Modularity in Grounded Interaction

Daniel Fried and Rudy Corona


Trevor Darrell


Dan Klein


Coline Devin


Anjali Kantharuban

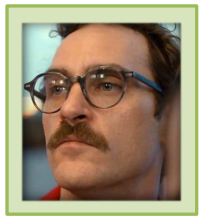

Justin Chiu

Interactive Systems

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Science Fiction…

Her, 2013



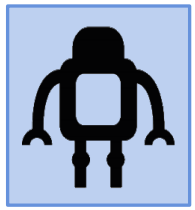
Let's start with your emails. You have several thousand emails regarding LA Weekly, but it looks like you haven't worked there in many years.

Oh yeah, I guess I was saving those because in some of them I thought I might have written some funny stuff.

Yeah, there are some funny ones. I'd say there are about 86 that we should save. We can delete the rest.

…and Reality

A Massive Neural Chatbot, 2020



Hi!

Hey there! What's up?

Same old, Monday blues… what are you up to?

I'm procrastinating on a paper I'm supposed to be writing.

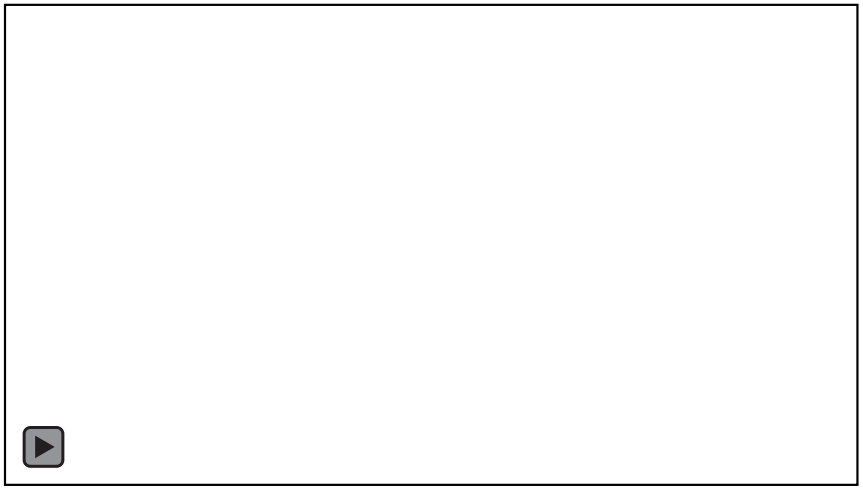
Are you thinking of pulling an all nighter?

Possibly. What's your paper on?

Grounding Systems

Neural grounding is effective…

Anderson et al. 2018, *inter alia.*

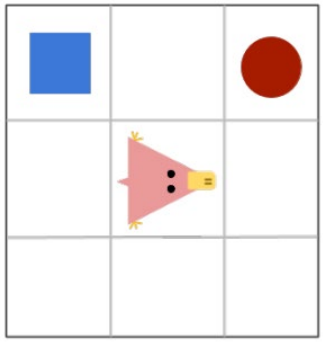


Turn left and take a right at the table. Take a left at the painting and then take your first right.

…but often not robust

Ruis et al. 2020. (See also Lake and Baroni 2018; Hill et al. 2020; Bahdanau et al. 2019 & 2020)

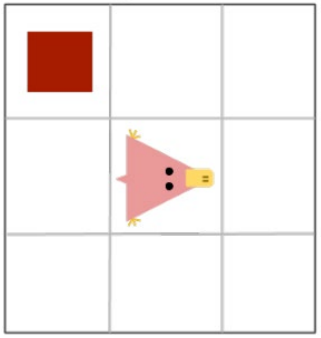
TRAIN



*"Walk to the **blue square**."* ✓

*"Walk to the **red circle**."*

TEST



*"Walk to the **red square**."* ✗

Better Systems Through Modularity

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

CS's main tool for building robust, understandable systems for complex tasks: break them down!

```
tr -sc 'A-Za-z' '\n' < shakes.txt
      | sort
      | uniq –c
```

Change all non-alpha to newlines

Sort in alphabetical order

Merge and count each type

```
1945 A          25 Aaron
  72 AARON       6 Abate
  19 ABBESS      1 Abates
   5 ABBOT       5 Abbess
 ... ...         6 Abbey
                 3 Abbot
                .... ...
```

(Example from Dan Jurafsky)

Brains vs. Blocks

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Neural Nets

Expressive, ground language to fuzzy world contexts

Modular Systems

Adaptable, understandable, compartmentalized

Modular Nets

All of these! (Hopefully)

Not a New Idea!

Figure 1. Conflict between theoretical extremes.

"To program today, we must describe things very carefully... But once we have modules that know how to learn, we won't have to specify nearly so much—**and we'll program on a grander scale, relying on learning to fill in details….**"

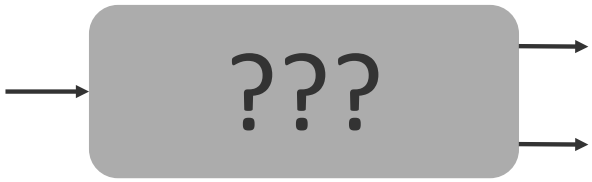
[Minsky, 1991. *Symbolic vs Connectionist or Neat vs Scruffy*]

Why Modularity?

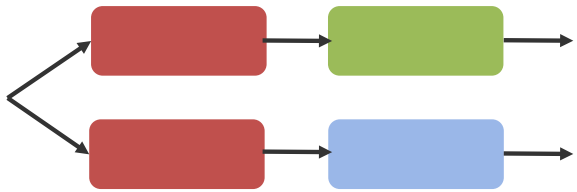
BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Systems

Monolithic

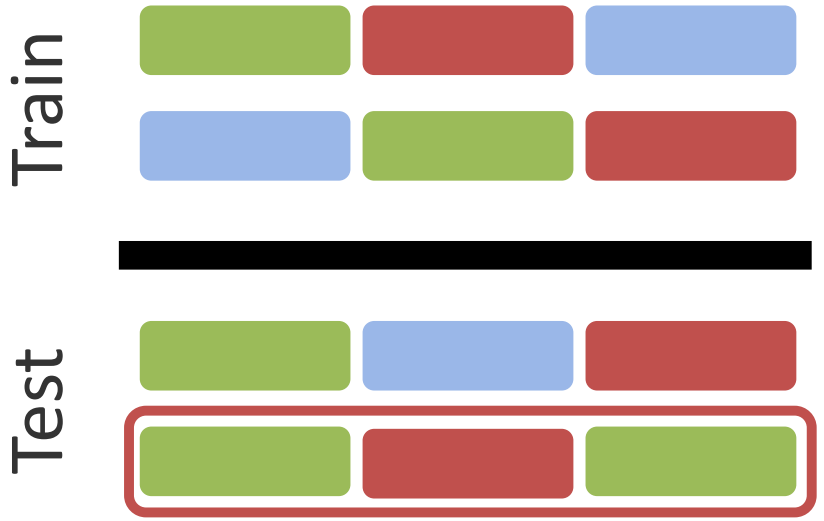


Modular



Modular models give more footholds for development

Systematicity



Modular models are often better at compositional generalization

Other NAACL Work on Systematicity!

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Latent Compositional Representations Improve Systematic Generalization in Grounded Question Answering

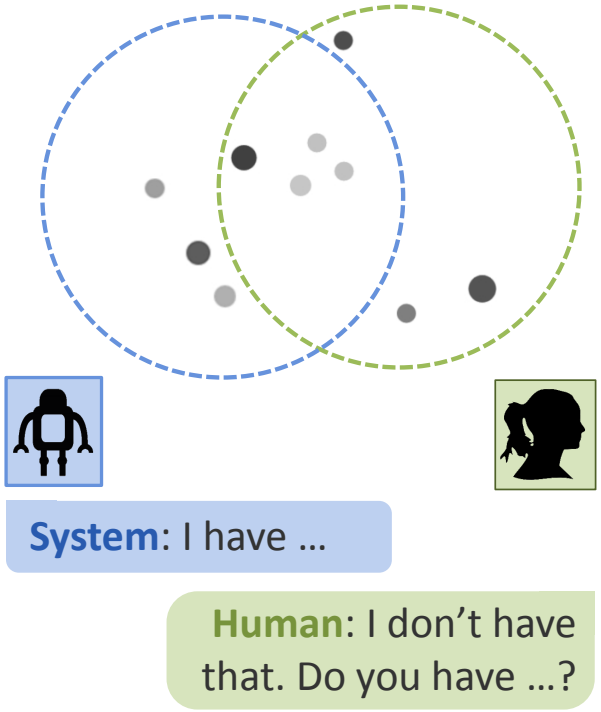
Ben Bogin[1] **Sanjay Subramanian**[2] **Matt Gardner**[2] **Jonathan Berant**[1,2]
[1]Tel-Aviv University [2]Allen Institute for AI

Compositional Generalization for Neural Semantic Parsing via Span-level Supervised Attention

Pengcheng Yin[♠,*] **Hao Fang**[♣], **Graham Neubig**[♠], **Adam Pauls**[♣],
Emmanouil Antonios Platanios[♣], **Yu Su**[♣], **Sam Thomson**[♣], **Jacob Andreas**[♣]
[♠]Carnegie Mellon University [♣]Microsoft Semantic Machines

Modularity in…

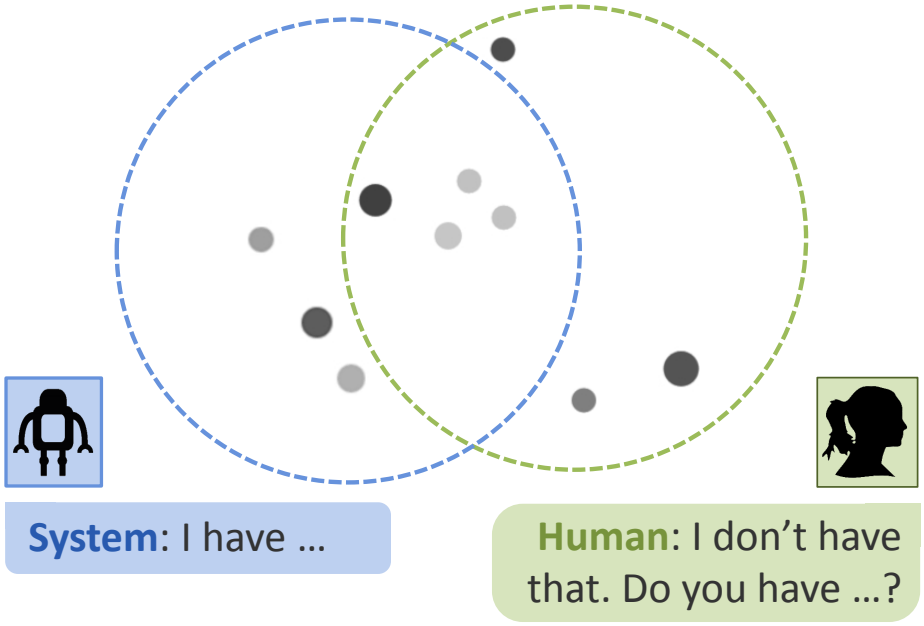
Dialogue



Instruction Following

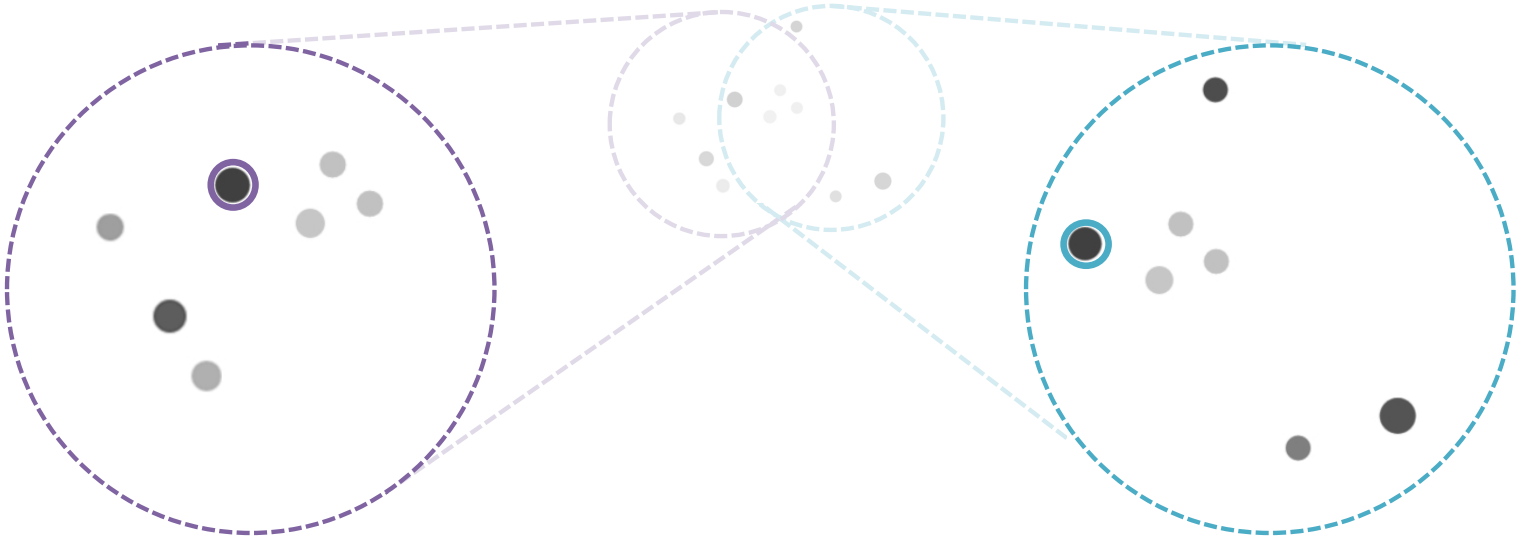
"Pick up the clock. Walk to the lamp. Then turn it on."

Modularity in Grounded Dialogue



[Fried, Chiu, and Klein. In submission]

Grounded Collaborative Dialogue



A: I have three dots in a line with a dark one in the center.

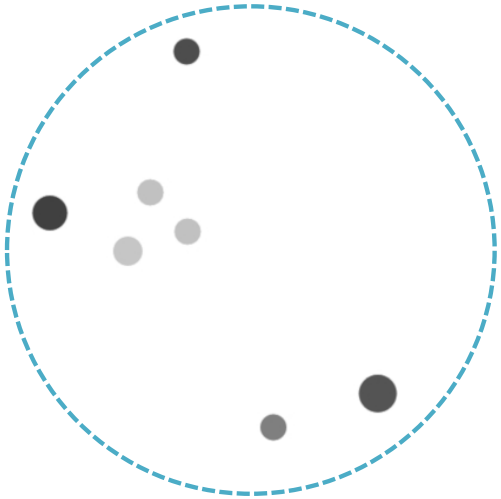
A: Is there a large black dot to the left of the three grey dots?

B: I don't have that. Do you have a cluster of three grey dots in a triangle?

B: Yes, let's select the black one.

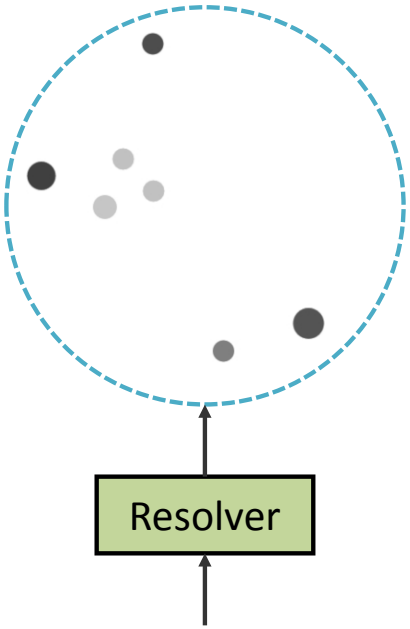
[Udagawa and Aizawa, 2019 & 2020]

Decomposing Into Subtasks



A: I have three dots in a line with a dark one in the center. → **B**: I don't have that. Do you have a group of three grey dots? → **A**: Is there a large black dot to the left of the three grey dots? → **B:???**

Decomposing Into Subtasks

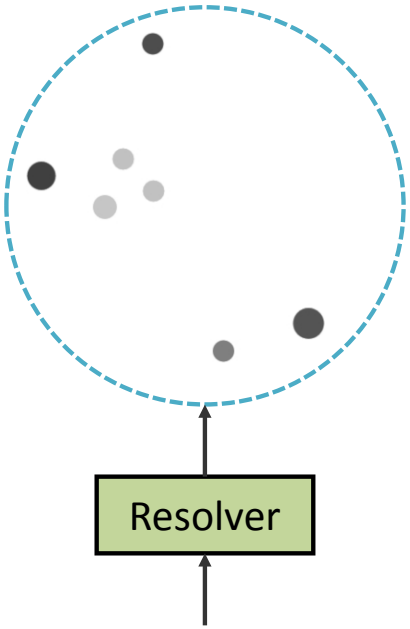


don't have that. Do
u have a group of
three grey dots?

A: Is there a large black dot to the left of the three grey dots?

B:???

Decomposing Into Subtasks



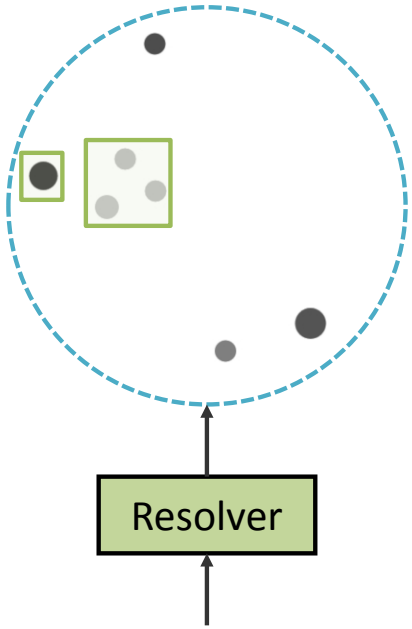
don't have that. Do
u have a group of
three grey dots?

A: Is there a large black dot to the left of the three grey dots?

B:???

