Towards an Indexical Model of Situated Language Comprehension for Real-World Cognitive Agents

Shiwali Mohan, Aaron Mininger, John E. Laird

Computer Science and Engineering
University of Michigan

December 13^{th}, 2013

Interactive Intelligent Agents

Interactive Intelligent Agents

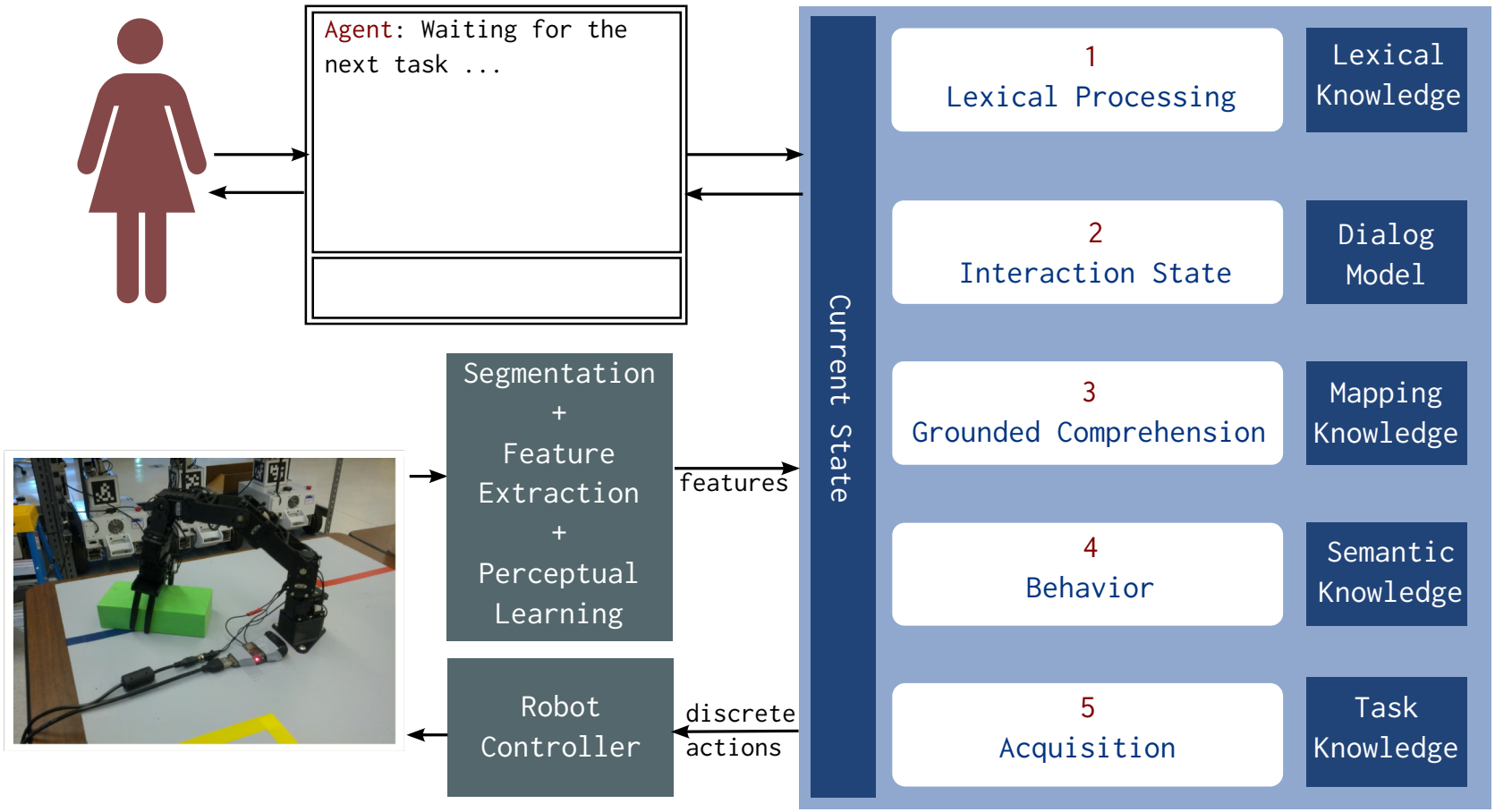
- Language for collaboration: Fong et al. 2003; Kollar et al. 2010
- Language for learning: Cantrell et al. 2011; Tellex et al. 2011; Matuszek et al. 2012; Mohan et al. 2012

Interactive Intelligent Agents

- Language for collaboration: Fong et al. 2003; Kollar et al. 2010
- Language for learning: Cantrell et al. 2011; Tellex et al. 2011; Matuszek et al. 2012; Mohan et al. 2012
- Need for a comprehensive language comprehension model

ROSIE

framework for learning with situated interactive instruction [Mohan et al. 2012]
based on Soar Cognitive Architecture [Laird 2012]



Situated Interactive Instruction

Human: Store the blue rectangle.

Agent: What is the goal of the action?

Human: The goal is the rectangle in the pantry and the pantry is closed.

Agent: I do not know the preposition in.

Human: The red circle is in the garbage.

Agent: OK.

Agent: Which action should I take?

Human: Open the pantry.

Agent: Which action should I take?

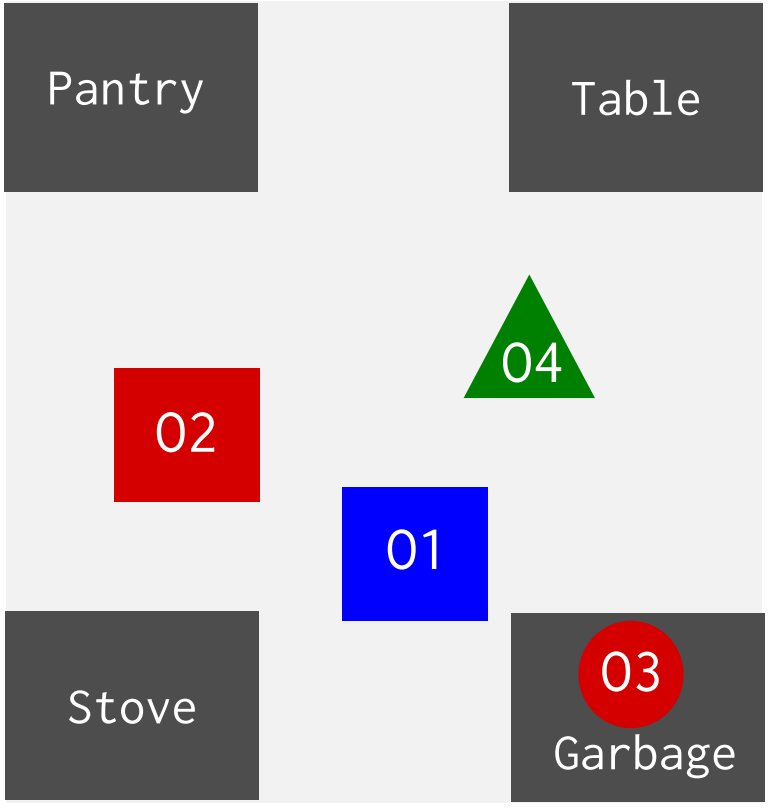
Human: Move it to the pantry.

Agent: Which action should I take?

Human: Close the pantry.

Agent: Which action should I take?

Human: You are done.



Communication is situated, contextual, interactive.

Situated Interactive Instruction

Human: Store the blue rectangle.

Agent: What is the goal of the action?

Human: The goal is the rectangle in the pantry and the pantry is closed.

Agent: I do not know the preposition in.

Human: The red circle is in the garbage.

Agent: OK.

Agent: Which action should I take?

Human: Open the pantry.

Agent: Which action should I take?

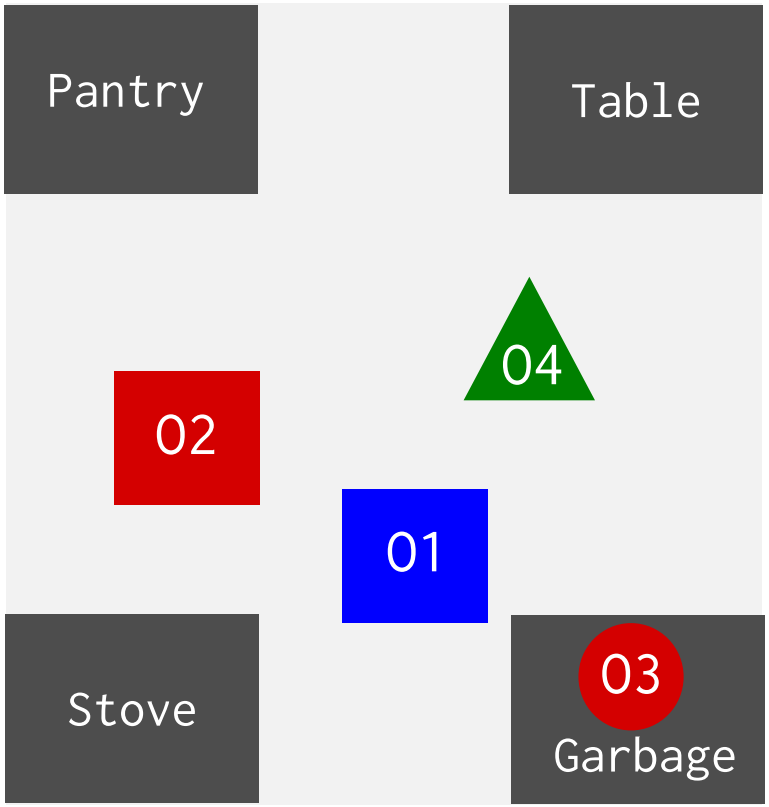
Human: Move it to the pantry.

Agent: Which action should I take?

Human: Close the pantry.

Agent: Which action should I take?

Human: You are done.



Communication is situated, contextual, interactive.

Situated Interactive Instruction

Human: Store the blue rectangle.

Agent: What is the goal of the action?

Human: The goal is the rectangle in the pantry and the pantry is closed.

Agent: I do not know the preposition in.

Human: The red circle is in the garbage.

Agent: OK.

Agent: Which action should I take?

Human: Open the pantry.

Agent: Which action should I take?

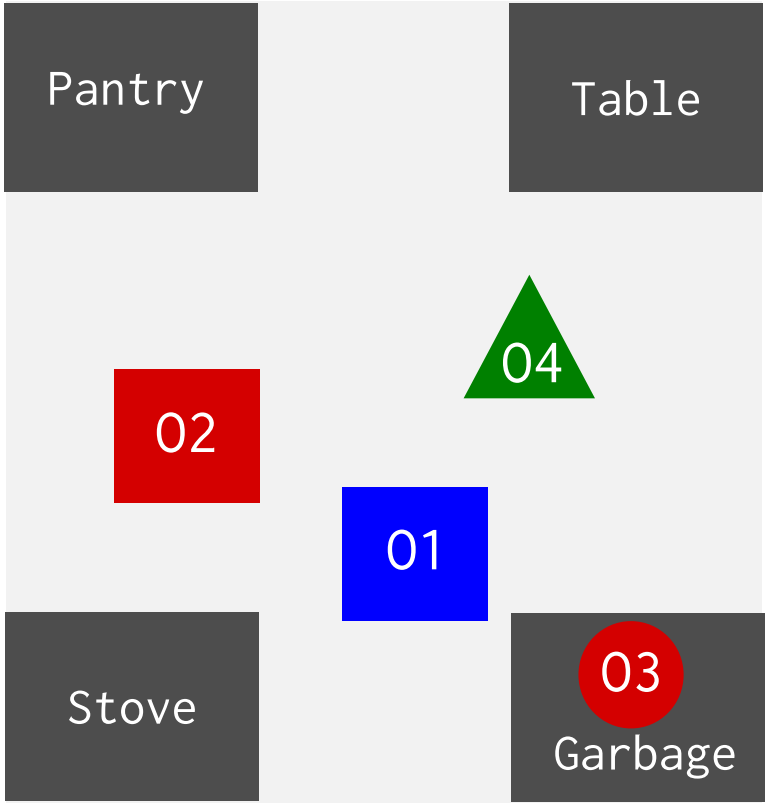
Human: Move it to the pantry.

