Meta-cognitive planning for learning representations

Tingke Shen[1], Peter Dayan[1,2] and Mihály Bányai[1,3]

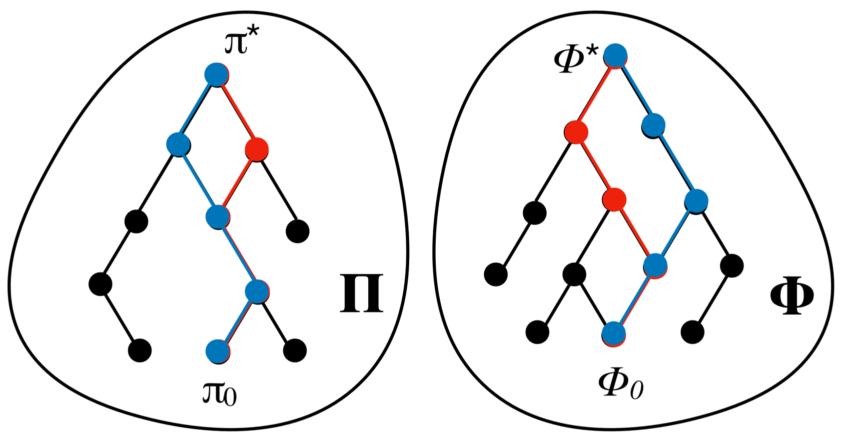
[1] Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[2] University Of Tübingen, Germany
[3] Center for Cognitive Computation, Central European University, Budapest, Hungary