Decomposing Into Subtasks

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

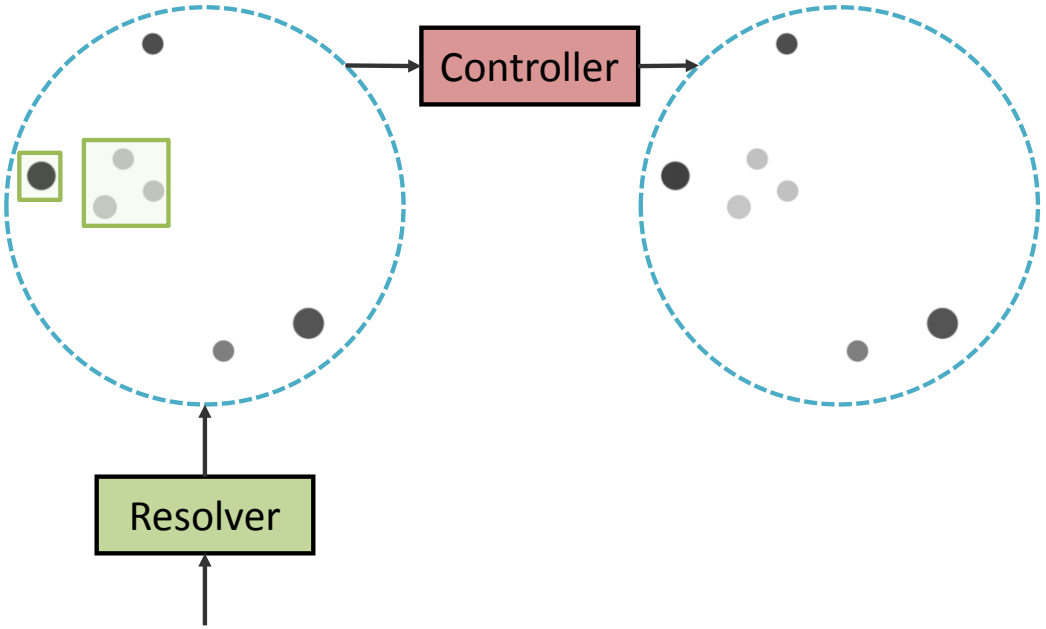


don't have that. Do
u have a group of
three grey dots?

A: Is there a large black dot to the left of the three grey dots?

B:???

Decomposing Into Subtasks

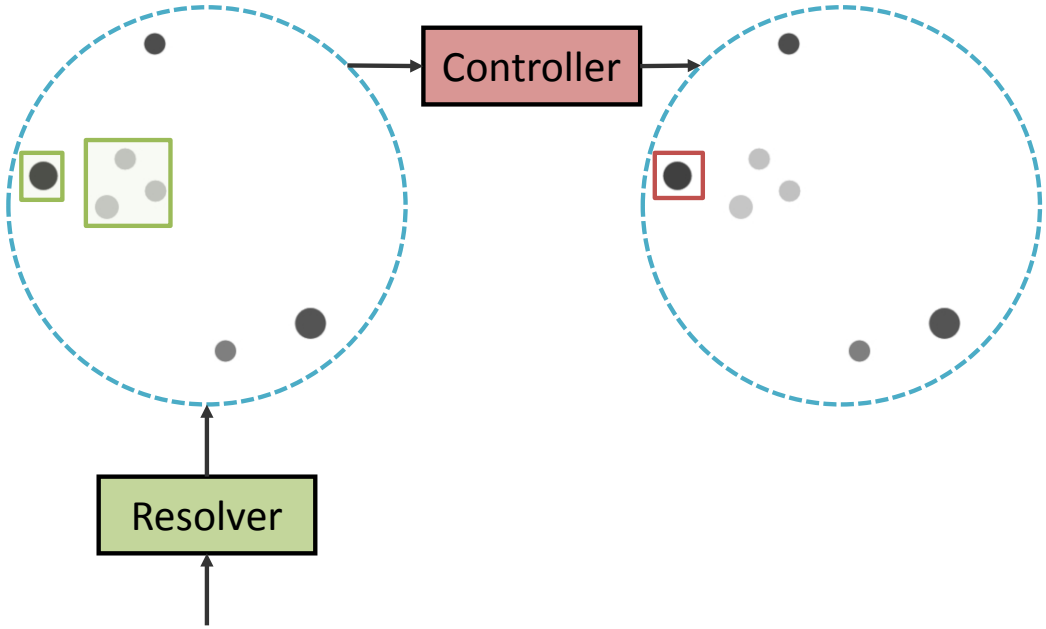


don't have that. Do
u have a group of
three grey dots?

A: Is there a large black dot to the left of the three grey dots?

B:???

Decomposing Into Subtasks

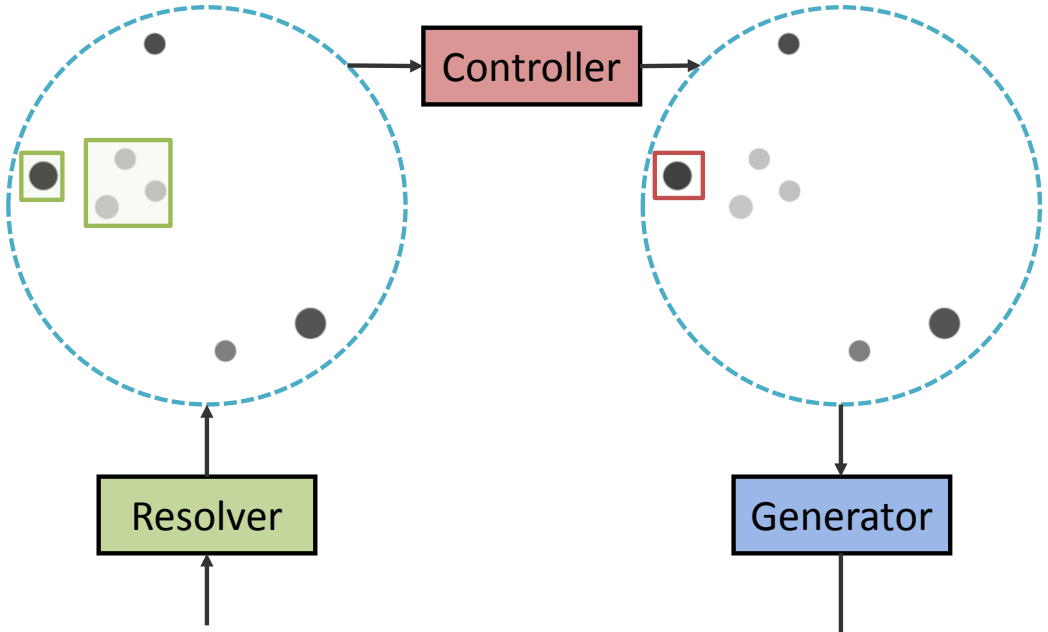


don't have that. Do
u have a group of
three grey dots?

A: Is there a large black dot to the left of the three grey dots?

B:???

Decomposing Into Subtasks

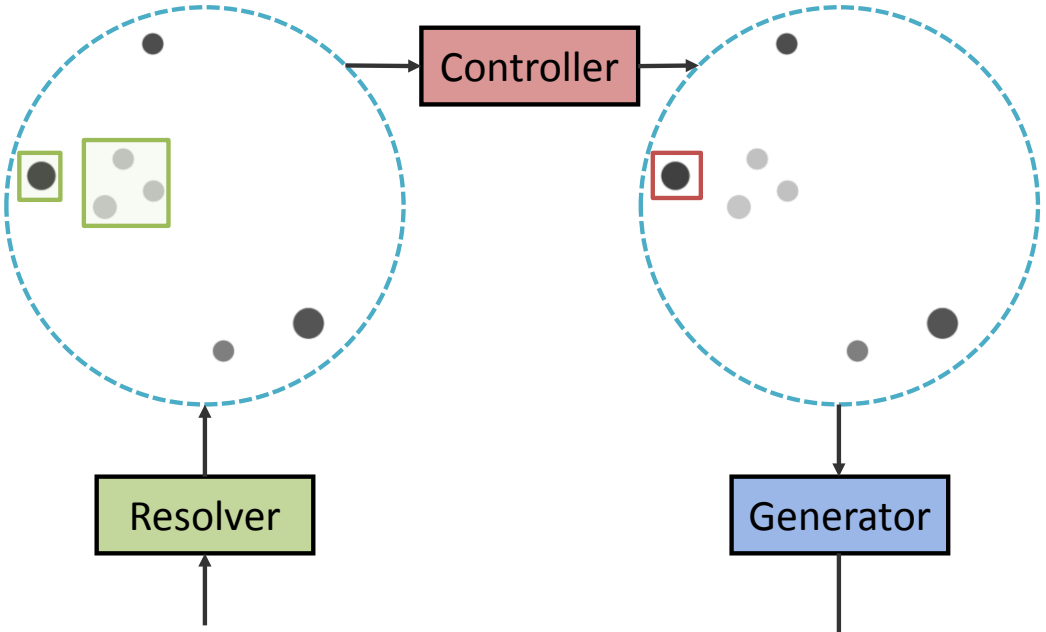


don't have that. Do
u have a group of
three grey dots?

A: Is there a large black dot to the left of the three grey dots?

B:Yes, let's select the black one.

Decomposing Into Subtasks

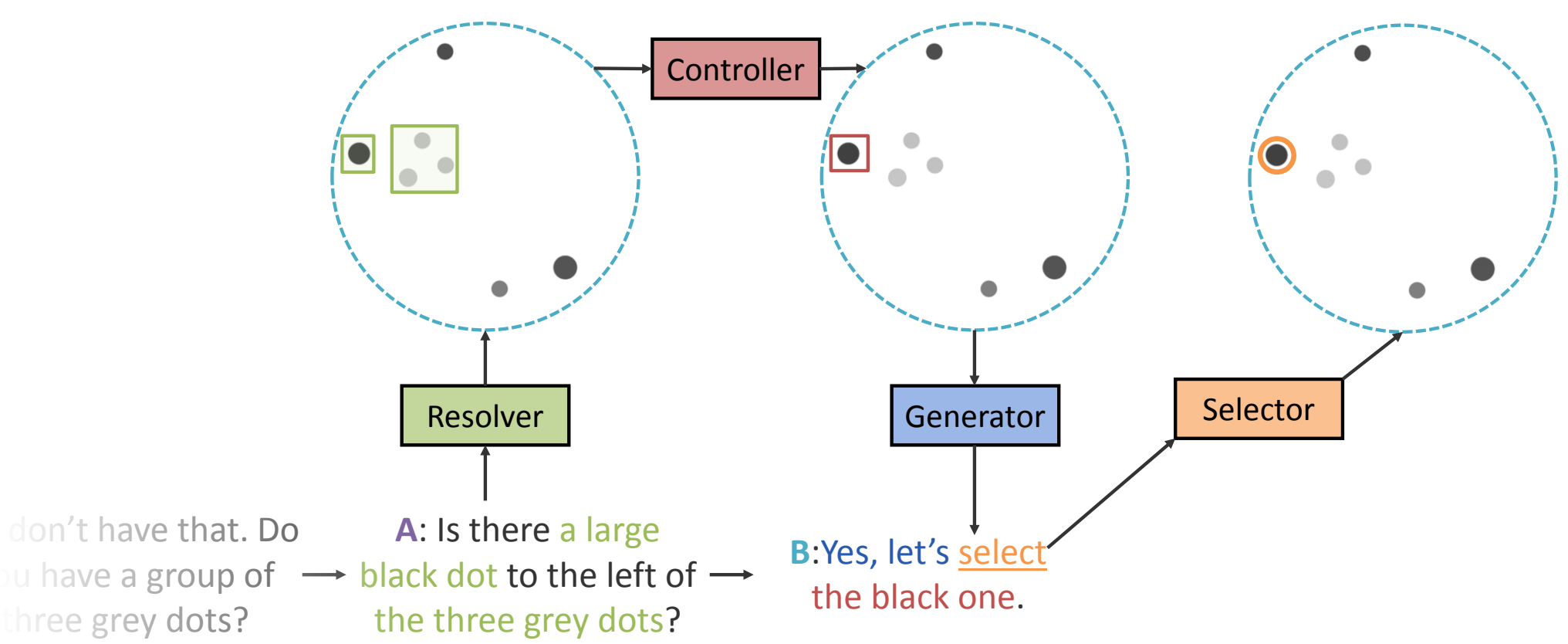


don't have that. Do you have a group of three grey dots?

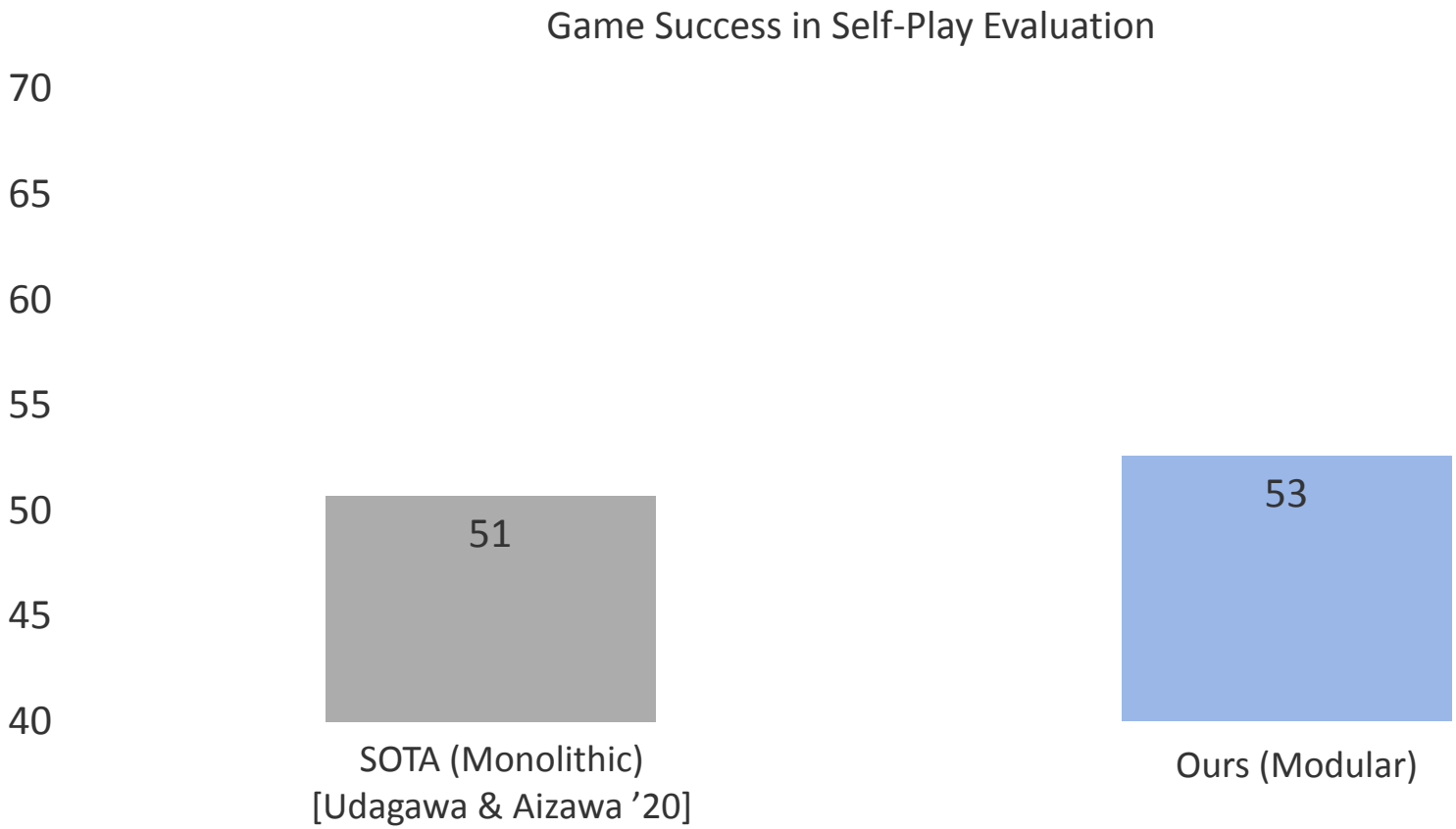
A: Is there a large black dot to the left of the three grey dots?

B:Yes, let's select the black one.

Decomposing Into Subtasks

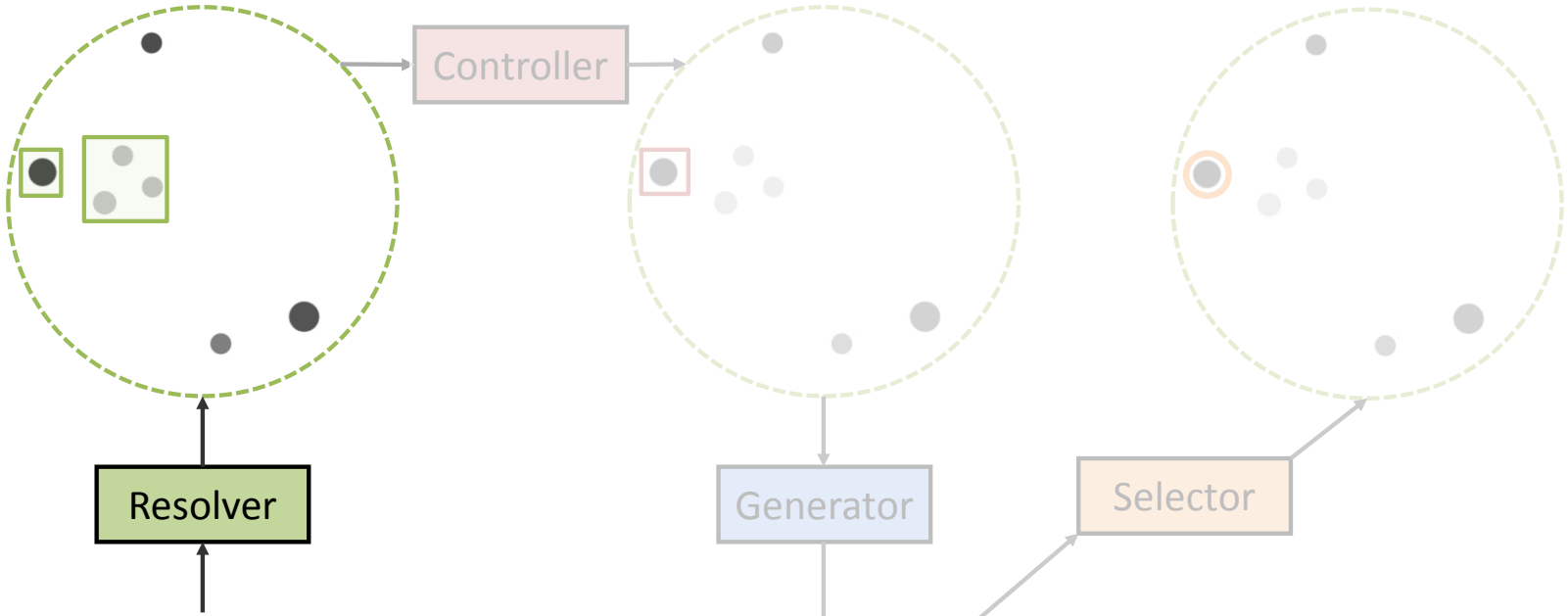


Effects of Modularity



But making the system modular is just the foundation! Now, we can iteratively improve it.