Agent: Which action should I take?

Human: Close the pantry.

Agent: Which action should I take?

Human: You are done.



Communication is situated, contextual, interactive.

Situated Interactive Instruction

Human: Store the blue rectangle.

Agent: What is the goal of the action?

Human: The goal is the rectangle in the pantry and the pantry is closed.

Agent: I do not know the preposition in.

Human: The red circle is in the garbage.

Agent: OK.

Agent: Which action should I take?

Human: Open the pantry.

Agent: Which action should I take?

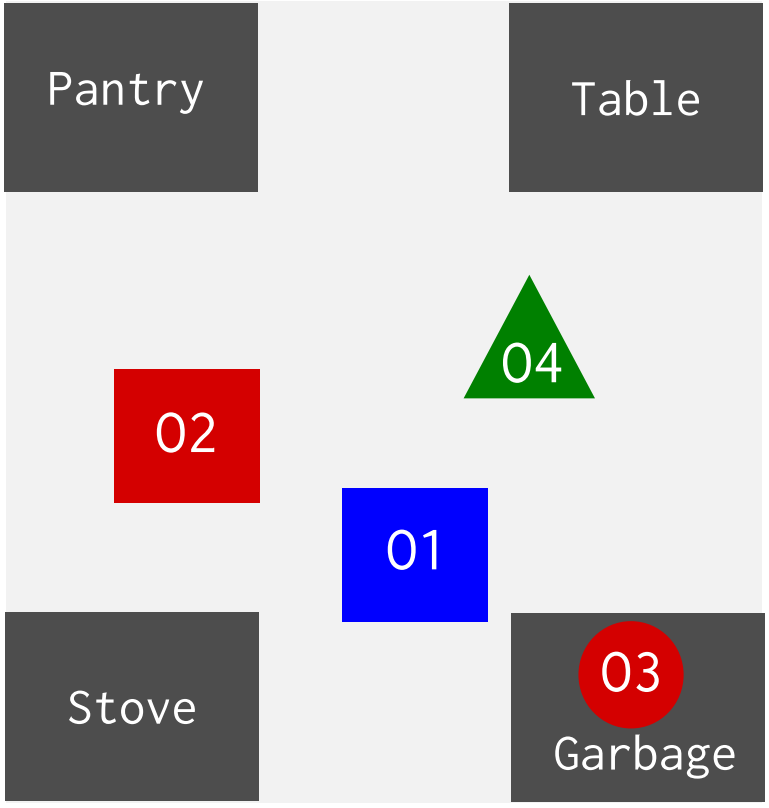
Human: Move it to the pantry.

Agent: Which action should I take?

Human: Close the pantry.

Agent: Which action should I take?

Human: You are done.



Communication is situated, contextual, interactive.

Situated Comprehension Model

- Desiderata
  - translate amodal symbols to modal knowledge
  - integrate non-linguistic context, knowledge
  - inform interaction
- Challenges
  - mixed representations
  - continual knowledge acquisition

Hypothesis

Embodied Psycholinguistics - Glenberg and Robertson (1999), Barsalou (1999), Zwaan (2003)
Cognitive Systems - Scheutz et al. (2004), Livingston and Riesbeck (2009)

- Linguistic communication is reference
- Speaker/hearer have a common ground
  - shared perceptions
  - common-sense knowledge
  - similar experiences
- Linguistic features are cues to search common ground
- Language specifies scene, knowledge fills up details

The Indexical Model

inspired by the Indexical Hypothesis: Glenberg and Robertson (1999)

imperative sentences: task-oriented comprehension
three steps

1. Index words and phrases to referents
   - *NN/ADJ* $\rightarrow$ perceptual classification
   - *NP* $\rightarrow$ set of objects
   - *V* $\rightarrow$ task

   ...
2. Extract domain-knowledge associated with referents
   - pre-encoded or learned (Mohan et al. 2012)
3. Mesh syntactical, environmental, knowledge-based constraints

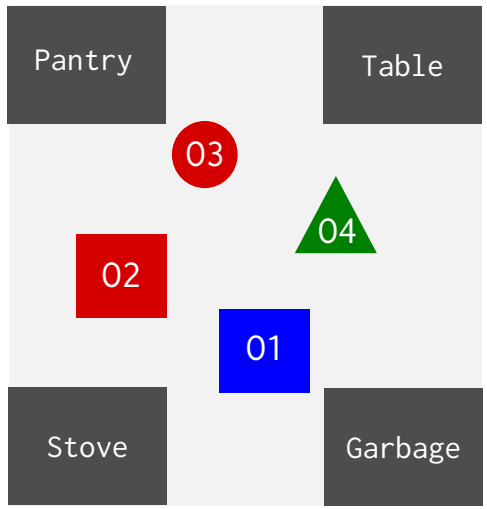
The Indexical Model

Step 1: Index components

Move the blue object to the right of the pantry.

indexical maps

perceptual knowledge



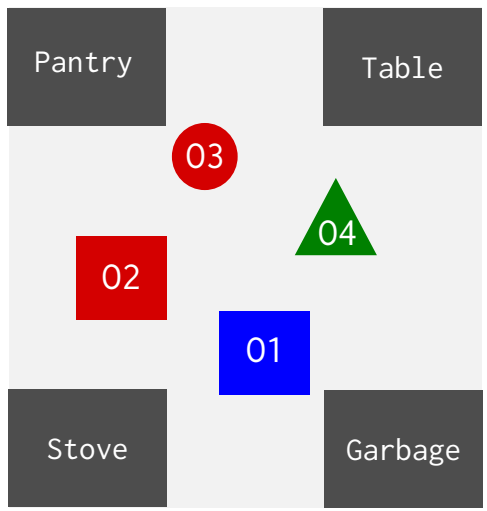
The Indexical Model

Step 1: Index components

Move the blue object to the right of the pantry.

indexical maps

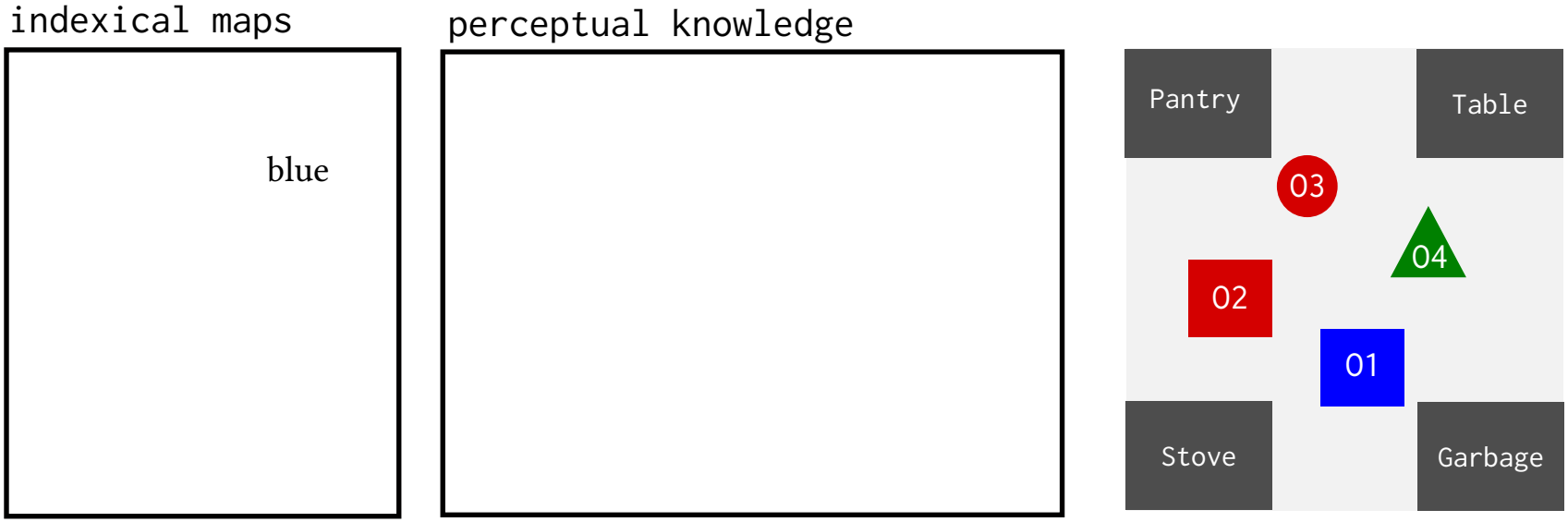
perceptual knowledge



The Indexical Model

Step 1: Index components

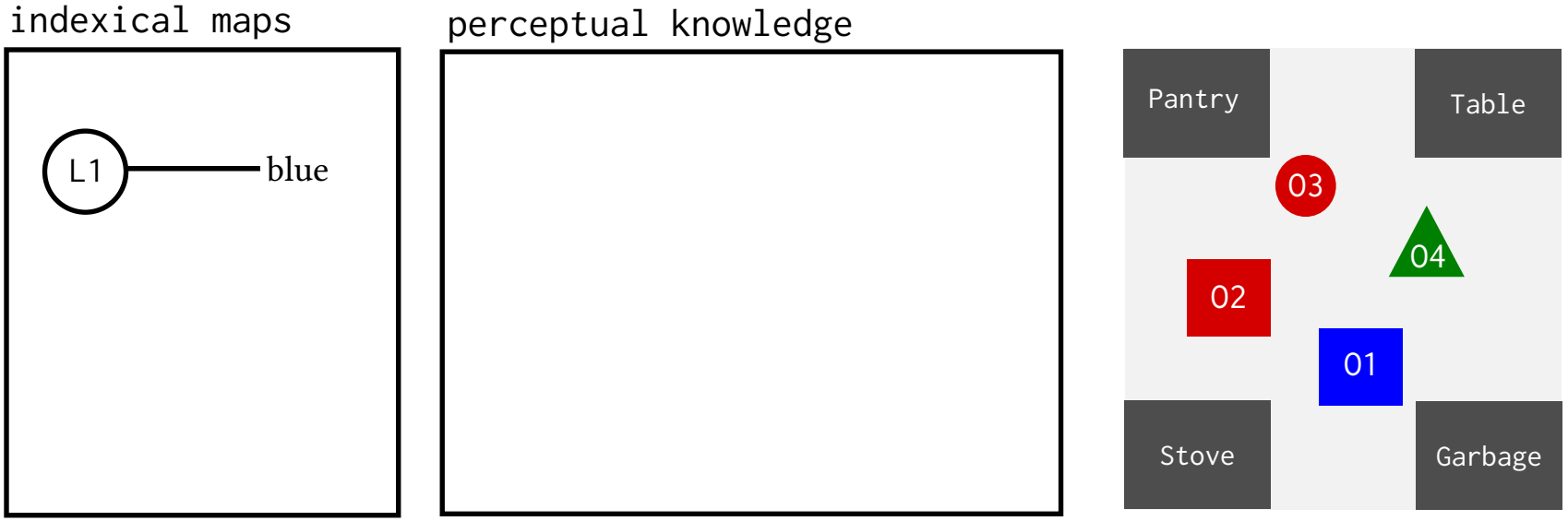
Move the blue object to the right of the pantry.



