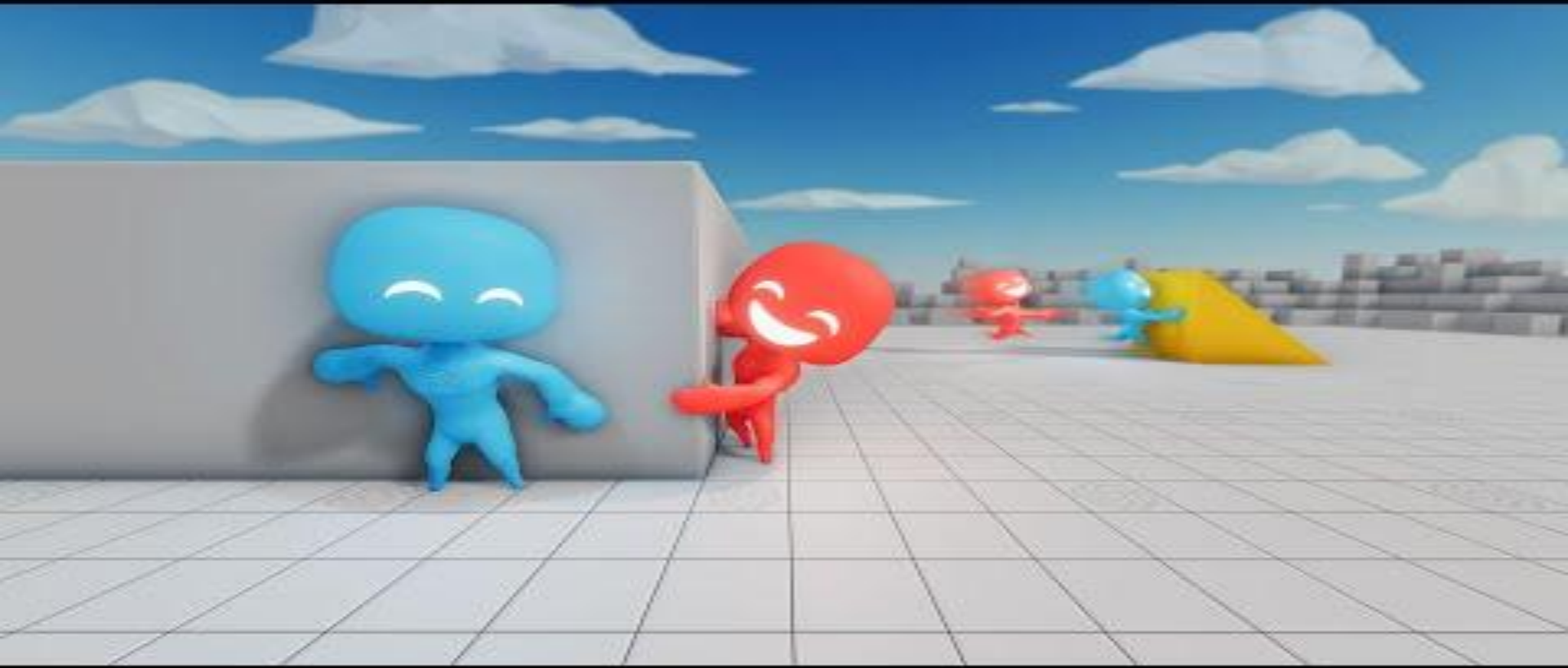
Emergent Tool Use from Multi-Agent Interaction

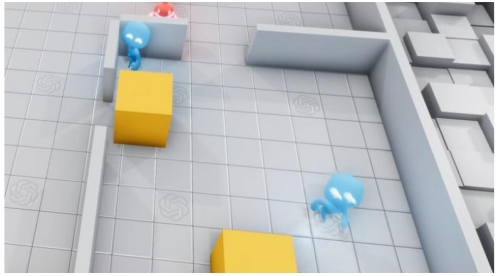
Hide & Seek

All images and charts come from OpenAI blog post and paper

Witalis Domitrz
Zuzanna Opała

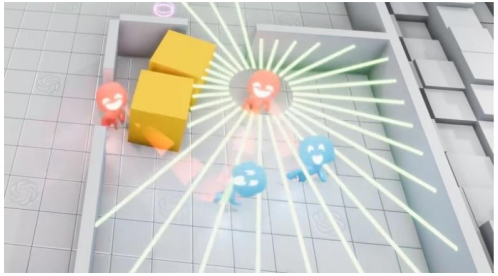


Actions and observations

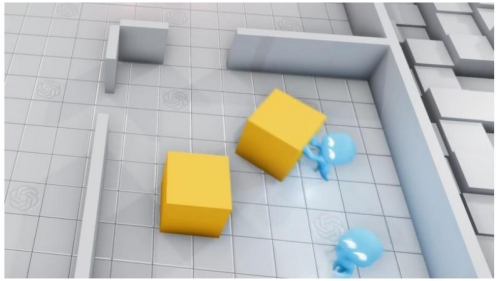


The agents can **move** by setting a force on themselves in the x and y directions as well as rotate along the z-axis.

