RL experiment report

1 Introduction and methods

The aim of this document is plot the results of the following RL experiment:

- Algorithms: A2C, ACKTR, DDPG, PPO2, SAC, TD3, TRPO
- Environment: Reacher2Dof-v0
- Number of time steps: 0.1M
- Number of initialisation seeds: 2
- Number of parallel environments: 8 for ACKTR and PPO2 and 1 for SAC and TD3 (parallelisation not supported).

The performance metrics are defined as follows:

- Train time (min) : Wall time to train.
- Success ratio : number of successful episodes / number of reachable episodes
  An episode is successful if the distance between the finger tip and the target is less than or equal to a threshold of 50mm, 20mm, 10mm or 5mm.
- Average reaching time : sum (number of time steps of all successful episodes) / number of successful episodes
  An episode has a maximum of 150 time steps.
- Efficiency: mean reward / mean training walltime.

2 Results

2.1 Raw results

2.2 Learning curves

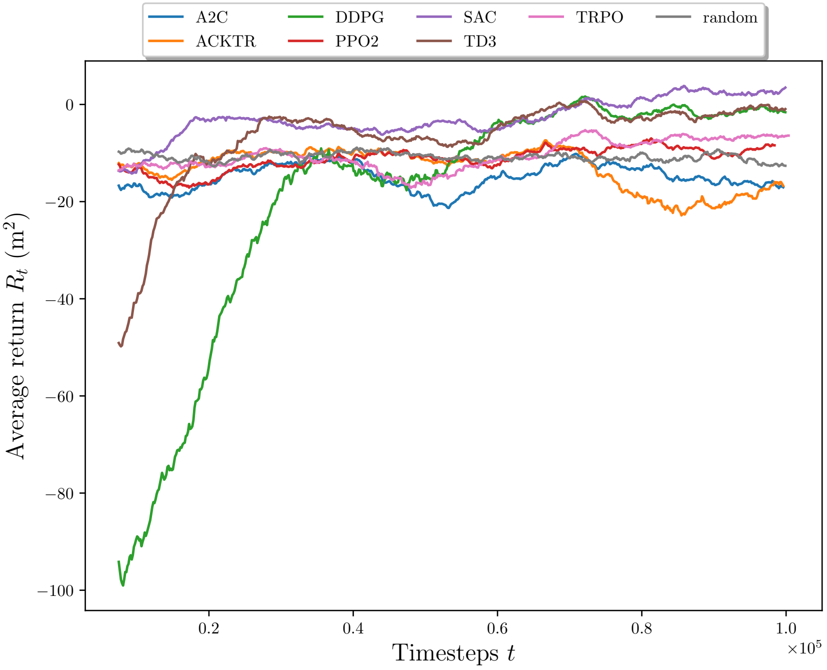


Figure 1: All learning curves.

