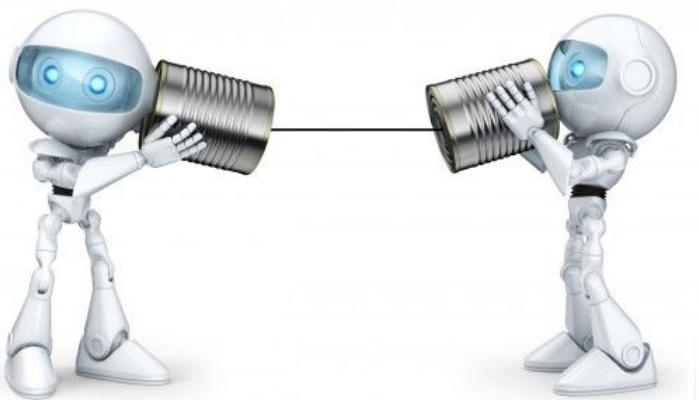
Training Multiple Intelligent Agents to Communicate

Varun Bhatt
MSc; supervised by Prof. Michael Buro



Picture from http://telecoms.com/430361/europe-leads-m2m-and-iot-growth-says-vodafone/

Examples of Human Communication

Sharing observations

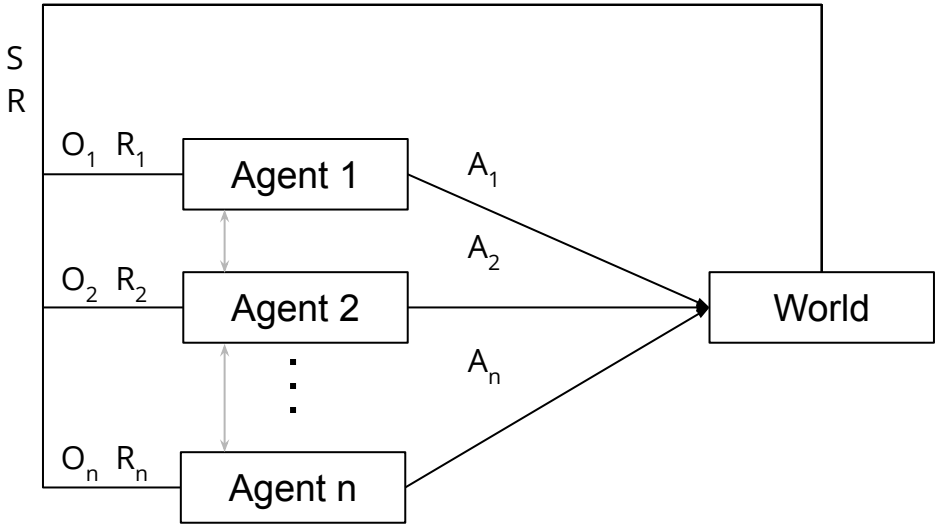
Suggesting action

Sharing action intention

Pictures from https://www.nps.gov/pefo/planyourvisit/safety.htm?fullweb=1, http://www.soccer-city.com/training-blog/2017/3/7/the-importance-of-on-field-communication, http://www.roadsafetyuae.com/changing-lanes-properly/

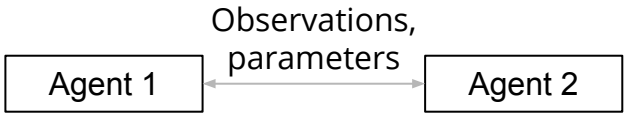
Formalism

Fully cooperative, partially observable markov game with cheap talk channel
(dec-POMDP)



Previous work

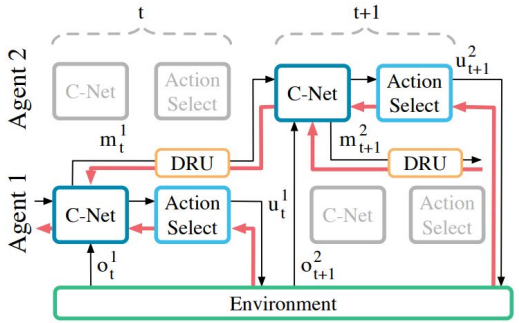
Pre-defined communication protocol (Tan, 1993)



Game theory, equilibrium (Claus & Boutilier, 1998)

