Northeastern University
Khoury College of
Computer Sciences

Toward Understanding Compositional Generalization in Object-Oriented World Modeling

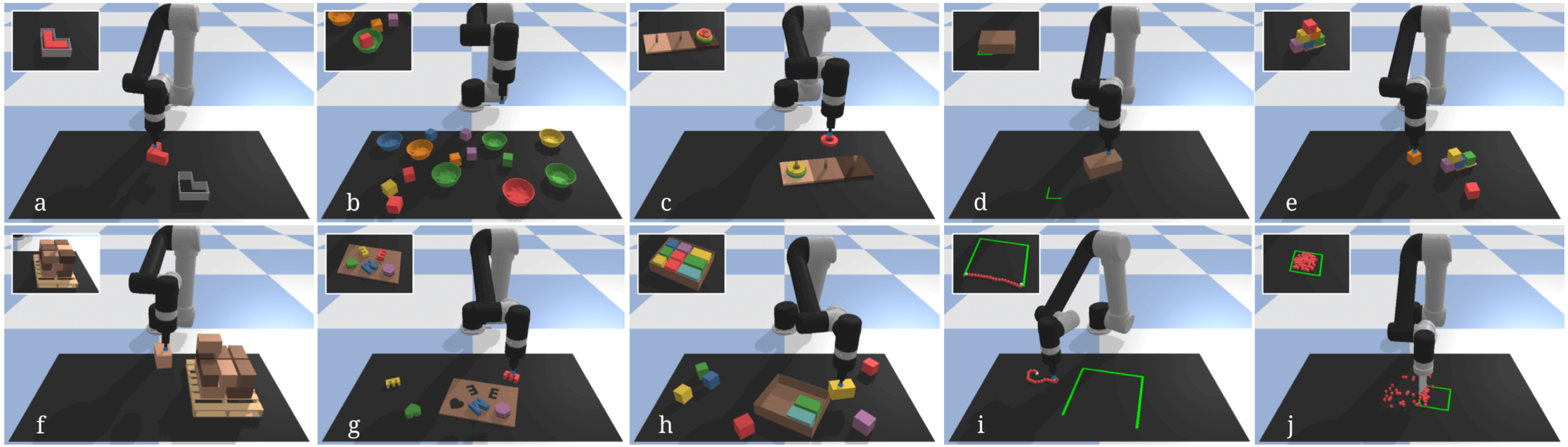
***Linfeng Zhao**, Lingzhi Kong, Robin Walters, Lawson Wong*
Khoury College of Computer Sciences, Northeastern University

International Conference on Machine Learning, 2022

Motivation

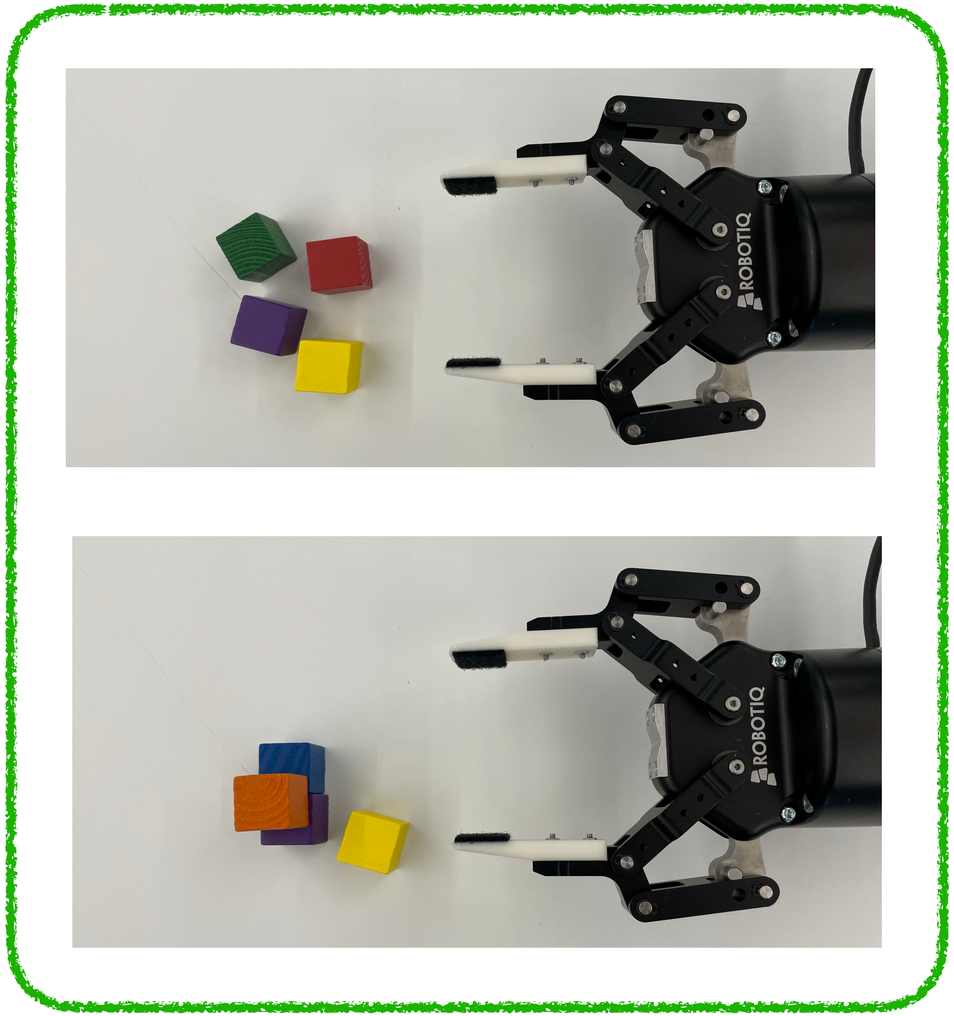
Motivation: Robotic Manipulation

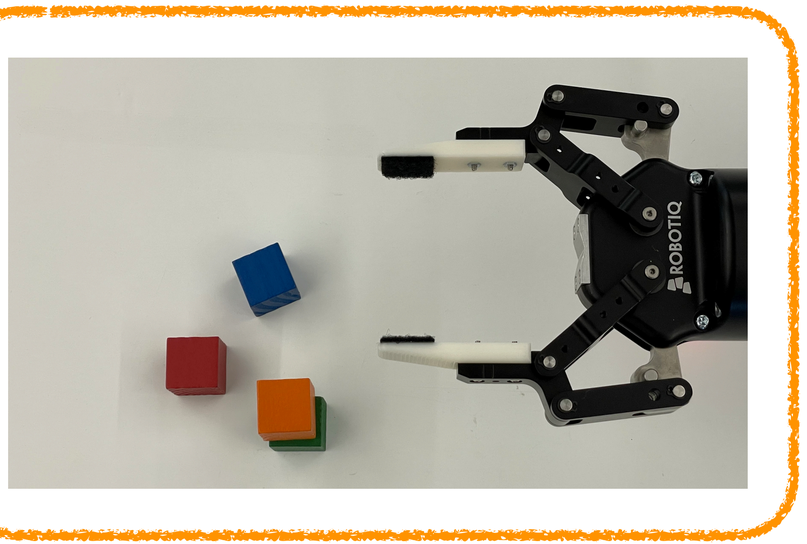
Example: Robotic Manipulation



Credit: Ravens

Motivation: Robotic Manipulation





Train

Generalization

Research Questions

Compositional Generalization (CG)

- How to measure compositional generalization (in world modeling)?
- Can we guarantee when compositional generalization is achieved?
- If possible, how can its implementation be efficient?

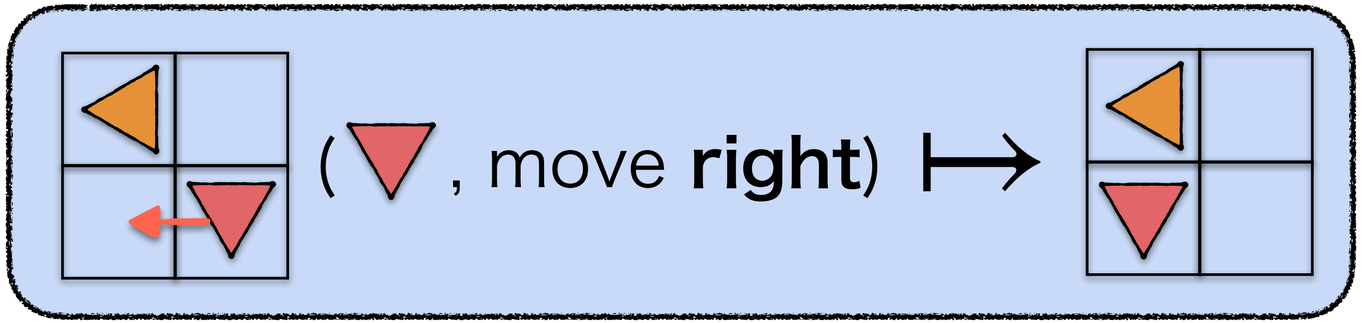
Outline

- Motivation
- Setup
- Defining Compositional Generalization
- Implementing Compositional Generalization
- Solving Binding Issue for Compositional Generalization
- Results

Setup

World Modeling

Learn a transition model $T : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{S}$



s_1 a_1 s_2

Actions are factorized by objects and relative to orientation

Object Library

Object Library $\mathbb{L}$

"Vocabulary"