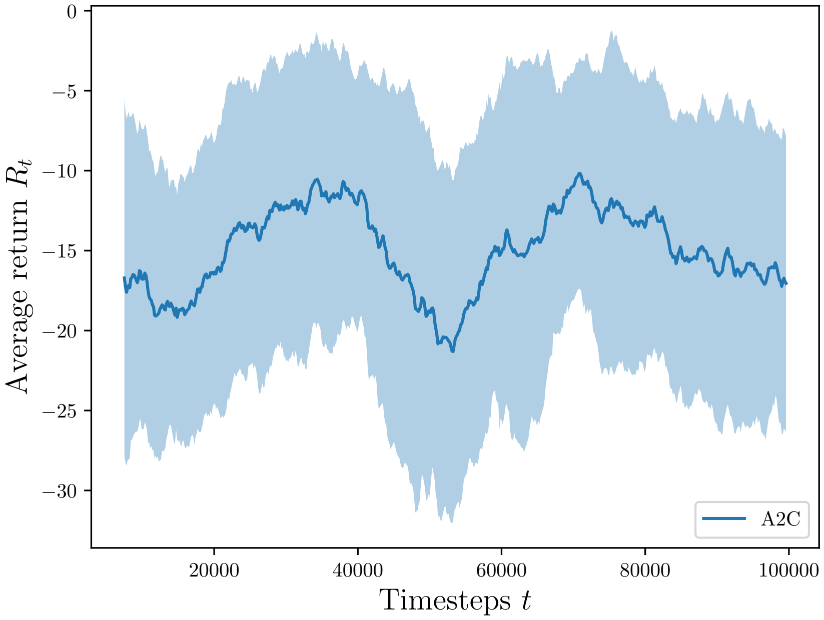


Figure 2: Learning curve A2C.

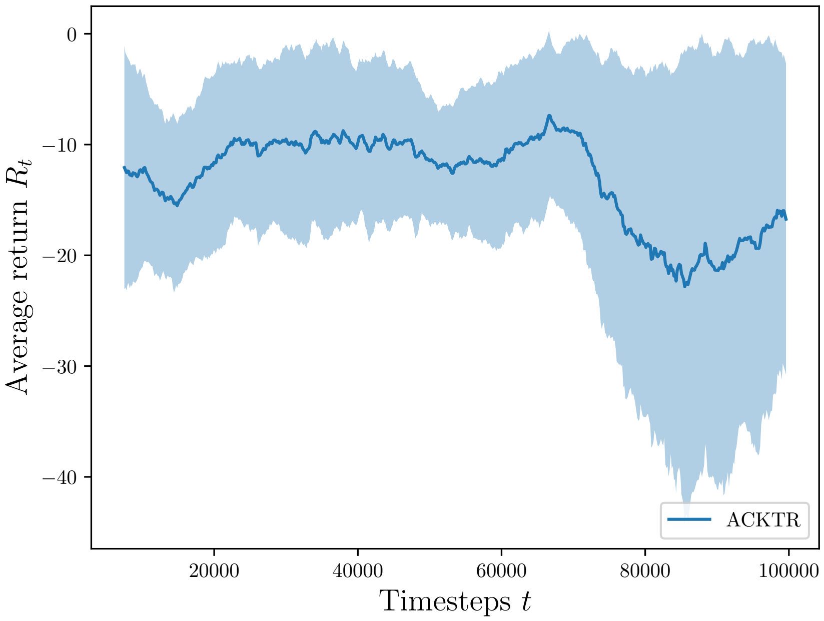


Figure 3: Learning curve ACKTR.

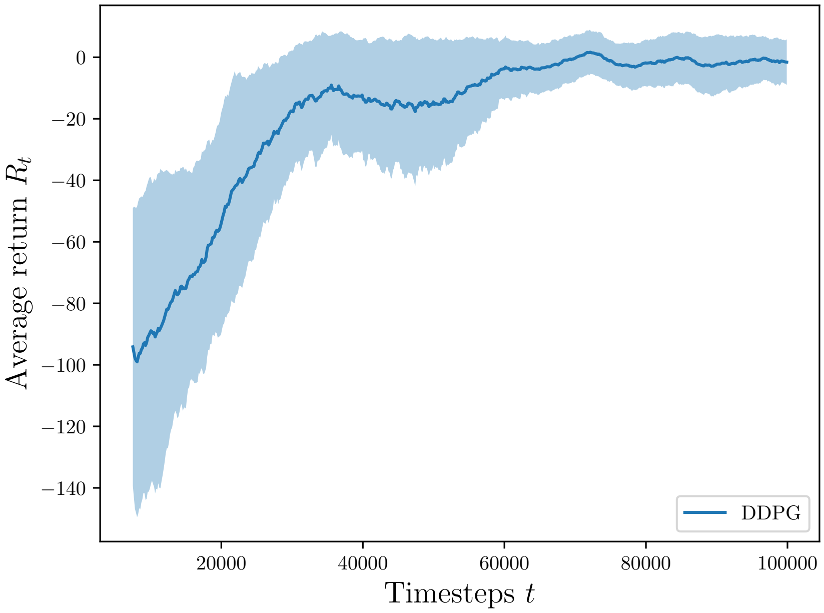


Figure 4: Learning curve DDPG.

