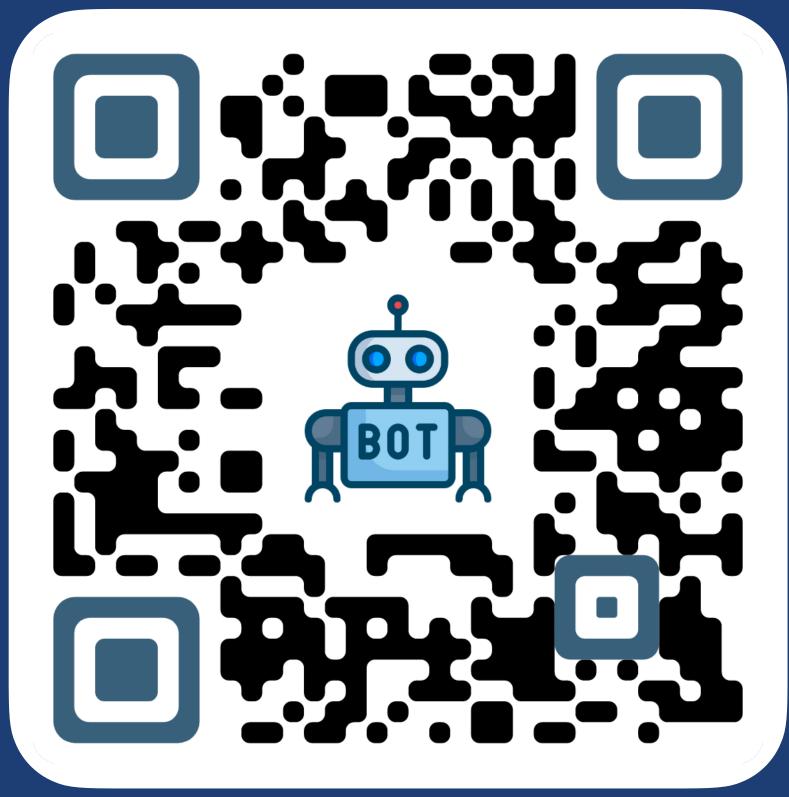
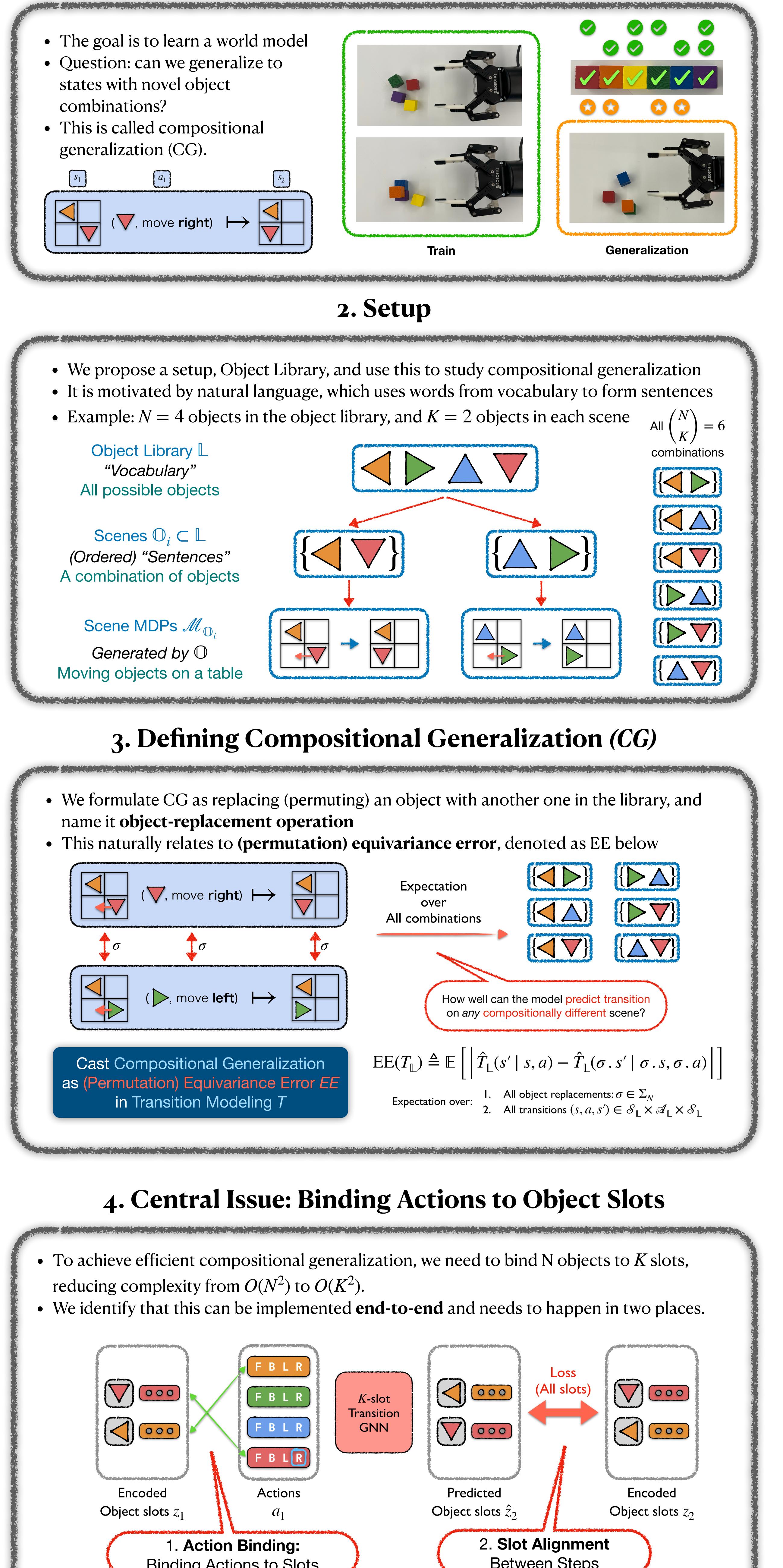
Compositional generalization in end-to-end world modeling requires correctly binding actions to object S O LS. We show Action Attention with Aligned Loss can provably achieve



