Learning Adaptive Language Interfaces through Decomposition

Stanford | NLP

iliad

Siddharth Karamcheti

Percy Liang

Dorsa Sadigh

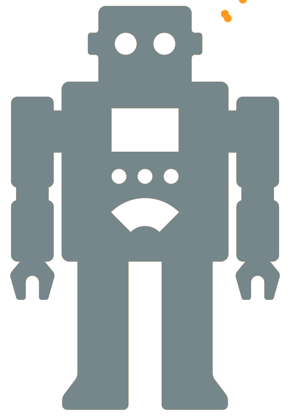

KIRKLAND

“Wash the coffee mug”

What can I do?

```
GOTO Mug;
PICKUP Mug;
PUT Mug;
…
```

I’m sorry – I don’t understand!
Please teach me!



Learning from Decomposition

Interaction

Wash the coffee mug

`I'm sorry - I don't understand!`

Go to the mug and pick it up

`GOTO Mug; PICKUP Mug`

Go to the sink and put it inside

`GOTO Sink; PUT Mug Sink`

Turn on the faucet

`TOGGLE Faucet`

Turn it off

`TOGGLE Faucet`

Pick up the mug

`PICKUP Mug`

...

Learning from Decomposition

Interaction

Wash the coffee mug

`I'm sorry - I don't understand!`

Go to the mug and pick it up

`GOTO Mug; PICKUP Mug`

Go to the sink and put it inside

`GOTO Sink; PUT Mug Sink`

Turn on the faucet

`TOGGLE Faucet`

Turn it off

`TOGGLE Faucet`

Pick up the mug

`PICKUP Mug`

...

Teaching

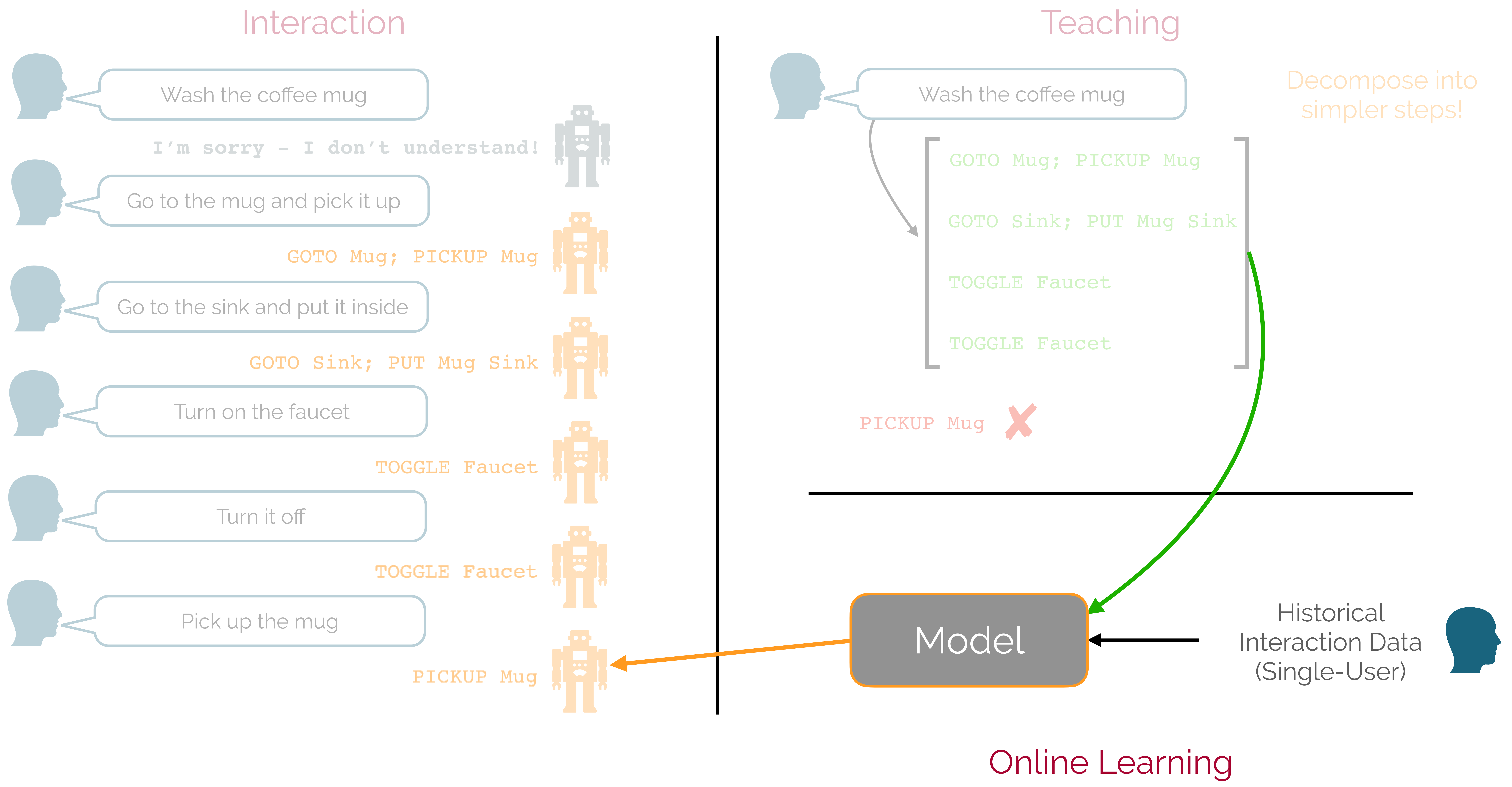
Wash the coffee mug

```
[ GOTO Mug; PICKUP Mug
  GOTO Sink; PUT Mug Sink
  TOGGLE Faucet
  TOGGLE Faucet ]
```