1. Finding the best representation for every step of the learning process

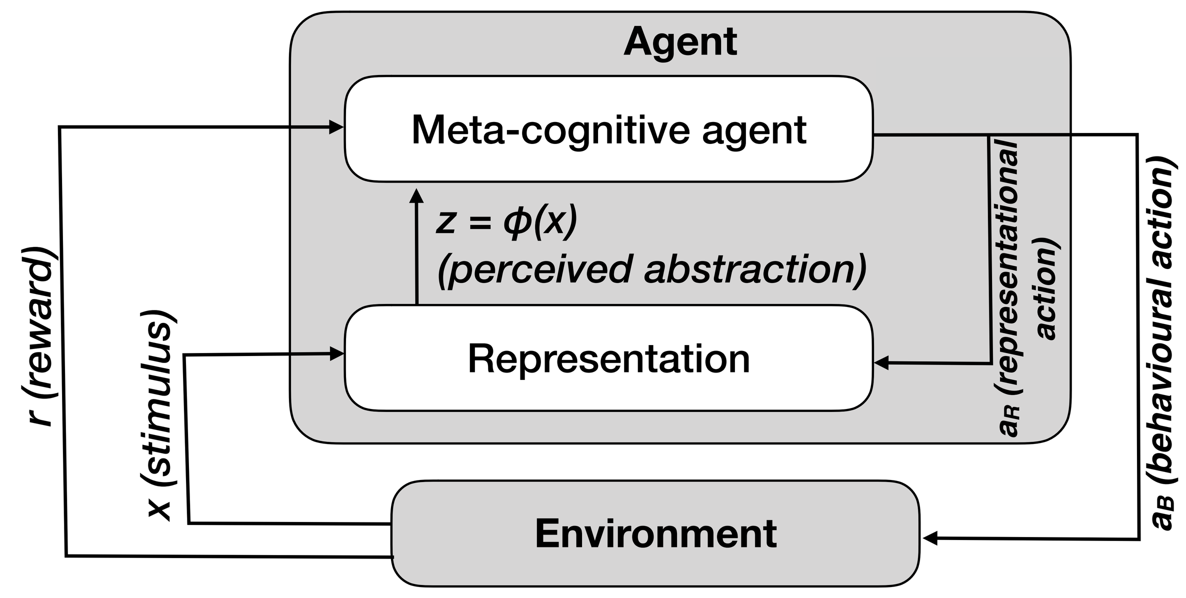
$z = \phi(x)$

$a \sim \pi(a \mid z)$

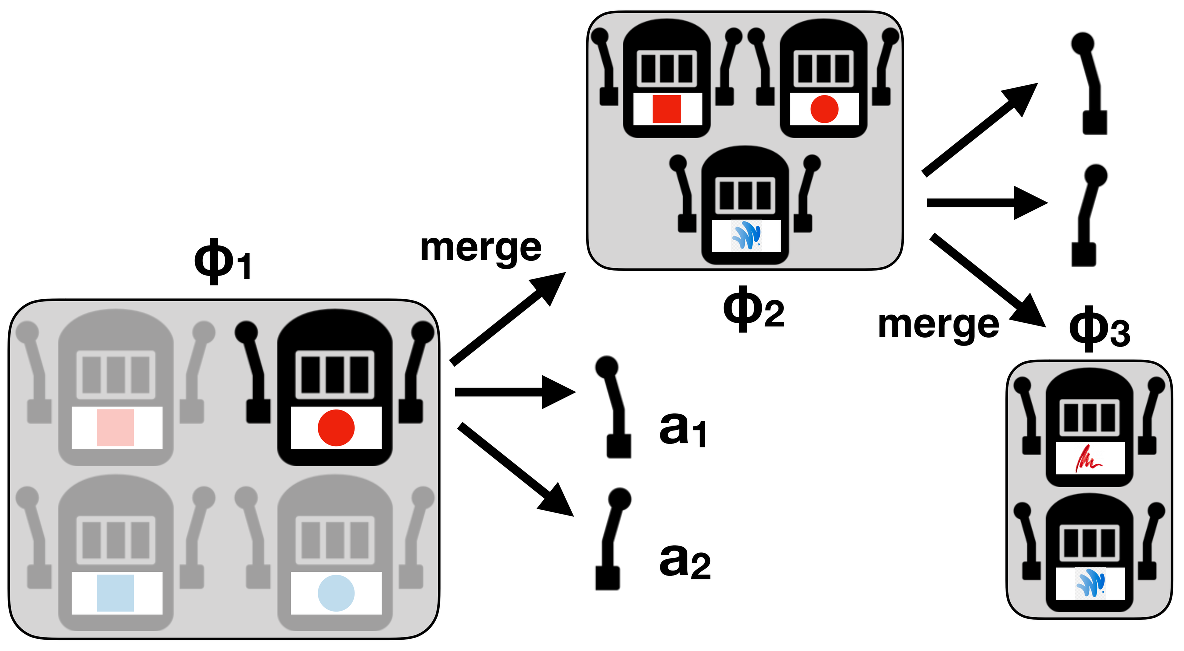
- We examine the effect of various inductive biases on representational decisions
- Learning delineates a joint path in the spaces of policies (π) and representations (ϕ)
- Representations must support reaching favourable points in both spaces in the future
- We use planning in the representational space to examine decision situations and how they depend on the parametrisation and the informational state of the agent
- The aim is to construct a normative theory of representation learning



2. Meta-cognitive planning



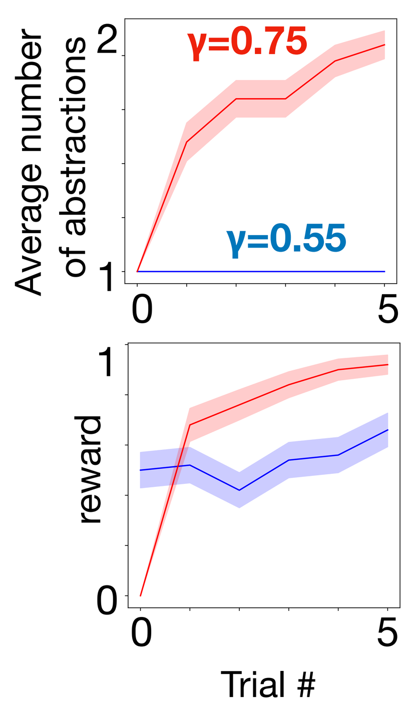
- Representations evolve during task learning via fine-graining and coarse-graining operations, I.e. splitting and merging abstractions
- Updating a representation is a meta-cognitive action
- The optimal representational update is given by planning in the meta-cognitive belief MDP, via the R-PLAN algorithm



