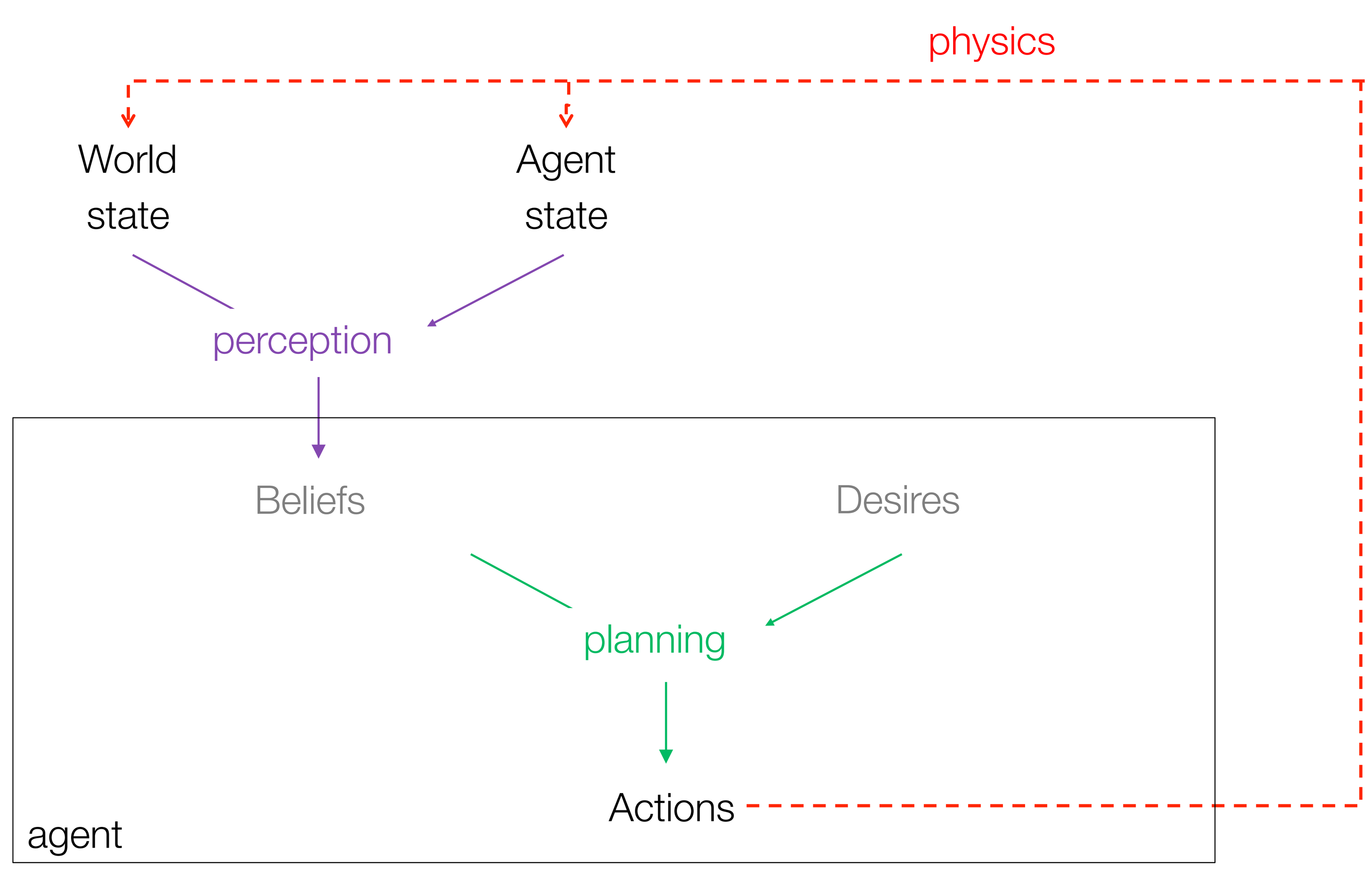
A Compositional Object-Based Approach to Learning Physical Dynamics

Michael Chang[1,2,4], Tomer Ullman[1,3], Antonio Torralba[1,4], Joshua B. Tenenbaum[1,3]

1 MIT
2 SuperUROP
3 brain+cognitive sciences
4 MIT EECS Electrical Engineering and Computer Science

Project Webpage: http://mbchang.github.io/npe Code: http://github.com/mbchang/dynamics

Motivation



A sense of intuitive physics endowed in an agent is a prior on the environment that the agent uses to accelerate higher-level learning.

