Northeastern University
Khoury College of
Computer Sciences

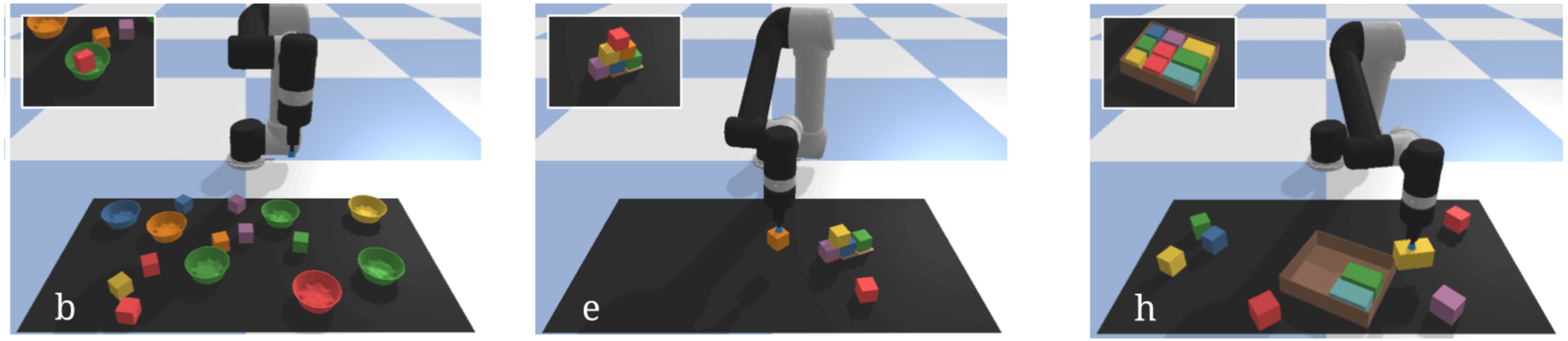
Toward Understanding Compositional Generalization in Object-Oriented World Modeling

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

ICML 2022 Long Presentation

Motivation

Motivation: Planning in robotic manipulation needs accurate transition model

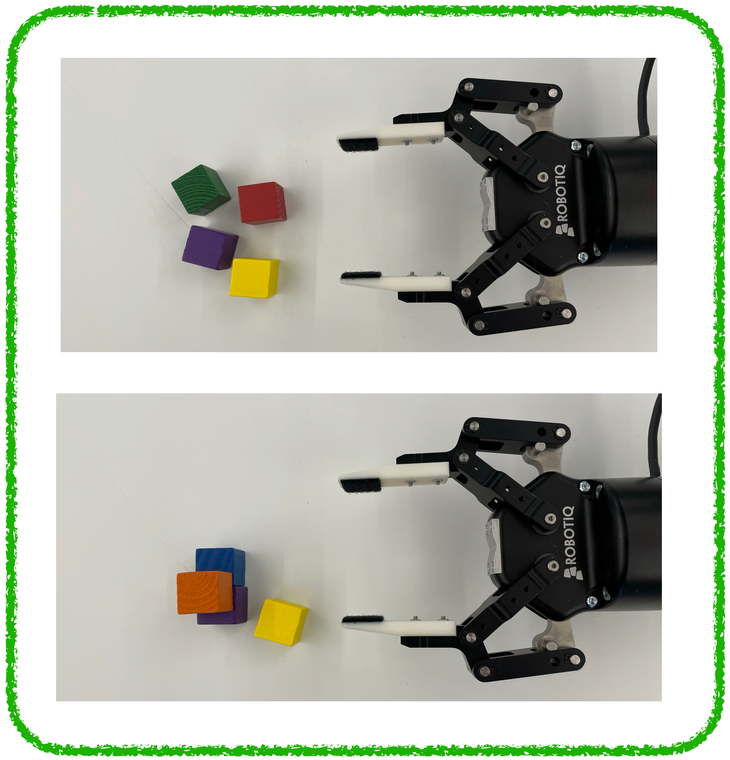


Issue: Current models struggle in generalization when multiple objects recombine

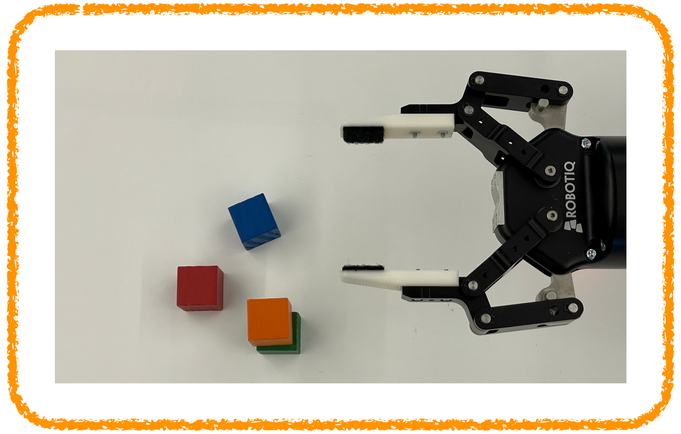
Credit: Ravens

Motivation: Compositional Generalization

How to learn a generalizable transition model when multiple objects recombine?



Train
Learning a world model on some combinations

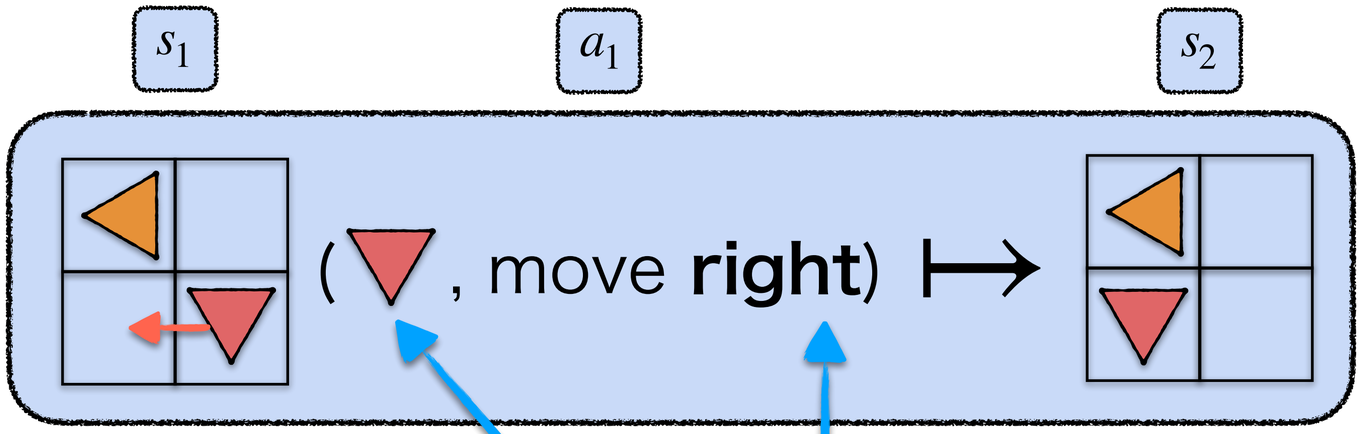


Generalization:
Use the world model on novel combinations

Background: World Modeling

Goal: Learn a world model (transition model) $T : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{S}$

4 Orientations



1) Image
2) Grid

1) Factorized
2) 4 directions
3) Relative

Research Questions

Compositional Generalization (CG)

- How to measure compositional generalization (in world modeling)?
- How to achieve guaranteed compositional generalization?
- How can its implementation be efficient?

Outline

- Motivation
- Setup: Object Library
- Defining Compositional Generalization
- Implementing Compositional Generalization
- Solving Binding Issue In End-to-end World Modeling for CG
- Results

Proposed Setup: Object Library

Motivation: sampling words from vocabulary to form sentences

Object Library $\mathbb{L}$

"Vocabulary"

All possible objects

Scenes $\mathbb{O}_i \subset \mathbb{L}$

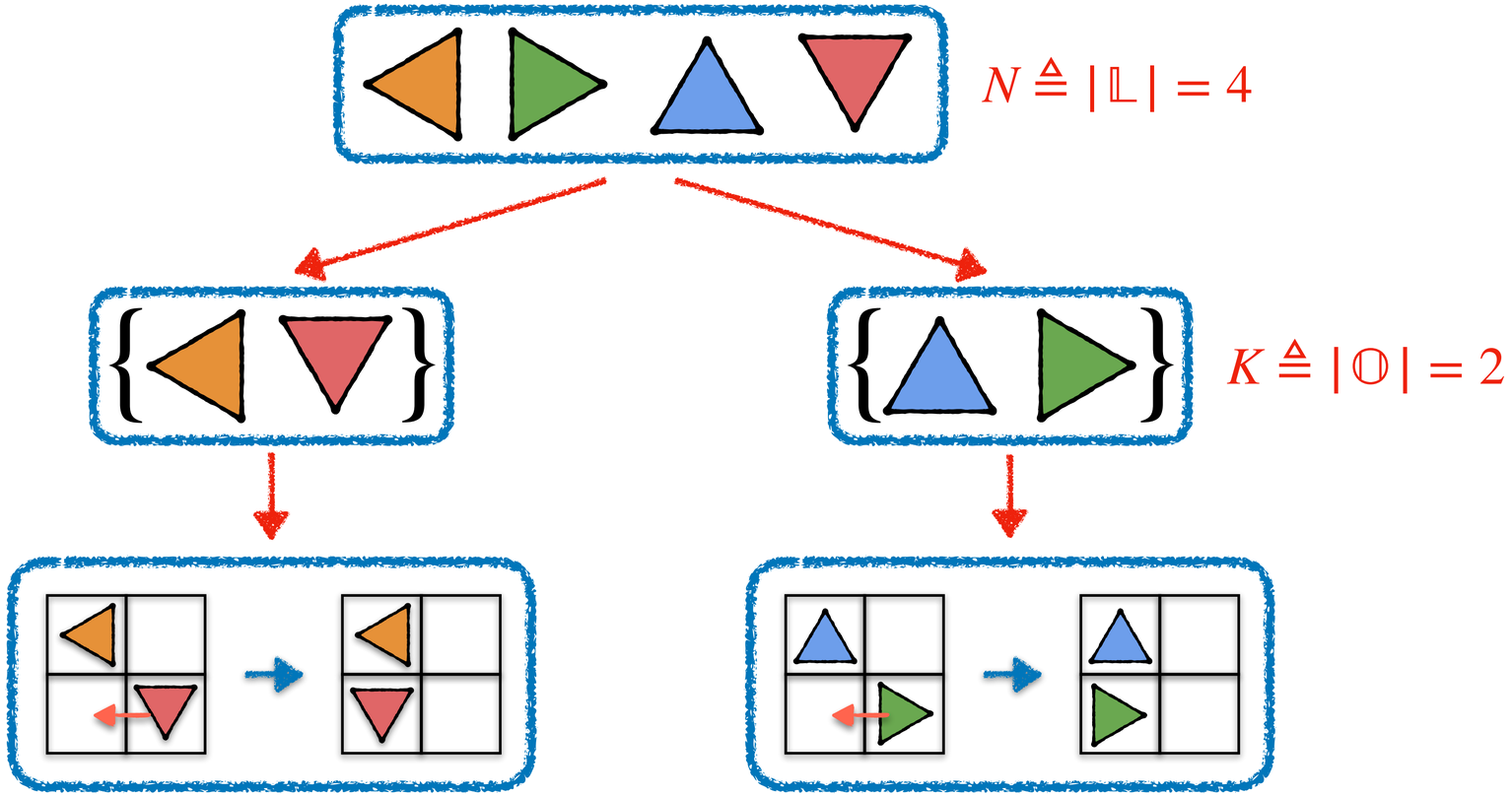
(Ordered) "Sentences"