$N \triangleq |\mathbb{L}| = 4$

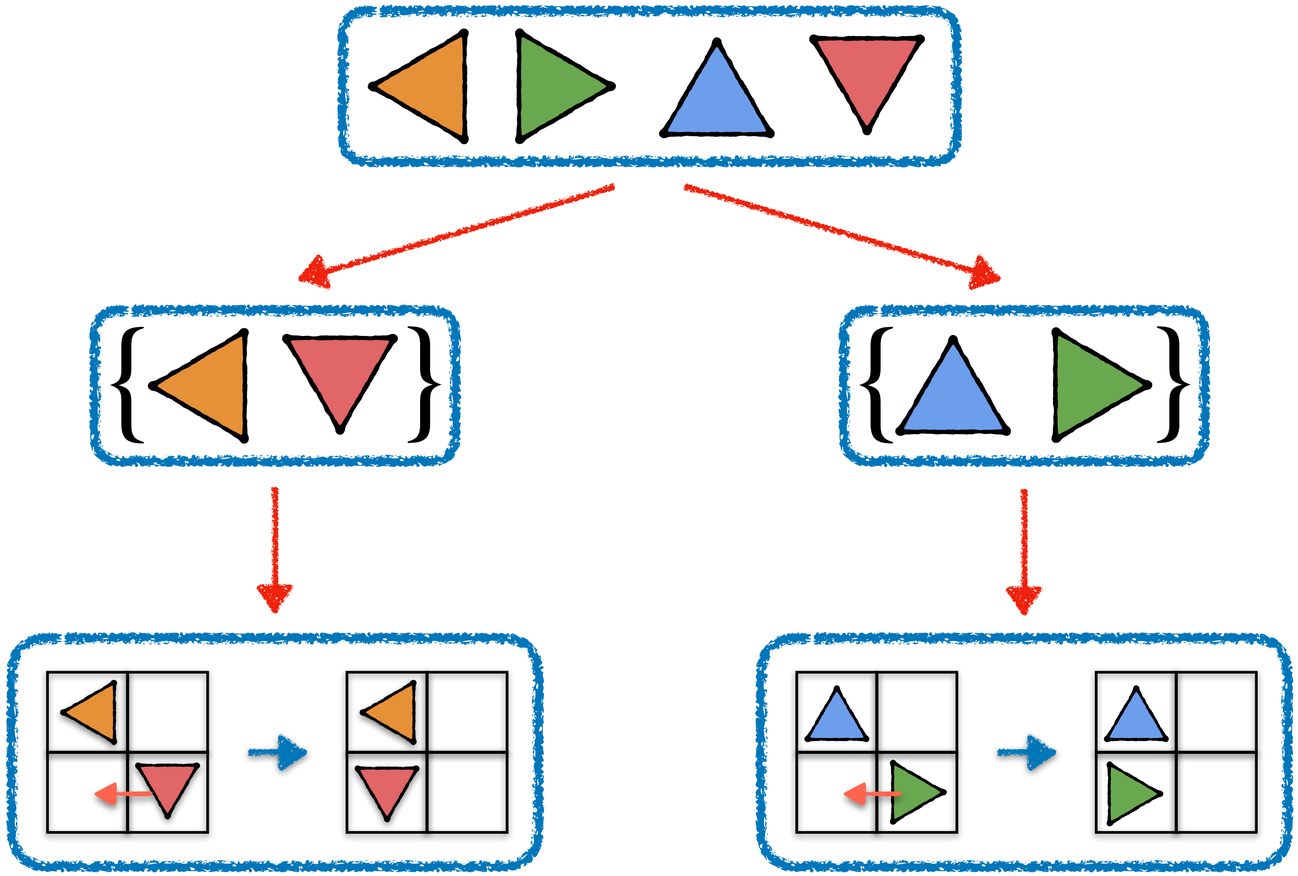
Scene Sets $\mathbb{O}_i \subset \mathbb{L}$

(Ordered) "Sentences"

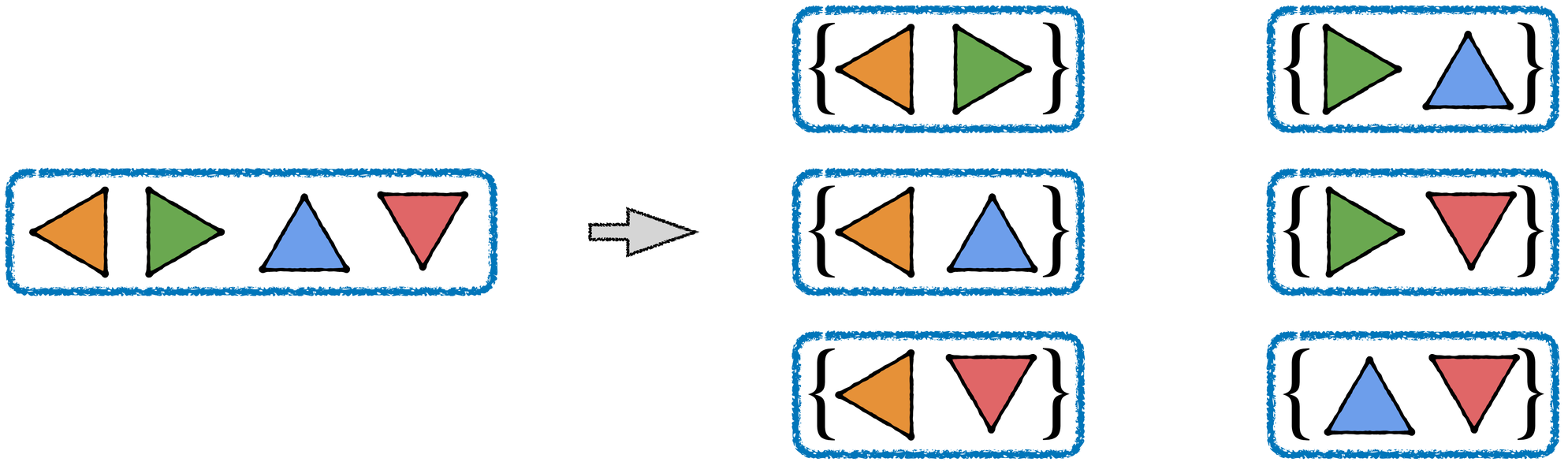
$K \triangleq |\mathbb{O}| = 2$

Scene MDPs $\mathscr{M}_{\mathbb{O}_i}$

Generated by $\mathbb{O}$



Object Library

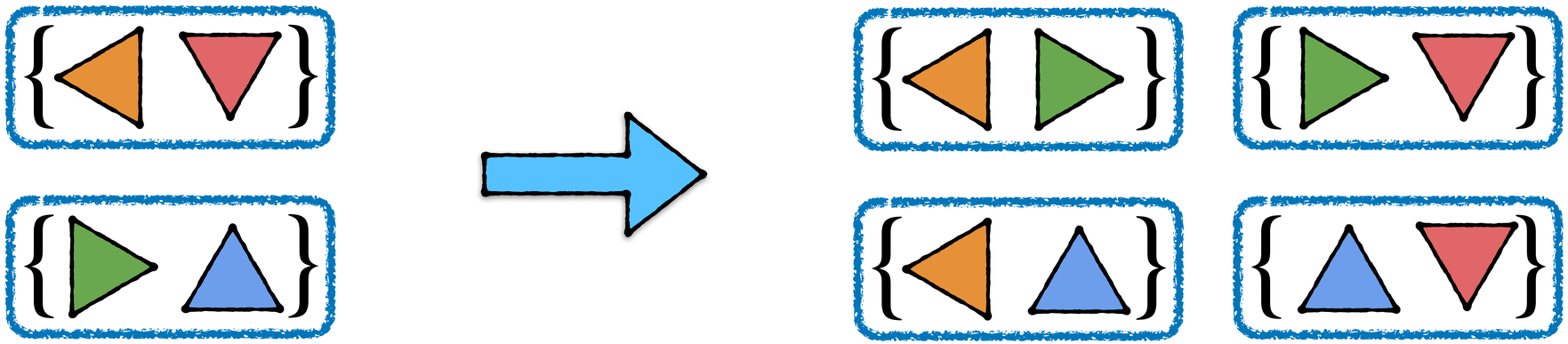


All $\binom{N}{K}$ combinations

Object Library

Training

Generalization



Learn a transition model $T : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{S}$

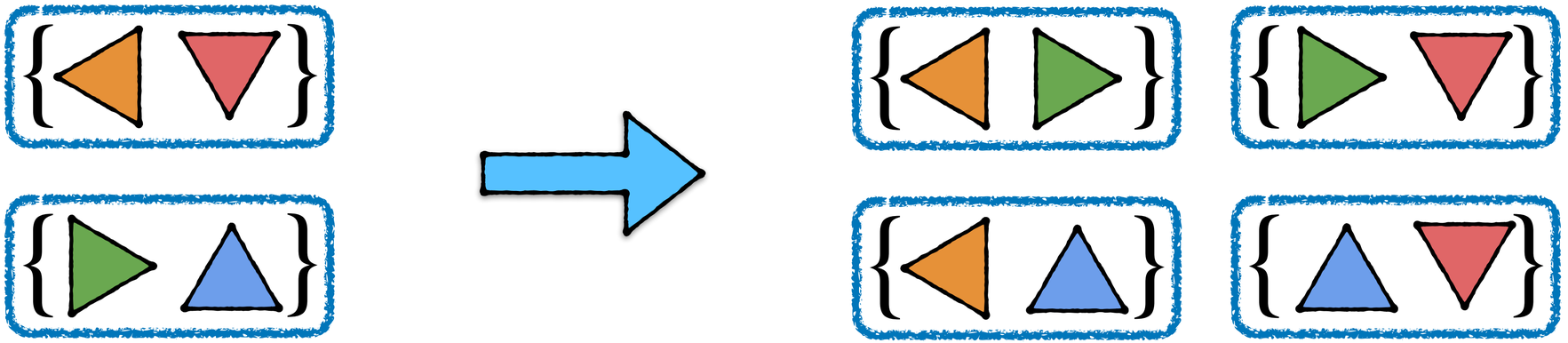
Outline

- Motivation
- Setup
- Defining Compositional Generalization
- Implementing Compositional Generalization
- Solving Binding Issue for Compositional Generalization
- Results

