PLANNING TO PLAN: A BAYESIAN MODEL FOR OPTIMIZING THE DEPTH OF DECISION TREE SEARCH

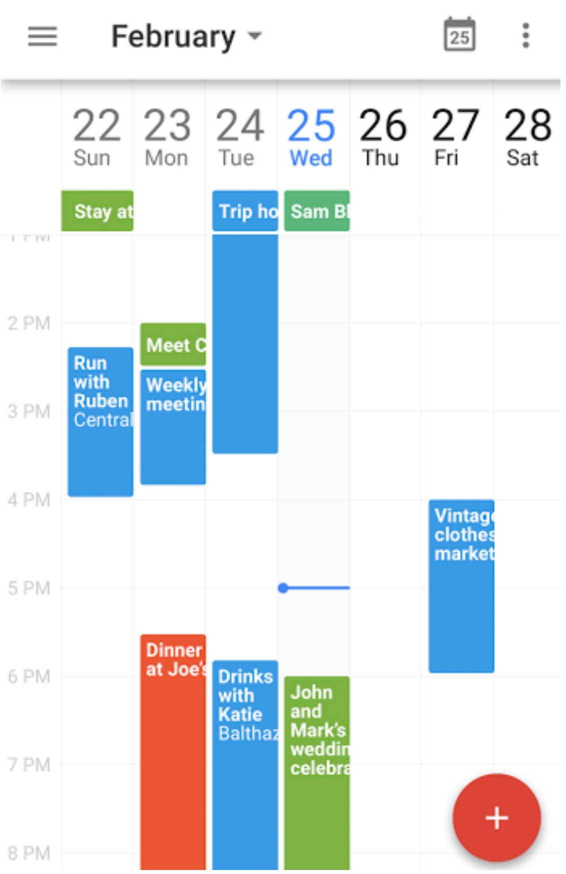
Ionatan Kuperwajs

Ma Lab
Center for Neural Science
New York University

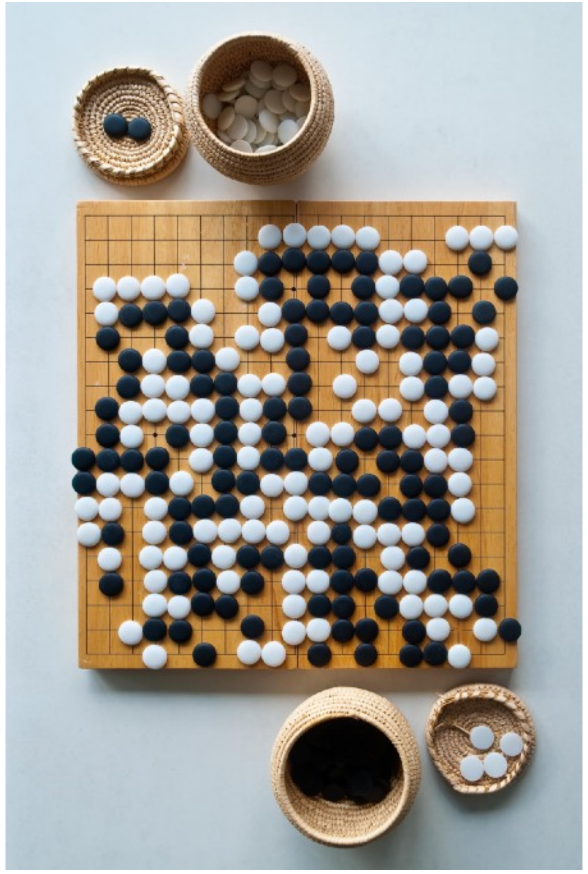
PLANNING

Planning is a hallmark of human intelligence

Spatial navigation



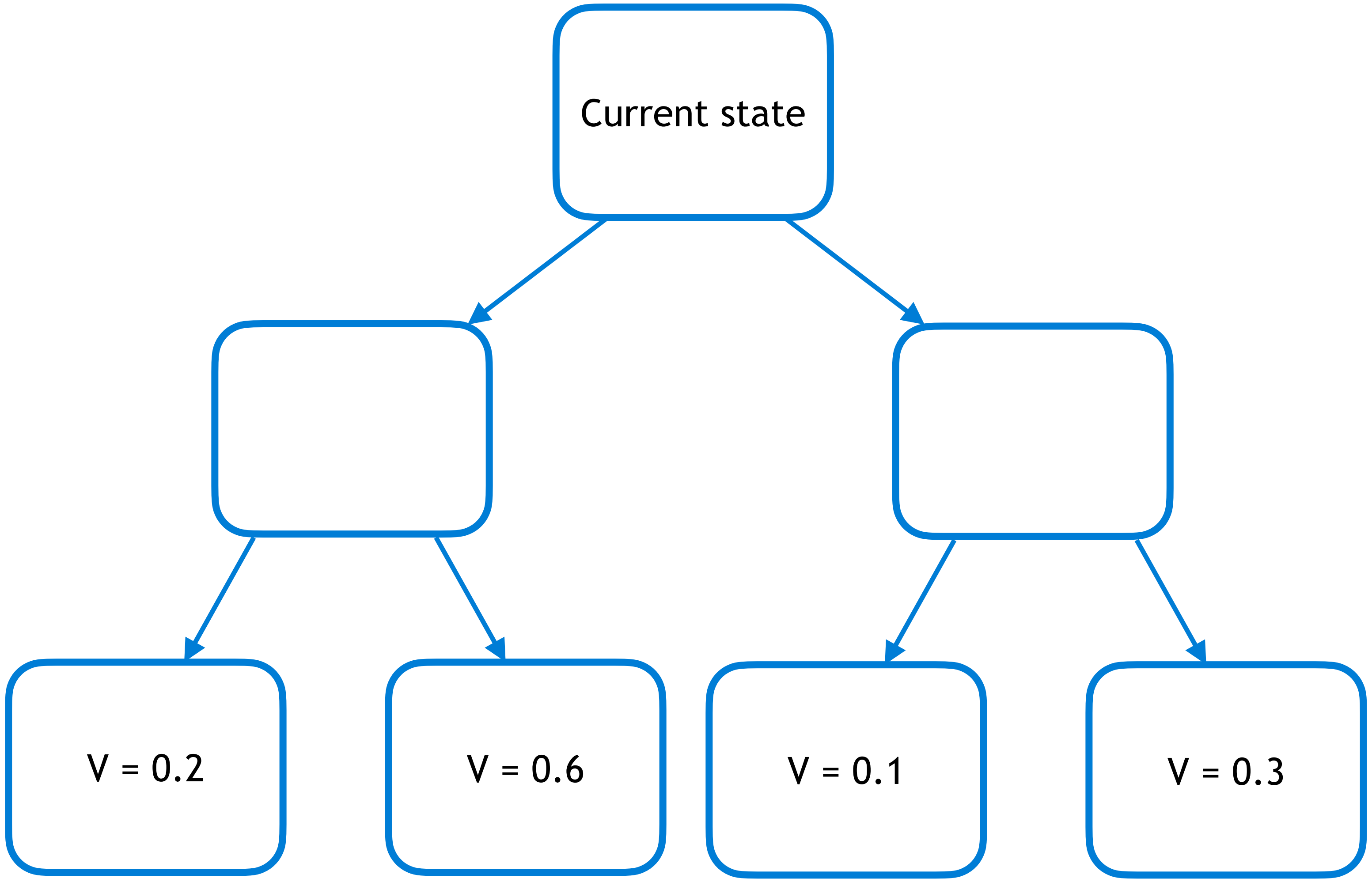
Organizational strategy



Playing Go

PLANNING

Planning is defined as the **mental simulation of future actions and their consequences**, typically formalized as search over a decision tree



