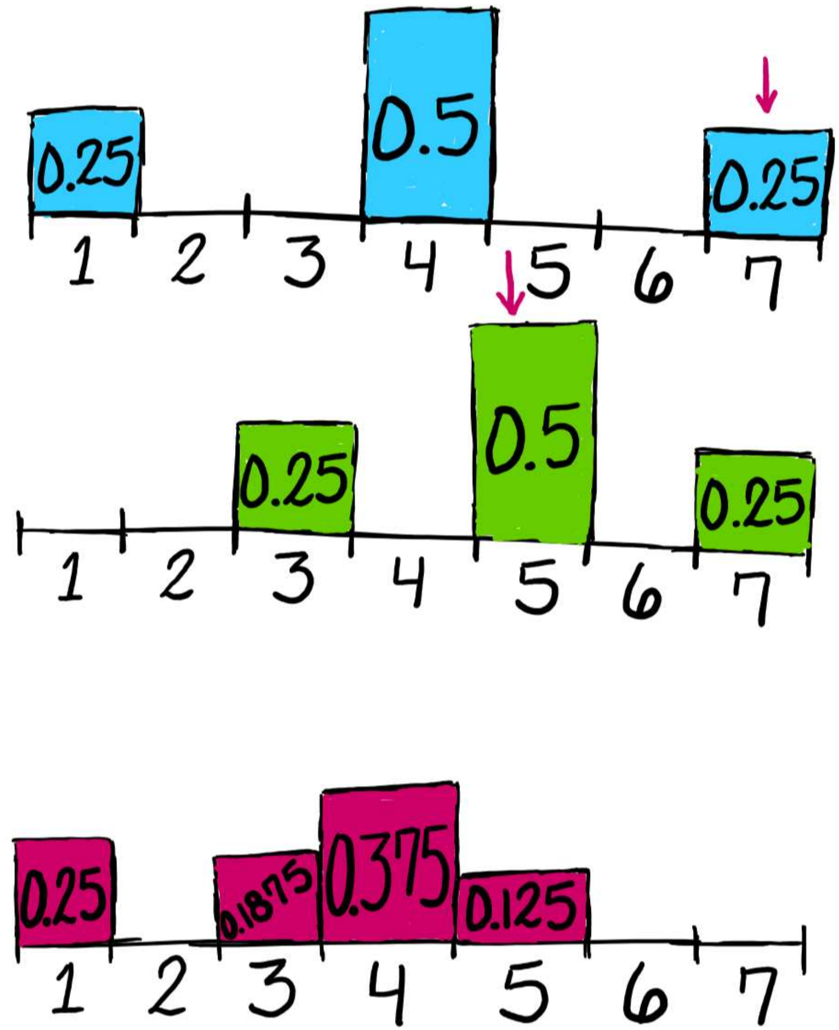
Backup Rules (4/5): Cserna Example

multiply probabilities and add to minimum value



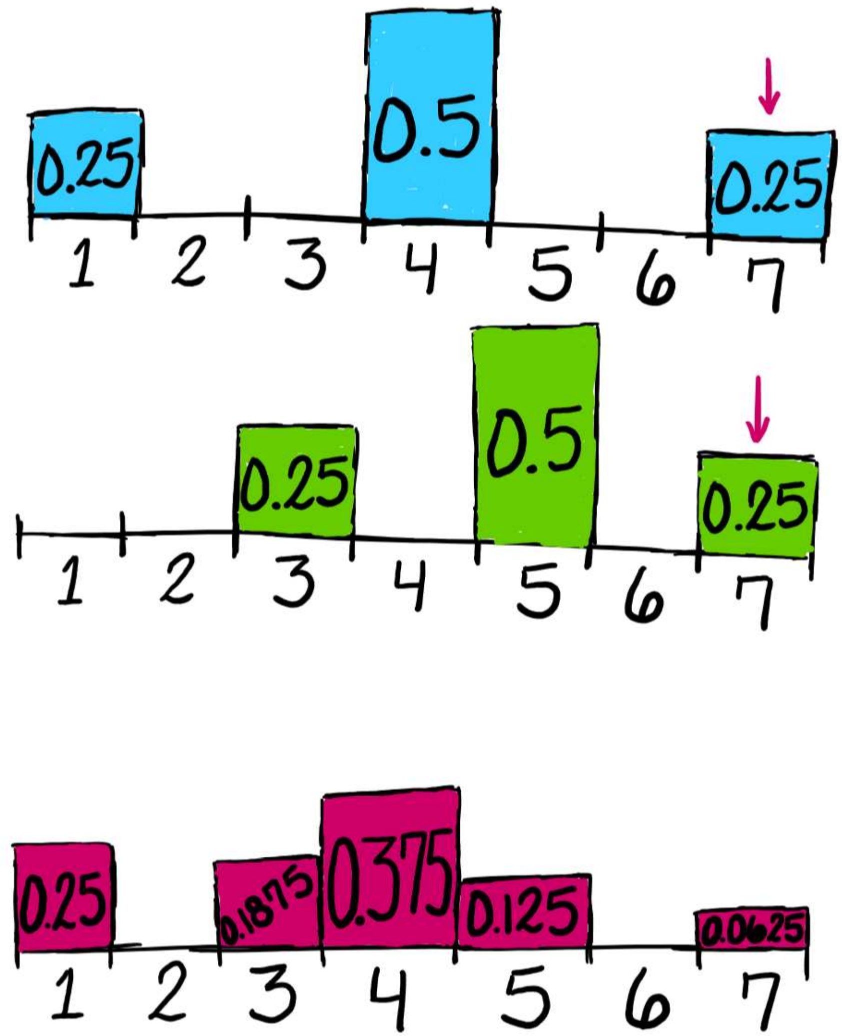
Backup Rules (4/5): Cserna Example

multiply probabilities and add to minimum value



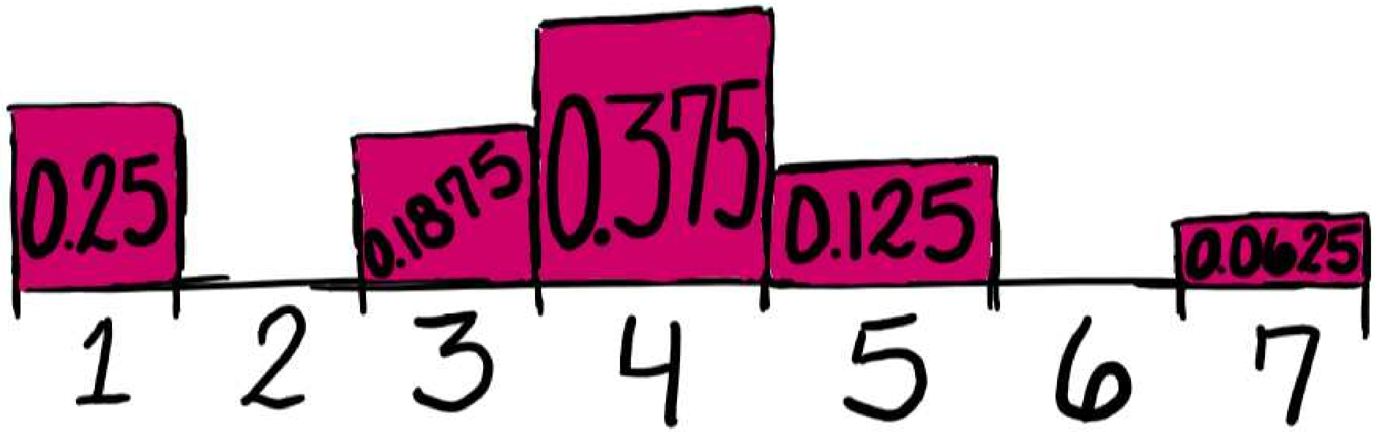
Backup Rules (4/5): Cserna Example

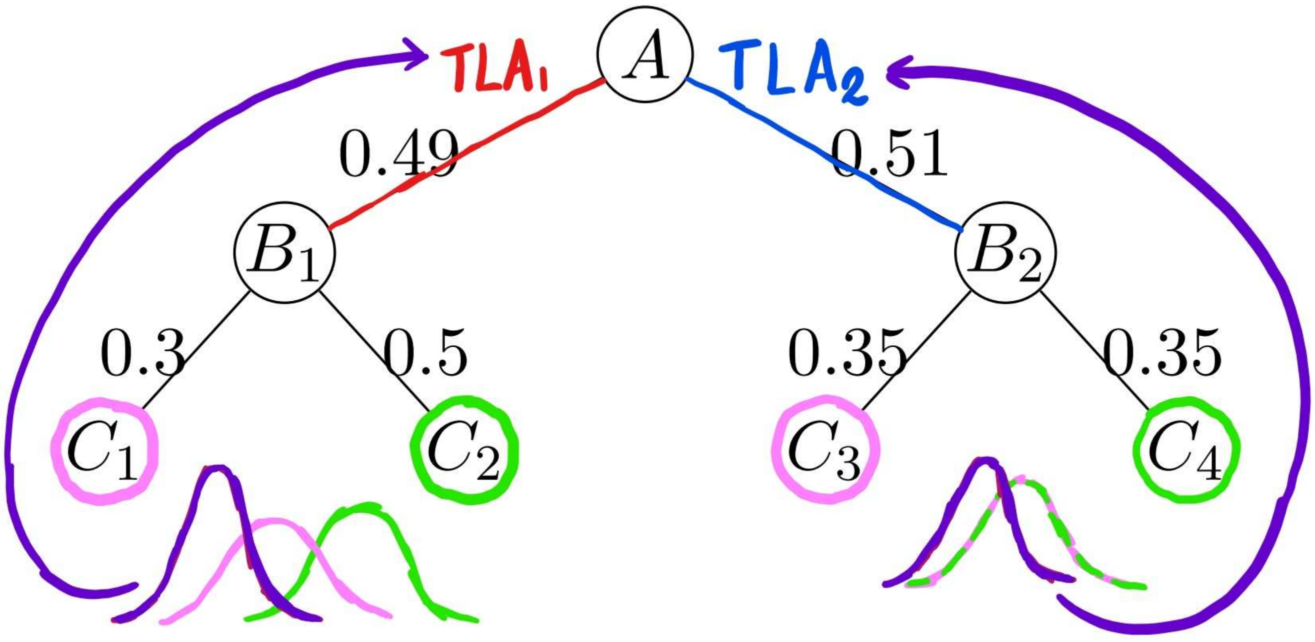
multiply probabilities and add to minimum value



Backup Rules (4/5): Cserna Example

return resulting belief distribution



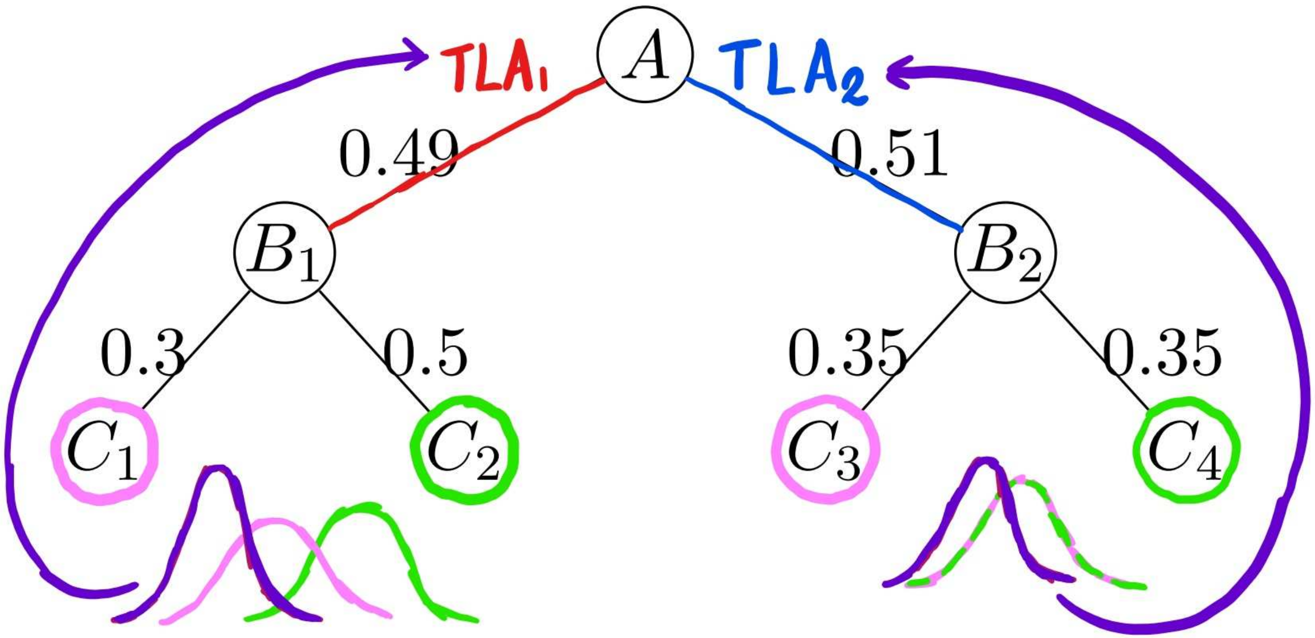


k-best (reformulated!):

parent $\leftarrow$ distribution of minimum values of k successors

class of backup rules ranging from Nancy to Cserna

Backup Rules (5/5): k-best



k-best (reformulated!):