Defining Compositional Generalization

Motivation

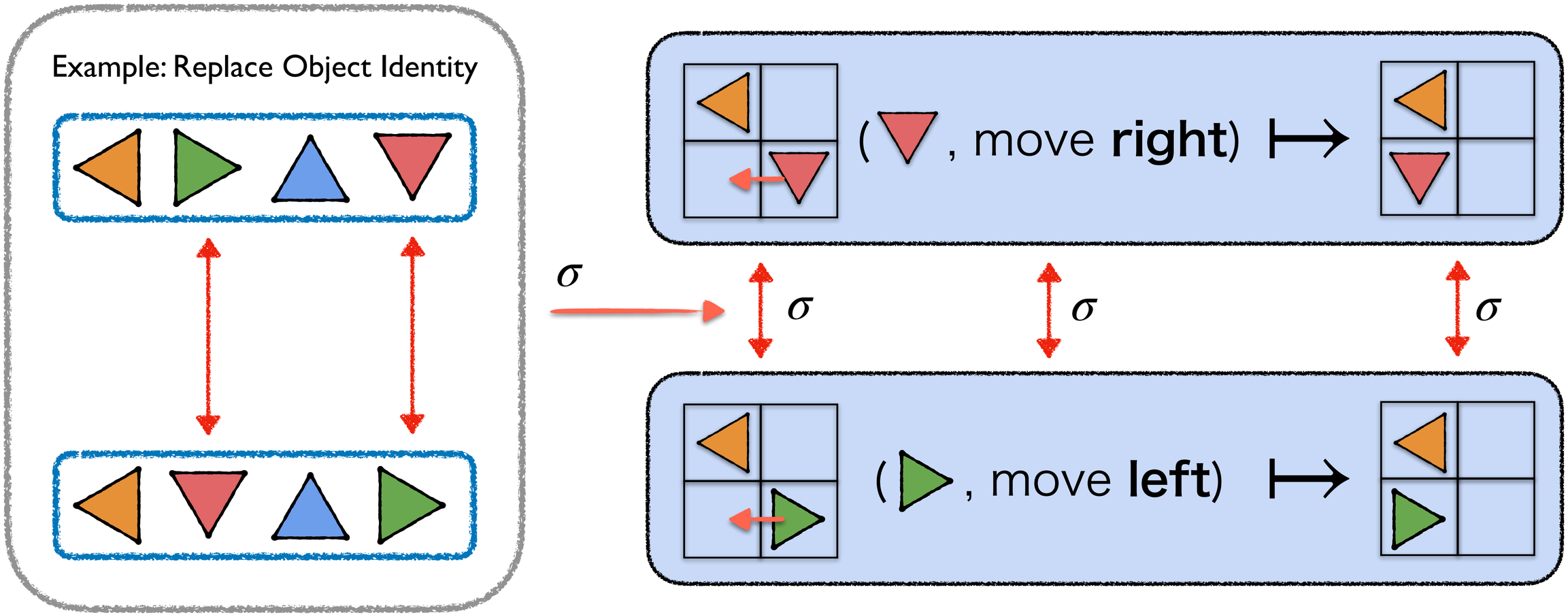
How to quantitively define and measure compositional generalization?



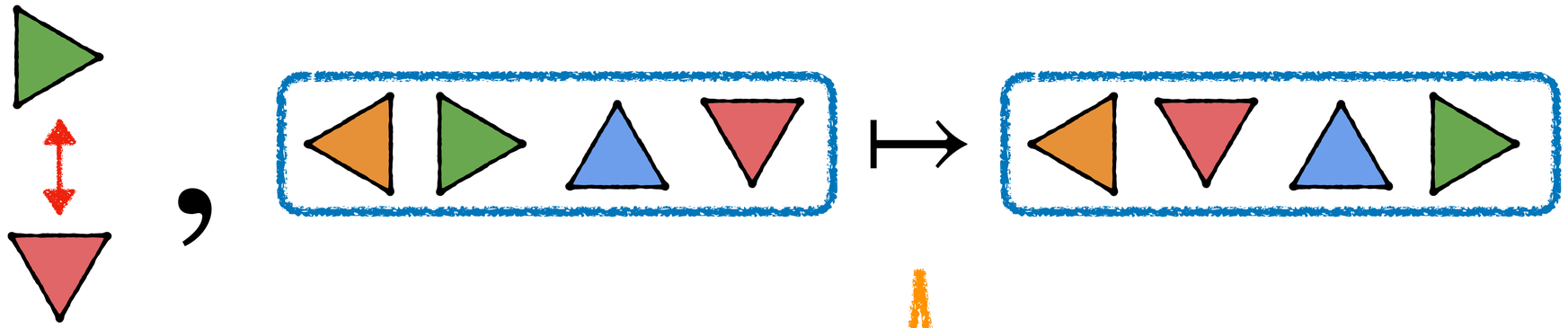
Training

Generalization

Object-replacement: Another View



Object-replacement Operation

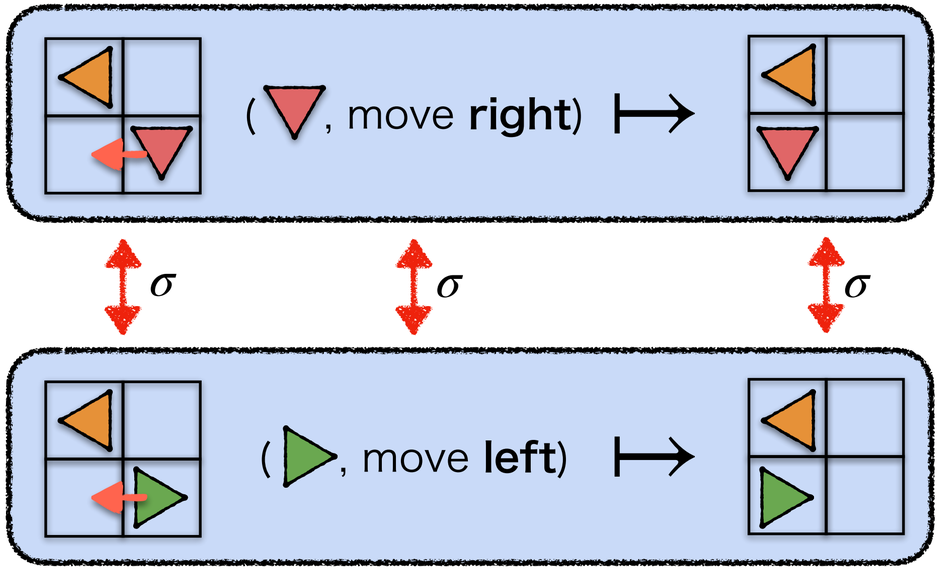


Formally define an operation:

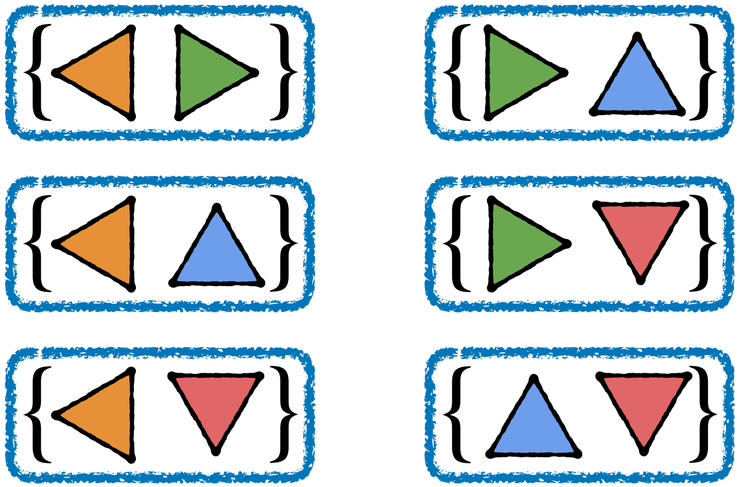
$$p_{\mathbb{L}} : \Sigma_N \times \mathbb{L} \rightarrow \mathbb{L}$$

Size-N Permutation Group

Measure CG with Equivariance Error



Expectation
over
All combinations



Cast **Compositional Generalization**
as **(Permutation) Equivariance Error**
in **Transition Prediction**

How well can the model predict
on *any compositionally different scene*?

$$\mathrm{EE}(T_{\mathbb{L}}) \triangleq \mathbb{E}\left[\left|\hat{T}_{\mathbb{L}}(s' \mid s, a) - \hat{T}_{\mathbb{L}}(\sigma \,.\, s' \mid \sigma \,.\, s, \sigma \,.\, a)\right|\right]$$

$$\sigma \in \Sigma_N, (s, a, s') \in \mathcal{S}_{\mathbb{L}} \times \mathcal{A}_{\mathbb{L}} \times \mathcal{S}_{\mathbb{L}}$$

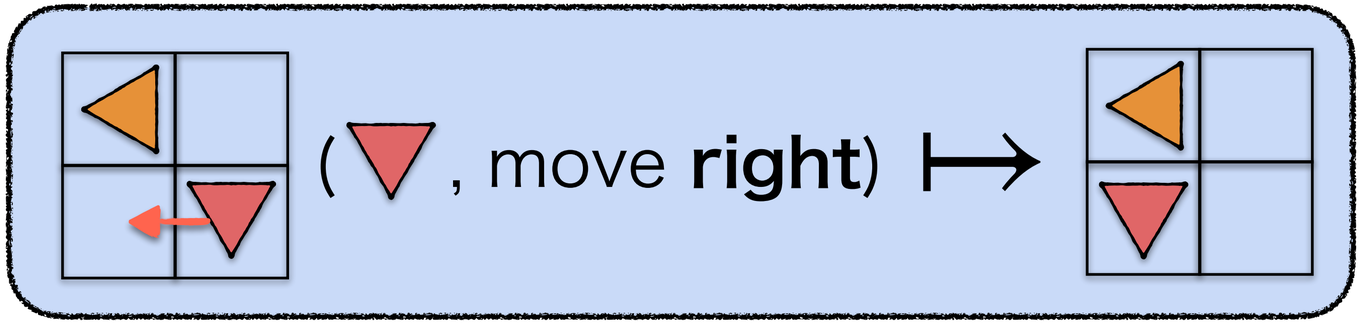
Outline

- Motivation
- Setup
- Defining Compositional Generalization
- Implementing Compositional Generalization
- Solving Binding Issue for Compositional Generalization
- Results