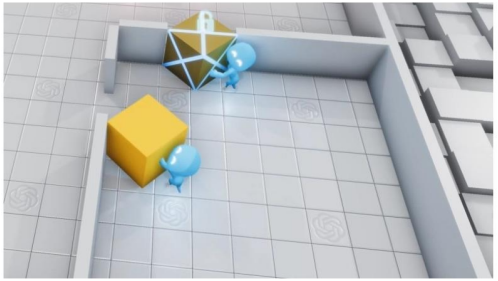
The agents can **see** objects in their line of sight and within a frontal cone.



The agents can **sense** distance to objects, walls, and other agents around them using a lidar-like sensor.

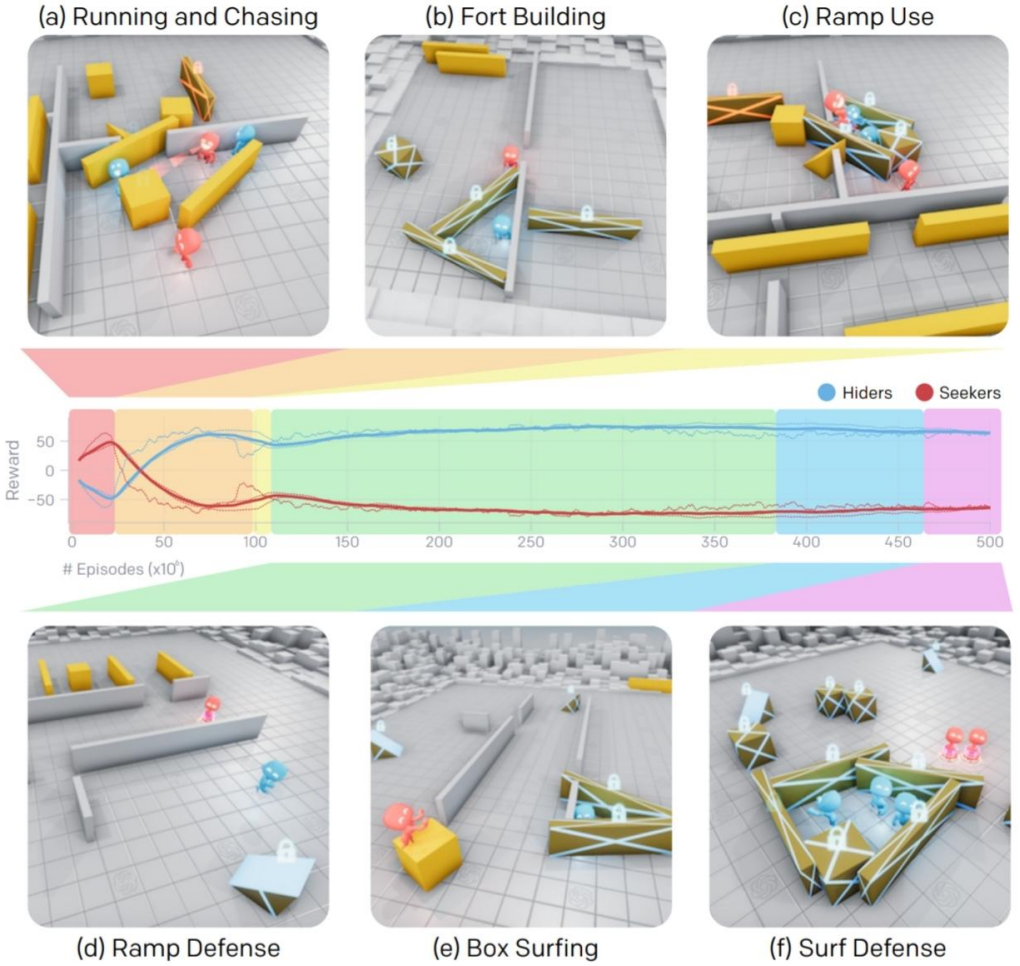