The Indexical Model

Step 1: Index components

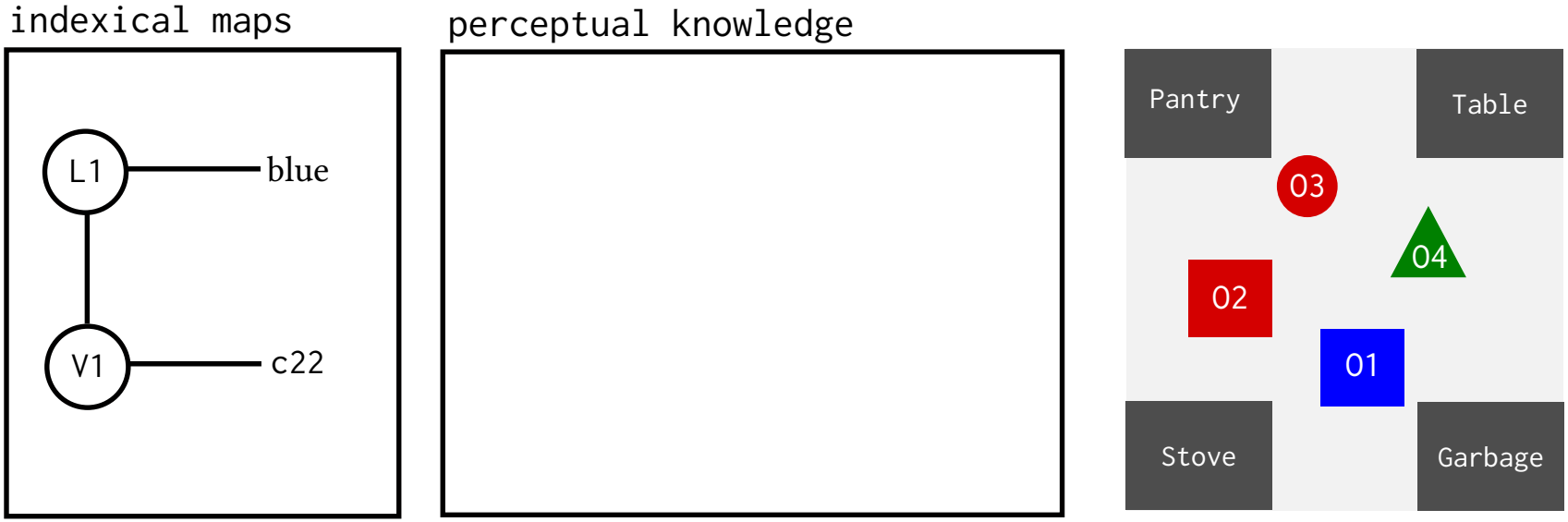
Move the blue object to the right of the pantry.



The Indexical Model

Step 1: Index components

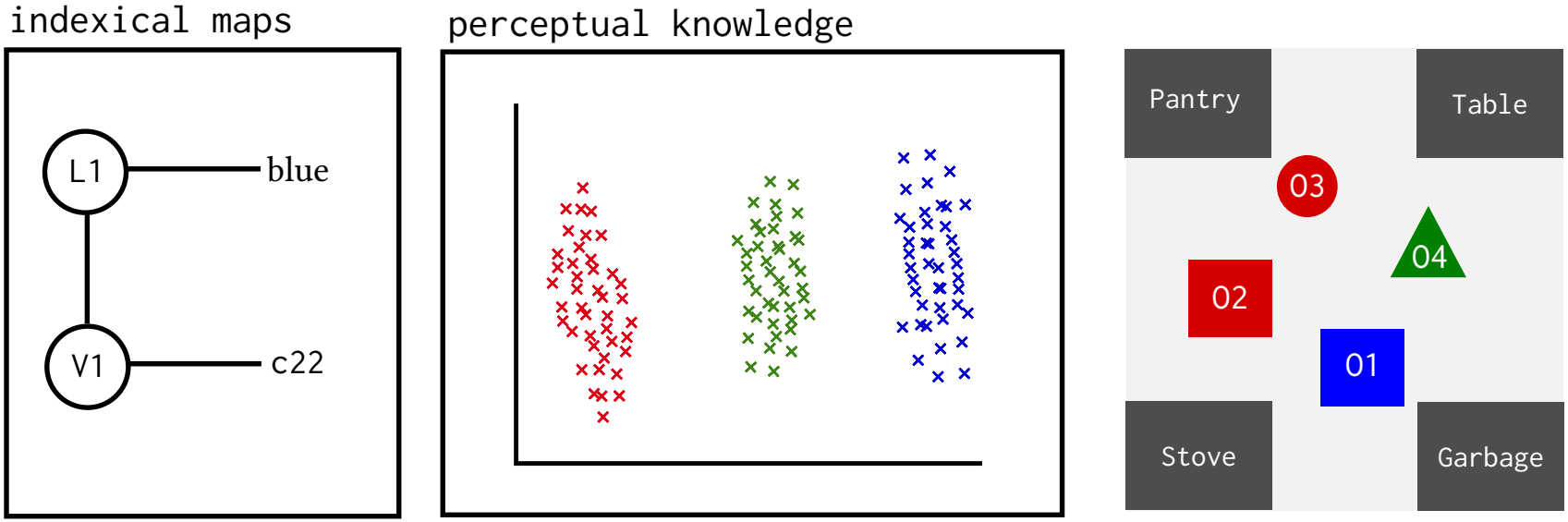
Move the blue object to the right of the pantry.



The Indexical Model

Step 1: Index components

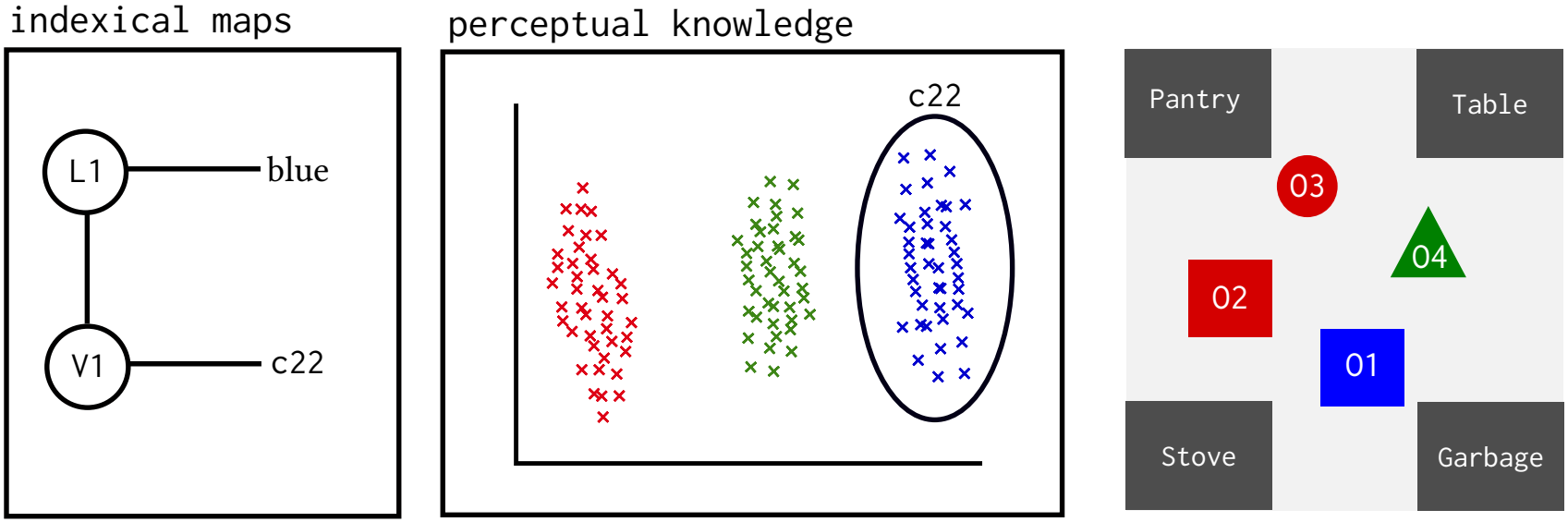
Move the blue object to the right of the pantry.



The Indexical Model

Step 1: Index components

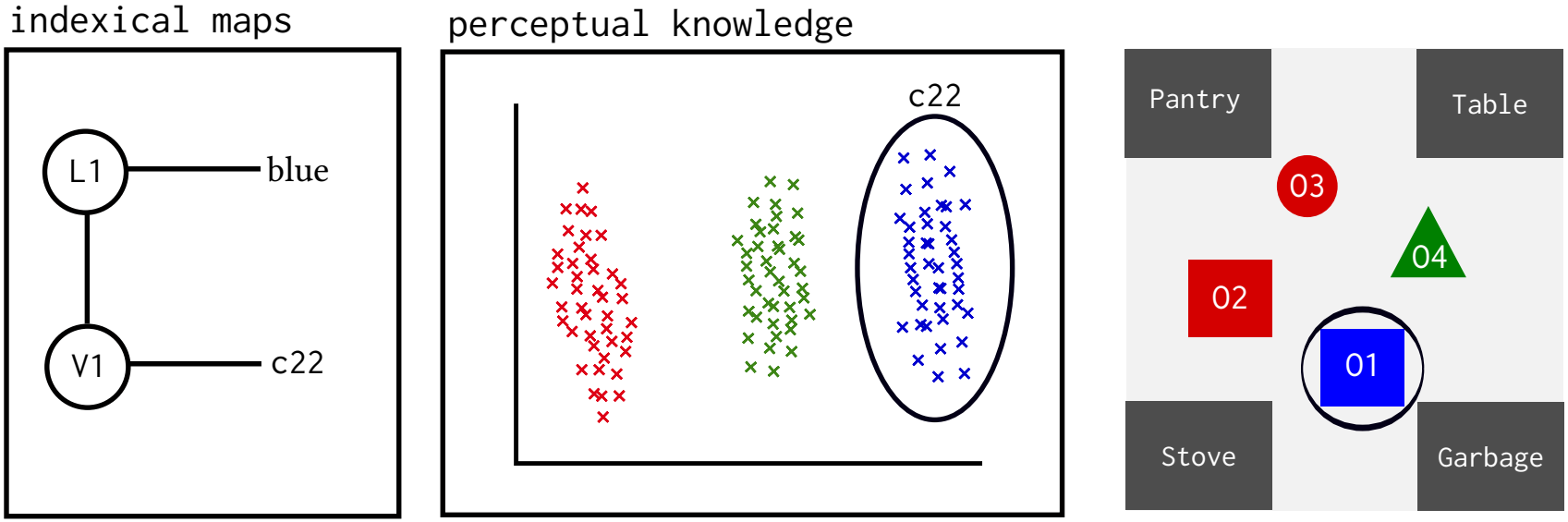
Move the blue object to the right of the pantry.