Implementing Compositional Generalization

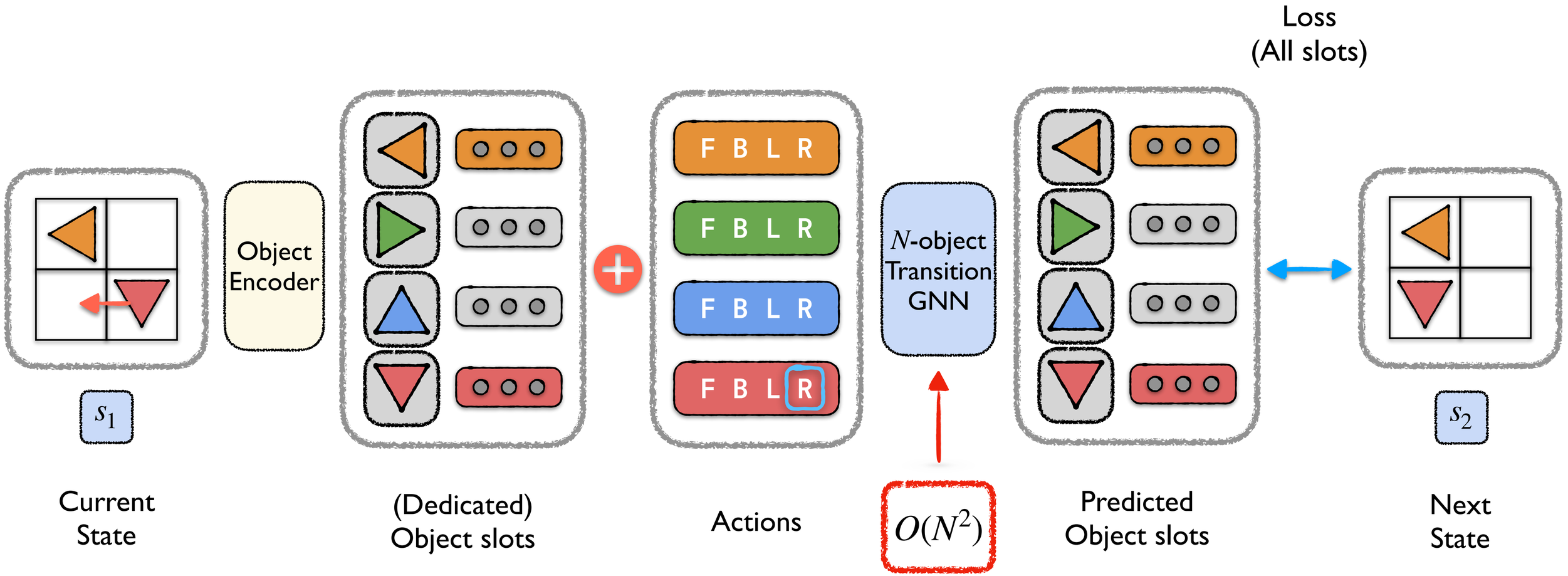
Recall: World Modeling

Learn a transition model $T : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{S}$



s_1 a_1 s_2

N-slot GNN: *A Naive Solution*



Research Questions

- How to measure compositional generalization (in world modeling)? ✓
- Can we guarantee when compositional generalization is achieved? ✓
- If possible, how can its implementation be efficient?

Three Necessary Components

1. Object Extraction

2. Actions Concatenated to Objects

3. Σ_N-equivariant transition model

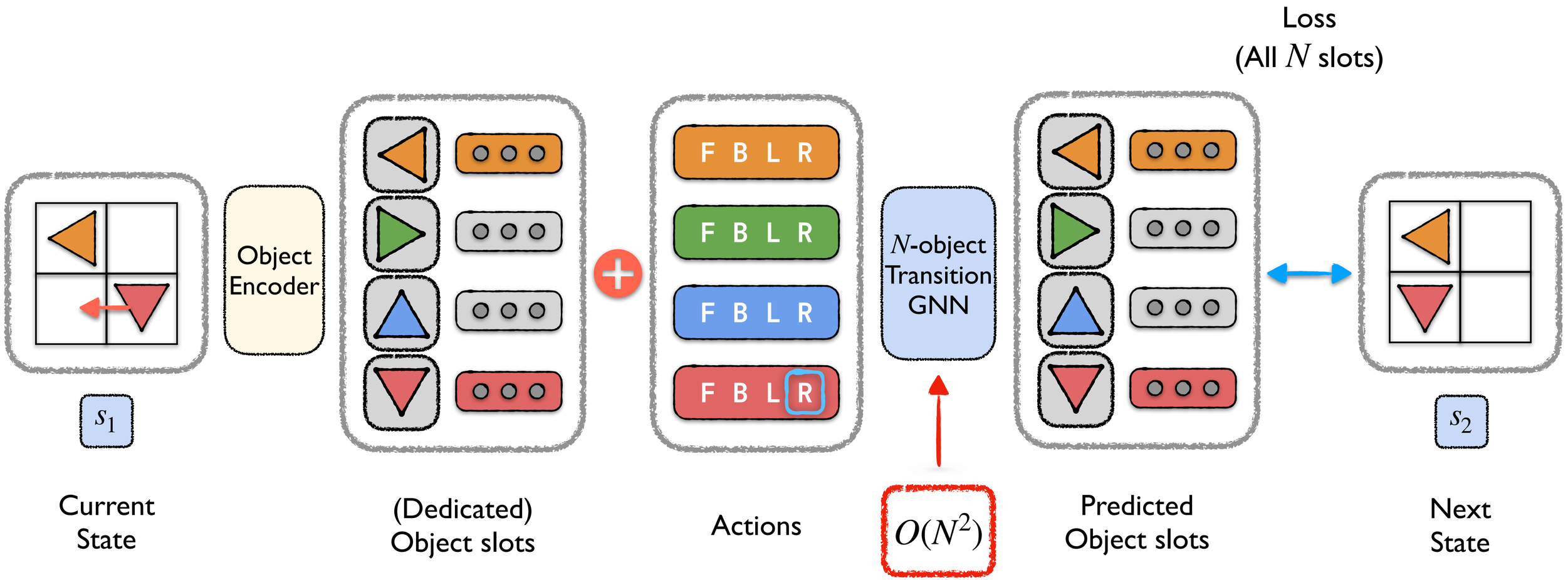
N-object GNN
"Library MDP"