A combination of objects

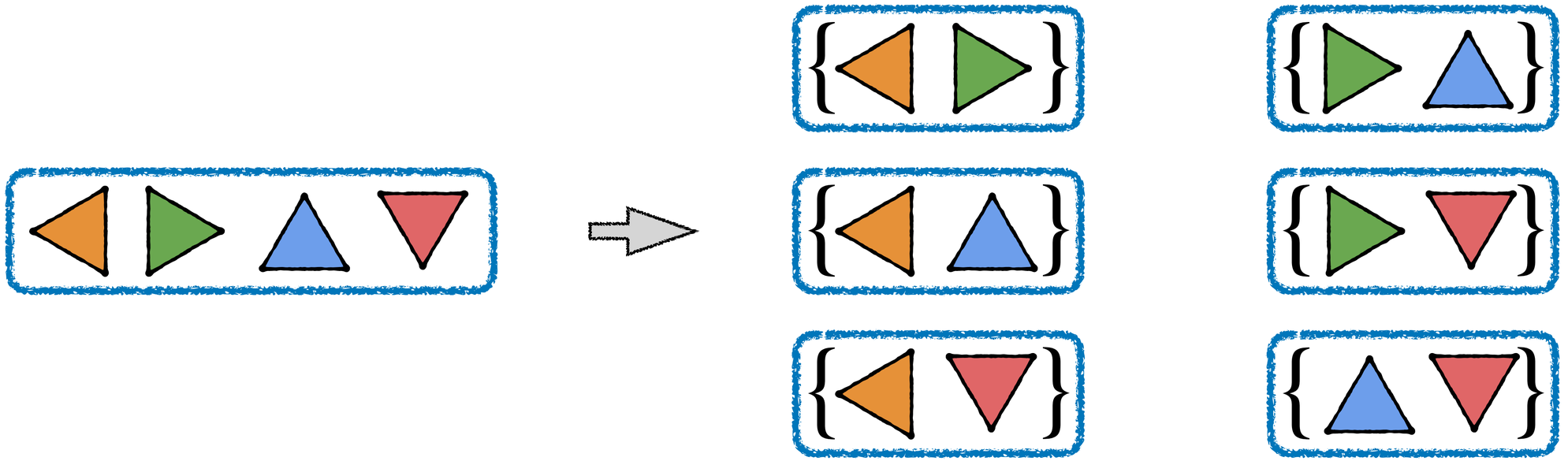
Scene MDPs $\mathscr{M}_{\mathbb{O}_i}$

Generated by $\mathbb{O}$

Moving objects on a table



Proposed Setup: Object Library

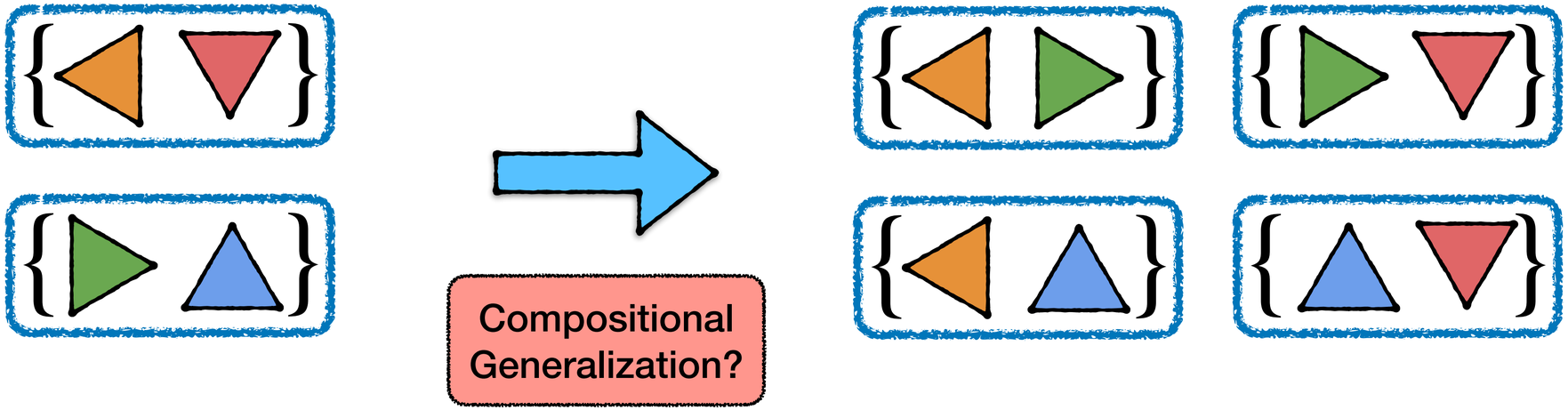


All $\binom{N}{K} = \binom{4}{2} = 6$ combinations

Object Library for CG

Training
Train the world model on some scenes

Generalization
Test the model on novel scenes



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

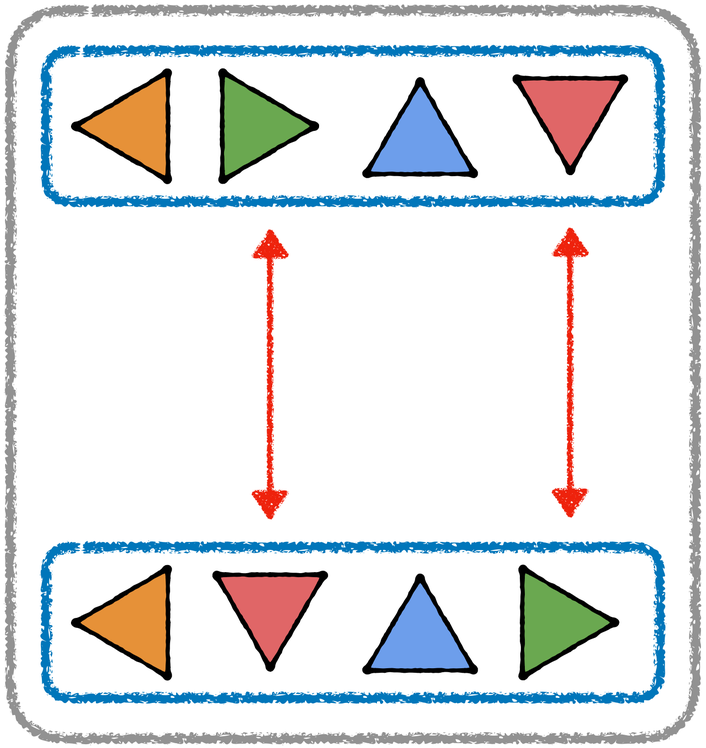
Outline

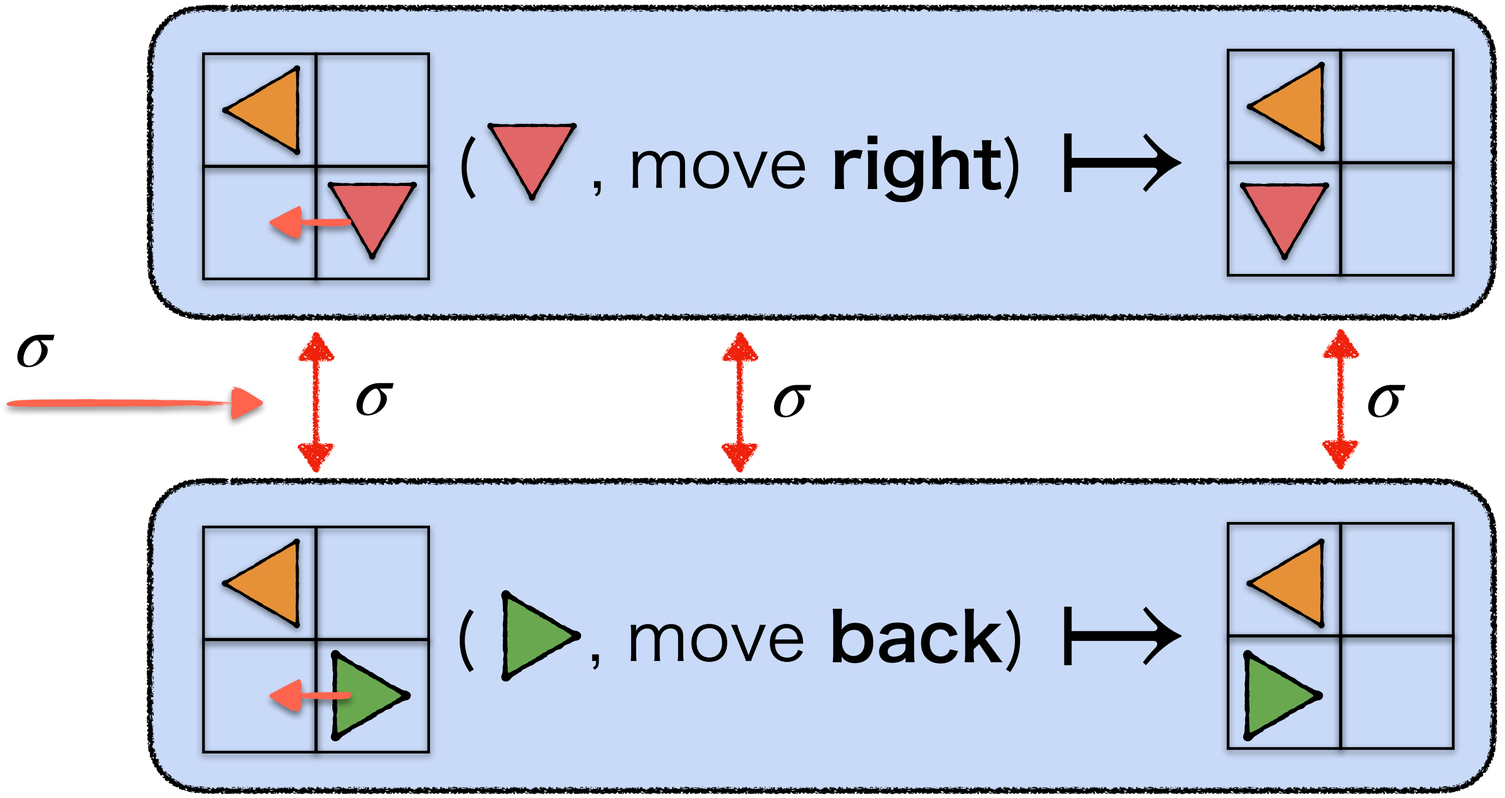
- Motivation
- Setup: Object Library
- Defining Compositional Generalization
- Implementing Compositional Generalization
- Solving Binding Issue In End-to-end World Modeling for CG
- Results