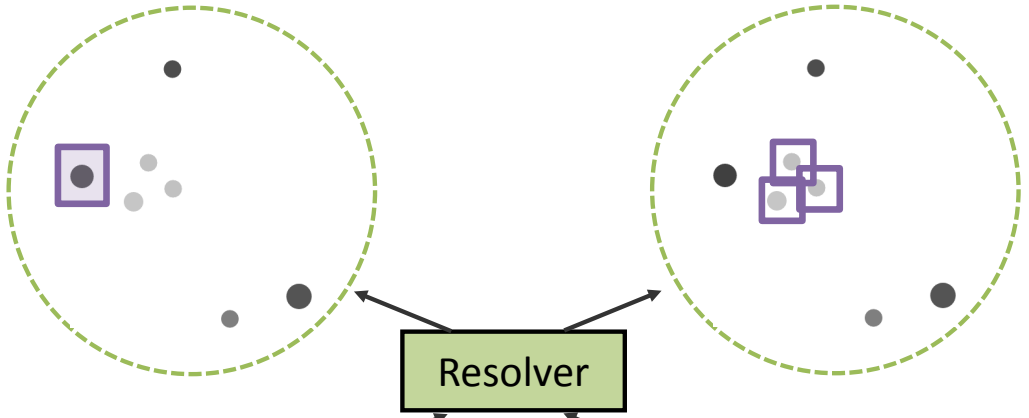
Focusing on the Resolver



Focusing on the Resolver

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

RelationNet [Santoro et al. 2017, Udagawa and Aizawa 2020]

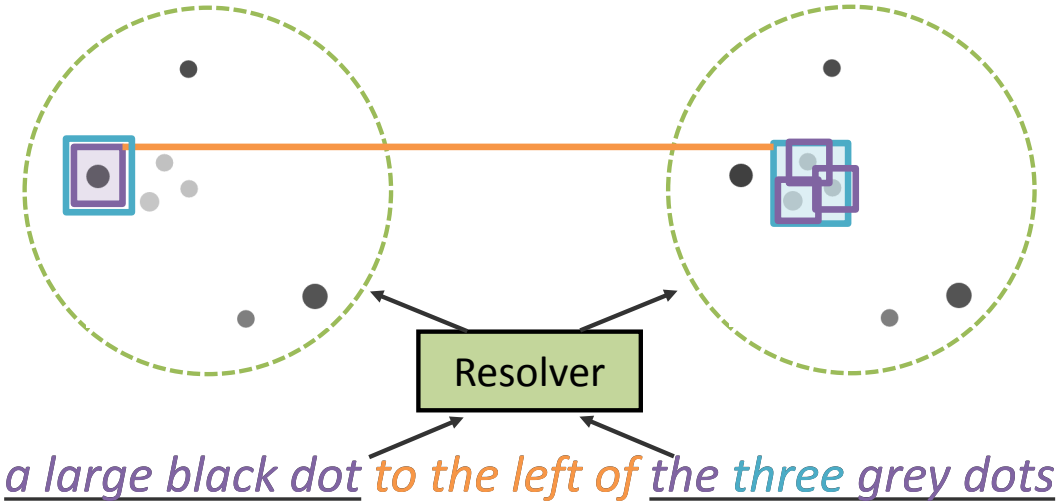


Is there a large black dot to the left of the three grey dots?

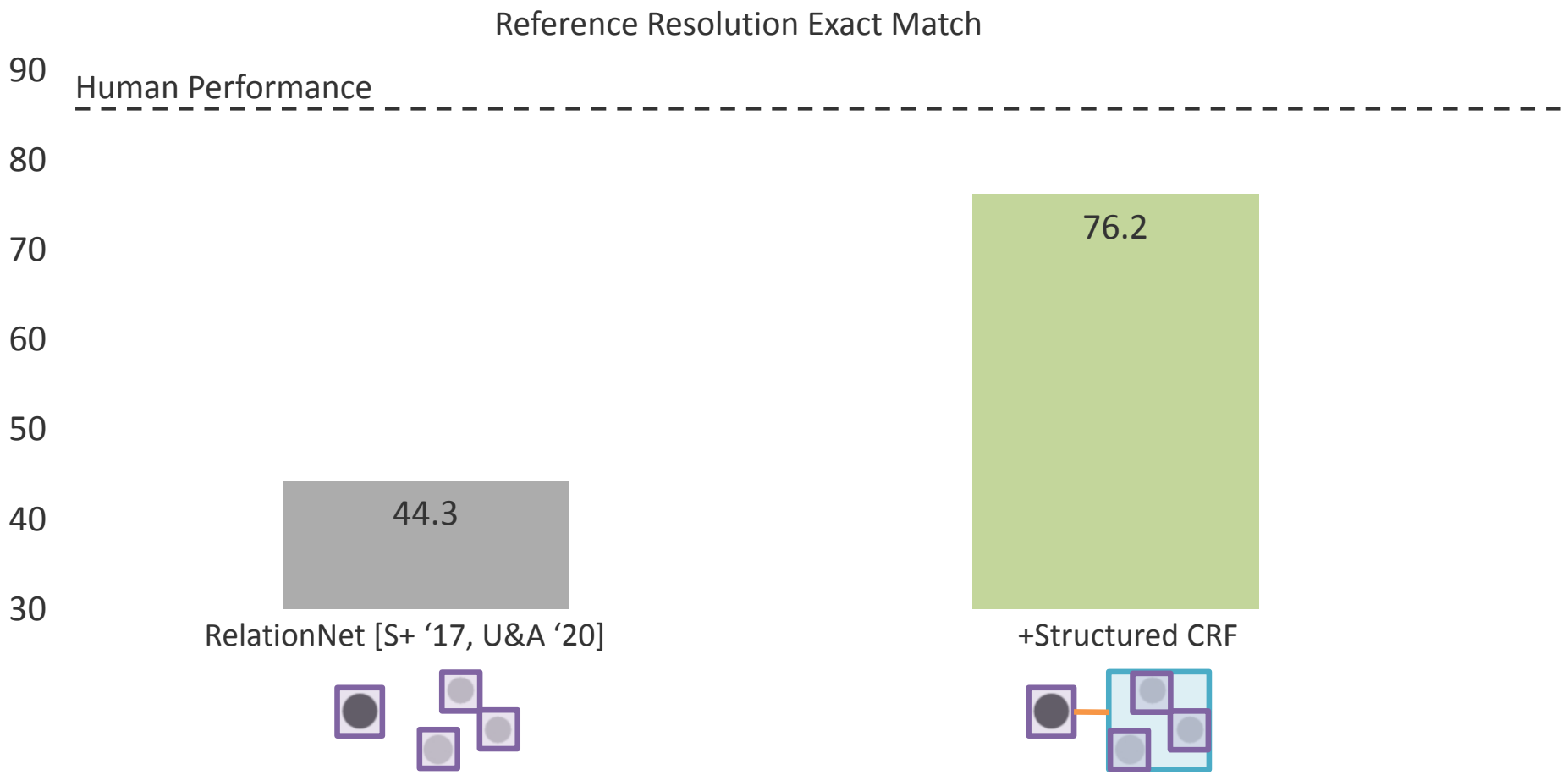
Focusing on the Resolver

RelationNet [Santoro et al. 2017, Udagawa and Aizawa 2020]

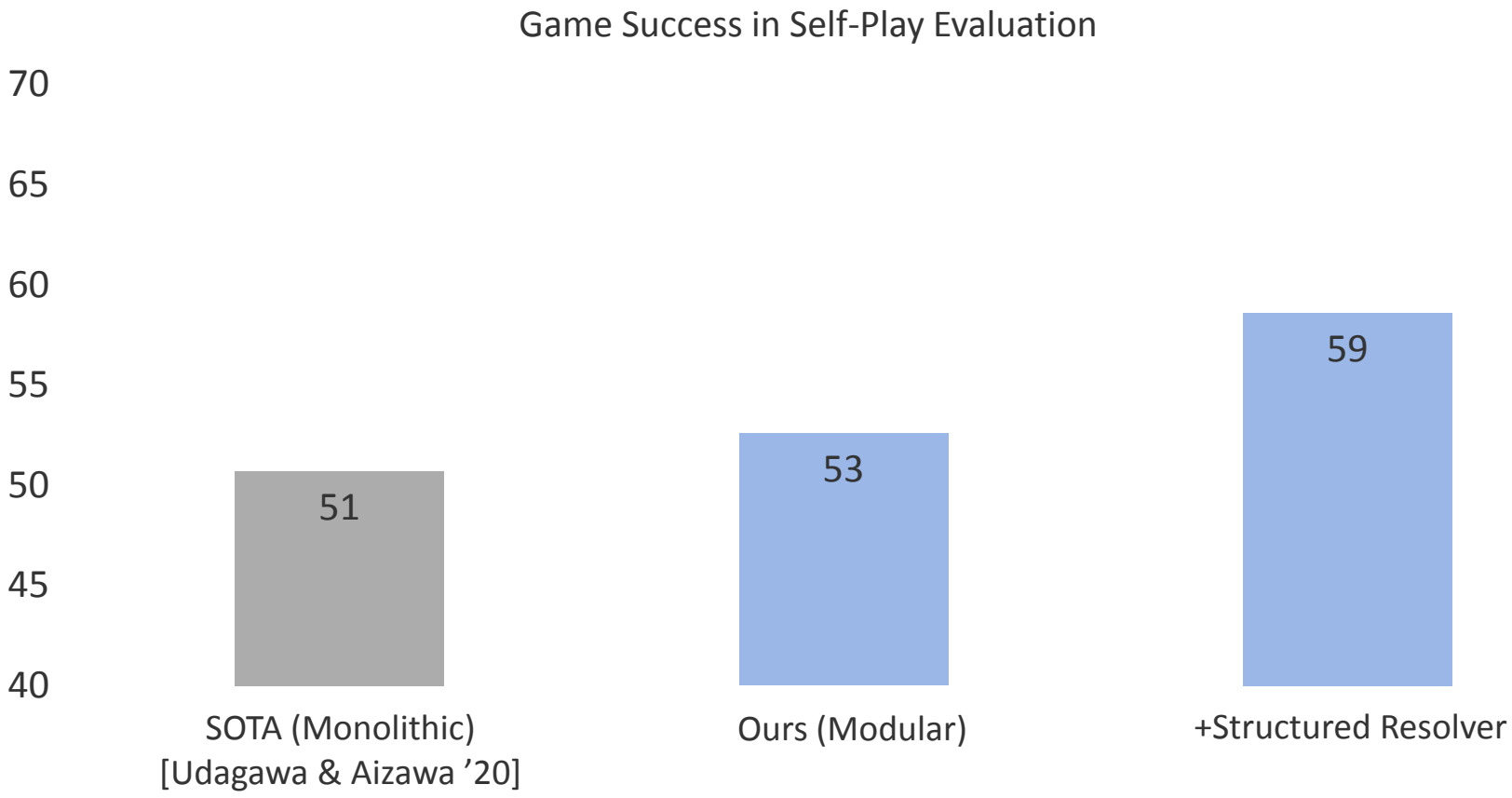
+ Structured Conditional Random Field (model groups and relations in the output structure)