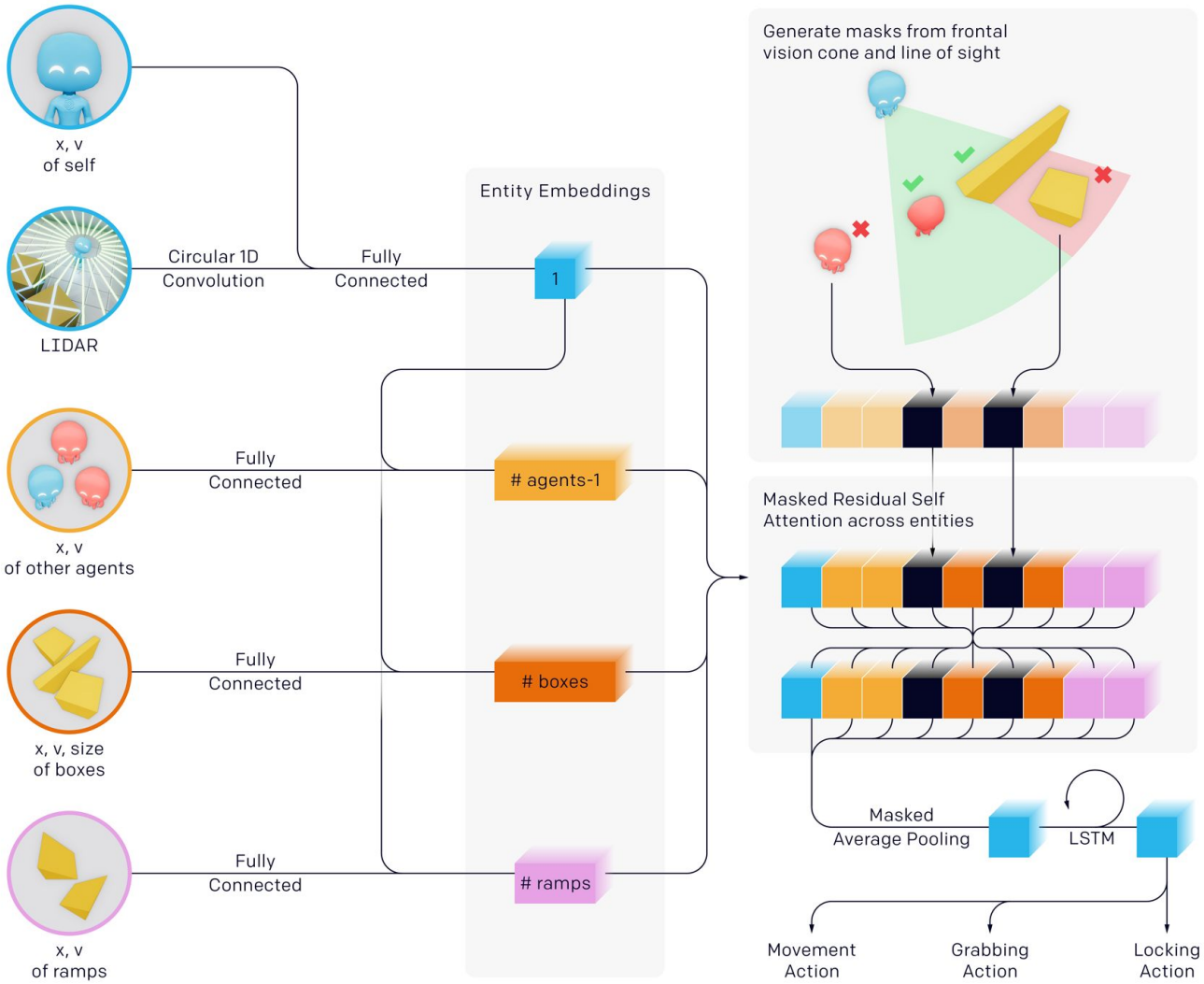
The agents can **grab and move** objects in front of them.



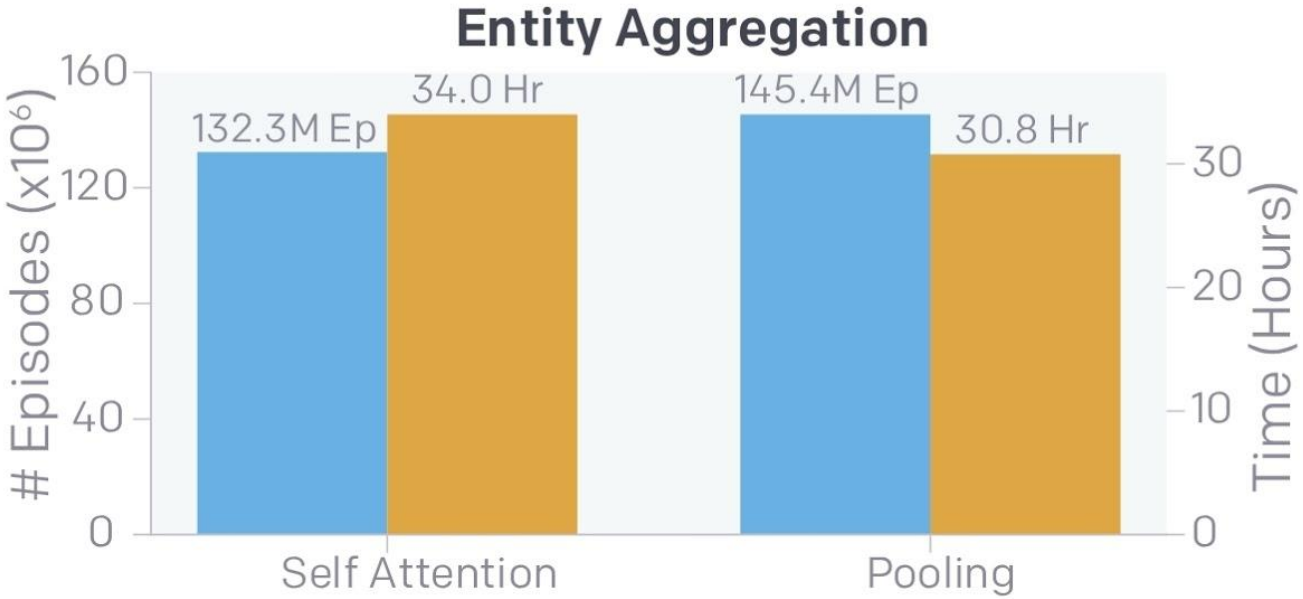
The agents can **lock** objects in place. Only the team that locked an object can unlock it.

