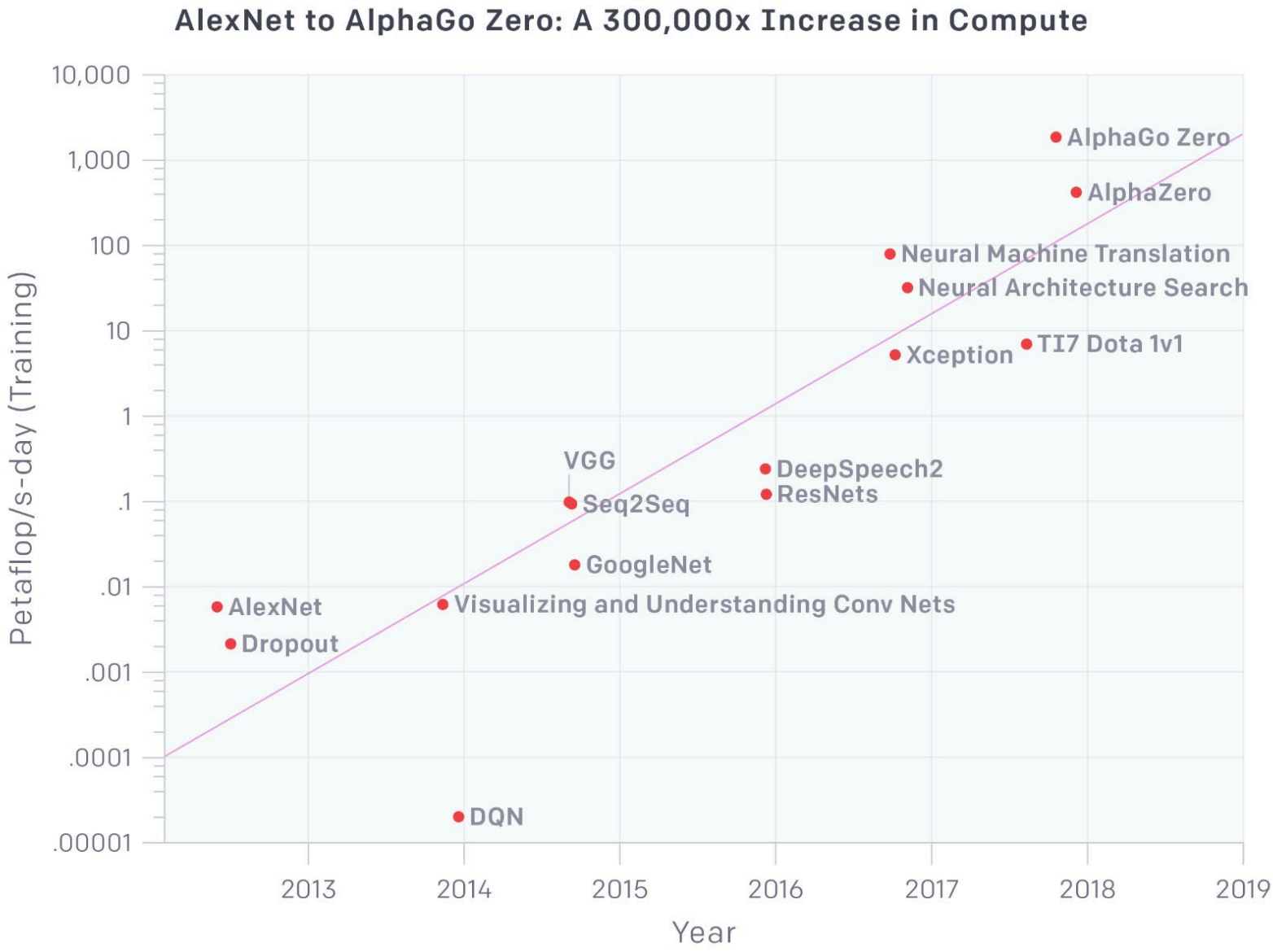
Discovering Optimal Strategies for Bounded Agents

(FA9550-18-1-0077)

PI: Thomas Griffiths (UC Berkeley/Princeton)

AFOSR Program Review:
Computational Cognition and Machine Intelligence Program
(10/7/20, Zoomville)



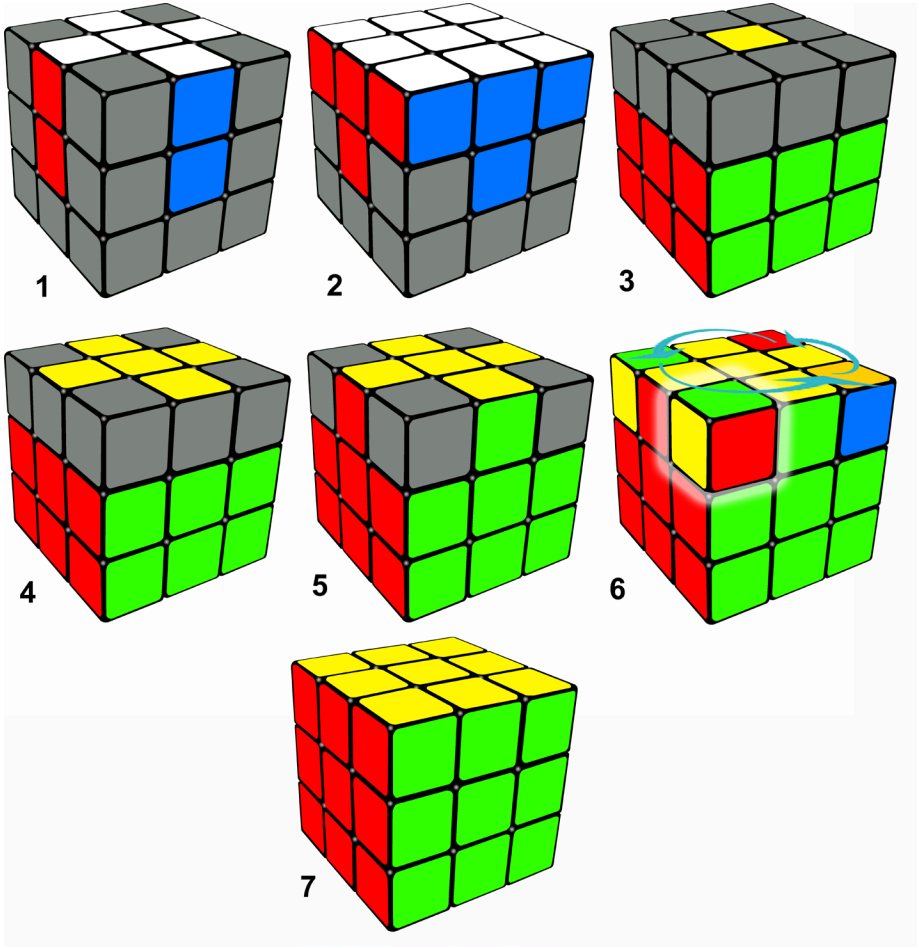
(OpenAI blog post)

Computational efficiency

~1 position/second

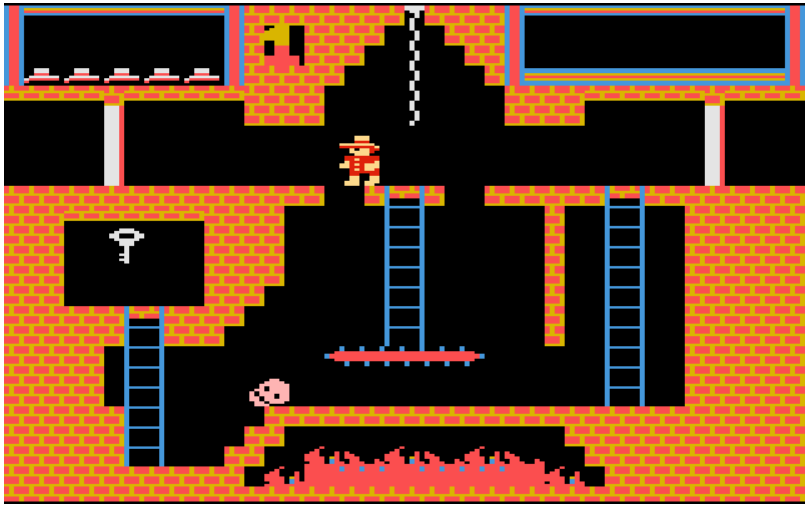
~100,000 positions/second

Representational efficiency



Distance	Count of Positions
0	1
1	18
2	243
3	3,240
4	43,239
5	574,908
6	7,618,438
7	100,803,036
8	1,332,343,288
9	17,596,479,795
10	232,248,063,316
11	3,063,288,809,012
12	40,374,425,656,248
13	531,653,418,284,628
14	6,989,320,578,825,358
15	91,365,146,187,124,313
16	about 1,100,000,000,000,000,000
17	about 12,000,000,000,000,000,000
18	about 29,000,000,000,000,000,000
19	about 1,500,000,000,000,000,000
20	about 300,000,000

Understanding human intelligence



Optimal Strategies for Bounded Agents (Griffiths)

Research Objectives:

Making decisions in real environments requires effective use of limited computational resources. This work explores how people identify efficient cognitive strategies for decision-making, problem-solving and reasoning, resulting in methods that can be used to improve automated systems.

Technical Approach:

Making effective use of limited resources can be formulated as a sequential decision problem, allowing us to use tools from reinforcement learning and hierarchical reinforcement learning to derive efficient strategies that we compare against human behavior.

Key Scientific Contributions:

This work builds connections between research on rational metareasoning in AI and understanding of human behavior, and aims to improve on the state of the art in meta-reasoning, hierarchical reinforcement learning, and meta-programming.

DoD Benefits:

The results of this research will include methods for determining how human decisions are affected by time pressure and cognitive load, as well as techniques for improving automated systems.