Object-replacement: Formalism

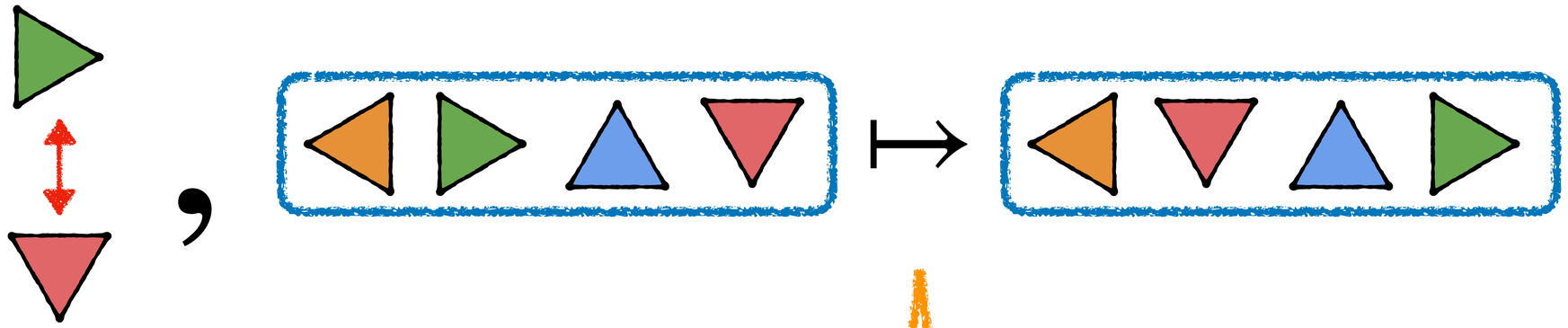
How to quantitively define and measure compositional generalization?

Example: Replace Object Identity





Object-replacement Operation

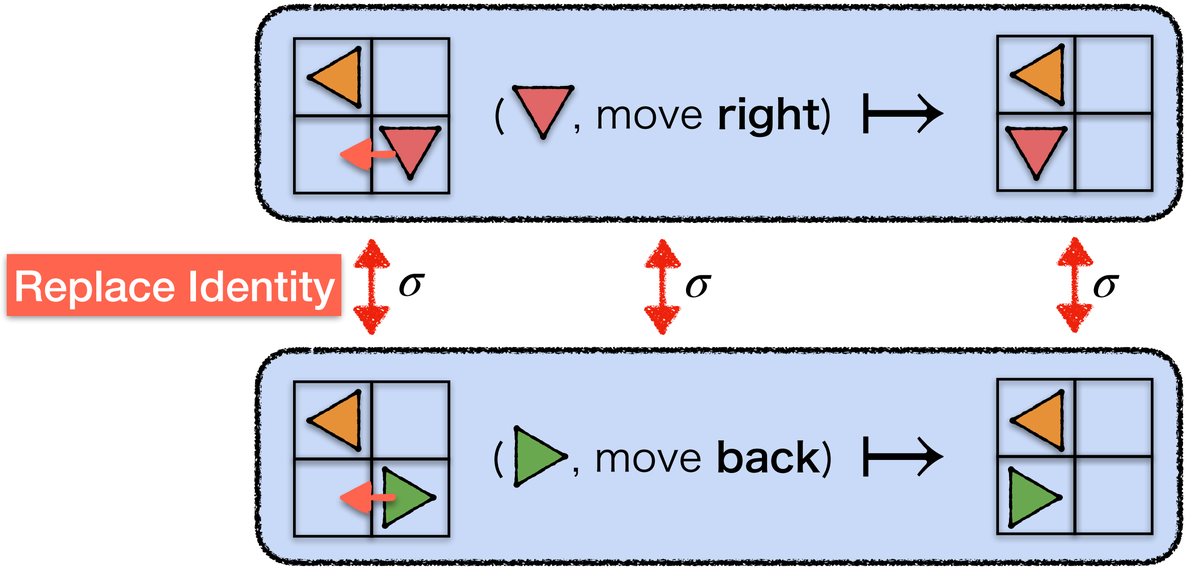


Object-replacement operation:

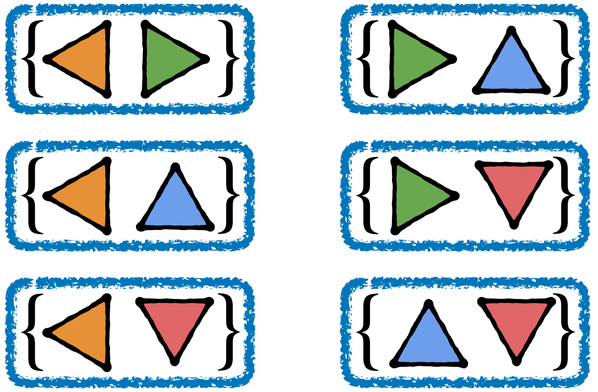
$$p_{\mathbb{L}} : \Sigma_N \times \mathbb{L} \to \mathbb{L}$$

Σ_N — order-N Permutation Group

Measure CG with Equivariance Error



Expectation over All combinations



How well can the model predict transition on scenes with different objects?

Cast Compositional Generalization as (Permutation) Equivariance Error *EE* in Transition Modeling *T*

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

Expectation over:

1. All object replacements: $\sigma \in \Sigma_N$
2. All transitions $(s, a, s') \in \mathcal{S}_{\mathbb{L}} \times \mathcal{A}_{\mathbb{L}} \times \mathcal{S}_{\mathbb{L}}$

Summary 1

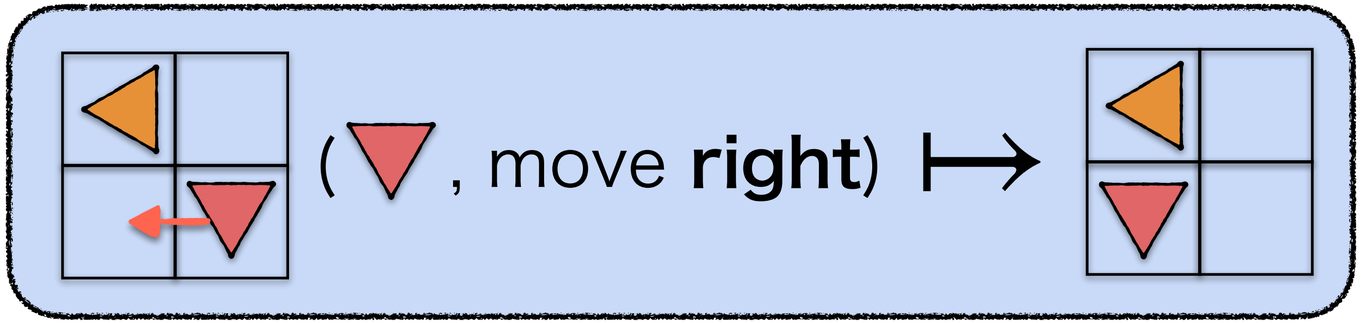
We propose a setup Object Library to define compositional generalization, using permutation equivariance error w.r.t. object-replacement operation.