The Indexical Model

Step 1: Index components

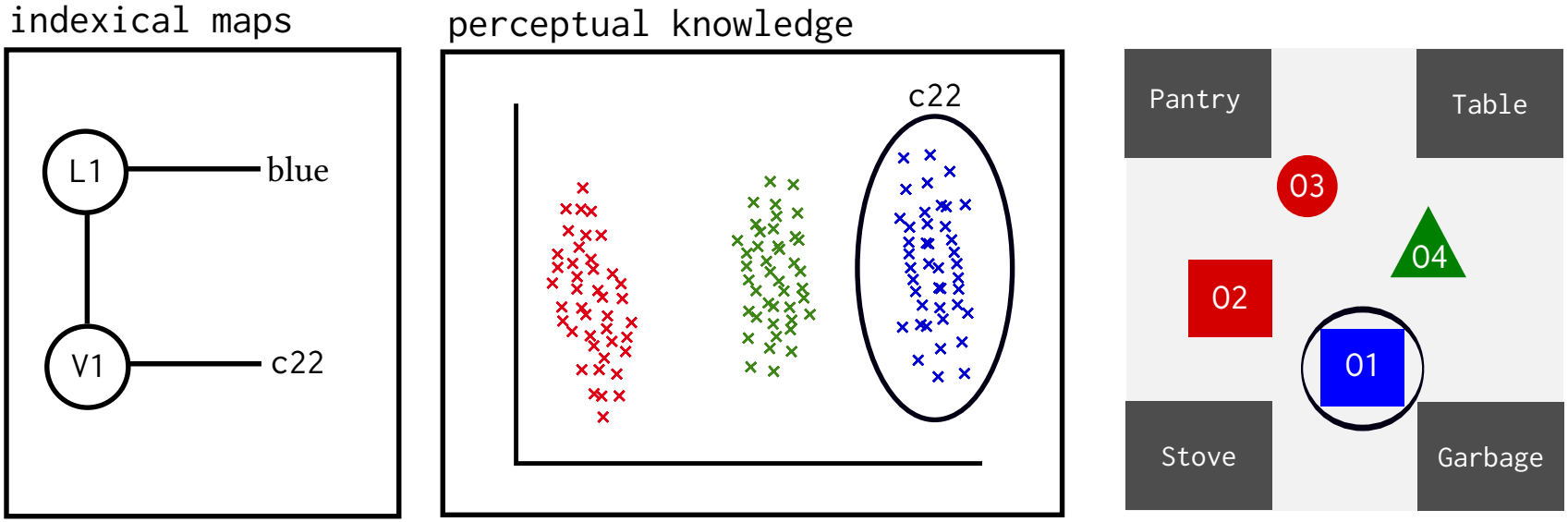
Move the blue object to the right of the pantry.



The Indexical Model

Step 1: Index components

Move the blue object to the right of the pantry.
01



The Indexical Model

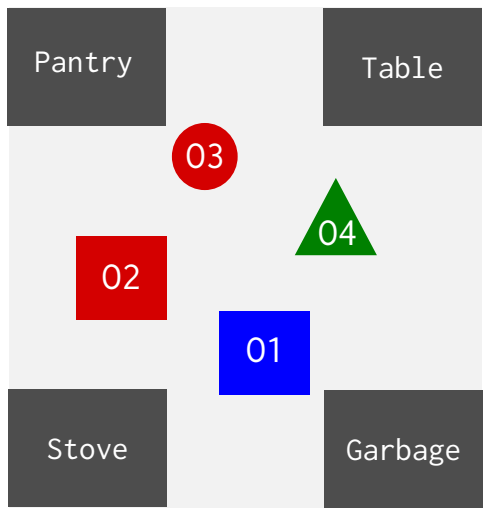
Step 1: Index components

Move the blue object to the right of the pantry.

01

indexical maps

spatial knowledge

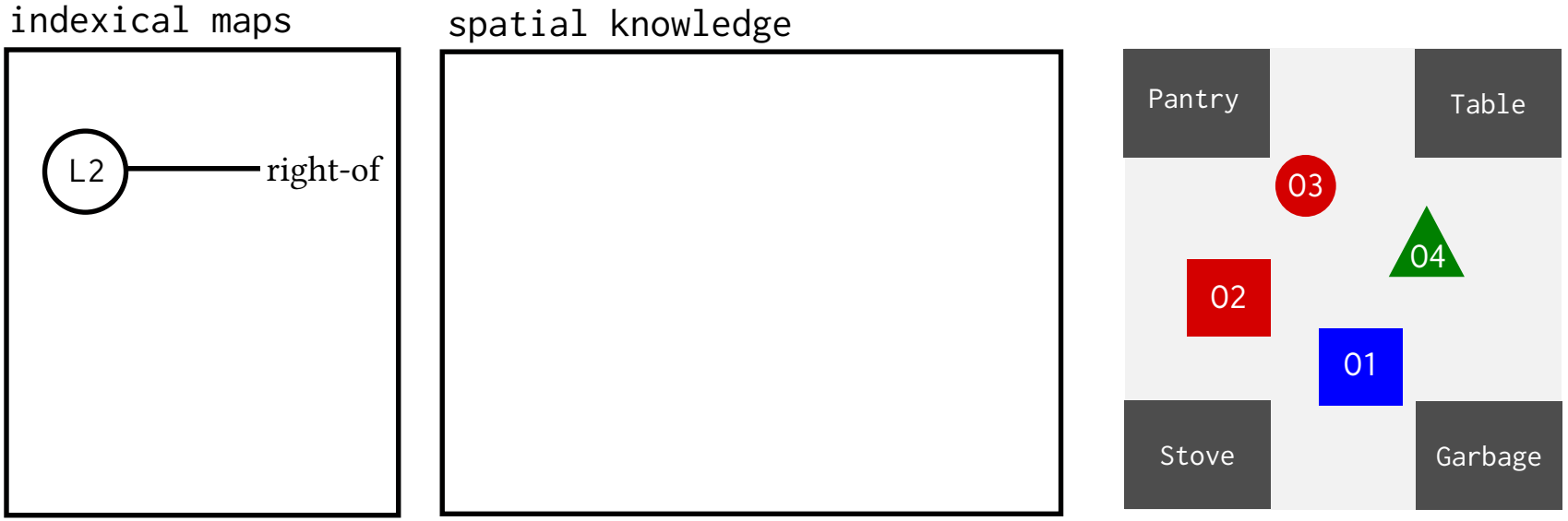


The Indexical Model

Step 1: Index components

Move the blue object to the right of the pantry.

01

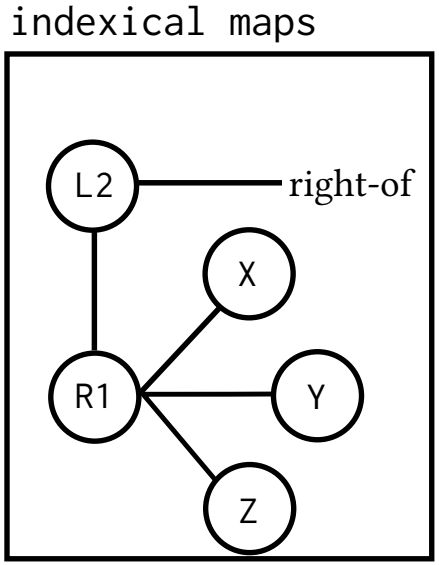


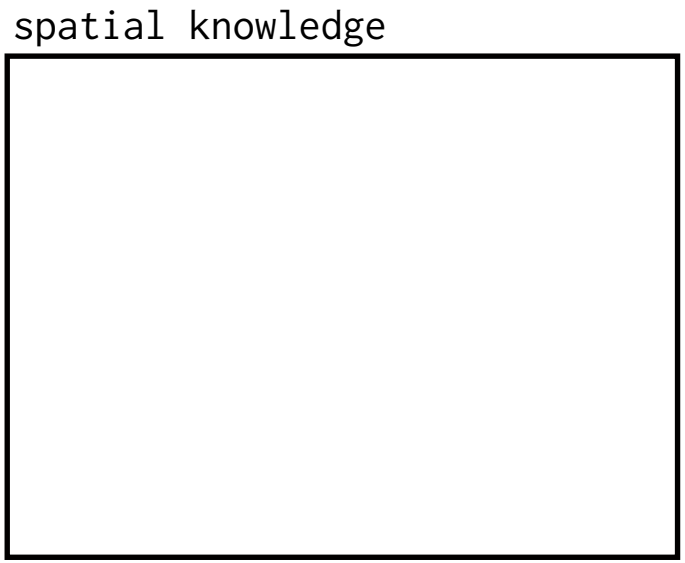
The Indexical Model

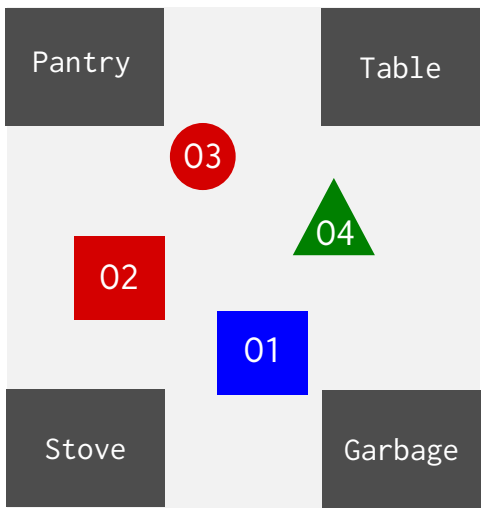
Step 1: Index components

Move the blue object to the right of the pantry.

01





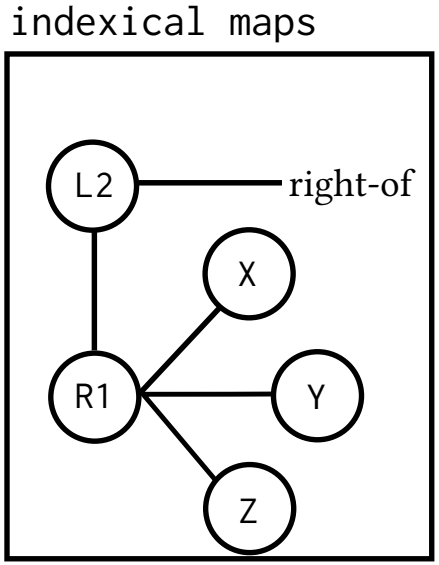


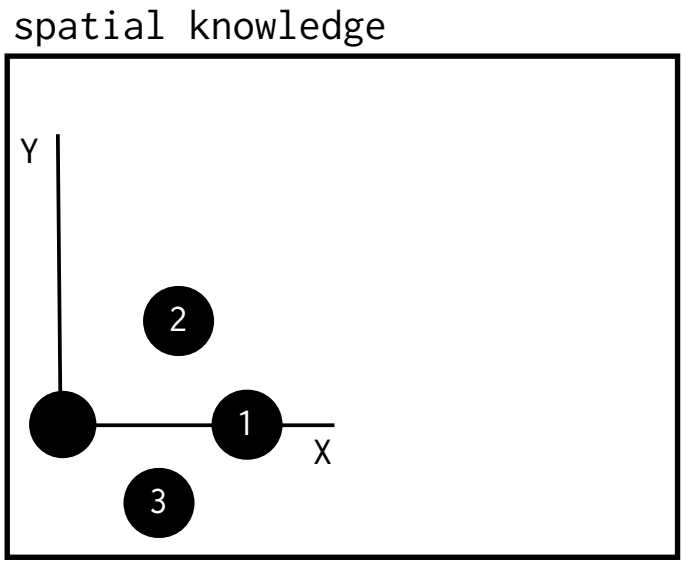
The Indexical Model

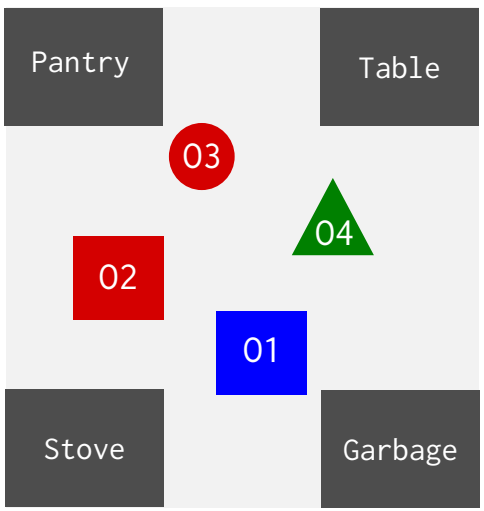
Step 1: Index components

Move the blue object to the right of the pantry.