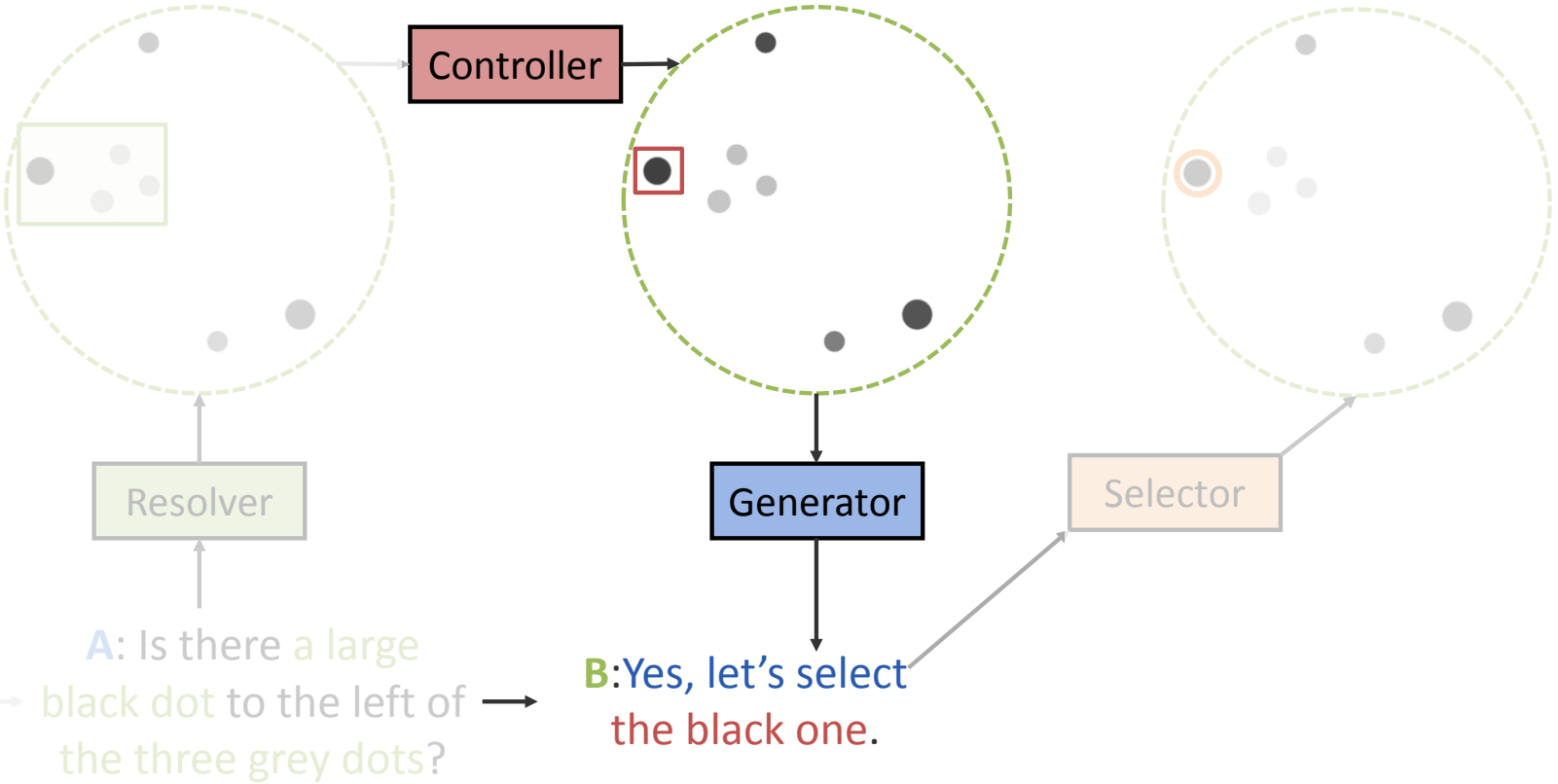
Reference Resolution Results



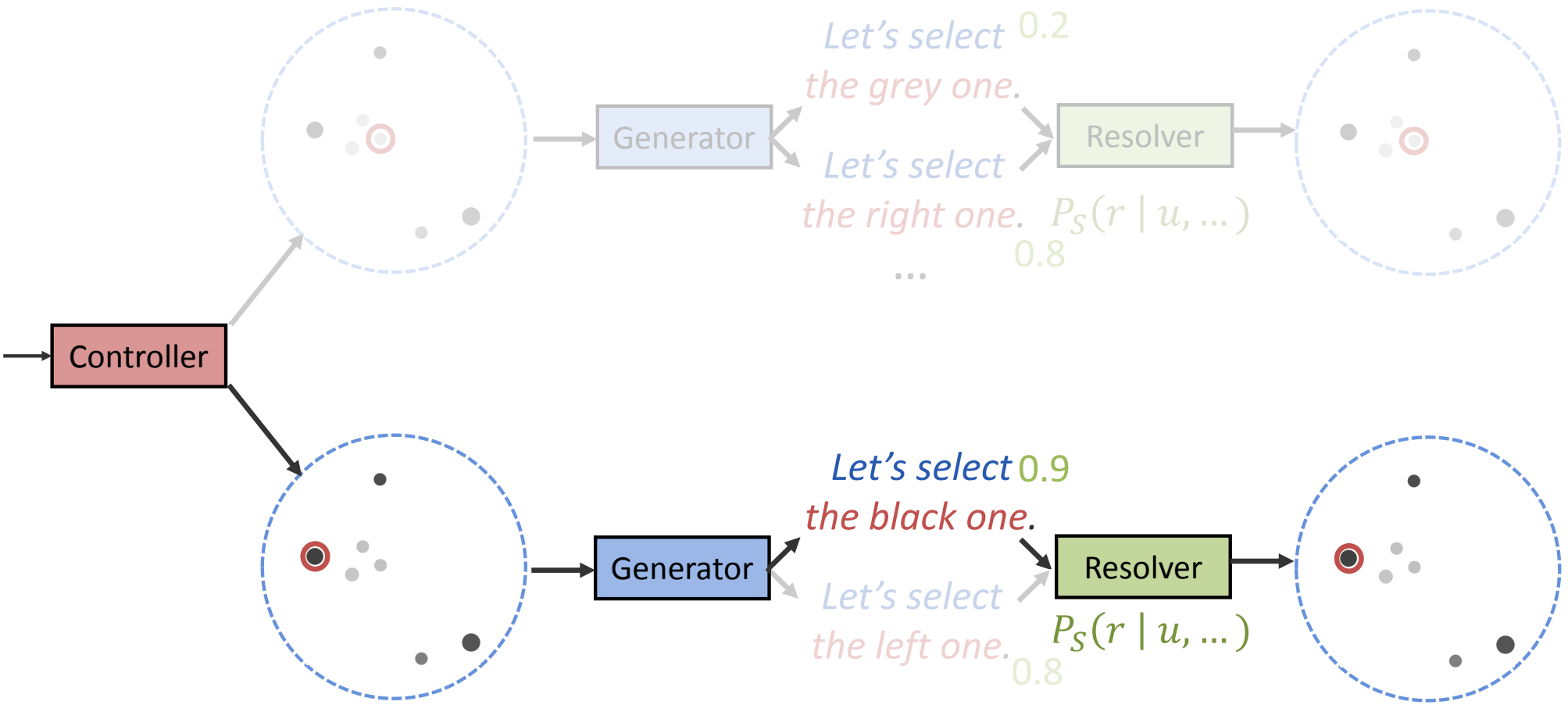
Automatic Evaluation Results



Focusing on Generation

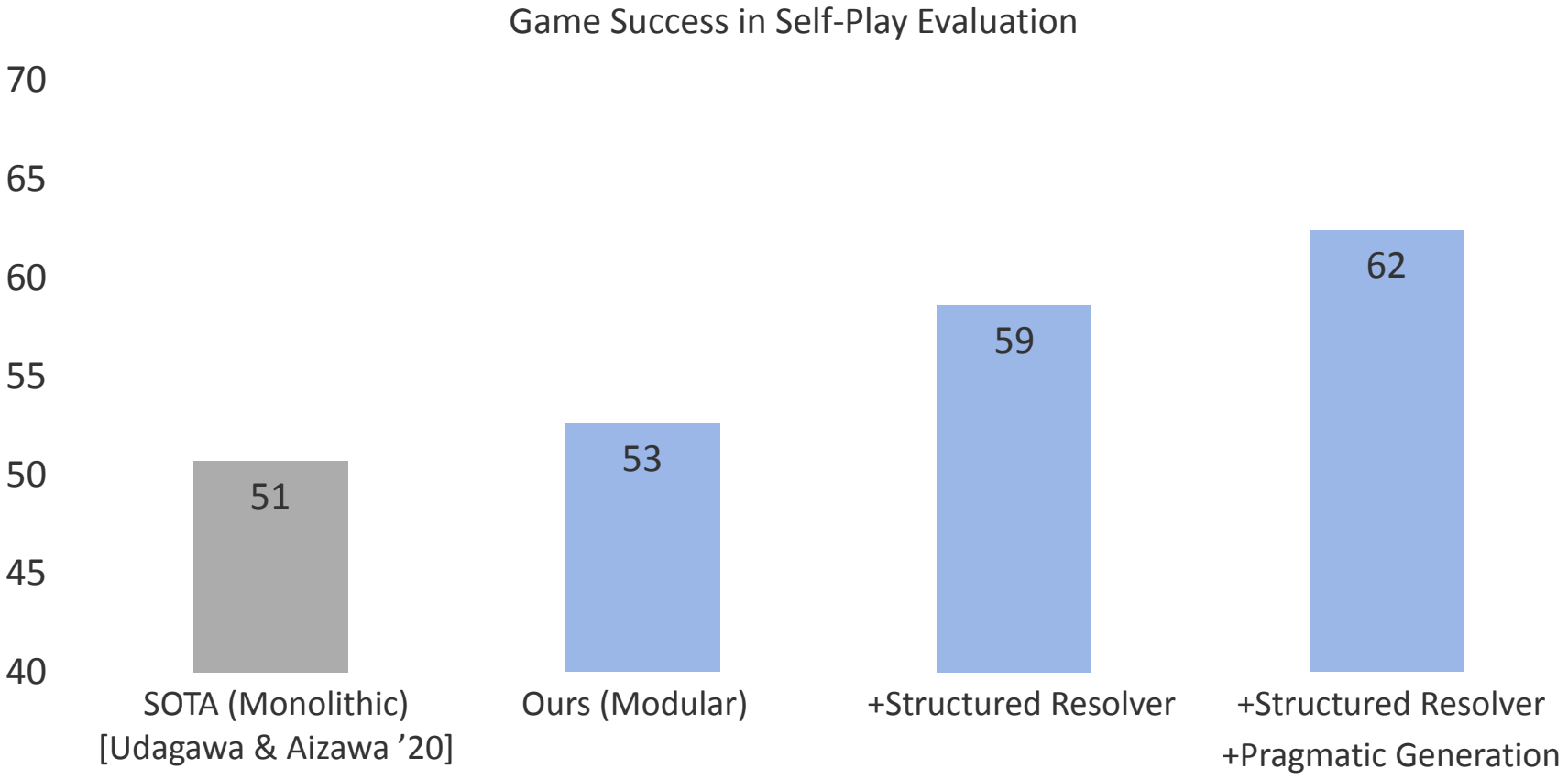


Module Reuse: Pragmatic Generation

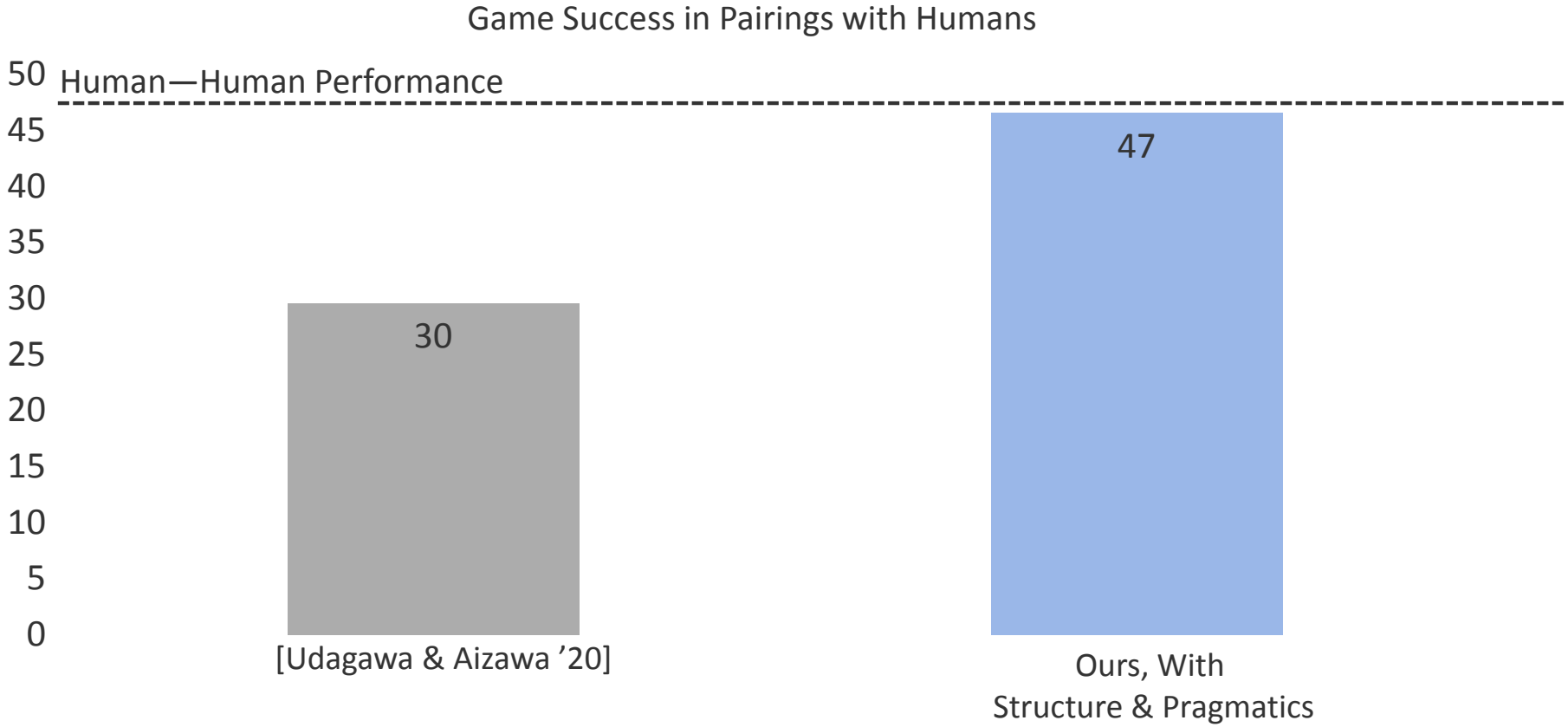


Inspired by Rational Speech Acts [Frank & Goodman 2012]

Automatic Evaluation Results

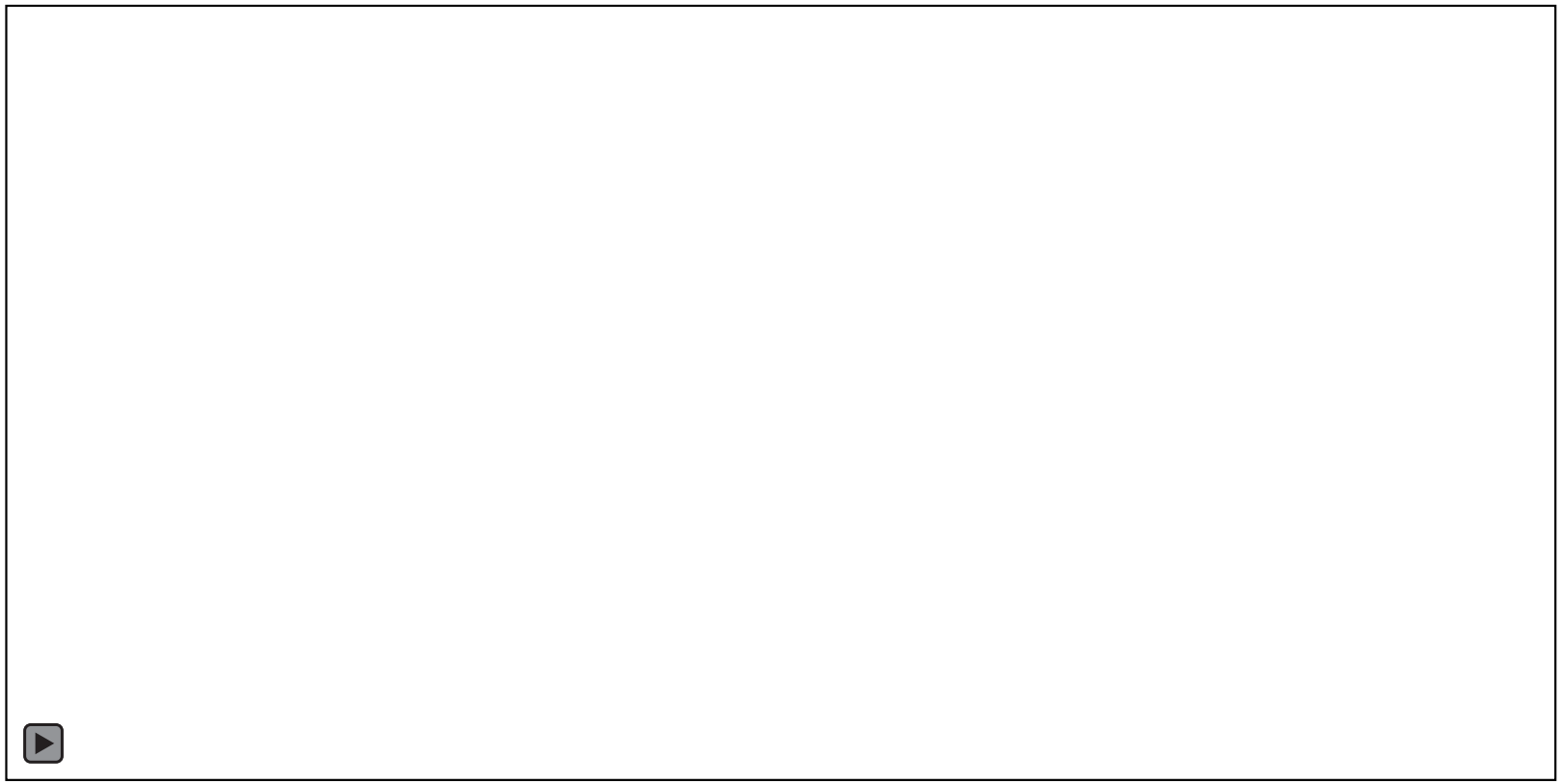


Human Evaluation Results



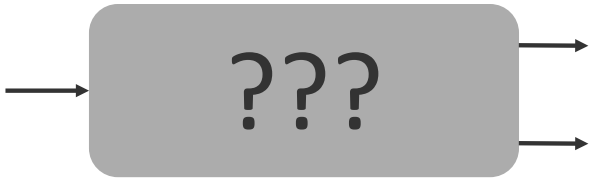
Demo

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

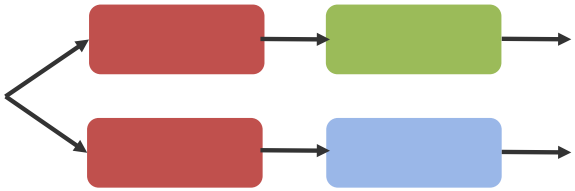


Takeaway

Monolithic



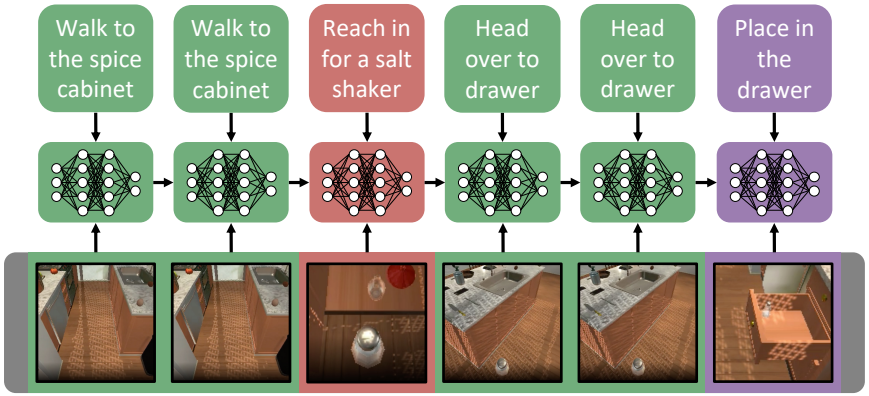
Modular



Decomposing a complex task into subtasks allows iterating on and repurposing the subtask components.

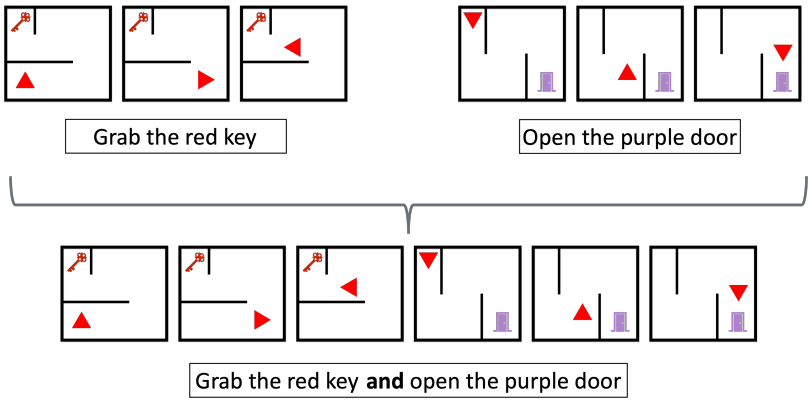
Modularity in Instruction Following

Modularize the Model



[Corona et al. 2021]

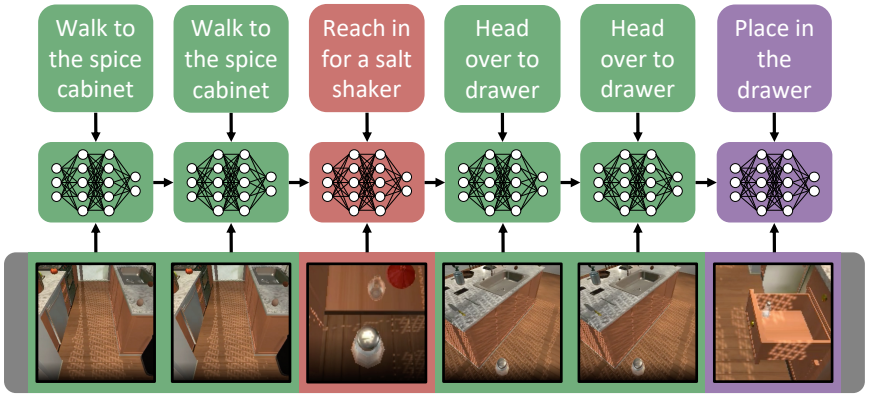
Modularize the Data



[Kantharuban et al. In submission]

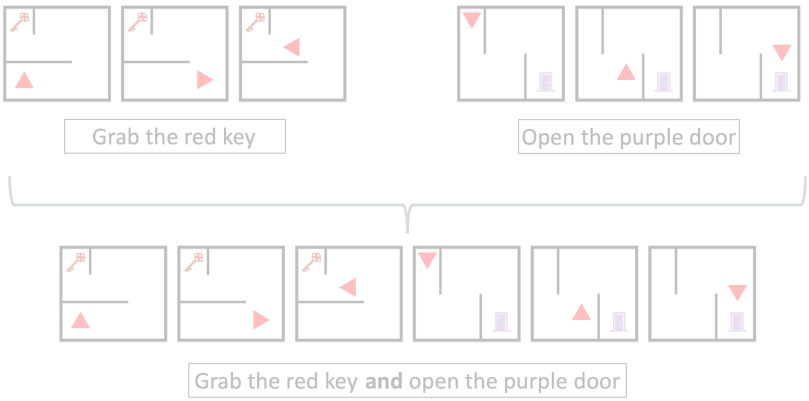
Modularity in Instruction Following

Modularize the Model



[Corona et al. 2021]

Modularize the Data



[Kantharuban et al. In submission]

Compositional Generalization

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Pickup a pen and put it in a box.