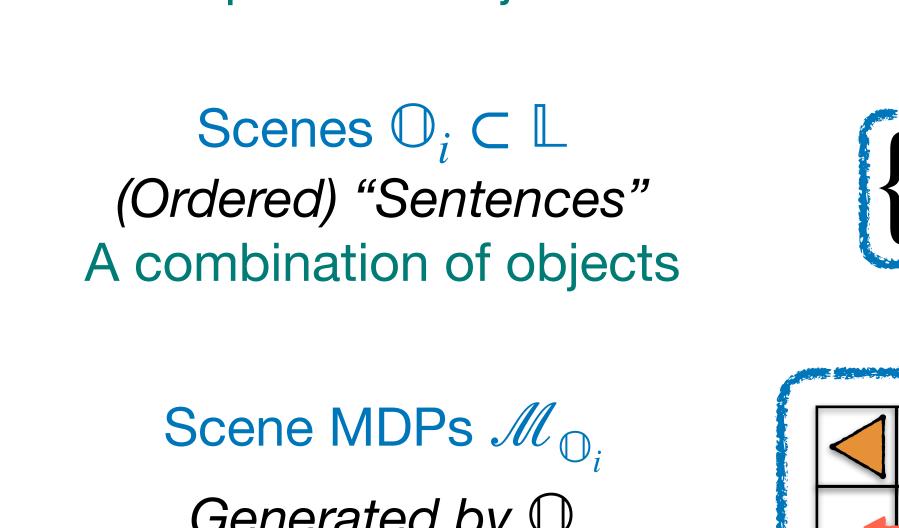


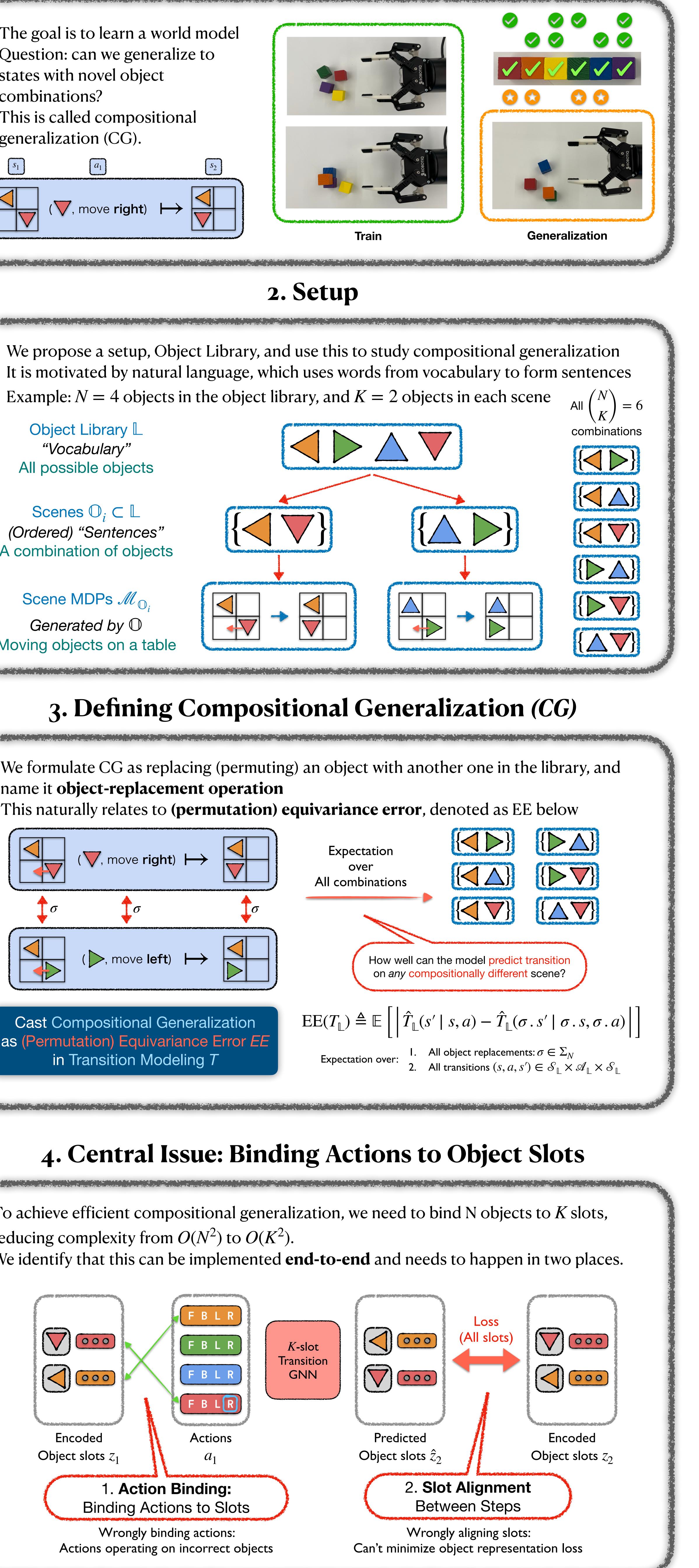
Toward Compositional Generalization in Object-oriented World Modeling Linfeng Zhao, Lingzhi Kong, Robin Walters, Lawson L.S. Wong Khoury College of Computer Sciences, Northeastern University