01





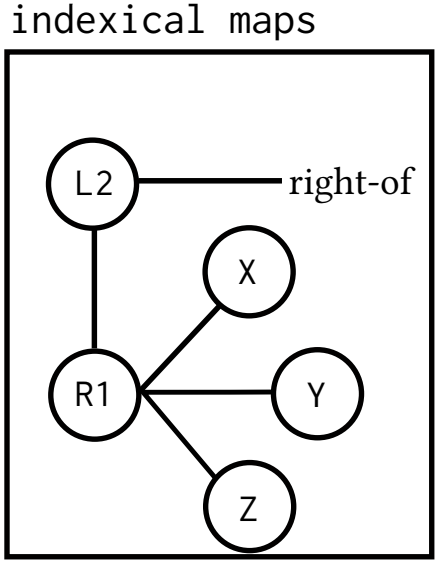


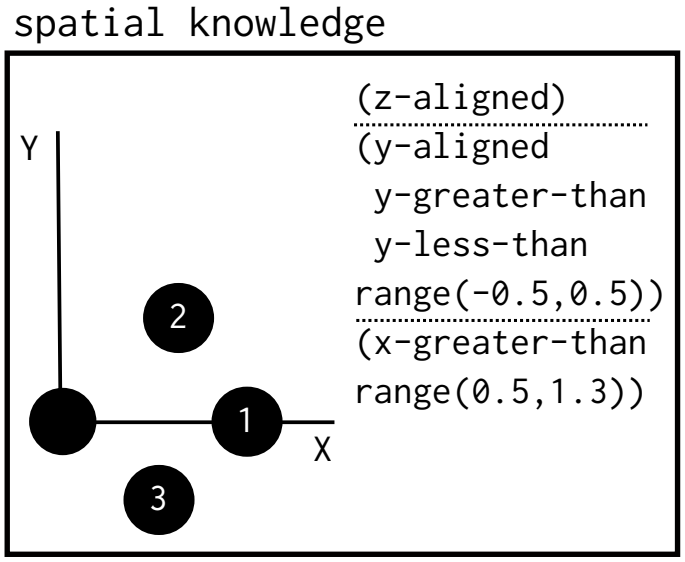
The Indexical Model

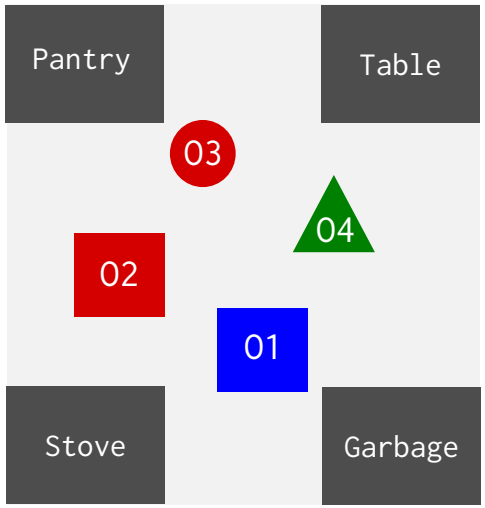
Step 1: Index components

Move the blue object to the right of the pantry.

01



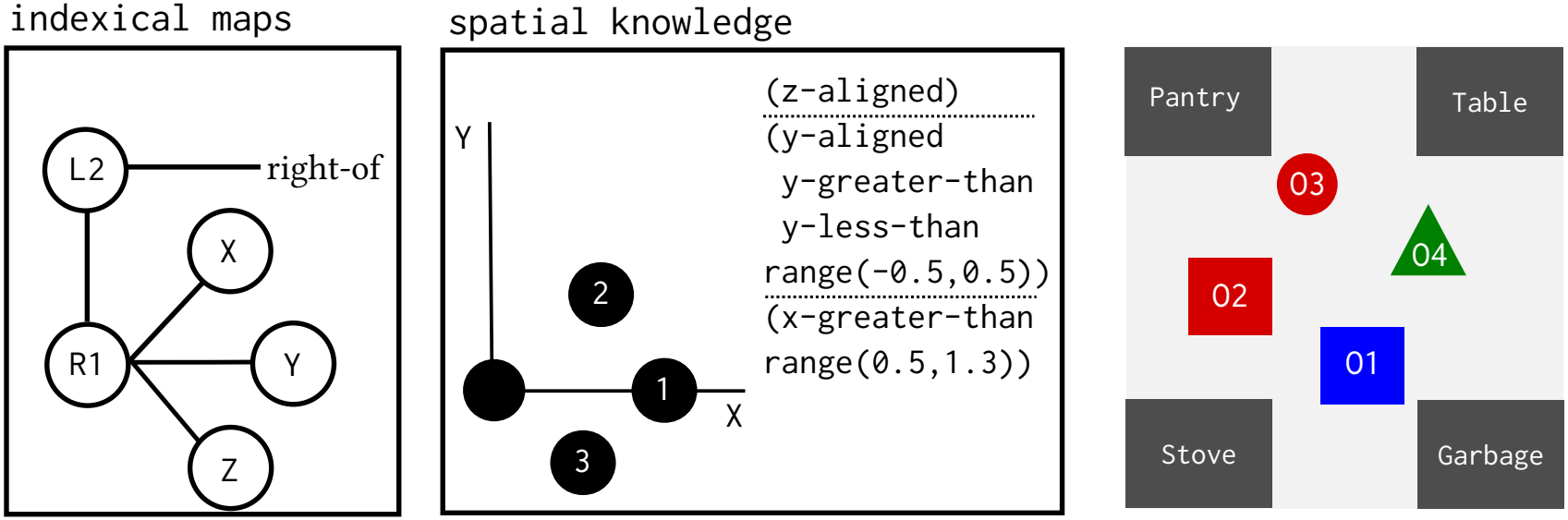




The Indexical Model

Step 1: Index components

Move the blue object to the right of the pantry.
O1 R1



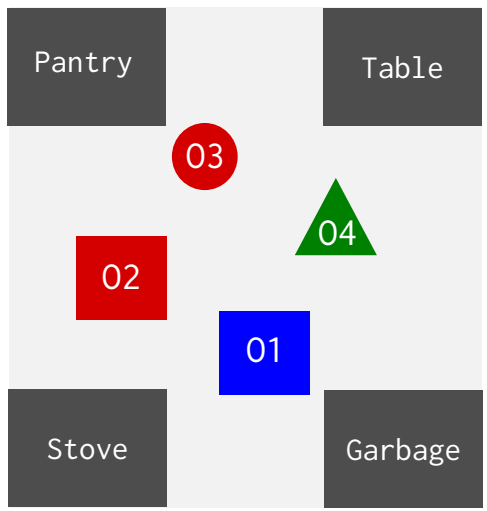
The Indexical Model

Step 1: Index components

Move *the blue object to the right of the pantry.*
O1 R1

indexical maps

task knowledge

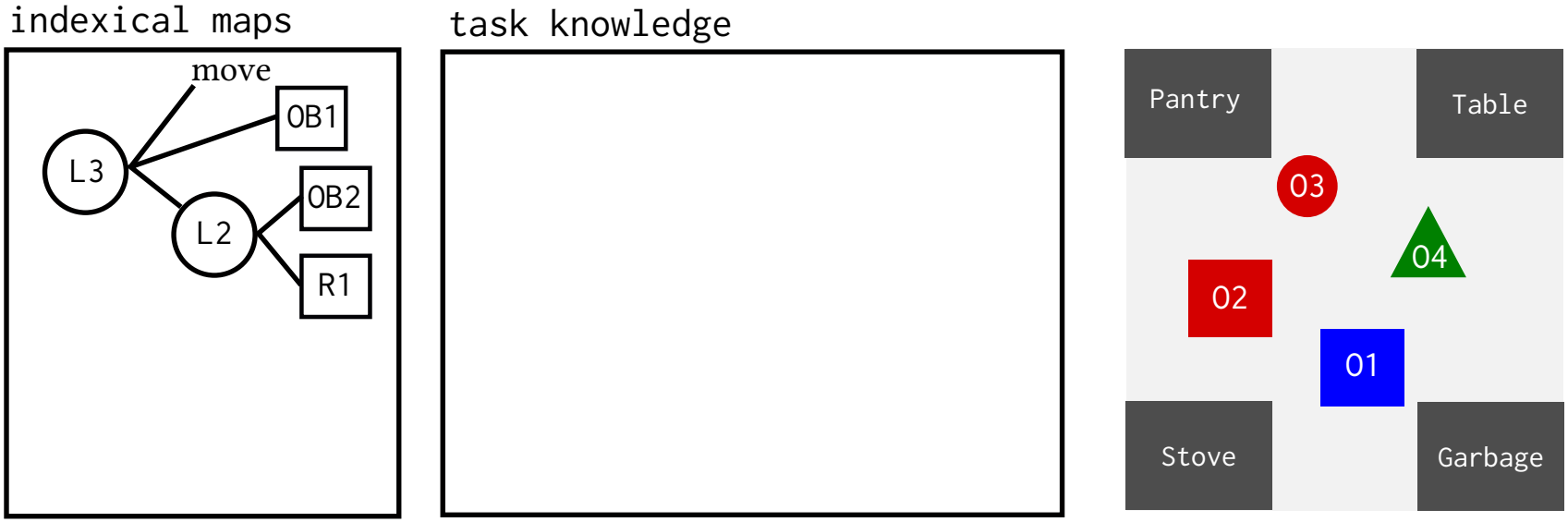


The Indexical Model

Step 1: Index components

Move the blue object to the right of the pantry.