Outline

- Motivation
- Setup: Object Library
- Defining Compositional Generalization
- Implementing Compositional Generalization
- Solving Binding Issue In End-to-end World Modeling for CG
- Results

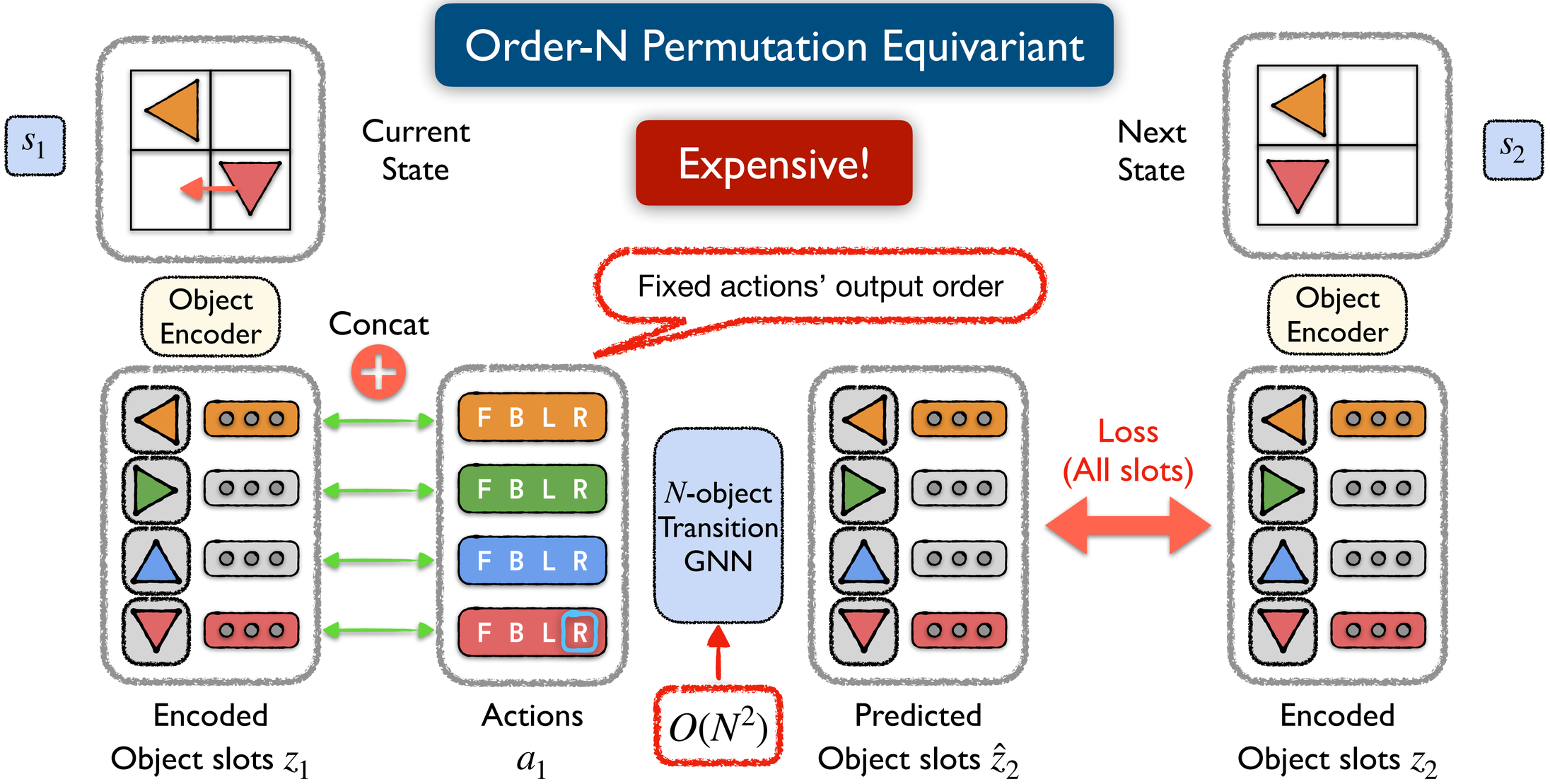
Recall: World Modeling

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



s_1 a_1 s_2

First Solution: *N*-slot GNN



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

Retrospect: Necessary Components

Formulation

(Permutation) Equivariance Error *EE* in Transition Modeling *T*

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

Expectation over:
1. All object replacements: $\sigma \in \Sigma_N$
2. All transitions $(s, a, s') \in \mathcal{S}_{\mathbb{L}} \times \mathcal{A}_{\mathbb{L}} \times \mathcal{S}_{\mathbb{L}}$

Implementation

1. Object Extraction

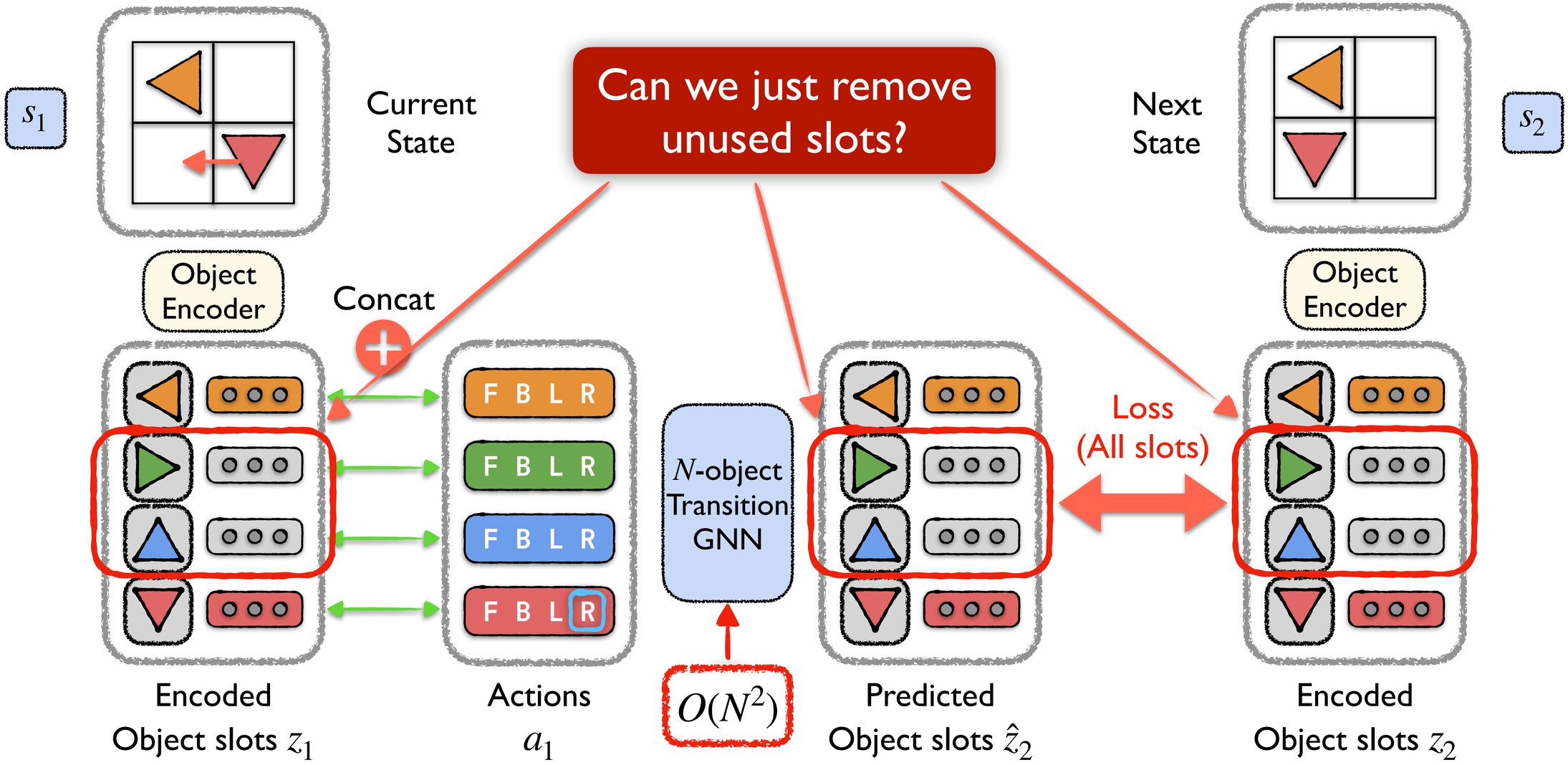
2. Actions Concatenated to Objects

3. Σ_N-equivariant transition model

1. It can achieve perfect compositional generalization

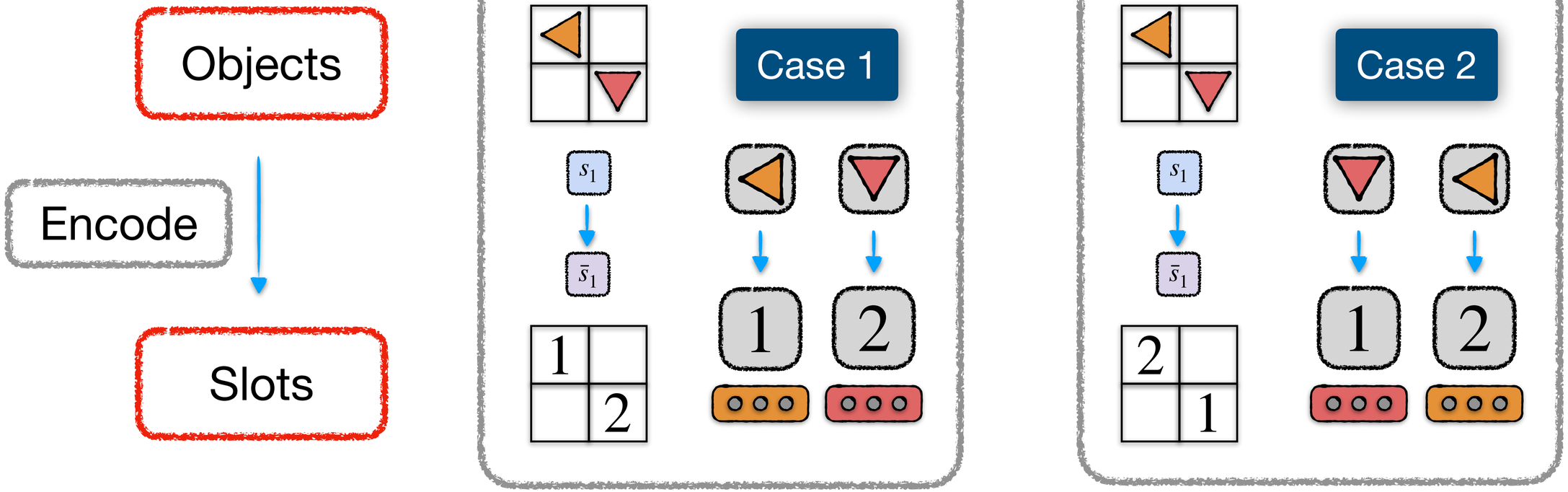
2. It cannot scale up well — $O(N^2)$ complexity (N = library size)

More efficient solution?