Evolution stages

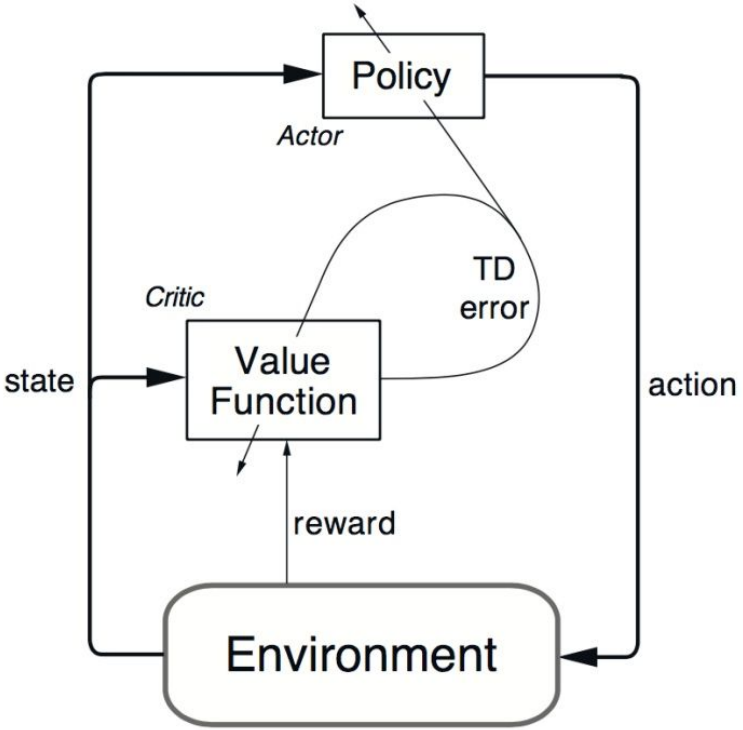


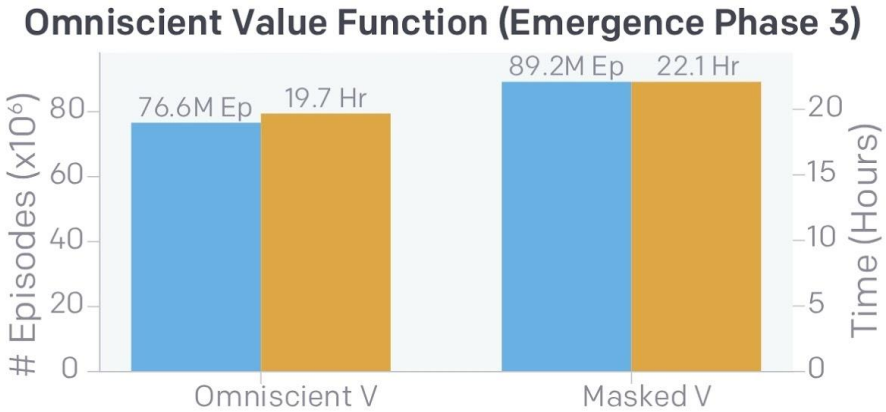

x, v
of self
LIDAR
Circular 1D
Convolution
Fully
Connected
Entity Embeddings
1
Fully
Connected
agents-1
x, v
of other agents
Fully
Connected
boxes
x, v, size
of boxes
Fully
Connected
ramps
x, v
of ramps
Generate masks from frontal
vision cone and line of sight
Masked Residual Self
Attention across entities
Masked
Average Pooling
LSTM
Movement
Action
Grabbing
Action
Locking
Action