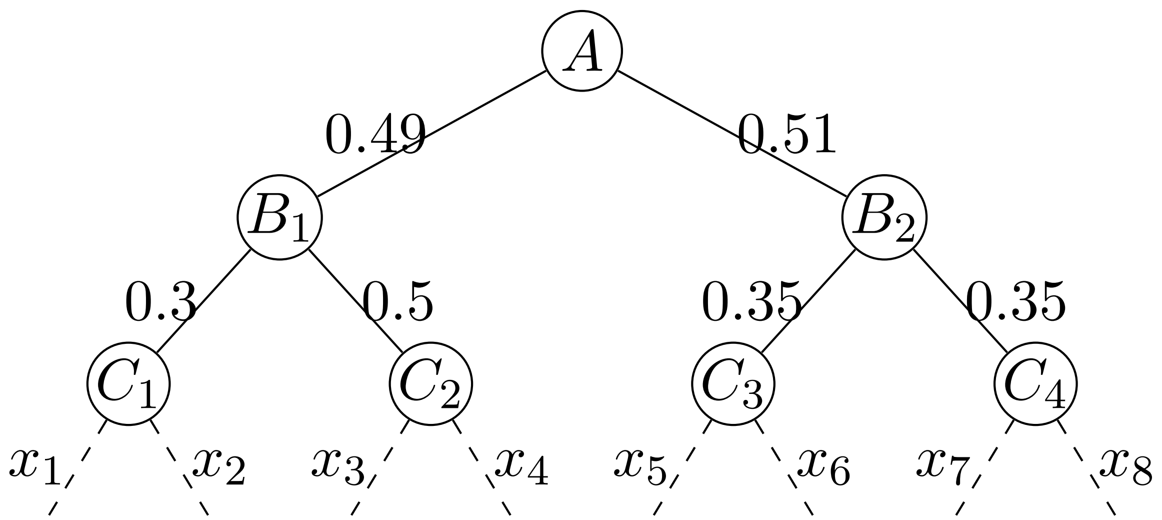
parent $\leftarrow$ distribution of minimum values of k successors

class of backup rules ranging from Nancy to Cserna

all incorrect, which works best?

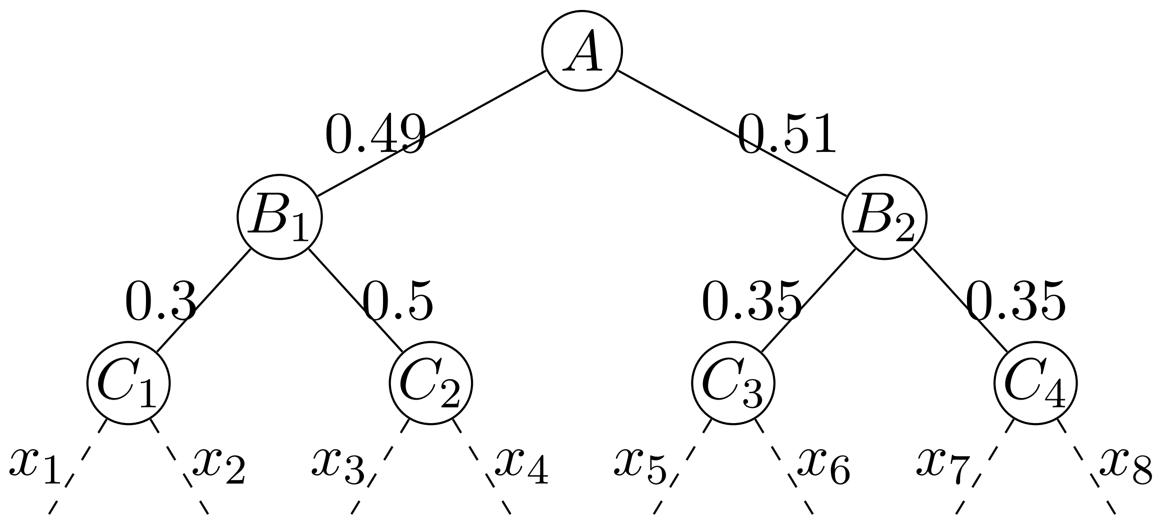
Benchmark: Last Incremental Decision

explore all but last level of search tree
backup information from frontier
agent picks first action
remaining path is optimally solved
why? used as a test by Pemberton in 1995



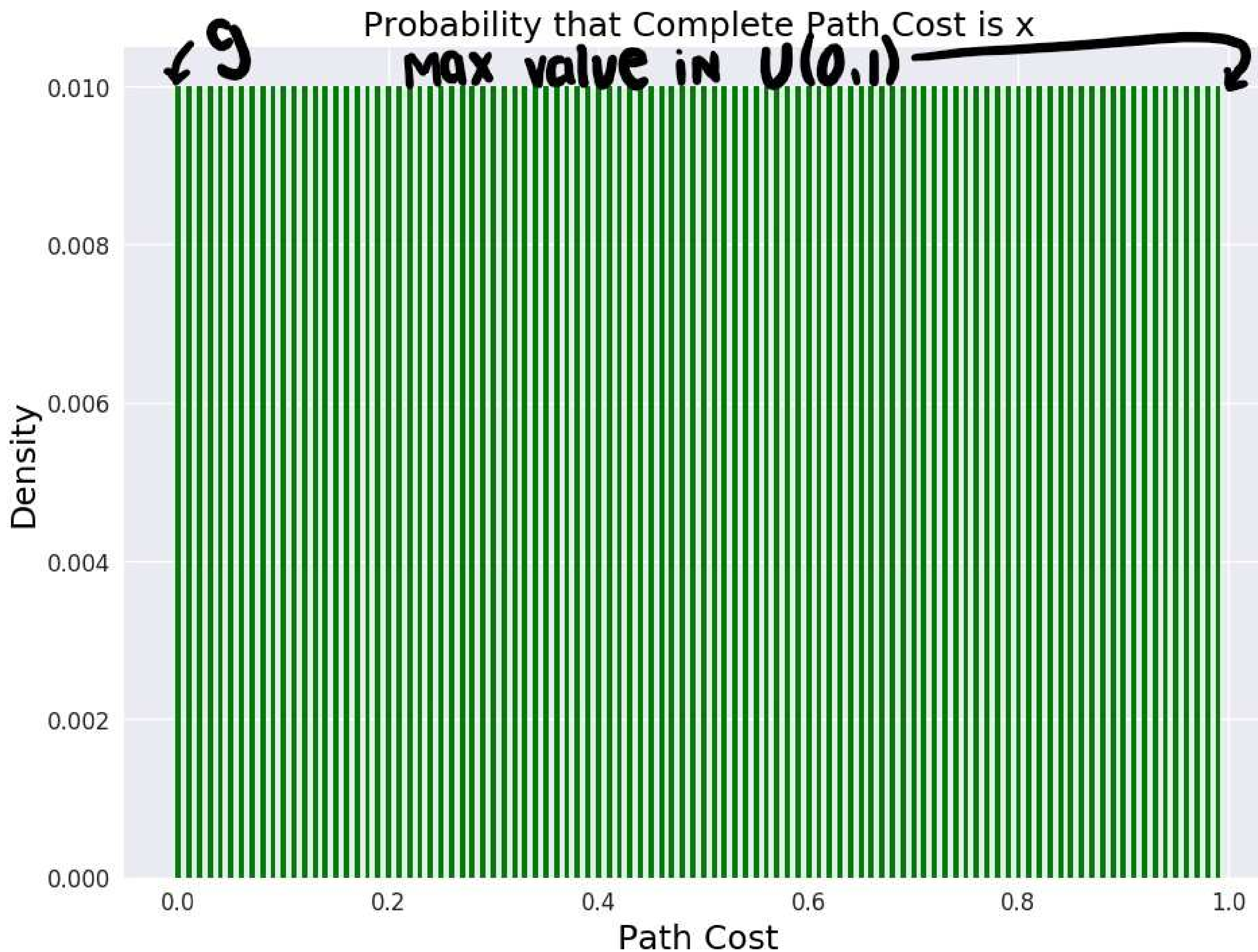
Benchmark: Last Incremental Decision

explore all but last level of search tree
backup information from frontier
agent picks first action
remaining path is optimally solved
why? used as a test by Pemberton in 1995



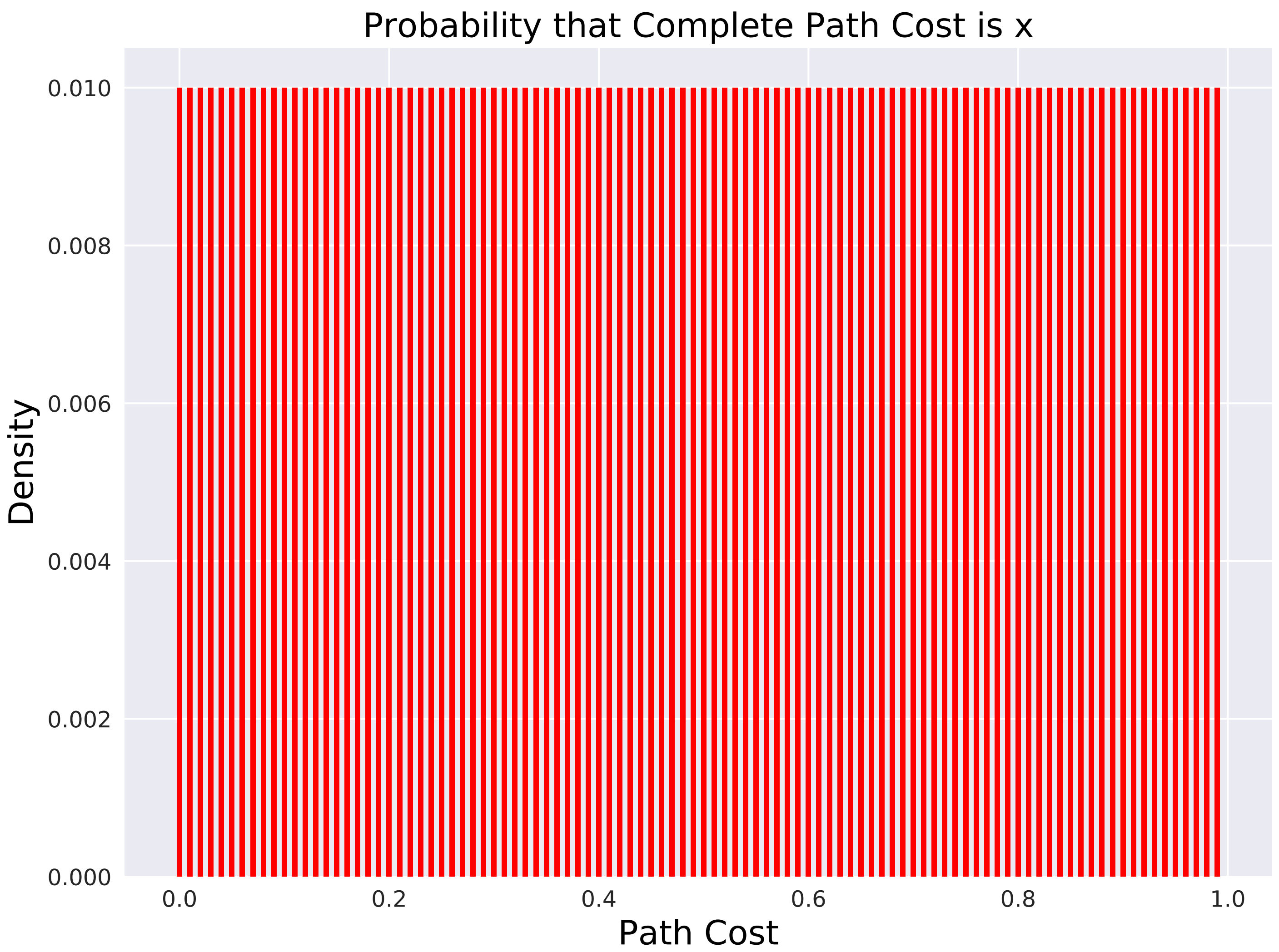
precise expected values of all successors are known!
expect Cserna to be optimal on average!