Overview

Where can this physics prior come from?

Symbolic Physics Engines

- Expressive
- Knowledge encoded in structure
- Difficult to adapt to scenarios outside description language

Neural Networks

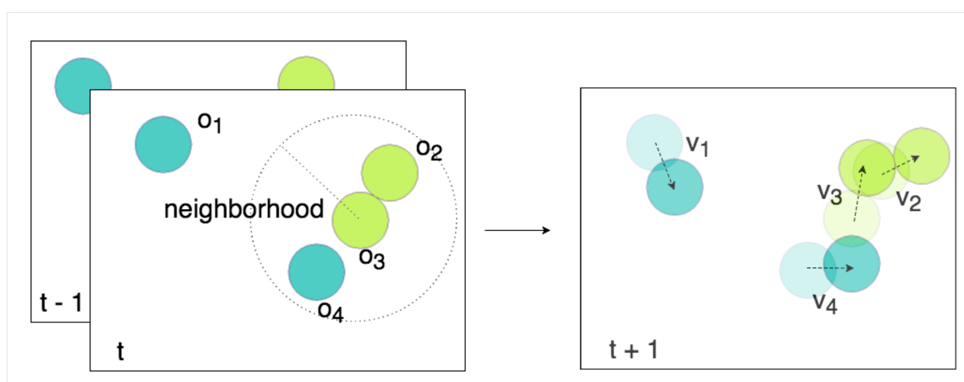
- Adaptive
- Knowledge emerges through training
- Requires retraining for new scenarios

Neural Physics Engine (NPE)

- *Generic* notions of objects and their interactions
- Trained to model *specific* object properties and dynamics of different worlds
- Transfers knowledge across different object counts and scene configurations

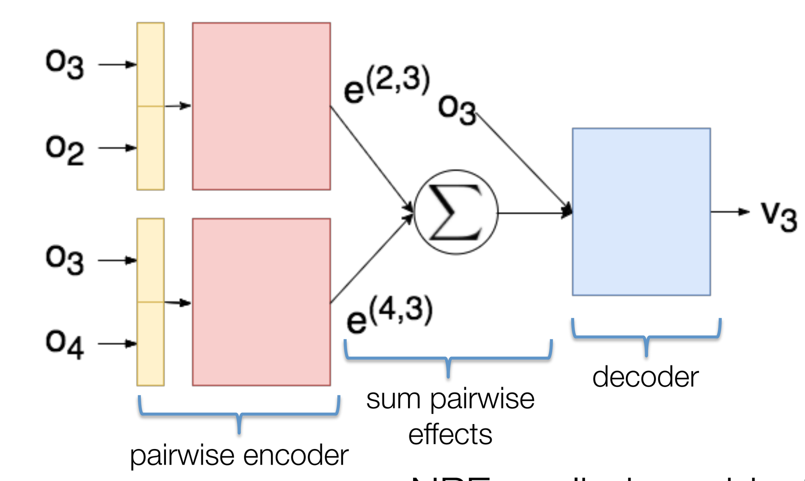
Neural Physics Engine (NPE)

Scenario



Predict the velocity of each object in turn, given the pairwise interactions with its neighborhood context objects.