Agent 1's action \ Agent 2's action		
	11	-30
	-30	7

End to end differentiable communication channel (Foerster et. al., 2016)

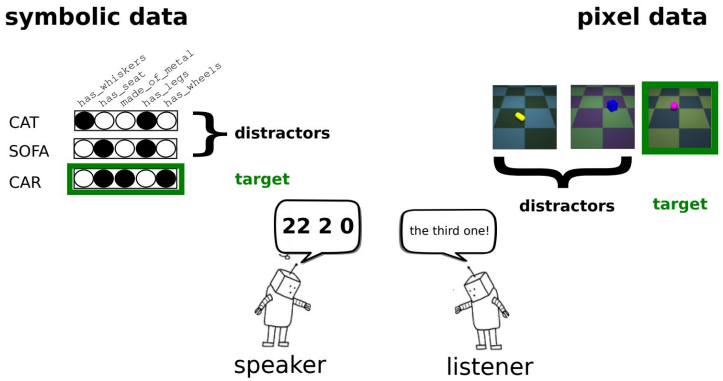


Tan, M. (1993). Multi-agent reinforcement learning: Independent vs. cooperative agents. In Proceedings of the Tenth International Conference on Machine Learning (pp. 330-337).
Claus, C., & Boutilier, C. (1998). The dynamics of reinforcement learning in cooperative multiagent systems. AAAI/IAAI, 1998, 746-752.
Foerster, J., Assael, I. A., de Freitas, N., & Whiteson, S. (2016). Learning to communicate with deep multi-agent reinforcement learning. In Advances in Neural Information Processing Systems (pp. 2137-2145).

Previous work

Communication only through actions (Bard et. al., 2019)

Emergent Communication (Lazaridou et. al., 2018)

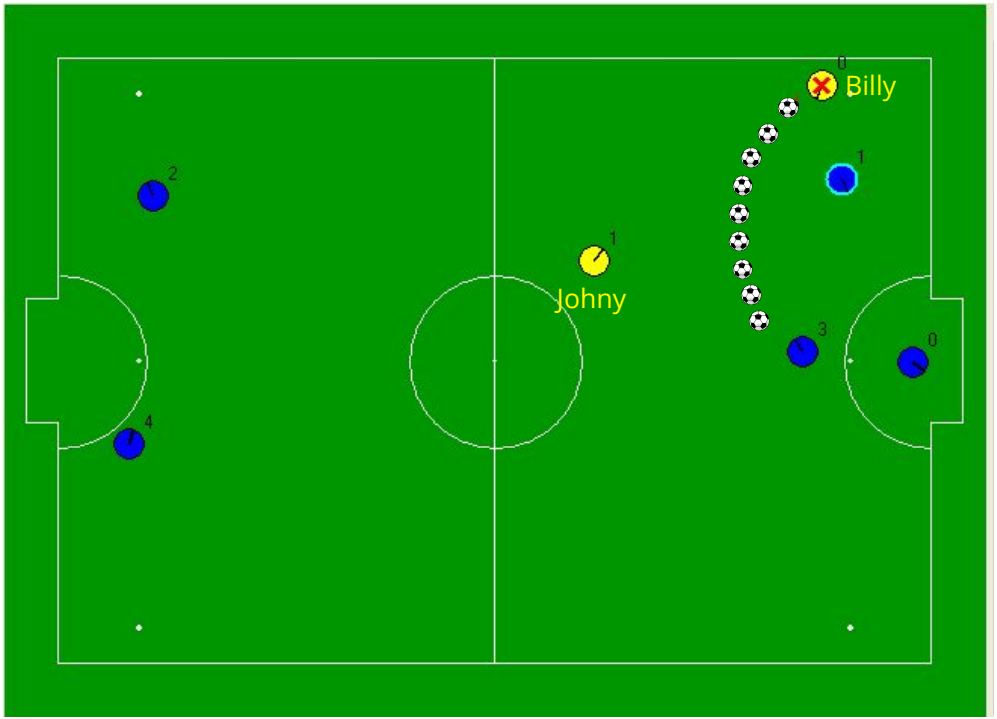


Bard, N., Foerster, J. N., Chandar, S., Burch, N., Lanctot, M., Song, H. F., ... & Dunning, I. (2019). The Hanabi Challenge: A New Frontier for AI Research. arXiv preprint arXiv:1902.00506.
Lazaridou, A., Hermann, K. M., Tuyls, K., & Clark, S. (2018). Emergence of linguistic communication from referential games with symbolic and pixel input. arXiv preprint arXiv:1804.03984.
Hanabi picture from https://rulesofplay.co.uk/products/hanabi-en?variant=824682117

Issues - Credit Assignment

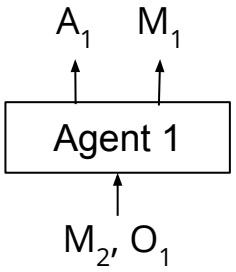
Who's to blame?