1. Slot Extraction

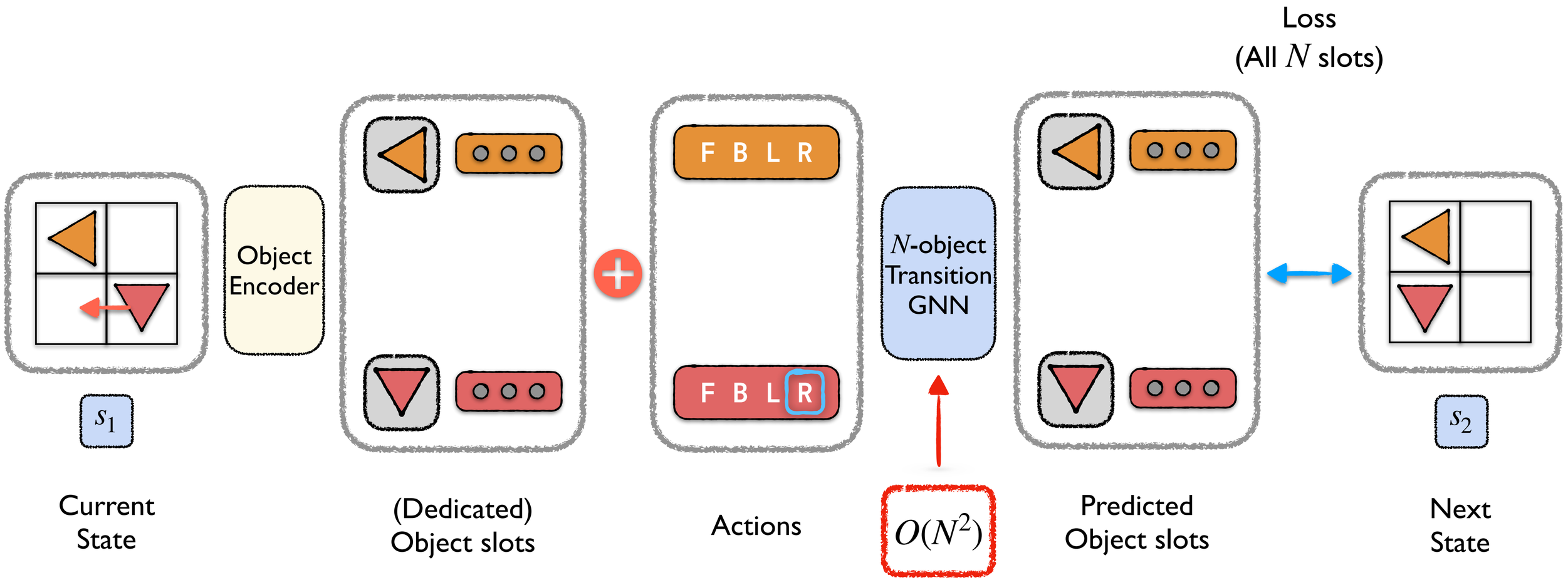
2. Action Binding to Slots

3. Σ_K-equivariant latent transition model

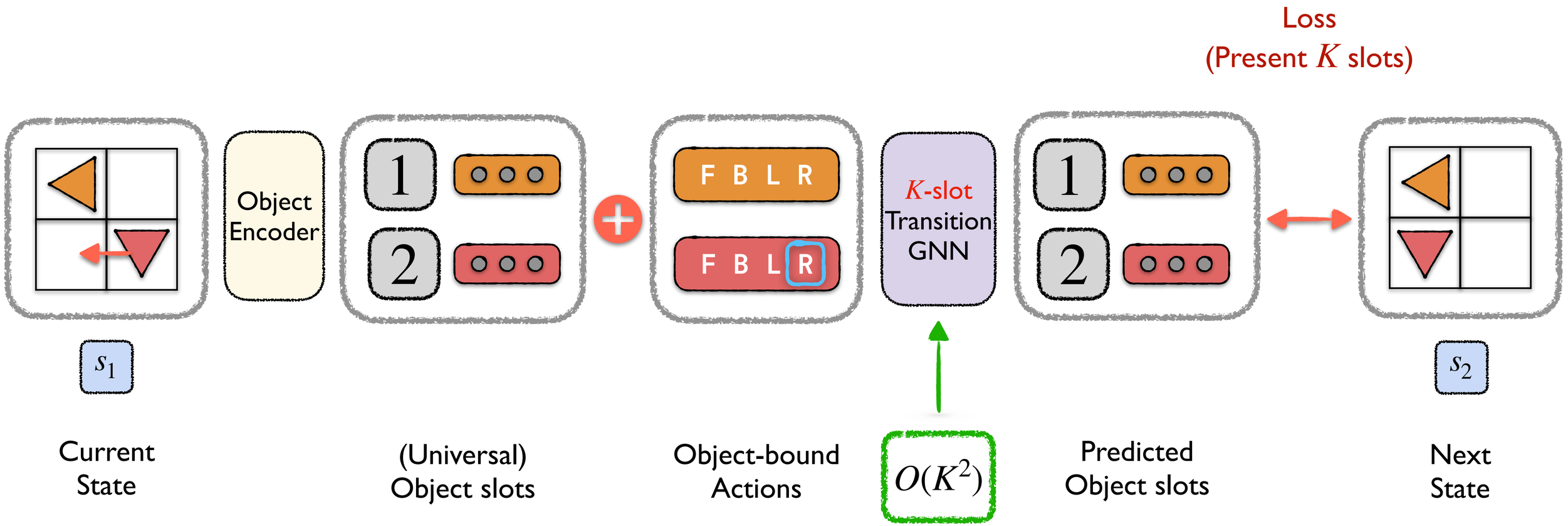
K-slot GNN: *More Efficient?*



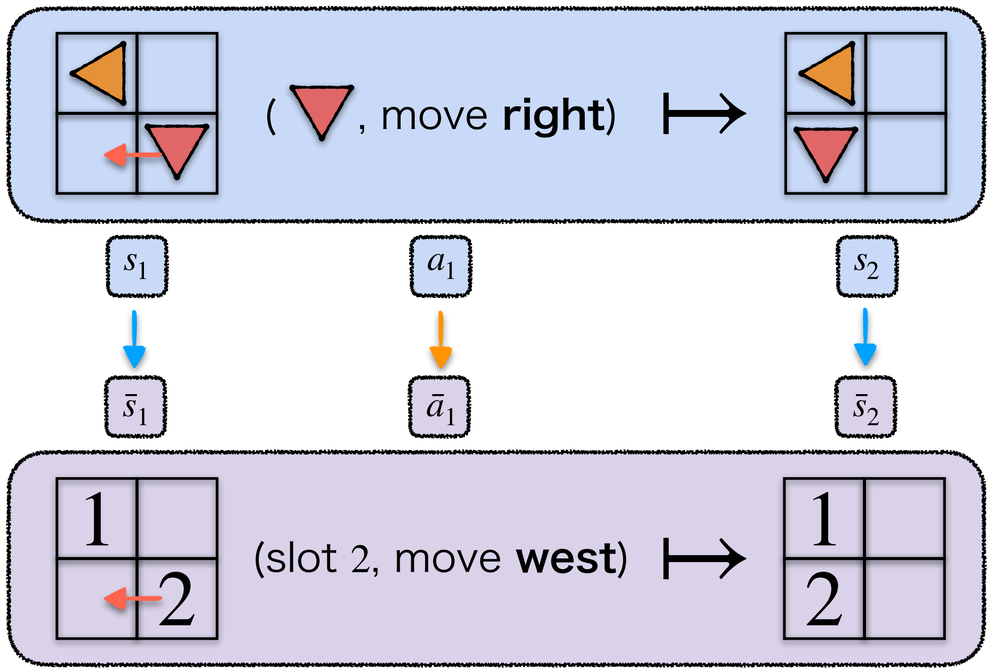
K-slot GNN: *More Efficient?*



K-slot GNN: *More Efficient?*



Slot MDP



Ideally: want a **canonical MDP** such that **all scenes** are **isomorphic** to it

"Slot MDP"
(Any slot can bind any object & action)

Challenge: *no canonical order* of objects or slots

Motivating Visualization

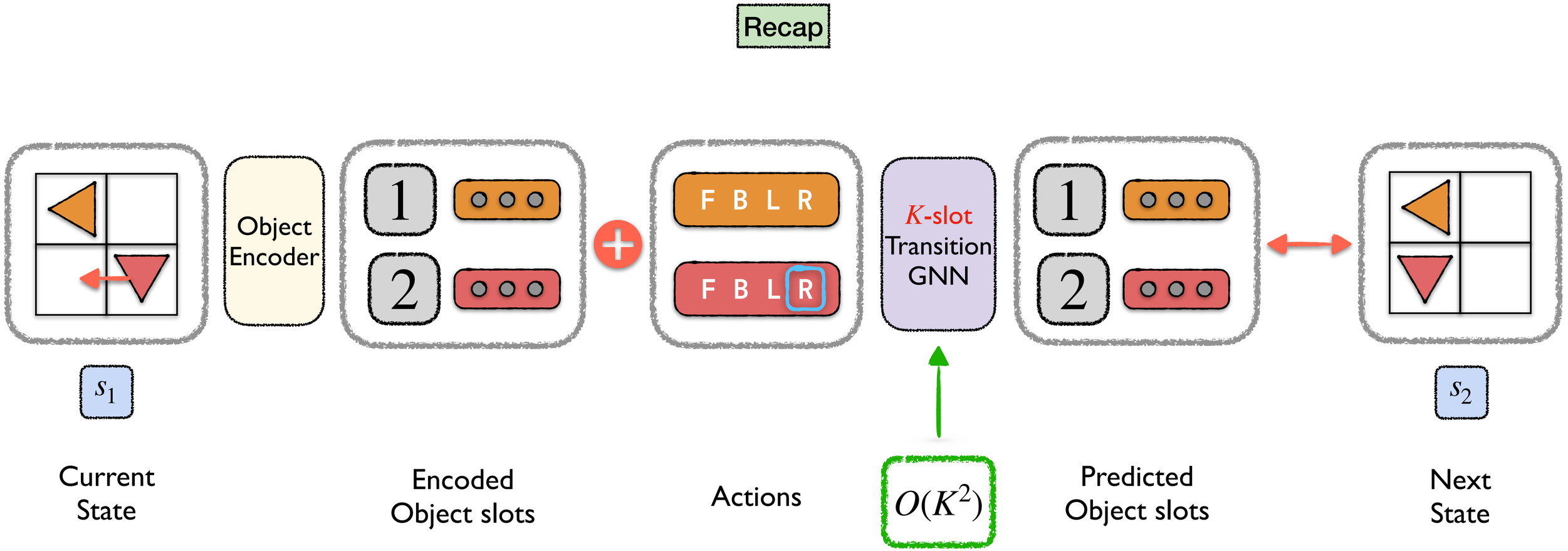
Challenge: *no canonical order* of objects or slots

Outline

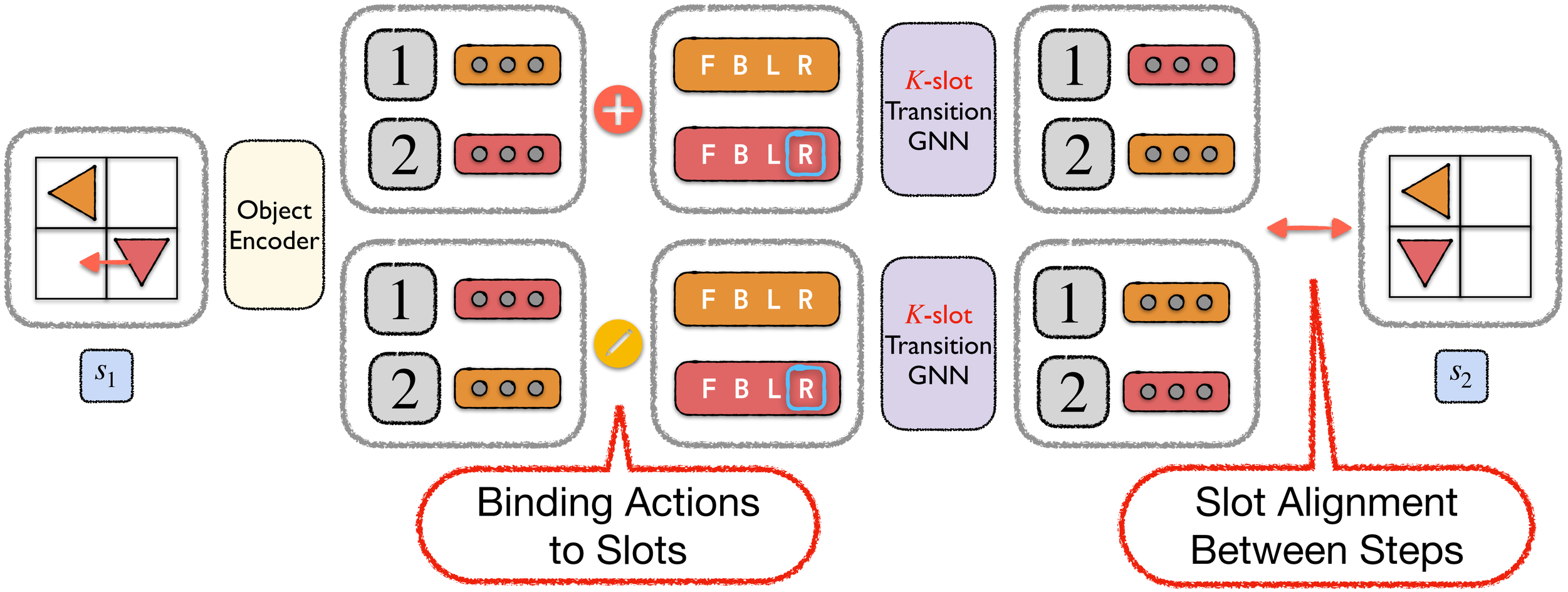
- Motivation
- Setup
- Defining Compositional Generalization
- Implementing Compositional Generalization
- Solving Binding Issue for Compositional Generalization
- Results

