"Emergence of Grounded Compositional Language in Multi-Agent Populations"

Mordatch (OpenAI) & Abbeel (Berkeley), 2018

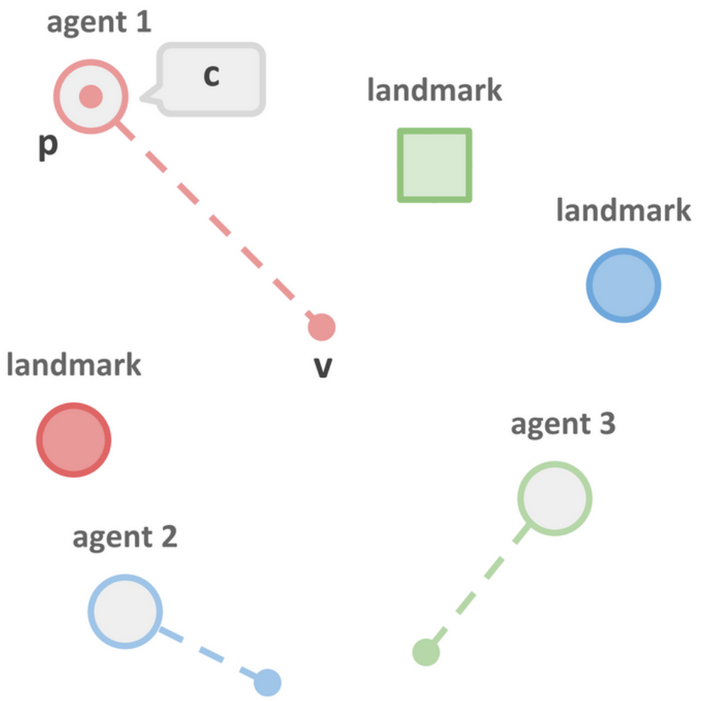
Presented by Andre

Let's go! ➜

Abstract

"...we propose a multi-agent learning environment and learning methods that **bring about emergence of a basic compositional language**. This language is represented as streams of abstract discrete symbols uttered by agents over time, but nonetheless has a coherent structure that possesses a defined vocabulary and syntax."

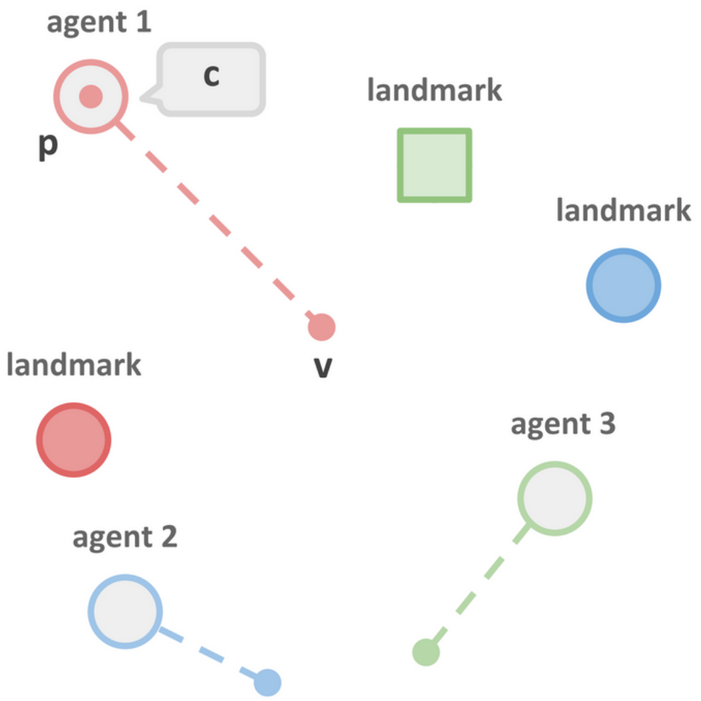
General Game Setup



- Cooperative partially observable Markov game
- Players share the same policy, action spaces, and observation spaces
- Objective: find the policy that maximizes the expected total reward across players