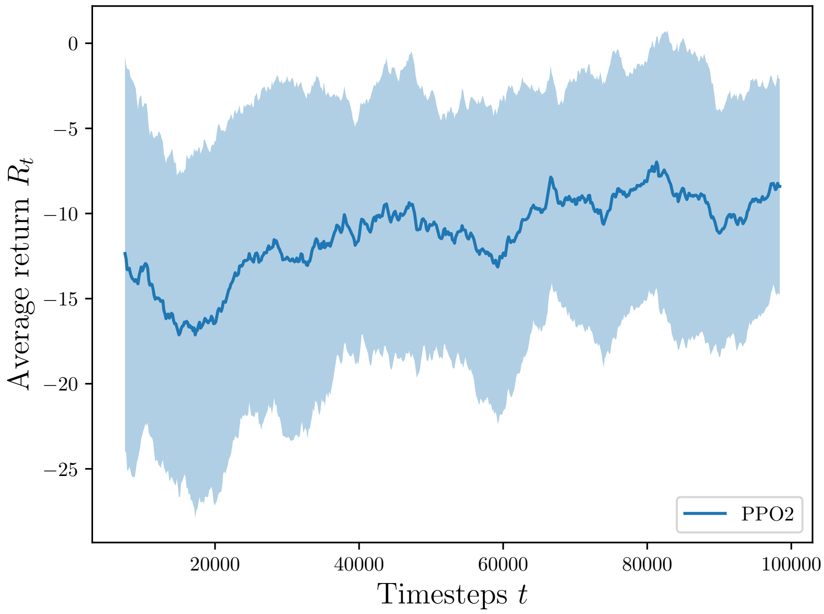


Figure 5: Learning curve PPO2.

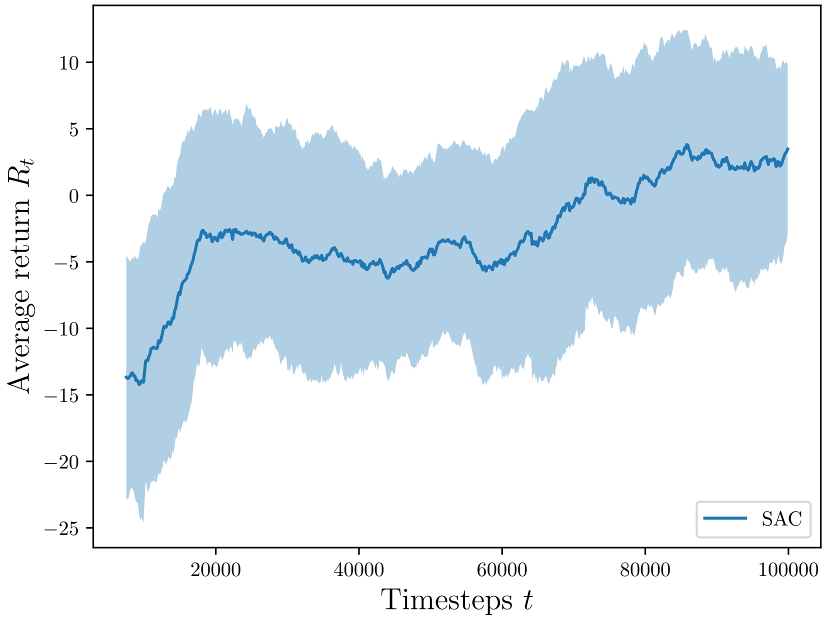


Figure 6: Learning curve SAC.

