man benchmark. In *International Conference on Machine Learning*, 2020.
[4] Frederik Ebert, Chelsea Finn, Sudeep Dasari, Annie Xie, Alex Lee, and Sergey Levine. Visual foresight: Model-based deep reinforcement learning for vision-based robotic control. *arXiv preprint arXiv:1812.00568*, 2018.
[5] K. Zhang, Zhuoran Yang, and Tamer Başar. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *ArXiv*, abs/1911.10635, 2019.
[6] Chao Yu, Akash Velu, Eugene Vinitsky, Yu Wang, Alexandre M. Bayen, and Yi Wu. The surprising effectiveness of mappo in cooperative, multi-agent games. *ArXiv*, abs/2103.01955, 2021.
[7] Zhen Zhang, Dongqing Wang, and Junwei Gao. Learning automata-based multiagent reinforcement learning for optimization of cooperative tasks. *IEEE Transactions on Neural Networks and Learning Systems*, 32:4639–4652, 2021.
[8] Xingyu Wang and Diego Klabjan. Competitive multi-agent inverse reinforcement learning with sub-optimal demonstrations. In *International Conference on Machine Learning*, 2018.
[9] Mikayel Samvelyan, Tabish Rashid, C. S. D. Witt, Gregory Farquhar, Nantas Nardelli, Tim G. J. Rudner, Chia-Man Hung, Philip H. S. Torr, Jakob N. Foerster, and Shimon Whiteson. The starcraft multi-agent challenge. *ArXiv*, abs/1902.04043, 2019.
[10] Jiechuan Jiang and Zongqing Lu. The emergence of individuality. In *International Conference on Machine Learning*, pages 4992–5001. PMLR, 2021.
[11] Kyunghwan Son, Daewoo Kim, Wan Ju Kang, David Earl Hostallero, and Yung Yi. Qtran: Learning to factorize with transformation for cooperative multi-agent reinforcement learning. *ArXiv*, abs/1905.05408, 2019.
[12] Yue Wang, Yao Wan, Chenwei Zhang, Lixin Cui, Lu Bai, and Philip S. Yu. Competitive multi-agent deep reinforcement learning with counterfactual thinking. *2019 IEEE International Conference on Data Mining*, pages 1366–1371, 2019.
[13] Michael L. Littman. Markov games as a framework for multi-agent reinforcement learning. In *Proceedings of the Eleventh International Conference on International Conference on Machine Learning*, page 157–163, 1994.
[14] Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, P. Abbeel, and Igor Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. In *Neural Information Processing Systems*, 2017.
[15] David Silver, Guy Lever, Nicolas Manfred Otto Heess, Thomas Degris, Daan Wierstra, and Martin A. Riedmiller. Deterministic policy gradient algorithms. In *International Conference on Machine Learning*, 2014.
[16] Vijay R Konda and John N Tsitsiklis. Actor-critic algorithms. In *Advances in neural information processing systems*, pages 1008–1014, 2000.
[17] Chenghao Li, Tonghan Wang, Chengjie Wu, Qianchuan Zhao, Jun Yang, and Chongjie Zhang. Celebrating diversity in shared multi-agent reinforcement learning. *Advances in Neural Information Processing Systems*, 34, 2021.
[18] Karol Kurach, Anton Raichuk, Piotr Stanczyk, Michal Zajac, Olivier Bachem, Lasse Espeholt, Carlos Riquelme, Damien Vincent, Marcin Michalski, Olivier Bousquet, and Sylvain Gelly. Google research football: A novel reinforcement learning environment. *ArXiv*, abs/1907.11180, 2020.
[19] Tim Franzmeyer, Mateusz Malinowski, and Joao F. Henriques. Learning altruistic behaviours in reinforcement learning without external rewards. In *International Conference on Learning Representations*, 2022.
[20] Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, Pieter Abbeel, and Igor Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. *Neural Information Processing Systems*, 2017.

付　　録

A　実験環境

本稿で実験を行った環境である PressurePlate において，エージェントの数やオブジェクトの配置などを変更した 2 つの環境について実験を行った．図 A·1 に PressurePlate_2 の環境を，図 A·2 に PressurePlate_3 の環境をそれぞれ示す．

	G			
W	W	D	D	W
A1	P			
A2				

図 **A·1**　PressurePlate_2 の環境．

	G			
W	W	D	D	W
	P			
W	W	D	D	W
A3	P			
A2	A1			

図 **A·2**　PressurePlate_3 の環境．

A1, A2, A3 はそれぞれエージェント, W は壁, G はゴール, D はドア, P はプレートの位置を示している．また PressurePlate_3 の環境ではプレートが二か所存在しているが, 上から 4 列目のプレートが 3 列目のドアに, 上から 7 列目のプレートが 6 列目のドアにそれぞれ対応している．またどちらの環境においてもエージェントの観測は変化せず, 自身を中心とした 5×5 の範囲を観測として獲得できる．