Fact: Object Slots are Unordered

Objects and Slots do not have canonical order



Motivating Visualization

Challenge: *no canonical order* of objects or slots

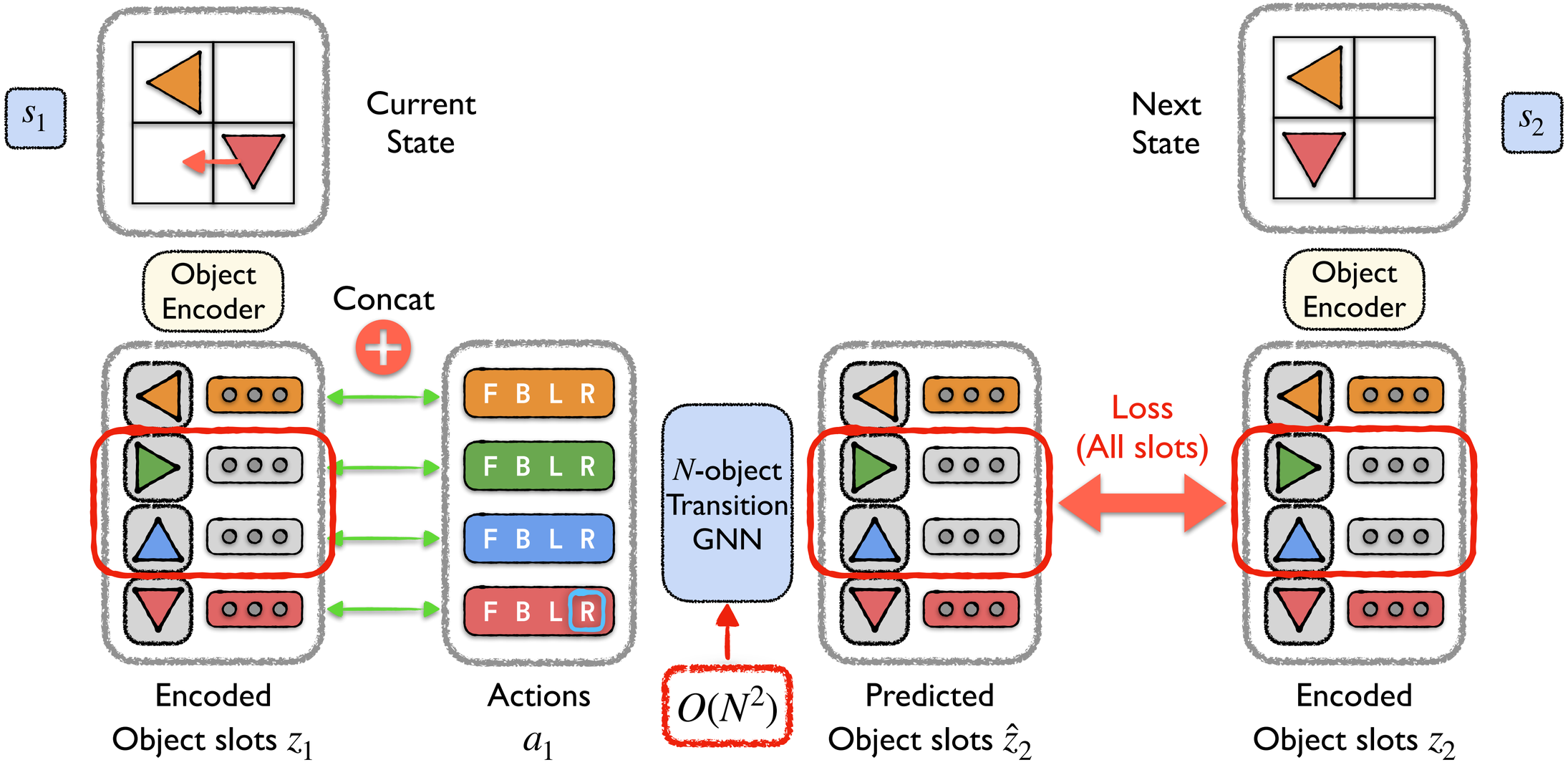
Summary 2

We give the first solution (using N-slot GNN) and highlight the key obstacle in end-to-end compositional generalization: *lack of canonical ordering*.

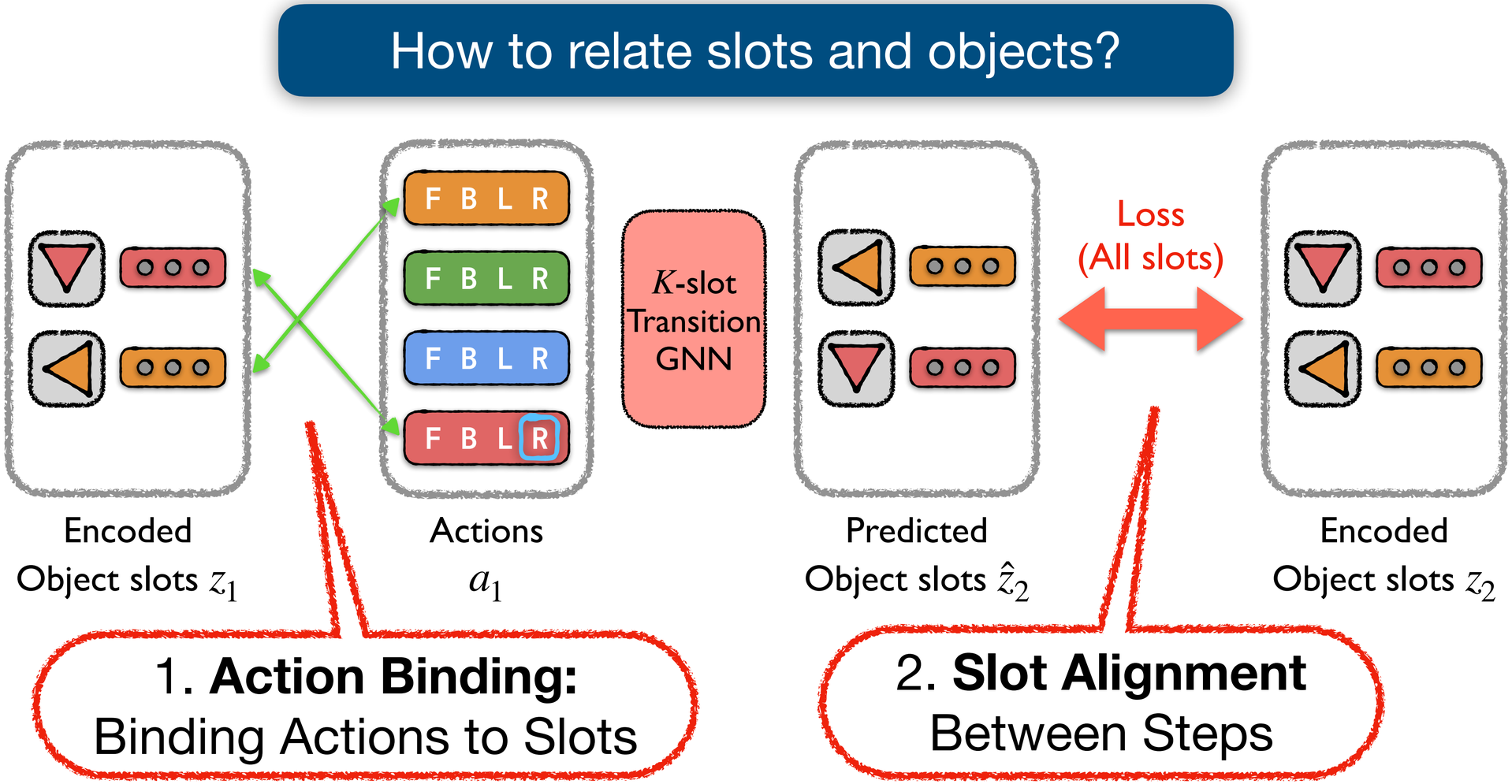
Outline

- Motivation
- Setup: Object Library
- Defining Compositional Generalization
- Implementing Compositional Generalization
- Solving Binding Issue In End-to-end World Modeling for CG
- Results