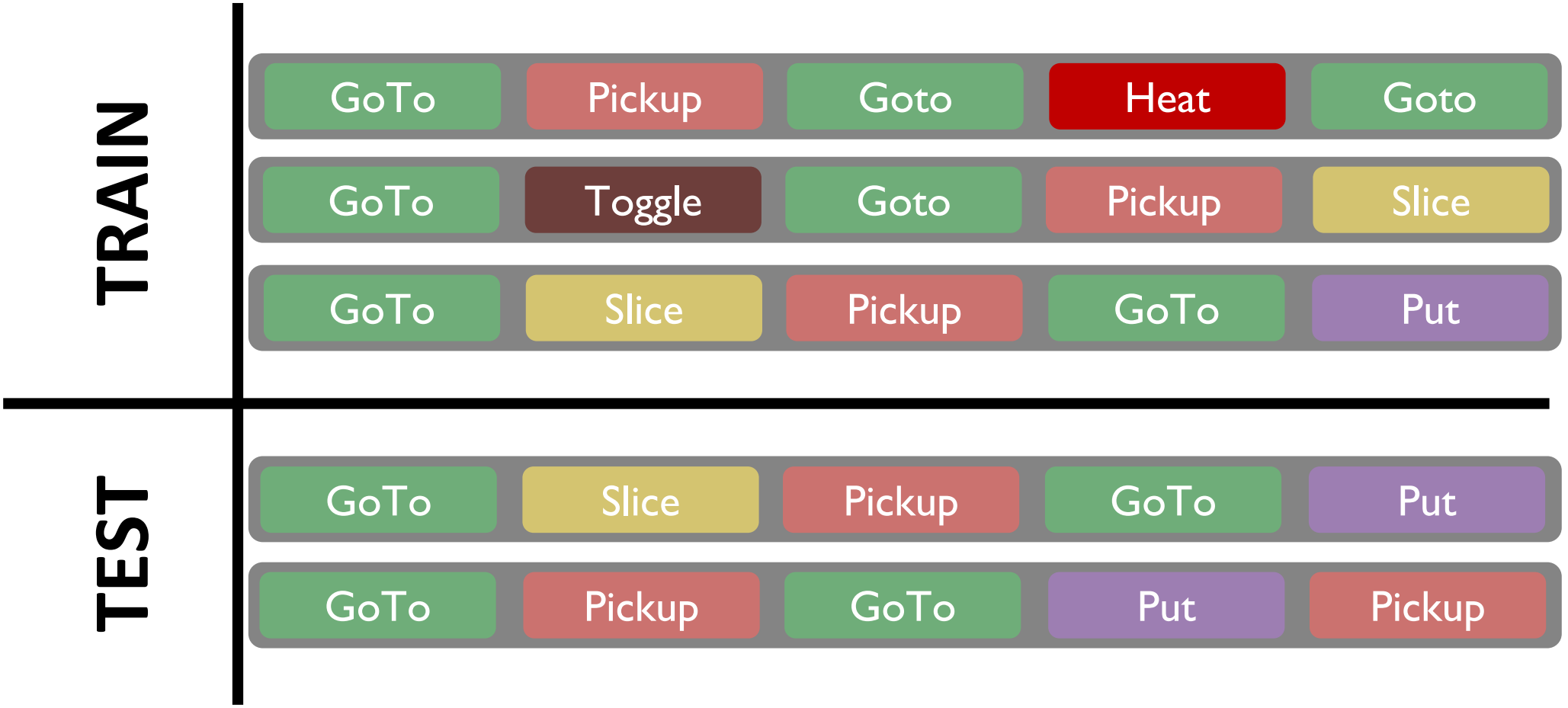
Pickup GoTo Put

Compositional Generalization

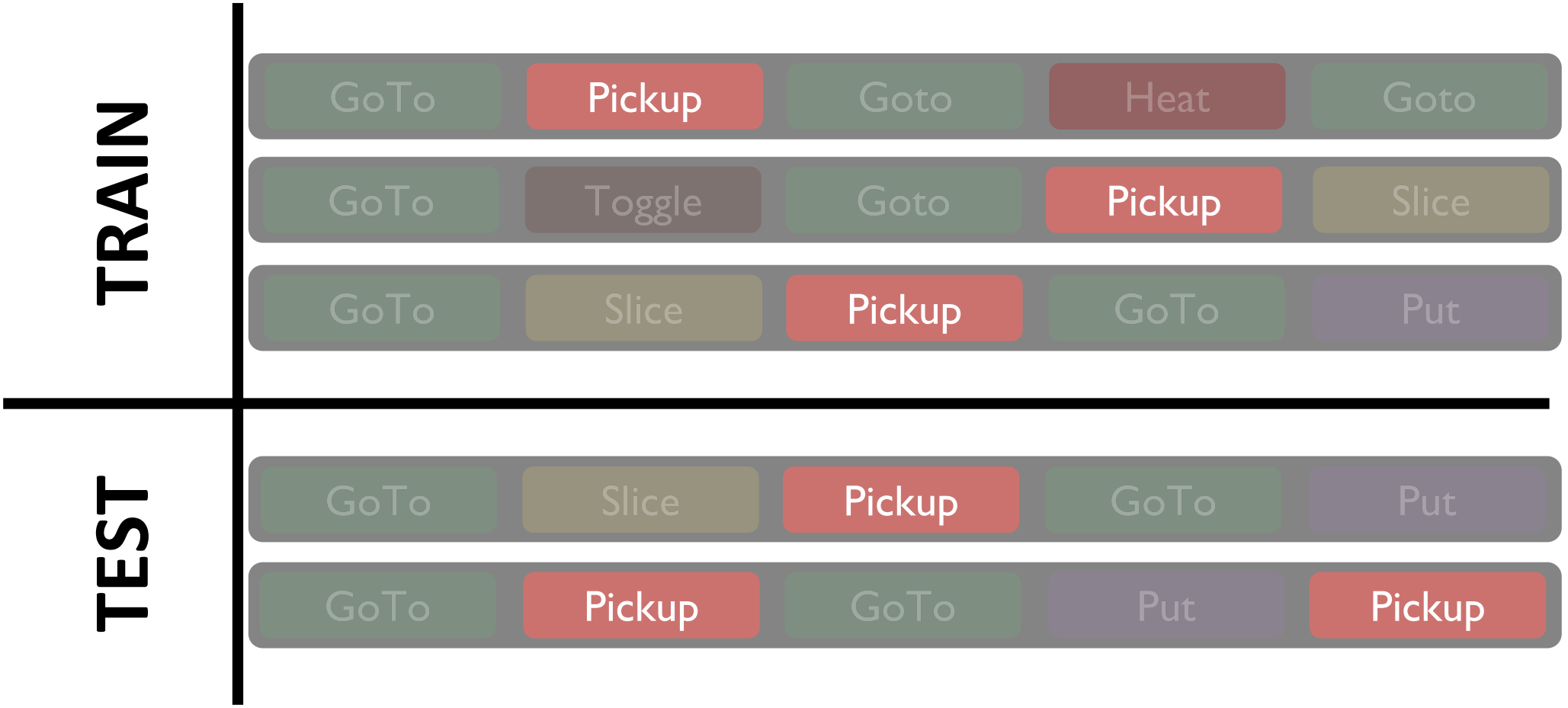
Pickup a watch and turn on the light.

Pickup GoTo Toggle

Compositional Generalization



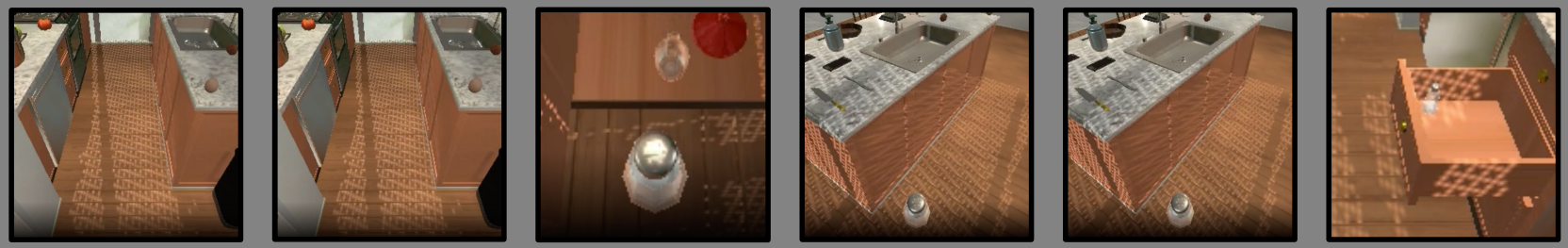
Compositional Generalization



Compositional Generalization

Action Compositionality

Walk to the spice cabinet. Reach in for a salt shaker. Head over to the drawer. Place the shaker inside the drawer.