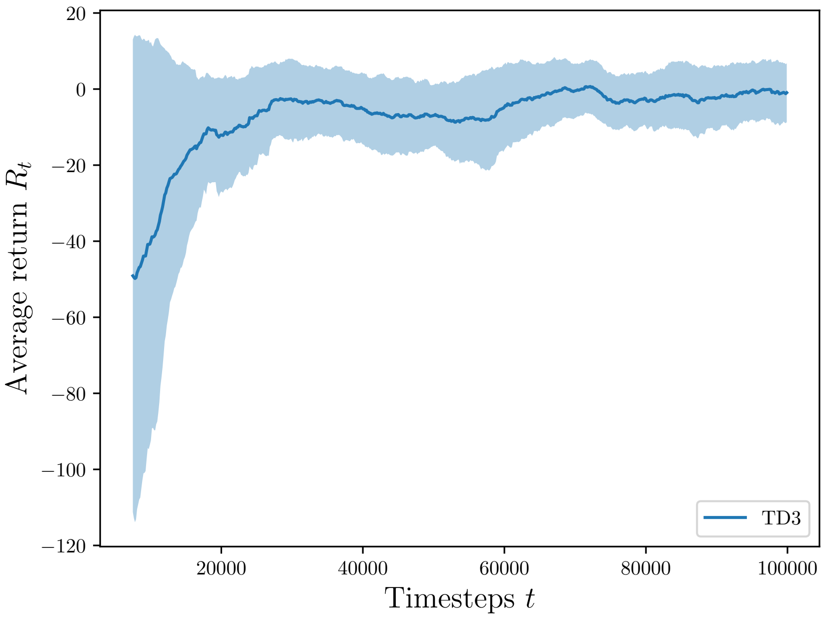


Figure 7: Learning curve TD3.

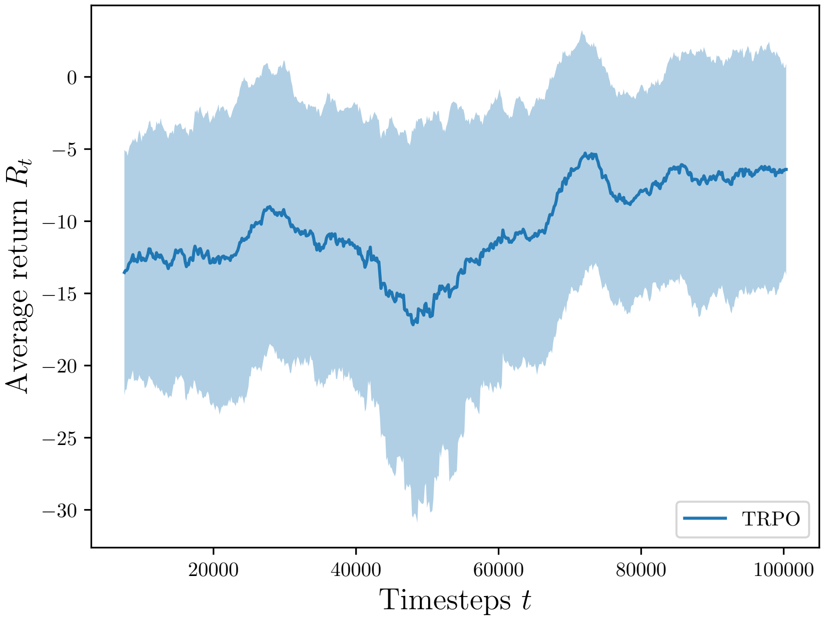


Figure 8: Learning curve TRPO.