Binding Issue in World Modeling

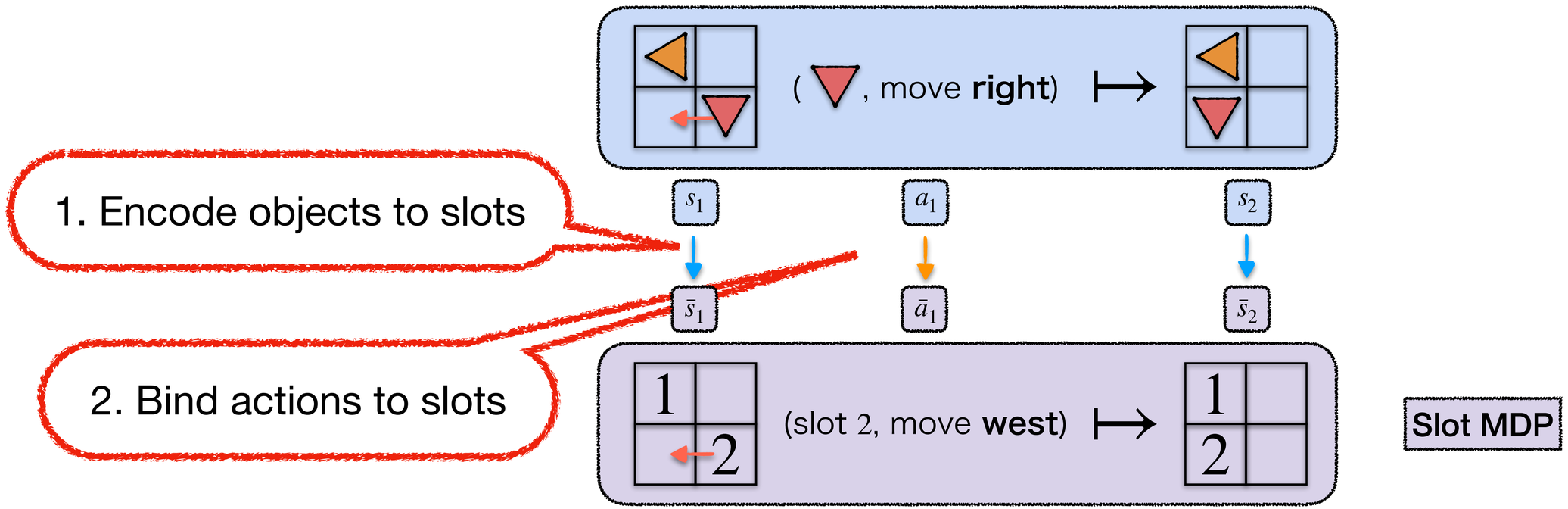


Binding Issue in World Modeling

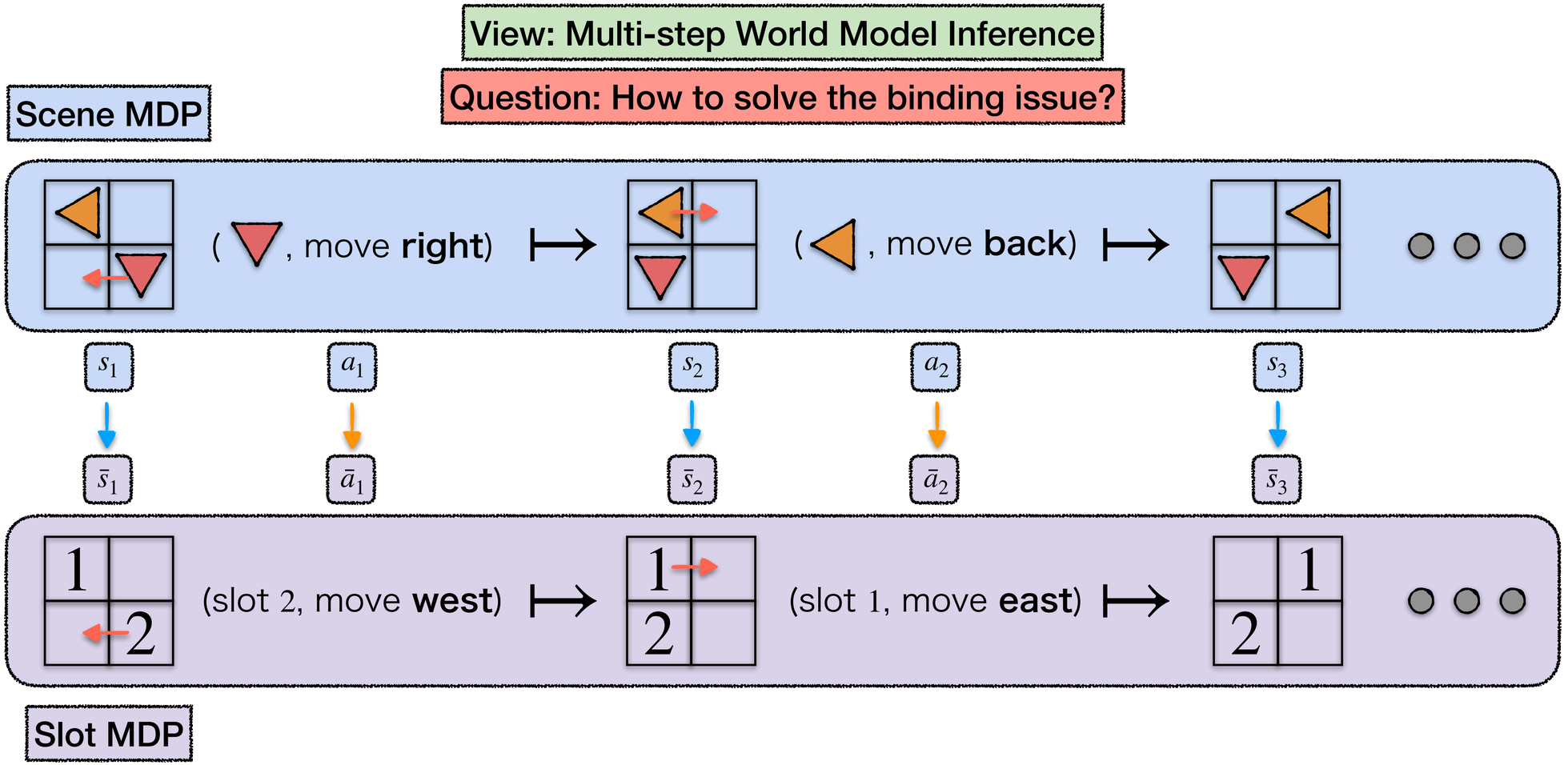


Slot MDP: Bind to Any Scene MDP

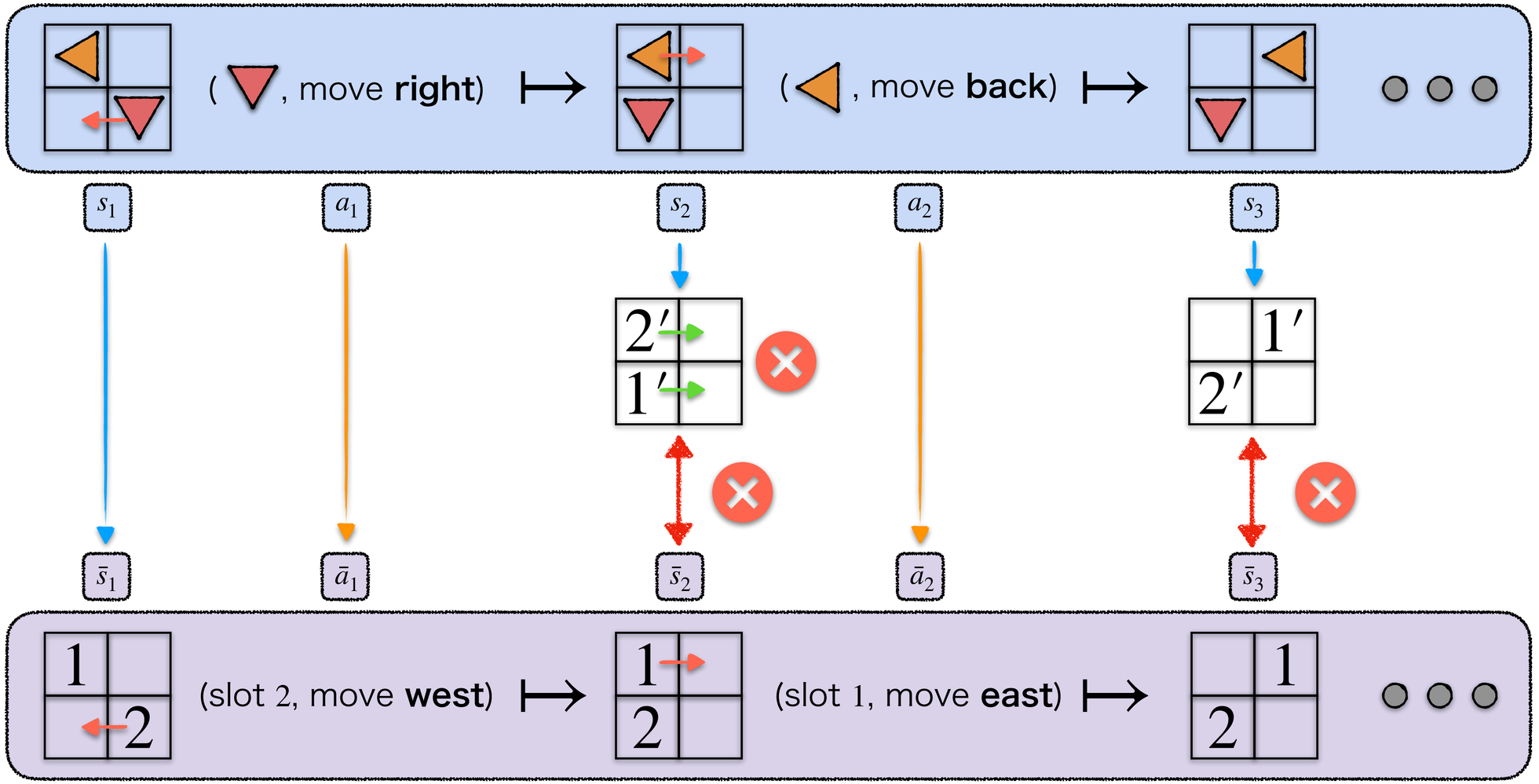
To solve the binding issue —
Learn a canonical MDP model that can correctly bind to any scene (object combination)