`PICKUP Mug` ✗

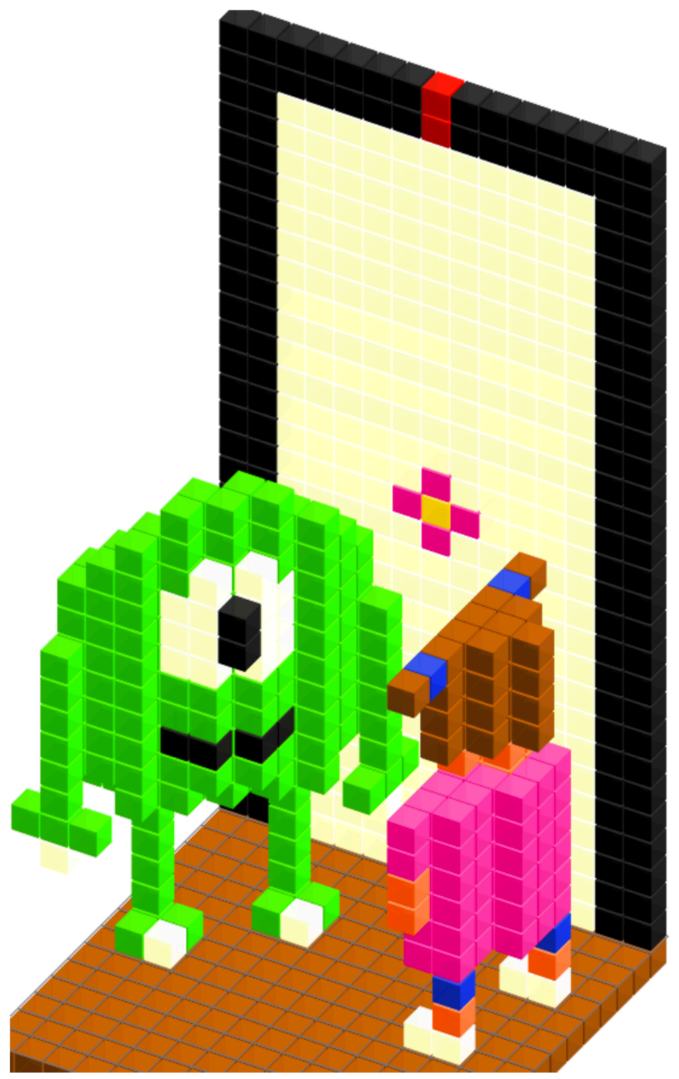
Decompose into simpler steps!