List of Project Goals

1. Formulate human metareasoning problems as sequential decision problems.

2. Determine how people identify high-level actions and use this to develop methods for discovering simple strategies.

3. Determine how people formulate algorithms and use this to develop methods for discovering complex strategies.

List of Project Goals

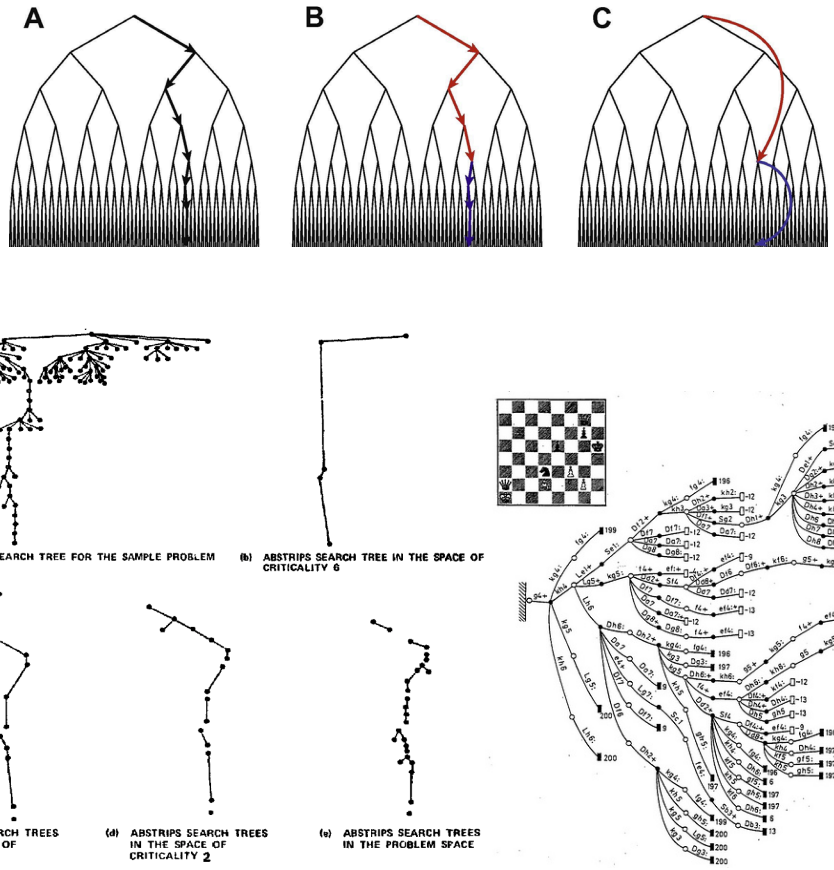
1. Formulate human metareasoning problems as sequential decision problems.

2. Determine how people identify high-level actions and use this to develop methods for discovering simple strategies.

3. Determine how people formulate algorithms and use this to develop methods for discovering complex strategies.

Strategies for Model-based Planning

- **Problem**:
  The curse of dimensionality
  - Time horizon
  - State/Transition Complexity
  - Uncertainty
  - Other people
- **Solutions:**
  - Hierarchies (Sacerdoti, 1974; Botvinick, Niv & Barto, 2009)
  - State Abstraction (Givan, Dean & Greig, 2003)
  - Heuristics (Newell & Simon, 1972)
  - Tree Search Strategies (Huys et al. 2012; Keramati et al., 2016)
- What makes a good solution?



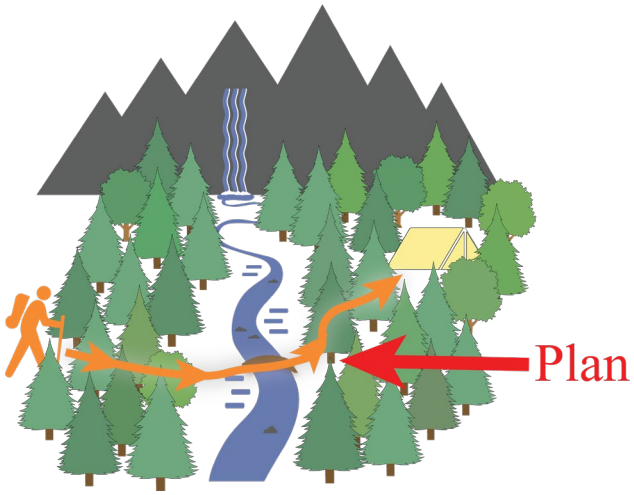
Meta-Reasoning and Planning

- Meta-reasoning → reasoning about a decision-making process (Russell & Wefald, 1992)
- Tractable planning as a **resource-rational problem** (Griffiths, Lieder & Goodman, 2015)
  - How should an agent allocate limited computational resources to achieve her goals?
- Goal: A normative, resource-rational account of partial planning over time

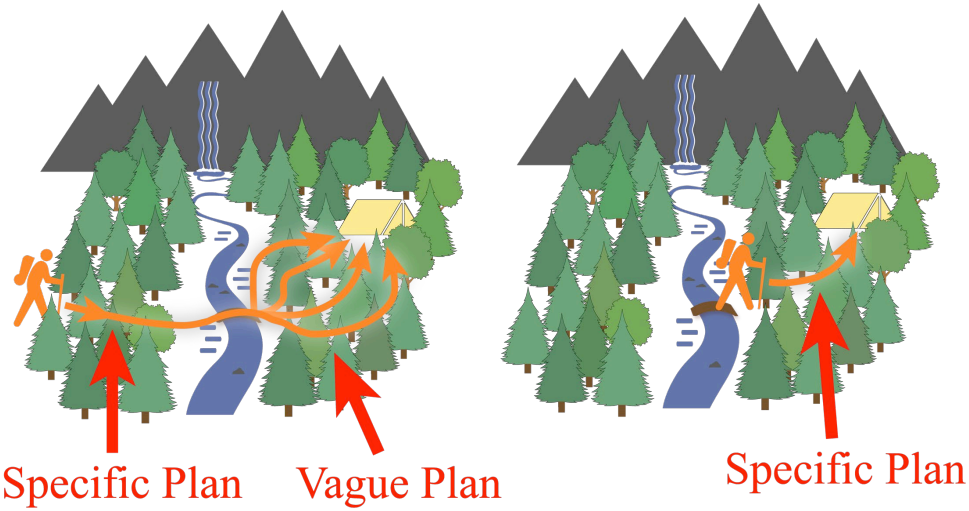
Meta-Reasoning and Planning

Proposal: Human planning involves adaptive construction of partial plans over time.

Standard Planning
How do I need to **act** at each moment in time?

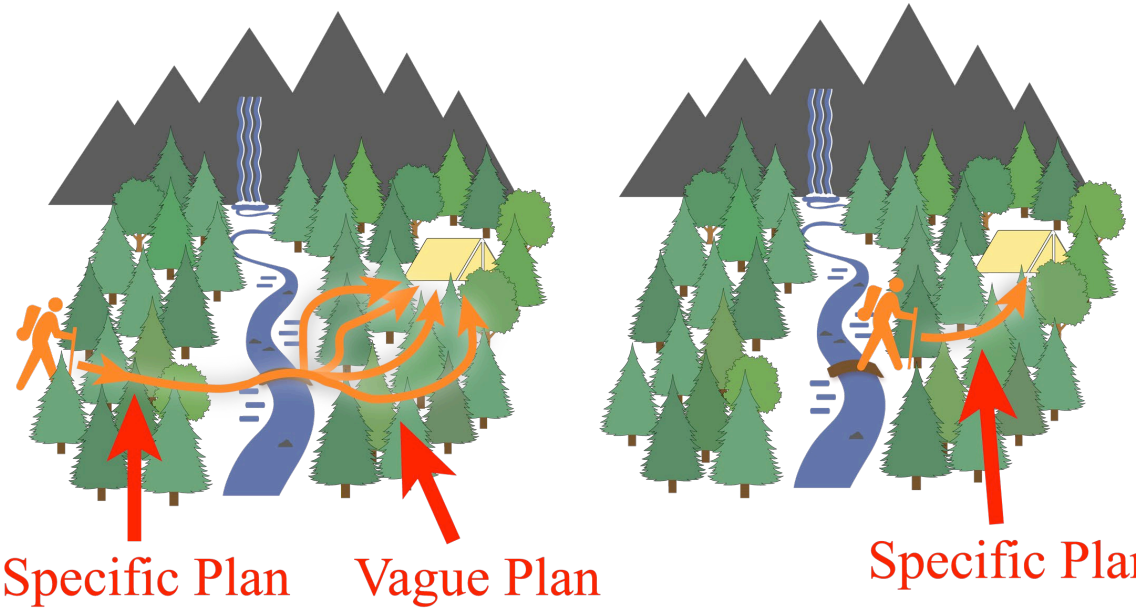


Meta-Planning
How do I need to **plan** at each moment in time?



Ho, Abel, Cohen, Littman, & Griffiths (2020)

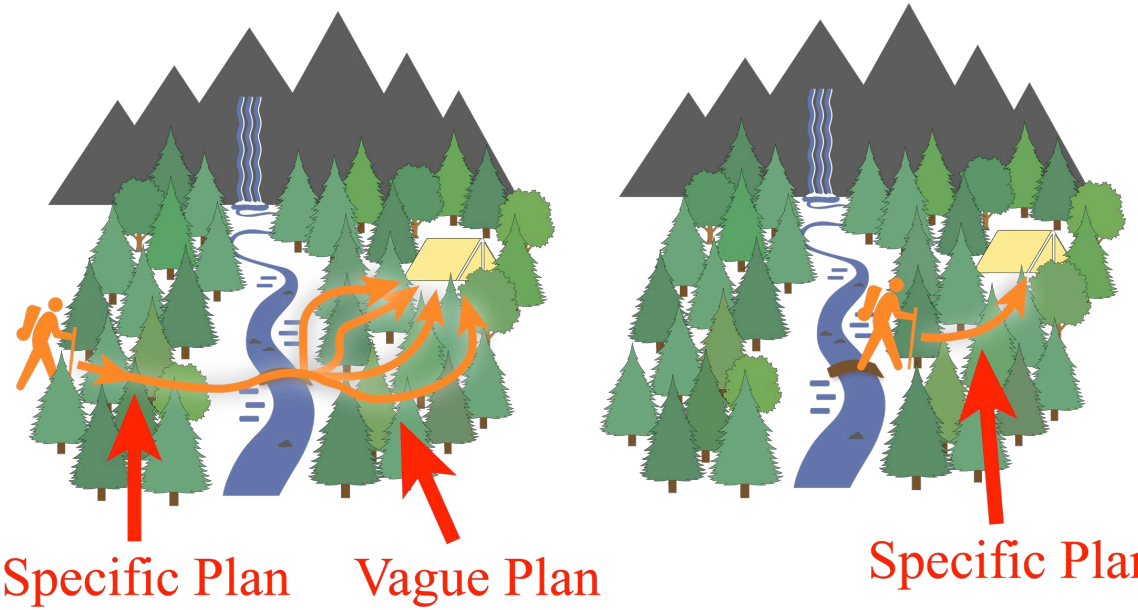
Meta-Reasoning and Planning



Requires formalizing two ideas:

1. *Partial plans* to control specificity/vagueness
2. *Costs* associated with partial plans

Meta-Reasoning and Planning



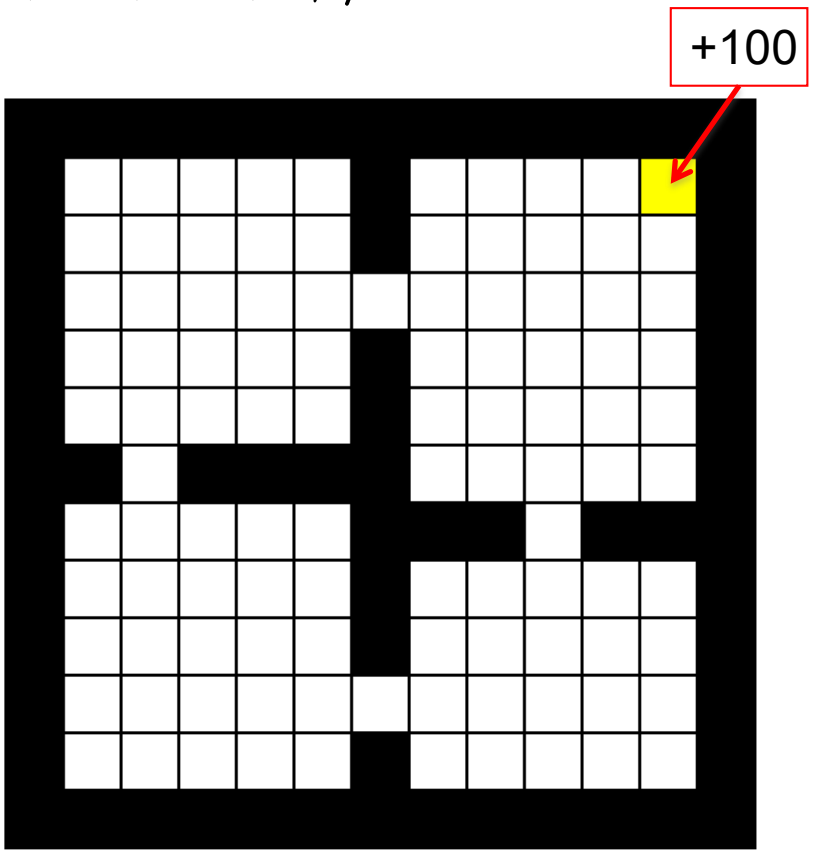
Requires formalizing two ideas:

1. *Partial plans* to control specificity/vagueness
2. *Costs* associated with partial plans

Markov Decision Processes

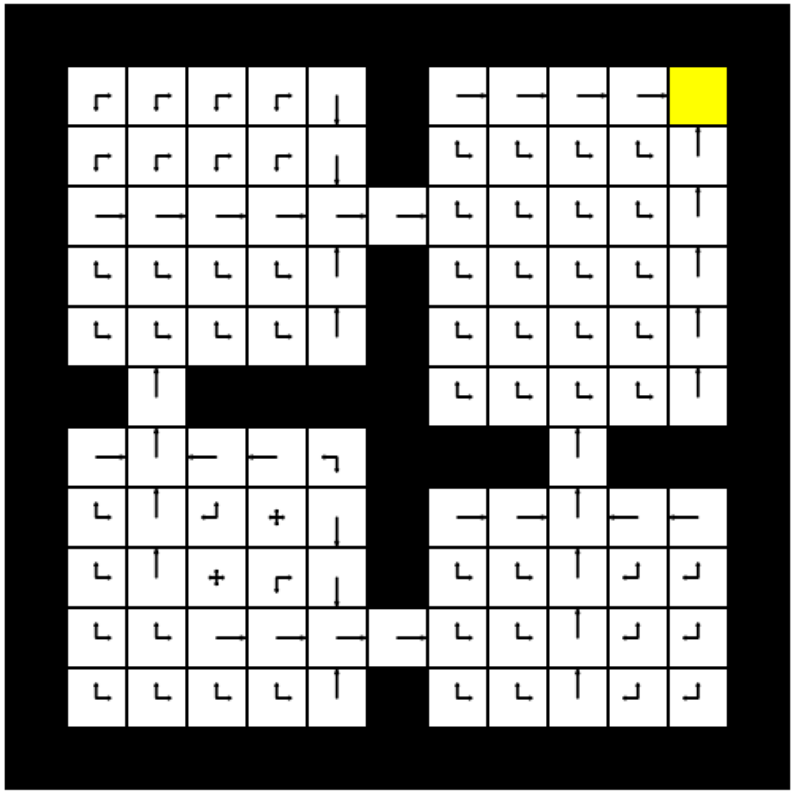
$$M = \langle S, A, T, R, \gamma \rangle$$

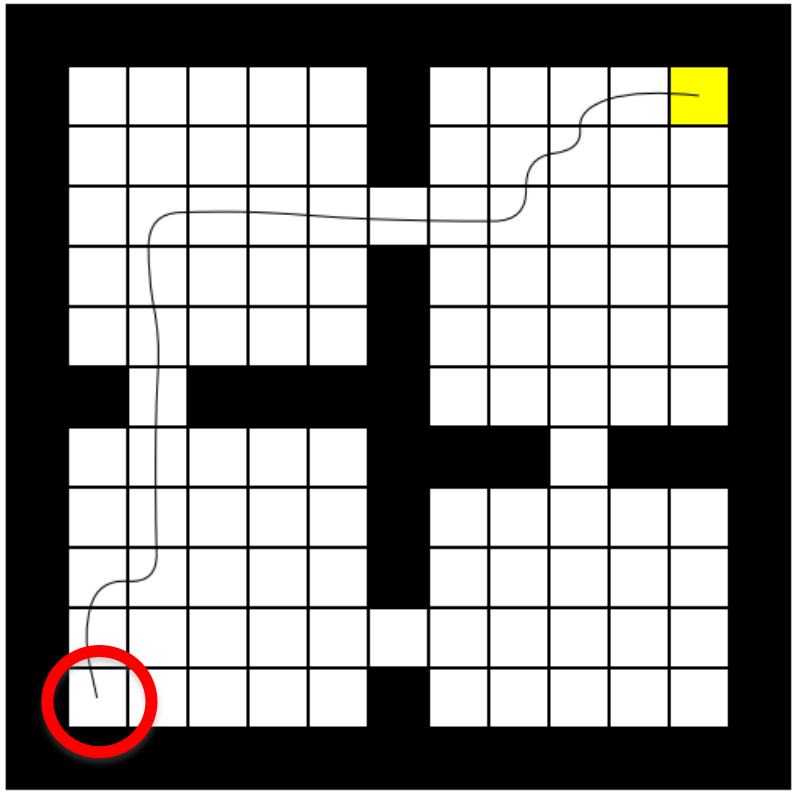
- S = States
- A = Actions
- T = Transitions
- R = Rewards
- γ = Discount Rate



Optimal Decision-Making

Solving an MDP = Finding the optimal policy





Optimal (Full) Plans

Bellman equations express the optimal value

$$V^*(s) = \max_a \left\{ R_{s,a} + \gamma \sum_{s'} P^a_{s,s'} V^*(s') \right\}$$

Best Future Reward From **current** state

Best Action

Immediate Reward

Best Future Reward From **possible next state**

- Best immediate action depends on best future actions in **all possible future states**
- Optimal planning involves:
  - Choosing the best action at a state
  - Identifying best actions at future states

Formalizing Partial Planning

Intuition: Relax maximization at future states

State-specific
Weight/Temperature

$$V^{\beta}(s) = \max_{\pi} \left\{ \sum_{a} \pi(a \mid s) Q^{\beta}(s,a) - \beta(s) D_{\mathrm{KL}} \left[\pi(\cdot \mid s) || \bar{\pi}(\cdot \mid s) \right] \right\}$$

Future Expected Value

Prior Policy

$$\pi^{\beta}(a \mid s) \propto \bar{\pi}(a \mid s) \exp \left\{ \frac{1}{\beta(s)} Q^{\beta}(s,a) \right\}$$