2.3 Evaluation

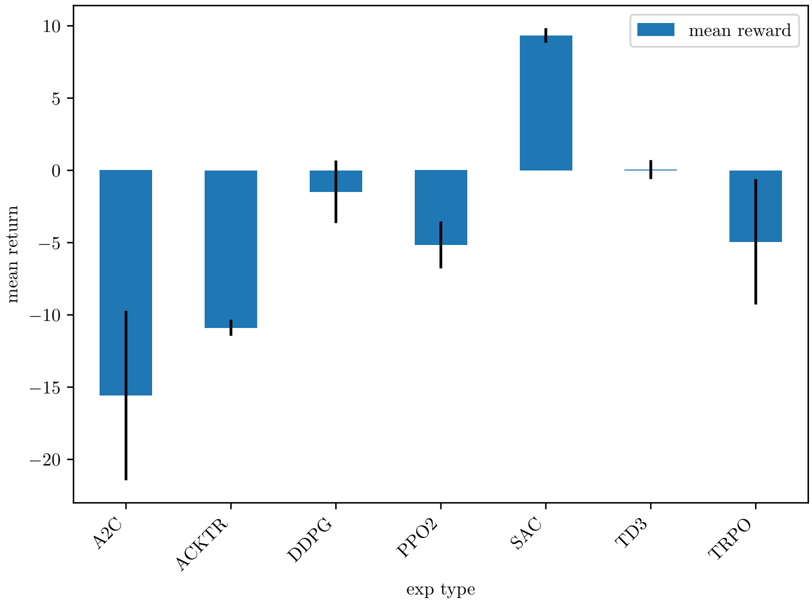


Figure 9: Mean reward vs algorithms.

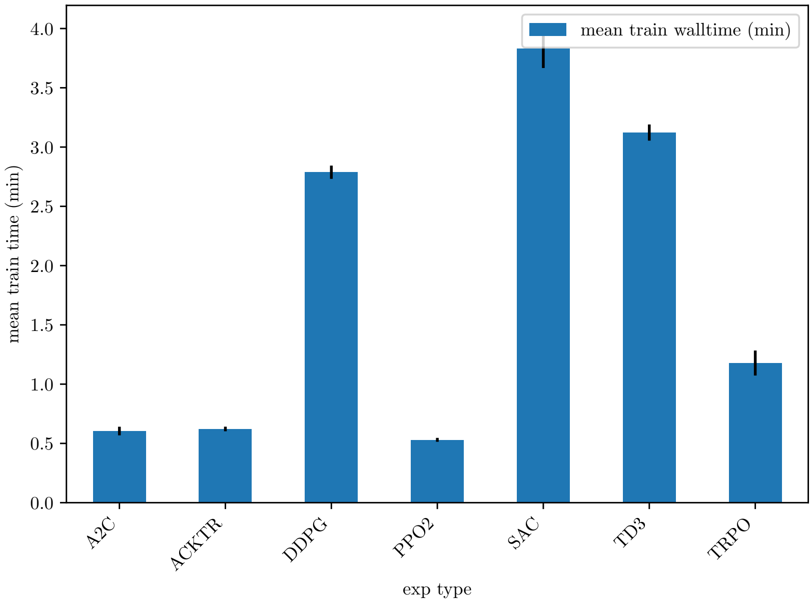


Figure 10: Mean walltime vs algorithms.