One-level Belief



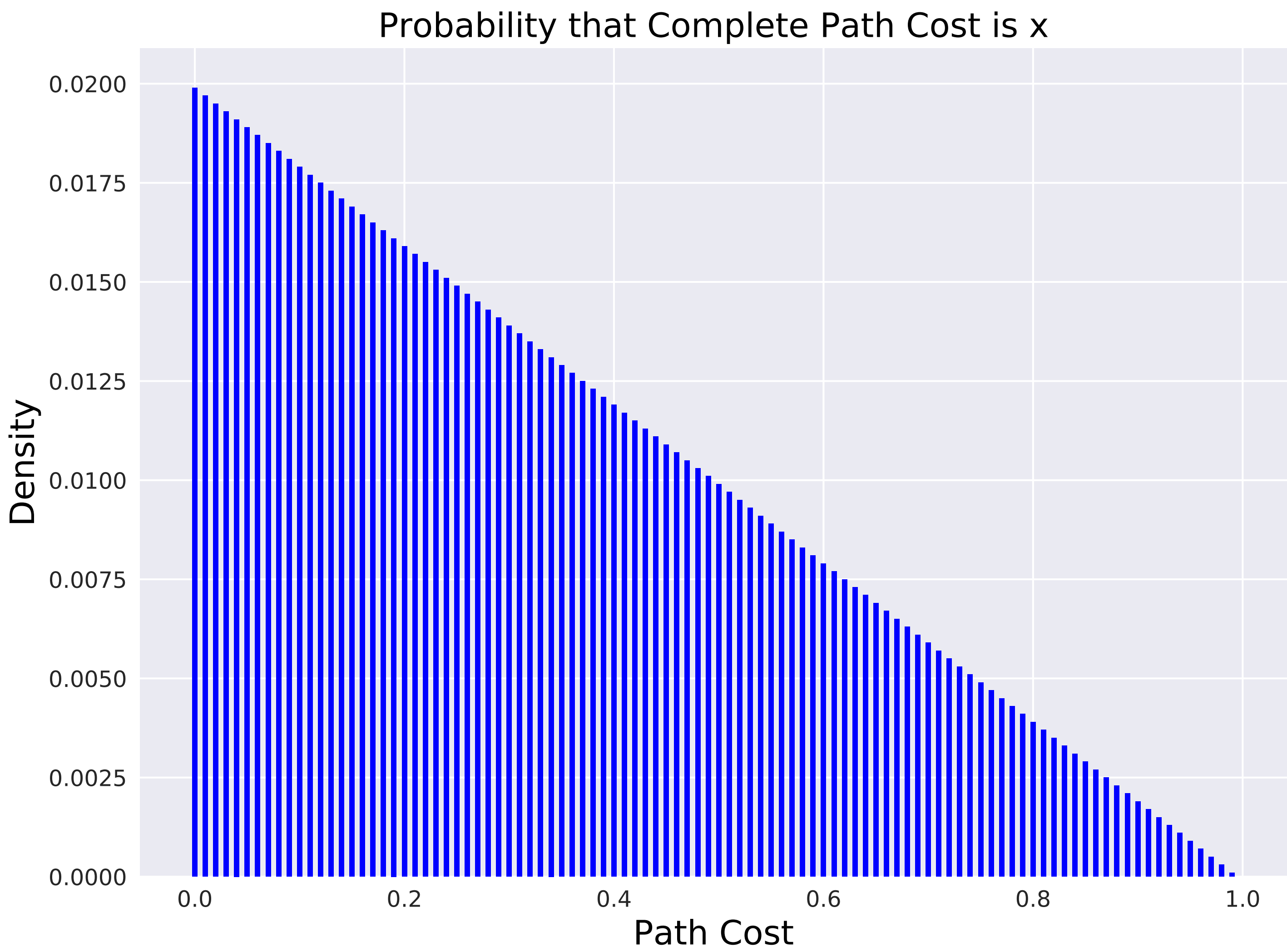
Take minimum of n U(0,1) beliefs shifted by g value

One-level Belief



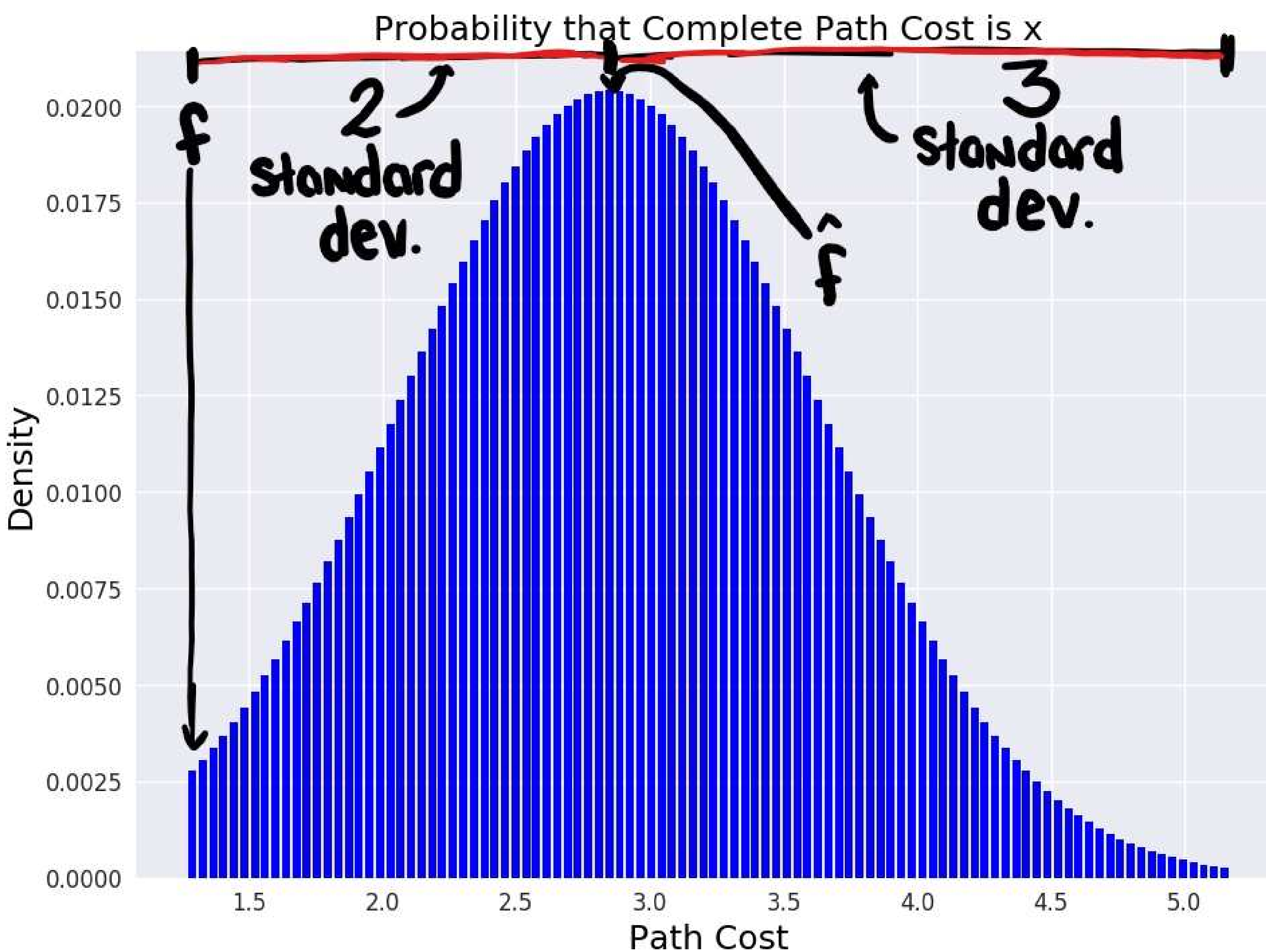
Take minimum of n U(0,1) beliefs shifted by g value

One-level Belief



One-level belief is result of this operation

Gaussian Belief



mean at $\hat{f}$, variance proportional to distance from goal

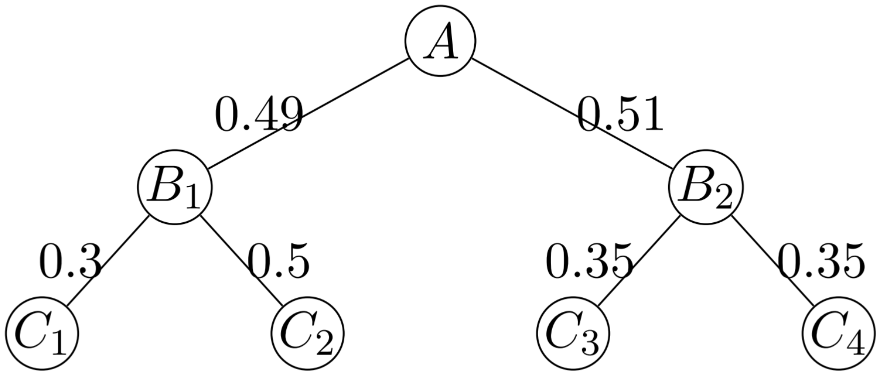
Benchmark: Random Trees

uniform n-ary tree with depth d
- binary with depth 10 for last incremental decision
- binary with depth 100 for other experiments
- implementation lazy but deterministic

all leaf nodes are goals

edge costs $\sim$ Uniform(0,1)
- so $h = 0$

$g =$ sum of edge costs



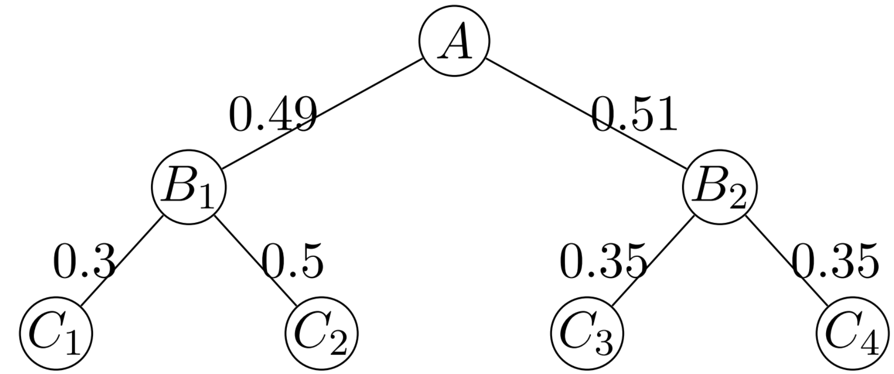
Benchmark: Random Trees

uniform n-ary tree with depth d
- binary with depth 10 for last incremental decision
- binary with depth 100 for other experiments
- implementation lazy but deterministic

all leaf nodes are goals

edge costs $\sim$ Uniform(0,1)
- so $h = 0$

$g =$ sum of edge costs



Real-time search using different backup rules

1. limited lookahead, then take action
   - single action commitment
2. sum edges costs until goal

Benchmark: Random Trees

uniform n-ary tree with depth d

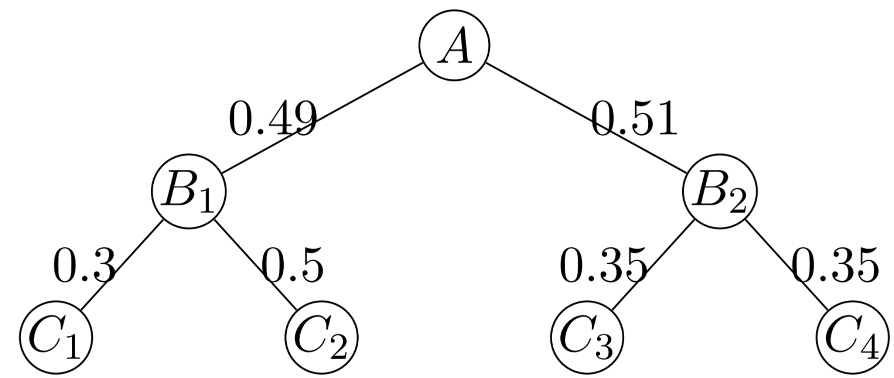
- binary with depth 10 for last incremental decision
- binary with depth 100 for other experiments
- implementation lazy but deterministic

all leaf nodes are goals

edge costs $\sim$ Uniform(0,1)