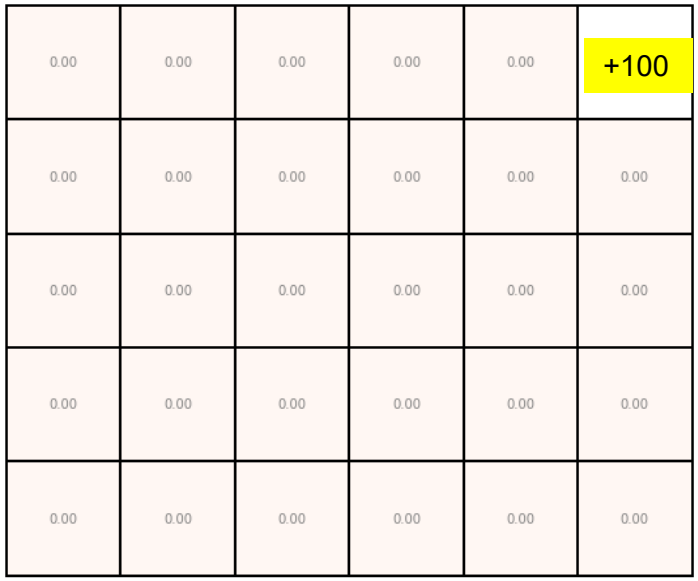
Prior Policy

State-specific
Inverse Temperature

Formalizing Partial Planning

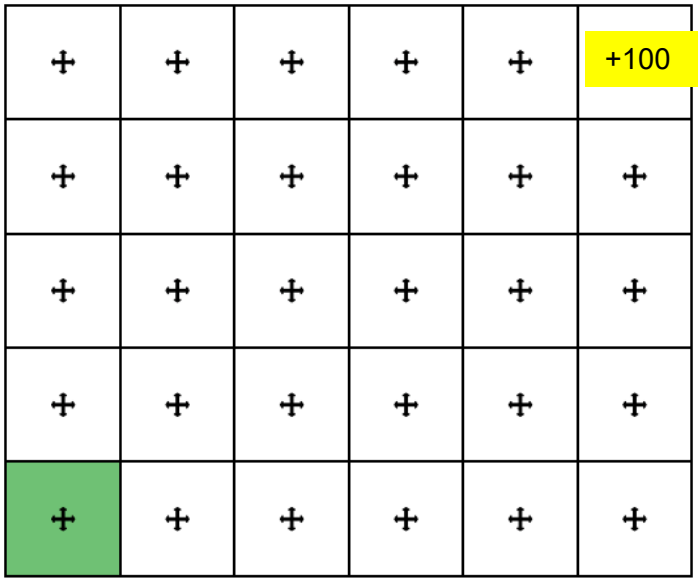
Inverse Temperature

$\beta(s)^{-1}$



Partial Policy

$\pi^{\beta}(a \mid s)$

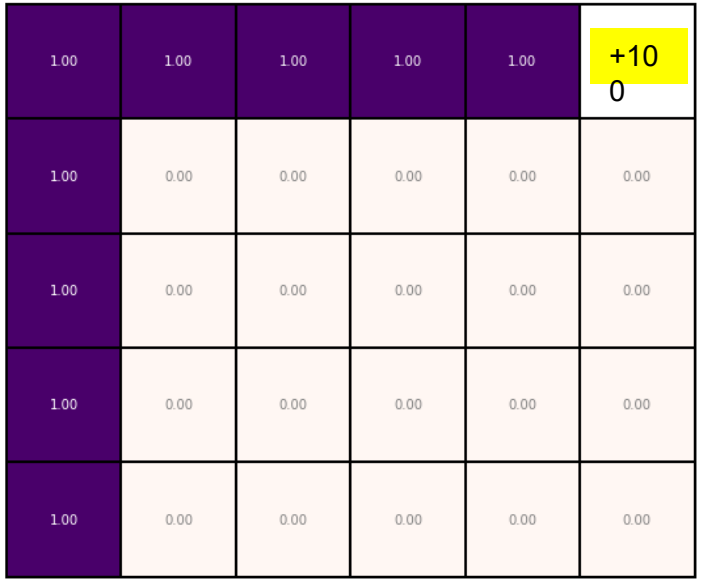


Prior policy (uniform over actions) completely dominates

Formalizing Partial Planning

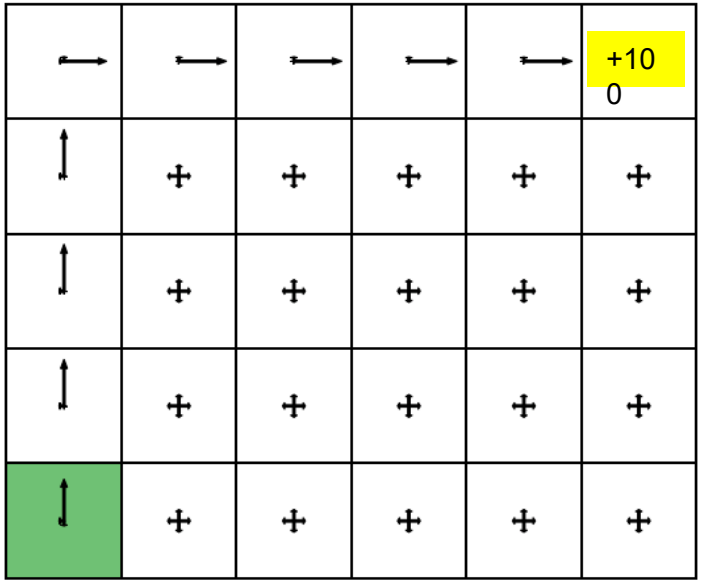
Inverse Temperature

$\beta(s)^{-1}$



Partial Policy

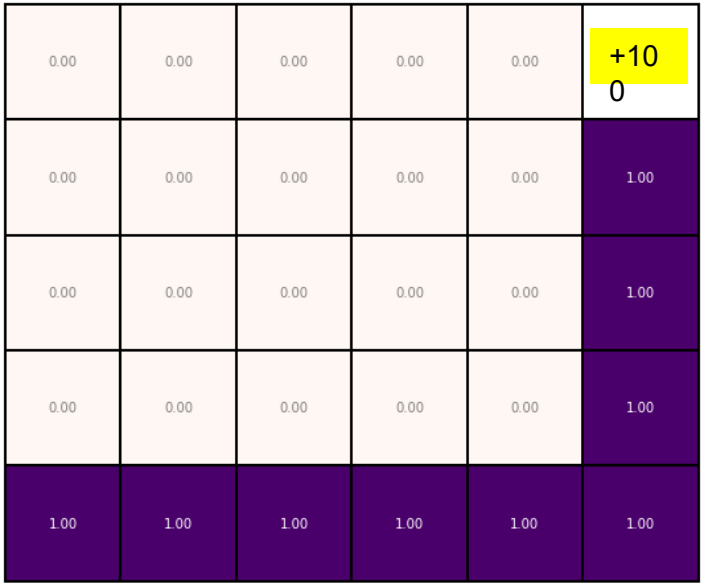
$\pi^{\beta}(a \mid s)$



Formalizing Partial Planning

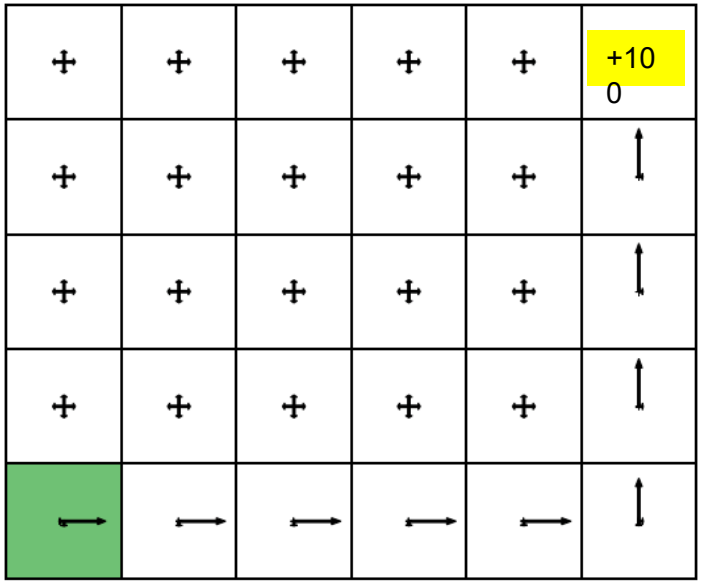
Inverse Temperature

$\beta(s)^{-1}$



Partial Policy

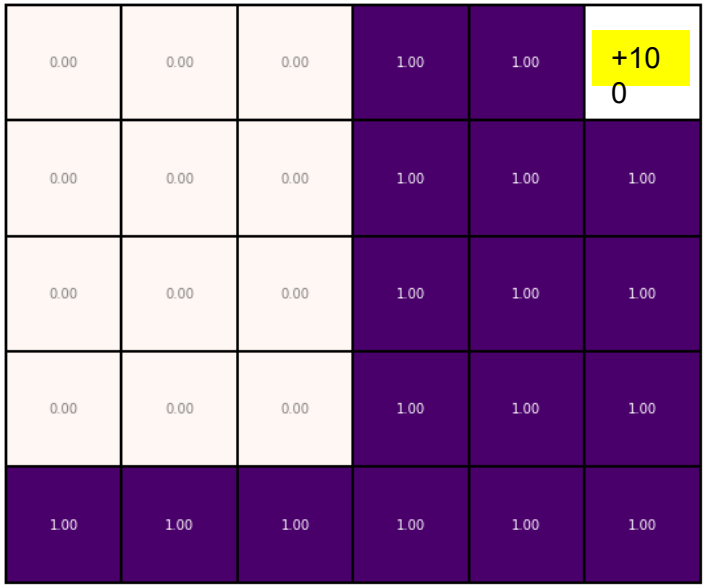
$\pi^{\beta}(a \mid s)$



Formalizing Partial Planning

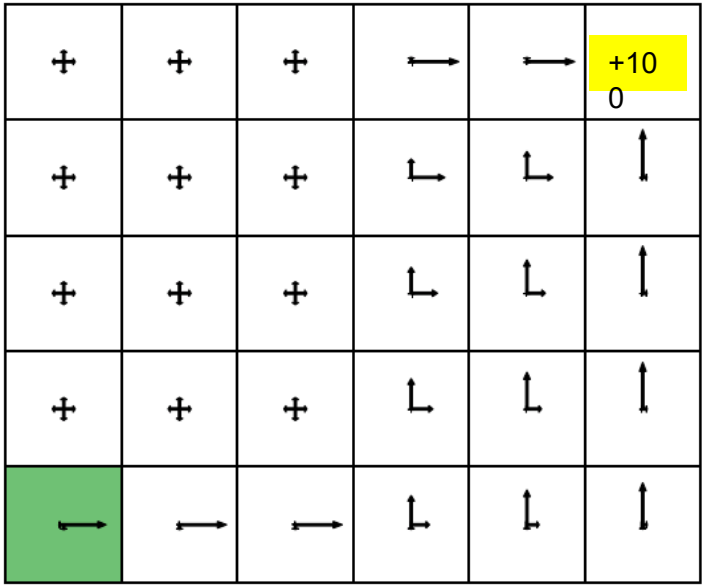
Inverse Temperature

$\beta(s)^{-1}$



Partial Policy

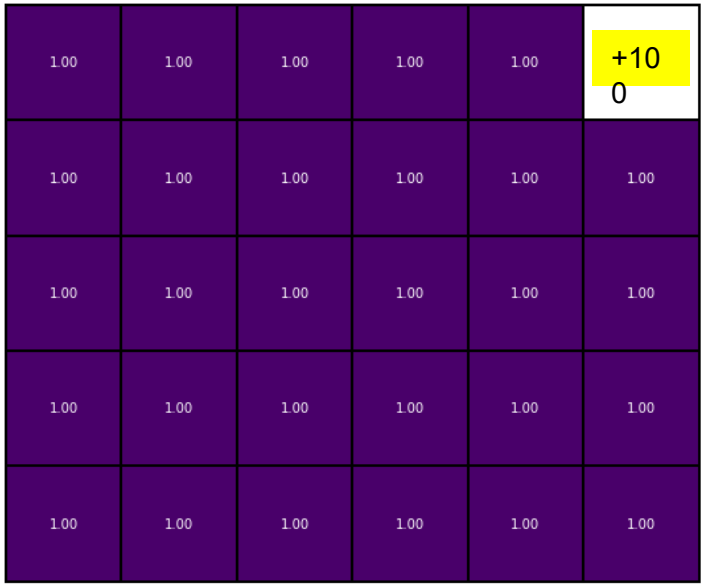
$\pi^{\beta}(a \mid s)$



Formalizing Partial Planning

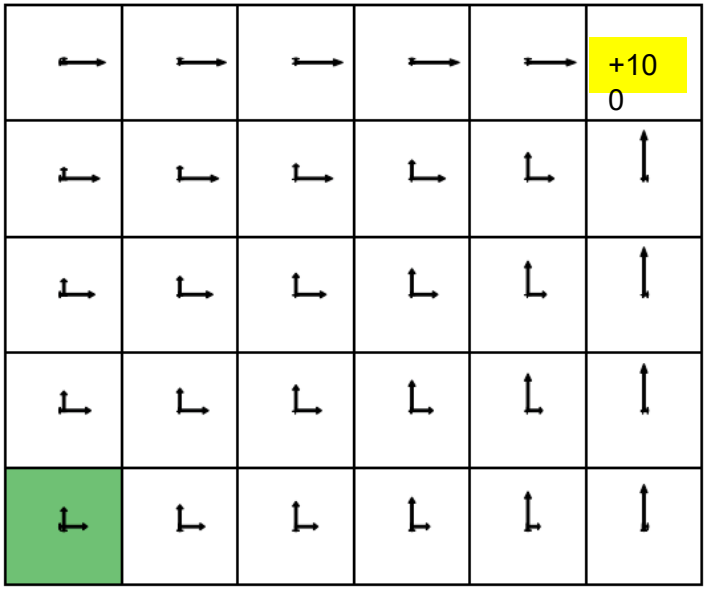
Inverse Temperature

$\beta(s)^{-1}$

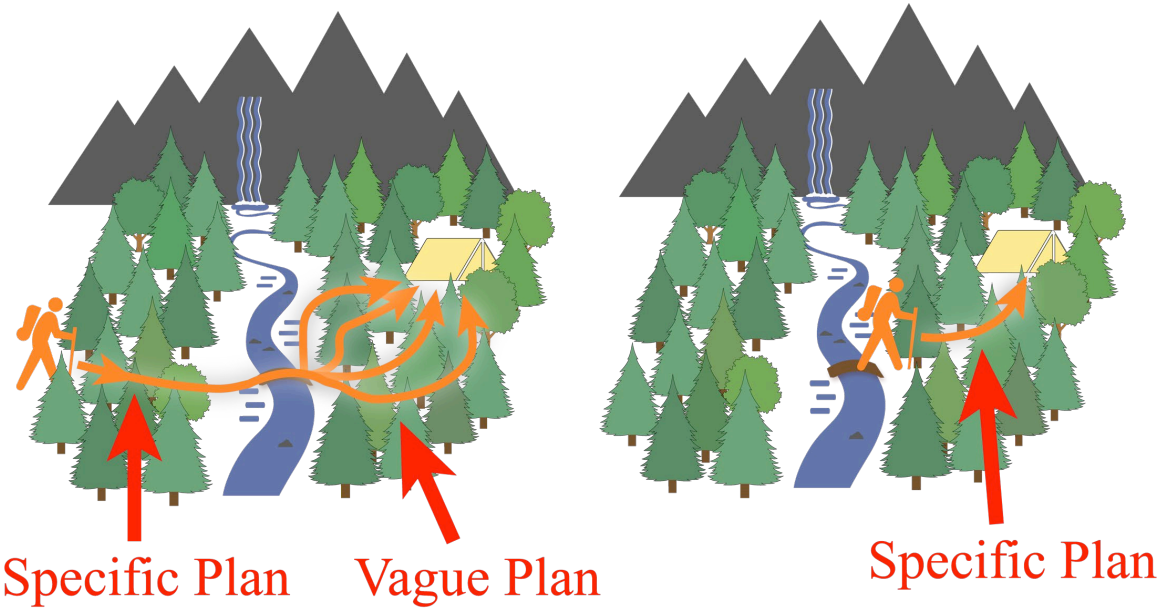