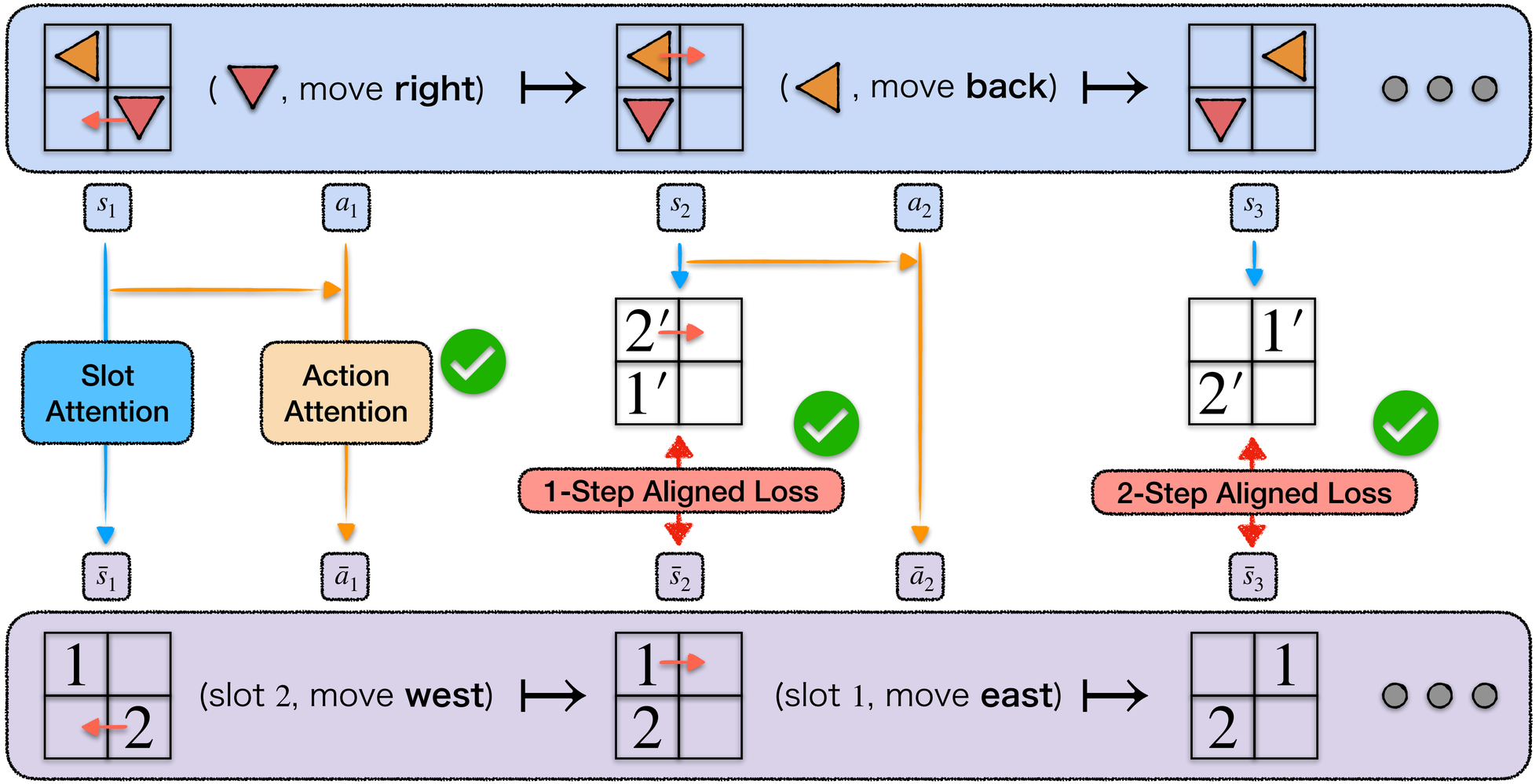
Solving *Binding Issue* in Slot MDP



Solving *Binding Issue* in Slot MDP



Method: Action Attention & Aligned Loss



Retrospect: Necessary Components

Expensive!	Efficient
1. Object Extraction	1. Slot Extraction
2. Actions Concatenated to Objects	2. Action Binding to Slots
3. Σ_N-equivariant transition model	3. Σ_K-equivariant latent transition model

Ours achieves efficient compositional generalization end-to-end:
Homomorphic Object-oriented World Model (HOWM)

Key Theorem: EE in Slot MDP

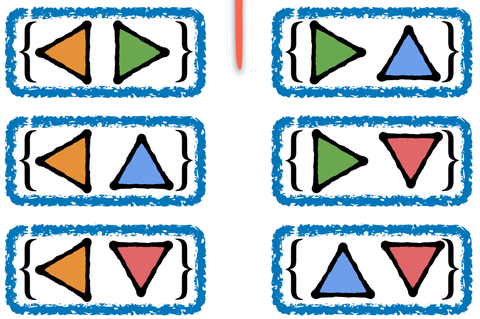
Theorem (informal):
If actions correctly bind to object slots,
the equivariance error is related by:

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

EE in Slot MDP
K-slot

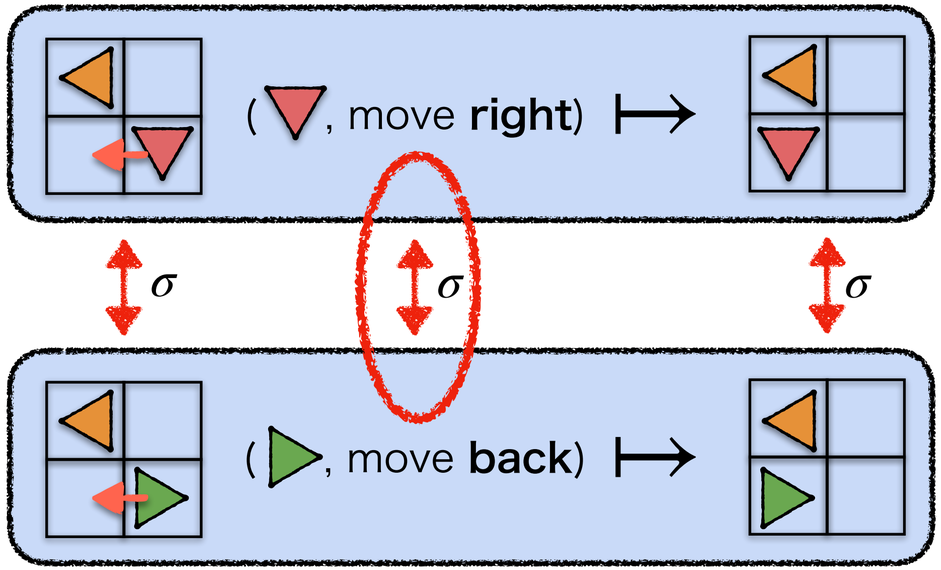
EE in Library MDP
N-object

Bind correctly for **all scenes**



Intuition: Binding = slot MDP can
correctly simulate any scene MDP

Corollary: CG in Slot MDP



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

EE in Slot MDP
K-slot

EE in Library MDP
N-object

Intuition: Learning perfectly in slot MDP
= Compositional Generalization

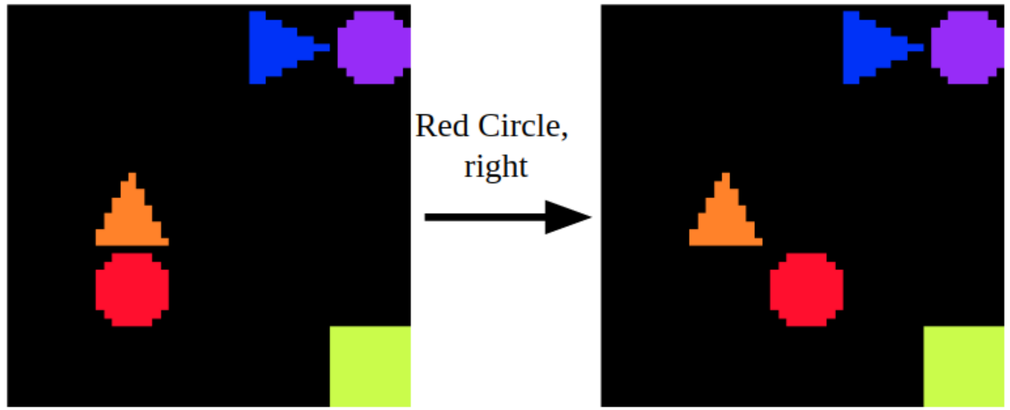
Summary 3

We investigate the binding issue and highlight the two places it appear, propose a method to solve it, and provide theoretical guarantees.

Outline

- Motivation
- Setup: Object Library
- Defining Compositional Generalization
- Implementing Compositional Generalization
- Solving Binding Issue In End-to-end World Modeling for CG
- Results

Environment (Block Pushing)



Environment: Block Pushing with Object Library

Random: locations, color, shapes

K = 5

N = 5,10,20,30

Experimental Setup

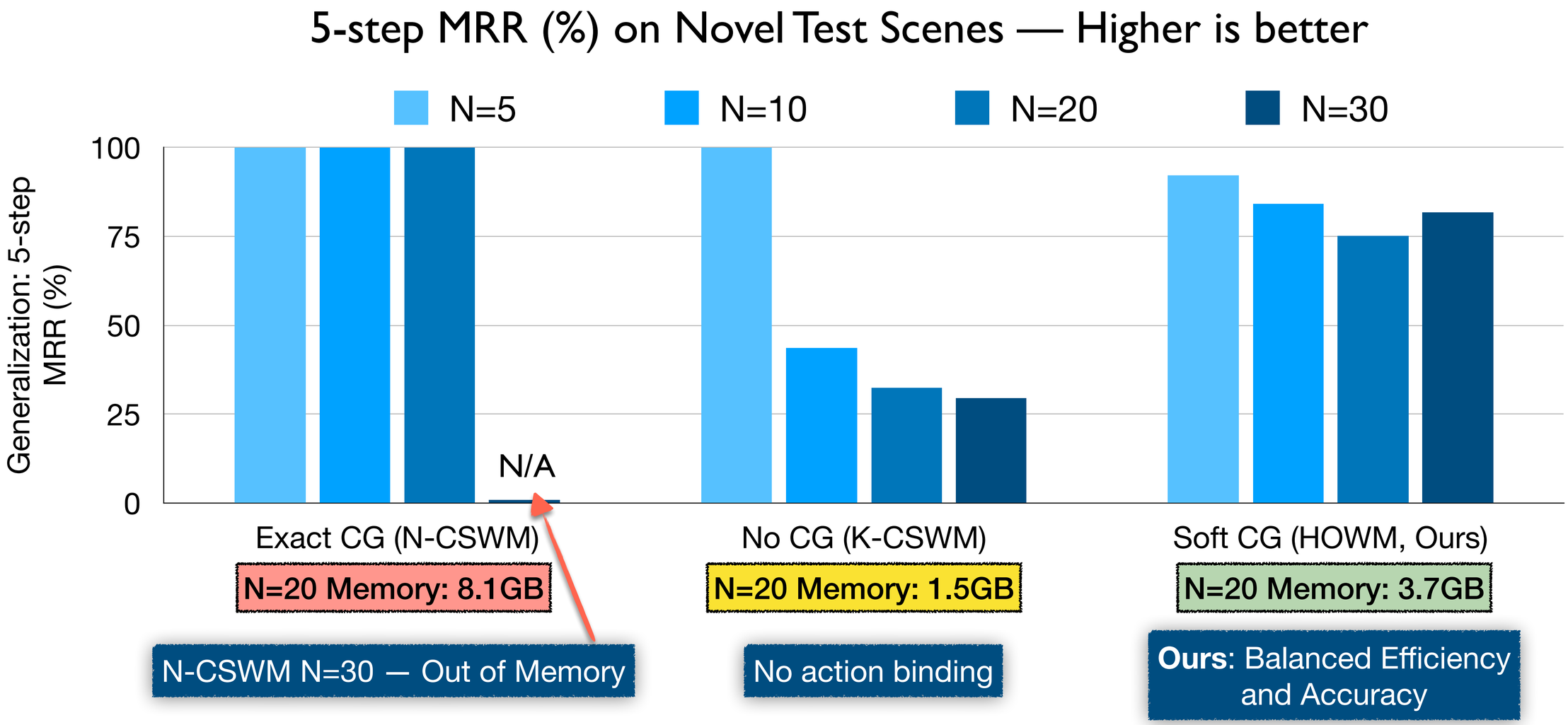
- **Exact CG**: N-slot WM
- **No CG**: break each component
- **Soft CG**: soft K-slot WM, **HOWM** (ours)