Self Attention vs Pooling



Omniscient Value Function

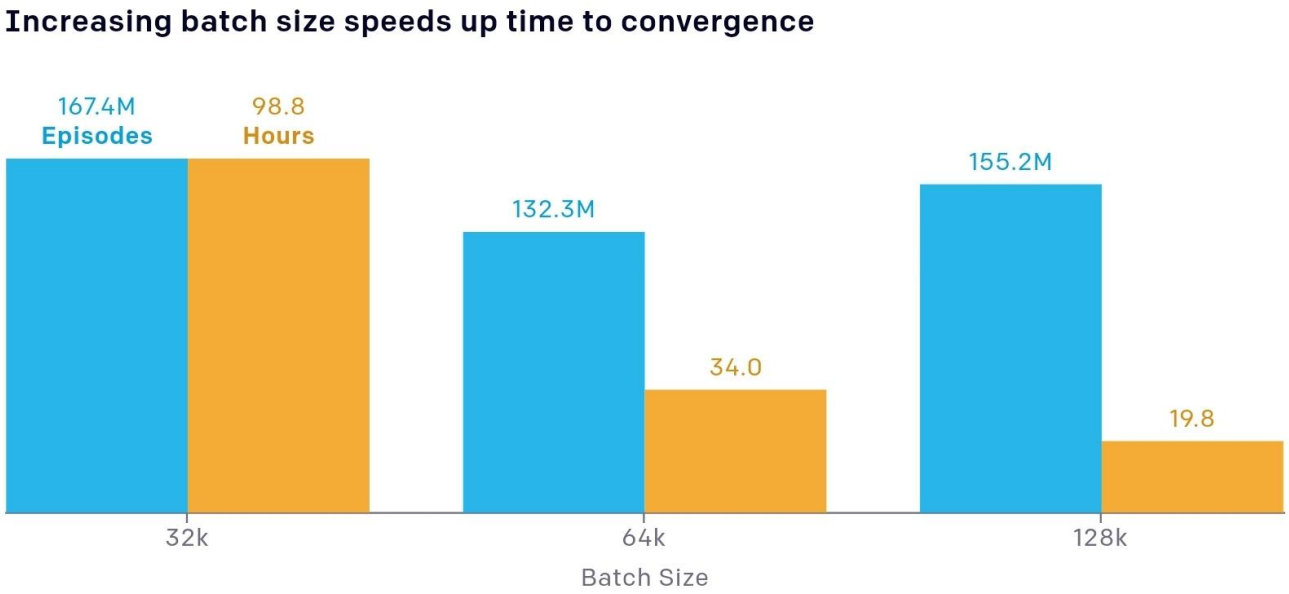




Masked value function does not reach *ramp defence* stage

source: http://incompleteideas.net/book/first/ebook/node66.html

The experiment scale



Note: We report batch size in number of contiguous chunks of transitions used in backpropagation through time, each of which contains 10 transitions, meaning that a reported batch size of 64k actually contains 640k transitions.

Batch sizes of 8k and 16k never reached ramp defence stage!

Importance of randomization

team size	# boxes	box shape	initial location	walls	emergence
1-3	3-9	random	random	random	6 stages
1	3-9	random	random	random	6 stages
1	7	fixed	random	random	6 stages
1-3	3-9	random	random	no walls	4 stages
1	3-9	random	random	no walls	2 stages*
1	7	fixed	random	no walls	2 stages*
1	7	fixed	fixed	no walls	2 stages

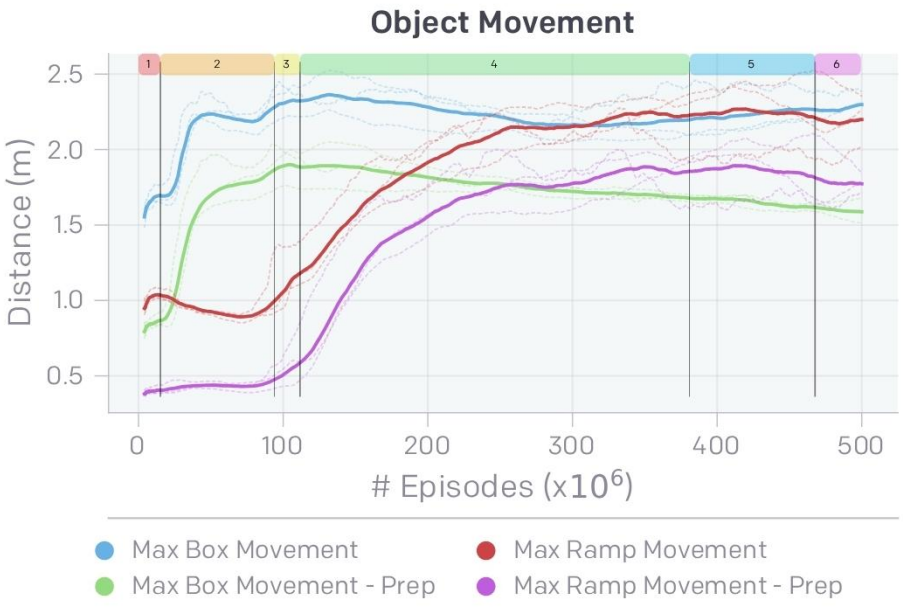
* Hiders run away and use boxes as movable shields.

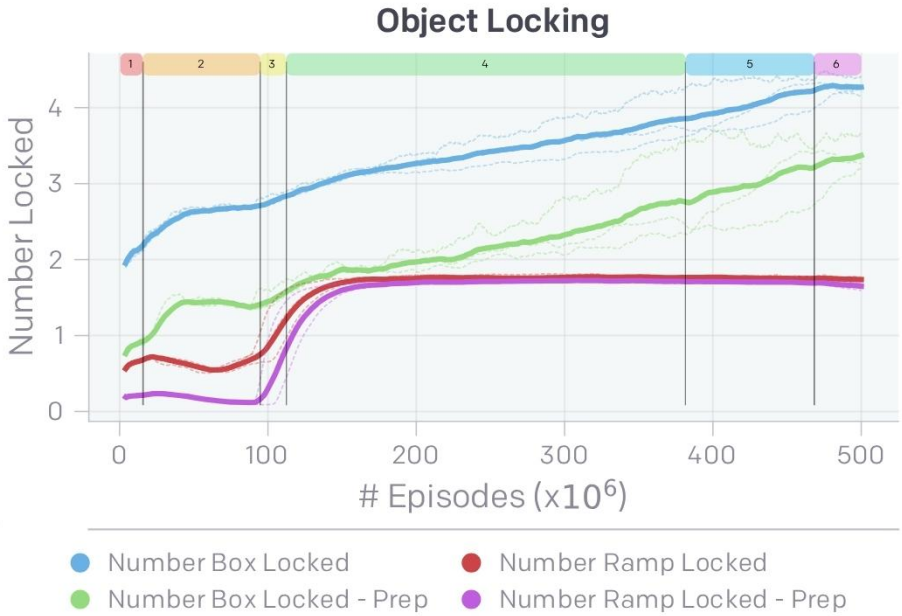
Table A.1: Number of stages of emergence for different levels of environmental randomization (batch size is held constant).