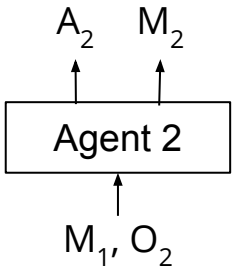
Johny for the call?
or Billy for the bad pass?



Picture from http://dcis.inf.fu-berlin.de/rojas/fu-fighters-autonomous-soccer-robots/fu-fighters-the-soccer-robots-of-freie-universitat-berlin/

Issues - Partial Observability



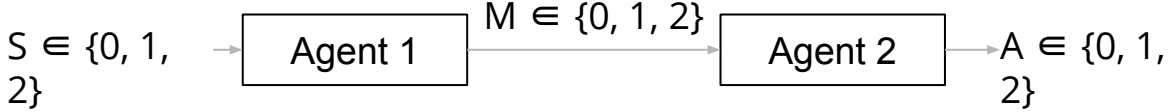


Agent 1

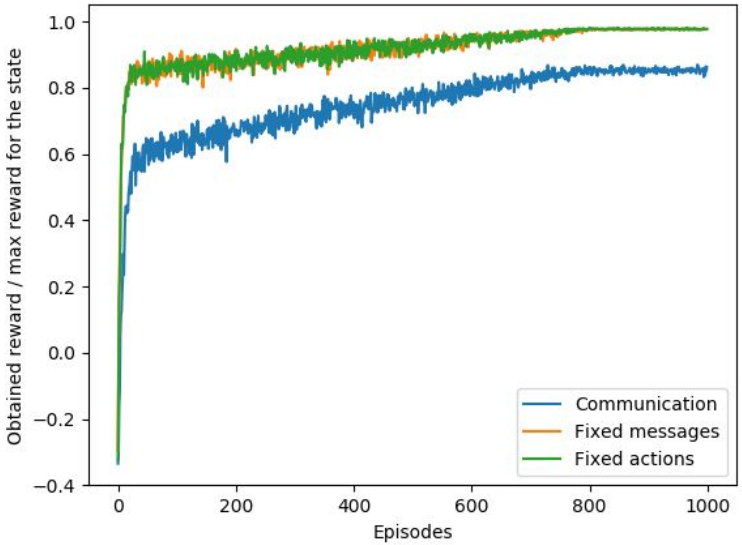
- Can’t see O_2
- Doesn’t know if M_1 was used or ignored

A Simple Experiment - Modified Climbing Game

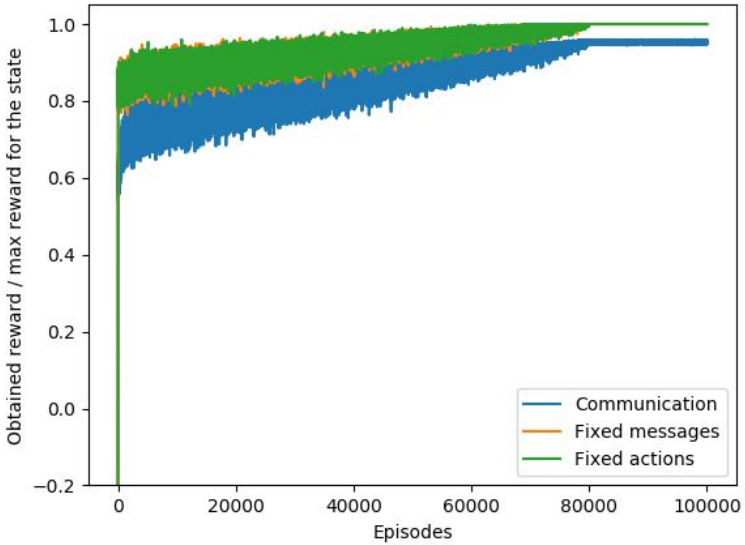
Agent 1's action \ Agent 2's action			
	11	-30	0
	-30	7	6
	0	0	5



