# Selectively Sharing Experiences Improves Multi-Agent Reinforcement Learning

Extended Abstract

Matthias Gerstgrasser
John A. Paulson School of Engineering and Applied Sciences, Harvard University
Cambridge, MA, United States
matthias@seas.harvard.edu

Tom Danino
The Taub Faculty of Computer Science, Technion - Israel Institute of Technology
Haifa, Israel
tom.danino@campus.technion.ac.il

Sarah Keren
The Taub Faculty of Computer Science, Technion - Israel Institute of Technology
Haifa, Israel
sarahk@cs.technion.ac.il

## ABSTRACT

We present a novel multi-agent RL approach, *Selective Multi-Agent Prioritized Experience Relay*, in which agents share with other agents a limited number of transitions they observe during training. The intuition behind this is that even a small number of relevant experiences from other agents could help each agent learn. Unlike many other multi-agent RL algorithms, this approach allows for largely decentralized training, requiring only a limited communication channel between agents. We show that our approach outperforms baseline no-sharing decentralized training and state-of-the art multi-agent RL algorithms. Further, sharing only a small number of highly relevant experiences outperforms sharing all experiences between agents, and the performance uplift from selective experience sharing is robust across a range of hyperparameters and DQN variants. A reference implementation is available under https://github.com/mgerstgrasser/super.

## KEYWORDS

Reinforcement Learning; Multi-Agent Reinforcement Learning; Cooperative AI

**ACM Reference Format:**
Matthias Gerstgrasser, Tom Danino, and Sarah Keren. 2023. Selectively Sharing Experiences Improves Multi-Agent Reinforcement Learning: Extended Abstract. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 3 pages.

## 1 INTRODUCTION

Multi-Agent Reinforcement Learning (RL) is often considered a hard problem: The environment dynamics and returns depend on the joint actions of all agents, leading to significant variance and non-stationarity in the experiences of each individual agent. Much recent work in multi-agent RL has focused on mitigating the impact of these [4, 6]. Our work goes in a different direction: leveraging the presence of other agents to collaboratively explore the environment more quickly.

We present a novel multi-agent RL approach that allows agents to share a small number of experiences with other agents. The intuition is that if one agent discovers something important in the environment, then sharing this with the other agents should help them learn faster. However, it is crucial that only important experiences are shared - we show that sharing all experiences indiscriminately will not improve learning. To this end, we make two crucial design choices: *selectivity* and *priority*. Selectivity means we only share a small fraction of experiences. Priority is inspired by a well-established technique in single-agent RL, *prioritized experience replay* (PER) [7]. With PER an off-policy algorithm such as DQN [5] will sample experiences not uniformly, but proportionally to "how far off" the current policy's predictions are in each state, formally the *temporal difference (td) error* . We use this metric to prioritize which experiences to share with other agents.

We dub the resulting multiagent RL approach *Selective Multi-Agent Prioritized Experience Relay* or *SUPER*. In this, agents independently use a DQN algorithm to learn, but with a twist: each agent relays its highest td-error experiences to the other agents, who insert them directly into their replay buffer, which is used for learning. This approach has several advantages:

(1) It consistently leads to faster learning and higher eventual performance, across hyperparameter settings.
(2) Unlike many "centralized training, decentralized execution" approaches, the SUPER learning paradigm allows for (semi-) decentralized training, requiring only a limited bandwidth communication channel between agents.
(3) The paradigm is agnostic to the underlying decentralized training algorithm, and can enhance many existing DQN algorithms. SUPER can be used together with PER, or without PER.[1]

In addition to the specific algorithm we develop, this work also introduces two key conceptual novelties.

(4) We show that communication can improve multi-agent RL even during training. Most prior work consider "learning to communicate", also known as emergent communication, which is a difficult problem but may improve coordination at convergence. We show that "communicate to learn" can also drastically improve performance during training.
(5) Related to this, we introduce the paradigm of "decentralized training with communication". This is a ‘middle ground between established approaches of decentralized and centralized training (including "centralized training, decentralized execution").

*Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.* © 2023 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

[1] Note that while the name "SUPER" pays homage to PER, SUPER is not related to PER other than using the same heuristic for relevance of experiences.

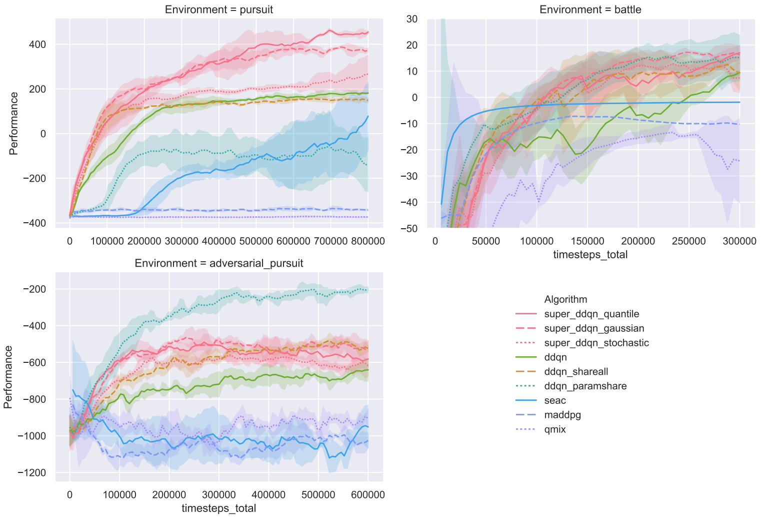

**Figure 1: Performance of SUPER variants and baselines on different domains.**

## 2 EXPERIMENTS AND RESULTS

We evaluate SUPER on a number of multiagent benchmark domains.