1. Slot Extraction

2. Action Binding to Slots

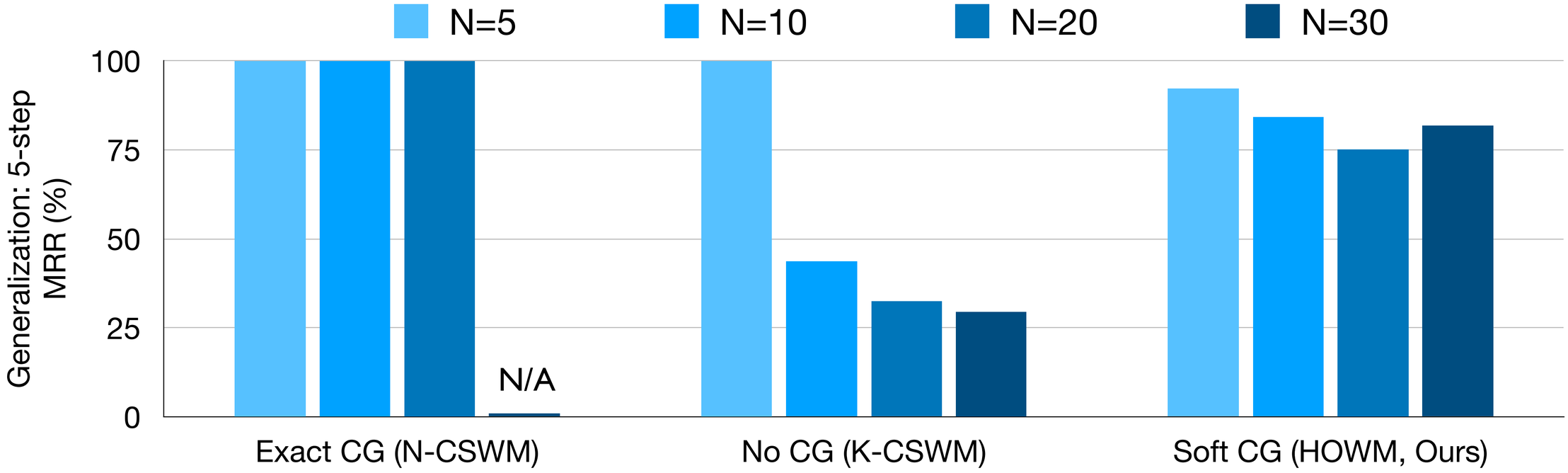
3. Σ_K-equivariant latent transition model

Quantitative Results



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

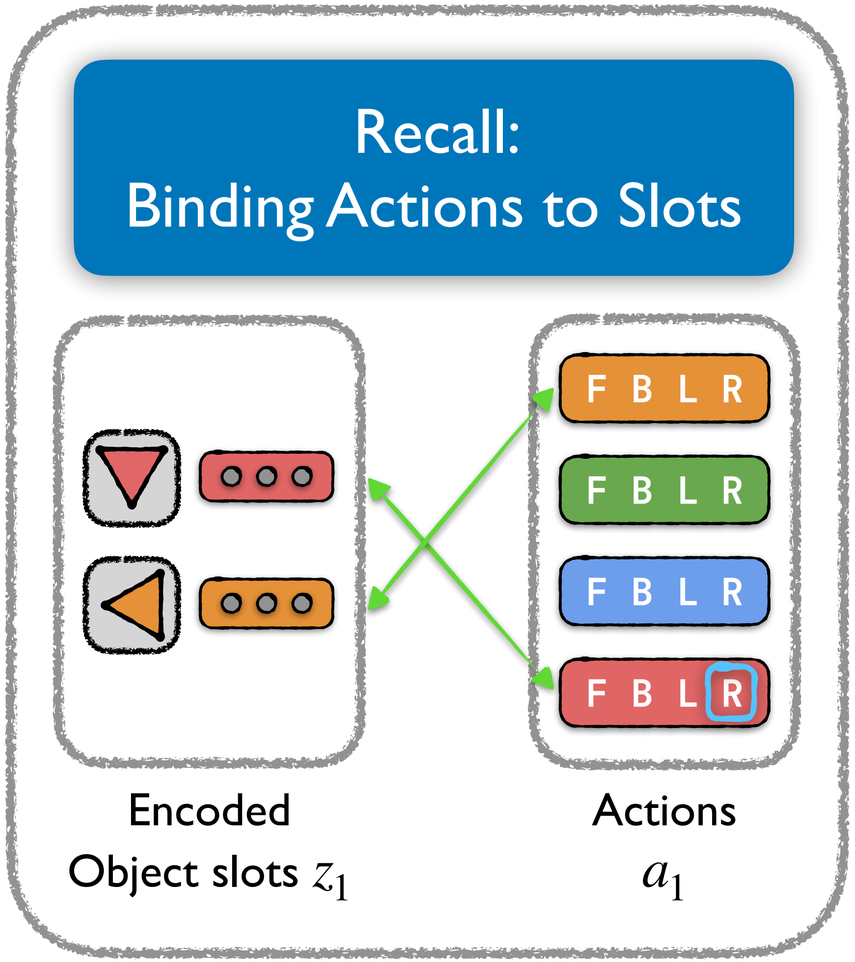
Quantitative Results

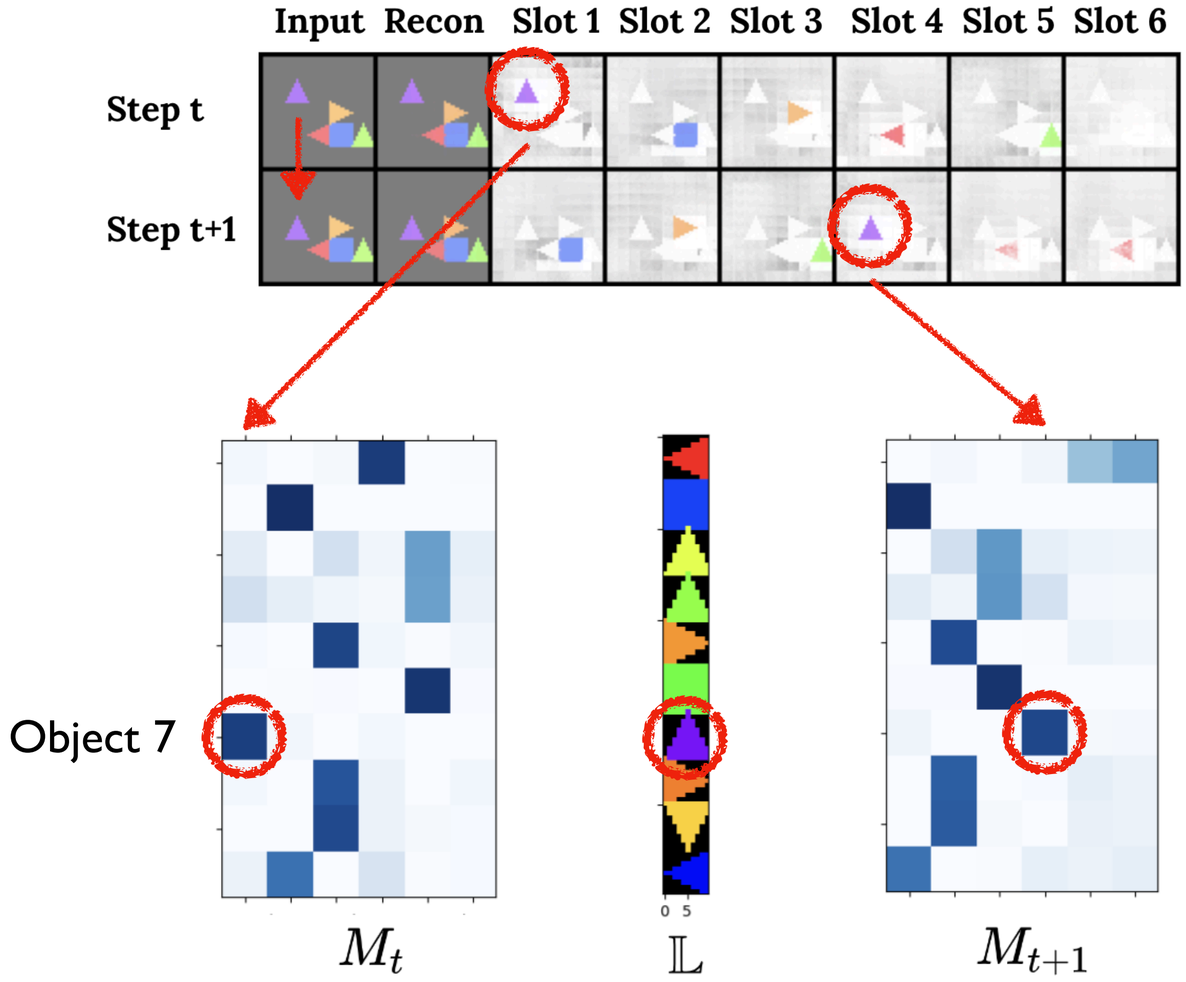


Takeaways of HOWM

It uses Action Attention with Aligned Loss to solve binding, so it can learn in the slot MDP for CG, and thus be more efficient

Binding Visualization





K=5 slots
5+1 rows
(+ 1 background)

N=10 objects
10 columns

Found object identity through actions (unknown identity)

Talk Summary

Study compositional generalization in object-oriented world modeling

- How we use “Object Library” to formulate compositional generalization
- How the formulation motivates our WM method and provides guarantees
- Understanding the central issue on end-to-end WM: binding problem
- Proposed Action Attention + Aligned Loss to solve CG in WM
- Representative results on our method, an oracle, and no-CG methods

More Information

Check out our project website:

http://lfzhao.com/oowm/

zhao.linf@northeastern.edu

Poster #412