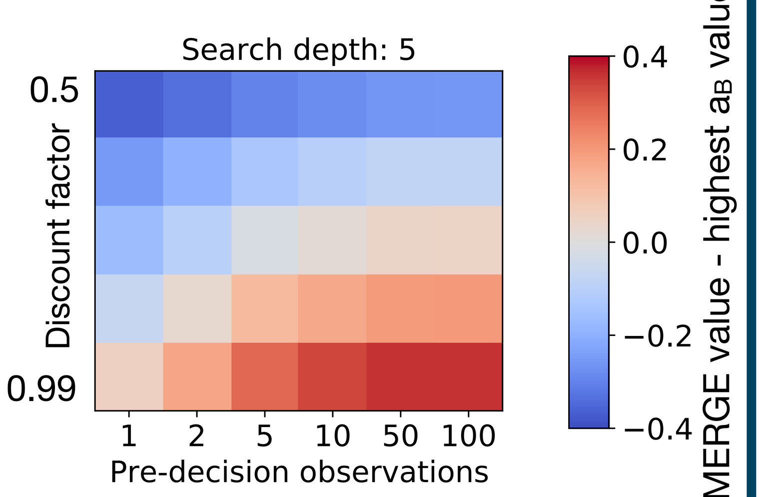
- Fully Bayesian belief updates are in general intractable, we use Beta distributions to approximate beliefs after representational actions
- The choice of belief approximation creates or unduly eliminates uncertainty
- We preserve the effective number of observations upon merging abstractions
- Merges and splits appear to be advantageous or disadvantageous based on **inductive biases** about the temporal structure of the task, the geometry of the reward function or the stochasticity of the environment

3. Effect of temporal preferences

- Temporal discounting itself has a similar role to an inductive bias
- An agent with infinite time to learn needs no generalisation via abstractions, and one with very little time will not want to merge or split any
- We use contextual bandits to analyse specific representational decisions
- First, we look at **when does the agent prefer to split an abstraction** instead of exploiting it
- Simulating full tree search planning with a fixed depth
- An agent with no prior experience will split its coarse-grained abstraction when its discount factor is higher

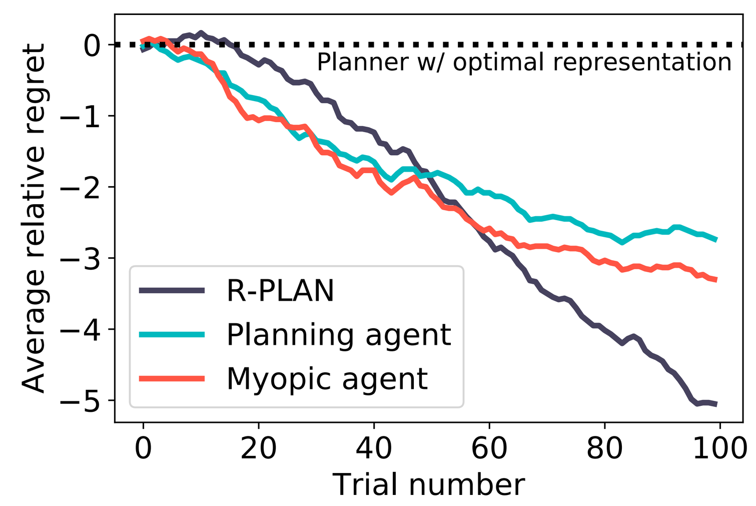


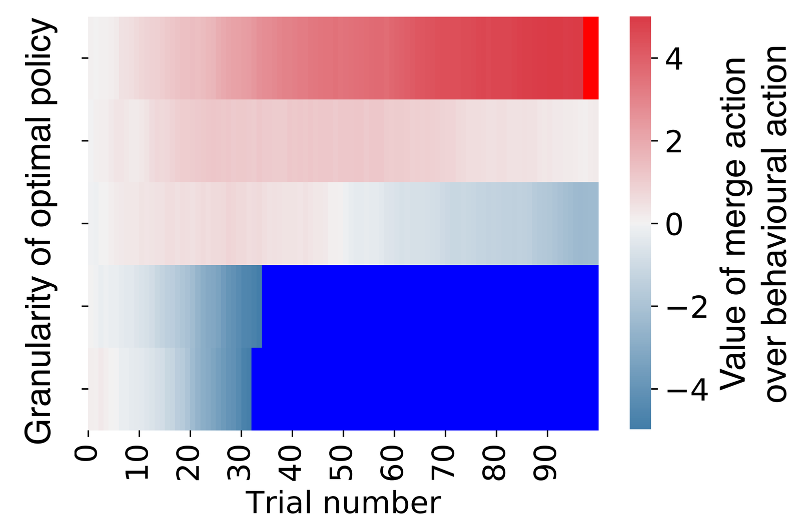
- Similarly, when faced with the option to **merge two abstractions**, an agent will do so when the discount factor is high enough
  - depending on the amount of information it already has about them
- This is the **belief** of the agent about the optimal choice, **not objective optimality**