O1 R1

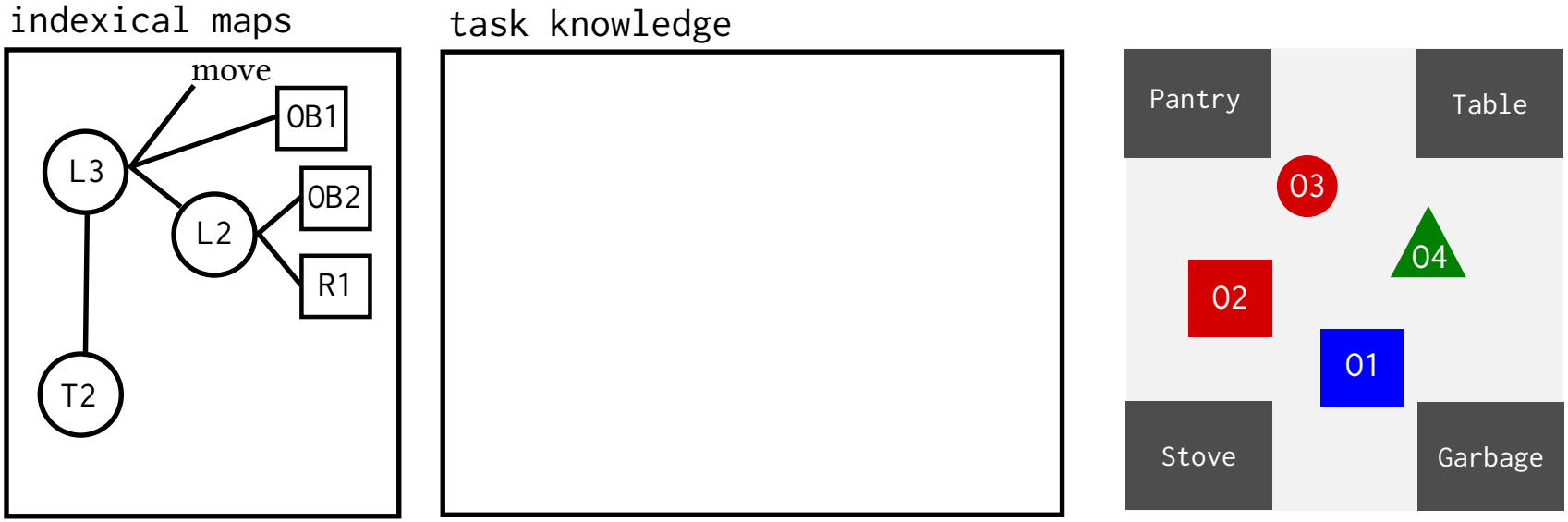


The Indexical Model

Step 1: Index components

Move the blue object to the right of the pantry.

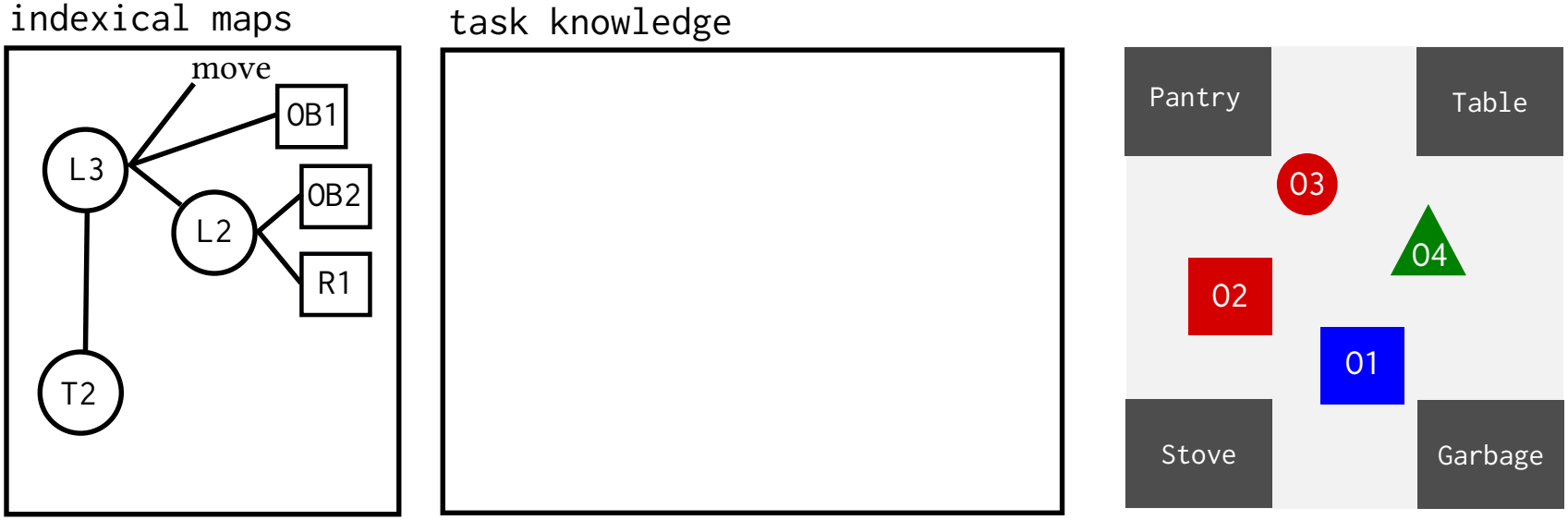
O1 R1



The Indexical Model

Step 1: Index components

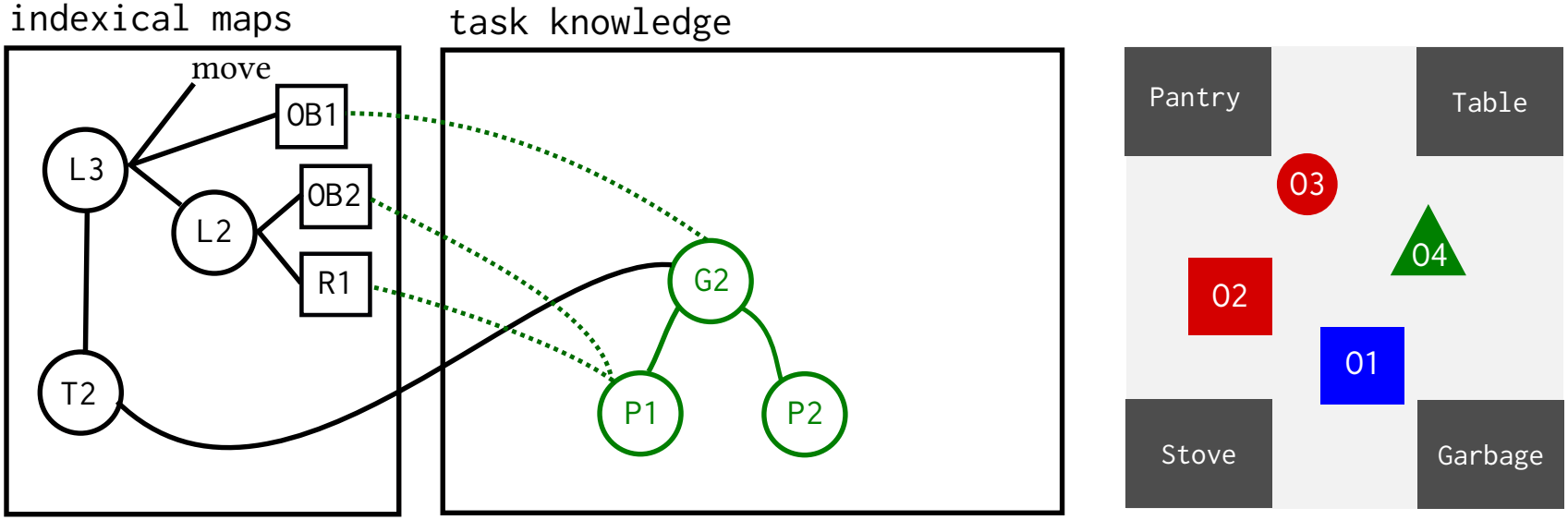
Move the blue object to the right of the pantry.
T2 O1 R1



The Indexical Model

Step 1: Index components

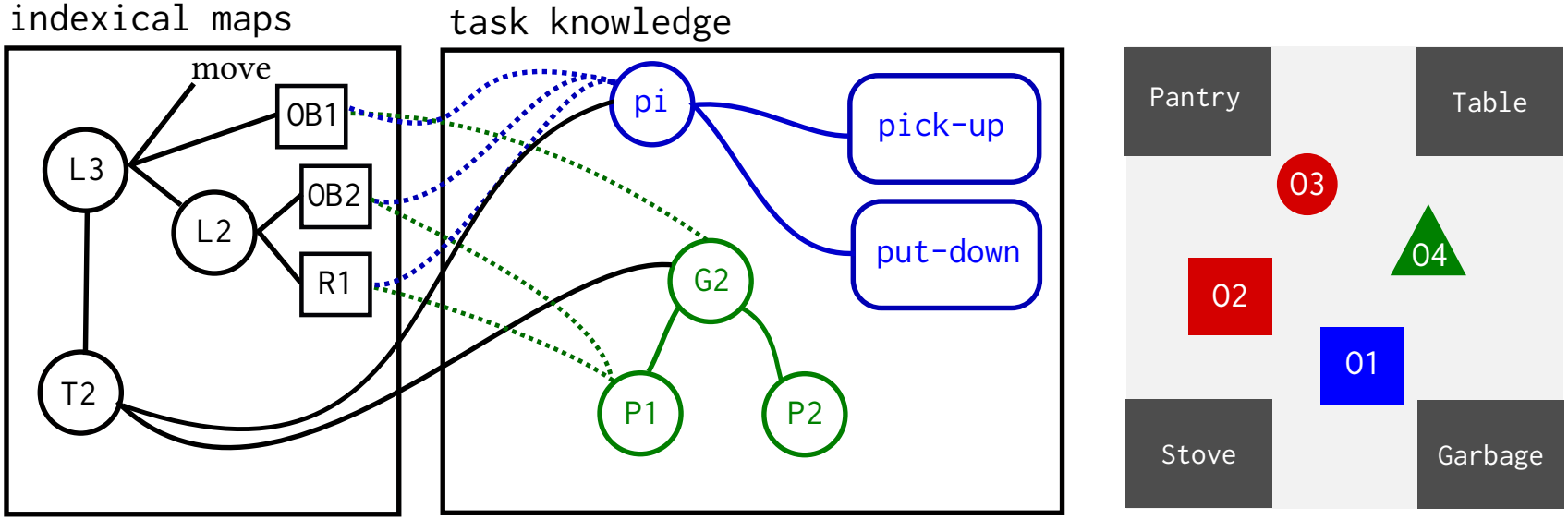
Move the blue object to the right of the pantry.
T2 O1 R1



