

# Meta-learning emotional control in bandit tasks

Prakhar Godara

TU Darmstadt, Darmstadt, Germany

## Abstract

In decision making scenarios, reasoning can be viewed as an agent executing an algorithm  $p \in \mathcal{P}$  that selects an action  $a \in A$ , aiming to optimize some outcome. Metareasoning extends this by selecting  $p$  itself through a meta-algorithm  $p_{\text{meta}}$ . Previous approaches to study metareasoning in humans have required that the transition/reward distributions are known by the agent, but the value function isn't. We extend these efforts to study metareasoning for agents acting in unknown environments by formalizing the meta problem as a meta Bayes adaptive Markov decision problem (meta-BAMDP). We formally investigate the theoretical consequences of this framework within the context of two-armed Bernoulli bandit (TABB) tasks. Not only do we make theoretical progress in making the (usually intractable) metareasoning problem tractable, but we also generate predictions for a resource rational account of human exploration in TABB tasks.