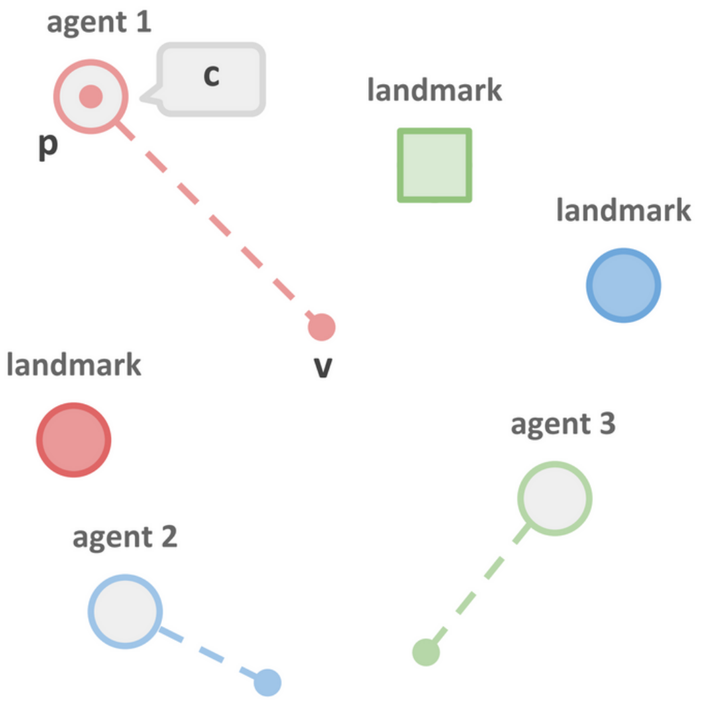
$$\max_{\boldsymbol{\pi}} R(\boldsymbol{\pi}), \quad \text{where} \quad R(\boldsymbol{\pi}) = \mathbb{E}\left[\sum_{t=0}^{T}\sum_{i=0}^{N} r(\mathbf{s}_i^t, \mathbf{a}_i^t)\right]$$

Environment Conditions



- Agents and landmarks possess intrinsic observable features, like color and shape
- Agents can gaze and move in a direction
- Each agent has internal tasks, either to gaze or move in a direction; not observed by others
    - Agent 1 may need Agent 2 to move in order to accomplish their task