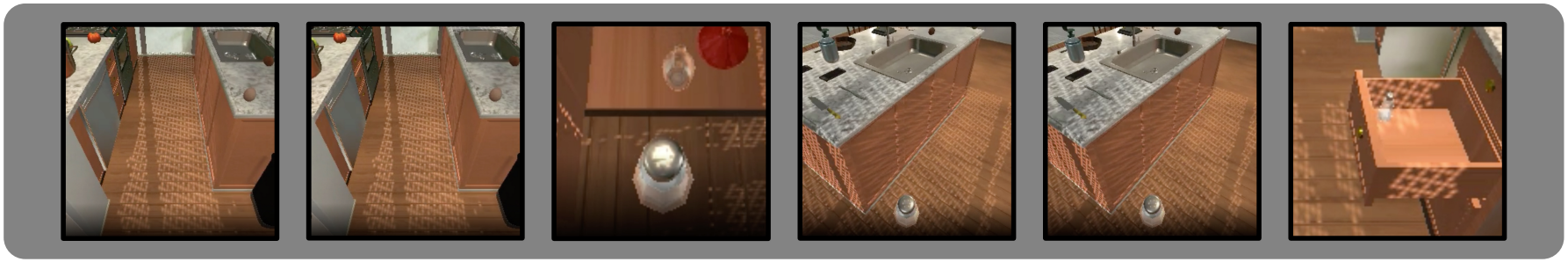
Action Compositionality

Walk to the spice cabinet

Reach in for a salt shaker

Head over to the drawer

Place the shaker inside the drawer



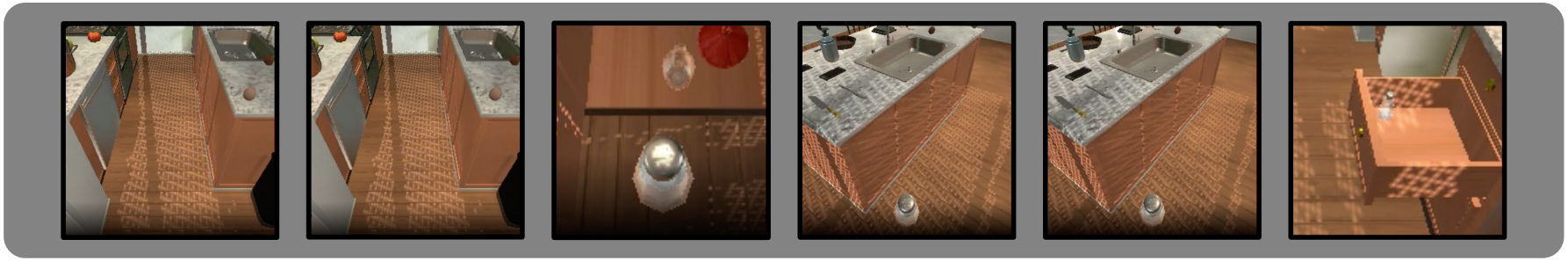
Action Compositionality

GoTo

Reach in for a salt shaker

Head over to the drawer

Place the shaker inside the drawer



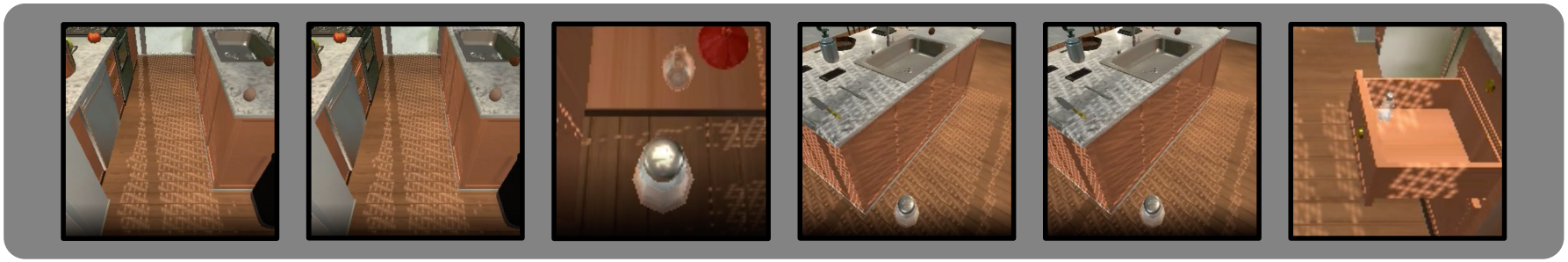
Action Compositionality

GoTo

Pickup

Head over to the drawer

Place the shaker inside the drawer



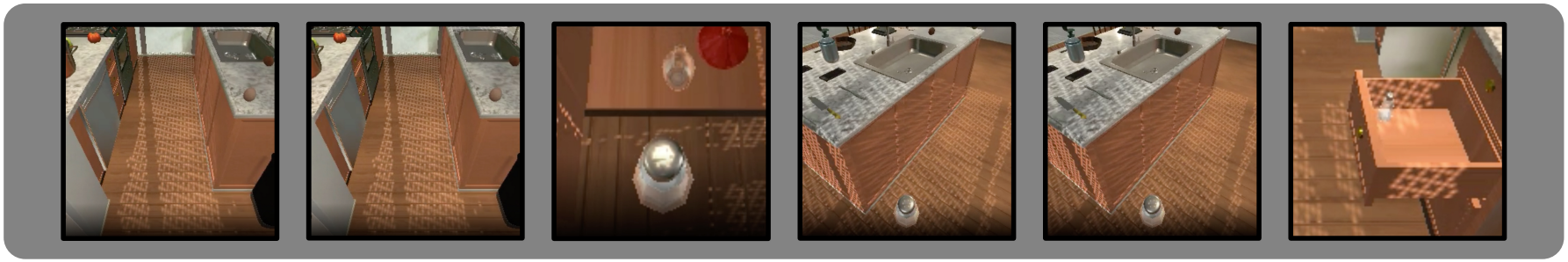
Action Compositionality

GoTo

Pickup

GoTo

Place the shaker inside the drawer



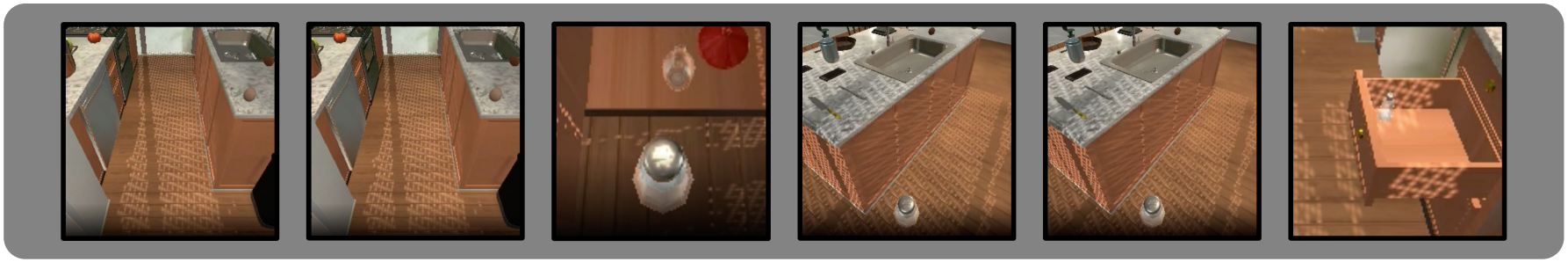
Action Compositionality

GoTo

Pickup

GoTo

Put



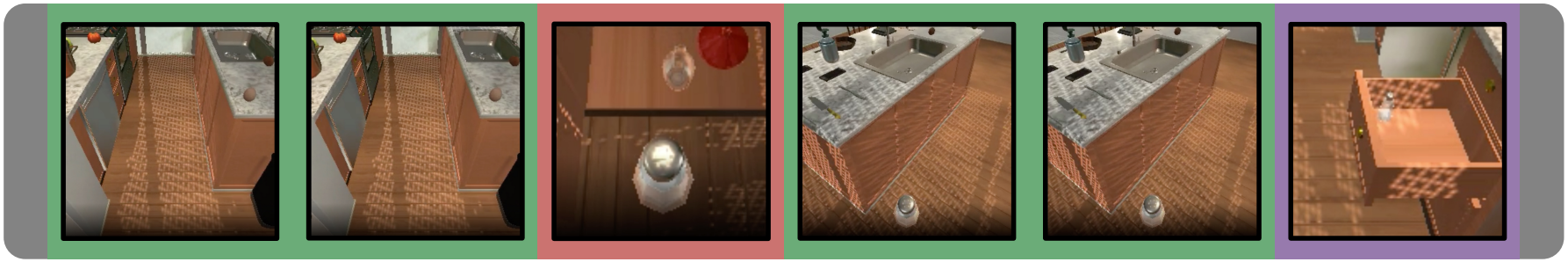
Action Compositionality

GoTo

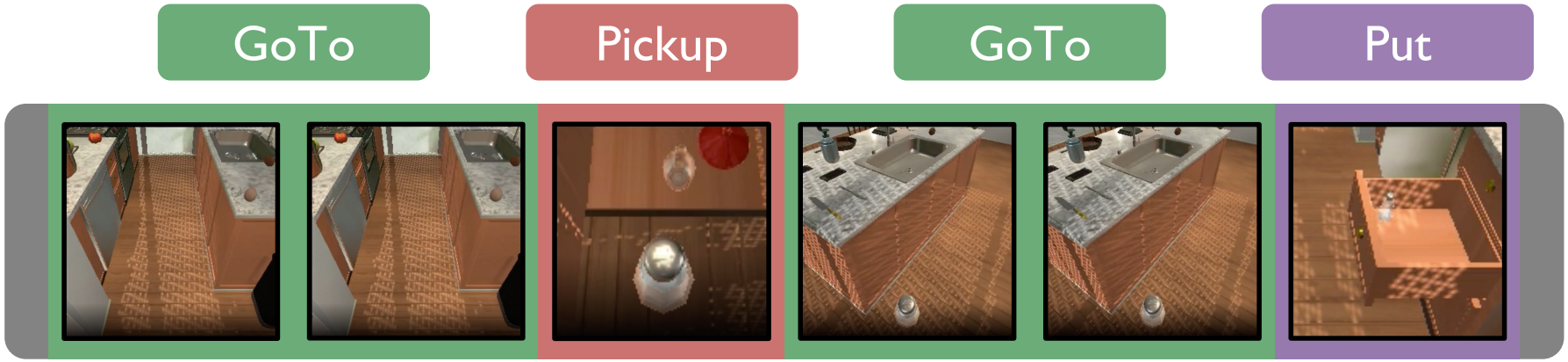
Pickup

GoTo

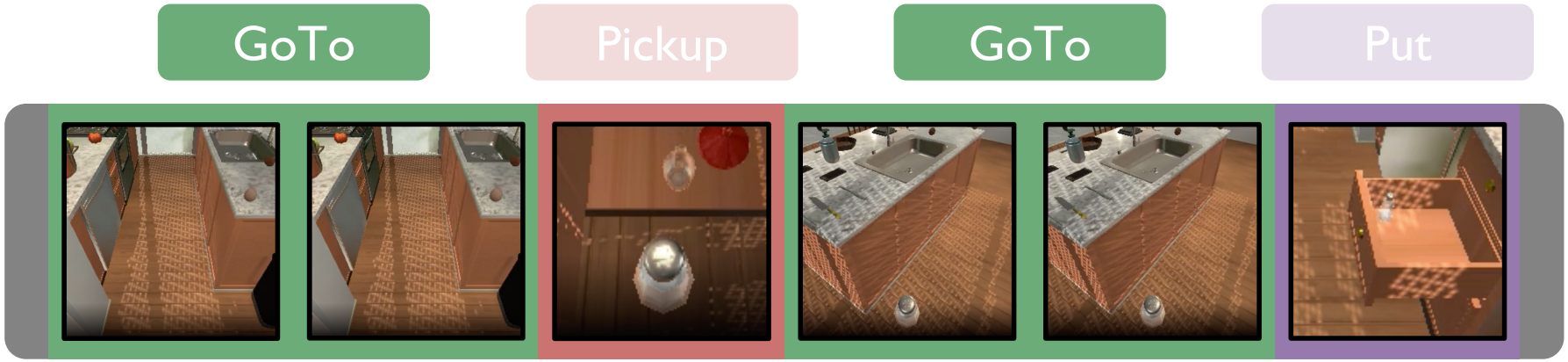
Put



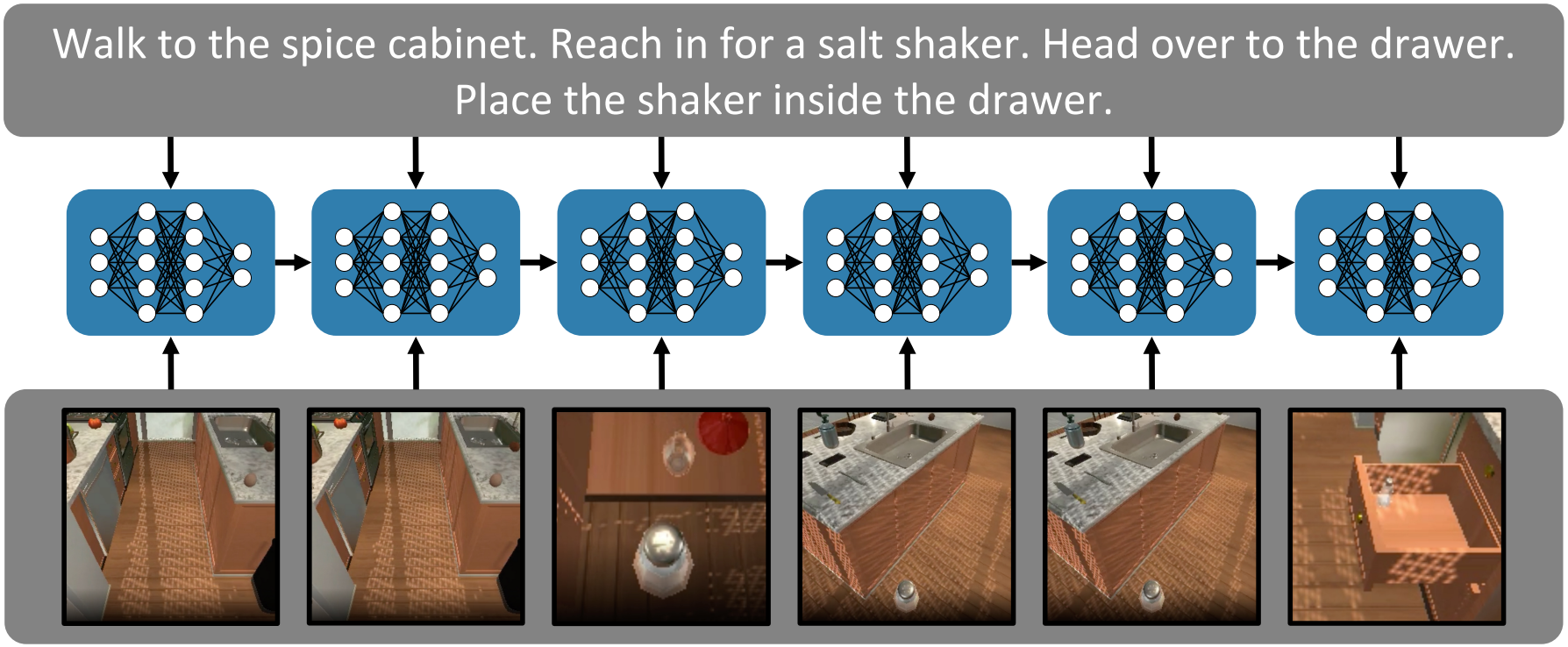
Action Compositionality



Action Compositionality



Monolithic Architectures



Modular Specialization

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Walk to the spice cabinet. Reach in for a salt shaker. Head over to the drawer. Place the shaker inside the drawer.

Modular Specialization

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Walk to the spice cabinet

GoTo

Reach in for a salt shaker. Head over to the drawer. Place the shaker inside the drawer.

Modular Specialization

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Walk to the spice cabinet

Reach in for a salt shaker

Head over to the drawer. Place the shaker inside the drawer.

GoTo

Pickup

Modular Specialization

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Walk to the spice cabinet

GoTo

Reach in for a salt shaker

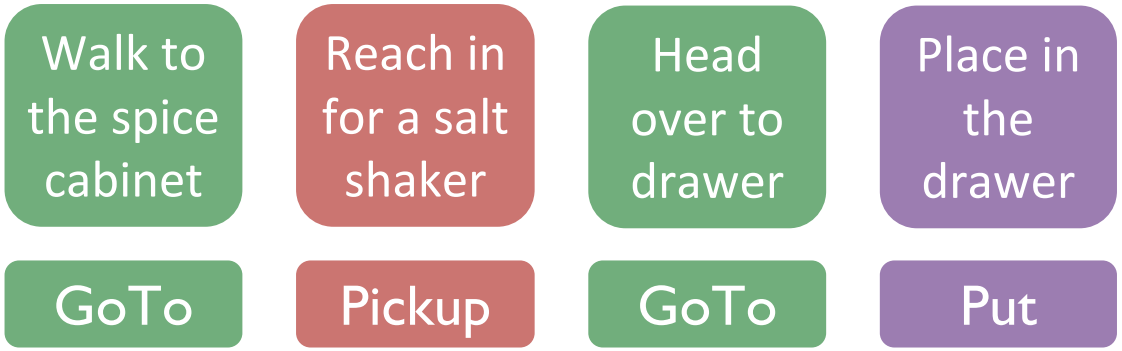
Pickup

Head over to drawer

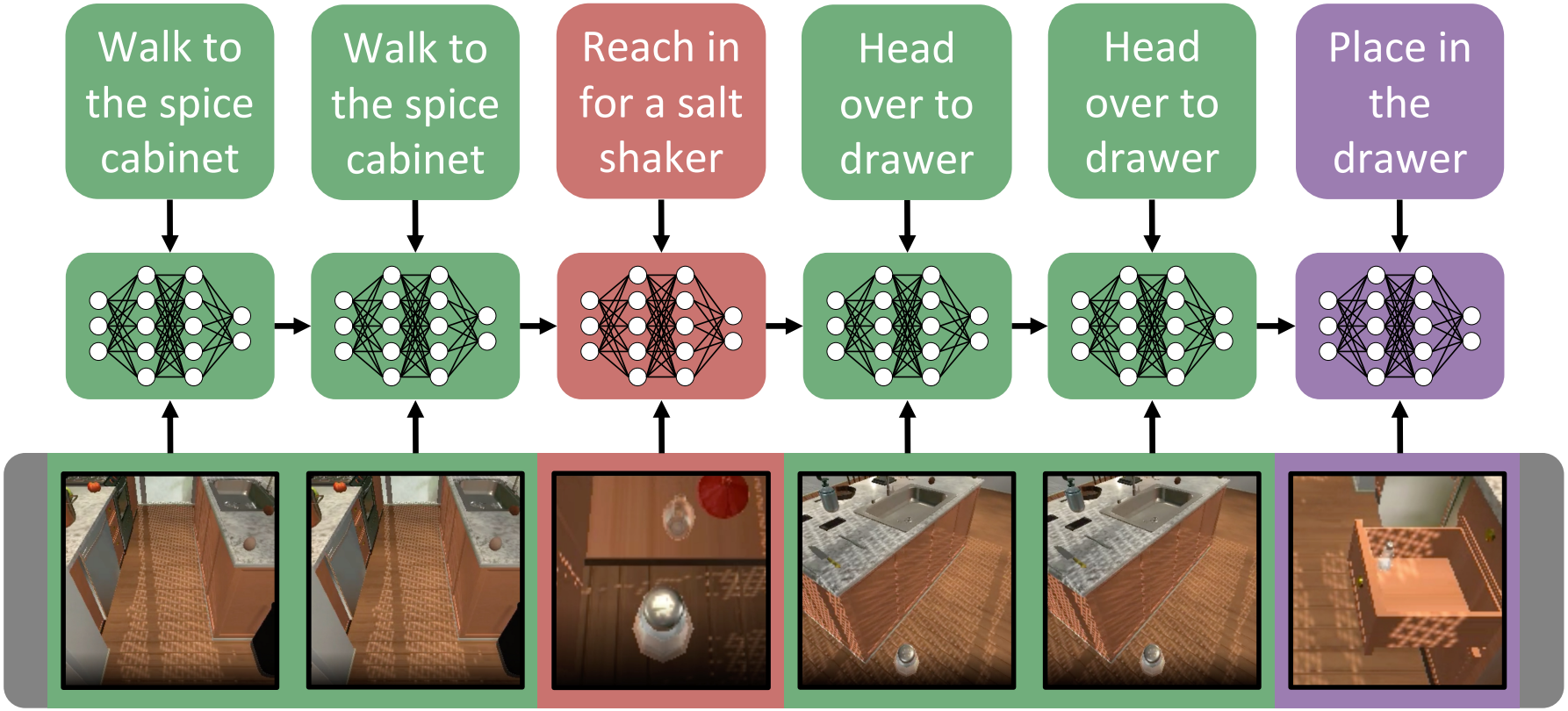
GoTo

Place the shaker inside the drawer.

Modular Specialization



Modular Specialization



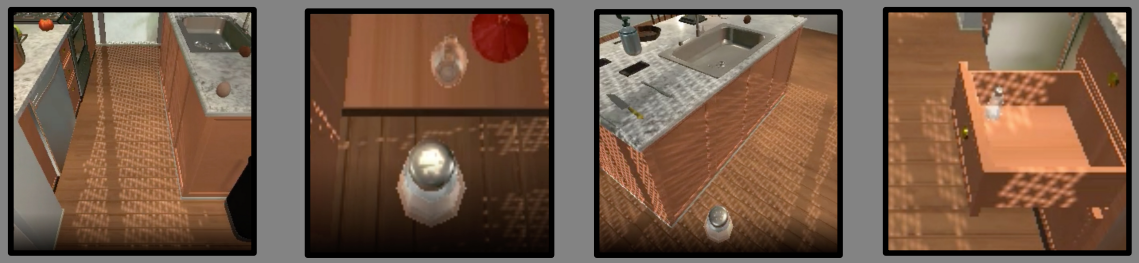
ALFRED

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Shridhar et al. 2020

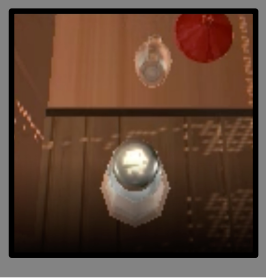
ALFRED

Embodiment



ALFRED

Embodiment

Language

Walk to the spice cabinet

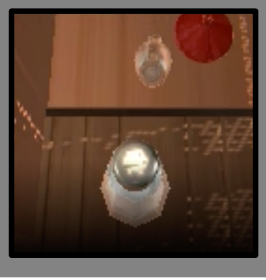
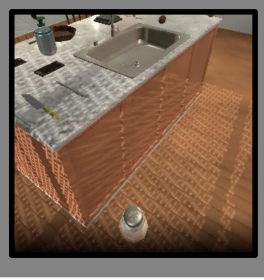
Reach in for a salt shaker

Head over to drawer

Place in the drawer

ALFRED

Embodiment


Language

Walk to the spice cabinet

Reach in for a salt shaker

Head over to drawer

Place in the drawer

Compositionality

Pickup

Put

Slice

GoTo

Toggle

Cool

Heat

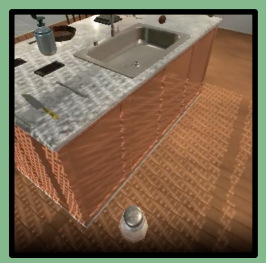
Clean

ALFRED

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Embodiment





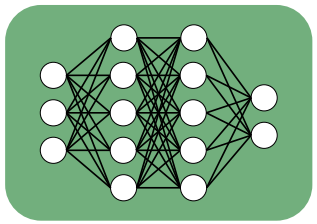

Language

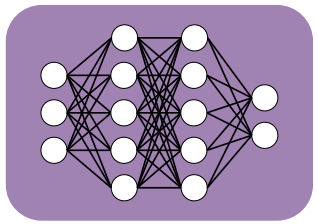
- Walk to the spice cabinet
- Reach in for a salt shaker
- Head over to drawer
- Place in the drawer

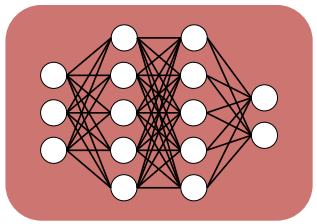
Compositionality

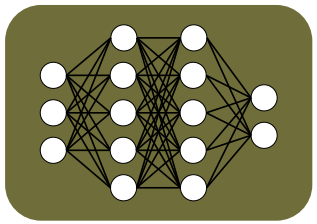
- Pickup
- Put
- Slice
- GoTo
- Toggle
- Cool
- Heat
- Clean

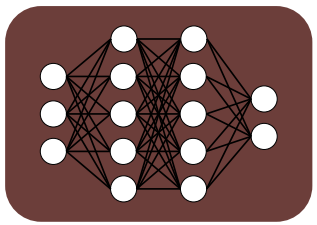
Modular Specialization

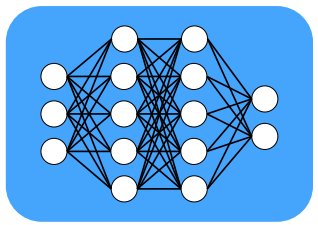

GoTo

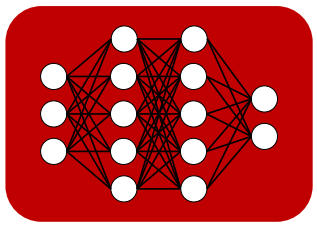

Put

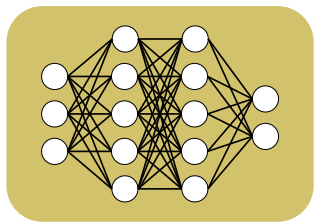

Pickup


Clean


Toggle

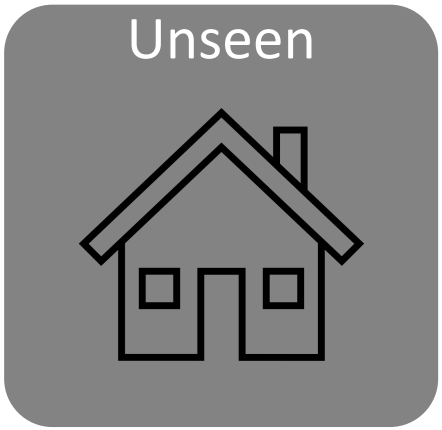

Cool


Heat


Slice

Standard Dataset Splits

Observed 3D Homes

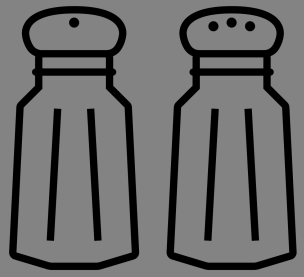


Novel 3D Homes

Compositional Splits

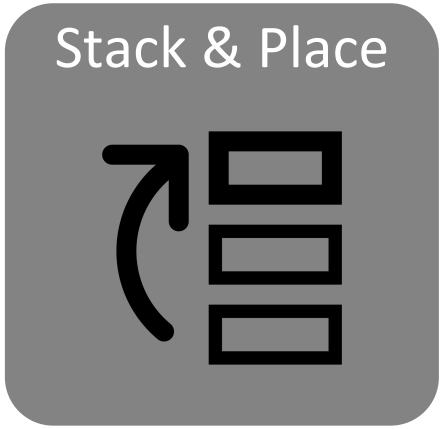
BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Pick Two