Partial Policy

$\pi^{\beta}(a \mid s)$



Meta-Reasoning and Planning



Requires formalizing two ideas:

1. *Partial plans* to control specificity/vagueness
2. *Costs* associated with partial plans

Information Theoretic Cost of Partial Plan

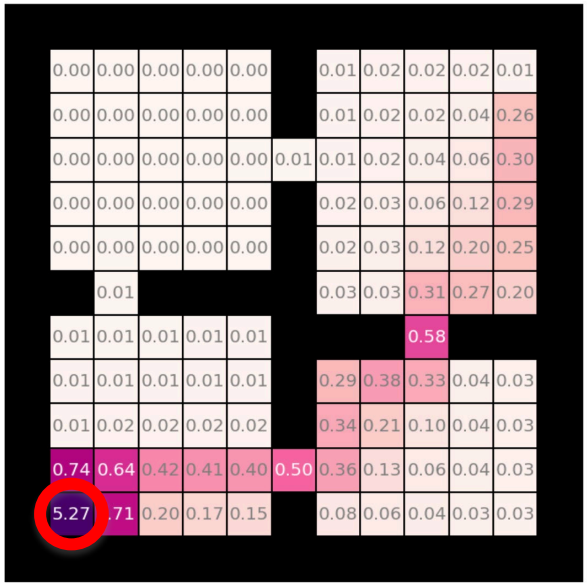
- Goal: Quantify the cost of a partial plan
  - The information theoretic cost in the existing formulation provides a natural cost
  - Sum of KL-divergences over the entire policy

$$C(\pi^{\beta}, \bar{\pi}) = \sum_{\tilde{s}} D_{\mathrm{KL}} \left[\pi^{\beta}(\cdot \mid \tilde{s}) || \bar{\pi}(\cdot \mid \tilde{s}) \right]$$

Partial Planning + Information Theoretic Costs

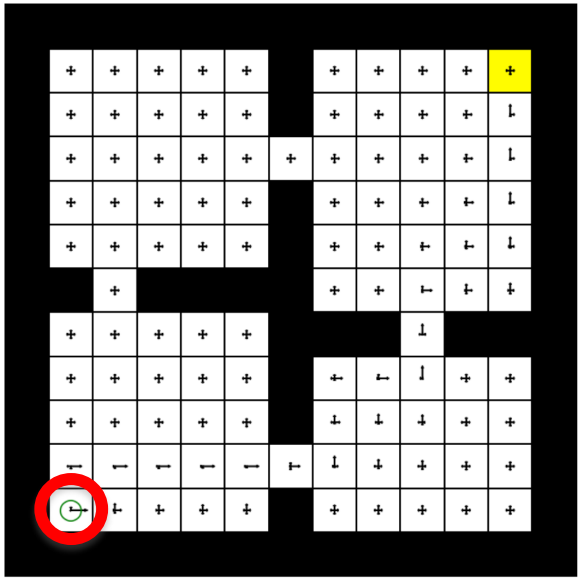
Inverse temperatures over simulated states

$\beta(\tilde{s})^{-1}$



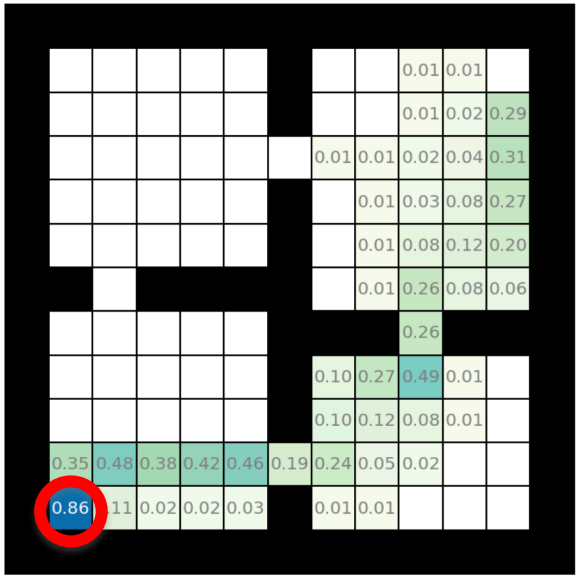
Partial Plan

$\pi^{\beta}(a \mid \tilde{s})$



Information Theoretic Cost

$D_{\mathrm{KL}}\left[\pi^{\beta}(\cdot \mid \tilde{s}) || \bar{\pi}(\cdot \mid \tilde{s})\right]$



$\bigcirc \rightarrow s = \tilde{s}$

(i.e., the state that planning is currently being done at)