4. Deep planning with MCTS

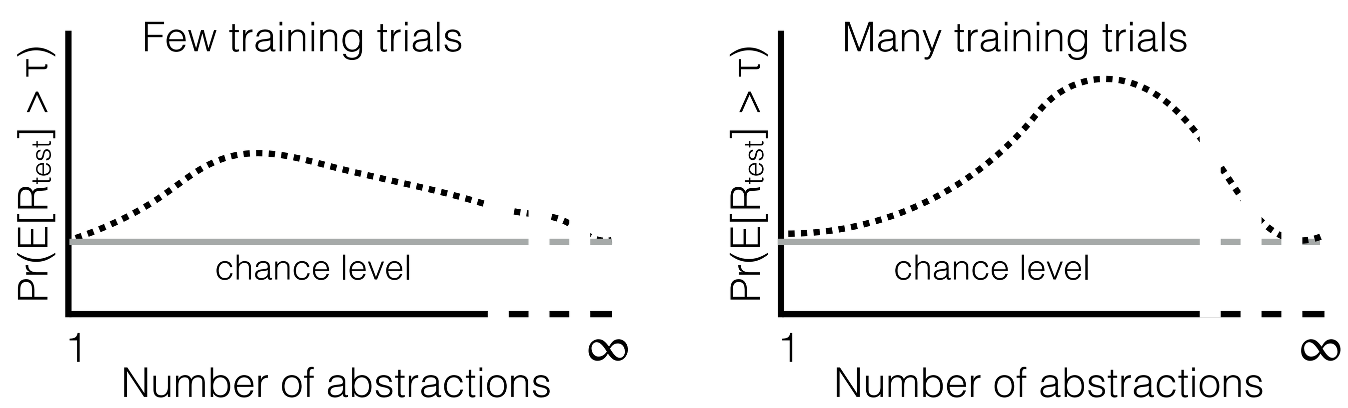
- When do the beliefs turn into objective advantage?
- We use a stochastic approximation (MCTS) to the full tree search in representational planning in order to increase depth
- We compare to agents that do not form abstractions of the state space
- All agents **lack any inductive bias about the geometry of the reward** function
- Planning has a slight advantage over myopic decision making
- Representational planning confers further advantage in terms of learning speed in more correlated environments
  - We did not include the possibility of splitting, thus the lower plateau





5. Non-Markovian reward functions

- What are **inductive biases that make a coarse-grained representation an optimal choice**?
  - Other than bluntly telling the agent something about reward geometry
- Having to learn a **good enough policy in finite time**
  - Instead of temporal discounting, we take a finite training phase and a test phase as a simplified setting
  - Continuous state space, so without abstractions no generalisation is possible
  - The objective we use is the probability of the expected reward being above a set **threshold** in the test phase
- The optimal choice of representation will mainly depend on how many observations can the agent learn from, as a form of **bias-variance trade-off**



6. Conclusions

- A normative theory of representation learning considers the sequential meta-cognitive decision making problem of forming abstractions that support all stages of learning
- The various inductive biases about the structure of the task and the environment co-determine optimal representation choices
- In order to make planning computations tractable, approximations are required to both the belief update and the tree search-based value estimation
- In order to construct representational trajectories we need to consider all the aspects above together
- Scaling to realistic problems will require amortising the planning computations into a reactive representation learning algorithm

Acknowledgements This work was funded by the Marie Skłodowska-Curie Action RELEARN-DLV-897042 of the European Commission, the Max Planck Society and the Humboldt Foundation.