Solving
Binding Issue
for Compositional Generalization

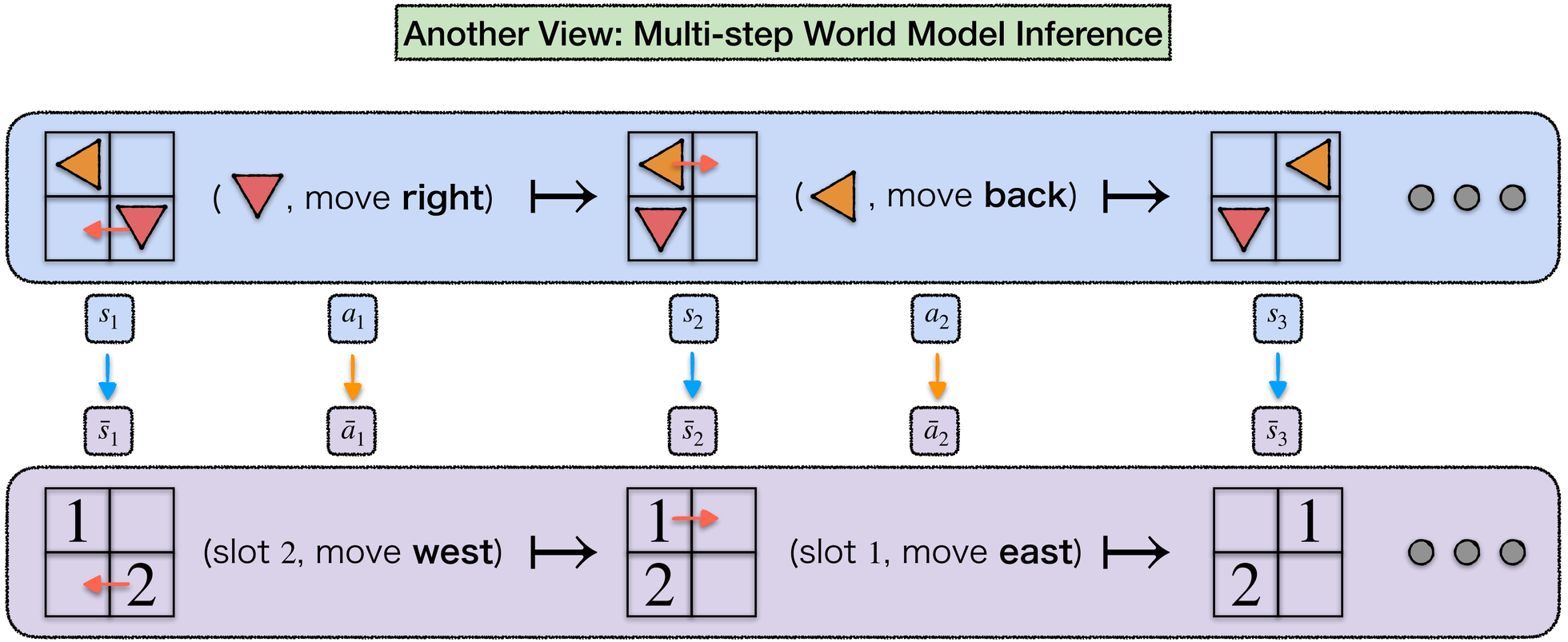
K-slot GNN: *Binding Issue*



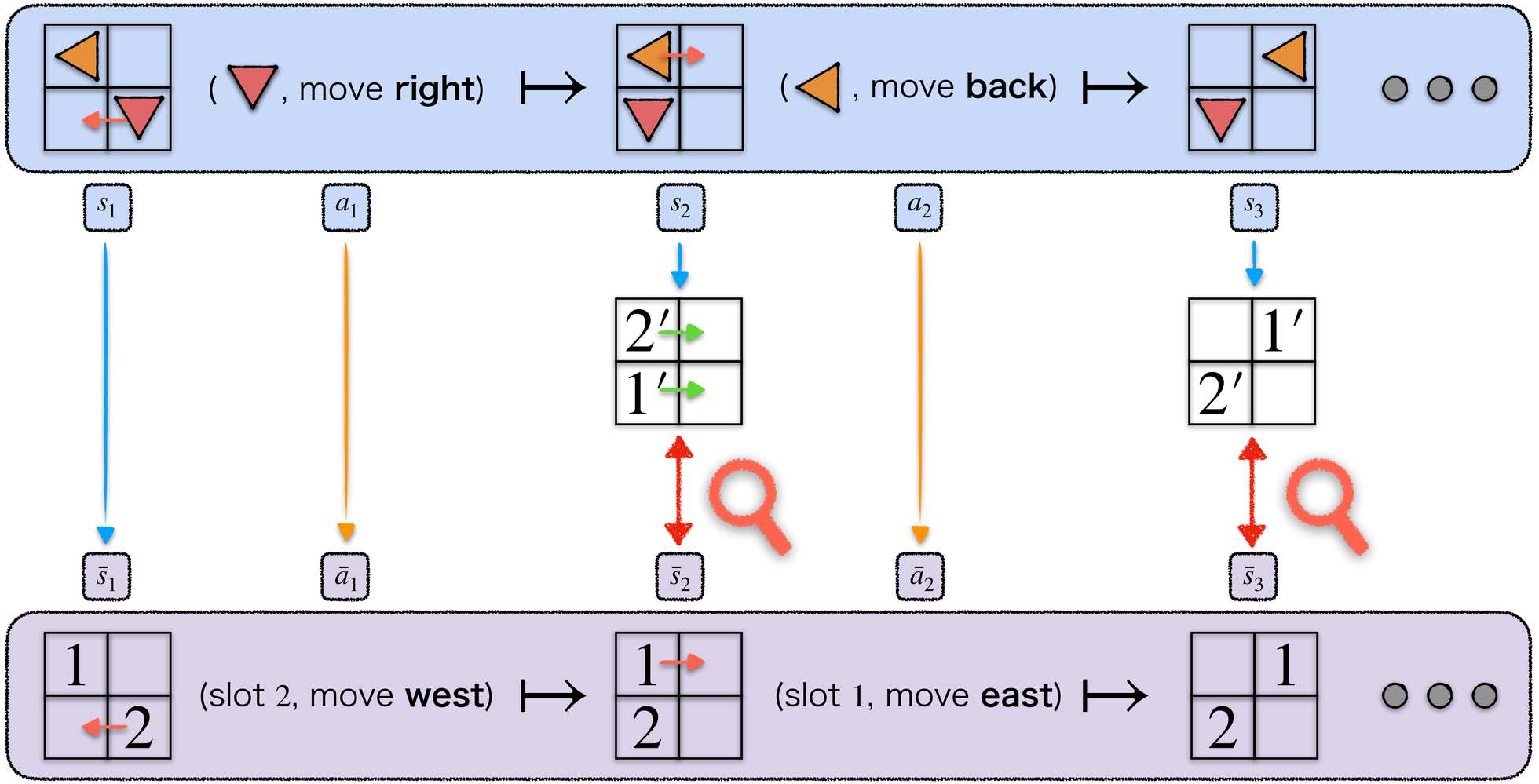
K-slot GNN: *Binding Issue*



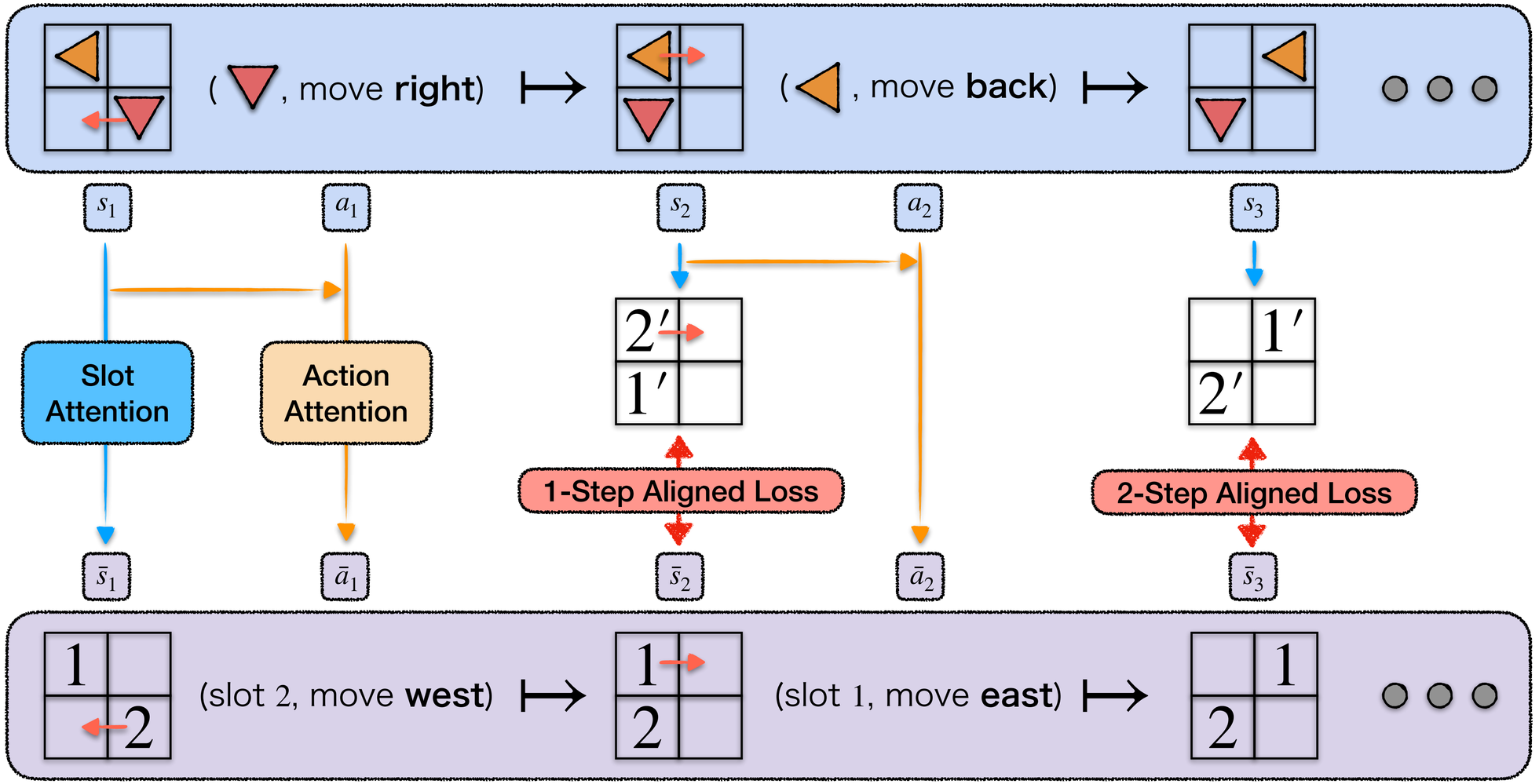
Binding Issue in Slot MDP



Binding Issue in Slot MDP



Method: Action Attention & Aligned Loss



Three Necessary Components

1. Object Extraction

2. Actions Concatenated to Objects

3. Σ_N-equivariant transition model

N-object GNN
"Library MDP"

1. Slot Extraction

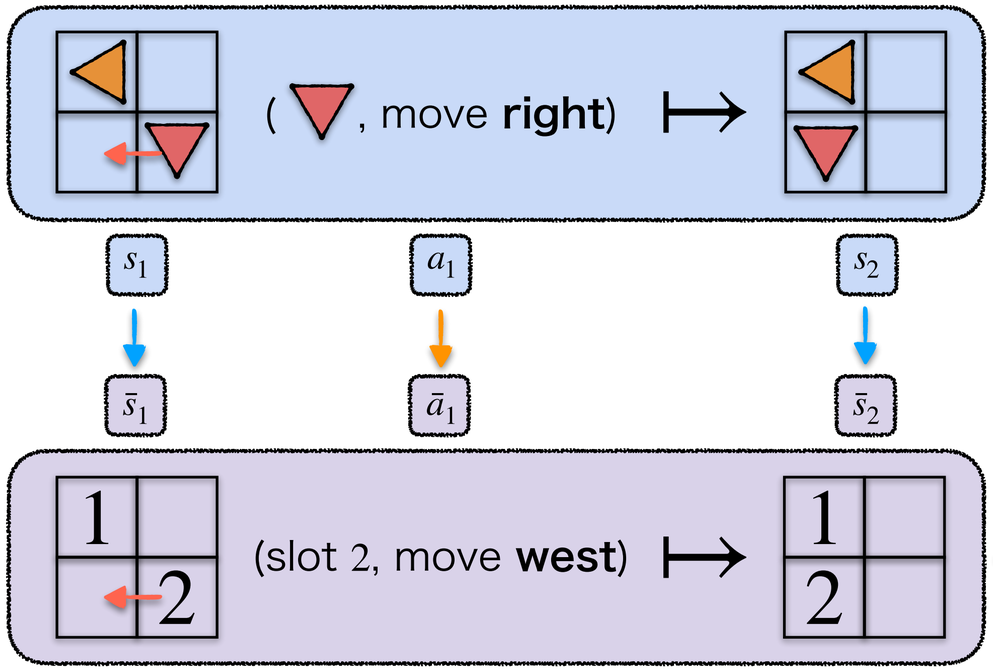
2. Action Binding to Slots

3. Σ_K-equivariant latent transition model

K-slot GNN
"Slot MDP"

Key Theorem: EE in Slot MDP

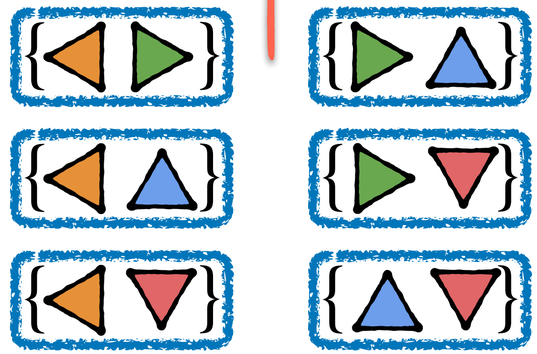
Theorem (informal): If object binding and action binding are correct, the equivariance error is related by:



$$\mathrm{EE}(T_{slot}) = \binom{N}{K} \cdot \mathrm{EE}(T_{lib})$$

EE in Slot MDP
K-slot

EE in Library MDP