KEY IDEAS IN THE PLANNING LITERATURE

Heuristics that mitigate the costs of planning

- Uncertainty arbitration
  Daw, Niv, and Dayan, 2005
- Pruning
  Huys et al., 2012
- Planning as evidence integration
  Solway and Botvinick, 2012; Solway and Botvinick, 2015
- Cost-benefit arbitration
  Kool et al., 2017
- Amortization
  Dasgupta et al., 2018; Hamrick et al., 2019

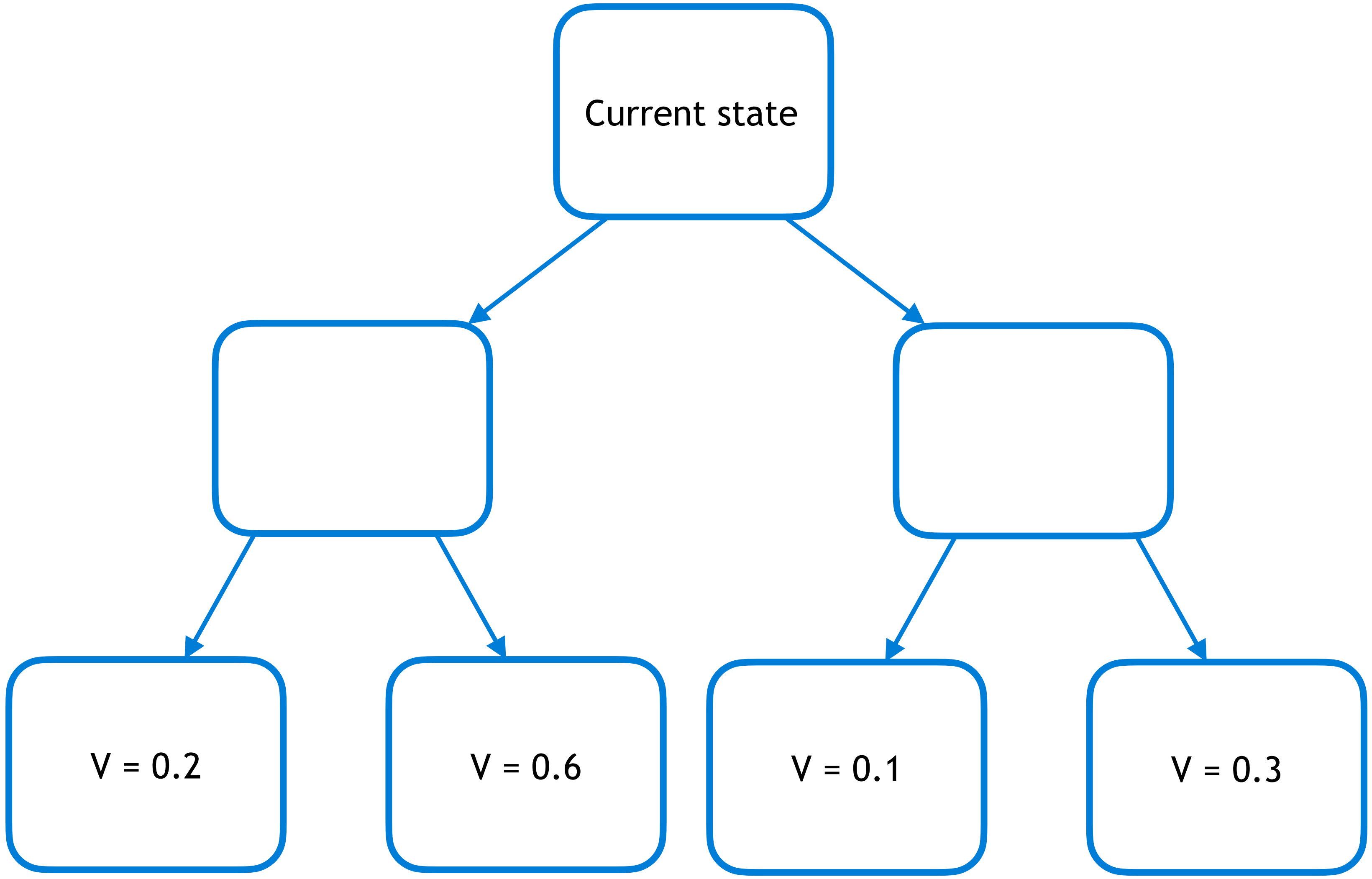
Normative models that are expensive or rely on prior experience

- Plan-until-habit
  Sezener et al., 2019
- Resource rationality
  Callaway et al., 2021

Lacking: a normative model that can derive useful heuristics to make planning more efficient

The goal is to develop a meta-planning model, where the agent can quickly **compute the statistical effects of prospection**