Model Architecture



NPE applied on object 3

Baselines

Pairwise layer

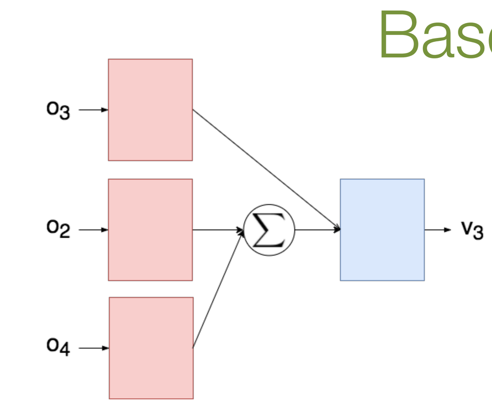
Multi-layer decoder

Multi-layer encoder

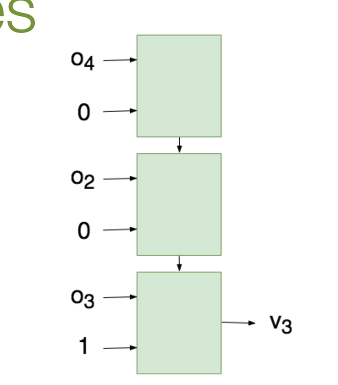
Multi-layer LSTM

No-Pairwise (NP) Baseline: no pairwise factorization

LSTM Baseline: no composition of independent effects



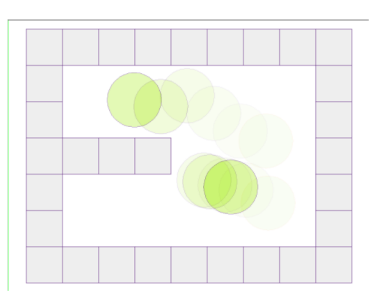
NP applied on object 3

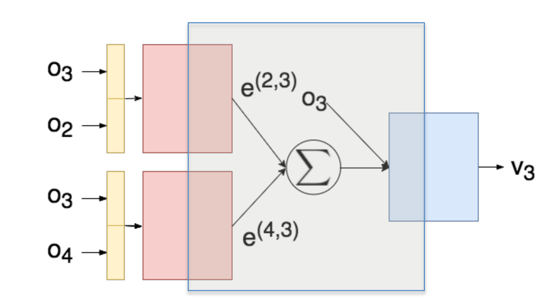


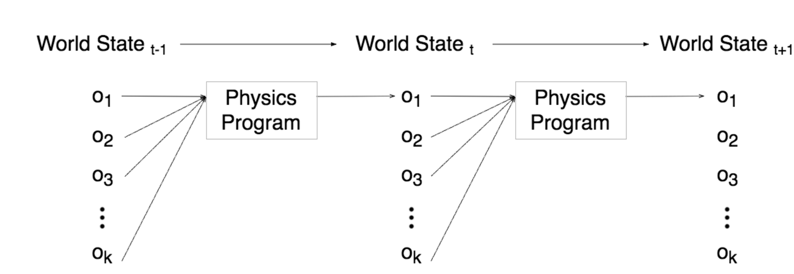
LSTM applied on object 3

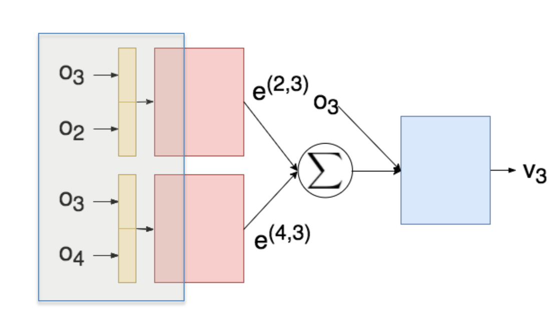
Composition and Factorization on the Levels of Both the Scene and Model

	Scene	Model
Composition	Compose larger objects from smaller building blocks	Object state prediction is a function composition between itself and other objects
Factorization	Object-based representations	Pairwise interactions