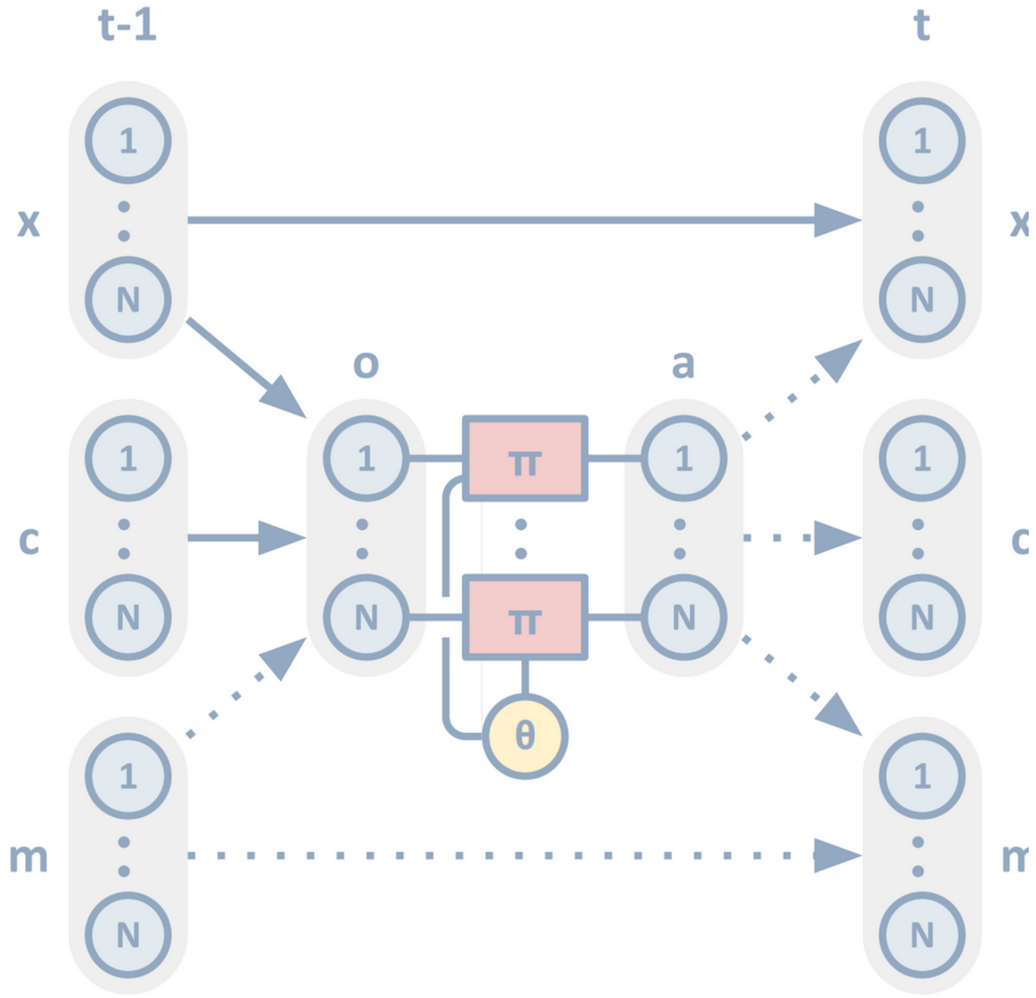
Communication Protocols



- Agents and communicate with other agents to coordinate action to accomplish tasks
- Communication is symbolic, sequential, not optimized (abstract), broadcasted to everyone, & executed at every time step
- Communicating is costly ('metabolic effort')
- Maximum vocabulary size 20
  - Larger vocabulary sizes are harder to train
  - Penalties to avoid repres. redundancy

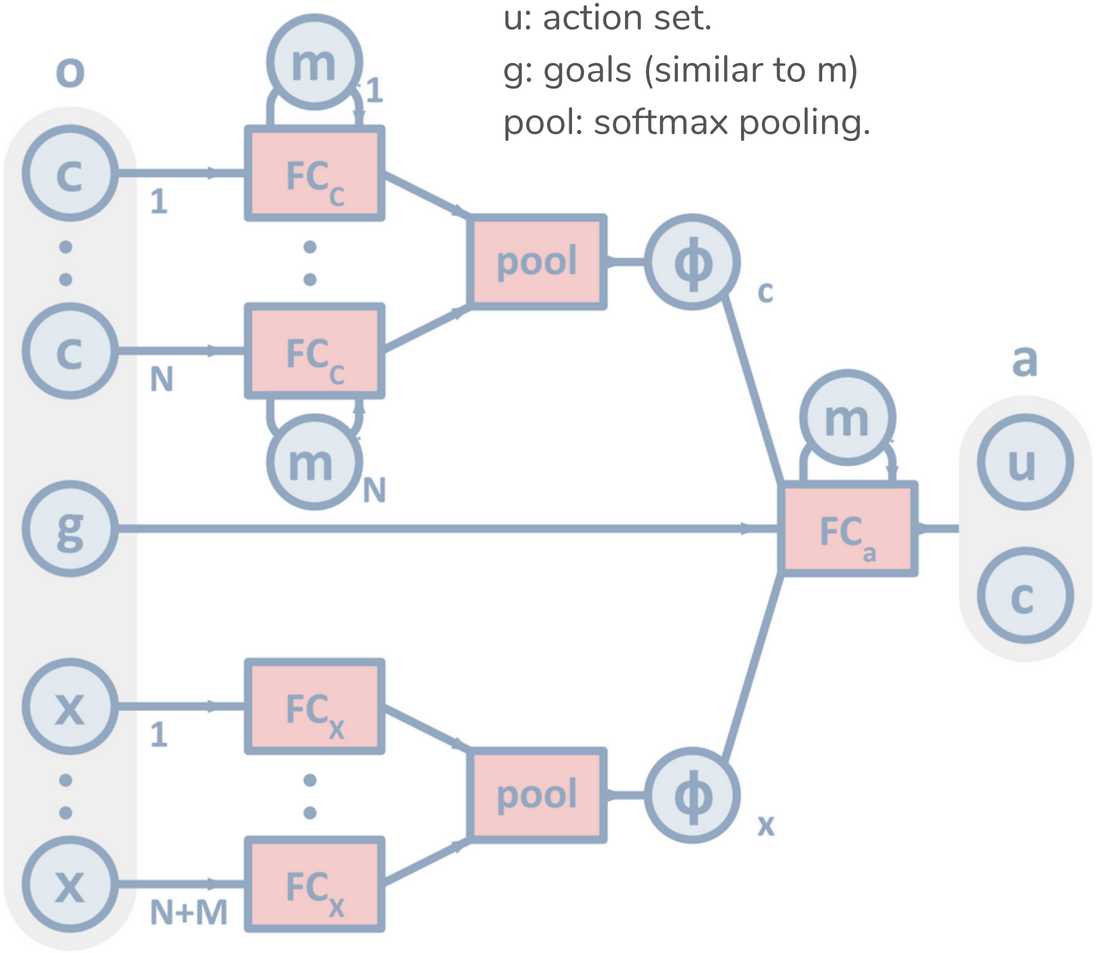
Evolution of the Environment

N: # agents
x: physical characteristics (color, shape, etc.)
t: time
c: communication/language
m: memory bank (stores individual tasks)
o: observation set
a: action set
pi: policy
dashed lines: one-to-one
bolded lines: all-to-all