$C(\tilde{\pi}, \bar{\pi}) = 6.22$

Meta-Reasoning about Partial Plans

Meta-planner optimizes Parameters of partial plan

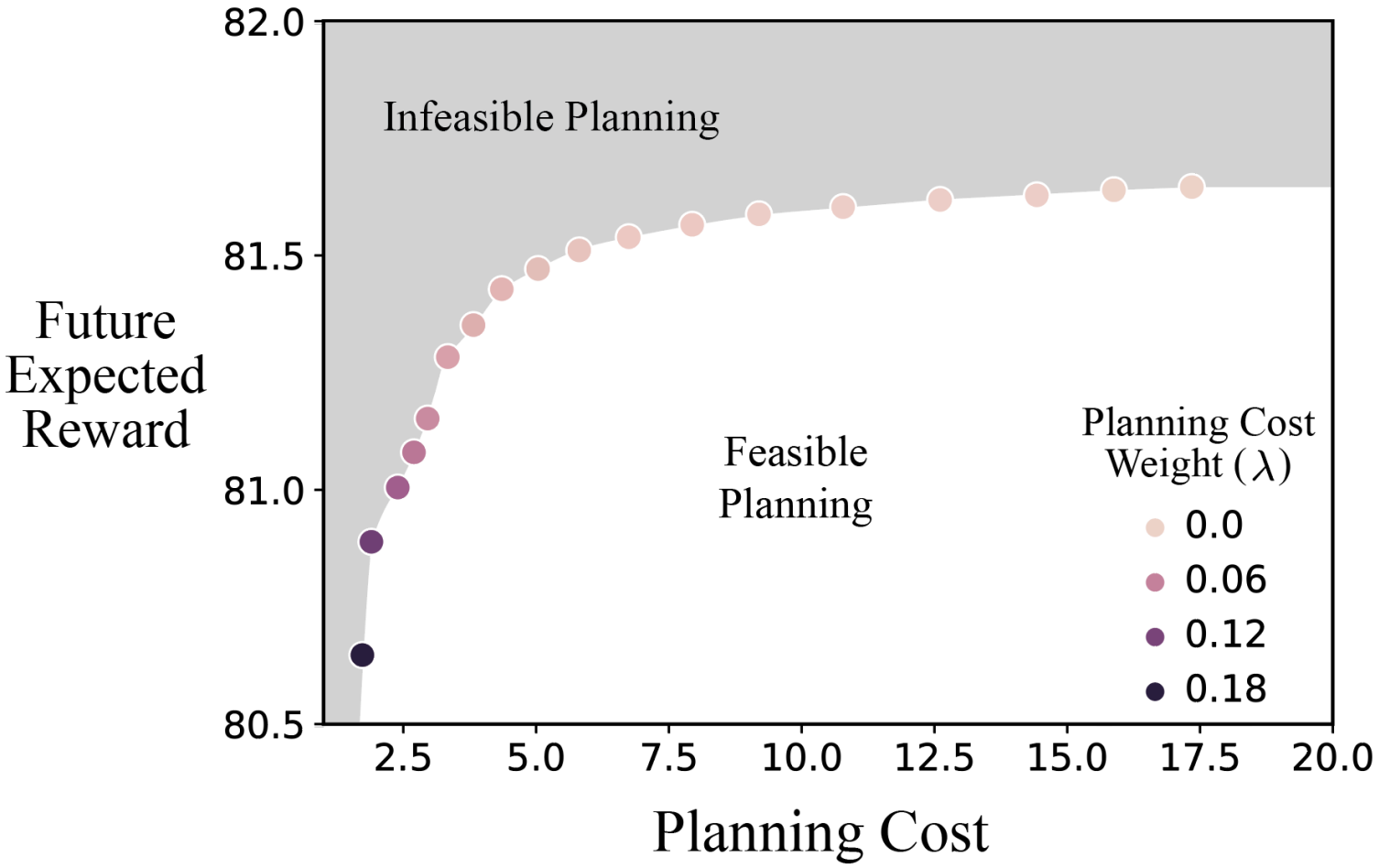
$$V^*_{\lambda,\bar{\pi}}(s) = \max_{\beta} \left\{ \sum_a \pi^{\beta}(a \mid \tilde{s} = s) Q^*_{\lambda,\bar{\pi}}(s,a) - \lambda C(\pi^{\beta}, \bar{\pi}) \right\}$$

Immediate action results from planning

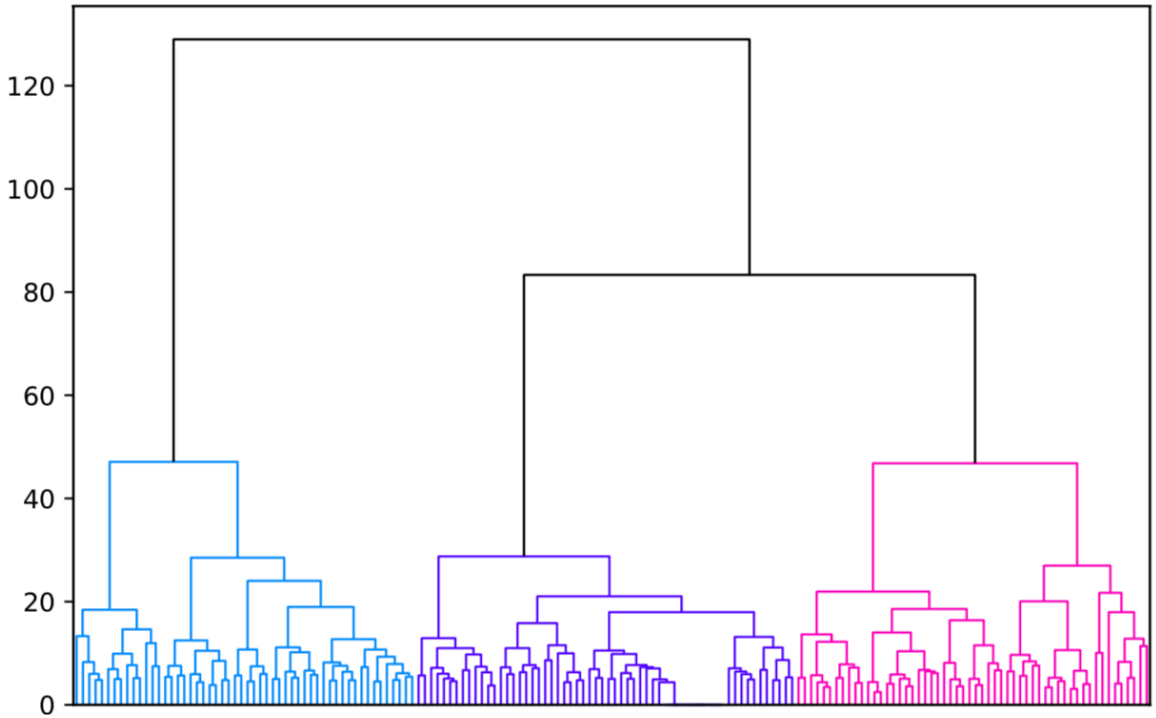
Information Theoretic cost of partial plan

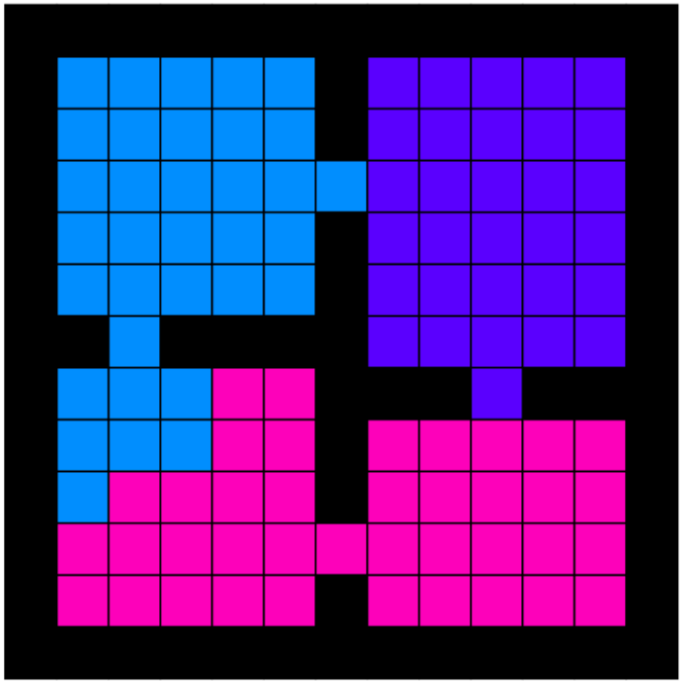
- λ trades off reward and information at the meta-planning level
- β trades off reward and information at the planning level
- Future planning and future planning costs are taken into account

Tradeoff between Behavioral Optimality and Partial Plan Cost



Clustering based on partial plan similarity

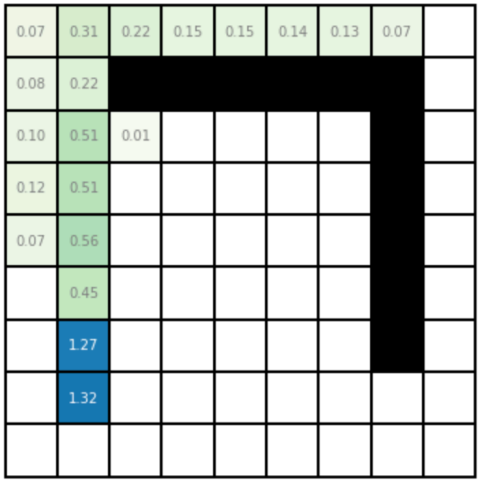


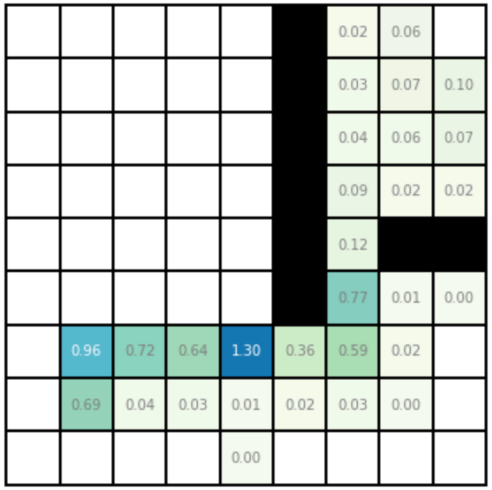


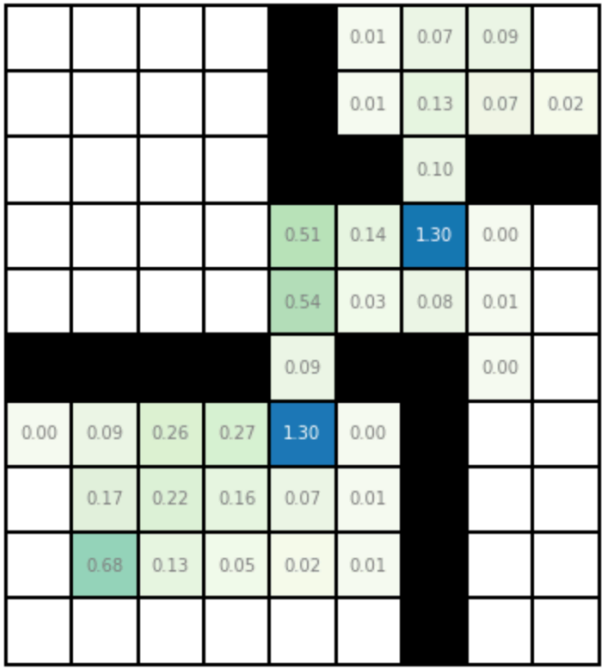
(Symmetric sum of KL-divergences)

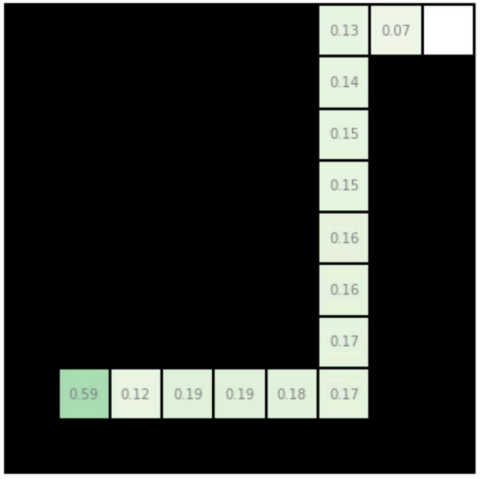
Not just bottlenecks

(Şimşek & Barto, 2009; Solway et al., 2014)