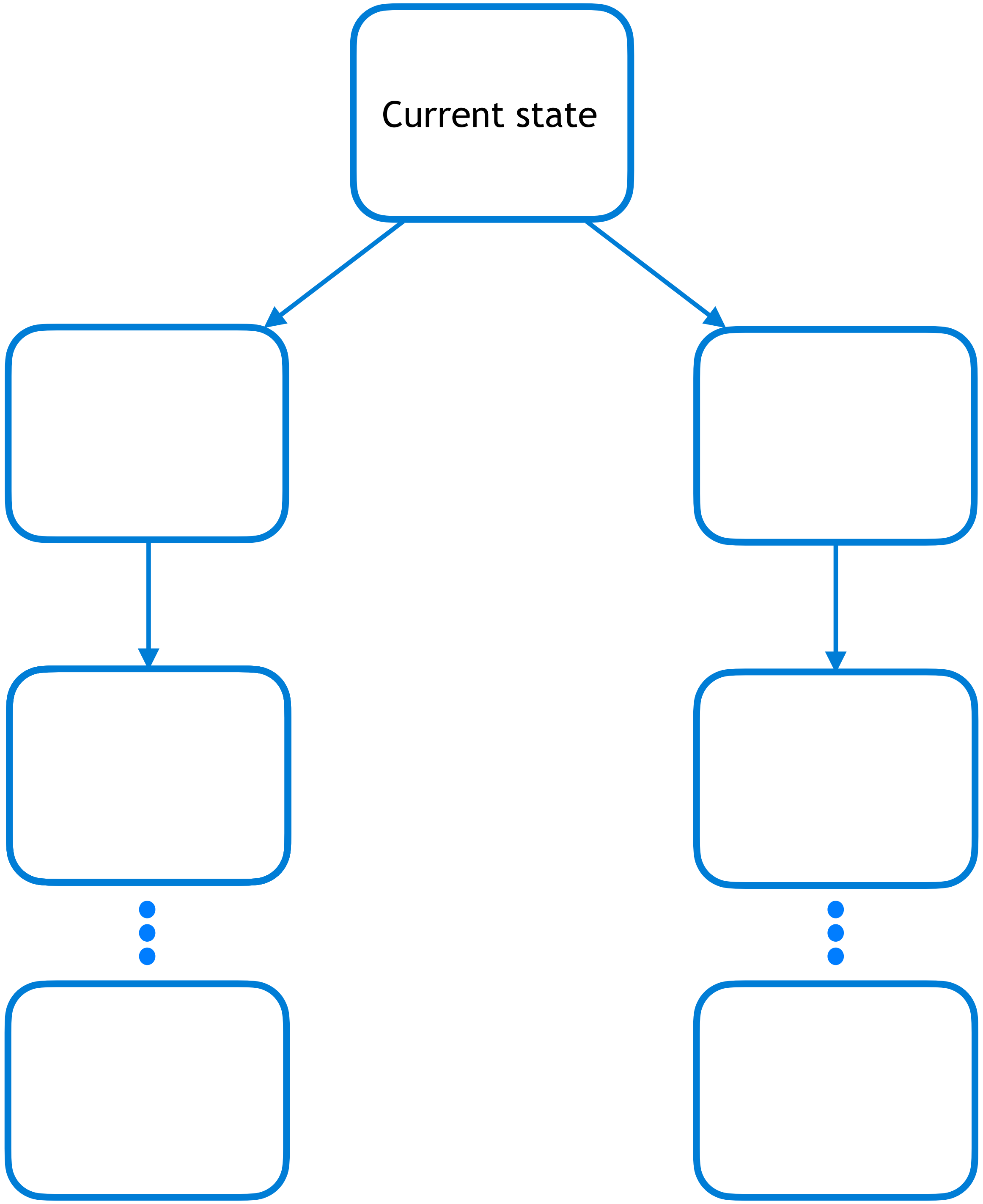
APPROXIMATING PLANNING WITH NOISY MEASUREMENTS



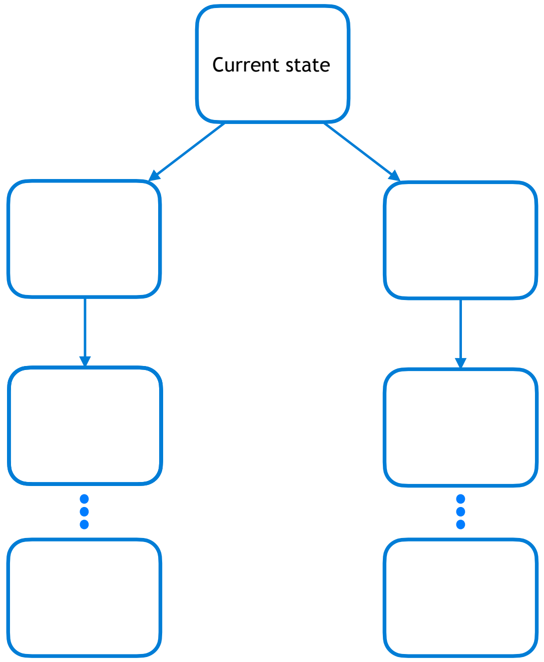
APPROXIMATING PLANNING WITH NOISY MEASUREMENTS

Myopic measurements are given

T noisy measurements



APPROXIMATING PLANNING WITH NOISY MEASUREMENTS



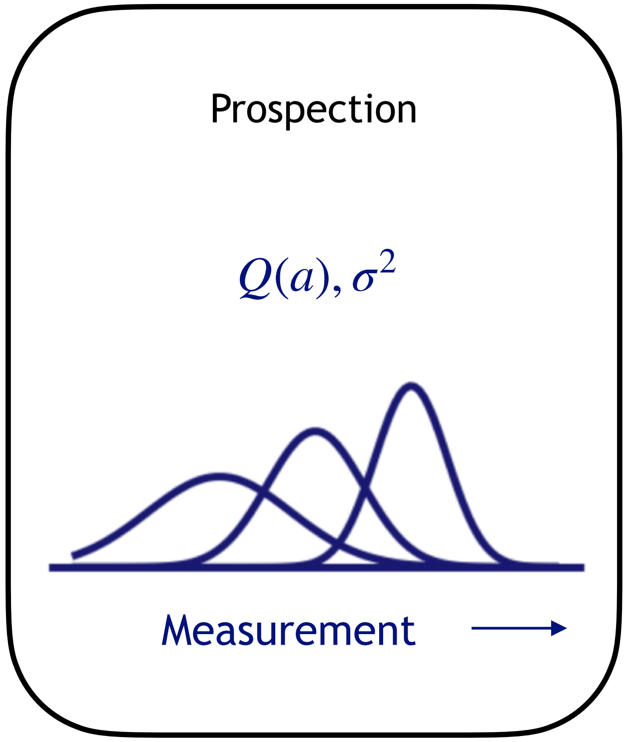
Features of the model

- Three stages: generative model, inference, and optimization
- Each state-action pairing has a theoretical expected reward, $Q(a)$, that the agent builds a probability distribution over
- Myopic evaluations are given, and each iteration of search produces a new, independent measurement of $Q(a)$
- The agent balances the expected reward gained by planning with the costs of planning
- Conceptually, this is an approximation of breadth-first search that decides how many layers of the tree to evaluate ahead of time

