

# Active Reinforcement Learning: Observing Rewards at a Cost

David Krueger

Jan Leike

John Salvatier

Owain Evans

What is Active RL?

MDP Algorithms

MDP Environments

Motivations

Bandit Algorithms

Bandit Experiments

MDP Experiments

This work was supported by Future of Life Institute grant 2015-144846 (JS, DK, OE). We thank Andreas Stuhlmüller, Tor Lattimore, Reimar Leike, Ryan Lowe, Akram Erraqabi, and Jessica Taylor for helpful discussions. The authors acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out this work.