- so $h = 0$

$g =$ sum of edge costs

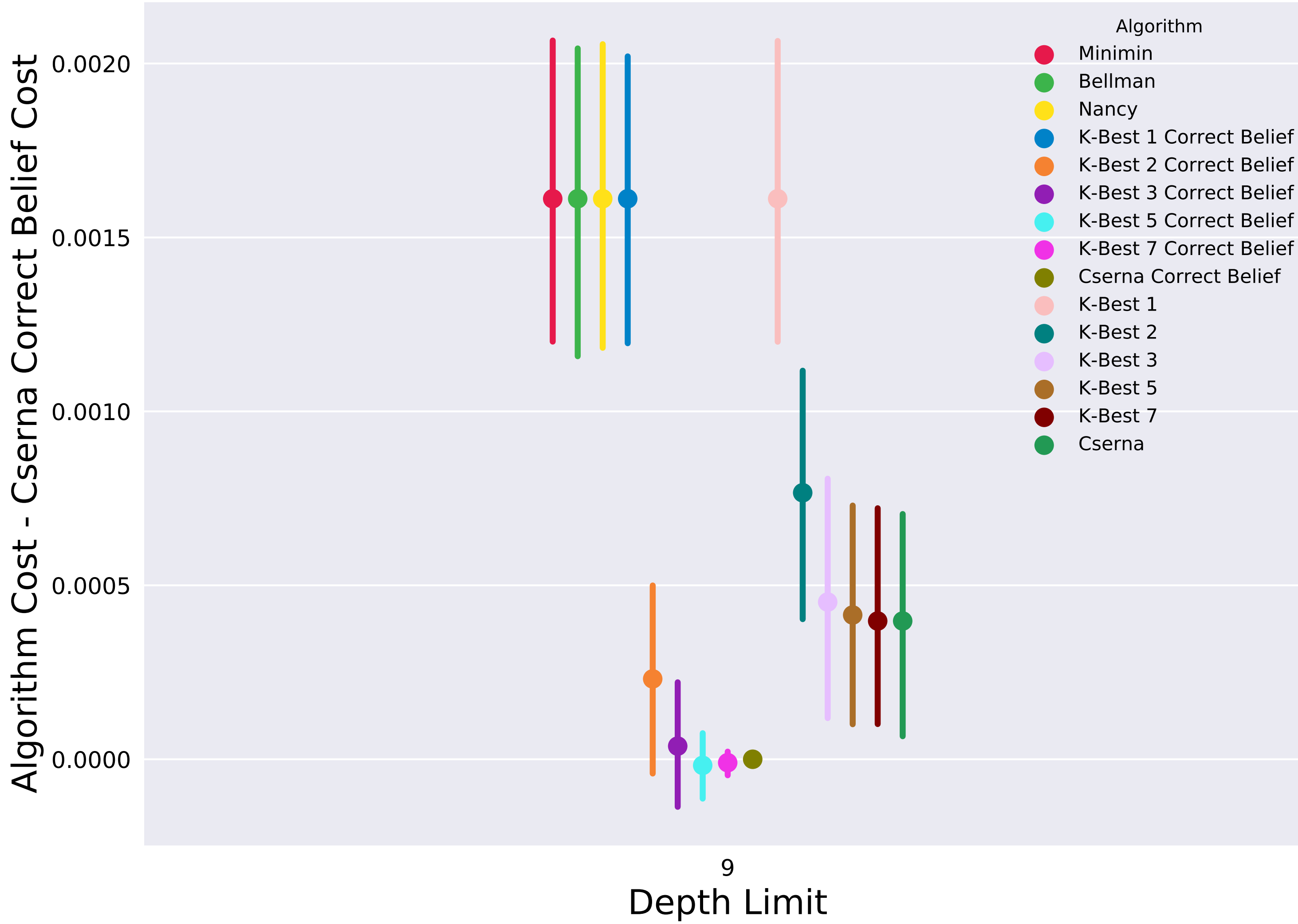


Real-time search using different backup rules

1. limited lookahead, then take action
   - single action commitment
2. sum edges costs until goal

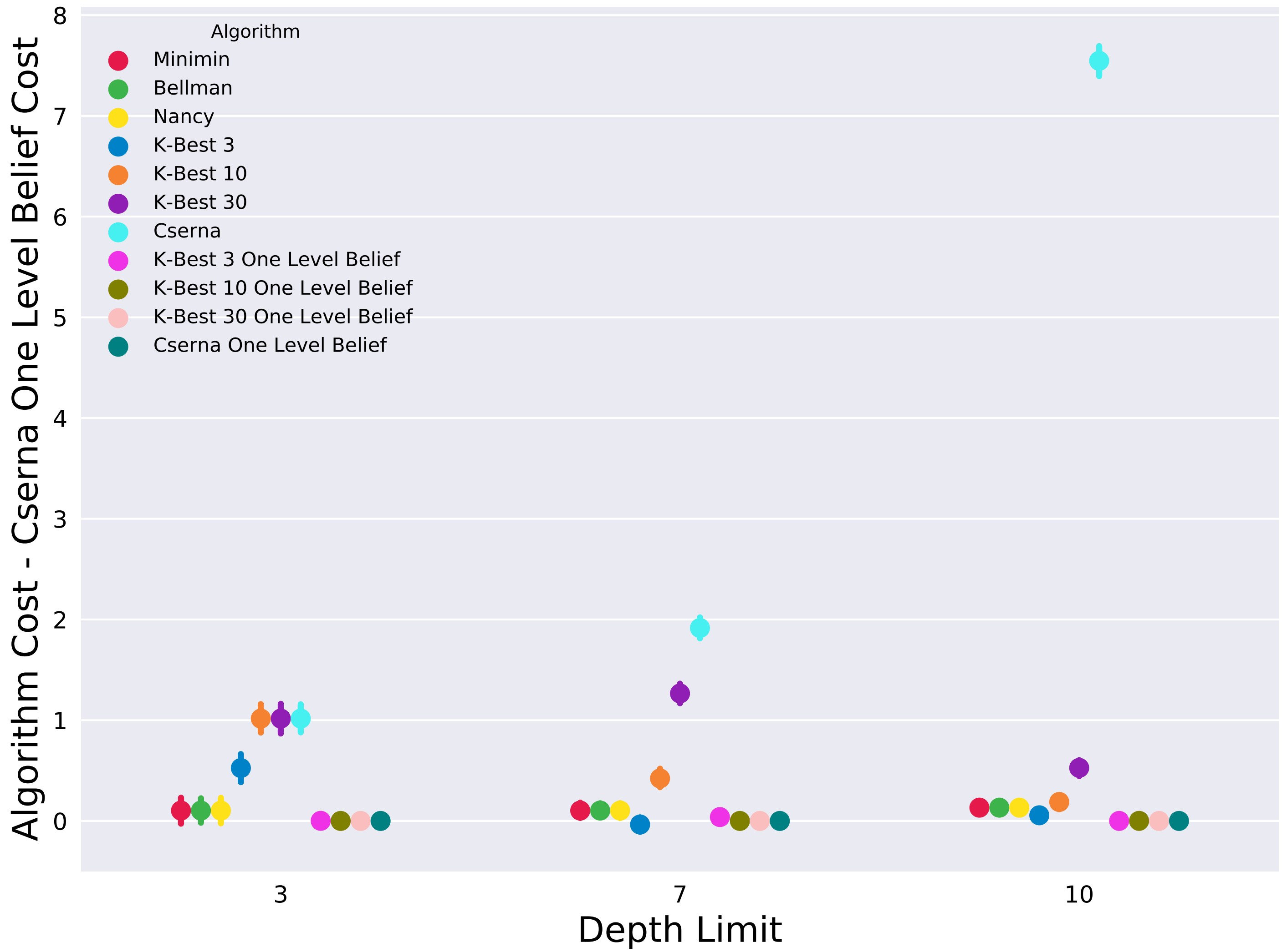
best backup rule is one with lowest average solution cost!

Backup Rules on Last Incremental Decision



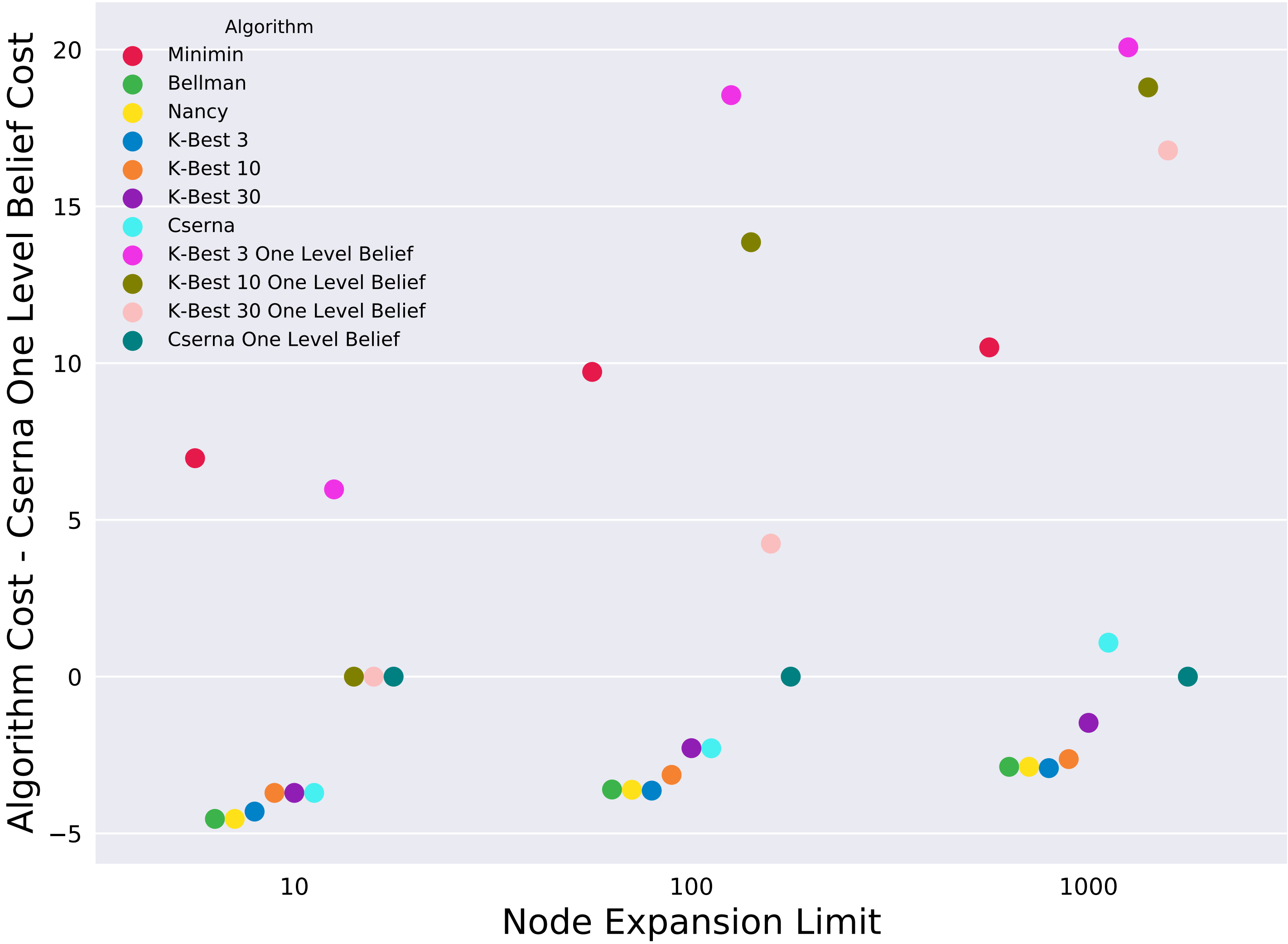
as expected, Cserna is optimal on average!