GENERATIVE MODEL

Generative model → Inference → Optimization

Sample noisy measurements of the true Q value for a



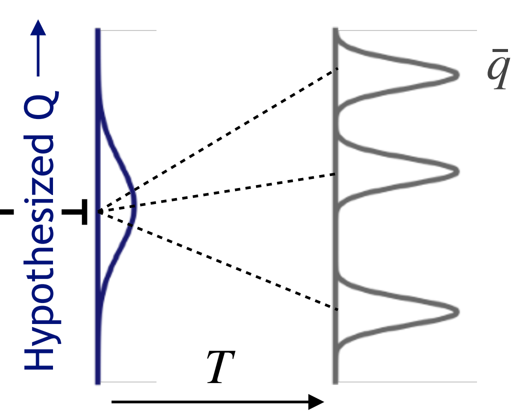
INFERENCE

Generative model → **Inference** → Optimization

Consider different futures for each action

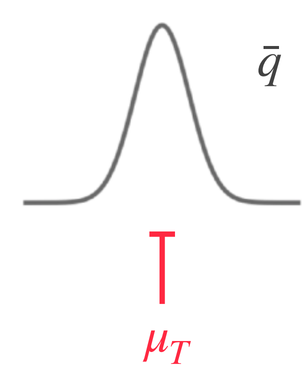
Simulate the evolution of the likelihood over Q

Given a single measurement of the true Q value for a



Integrate information for each future

Compute per action



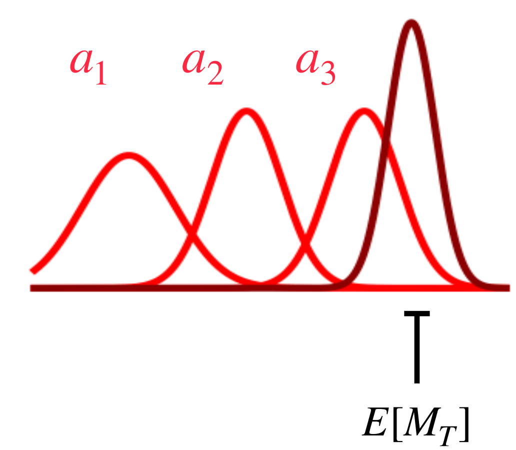
Combine across actions

$$M_T = \max_a \mu_T(a)$$

Marginalize over all possible futures

Distribution of $\mu_T(a_i)$

Distribution of M_T

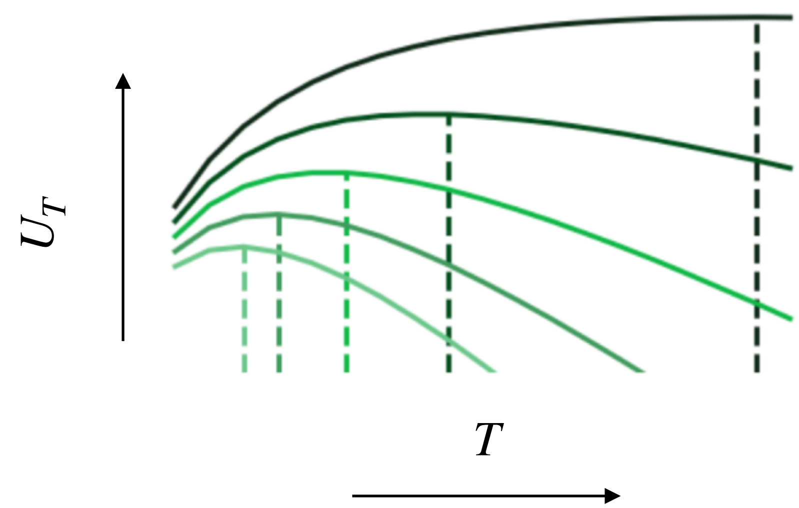


OPTIMIZATION

Generative model → Inference → **Optimization**

Optimize the depth of tree search independently of any action

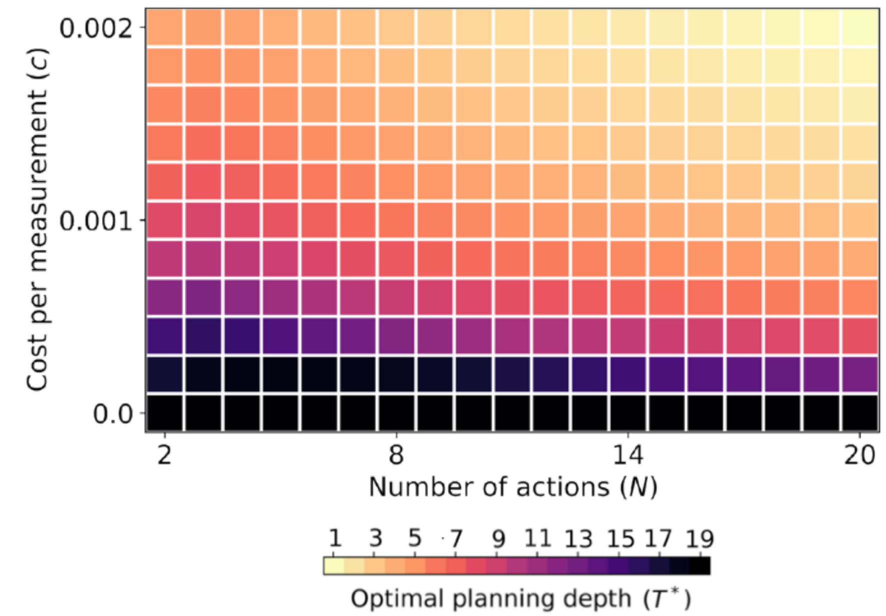
Plan for T^* iterations by balancing expected reward and a fixed cost per iteration