The Indexical Model

Step 1: Index components

Move the blue object to the right of the pantry.
T2 O1 R1

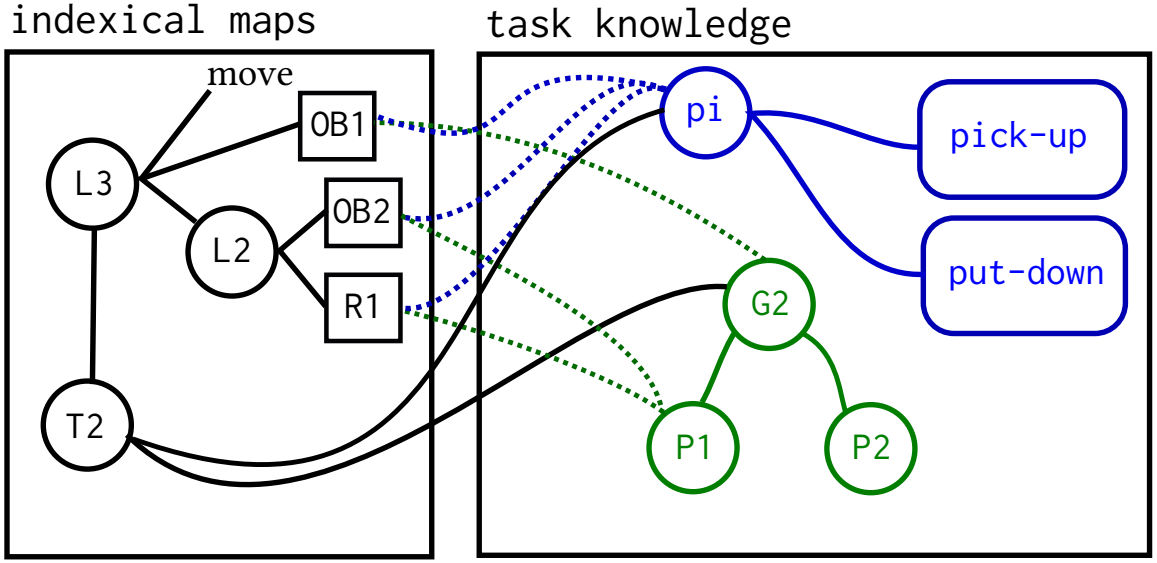


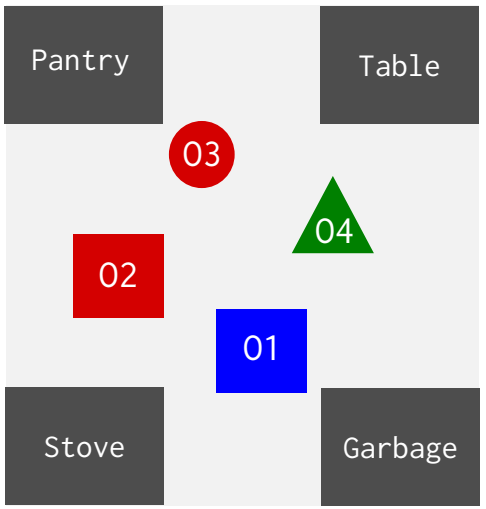
The Indexical Model

Step 2: Extract and instantiate domain knowledge

Move the blue object to the right of the pantry.
T2 O1 R1

```
I  : T2(O1,(R1,pantry); G2(O1,(R1,pantry)); pi(O1,R1,pantry))
```





The Indexical Model

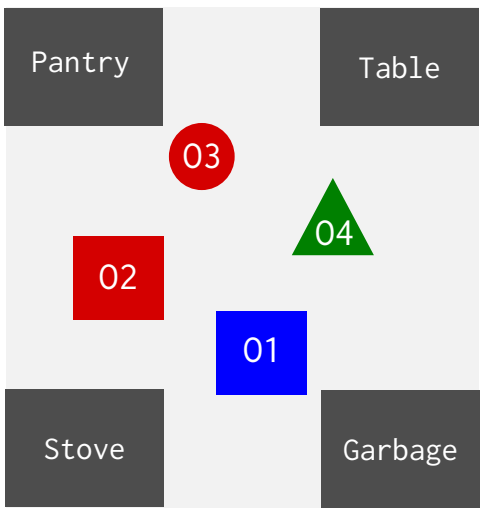
Step 3: Mesh constraints

Move the blue object to the right of the pantry.

T2 O1 R1

```
I  : T2(O1,(R1,pantry); G2(O1,(R1,pantry)); pi(O1,R1,pantry))
```

```
available [A]:
T2(O1,(R1,pantry); G2(O1,(R1,pantry)); pi(O1,R1,pantry))
T3(O1; G3(O1,(IN,pantry)); pi(O1,IN,pantry))
T4 ...
...
```



The Indexical Model

Step 3: Mesh constraints

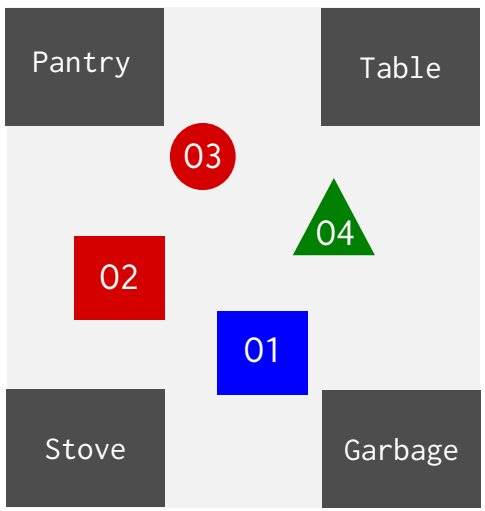
Move the blue object to the right of the pantry.

T2 O1 R1

```
I  : T2(O1,(R1,pantry); G2(O1,(R1,pantry)); pi(O1,R1,pantry))
```

```
available [A]:
T2(O1,(R1,pantry); G2(O1,(R1,pantry)); pi(O1,R1,pantry))
T3(O1; G3(O1,(IN,pantry)); pi(O1,IN,pantry))
T4 ...
...


execute [I ∩ A]
T2(O1,(R1,pantry); G2(O1,(R1,pantry)); pi(O1,R1,pantry))
```



Addressing Complexities

1. Ambiguity: referring expressions
2. Incomplete specification: unexpressed verb arguments

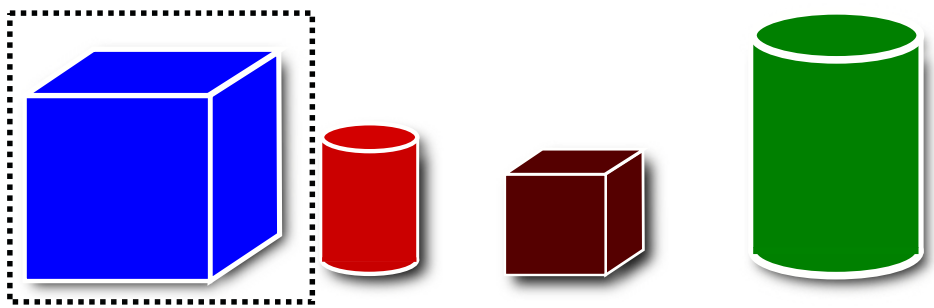
Referring Expressions

Referring expressions are situational {*it*, *this cube*, *that*, *the large cube* }

Referring Expressions

Referring expressions are situational {*it*, *this cube*, *that*, *the large cube*}

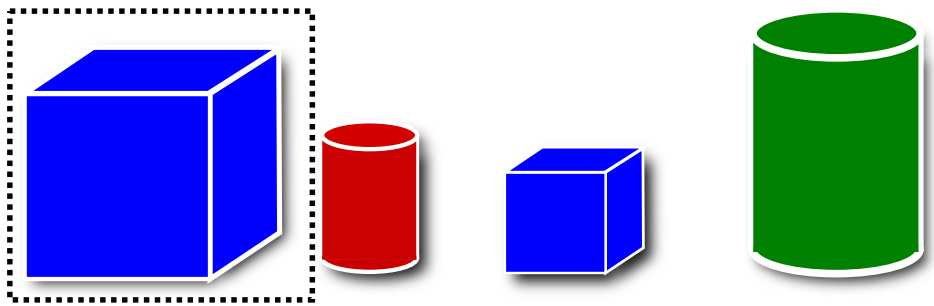
Pick up the blue cube.



Referring Expressions

Referring expressions are situational {*it*, *this cube*, *that*, *the large cube*}

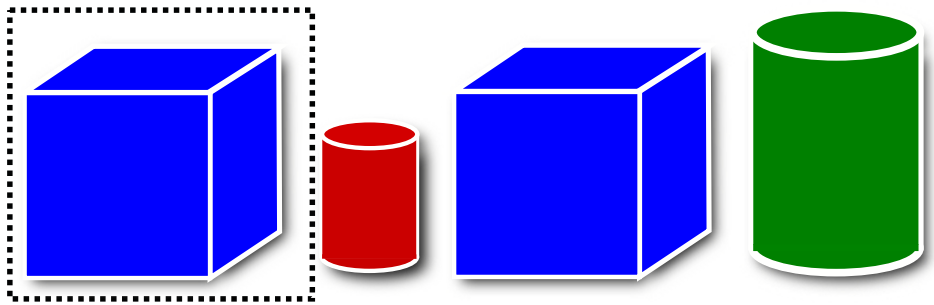
Pick up the large, blue cube.



Referring Expressions

Referring expressions are situational {*it*, *this cube*, *that*, *the large cube*}

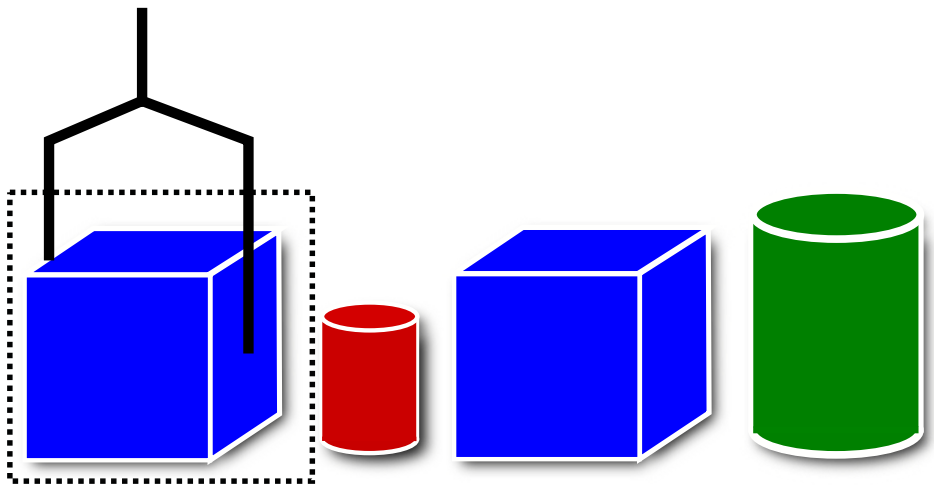
Pick up the cube on the left of the red cylinder.



Referring Expressions

Referring expressions are situational {*it*, *this cube*, *that*, *the large cube*}

Put it down.