How do we know that agent learn something?

- some environment metrics
- looking at the emerged behaviours

It won't scale on more complex environments...





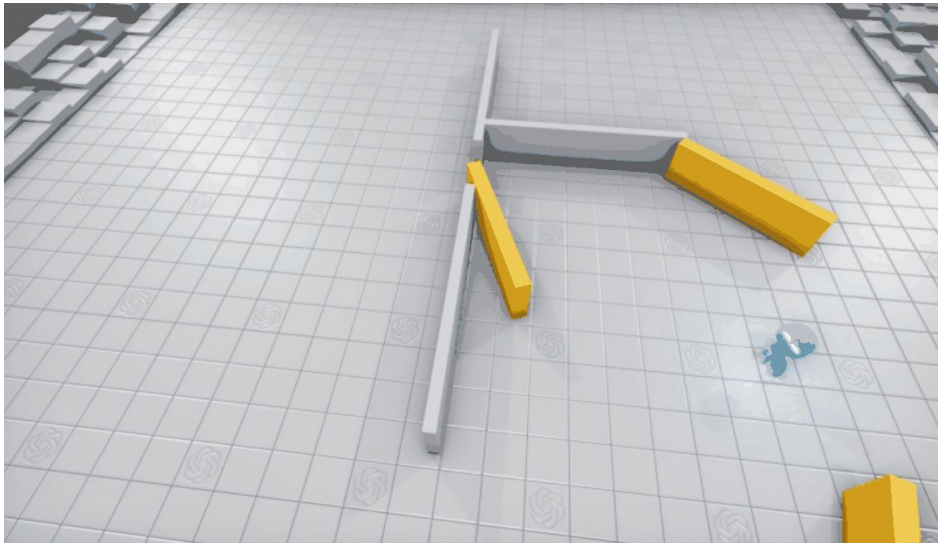
What’s the goal of RL?

to learn skills while interacting with an environment and gain the ability to generalise on unseen situation

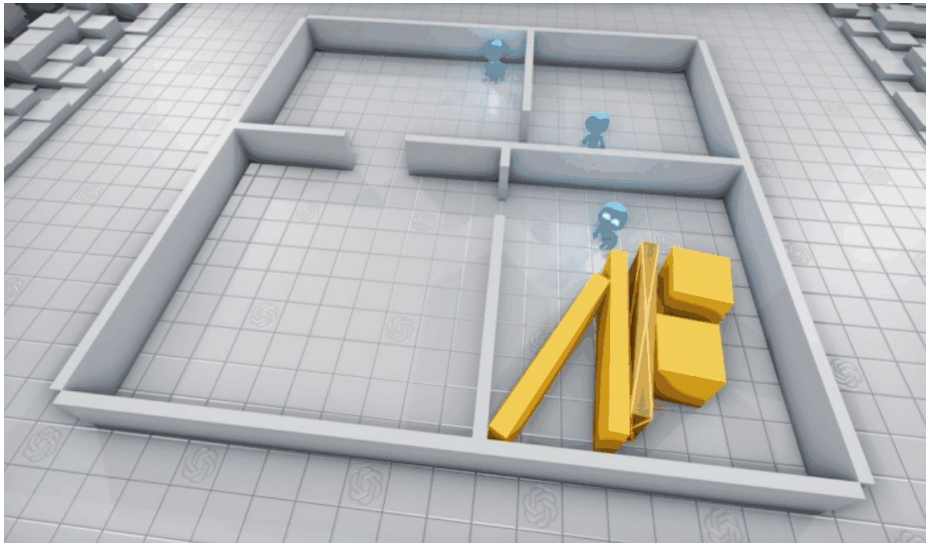
- With one agent we are limited by the constraints of the task.
- Multi-agent competition can lead to supposingly endless learning possibilities.

- This paper shows that indeed agents not only explore but learn to use tool and evolve human-like behaviours.