Results - Reward



Results - Reward

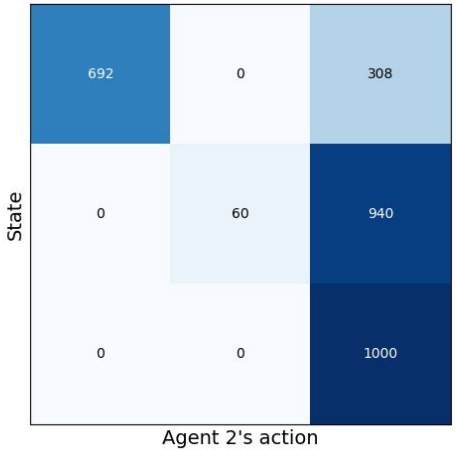


Results - Convergence points

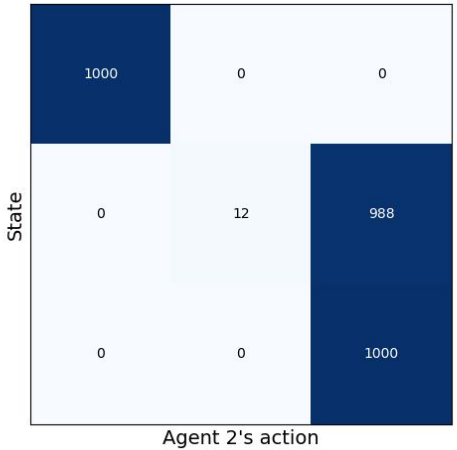
Agent 2's action

State			
	11	-30	0
	-30	7	6
	0	0	5

Short run

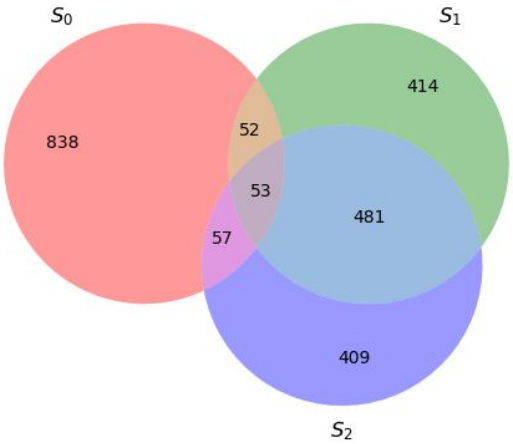


Long run

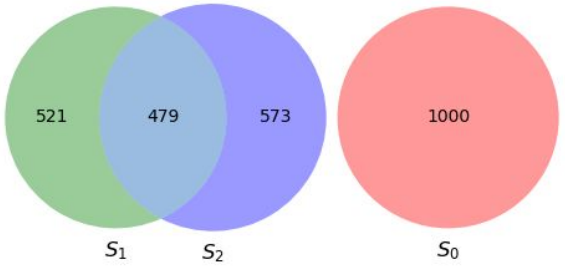


Results - Message protocol

Short run



Long run

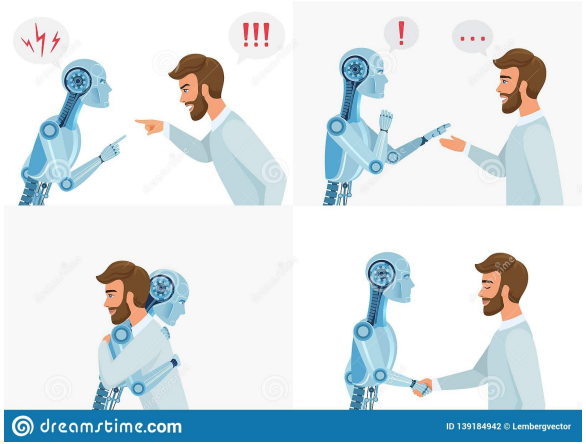


Message is not unique to state

Ideas

Two timescale optimization/exploration

Two way communication



Questions?

Picture from https://www.dreamstime.com/artific-intelligence-interaction-concept-human-robot-modern-communication-business-technology-vector-illustration-image139184942