gameplay becomes extreme in learning between those who can use equally sophisticated (i.e., multi-memory) strategies. We also found a novel problem that Nash equilibrium is difficult to reach in multi-memory zero-sum games. Here, note that convergence to Nash equilibrium, either as a last-iterate [32, 33, 34, 35, 36, 37, 38] or as an average of trajectories [39, 40, 41], is a frequently discussed topic. In general, heteroclinic cycles fail to converge even on average. What algorithm can converge to Nash equilibrium in multi-memory zero-sum games would be interesting future work.

References

[1] Drew Fudenberg and Jean Tirole. *Game theory.* MIT press, 1991.

[2] John F Nash Jr. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49, 1950.

[3] John G Cross. A stochastic learning model of economic behavior. *The Quarterly Journal of Economics*, 87(2):239–266, 1973.

[4] Tilman Börgers and Rajiv Sarin. Learning through reinforcement and replicator dynamics. *Journal of Economic Theory*, 77(1):1–14, 1997.

[5] Josef Hofbauer, Karl Sigmund, et al. *Evolutionary games and population dynamics.* Cambridge university press, 1998.

[6] Satinder Singh, Michael J Kearns, and Yishay Mansour. Nash convergence of gradient dynamics in general-sum games. In *UAI*, pages 541–548, 2000.

[7] Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *ICML*, pages 928–936, 2003.

[8] Michael Bowling and Manuela Veloso. Multiagent learning using a variable learning rate. *Artificial Intelligence*, 136(2):215–250, 2002.

[9] Michael Bowling. Convergence and no-regret in multiagent learning. In *NeurIPS*, pages 209–216, 2004.

[10] Christopher JCH Watkins and Peter Dayan. Q-learning. *Machine learning*, 8(3):279–292, 1992.

[11] Michael Kaisers and Karl Tuyls. Frequency adjusted multi-agent q-learning. In *AAMAS*, pages 309–316, 2010.

[12] Sherief Abdallah and Michael Kaisers. Addressing the policy-bias of q-learning by repeating updates. In *AAMAS*, pages 1045–1052, 2013.

[13] Panayotis Mertikopoulos and William H Sandholm. Learning in games via reinforcement and regularization. *Mathematics of Operations Research*, 41(4):1297–1324, 2016.

[14] Panayotis Mertikopoulos, Christos Papadimitriou, and Georgios Piliouras. Cycles in adversarial regularized learning. In *SODA*, pages 2703–2717, 2018.

[15] Karl Tuyls and Ann Nowé. Evolutionary game theory and multi-agent reinforcement learning. *The Knowledge Engineering Review*, 20(1):63–90, 2005.

[16] Karl Tuyls, Pieter Jan'T Hoen, and Bram Vanschoenwinkel. An evolutionary dynamical analysis of multi-agent learning in iterated games. *Autonomous Agents and Multi-Agent Systems*, 12(1):115–153, 2006.

[17] Daan Bloembergen, Karl Tuyls, Daniel Hennes, and Michael Kaisers. Evolutionary dynamics of multi-agent learning: A survey. *Journal of Artificial Intelligence Research*, 53:659–697, 2015.

[18] Wolfram Barfuss. Towards a unified treatment of the dynamics of collective learning. In *Challenges and Opportunities for Multi-Agent Reinforcement Learning, AAAI Spring Symposium*, 2020.