1. Motivation



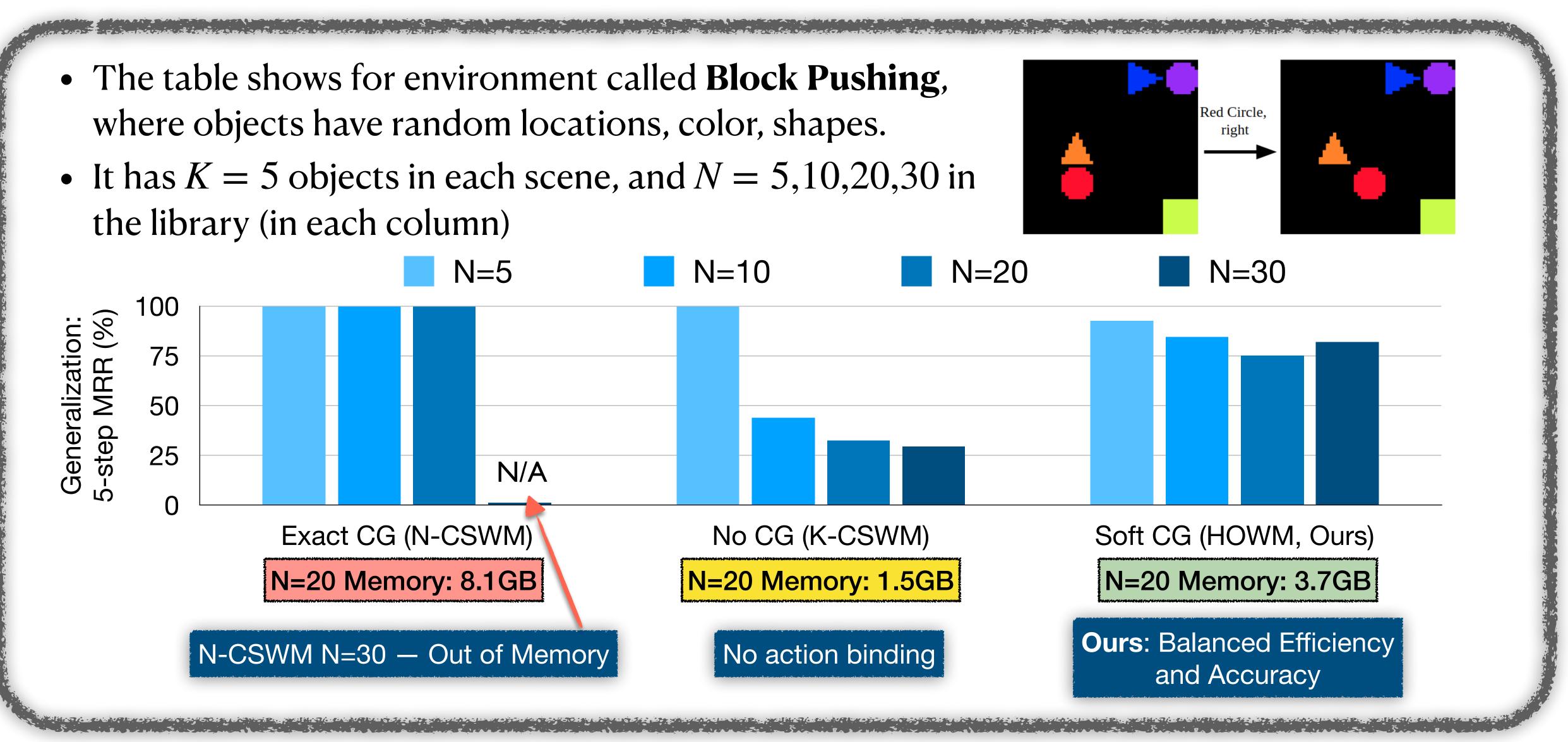
elenanderena





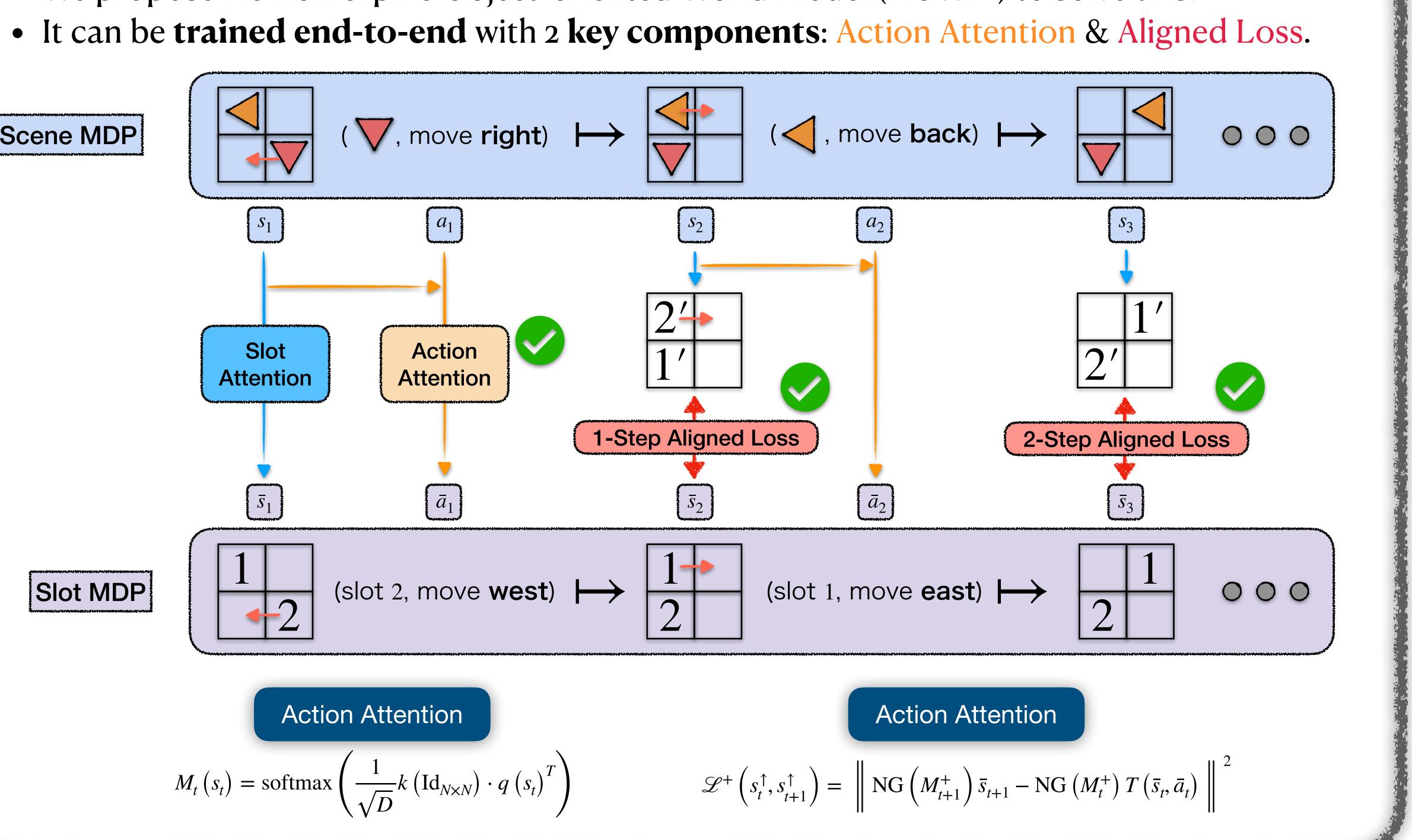


Scene MDP Slot MDP



- The action is to move the purple triangle to south (while other
- objects are fixed). • The action is a fixed interaction protocol. The environment and the model assume the same order for all scenes, but unknown to model.
- Our HOWM learns the order for step t and t + 1, thus can align then into a canonical order.

5. Our Method: Homomorphic Object-oriented World Model • We propose Homomorphic Object-oriented World Model (HOWM) to solve this.



6. Theoretical Guarantees

• We formally show when it is possible to use the slot MDP to achieve compositional generalization: zero equivariance error, as long as actions correctly bind to object slots. Theorem (informal): $\cdot \text{EE}(T_{lib})$ $EE(T_{slot}) =$ If actions correctly bind to object slots, the equivariance error is related by: EE in Library MDF EE in Slot MDP N-object K-slot Intuition: Binding = slot MDP can Bind correctly for all scenes correctly simulate any scene MDP

7. Representative Results

8. Visualization & Takeaways

• The visualization shows the binding matrix learned by Action Attention, purely from interaction (purple triangle moves when the action operates that slot) • In summary, our method (HOWM) uses Action Attention with Aligned Loss to solve the binding problem, so it can learn in the slot MDP end-to-end, and thus be more efficient.

