Learning to Communicate with Deep Multi-Agent RL

Jakob N. Foerster*, Yannis M. Assael*, Nando de Freitas, Shimon Whiteson

https://arxiv.org/abs/1605.06676



UNIVERSITY OF OXFORD

Motivation

"*In the beginning was the Word, and the Word was with God...*" [1]

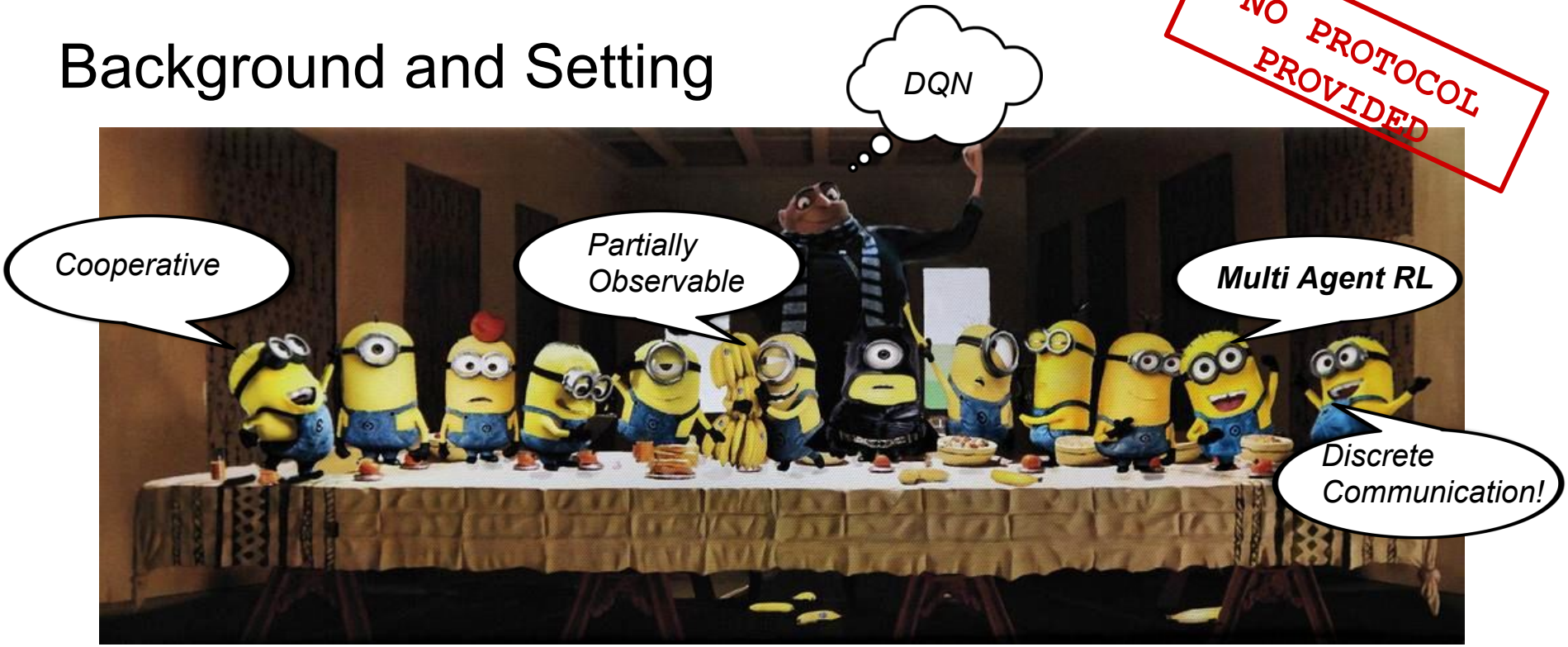
> "*..as one people speaking the same language they have begun to do this, then nothing they plan to do will be impossible..*" [2]

"*..debates on these questions have been so fiery that in 1866 the French Academy of Sciences banned publications about the origin of human language.*" [3]

http://openprtspecs.blogspot.com

[1] John et al., The Bible, Judea, 90-110 A.D.
[2] Yawest et al., Genesis, Judea, 600-500 B.C.
[3] Christiansen and Kirby, Language evolution. OUP Oxford, 2003.