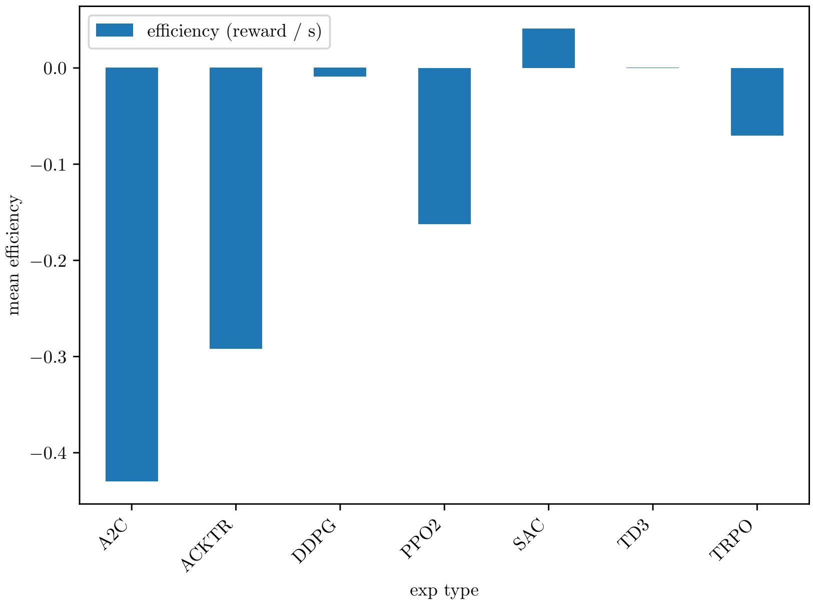
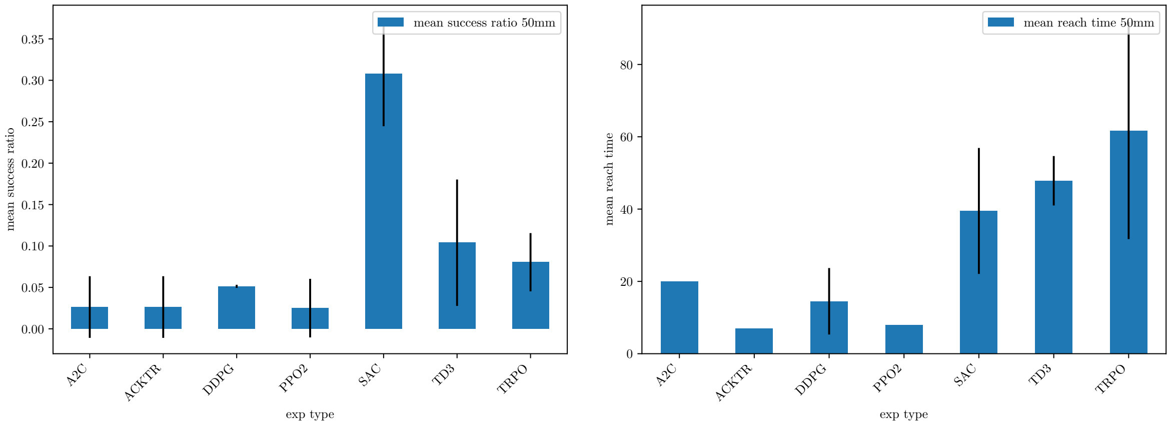


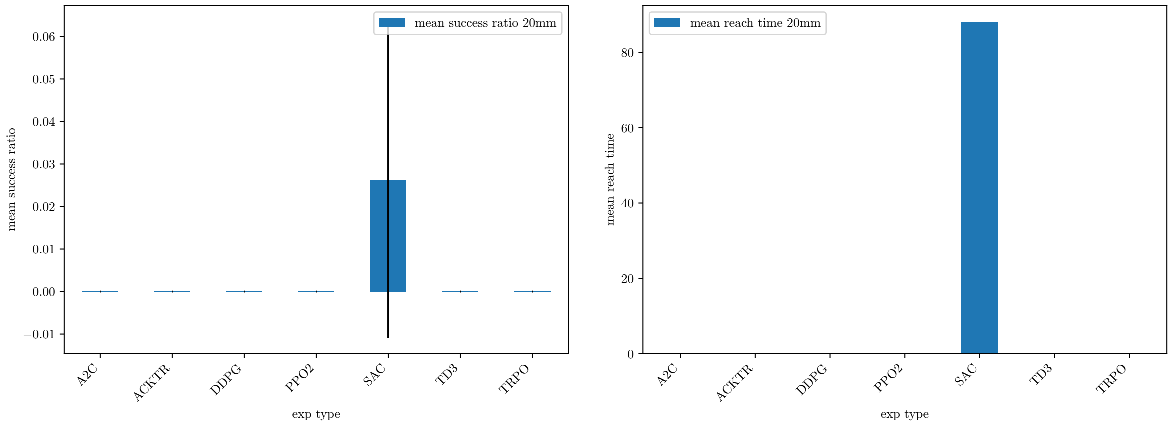
Figure 11: Efficiency vs algorithms.