Figure 1 shows learning curves of SUPER implemented on dueling DDQN ("SUPER DDQN"), with stochastic, Gaussian and quantile experience selection and target bandwidth 0.1, compared to standard no-sharing dueling DDQN [9], share-all SUPER-DDQN, parameter-sharing dueling DDQN, as well as MADDPG [4], QMIX [6] and SEAC [1].

We find that SUPER (red curves) consistently outperforms the baseline DDQN algorithm (solid green curve), often significantly. In Pursuit, for instance, SUPER-DDQN achieves over twice the reward of no-sharing DDQN at convergence, increasing from 180.7 (std.dev 2.8) to 450.5 (std.dev 9.9) for quantile SUPER-DDQN measured at 800k training steps. Similar results hold in other environments, and when combining SUPER with a (non-double, non-dueling) DQN algorithm. SUPER also performs significantly better than SEAC (solid blue), MADDPG (dashed violet) and QMIX (dotted purple).

## 3 CONCLUSION & DISCUSSION

We present selective multiagent PER, a selective experience-sharing mechanism that can improve DQN-family algorithms in multiagent settings. Conceptually, our approach is rooted in the same intuition that Prioritized Experience Replay is based on, which is that td-error is a useful approximation of how much an agent could learn from a particular experience. In addition, we introduce the a second key design choice of selectivity, which allows semi-decentralized learning with small bandwidth, and drastically improves performance in some domains.

Our selective experience approach improves performance of both DQN and dueling DDQN baselines, and does so across a range of environments and hyperparameters. It outperforms state-of-the-art multi-agent RL algorithms, in particular MADDPG, QMIX and SEAC. The only pairwise comparison that SUPER loses is against parameter sharing in Adversarial-Pursuit, in line with a common observation that in practice parameter sharing often outperforms sophisticated multi-agent RL algorithms. However, we note that parameter sharing is an entirely different, fully centralized training paradigm. Furthermore, parameter sharing is limited in its applicability, and does not work well if agents need to take on different roles or behavior to successfully cooperate. We see this in the Pursuit domain, where parameter sharing performs poorly, and SUPER outperforms it by a large margin. The significantly higher performance than QMIX, MADDPG and SEAC is somewhat expected given that baseline non-sharing DQN algorithms often show state-of-the-art performance in practice, especially with regard to sample efficiency.

Our algorithm is different from the "centralized training, decentralized execution" baselines we compare against in the sense that it does not require fully centralized training. Rather, it can be implemented in a decentralized fashion with a communication channel between agents. We see that performance improvements scale down even to very low bandwidth, making this feasible even with limited bandwidth. We think of this scheme as "decentralized training with communication" and hope this might inspire other semi-decentralized algorithms. In addition to training, we note that such a "decentralized with communication" approach could potentially be deployed during execution, if agents keep learning. While this is beyond the scope of the current paper, in future work we would like to investigate if this could help when transferring agents to new domains, and in particular with adjusting to a sim-to-real gap.

We focus on the DQN family of algorithms in this paper. In future work, we would like to explore SUPER in conjunction with other off-policy RL algorithms such as SAC [2, 3] and DDPG [8]. If the improvements we see in this work hold for other algorithms and domains as well, this could improve multi-agent RL performance in many settings.

## REFERENCES

[1] Filippos Christianos, Lukas Schäfer, and Stefano Albrecht. Shared experience actor-critic for multi-agent reinforcement learning. *Advances in Neural Information Processing Systems*, 33:10707–10717, 2020.

[2] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning*, pages 1861–1870. PMLR, 2018.

[3] Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. In Yoshua Bengio and Yann LeCun, editors, *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, 2016.

[4] Ryan Lowe, Yi I Wu, Aviv Tamar, Jean Harb, OpenAI Pieter Abbeel, and Igor Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. *Advances in neural information processing systems*, 30, 2017.

[5] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. In *NIPS Deep Learning Workshop*. 2013.

[6] Tabish Rashid, Mikayel Samvelyan, Christian Schroeder de Witt, Gregory Farquhar, Jakob N Foerster, and Shimon Whiteson. Monotonic value function factorisation for deep multi-agent reinforcement learning. *Journal of Machine Learning Research*, 21, 2020.

[7] Tom Schaul, John Quan, Ioannis Antonoglou, and David Silver. Prioritized experience replay. In Yoshua Bengio and Yann LeCun, editors, *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, 2016.

[8] David Silver, Guy Lever, Nicolas Heess, Thomas Degris, Daan Wierstra, and Martin Riedmiller. Deterministic policy gradient algorithms. In *International conference on machine learning*, pages 387–395. PMLR, 2014.

[9] Hado Van Hasselt, Arthur Guez, and David Silver. Deep reinforcement learning with double q-learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30, 2016.