N-object



Bind **objects** and **actions** correctly for **all scenes**

Corollary: CG in Latent Space

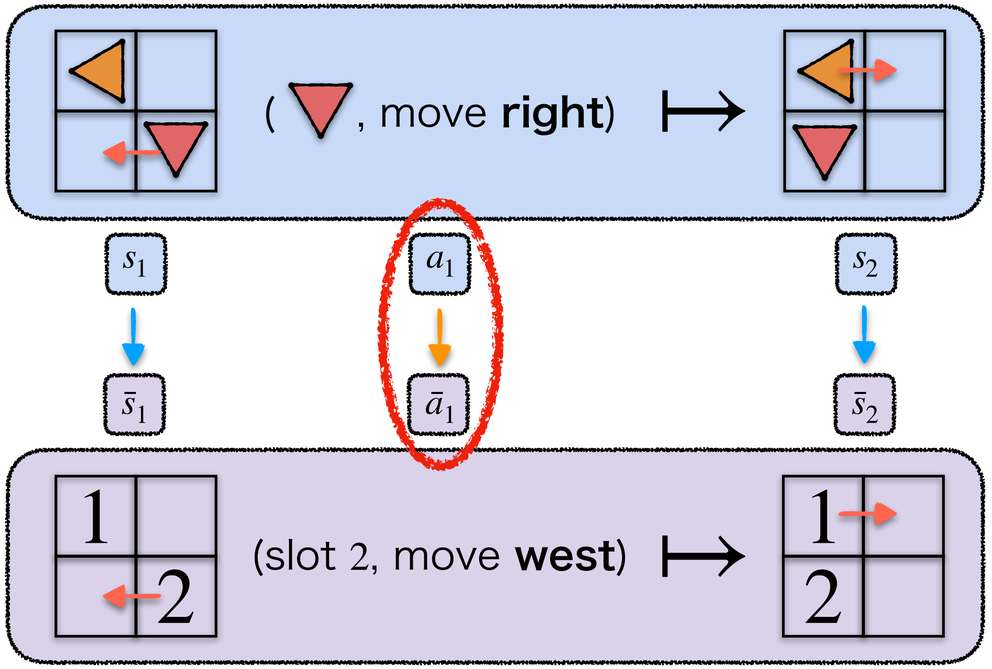
$$0 = \mathrm{EE}(T_{slot}) = \binom{N}{K} \cdot \mathrm{EE}(T_{lib})$$

EE in Slot MDP
K-slot

EE in Library MDP
N-object

If object binding and action binding are correct:

$\bar{\sigma} \,.\, \phi(s) = \phi(\sigma \,.\, s)$ **and** $\bar{\sigma} \,.\, \alpha_s(a) = \alpha_{\sigma.s}(\sigma \,.\, a)$

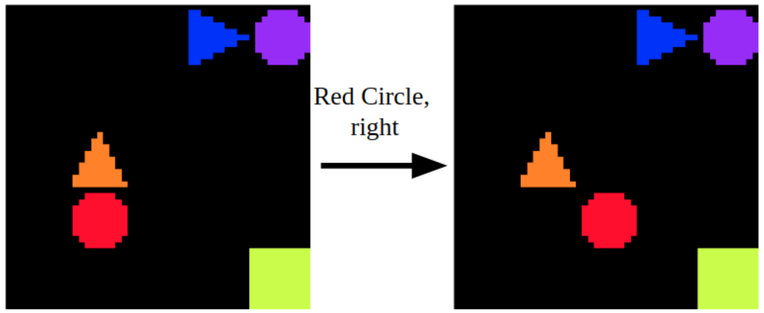


Outline

- Motivation
- Setup
- Defining Compositional Generalization
- Implementing Compositional Generalization
- Solving Binding Issue for Compositional Generalization
- Results

Results

Environment (Block Pushing)



Object Library: Block Pushing

Random objects (locations, color, shapes)
K = 5, N = 5, 10, 20, 30

Experimental Setup

- **Exact CG**: Σ_N-equivariant WM
- **No CG**: break each component
- **Soft CG**: Learned Σ_K-equivariant WM