Background and Setting

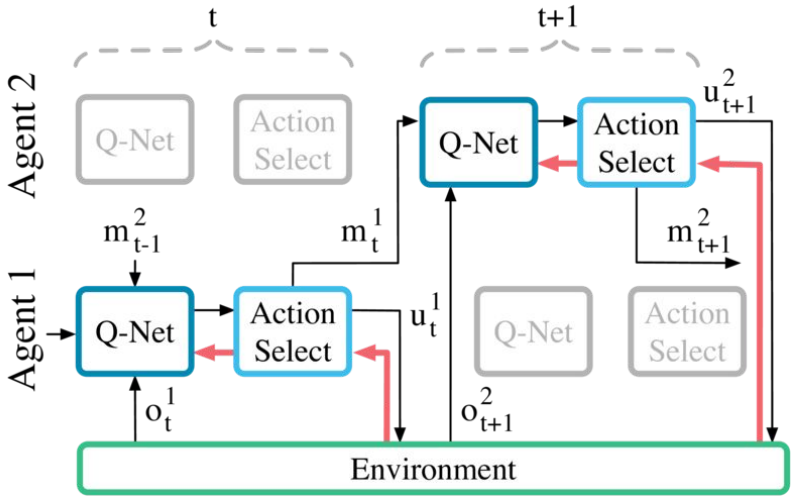


Also: Centralised Learning vs Decentralised Execution

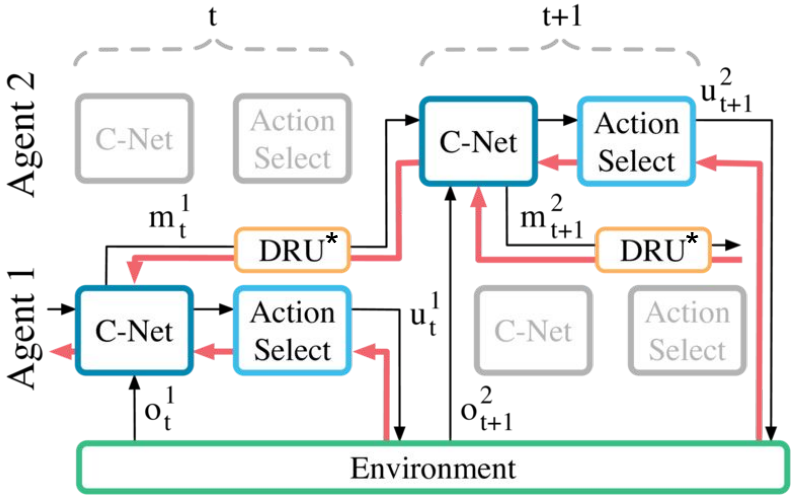
Methods

***D**ifferentiable **I**nter-**A**gent **L**earning & **R**einforced **I**nter-**A**gent **L**earning*

Methods



a) RIAL - RL based Communication

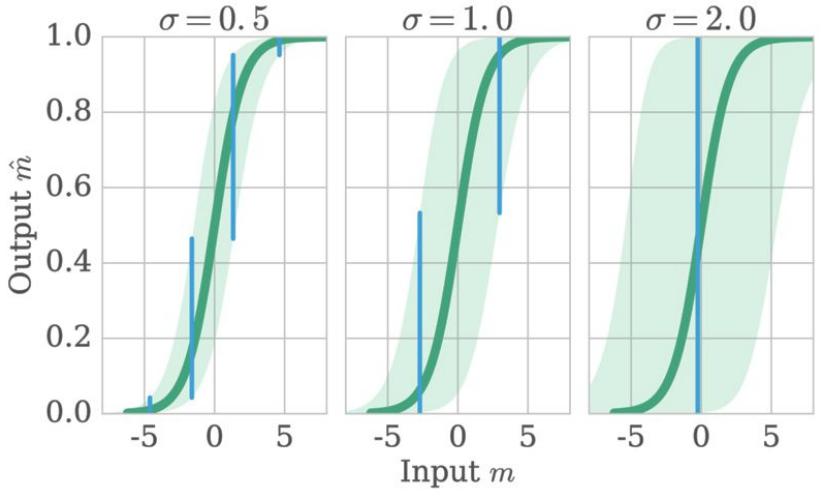


b) DIAL - Differentiable Communication

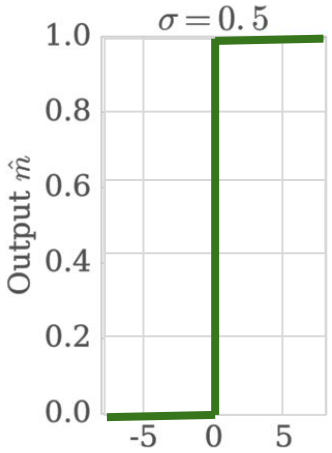
$$^{*}\mathrm{DRU}(m) = \begin{cases} \mathrm{Logistic}(\mathcal{N}(m, \sigma)), \text{ if training, else} \\ \mathbb{1}\{m > 0\} \end{cases}$$