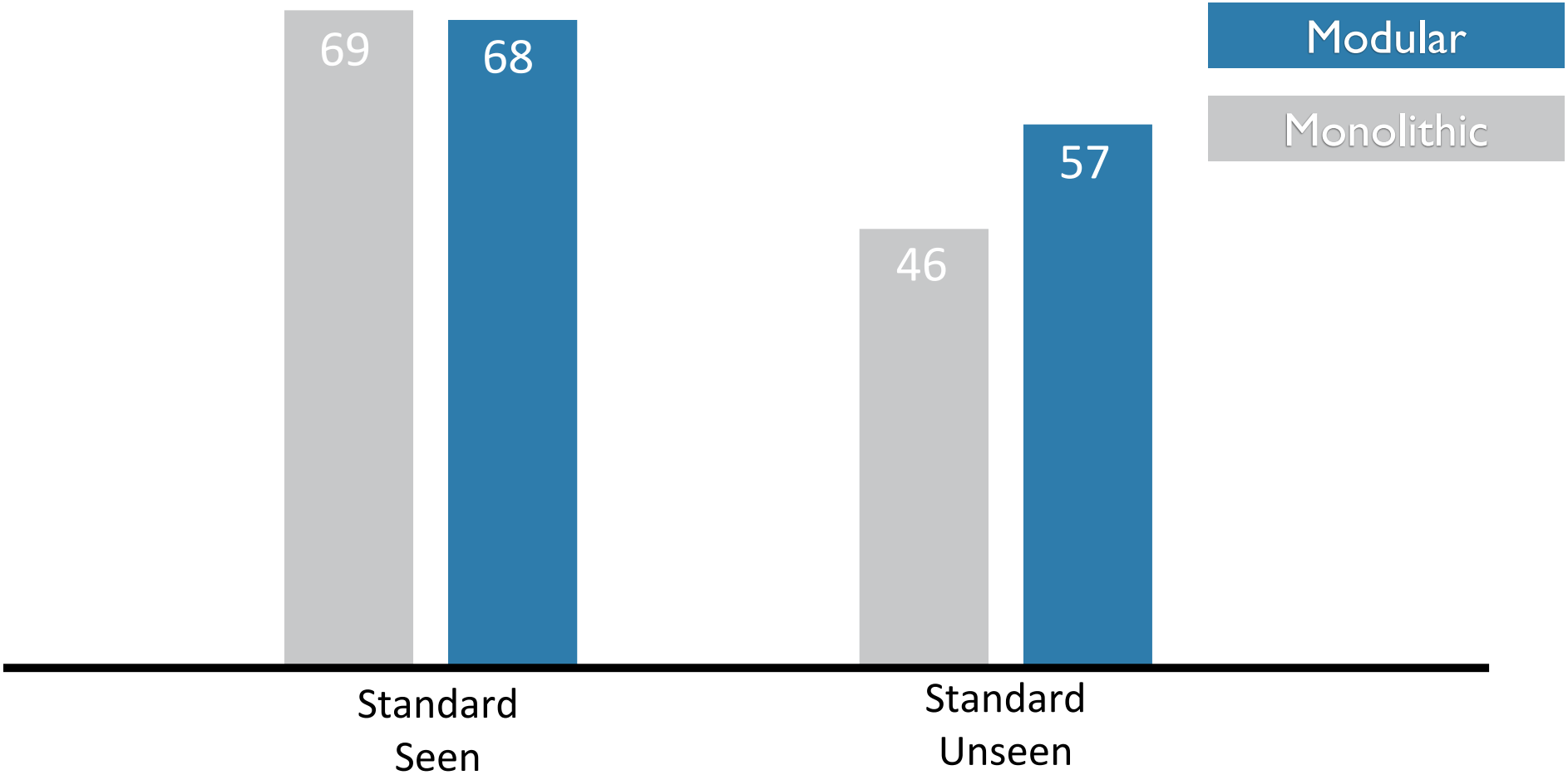
GoTo | Pickup | GoTo | Put | Pickup

Compositional Splits

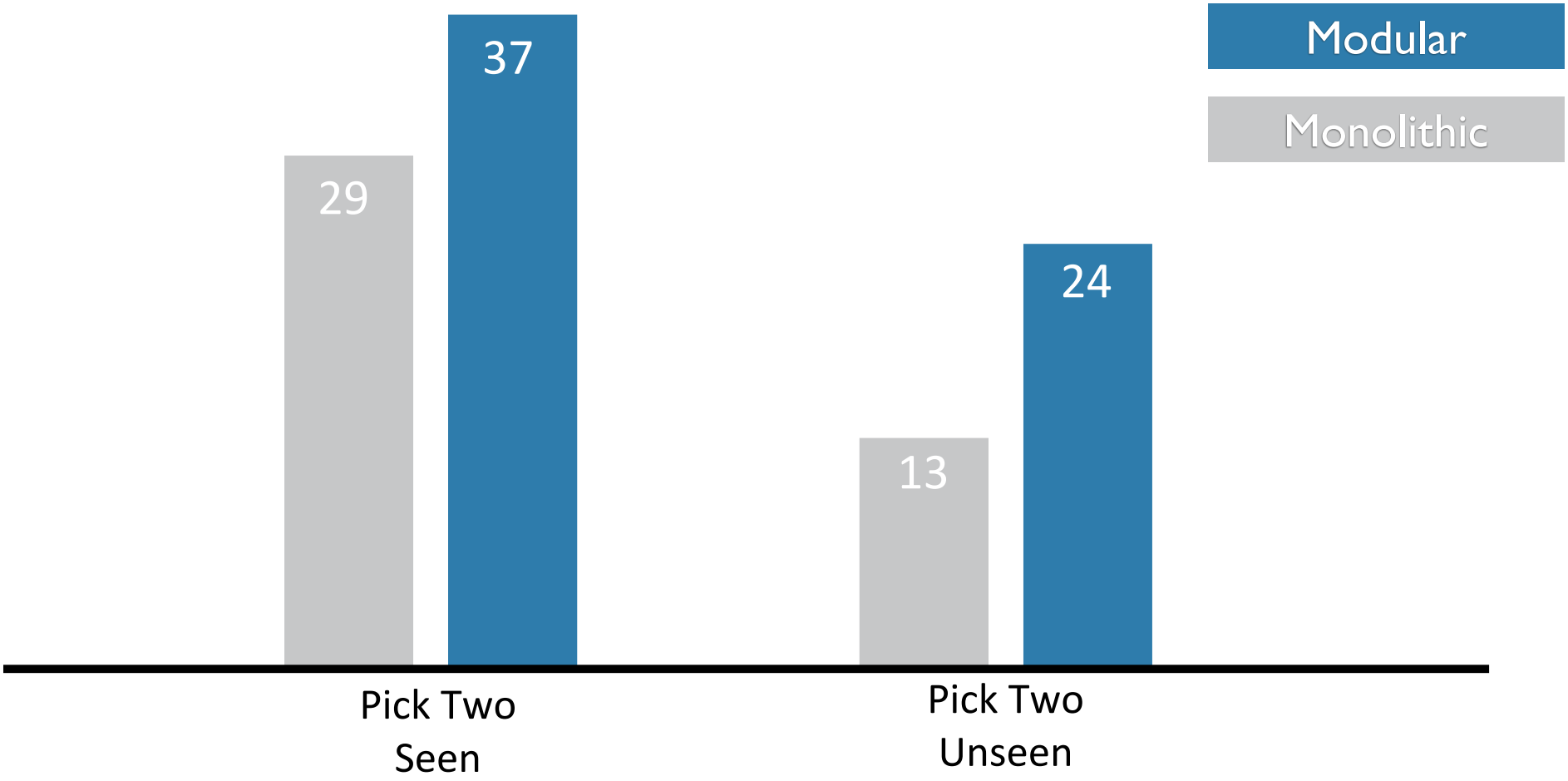


Pickup | GoTo | Put | Pickup | GoTo | Put

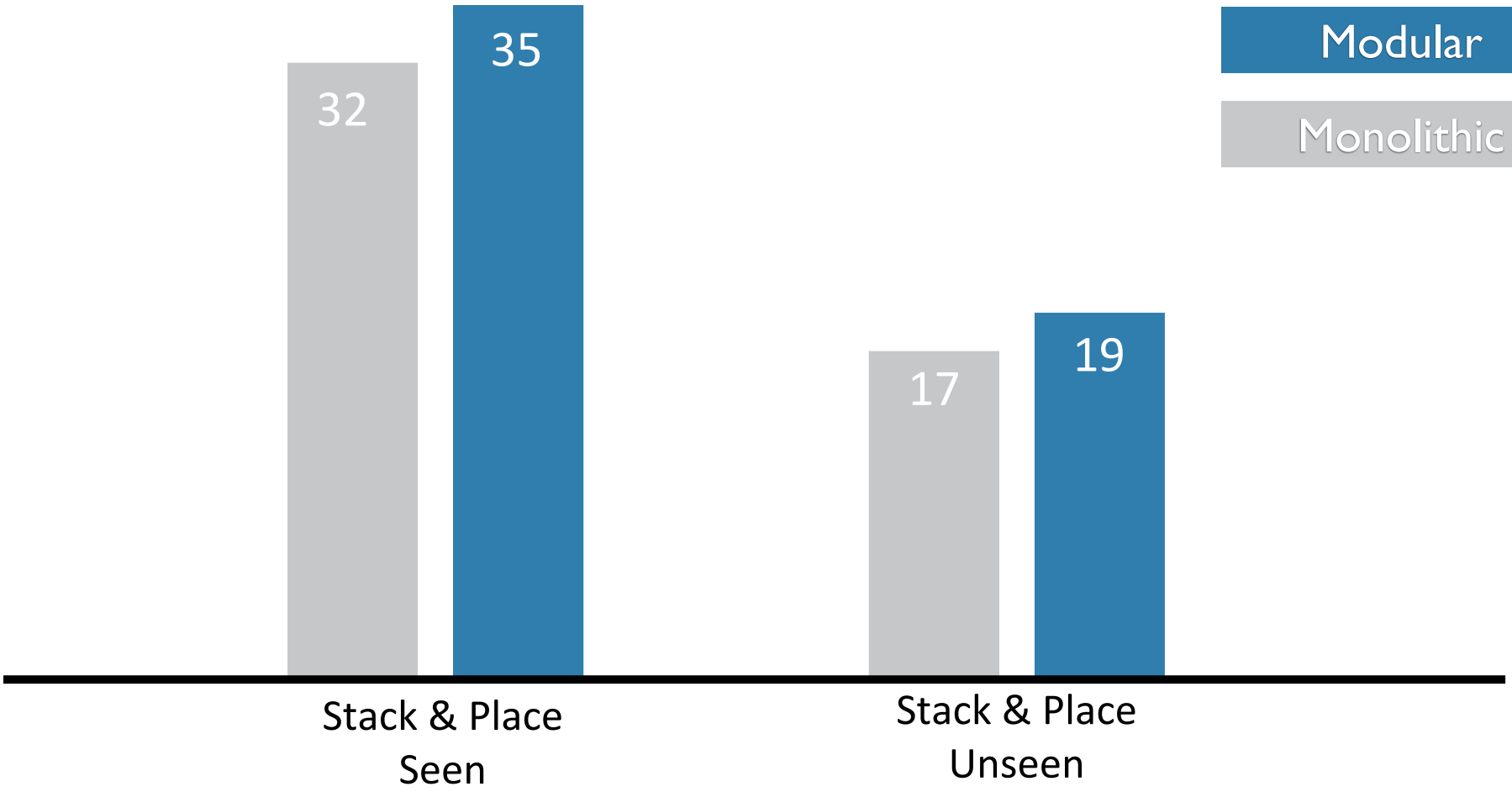
Results


Modular
Monolithic
69
68
46
57
Standard
Seen
Standard
Unseen

Results

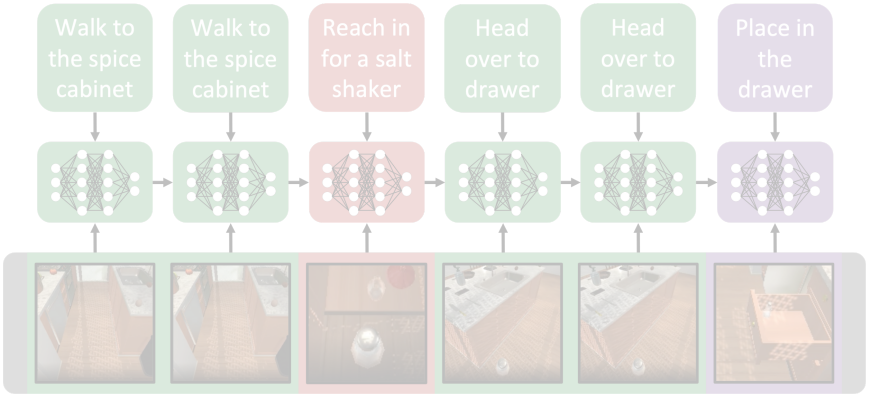


Results


Modular
Monolithic
32
35
17
19
Stack & Place Seen
Stack & Place Unseen

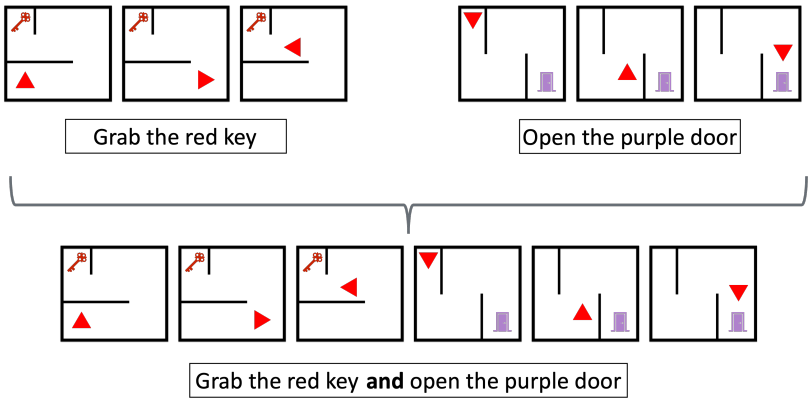
Modularity in Instruction Following

Modularize the Model



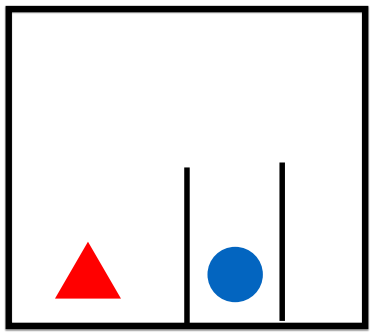
[Corona et al. 2021]

Modularize the Data

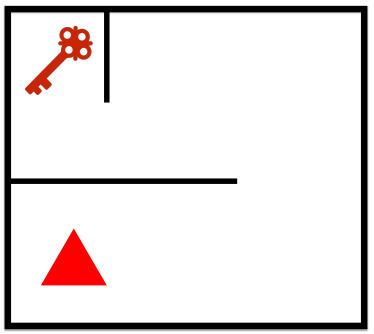


[Kantharuban et al. In submission]

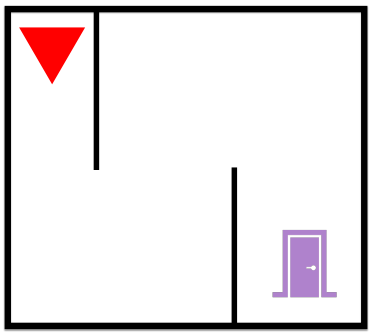
Common Building Blocks



Go to the blue circle

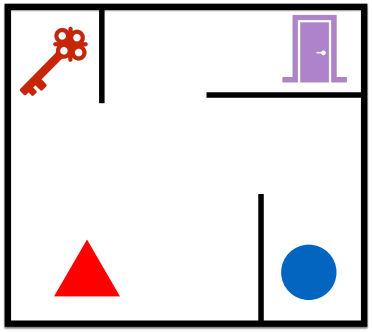


Grab the red key



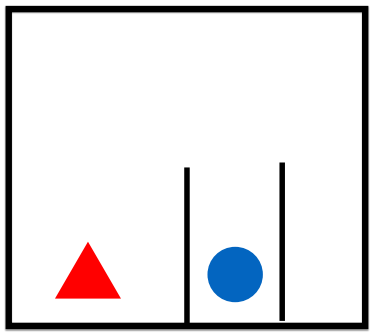
Open the purple door

Sparsity of Compositions

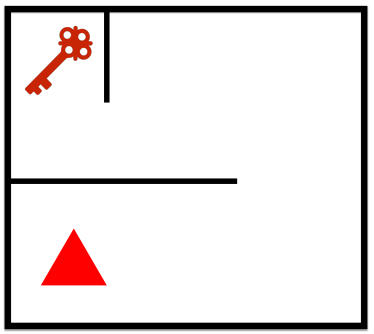


Go to the blue circle, grab the red key, and open the purple door.