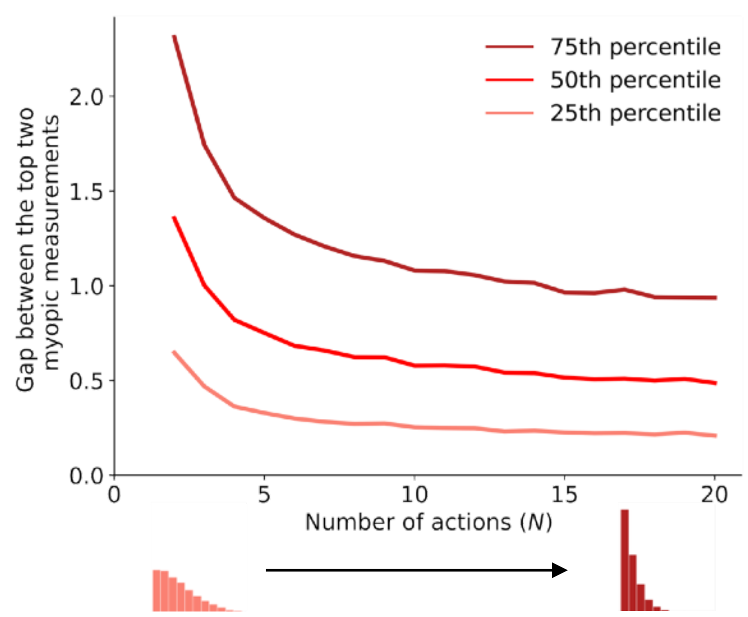
$$U_T = (E[M_T] - E[M_1]) - cNT$$

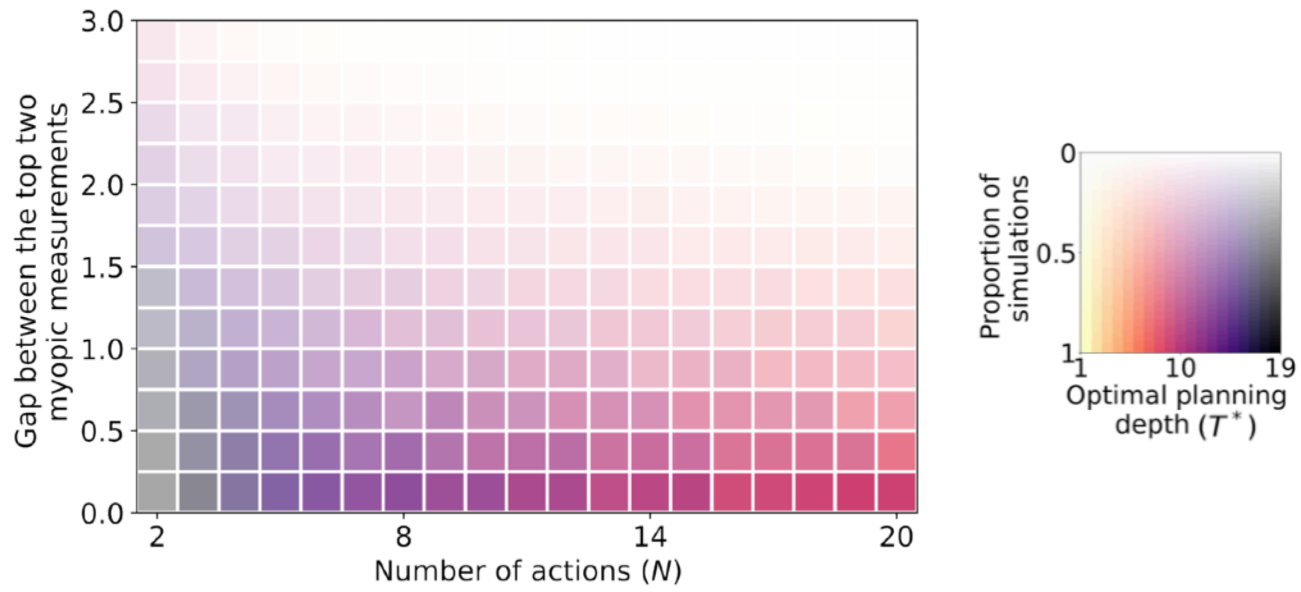
MODEL PREDICTIONS

Deeper planning is more beneficial at lower costs with less alternatives



How does this effect interact with the gap between the top myopic actions?





This dependency can act as a basis for a heuristic of the form $T^* = F(N, \text{gap})$

Our model can also take into account **tradeoffs between prospective and retrospective systems**

AMENDING THE INFERENCE MODEL FOR RETROSPECTION

Generative model

Sample noisy measurements of the true Q value for a

Retrospection

$Q(a), \sigma^2_{\text{retro}}$

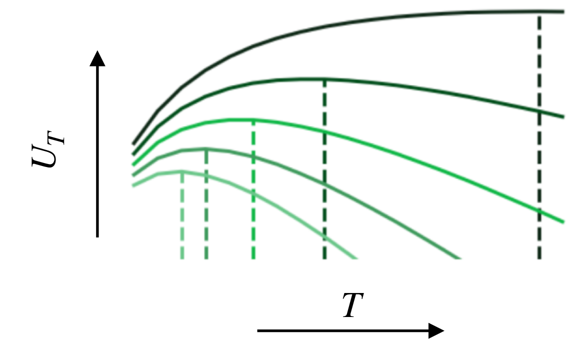
Prospection

$Q(a), \sigma^2$

Optimization

Optimize the depth of tree search

Plan for T^* iterations by balancing expected reward and a fixed cost per iteration



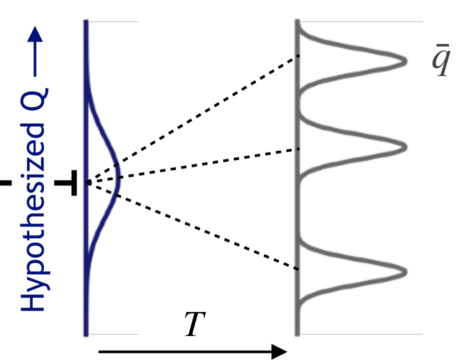
$$U_T = (E[M_T] - E[M_1]) - cNT$$

Inference

Consider different futures for each action

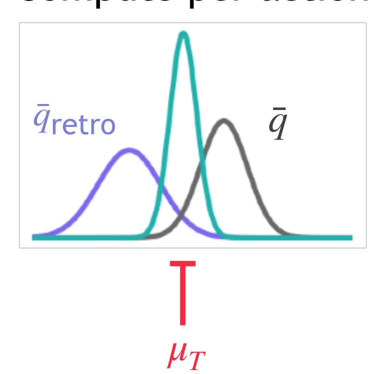
Given a single measurement of the true Q value for a