Learning from Decomposition



Related Work — Semantic Parsing & Interaction

- Closest to our work is Voxelurn [1]
  - Grammar-based semantic parsers:
    - *Reliable one-shot generalization* ✓
    - *Lexical flexibility* ✗
      - “Add palm tree” —> “Create a palm tree”



- Separately: Neural sequence-to-sequence models [2, 3].
  - *Lexical flexibility* ✓
  - *Reliable one-shot generalization* [4] ✗

```
def: add palm tree
    def: brown trunk height 3
        def: add brown top 3 times
            repeat 3 [add brown top]
    def: go to top of tree
        select very top of has color brown
    def: add leaves here
        def: select all sides
            select left or right or front or back
        add green
```

[1] Naturalizing a Programming Language via Interactive Learning — Wang et. al. 2017
[2] Data Recombination for Neural Semantic Parsing — Jia and Liang 2016
[3] From Language to Programs: Bridging Reinforcement Learning and Maximum Marginal Likelihood — Guu et. al. 2017
[4] Six Challenges for Neural Machine Translation — Koehn and Knowles 2017

This Work:

Applies interactive learning from decomposition, and introduces a *neural "exemplar-based" parser* that is *lexically flexible* and can *reliably generalize from limited examples*.

Critical Point:

Trust during Inference — Given a novel utterance, output *"I don't understand."*


KIRKLAND

This Work:

Applies interactive learning from decomposition, and introduces a *neural "exemplar-based" parser* that is *lexically flexible* and can *reliably generalize from limited examples*.

Critical Point:

Trust during Inference — Given a novel utterance, output *"I don't understand."*

Roadmap:

1. **Exemplar-Based Semantic Parsing**
2. Experiments
3. Limitations & Discussion