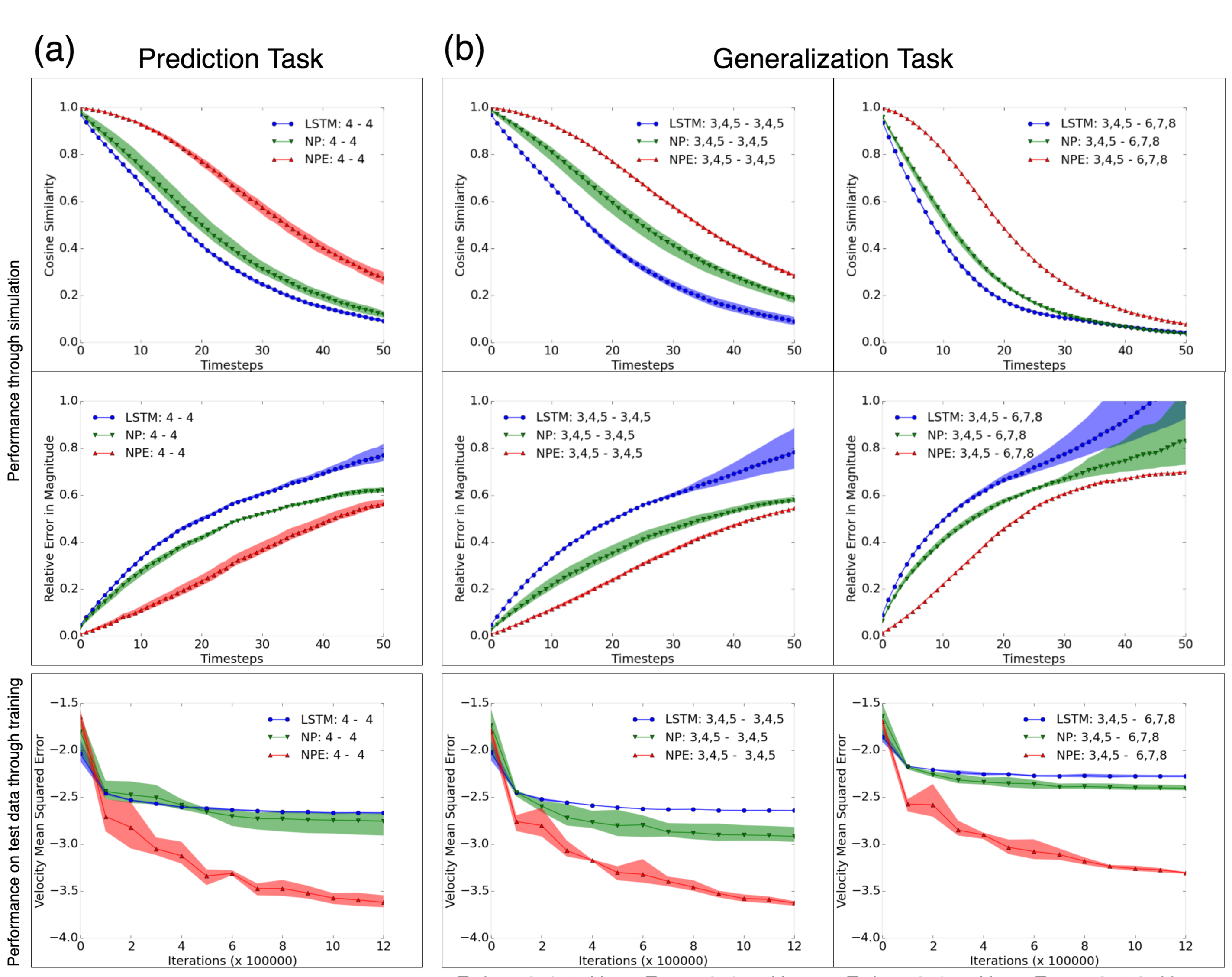


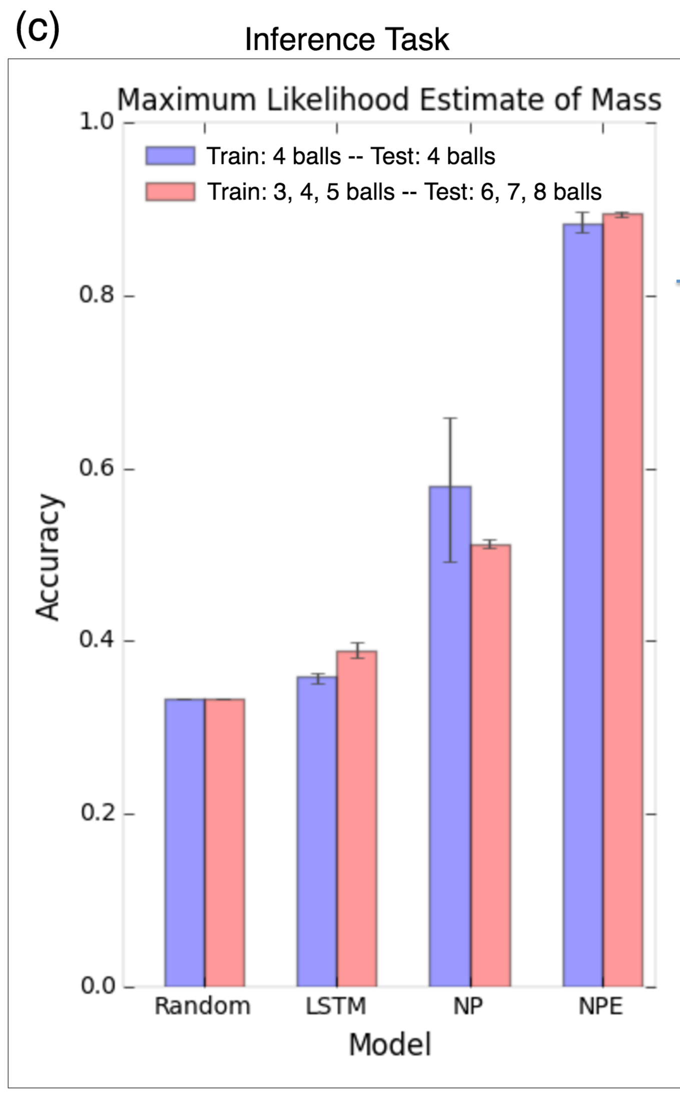






Generalization: different numbers of objects





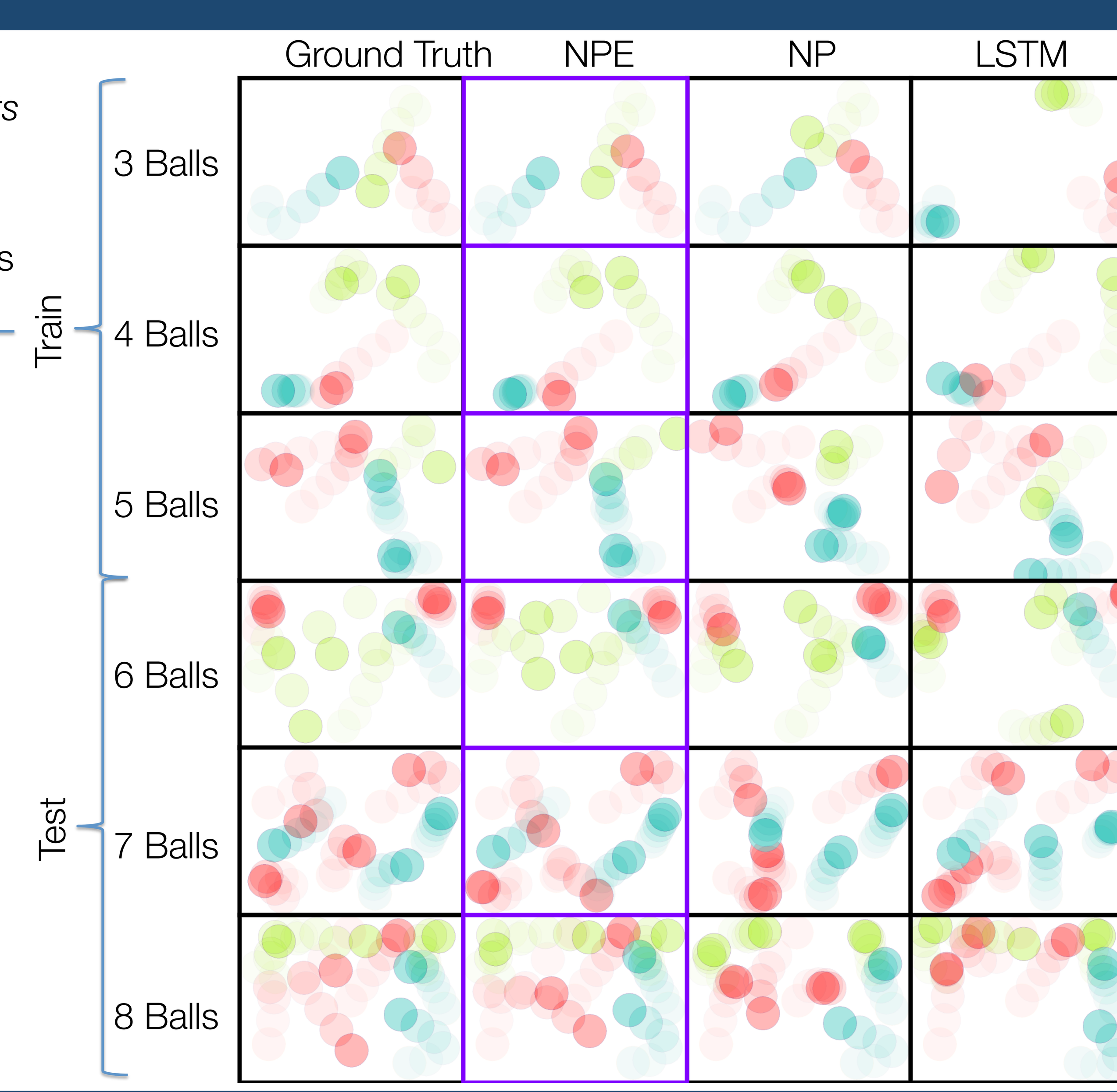
3, 4, 5: *worlds with fewer objects*

6, 7, 8: worlds with *more* objects

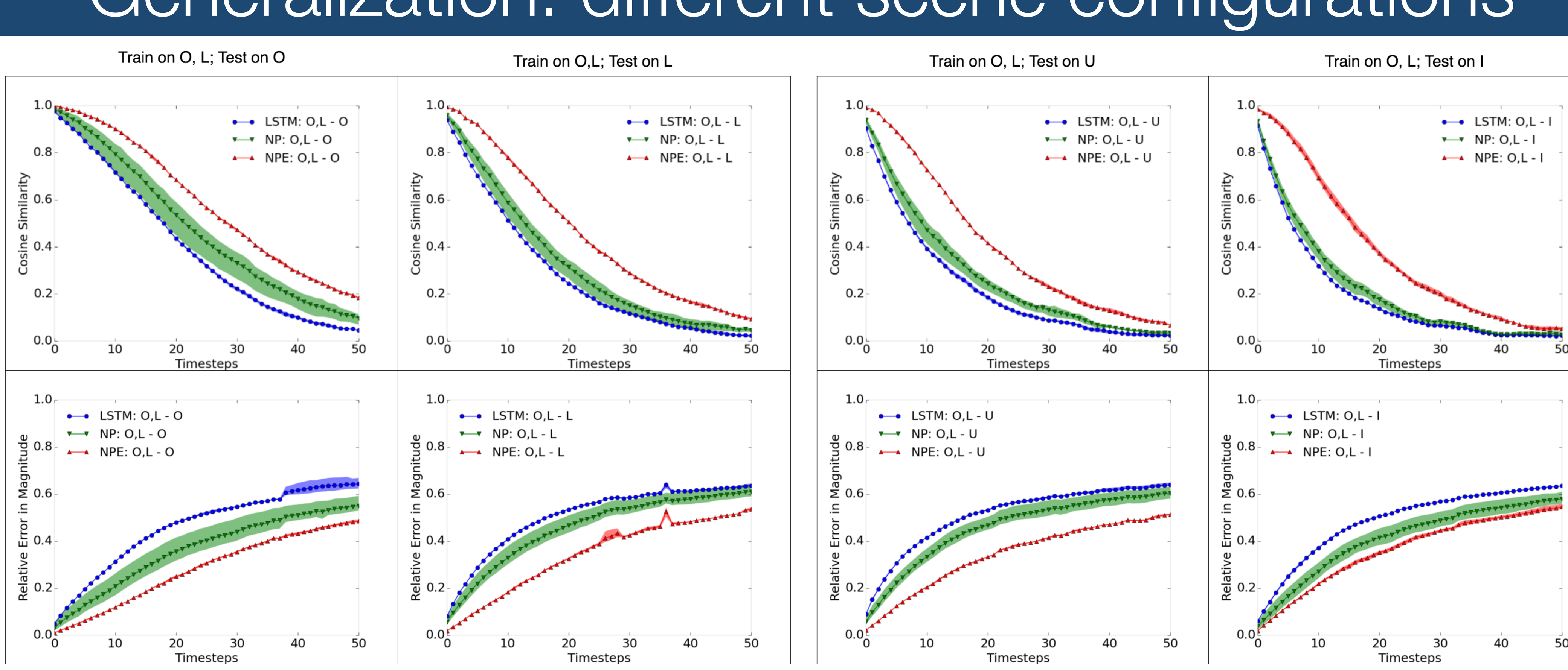
a) *Prediction Task*: Cosine similarity (top) and relative magnitude (bottom) of prediction vs ground truth over 50 steps of simulation. Log MSE on velocity (bottom) over the course of training

b) *Generalization Task*: Same metrics as prediction, but train on 3, 4, 5; test on 6, 7, 8.

c) *Inference Task*: NPE gets about 90% accuracy in prediction and generalization setting.



Generalization: different scene configurations



"O", "L": worlds without internal obstacles

"U", "I": worlds with internal obstacles

NPE does not overlap with internal obstacles, while the NP and LSTM do. This shows the NPE is invariant to position and scene configuration, while NP and LSTM memorize the training wall configuration.