Simulate the evolution of the likelihood over Q



Integrate information for each future

Compute per action



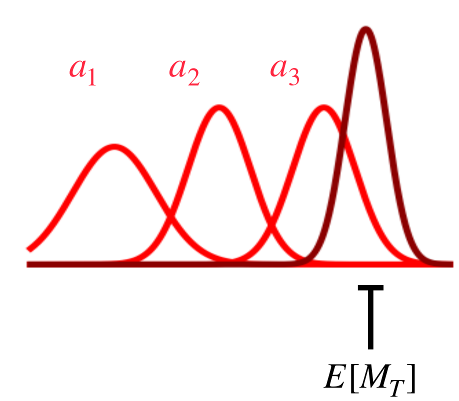
Combine across actions

$$M_T = \max_a \mu_T(a)$$

Marginalize over all possible futures

Distribution of $\mu_T(a_i)$

Distribution of M_T



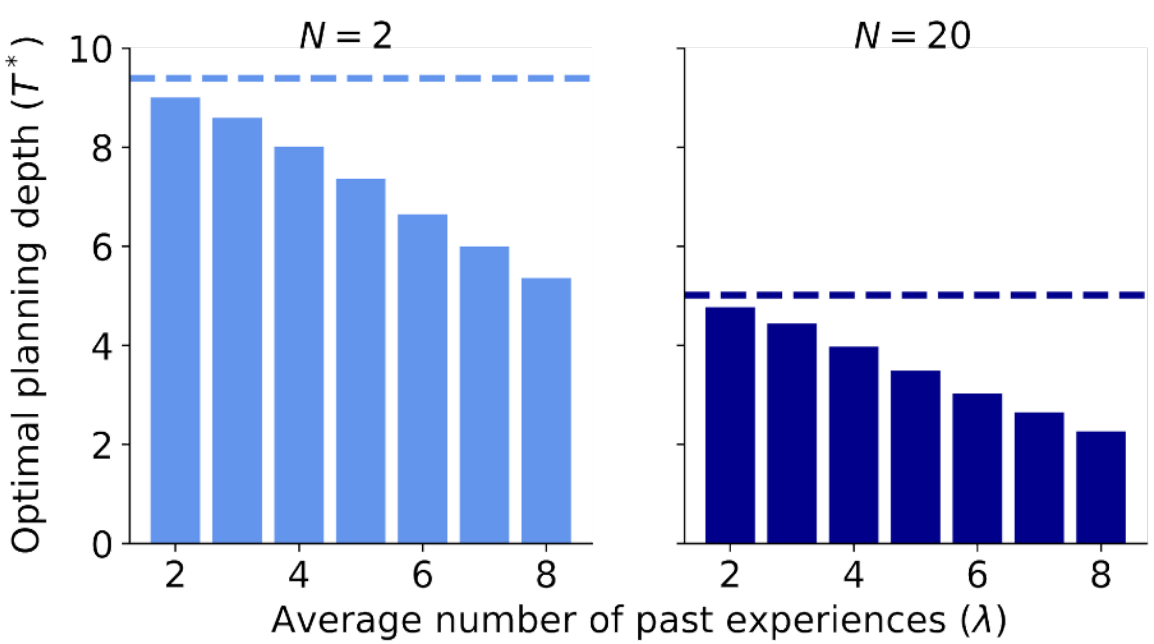
THE EFFECTS OF RETROSPECTION

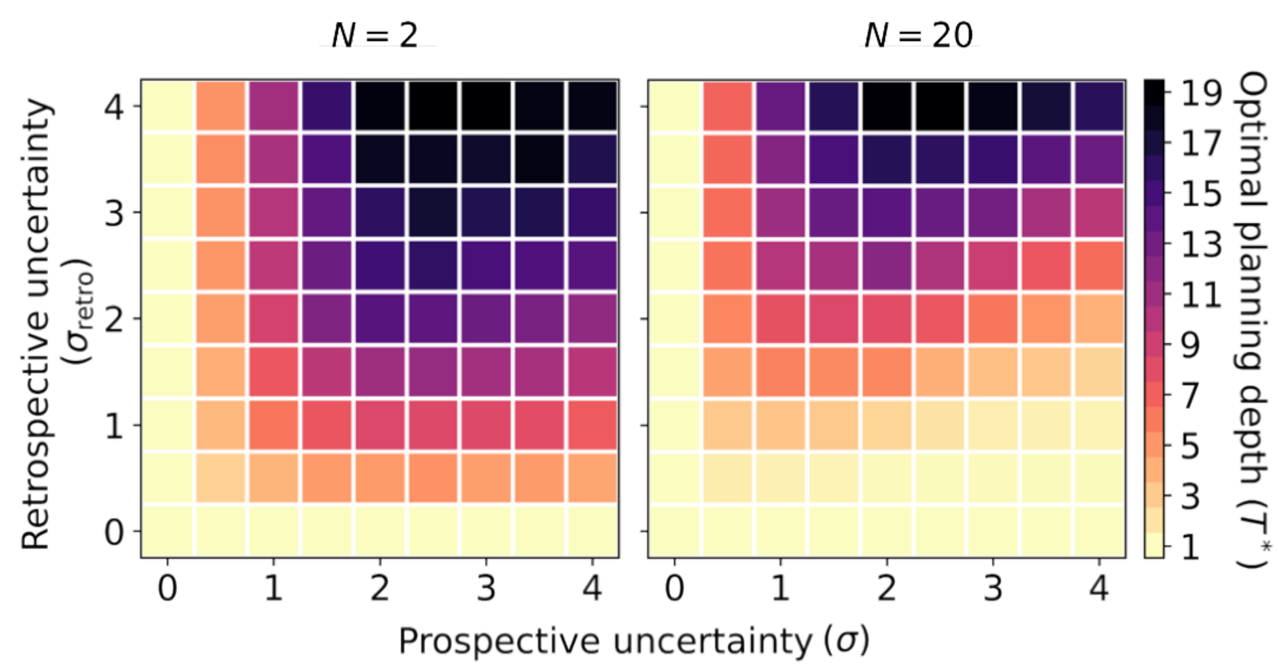
Result #1: the transition from model-based to model-free control over decision-making with time

Compare to *Dickinson, 1985*

Result #2: uncertainty-based weighting of prospective and retrospective information

Compare to *Daw, Niv, and Dayan, 2005*





Now, the heuristic can take the form $T^* = F(N, \text{gap}, \lambda, \sigma_{\text{retro}}, \sigma)$

Deriving an **online variant of our model** allows it to interact with a planning algorithm while planning is taking place