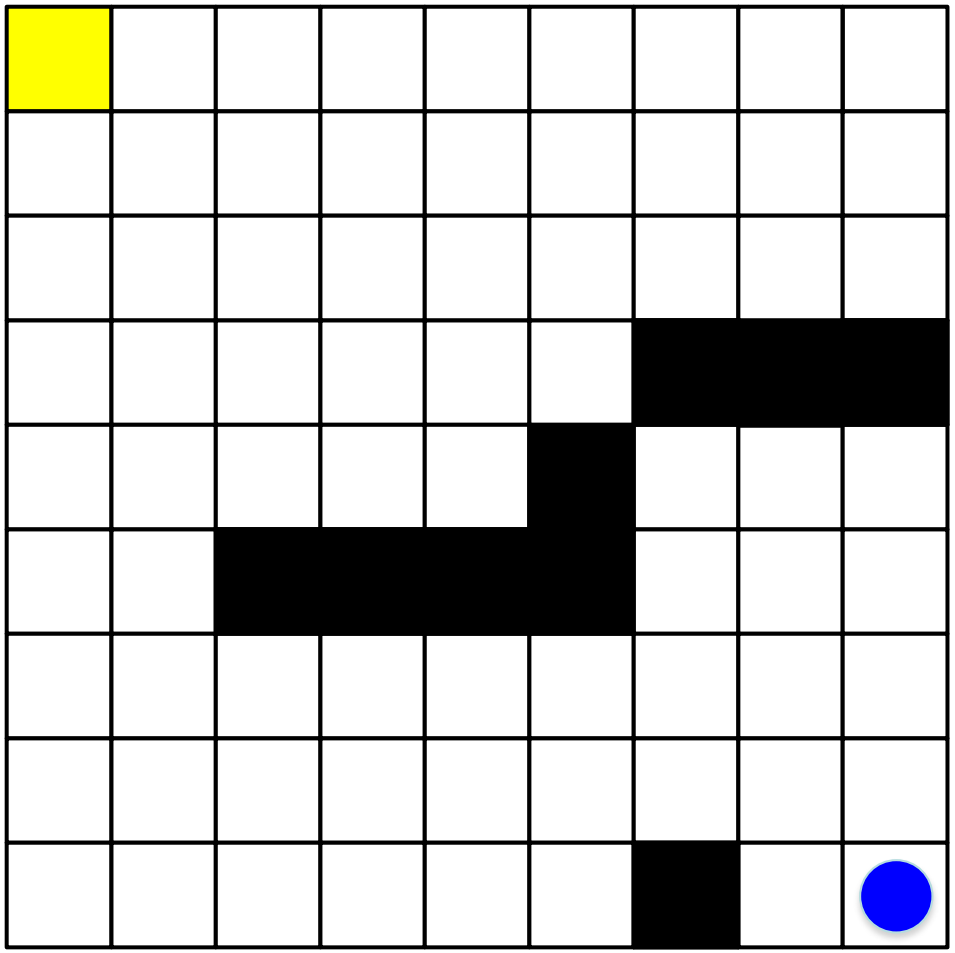






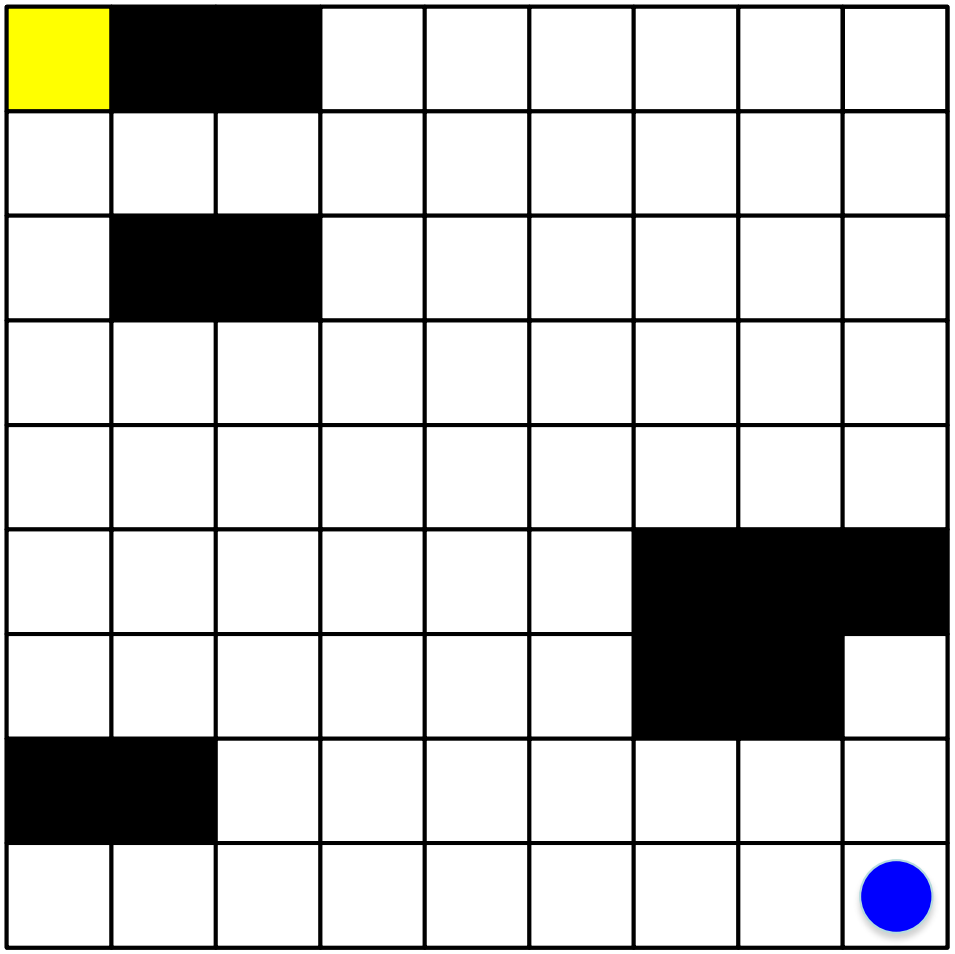


Experiment: Parametric Mazes

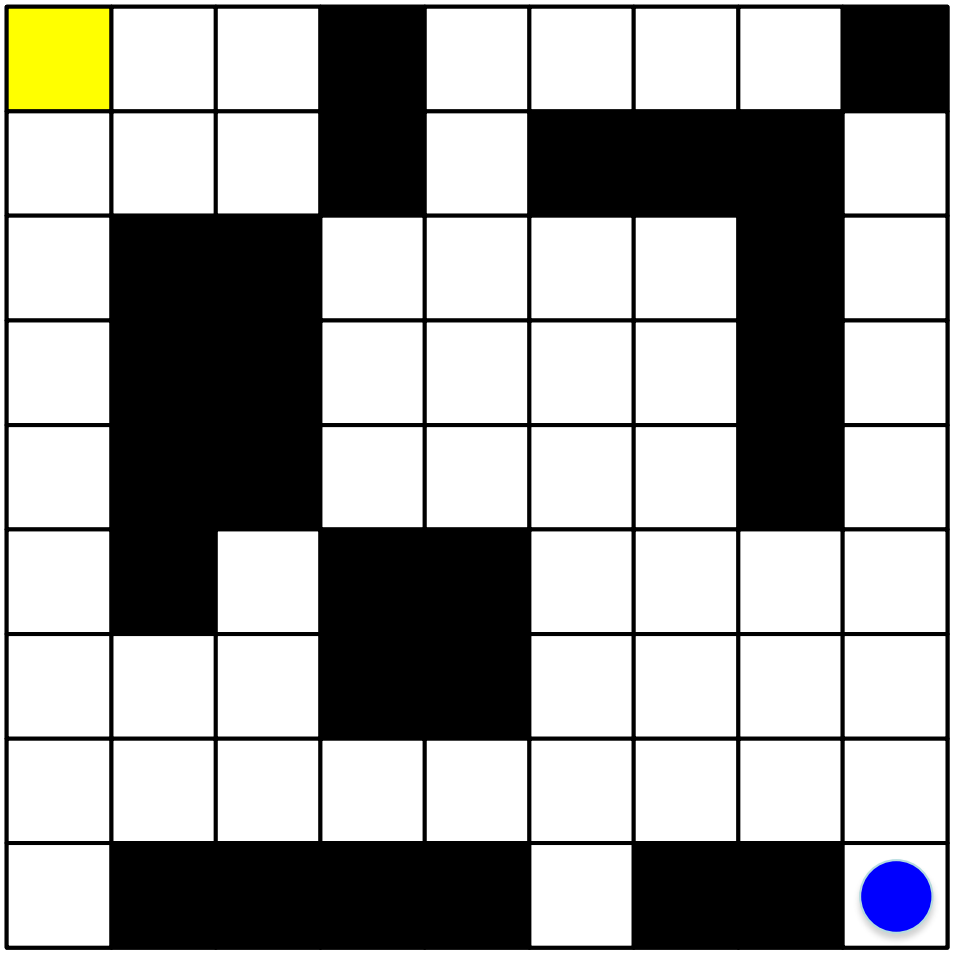


Get to the goal as quickly as possible

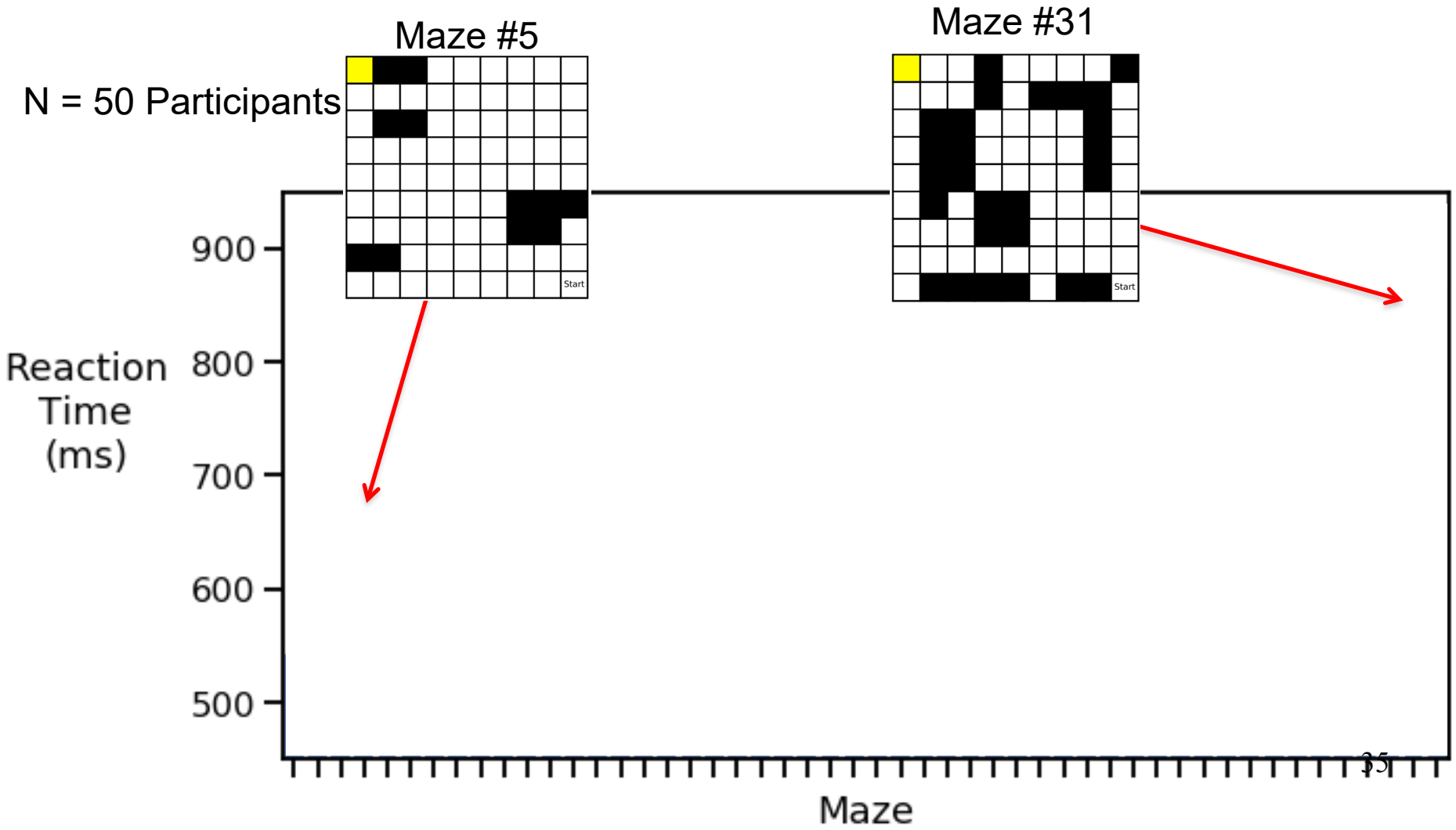
Maze #5



Maze #31

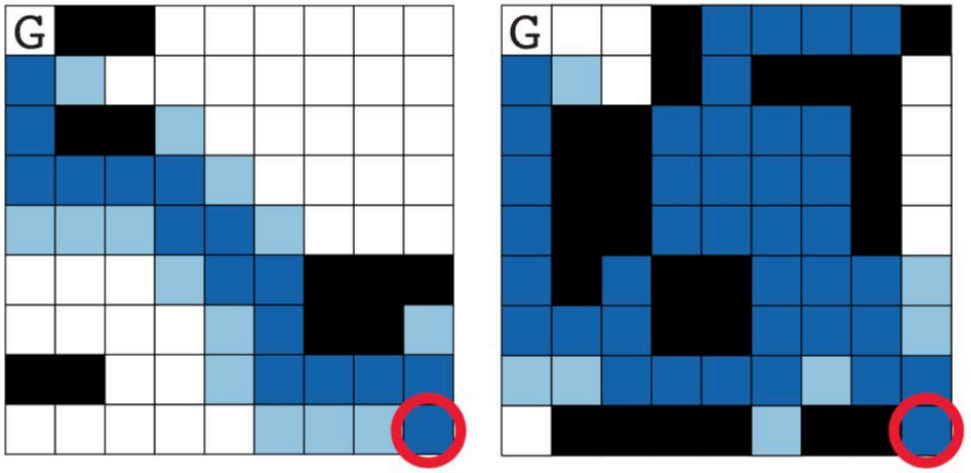


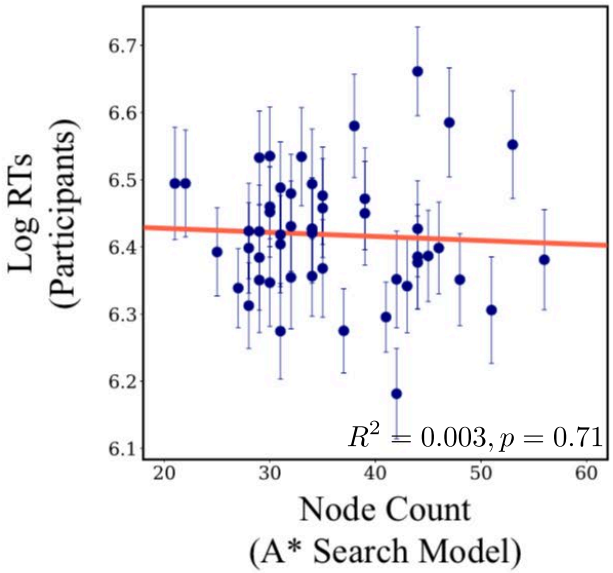
How long does the **first action** take?



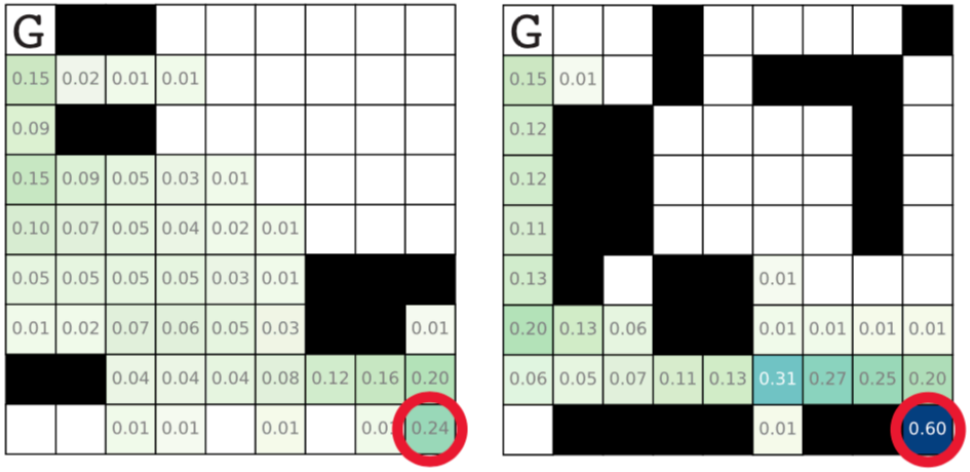
Simple Planning?

- A* Search with Manhattan distance heuristic
  - Opened nodes = reaction time?
- Captures some intuitions, but overall does not explain human reaction times very well

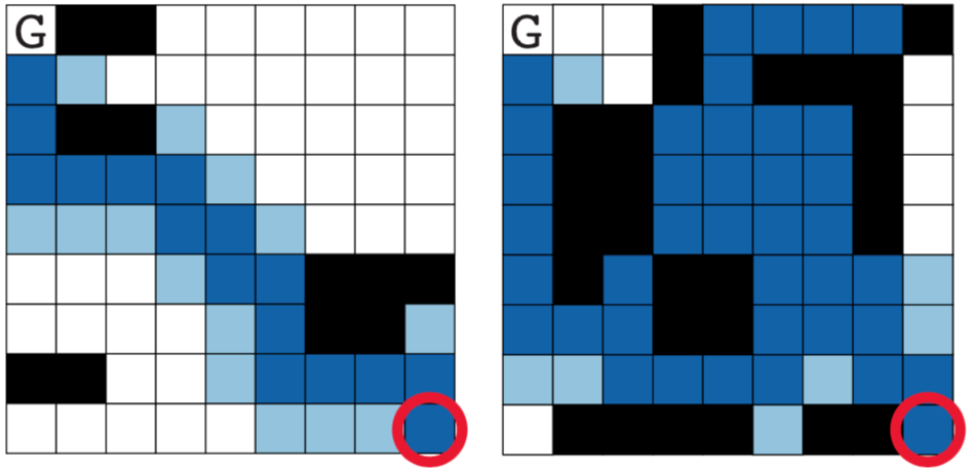




Partial Planning Model



Planning to Plan



Nodes opened by A* with Euclidean Distance Heuristic

Ongoing work

- Identifying subgoals based on minimizing computational costs (Correa et al., 2020)
- Identifying representations of problems that support efficient solutions
- Extending the “resource rational” approach to identifying subroutines in simple algorithms

Conclusions

- Finding computationally and representationally efficient strategies is key to human intelligence
- Resource rationality gives us a way to formalize the kinds of solutions people seem to find
- Capturing this capacity can help us understand what computationally/representationally efficient policies look like as a target for machines

