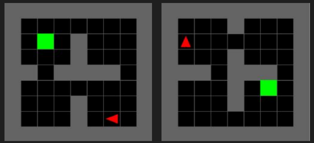
# Zero-Shot Generalization in Offline Reinforcement Learning with WSAC-N

**Maxime Museur** (M.D.I.Museur@student.tudelft.nl)

Supervisors: Dr. Matthijs Spaan, Max Weltevrede

## (1) Introduction

- Offline reinforcement learning (RL) = RL where agent cannot perform actions in environment, only has access to static dataset.
- Recent work has shown that offline RL does not generalize as well as behavioral cloning (BC). [1]
- We aim to:
  - **Compare generalization abilities between WSAC-N and baseline BC**
  - **Investigate effect of dataset size and quality on generalization**
- Environment from [2] (see figure 1)



## (2) Method

- Propose and implement **WSAC-N**
- = SAC-N [3] weighted with weights from SUNRISE DQN [4] to downweight actions with high variance
- Generate datasets with varying quality of policies: **expert**, **mixed suboptimal-expert**, **suboptimal**, **random**
- Compare generalization of WSAC-N with baseline BC
- Compare effect of dataset size and quality on generalization

TU Delft

CSE3000 Research Project
Final Presentation Poster

## (3) Generalization

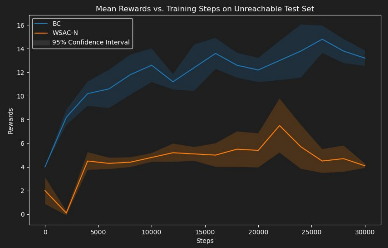
- Zero-shot generalization [5]
- Test sets from [2]:
  - Reachable = unseen agent location and direction, seen topology and goal location
  - Unreachable = unseen agent location, direction, topology, seen goal location

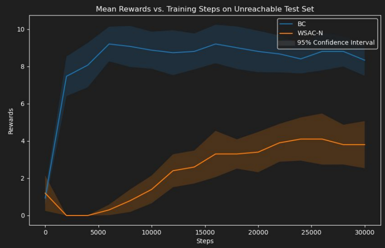
## References

[1] Qingfei You Ishita Mediratta, Minqi Jiang, and Roberta Raileanu. The generalization gap in offline reinforcement learning, 2024.
[2] Max Weltevrede, Matthijs T. J. Spaan, and Wendelin B öhmer. The role of diverse replay for generalization in reinforcement learning, 2023.
[3] Gaon An, Seungyong Moon, Jang-Hyun Kim, and Hyun Oh Song. Uncertainty-based offline reinforcement learning with diversified q-ensemble, 2021.
[4] Kimin Lee, Michael Laskin, Aravind Srinivas, and Pieter Abbeel. Sunrise: A simple unified framework for ensemble learning in deep reinforcement learning, 2021.
[5] Robert Kirk, Amy Zhang, Edward Grefenstette, and Tim Rockt äschel. A survey of zero-shot generalisation in deep reinforcement learning. Journal of Artificial Intelligence Research, 76:201–264, January 2023. ISSN 1076-9757. doi: 10.1613/jair.1.14174.

## (4) Experiments & Conclusions

a) Generalization of **WSAC-N** and **BC** with expert and non-expert datasets as training

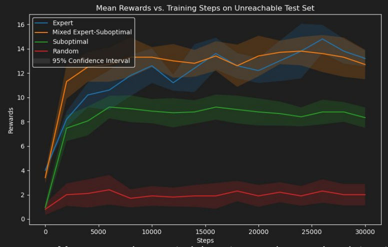


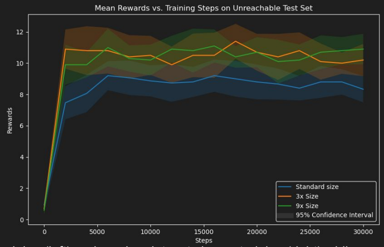




Mean rewards over various training steps, using expert datasets (left) and suboptimal datasets (right) for training and testing on the unreachable test set.

Conclusion: BC generalizes better than WSAC-N with both expert and non-expert datasets.

b) Effect of dataset quality and size on generalization




Mean rewards over training steps, using varying dataset quality as training (left) and varying dataset size as training (right) while testing on the unreachable test set.

Conclusion: Quality of data generally has a positive impact on generalization, and dataset size has negligible impact on generalization.