Methods - DRU

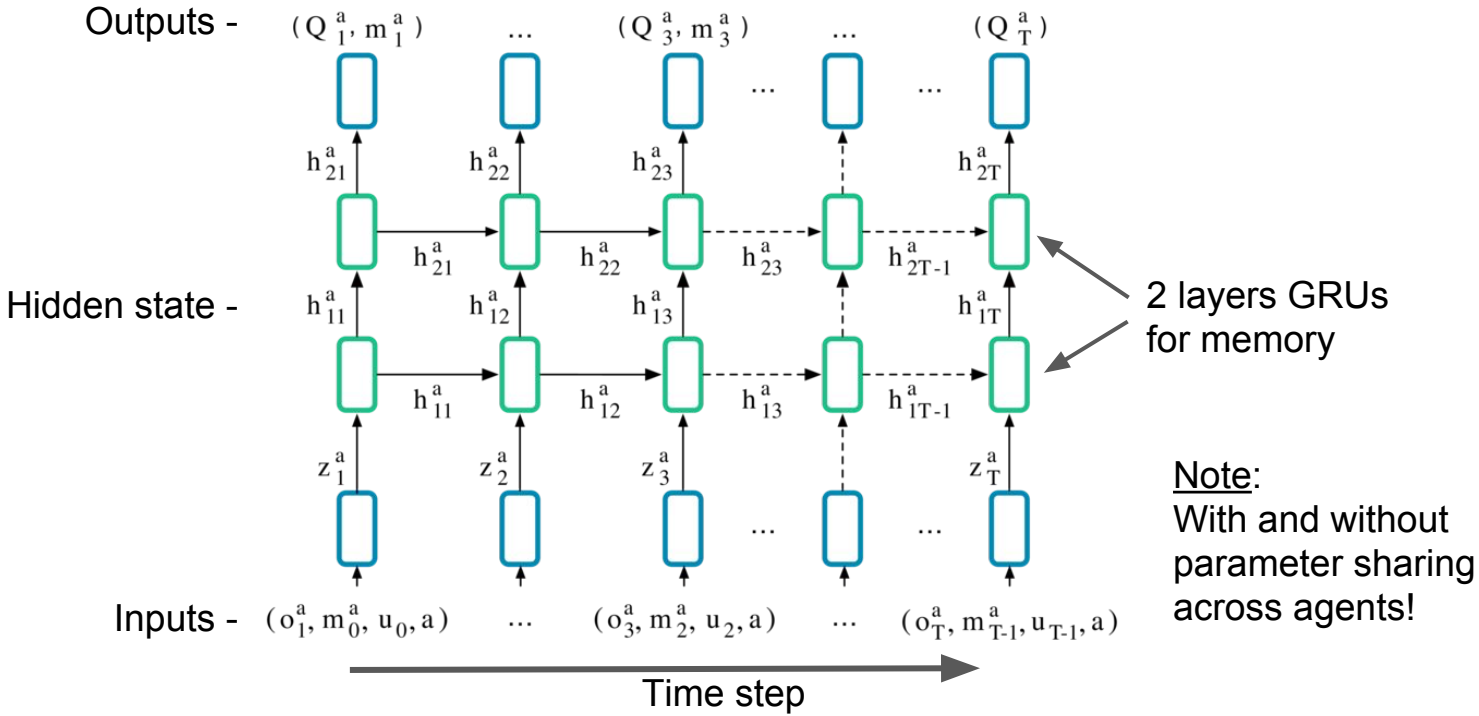
Centralised Training



Decentralised Execution

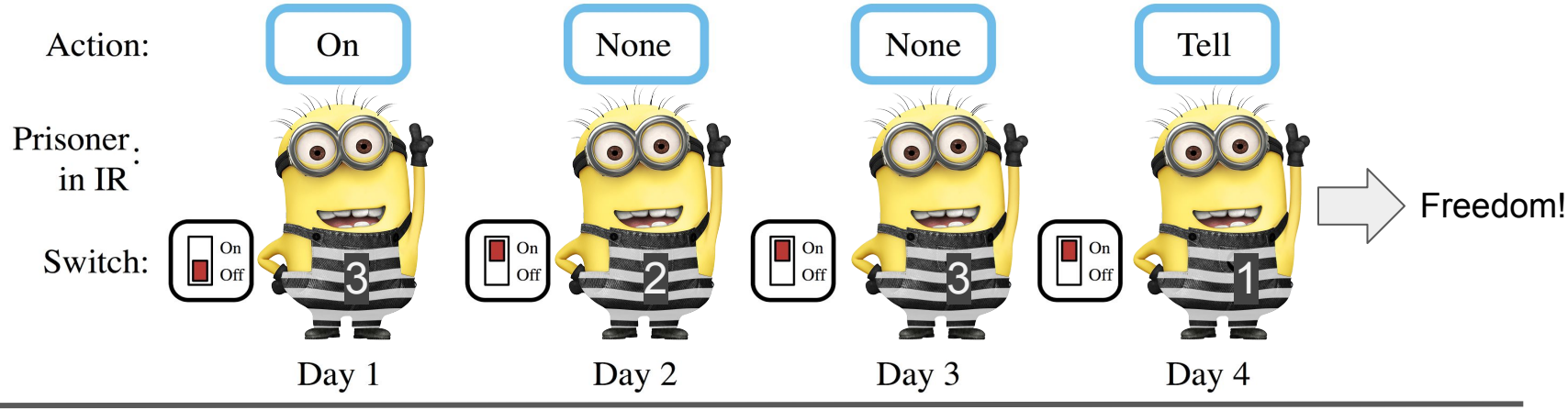


Methods - Architecture



Experiments - Switch

Riddle:



RL:

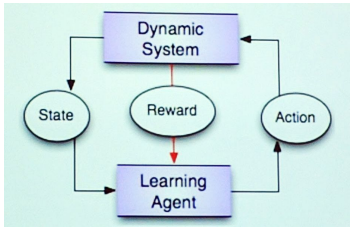
Multi-Agent: N agents with 1-bit communication channel