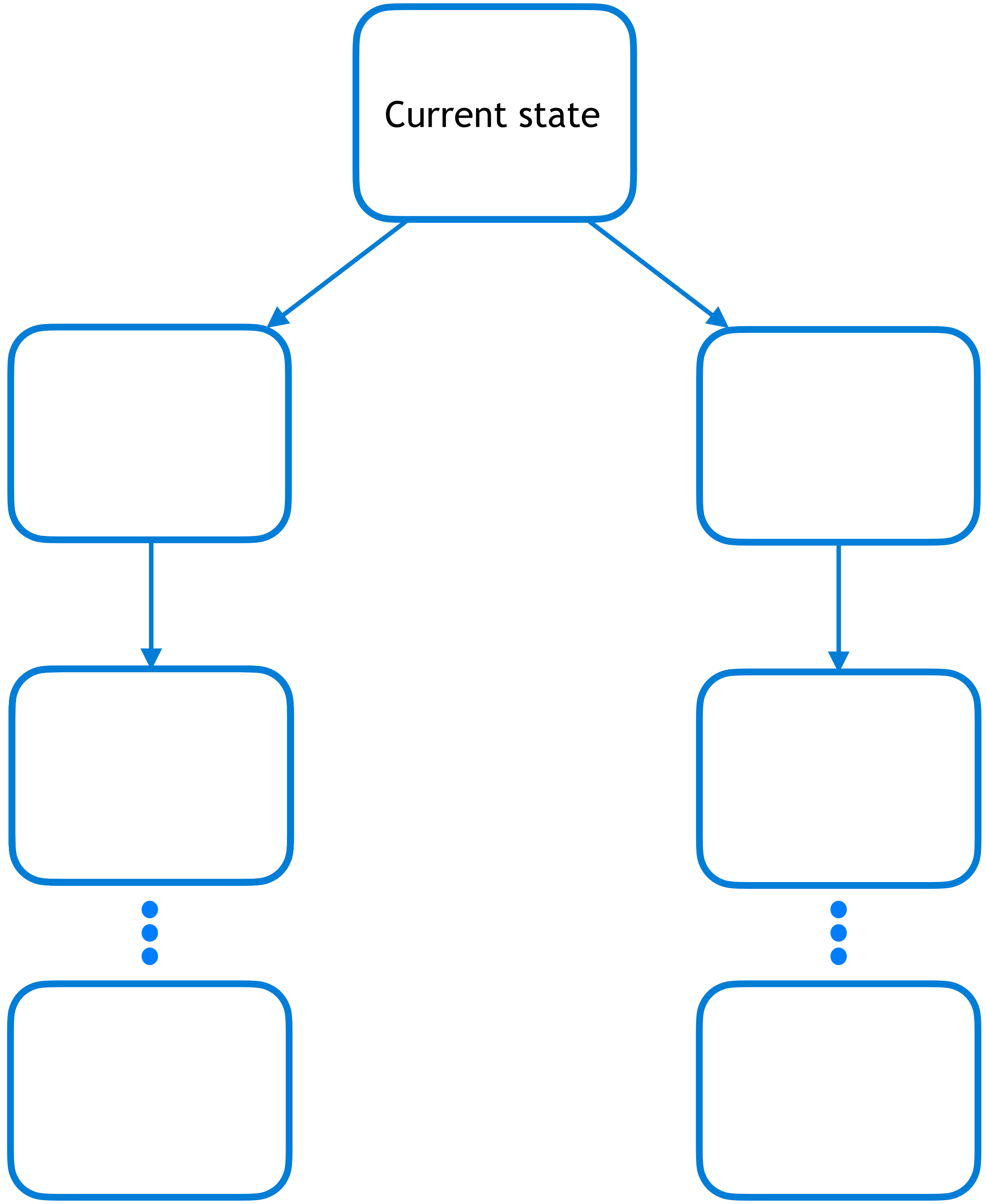
ONLINE META-PLANNING

Myopic measurements are given

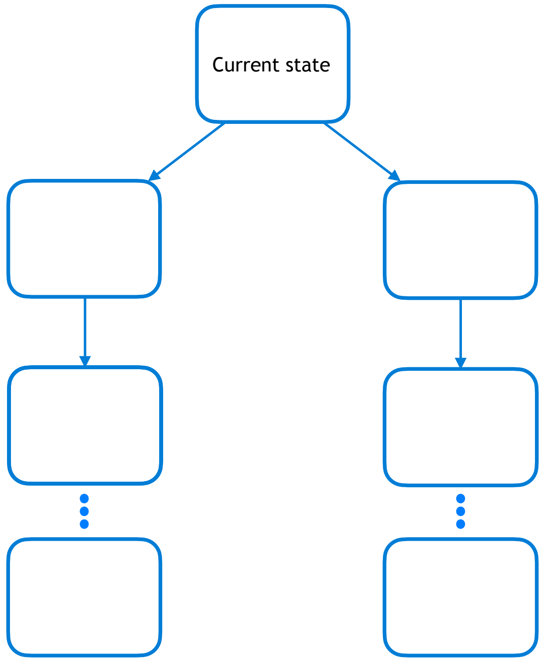
Is planning another step worthwhile?

Draw another set of myopic measurements

T noisy measurements



ONLINE META-PLANNING



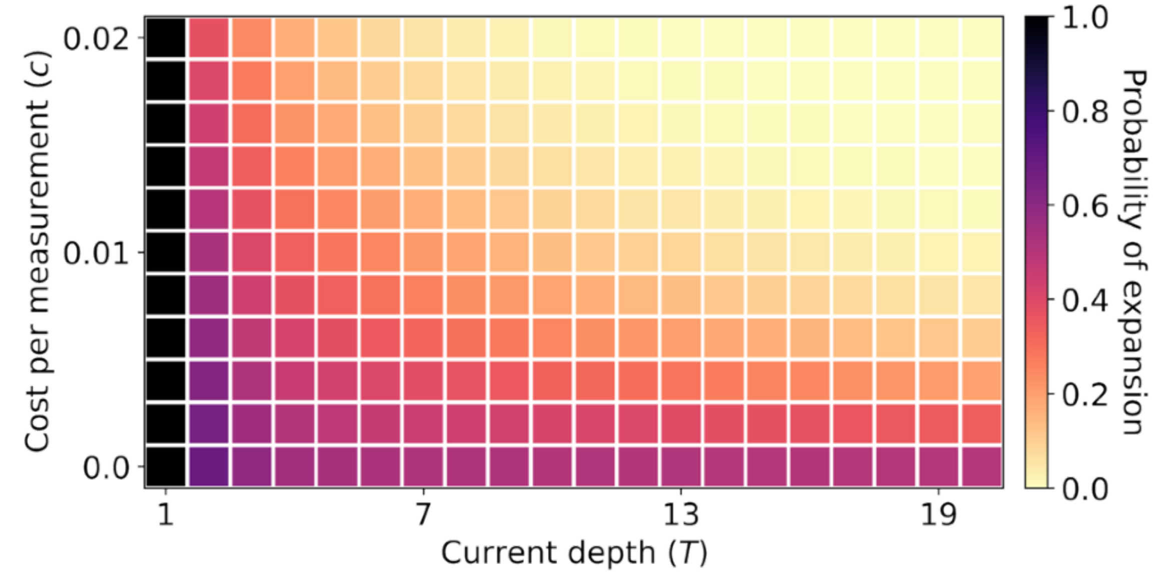
Features of the online model

- Iteratively approximate the information gained by planning another iteration
- Still breadth-first search, but the model receives additional information at each step of search
- Derivation for the model is the same, but now there is a sequence of T past measurements for each action and the future consists of a single time step, from T to $T + 1$