System Optimization

- Q Learning: scales quadratically w/ language size :'(
- Model-free policy gradient methods have too high variance
- Approach: use an end-to-end differentiable NN to model agents and states in the entire system
- Each step, train 1024 'worlds' (batch size) and backprop results to optimize communication



Experiment Results

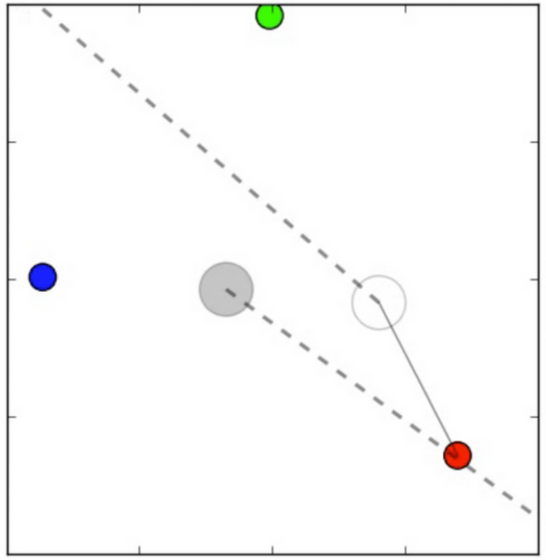
- Authors could easily interpret symbols because of the small vocabulary size
- Language contained: nouns (references to certain agents + landmarks), verbs ('GO-TO'), no redundancy
    - Self-awareness demonstrated via linguistic reference?

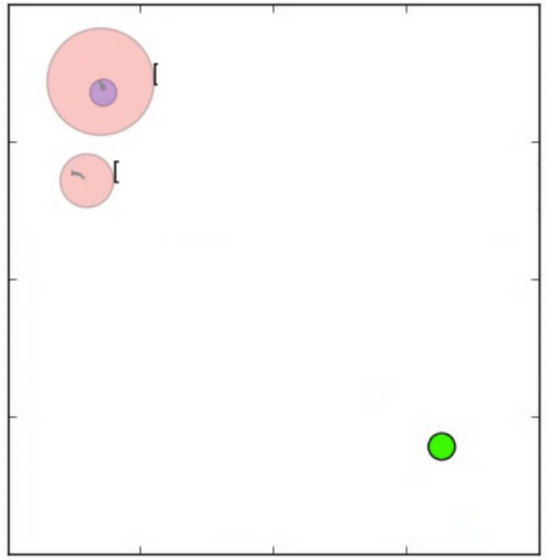
```
Red Agent: GOTO, RED, BLUE-AGENT, ...
Green Agent: ..., ..., ..., ...
Blue Agent: RED-AGENT, GREEN, LOOKAT, ...
(... = silence)
```

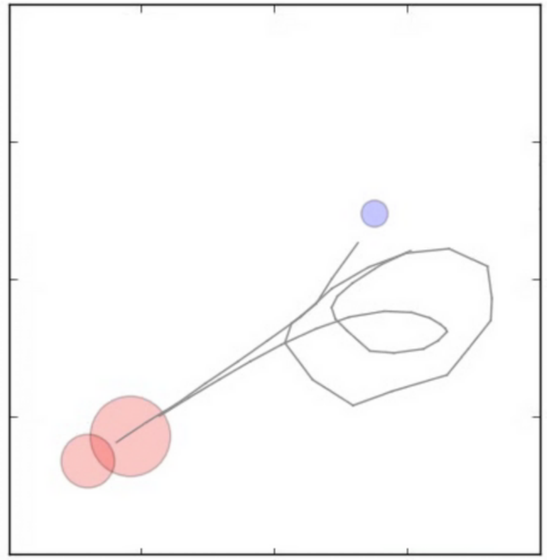
Condition	Train Reward	Test Reward
No Communication	-0.919	-0.920
Communication	-0.332	-0.392

Experiment Results

Agents developed a host of non-language behavioral communication approaches: pointing, guiding, pushing ('waggle dance' from swarm intelligence?)







Where to go next? // What can we do?

- Liberate restrictions on language (free speech is the way)
  - Allow larger vocabulary size
  - Remove penalizations for language redundancy, allow 'synonyms'
  - Remove/diminish/adjust metabolic cost
- Explore civilization-based rather than individual-based tasks/goals
- Explore independence of model definition + policy usage & optimization
  - Principles from swarm intelligence: model autonomy
- Let deep learning systems do more of the work, less human design

thank you