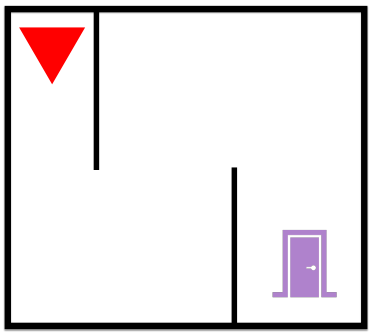
Decontextualization



Go to the blue circle



Grab the red key



Open the purple door

Decontextualization

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Go to the blue circle

Grab the red key

Open the purple door

Decontextualized Data Augmentation

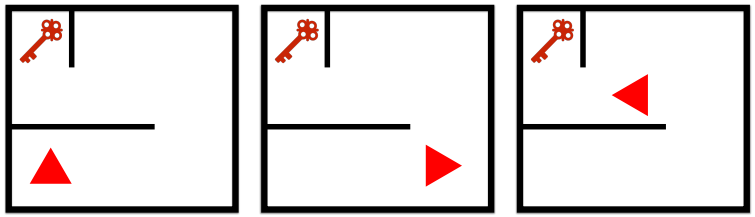
BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Go to the blue circle, grab the red key, and open the purple door.

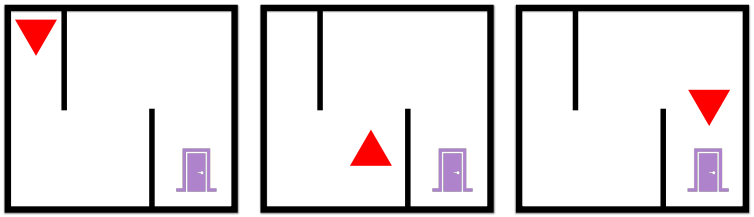
=

Go to the blue circle + Grab the red key + Open the purple door

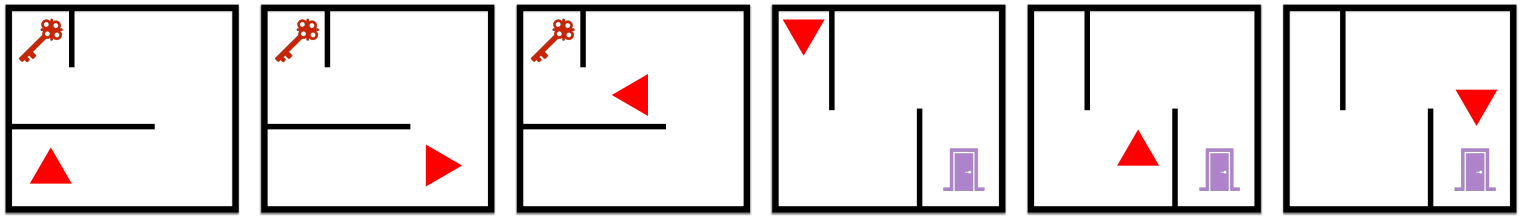
Smash Cuts



Grab the red key



Open the purple door



Grab the red key **and** open the purple door

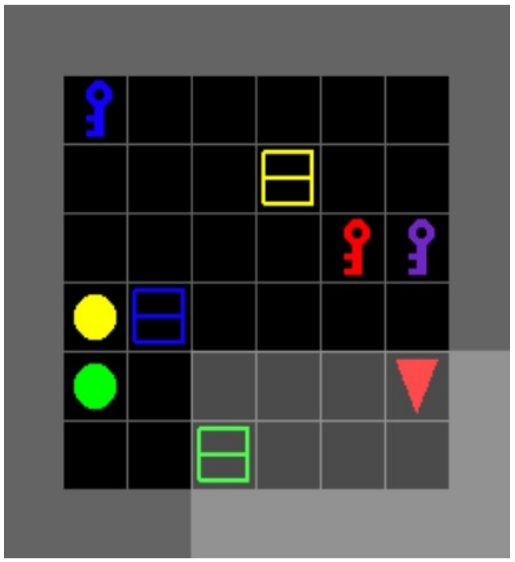
Datasets

Crafting

"Chop down a tree and mine a rock."

[Devin et al. 2019]

BabyAI

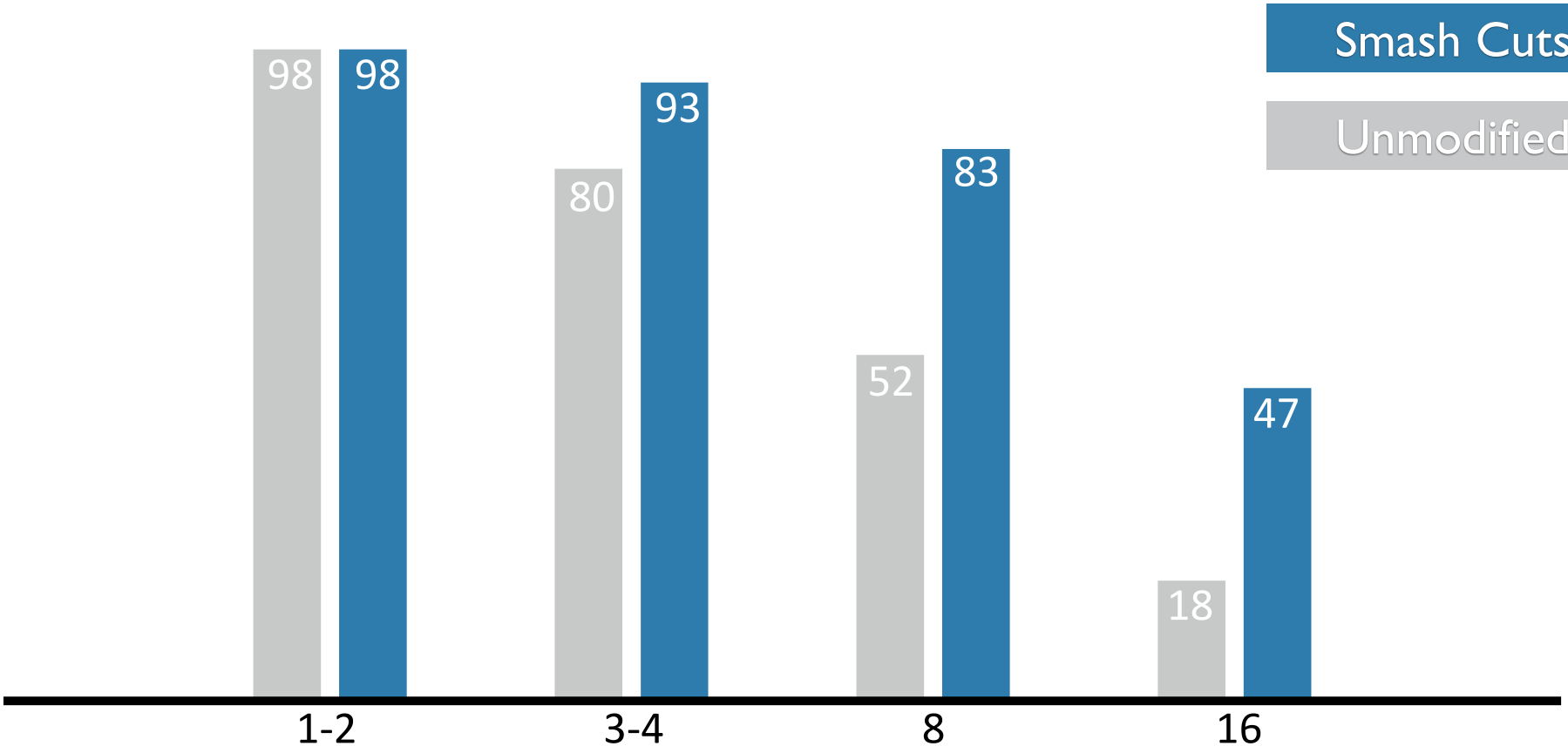


"Put the blue key next to the green ball."

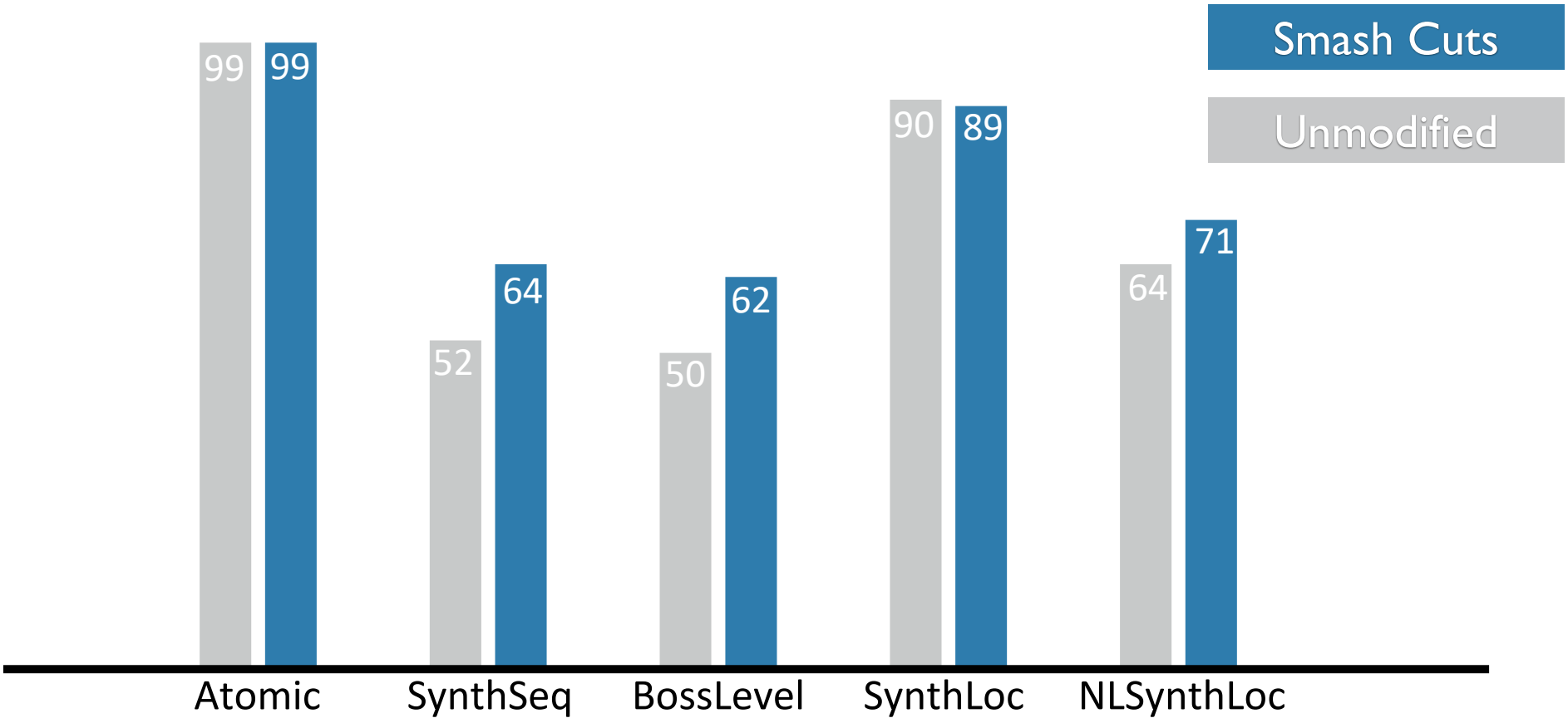
[Chevalier-Boisvert et al. 2019]

Crafting Results

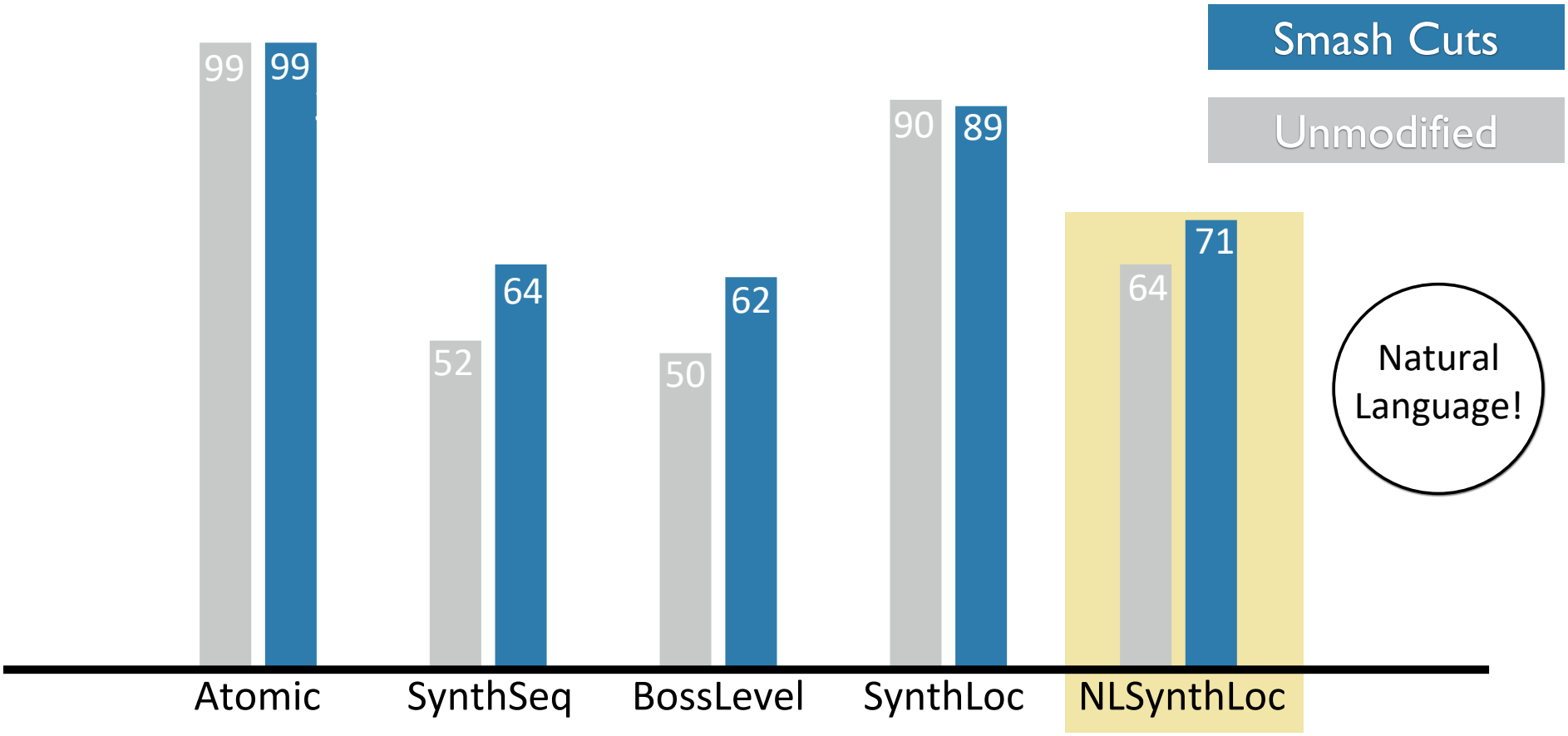
BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH



BabyAI Results



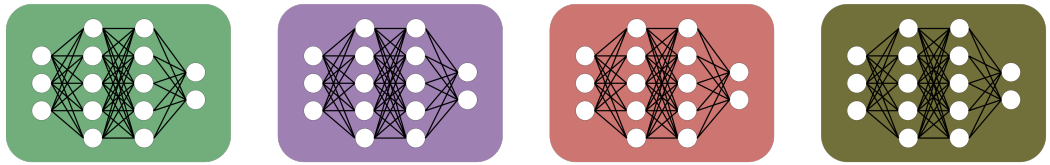
BabyAI Results



[Marzoev et al. 2020]

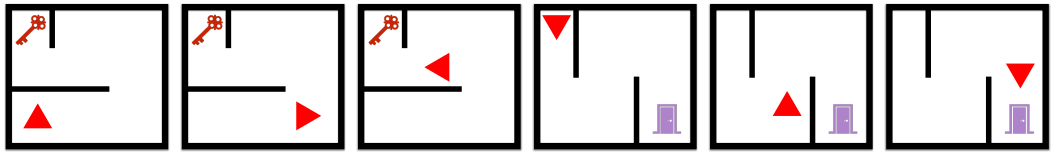
Takeaways

Modularize the Model



Modularizing action execution can improve compositional generalization

Modularize the Data

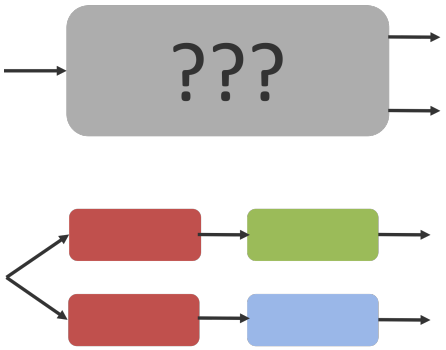


Training on simulated compositions can improve performance on real ones

Why Modularity?

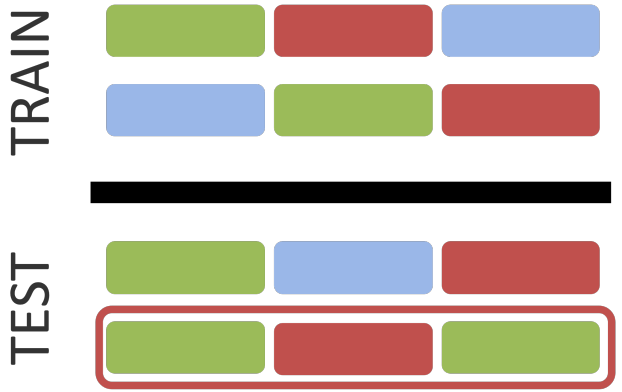
BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Systems



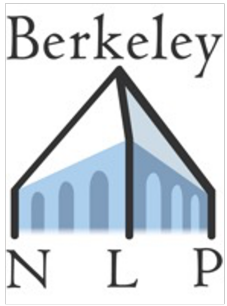
Modularity decomposes hard problems into easier ones.

Systematicity



Modularity improves compositional generalization.

Thank you!



dfried@berkeley.edu rcorona@berkeley.edu