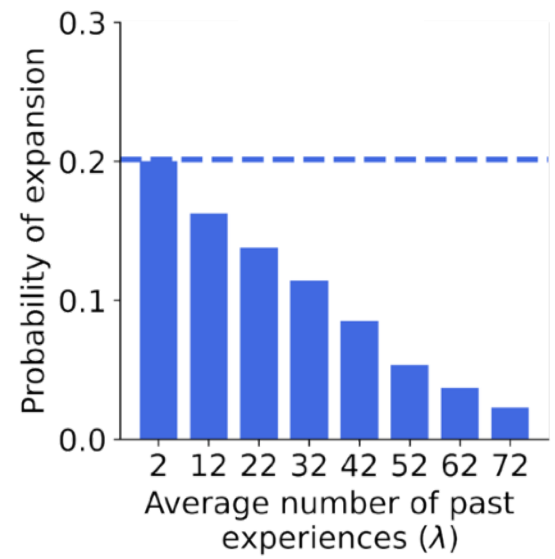
ONLINE MODEL PREDICTIONS

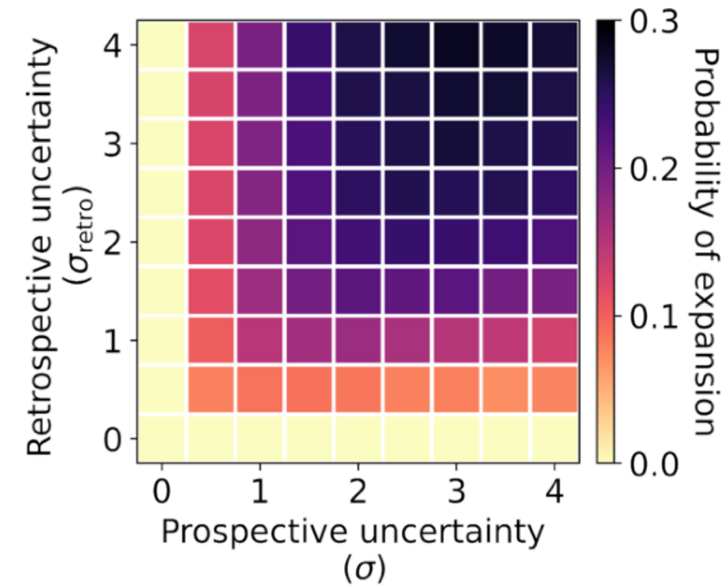
The online variant of the model replicates the results from its offline counterpart

Myopic predictions



Effects of retrospection



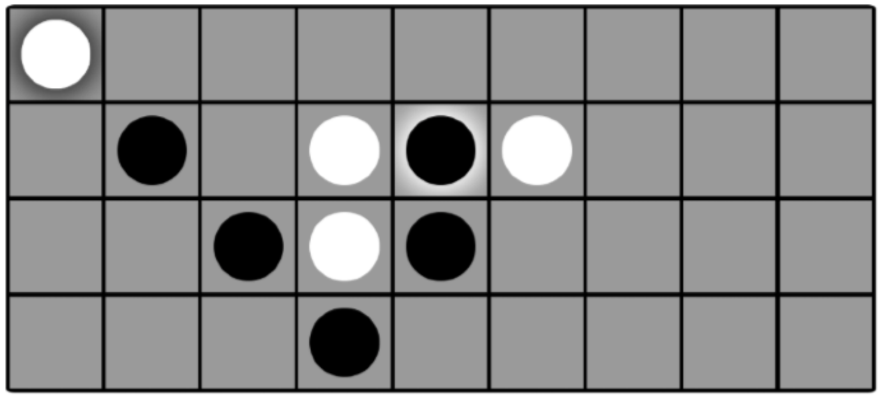


Here, the heuristic is formulated in terms of probability of expansion as

$$P_{\text{expansion}} = F(N, \text{gap}, \lambda, \sigma_{\text{retro}}, \sigma)$$

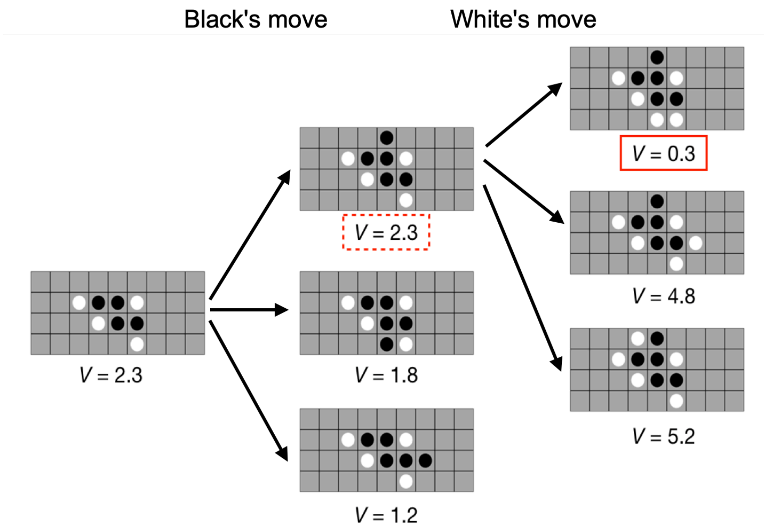
One method for applying this theoretical framework to data is to make use of heuristics to **inform components of a tree search model**, such as a stopping rule or acquisition function

APPLYING THE MODEL TO A COMPLEX PLANNING TASK

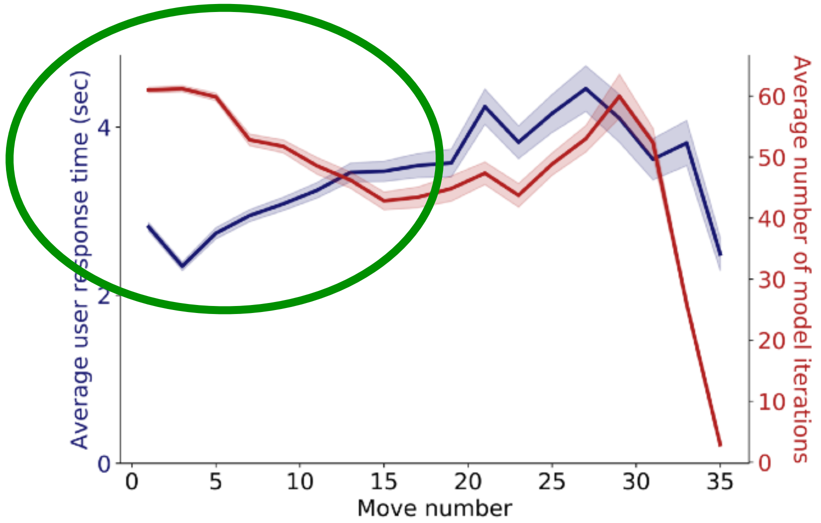


$1.2*10^{16}$ possible states

Planning model



Discrepancies with data



1. More nuanced planning components would lead to a better fit to data
2. Initial evidence for shorter response times in the opening can be better explained if the heuristic suggests shallower planning

SUMMARY

- Presented a normative framework for optimizing the depth of decision tree search
- The model uses Bayesian inference to combine myopic estimates and retrospective samples and computes the value of planning
- The model makes intuitive predictions about planning depth over a range of parameters
- These parameters can in turn be used to derive useful heuristics about when it is beneficial to plan, which can in practice improve a planning algorithm
- An online implementation of the model shows consistent results

Ma Lab
Wei Ji Ma
Bas van Opheusden
Heiko Schütt
Hsin-Hung Li
Dongjae Kim
Jenn Lee
Peiyuan Zhang
Xiang Li
Daisy Lin
Luísa Leonelli
Sammy Tavassoli

NYU/CNS Center for Neural Science

MA LAB HOMINES STERCVS NESCIVNT MMXIII

NYU NEUROSCIENCE