Backup Rules + Uniform Bounded Depth Expansion



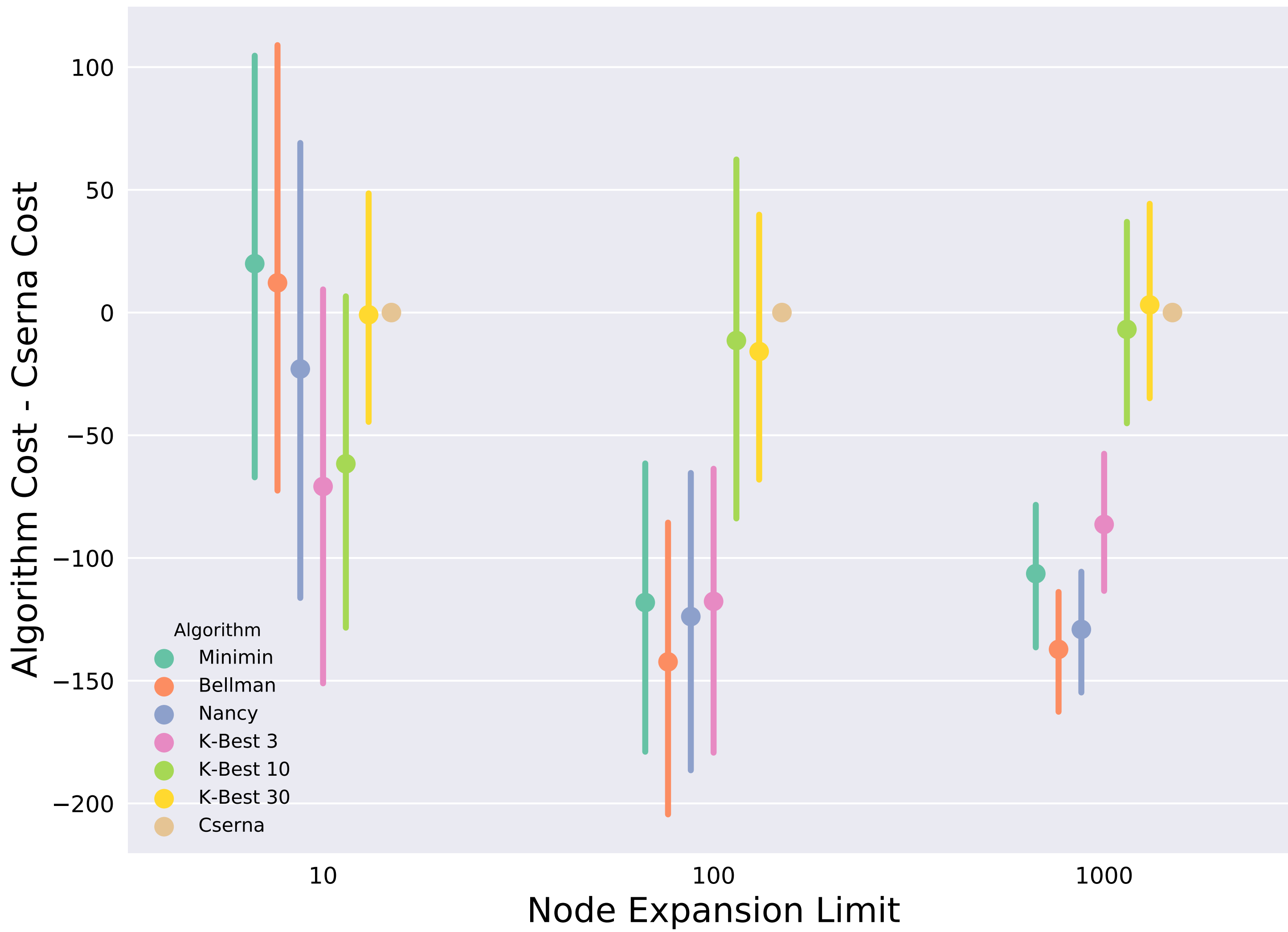
Minimin/Bellman/Nancy equivalent, Cserna Gaussian worse

Backup Rules + f Expansion on Random Trees



Minimin has no h, Bellman/Nancy equivalent, Cserna worse

Backup Rules + f Expansion on 15-Puzzle



Bellman and Nancy best overall

Answering the First Question

1. Which action to select?
   minimum $\hat{f}$ by principle of rationality

Answering the First Question

1. Which action to select?
   minimum $\hat{f}$ by principle of rationality
2. How to backup from frontier?
   Bellman or Nancy (Nancy has added benefit of belief)

Answering the First Question

1. Which action to select?
   minimum $\hat{f}$ by principle of rationality
2. How to backup from frontier?
   Bellman or Nancy (Nancy has added benefit of belief)
3. Which nodes to expand?

Answering the First Question

1. Which action to select?
   minimum $\hat{f}$ by principle of rationality
2. How to backup from frontier?
   Bellman or Nancy (Nancy has added benefit of belief)
3. Which nodes to expand?

setting:

- search expands frontier nodes under top-level actions (TLAs)
- only lowest $\hat{f}$ node under TLA is important (Bellman/Nancy backup)
- how to select node to expand?

Expansion Strategies

f expansion:

expand node on frontier with lowest lower bound

ignores uncertainty in heuristic

f expansion:

expand node on frontier with lowest lower bound

ignores uncertainty in heuristic

$\hat{f}$ expansion:

expand node on frontier with lowest expected value

accounts for uncertainty, but with scalar

f expansion:

expand node on frontier with lowest lower bound

ignores uncertainty in heuristic

$\hat{f}$ expansion:

expand node on frontier with lowest expected value

accounts for uncertainty, but with scalar

Breadth-first expansion:

expand nodes in order of generation

literally just brute force

Expansion Strategies

f expansion:

expand node on frontier with lowest lower bound

ignores uncertainty in heuristic

$\hat{f}$ expansion:

expand node on frontier with lowest expected value

accounts for uncertainty, but with scalar

Breadth-first expansion:

expand nodes in order of generation

literally just brute force

alternatives?

Expansion Strategies

f expansion:

expand node on frontier with lowest lower bound

ignores uncertainty in heuristic

$\hat{f}$ expansion:

expand node on frontier with lowest expected value

accounts for uncertainty, but with scalar

Breadth-first expansion:

expand nodes in order of generation

literally just brute force

alternatives?

Risk-based expansion (new!):

expand nodes which minimize expected regret

acknowledges uncertainty, relies on belief of values

Expansion Strategies

f expansion:

expand node on frontier with lowest lower bound

ignores uncertainty in heuristic

$\hat{f}$ expansion:

expand node on frontier with lowest expected value

accounts for uncertainty, but with scalar

Breadth-first expansion:

expand nodes in order of generation

literally just brute force

alternatives?

Risk-based expansion (new!):

expand nodes which minimize expected regret

acknowledges uncertainty, relies on belief of values

which of these performs best?

Risk-based Expansion

Risk: expected regret if a suboptimal action is selected

$$\mathbb{E}\left[f^*(\alpha) - f^*(\beta) \;\middle|\; f^*(\beta) < f^*(\alpha)\right]$$

expectation over possible values for TLAs

exploit full belief given by Nancy backups

Risk-based Expansion

Risk: expected regret if a suboptimal action is selected

$$\mathbb{E}\left[f^*(\alpha) - f^*(\beta) \mid f^*(\beta) < f^*(\alpha)\right]$$

expectation over possible values for TLAs

exploit full belief given by Nancy backups

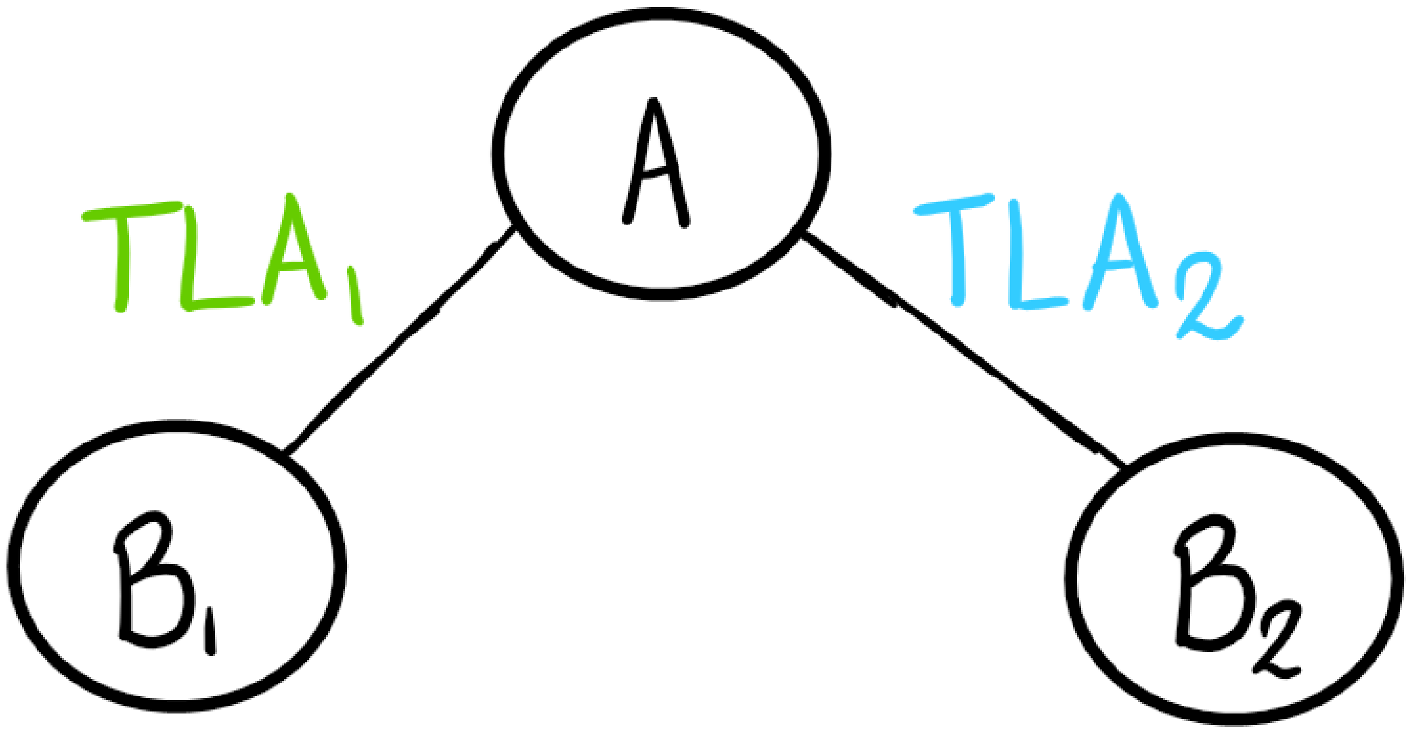
in discrete case, where α is TLA with lowest expected value, all other TLAs are β_i, and a and b_i are potential values from their beliefs:

$$\sum_{\beta_i} \left(\sum_{a} \sum_{b_i < a} P(a) P(b_i)(a - b_i) \right)$$

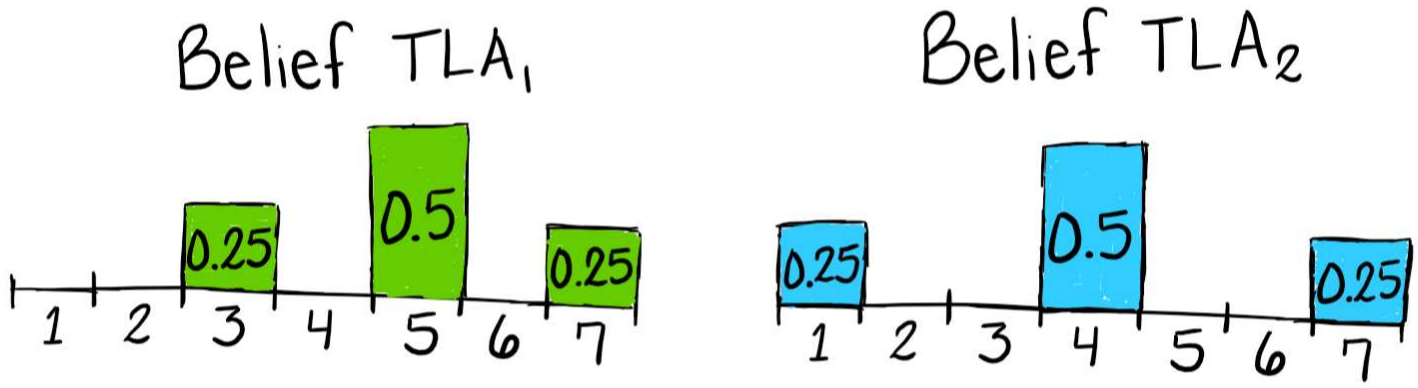
expand under the TLA that minimizes risk!

Risk-based Expansion Example

expand under TLA_1 or TLA_2?