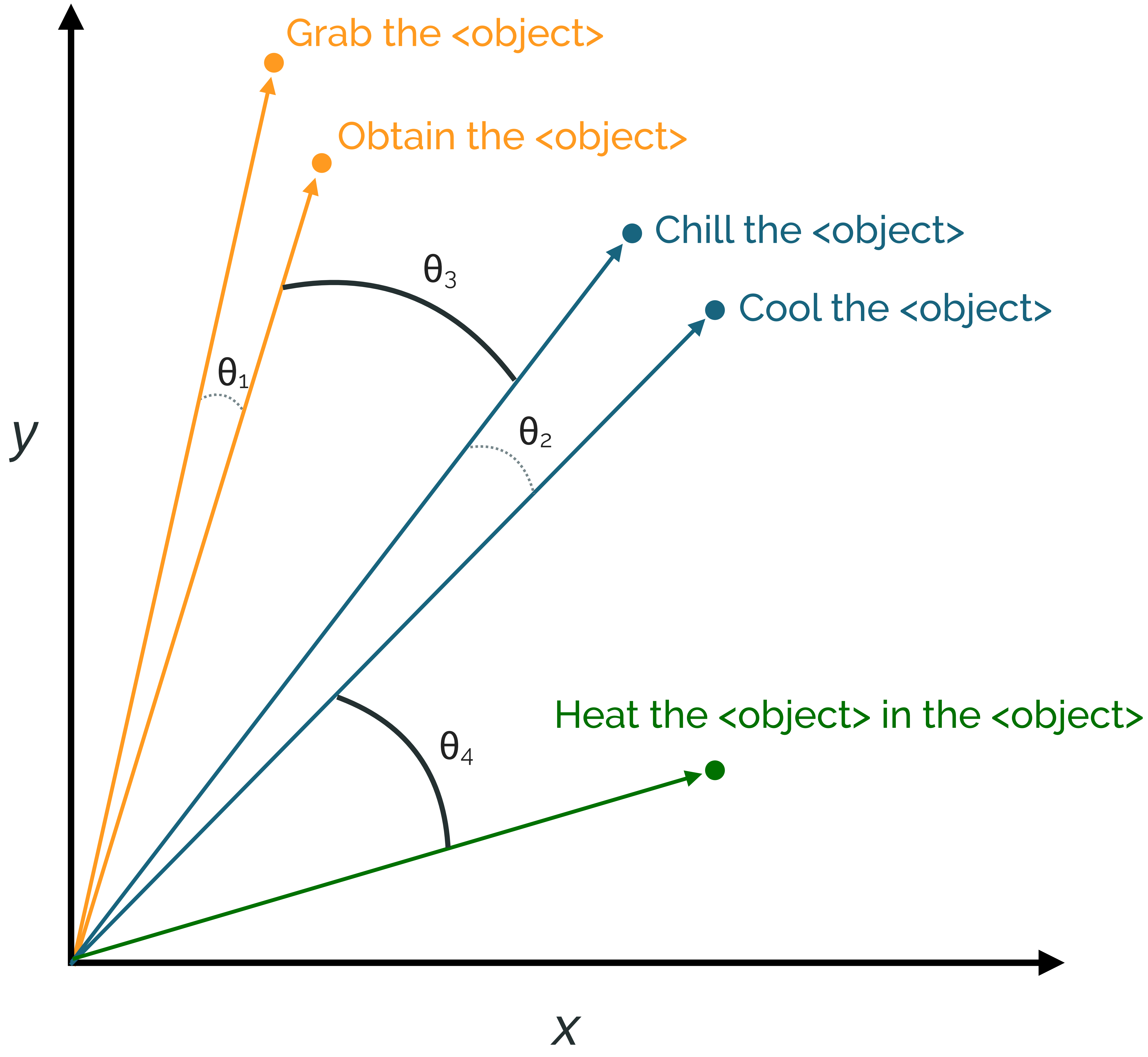
Exemplar-Based Semantic Parsing

Overview: Treat each (utterance, program) pair as a single point in a learned latent space.

Inference: Given utterance *u, embed u and retrieve closest "exemplar."*

Reliable Generalization: Decouple "functions"/arguments" and operate on "lifted" utterances.

Trust during Inference



2D Visualization of our learned latent space.

Trust during Inference:

Given a novel utterance, how to output *"I don't understand!"*

Intuition:

Set a "threshold" τ in embedding space!

if **CosineDistance**$(\phi(u), \phi(i)) \geq \tau \quad \forall_i$

return `"I don't understand"`

Roadmap:

1. Exemplar-Based Semantic Parsing
2. **Experiments**
3. Limitations & Discussion

Environments & Tasks

Environment: We use a simplified (2D) version of the AI2-THOR Simulation Environment [1]

Tasks: We borrow from ALFRED [2]:

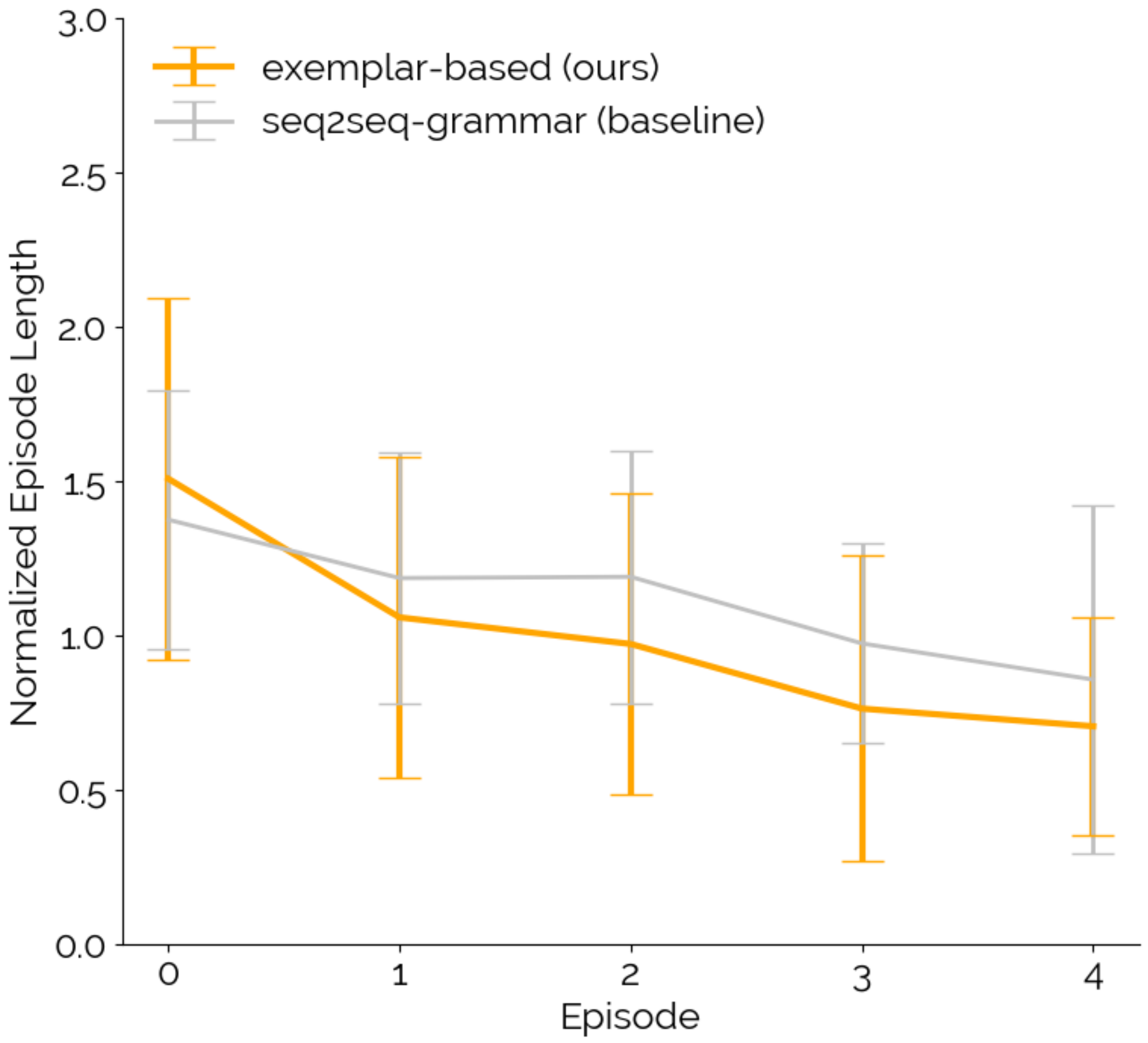
- **Pick and Place**
- **Examine in Light**
- **Nested Pick and Place**
- **Pick Two and Place**
- **Pick, Heat, and Place**
- **Pick, Clean, and Place**
- **Pick, Cool, and Place**

Simple Primitives:

- GOTO (object)
- PICKUP (object)
- TOGGLE (object)
- PUT (object, receptacle)
- OPEN/CLOSE (receptacle)

[1] AI2-THOR: An Interactive 3D Environment for Visual AI — Kolve et. al. 2017
[2] ALFRED: A Benchmark for Interpreting Grounded Instructions for Everyday Tasks — Shridhar et. al. 2020

Results — Multi-Task [20 Users x 7 Task Types]



Normalized Episode Length
How many utterances does it take a user to complete a task?

(Lower is better!)

Baseline: Seq2Seq + Backoff Grammar

Seq2seq models are unpredictable when trained with limited data — leverage backoff grammar:

- Grammar-based parser for *"simple" instructions.*
- Seq2Seq responsible for "high-level" language!