Alternatives - Intrinsic Motivation



Count-based exploration with selected observations



Count-based exploration with full observations

Evaluation through transfer

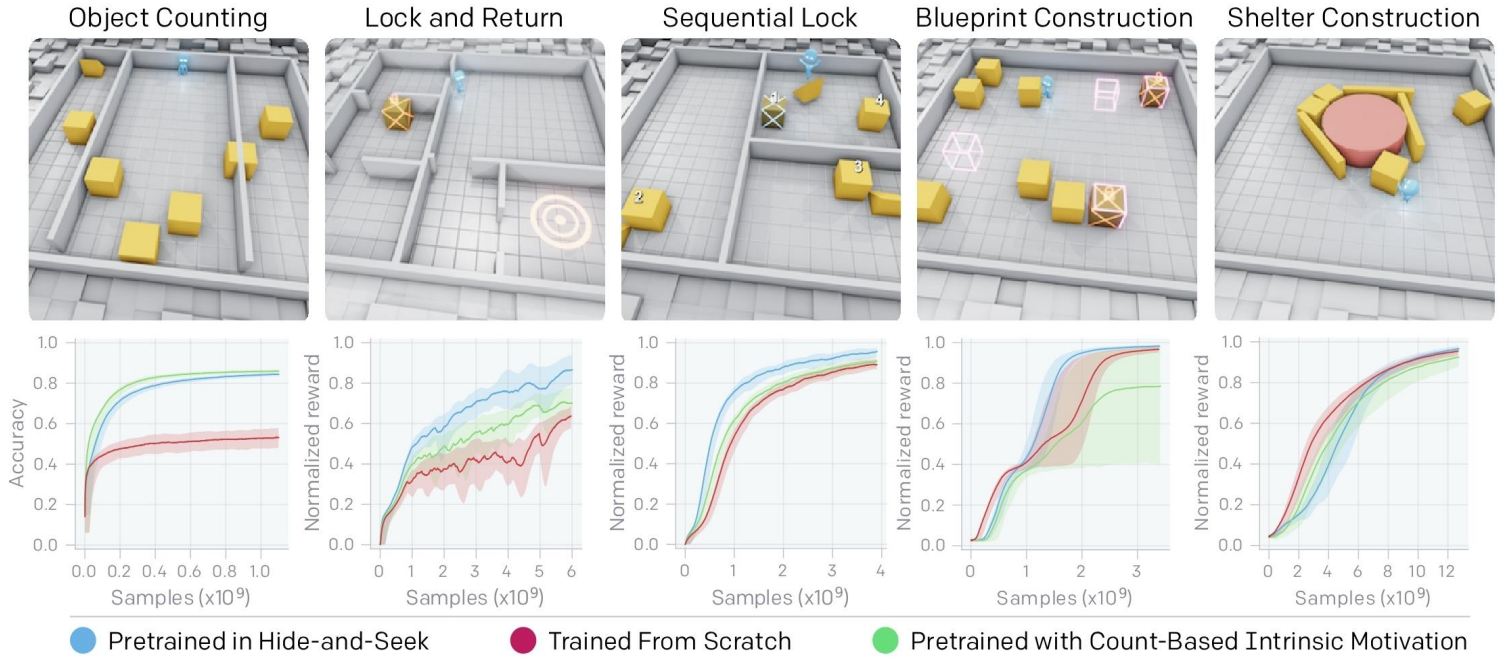
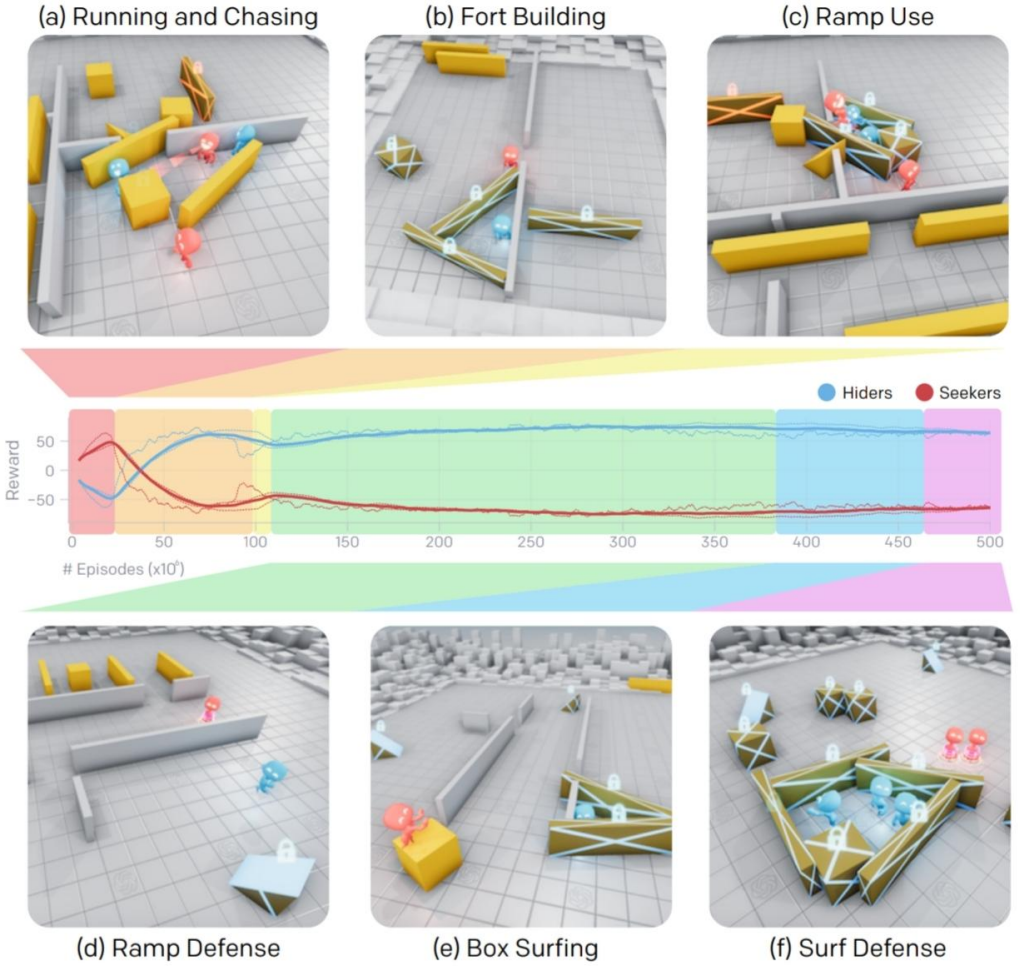
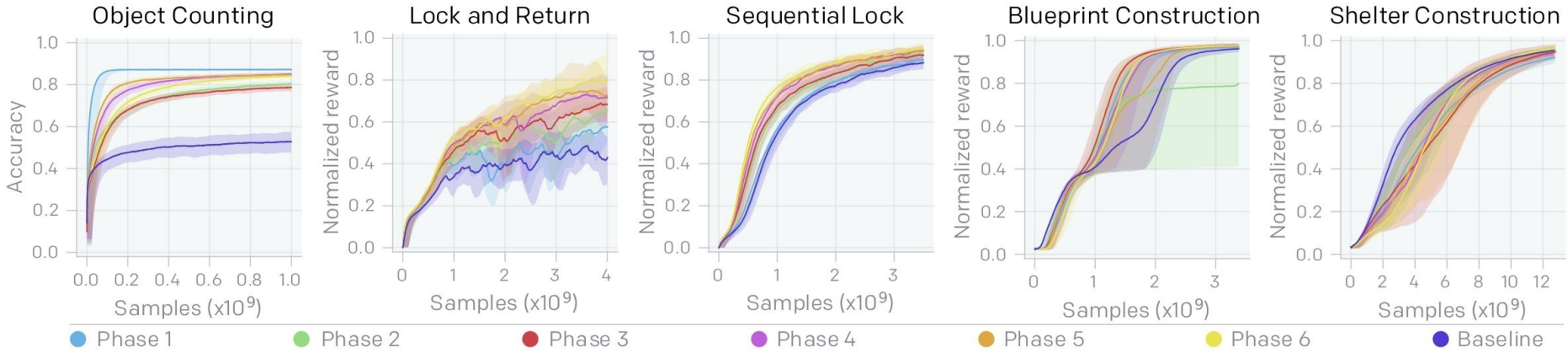