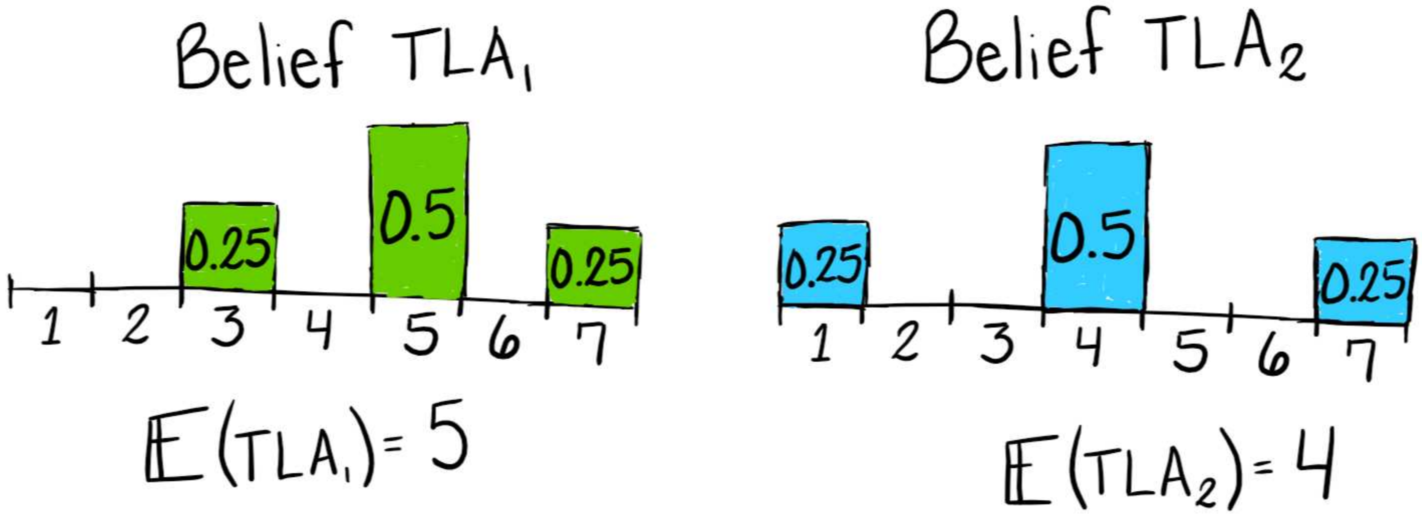


obtain beliefs of TLA_1 and TLA_2



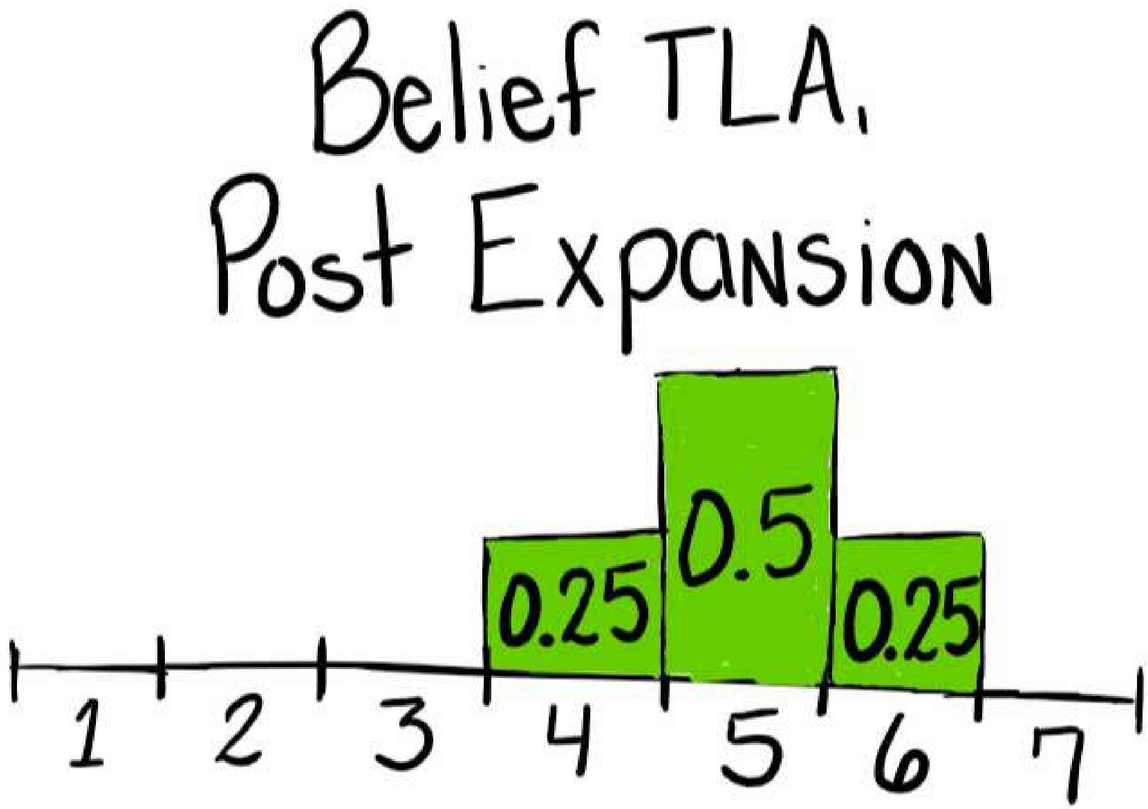
Risk-based Expansion Example

identify TLA_2 as α TLA

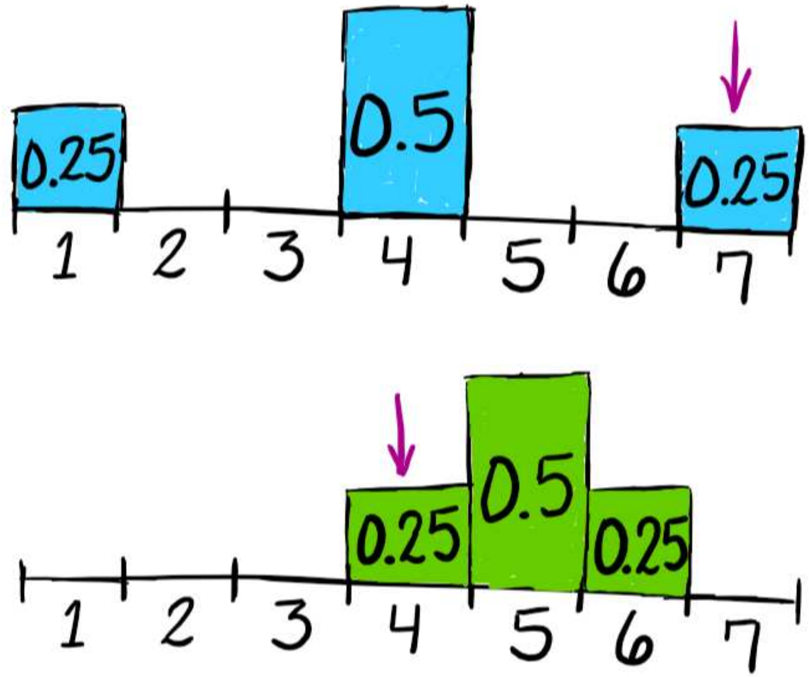


Risk-based Expansion Example

calculate post expansion belief for TLA_1



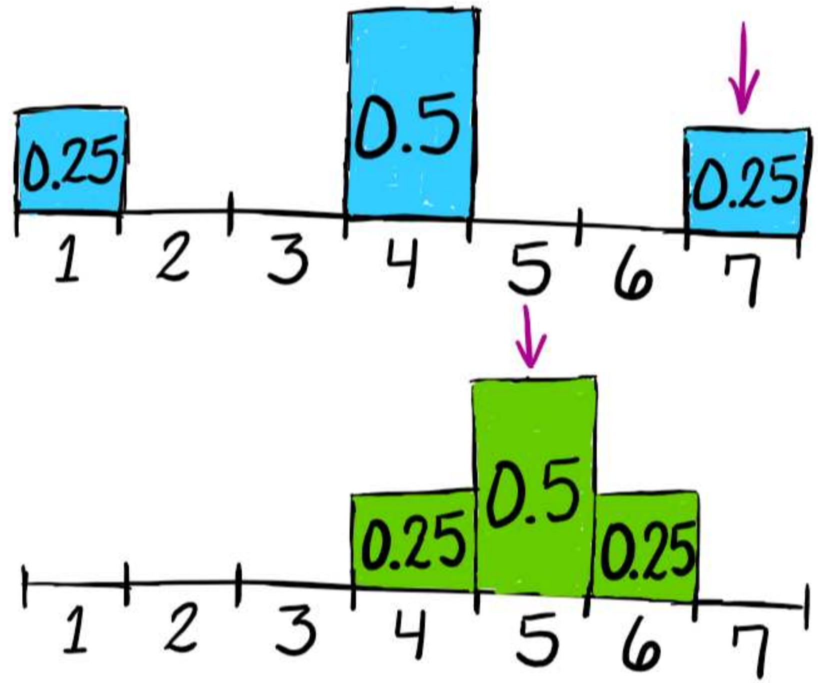
use post expansion belief to calculate risk of expansion under TLA_1



$$\text{risk}_{\text{TLA}_1} = 0.25 \times 0.25 \times (7-4)$$

Risk-based Expansion Example

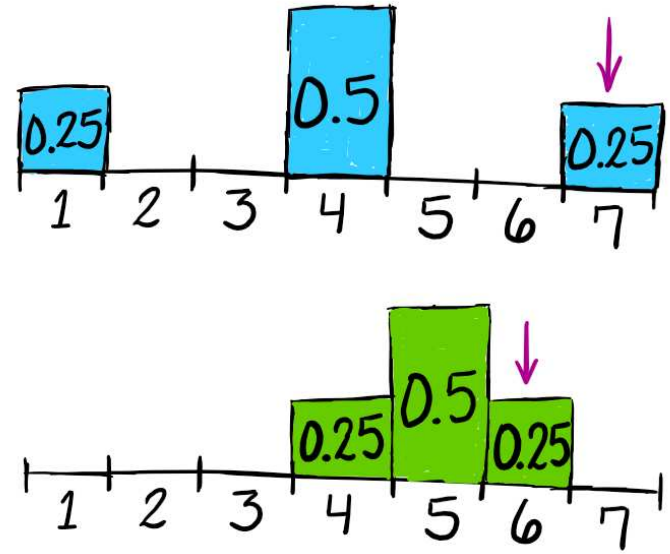
use post expansion belief to calculate risk of expansion under TLA_1



$$risk_{TLA_1} = 0.1875 + 0.25 \times 0.5 \times (7-5)$$

Risk-based Expansion Example

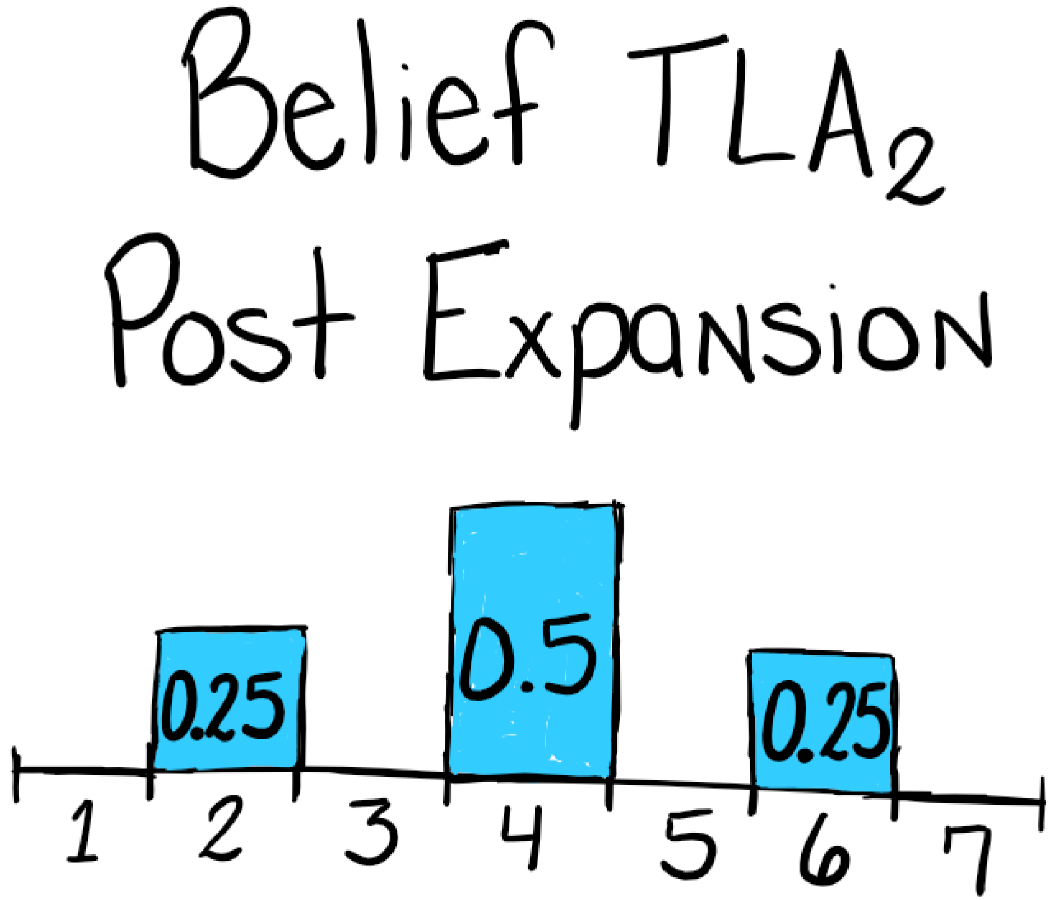
risk of expansion under TLA_1 is 0.5!