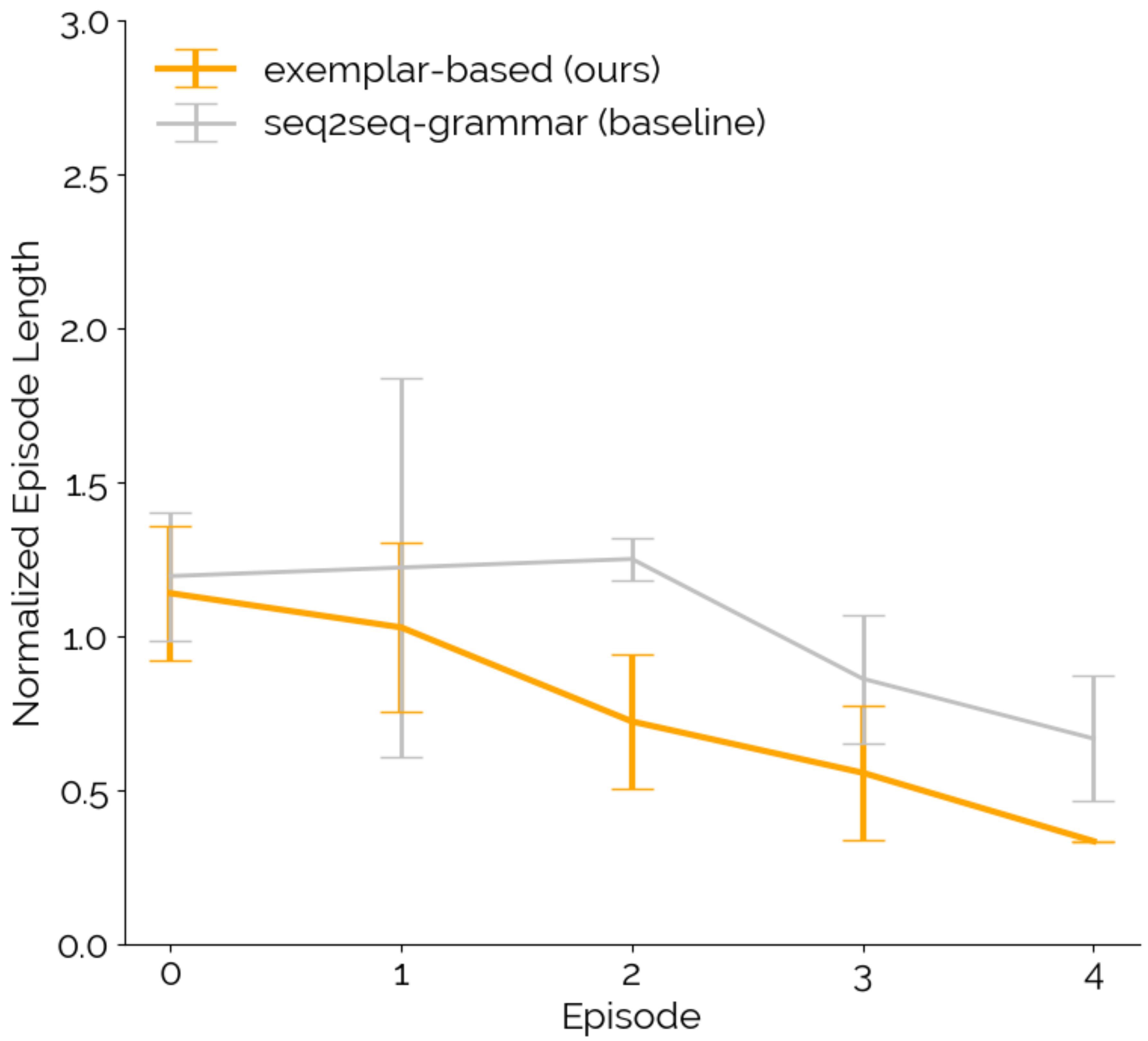
Takeaway: Users don't seem to be *teaching or re-using* new high-level language!

- For seq2seq-grammar baseline —> *89.9% of all utterances handled by grammar!*

Users are not incentivized to teach …

… in simple tasks.

Results — Pick, Cool, and Place [3 Users]



Normalized Episode Length
How many utterances does it take a user to complete a task?

Takeaways:

- Normalized Length ~0.3
- Users are able to:
  - Reuse high-level abstractions.
  - Complete tasks in 1/3 of the time!
  - Less reliance on simple utterances.

Roadmap:

1. Exemplar-Based Semantic Parsing
2. Experiments
3. **Limitations & Discussion**

Discussion, Limitations, and Looking Forward

Towards More Complex Settings

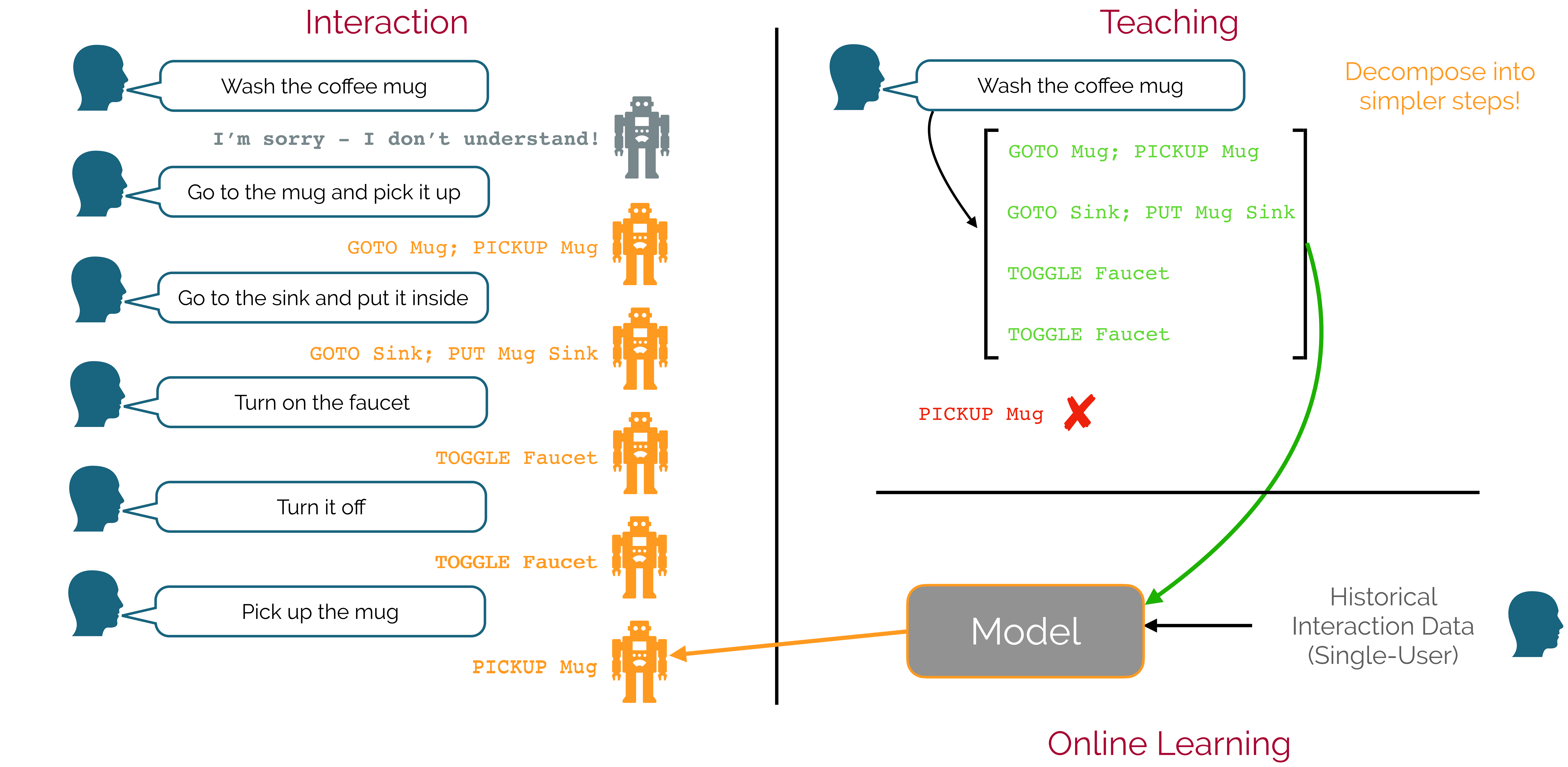
- Simple settings, users take the "shortest path."
- We need environments where there is a "natural incentive" to teach nested abstractions!
    - Minecraft
    - Cooking (a la EPIC-KITCHENS [1])

On Trusting Interactive Learning

- What is the system learning?
    - "Wash the mug" —> "Wash the countertop?"
- We need tools for *transparency and reliability!*

Setup for EPIC-KITCHENS [1], a free-form cooking domain where the use and definition of nested abstractions ("peel the apples," "make some pie crust") are naturally incentivized.

[1] The EPIC-KITCHENS Dataset — Damen et. al. 2018



Thanks so much!

If you have questions/comments/helpful tips, feel free to email me — skaramcheti@cs.stanford.edu