List of Publications, Awards, Honors, etc. Attributed to the Grant

- Sanborn, S., Bourgin, D. D., Chang, M., & Griffiths, T. L. (2018) Representational efficiency outweighs action efficiency in human program induction. *Proceedings of the 40th Annual Conference of the Cognitive Science Society.*
- Chang, M. B., Gupta, A., Levine, S., & Griffiths, T. L. (2018). Automatically composing representation transformations as a means for generalization. *International Conference on Learning Representations (ICLR)*.
- Griffiths, T. L., Callaway, F., Chang, M. B., Grant, E., Krueger, P. M., & Lieder, F. (2019). Doing more with less: meta-reasoning and meta-learning in humans and machines. *Current Opinion in Behavioral Sciences, 29*, 24-30
- Ho, M. K., Abel, D., Cohen, J. D., Littman, M. L., & Griffiths, T. L. (2020). The Efficiency of Human Cognition Reflects Planned Information Processing. *Proceedings of the 34th AAAI Conference on Artificial Intelligence.*
- Correa, C. G.*, Ho, M. K.*, Callaway, F., & Griffiths, T. L. (2020). Resource-rational Task Decomposition to Minimize Planning Costs. *Proceedings of the 42nd Annual Conference of the Cognitive Science Society.*
- Griffiths, T.L. (in press). Understanding human intelligence via human limitations. *Trends in Cognitive Sciences.*