$$risk_{TLA_1} = 0.4375 + 0.25 \times 0.25 \times (7-6)$$

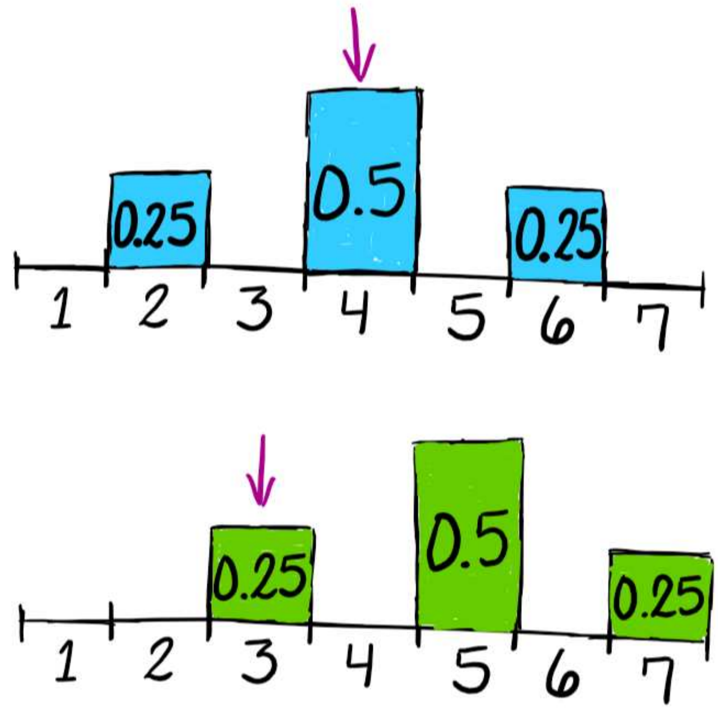
$$risk_{TLA_1} = 0.5$$

calculate post expansion belief for TLA_2



Risk-based Expansion Example

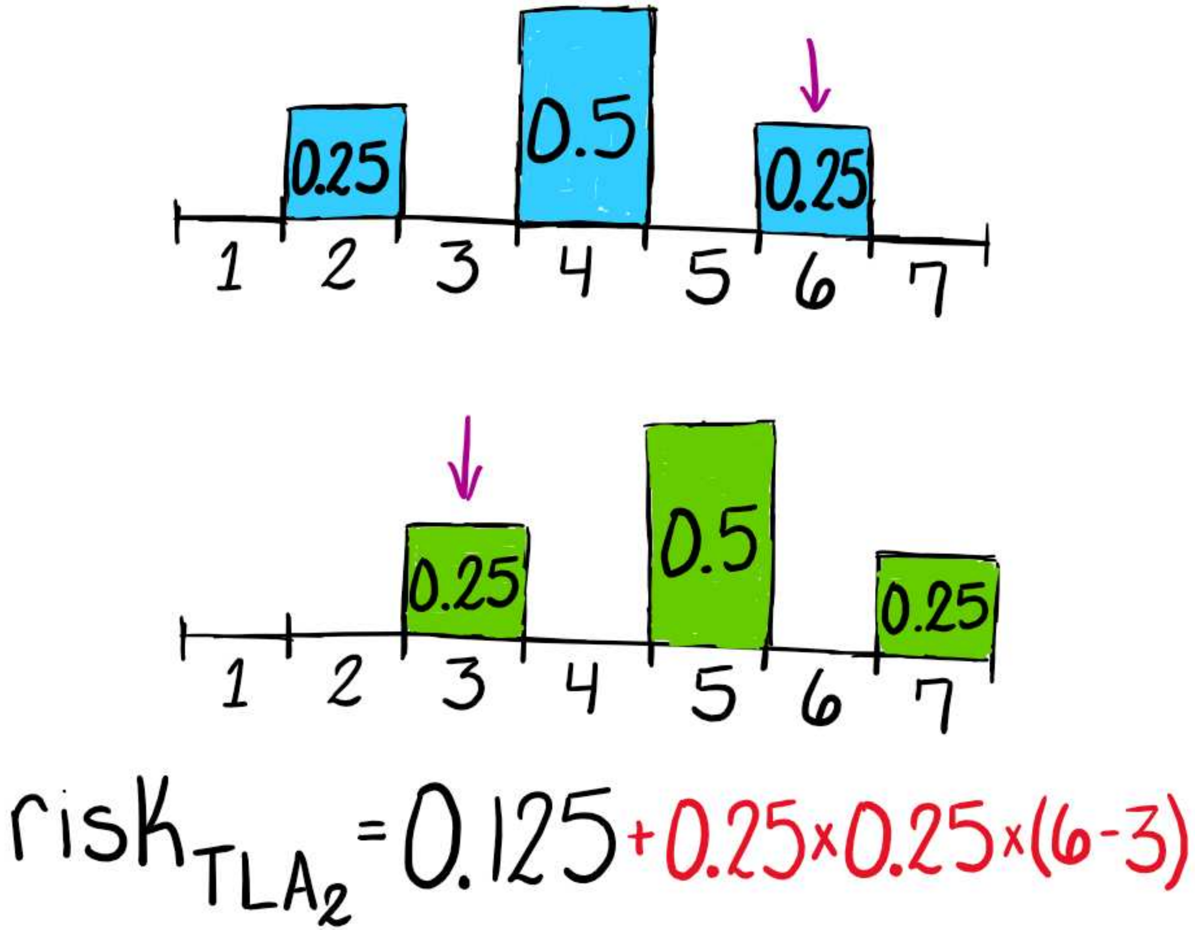
use post expansion belief to calculate risk of expansion under TLA_2