State: N-bit array: has i-th prisoner been to IR

Action: ‘Tell’ / ‘None’

Reward: + 1 (*freedom*) / 0 (*episode expires*) / -1 (*all die*)

Observation: ‘None’ OR switch



Experiments - Switch Complexity Analysis

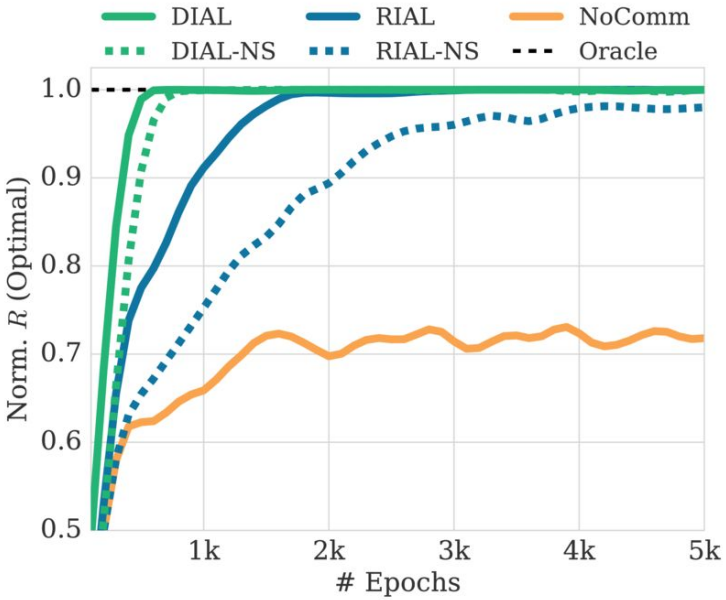
For n Minions:

$$4^{n(3^{T+1}-3)/2}$$

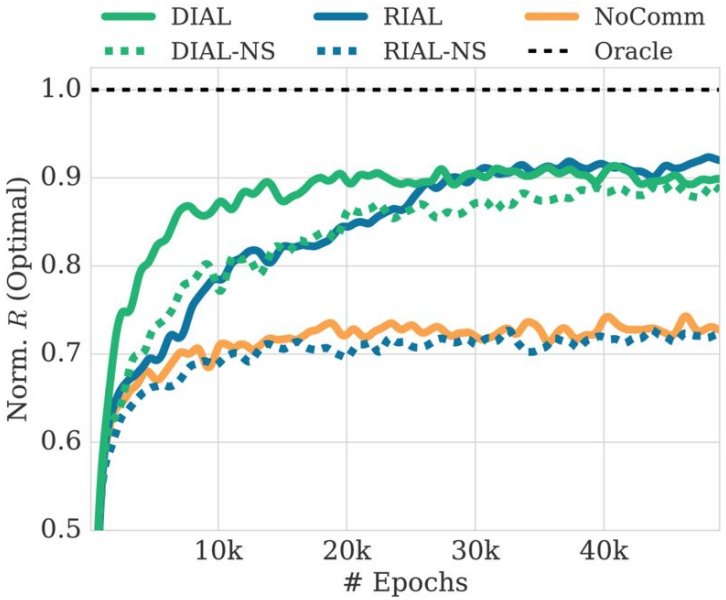
For 4 Minions:

$$4^{354288}$$

Experiments - Switch

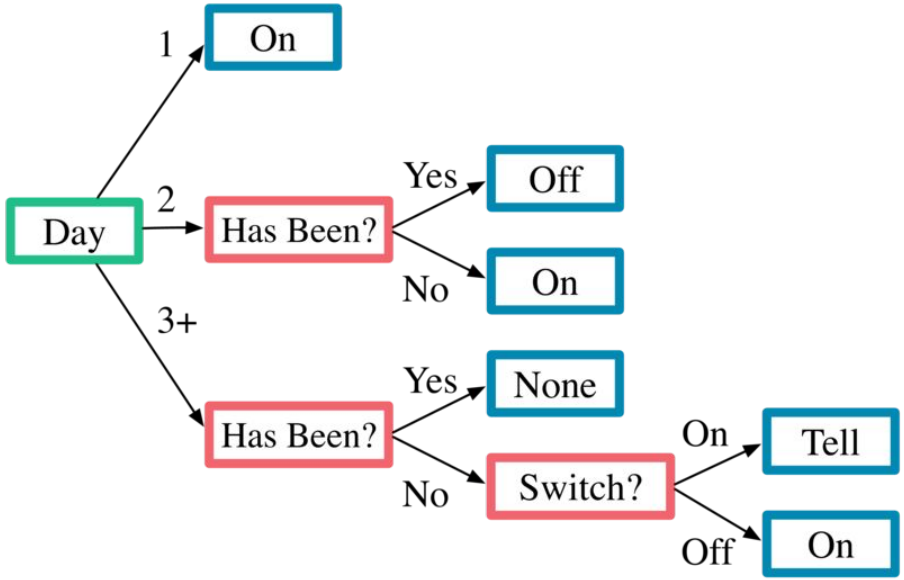


3 Minions



4 Minions

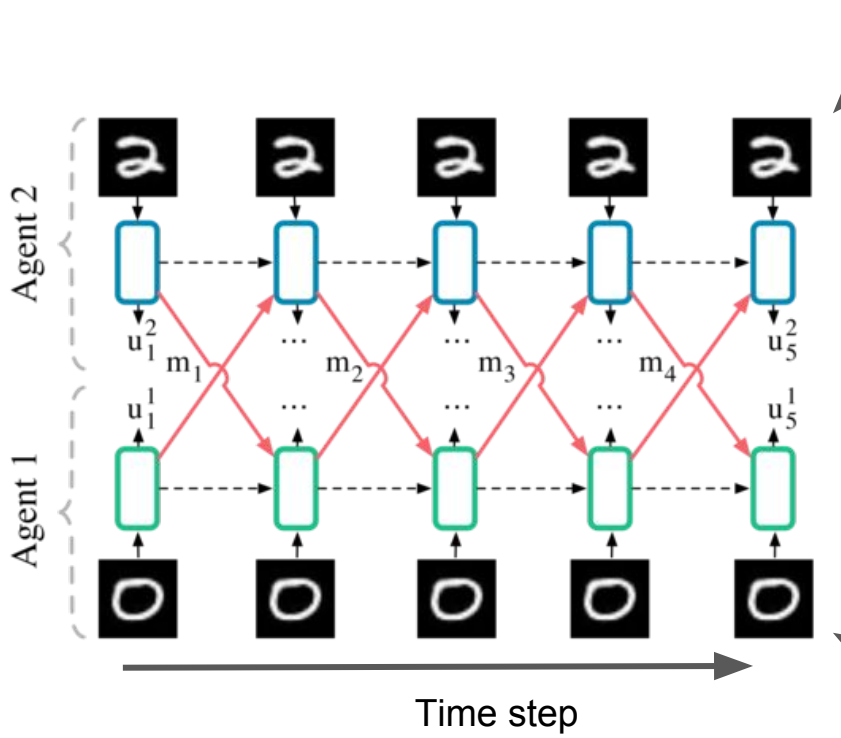
Experiments - Switch Strategy



3 Minions Strategy

Experiments - MNIST Games

Experiments - MNIST Game



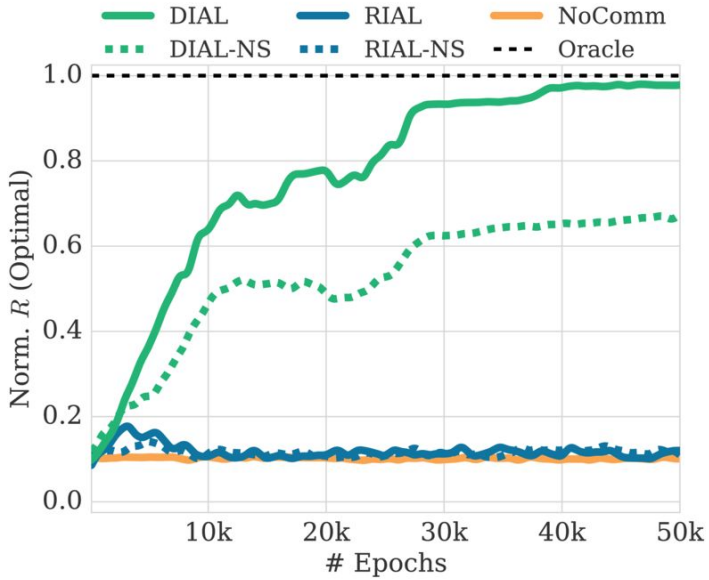
Reward:

- +1 (correct guess)
- 0 (wrong guess)

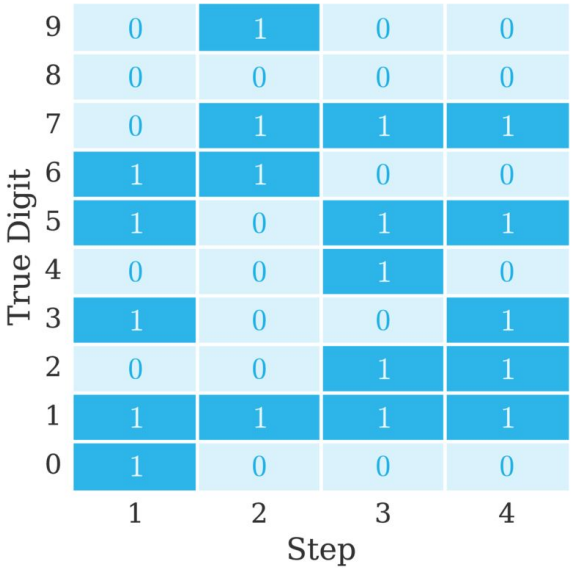
Communication:

- 1-bit per step (*m*)

Experiments - MNIST Results and Strategy

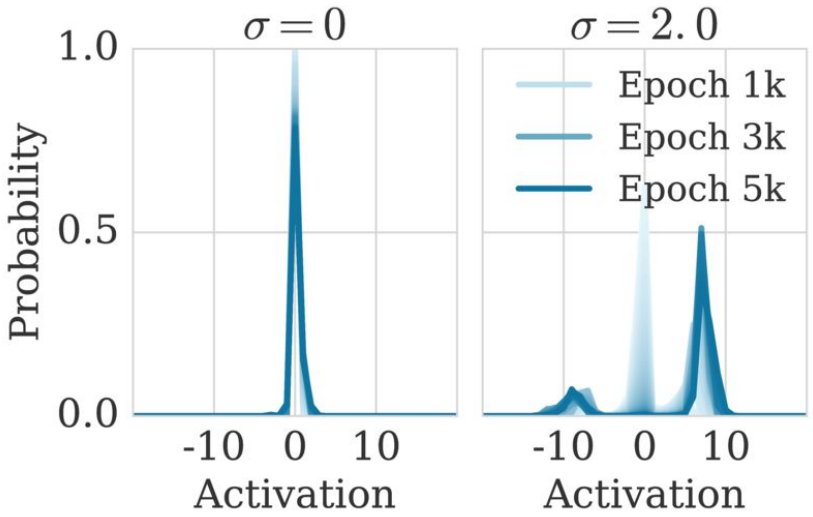


Results



Strategy

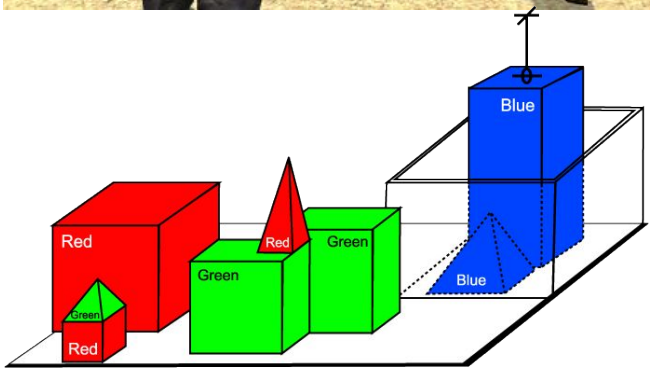
Experiments - Impact of Noise



Future Work

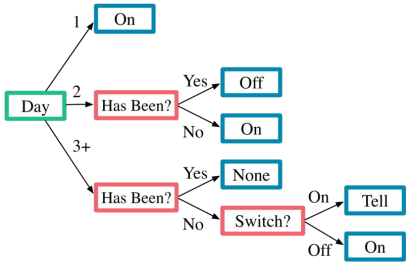

ROYAL FLUSH
STRAIGHT FLUSH

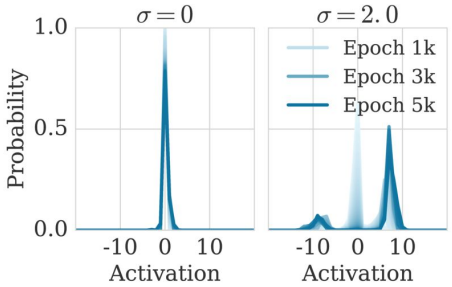

FULL HOUSE
FLUSH


Blue
Red
Red
Green
Green
Blue
Green
Red

Conclusions

- AI agents can discover communication protocols through Deep RL
- Gradients used in DIAL allow for faster learning
- Parameter sharing can accelerate learning
- Adding noise to the channel => bimodal distribution
- Protocols can be extracted and understood
- A lot of exciting work to be done here





PS: Code will be published online

thank you!

Q&A