1. Slot Extraction

2. Action Binding to Slots

3. Σ_K-equivariant latent transition model

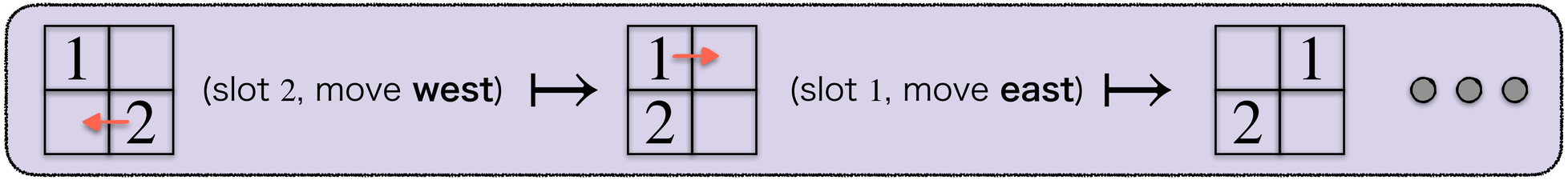
Measuring CG in Practice

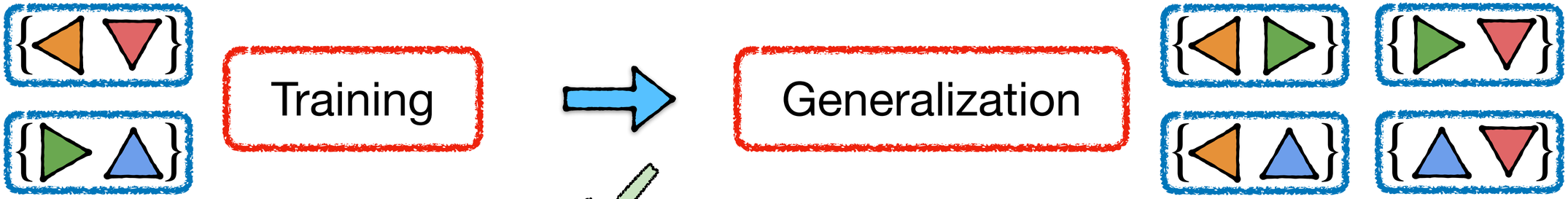
$$\mathrm{EE}(T_{lib}) \triangleq \mathbb{E}\left[\left|\hat{T}_{\mathbb{L}}(s' \mid s, a) - \hat{T}_{\mathbb{L}}(\sigma . s' \mid \sigma . s, \sigma . a)\right|\right]$$

$$\sigma \in \Sigma_N, (s, a, s') \in \mathcal{S}_{\mathbb{L}} \times \mathcal{A}_{\mathbb{L}} \times \mathcal{S}_{\mathbb{L}}$$

$$\mathrm{EE}(T_{slot}) = \binom{N}{K} \cdot \mathrm{EE}(T_{lib})$$

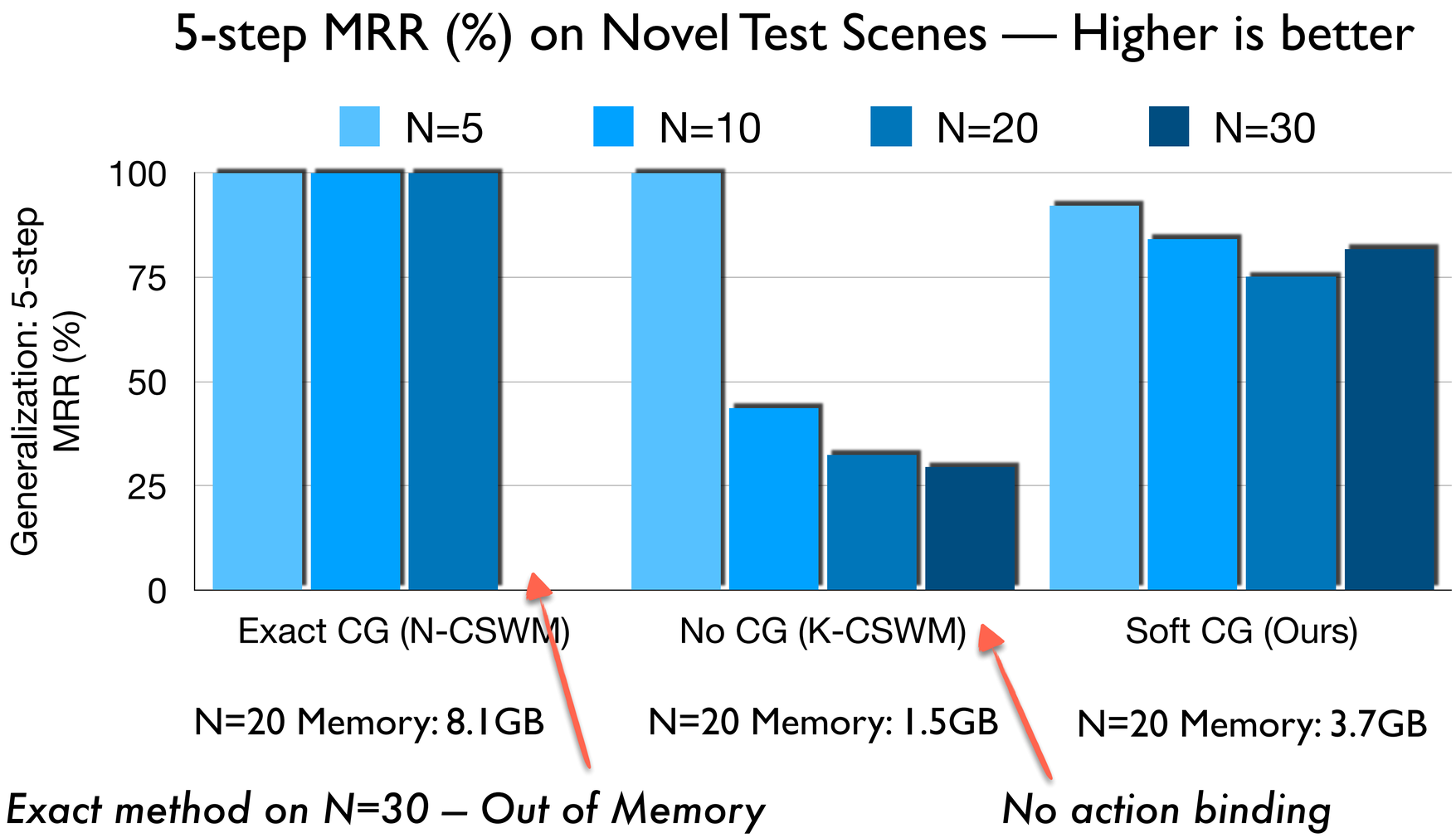
Binding





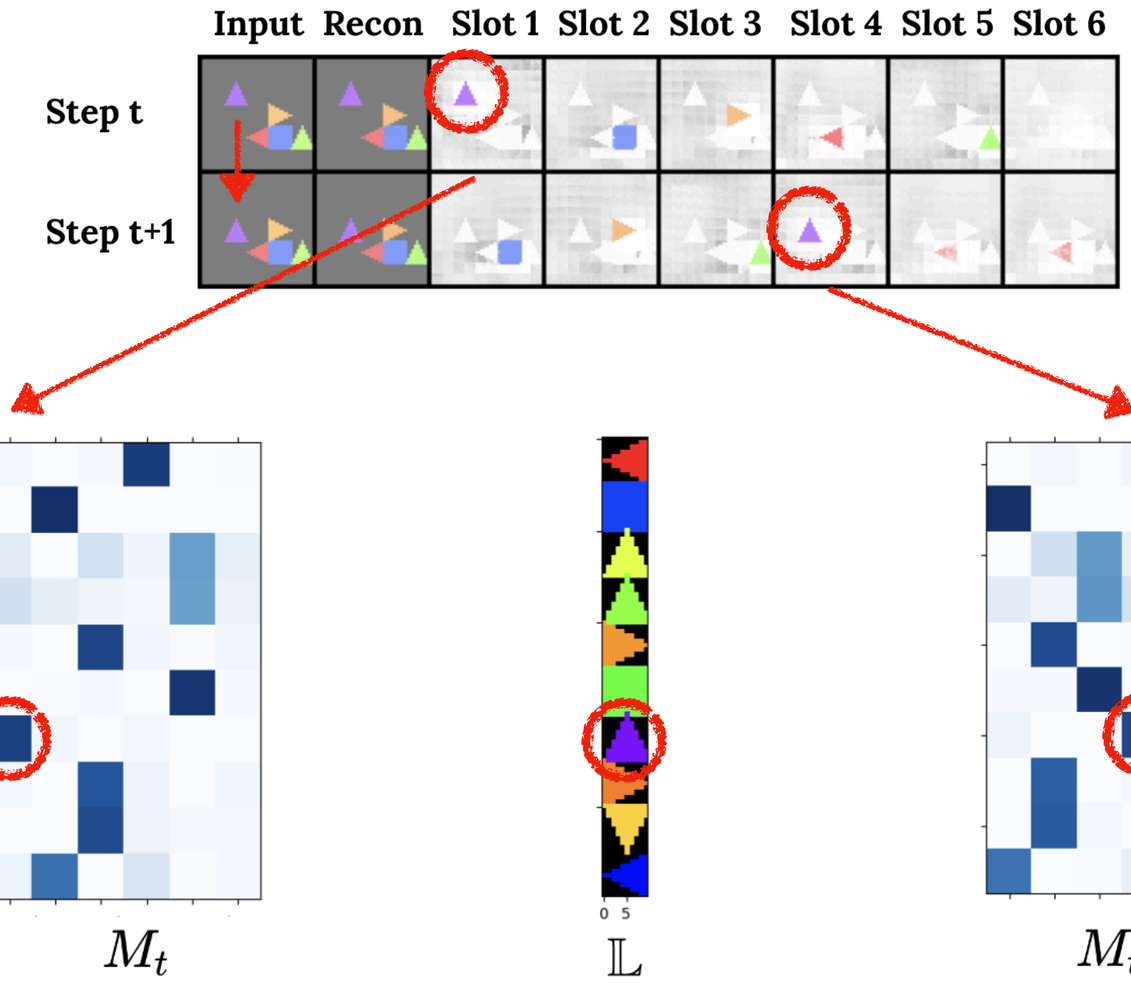
[Keysers et al., ICLR 2020, "Measuring Compositional Generalization: A Comprehensive Method on Realistic Data"]

Quantitative Results



[Kipf et al., ICLR 2020, Contrastive Learning of Structured World Models]

Binding Visualization



5+1 rows:
K=5 slots + 1 background

10 columns:
N=10 objects

Found object identity through actions (Objects unknown)

Summary

Study compositional generalization in object-oriented world modeling

- Introduce a conceptual environment (Object Library)
- Formalize compositional generalization (by equivariance error)
- Propose a soft approach with guarantees
- Practically measure compositional generalization

More Information

Check out our project website:

http://lfzhao.com/oowm/

zhao.linf@northeastern.edu