$$risk_{TLA_2} = 0.5 \times 0.25 \times (4-3)$$

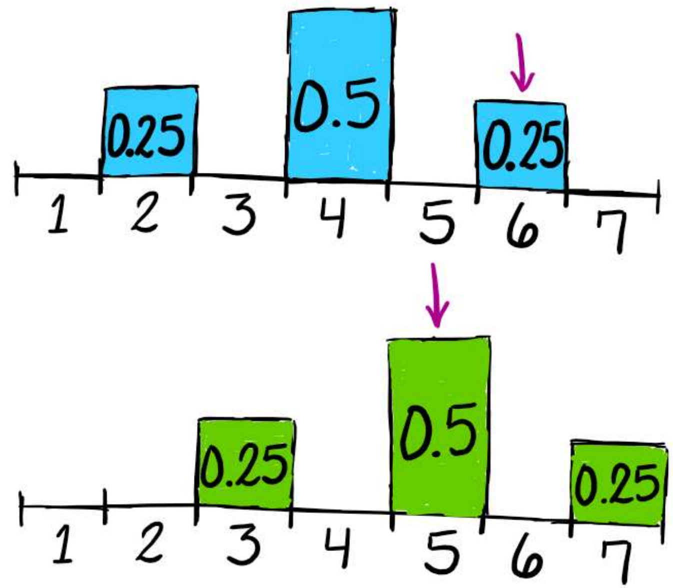
Risk-based Expansion Example

use post expansion belief to calculate risk of expansion under TLA_2



Risk-based Expansion Example

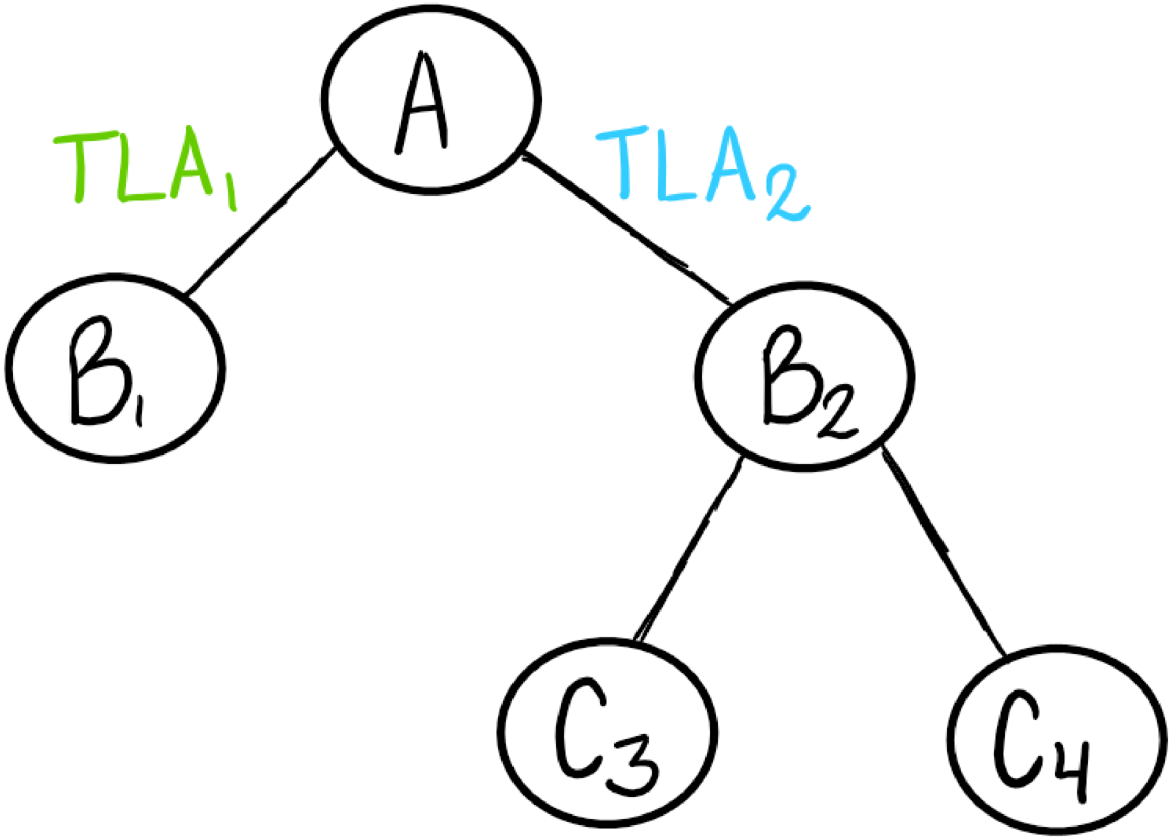
risk of expansion under TLA_2 is



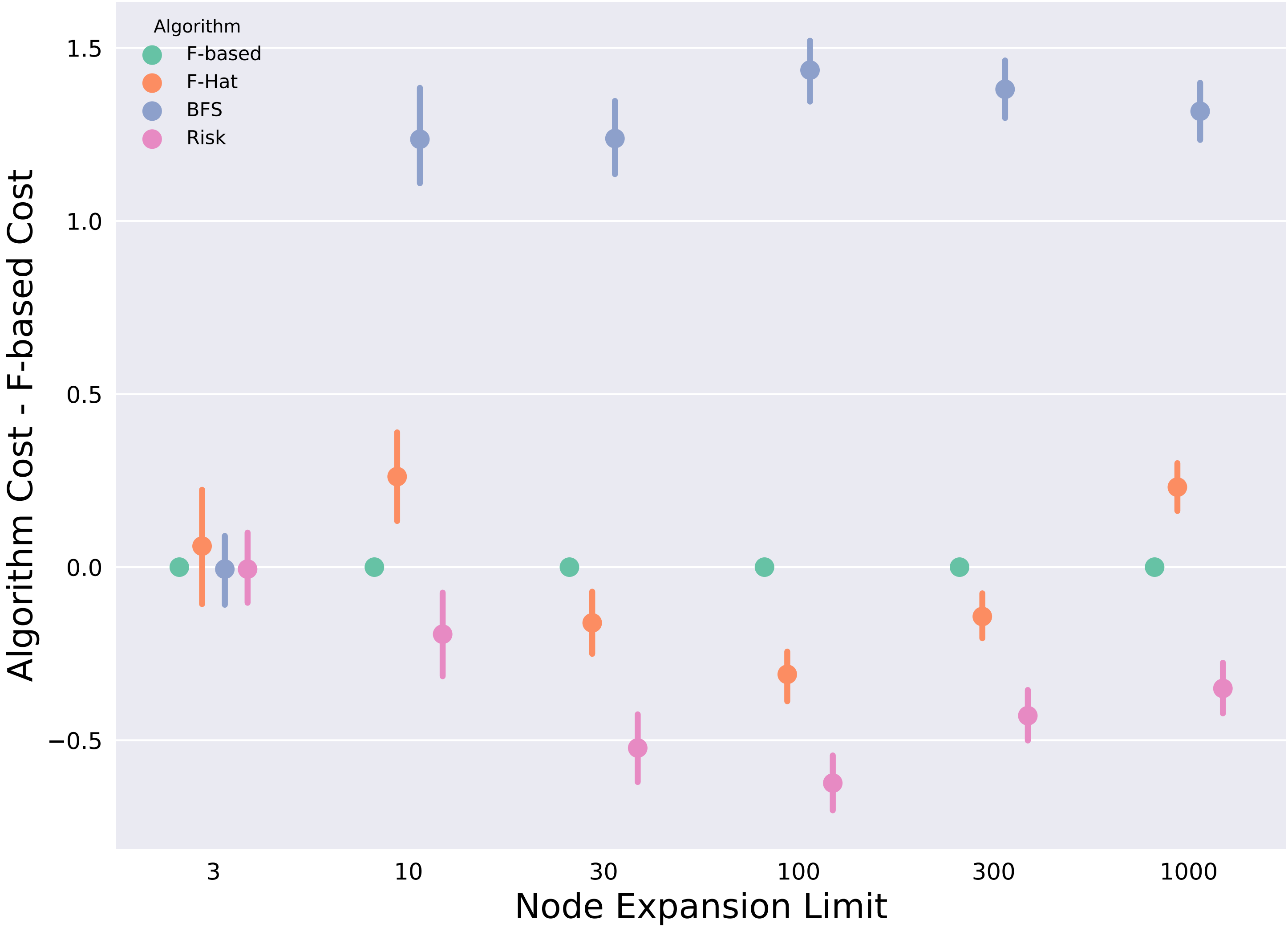
$$risk_{TLA_2} = 0.3125 + 0.25 \times 0.5 \times (6-5)$$

$$risk_{TLA_2} = 0.4375$$

expand under TLA_2!

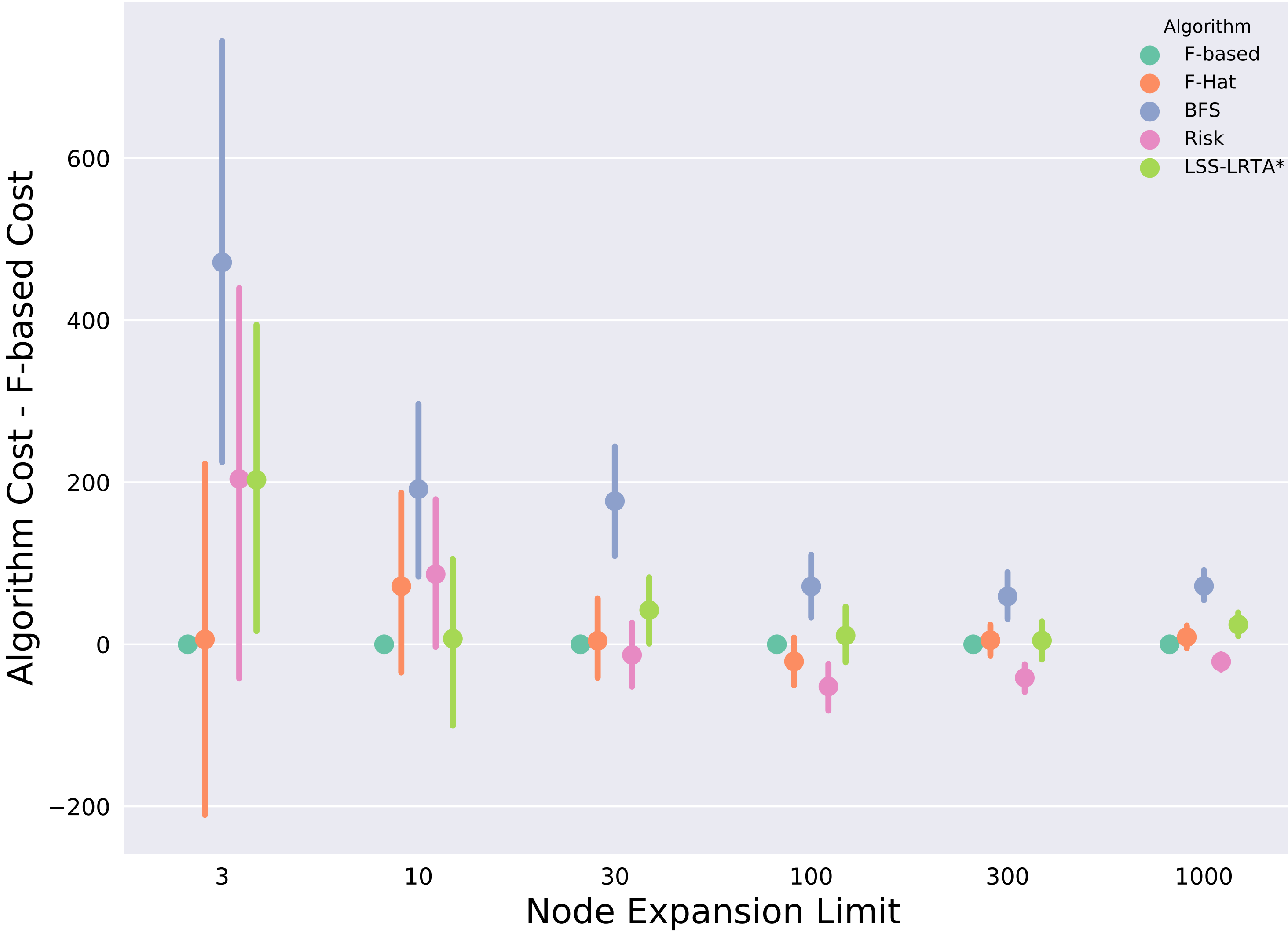


Expansion Strategies on Random Trees



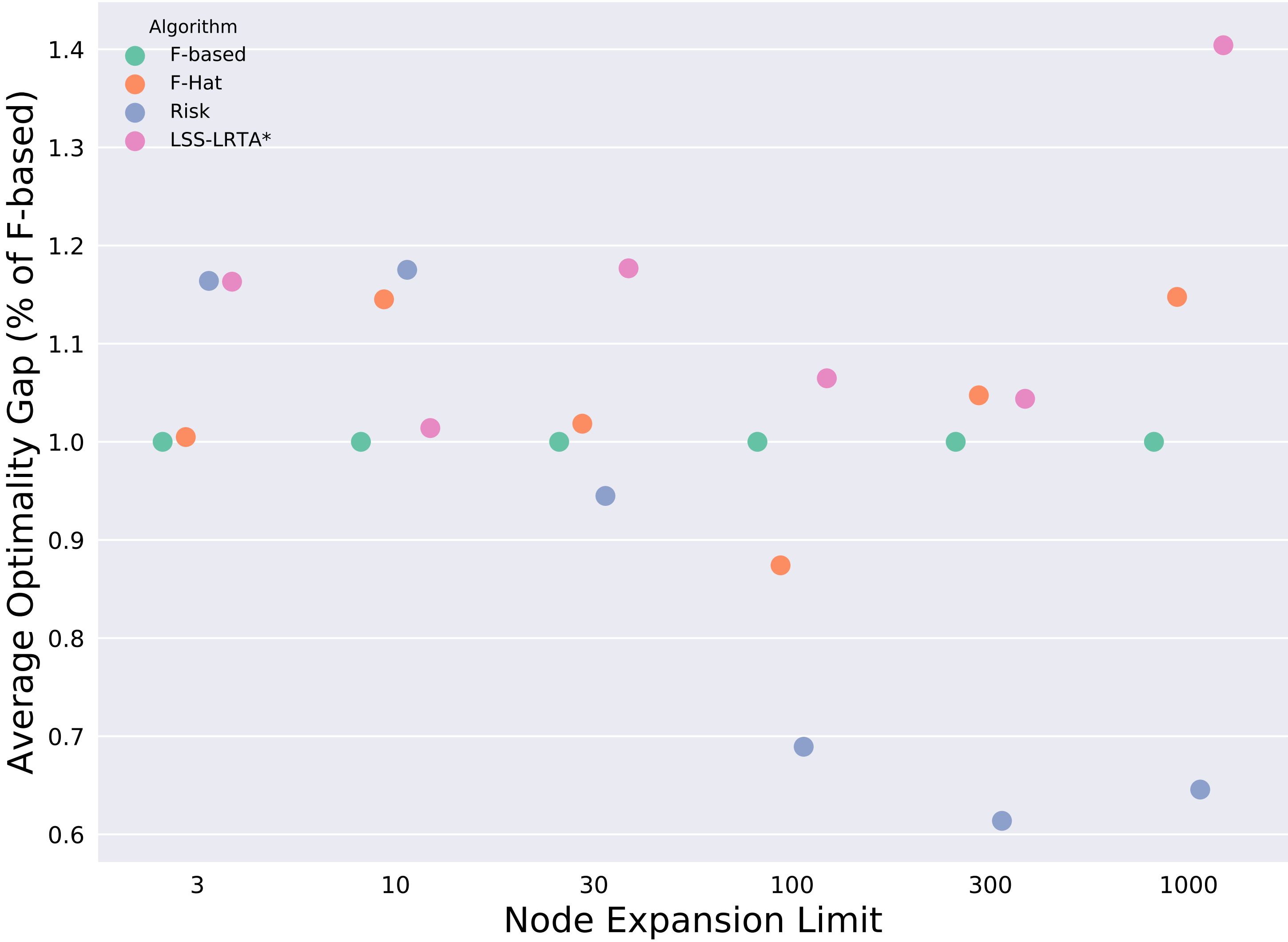
risk performs best!

Expansion Strategies on 15-Puzzle



risk performs best!

Optimality Gap Relative to f-expansion on 15-Puzzle



risk has a fraction of f-based expansion's optimality gap!

Answering the Second Question

1. Which action to select?
   minimum $\hat{f}$ by principle of rationality

Answering the Second Question

1. Which action to select?
   minimum $\hat{f}$ by principle of rationality
2. How to backup from frontier?
   Bellman or Nancy (Nancy has added benefit of belief)

Answering the Second Question

1. Which action to select?
   minimum $\hat{f}$ by principle of rationality
2. How to backup from frontier?
   Bellman or Nancy (Nancy has added benefit of belief)
3. Which nodes to expand?
   those which minimize risk

Answering the Second Question

1. Which action to select?

   minimum $\hat{f}$ by principle of rationality
2. How to backup from frontier?

   Bellman or Nancy (Nancy has added benefit of belief)
3. Which nodes to expand?

   those which minimize risk

Nancy (Nancy backups + risk-based expansion) addresses these questions!

Conclusions

Conclusion

viewed real-time planning as decision-making under uncertainty

- studied 5 backup rules (2 new + 1 reformulated)
- Bellman and Nancy backups performed best
- discussed 4 expansion strategies (1 new)
- risk-based expansion performed best

Nancy (risk-based expansion + Nancy backups) outperforms conventional LSS-LRTA*

Conclusion

viewed real-time planning as decision-making under uncertainty

- studied 5 backup rules (2 new + 1 reformulated)
- Bellman and Nancy backups performed best
- discussed 4 expansion strategies (1 new)
- risk-based expansion performed best

Nancy (risk-based expansion + Nancy backups) outperforms conventional LSS-LRTA*

future directions:

- broader testing
- efficient implementation
- explore similarities with UCT

Conclusion

viewed real-time planning as decision-making under uncertainty

- studied 5 backup rules (2 new + 1 reformulated)
- Bellman and Nancy backups performed best
- discussed 4 expansion strategies (1 new)
- risk-based expansion performed best

Nancy (risk-based expansion + Nancy backups) outperforms conventional LSS-LRTA*

future directions:

- broader testing
- efficient implementation
- explore similarities with UCT

More broadly, metareasoning about uncertainty pays off, even for deterministic domains!