Figure 6: Fine-tuning Results. We plot the mean normalized performance and 90% confidence interval across 3 seeds smoothed with an exponential moving average, except for Blueprint Construction where we plot over 6 seeds due to higher training variance.

Evolution stages



Evaluation in different phrases



Zero-shot generalization

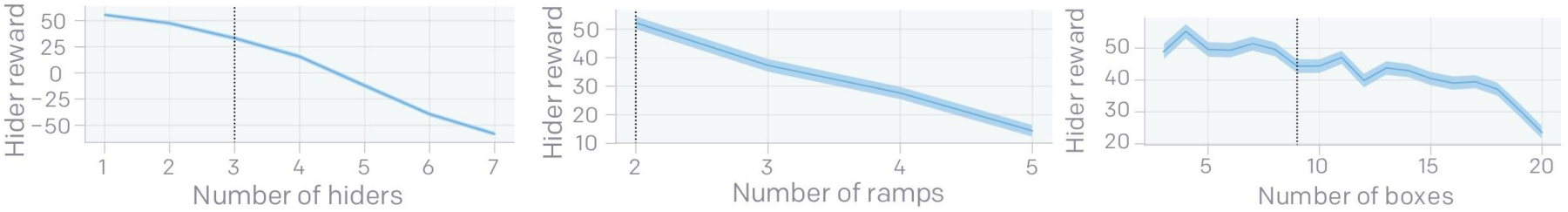


Figure A.8: Zero-shot generalization to a larger number of hiders (left), ramps (center) and boxes (right). The dotted line denotes the boundary of the training distribution (1-3 hiders, 2 ramps, 1-9 boxes). Error bars denote standard error of the mean.

Summary

- Based on the simple game rules agents learned fairly complicated behaviours such as building a shelter.
- Tool use emerged through it wasn’t directly incentivised .
- It’s proof of concept for unsupervised evolution through the competition.
- The agents learned to cooperate.
- Their behaviours were human-like and easy to interpret.

Resources and other interesting papers

- Emergent Tool Use From Multi-Agent Autocurricula by Bowen Baker and Ingmar Kanitscheider and Todor Markov and Yi Wu and Glenn Powell and Bob McGrew and Igor Mordatch, 2019
- OpenAi blog post Emergent Tool Use From Multi-Agent Autocurricula
- Autocurricula and the Emergence of Innovation from Social Interaction: A Manifesto for Multi-Agent Intelligence Research by Joel Z. Leibo and Edward Hughes and Marc Lanctot and Thore Graepel, 2019
- #Exploration: A Study of Count-Based Exploration for Deep Reinforcement Learning by Haoran Tang and Rein Houthooft and Davis Foote and Adam Stooke and Xi Chen and Yan Duan and John Schulman and Filip De Turck and Pieter Abbeel, 2016