(a) Success ratios

(b) Reach time

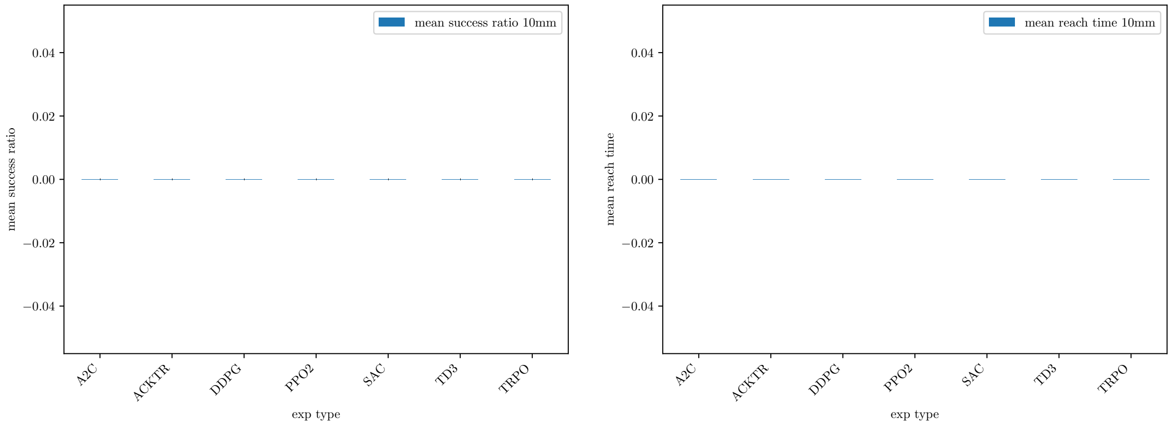
Figure 12: Success threshold: 50mm.



(a) Success ratios

(b) Reach time

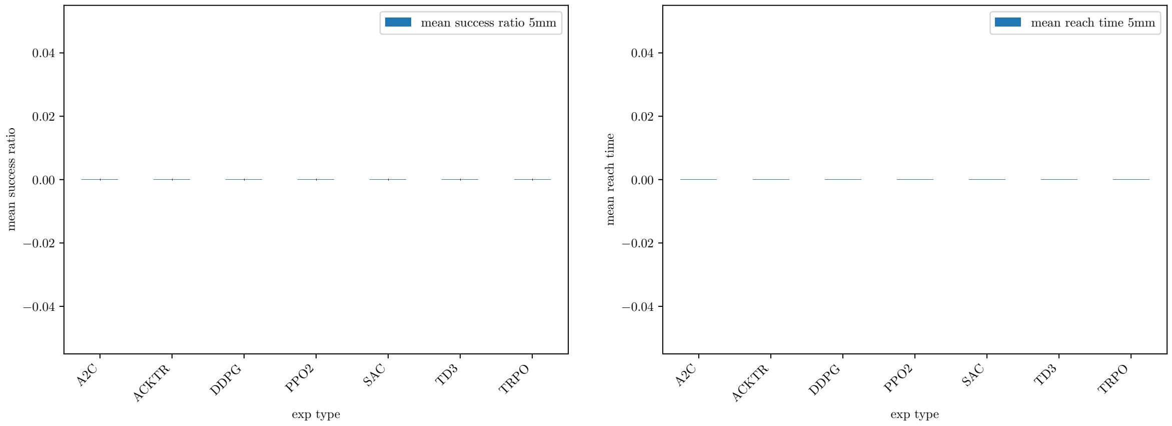
Figure 13: Success threshold: 20mm.



(a) Success ratios

(b) Reach time

Figure 14: Success threshold: 10mm.



(a) Success ratios

(b) Reach time

Figure 15: Success threshold: 5mm.

3 Findings summary