Use of Referring Expressions

The Givenness Hierarchy: Gundel (1993)

- Salience based cognitive-status of referent objects
  - discourse, action, surprise
- Use of referring expression (RE) is dependent on salience
  - more informative RE for less salient objects and vice-versa
  - efficient communication
- RE and cognitive status

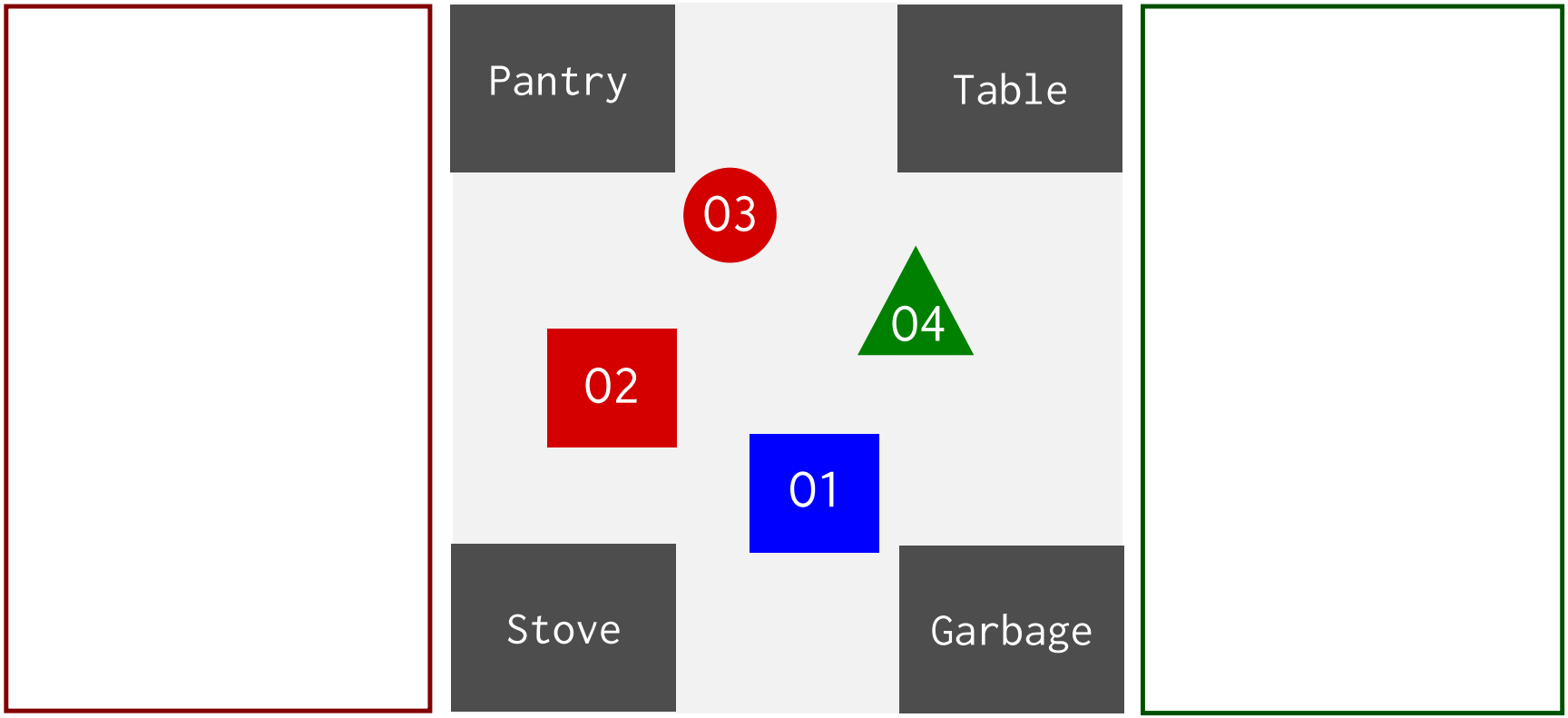
  `in-focus` (*it*) >
  `activated` (*this*, *that rectangle*) >
  `uniquely-identifiable`(*the green rectangle*) >
  `type-identifiable` (*a green rectangle*)

RE Resolution Model

Non-linguistic contexts: 2 dimensions

- informational: perceptual, spatial, task knowledge
- temporal: recency-based activation, dialog attentional focus

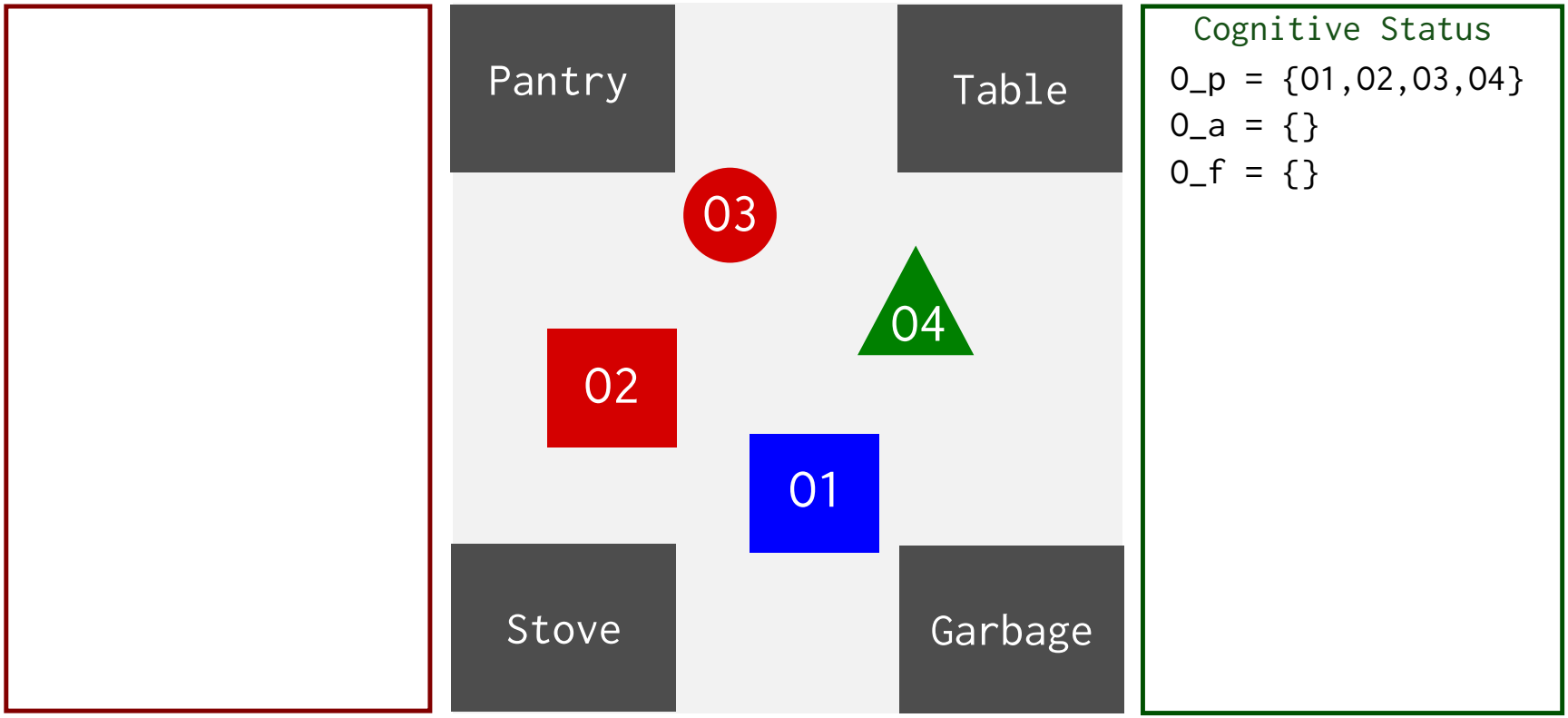
RE Resolution in the Indexical Model



O_a: Soar's architectural activation; O_f: Dialog model based on Grosz and Sidner (1986)

RE Resolution in the Indexical Model

maintain cognitive status

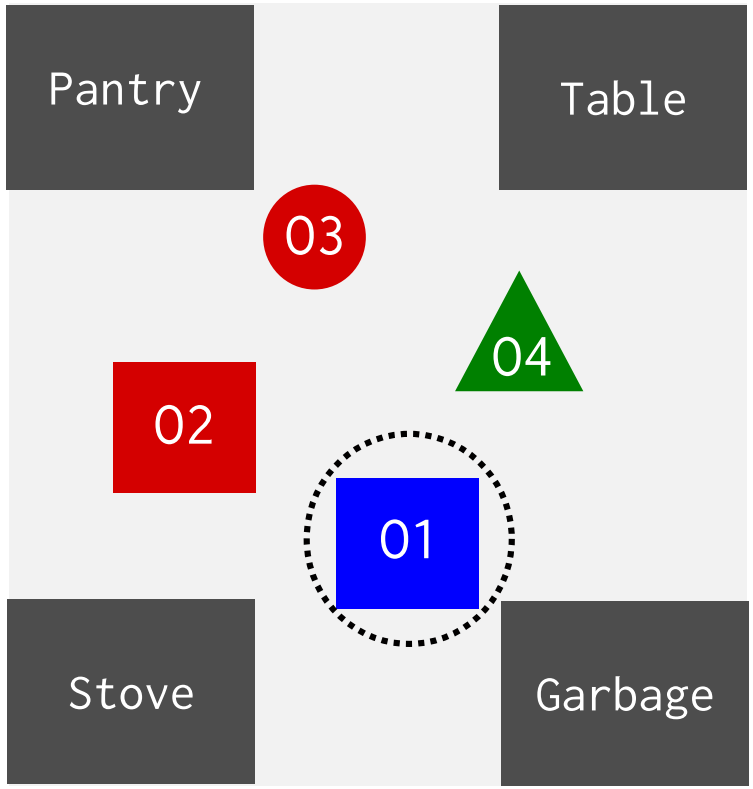


O_a: Soar's architectural activation; O_f: Dialog model based on Grosz and Sidner (1986)

RE Resolution in the Indexical Model

maintain cognitive status

Instructor: This is a blue rectangle.



Cognitive Status

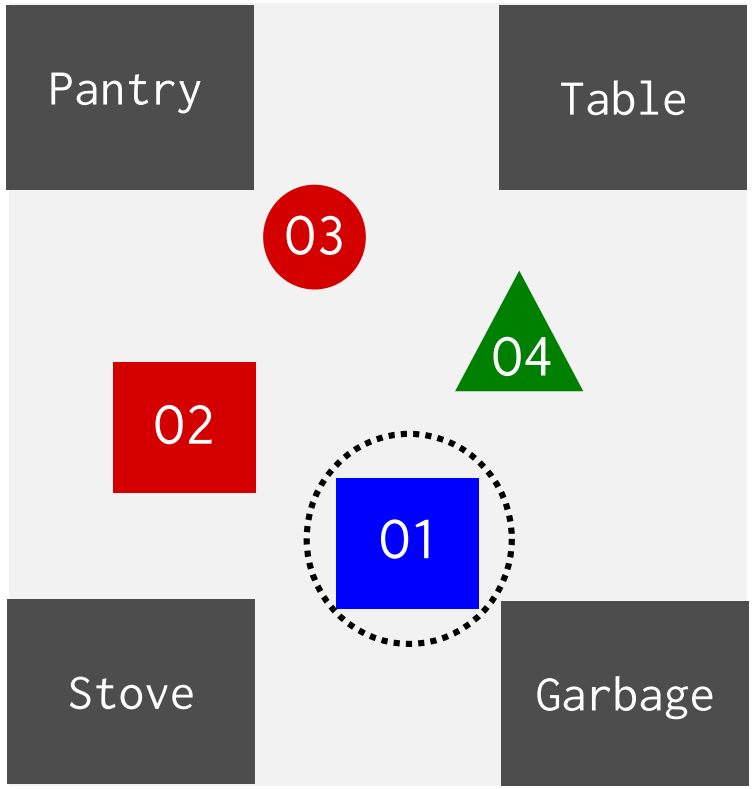
```
O_p = {O1,O2,O3,O4}
O_a = {}
O_f = {}
```

O_a: Soar's architectural activation; O_f: Dialog model based on Grosz and Sidner (1986)

RE Resolution in the Indexical Model

maintain cognitive status

Instructor: This is a blue rectangle.



Cognitive Status

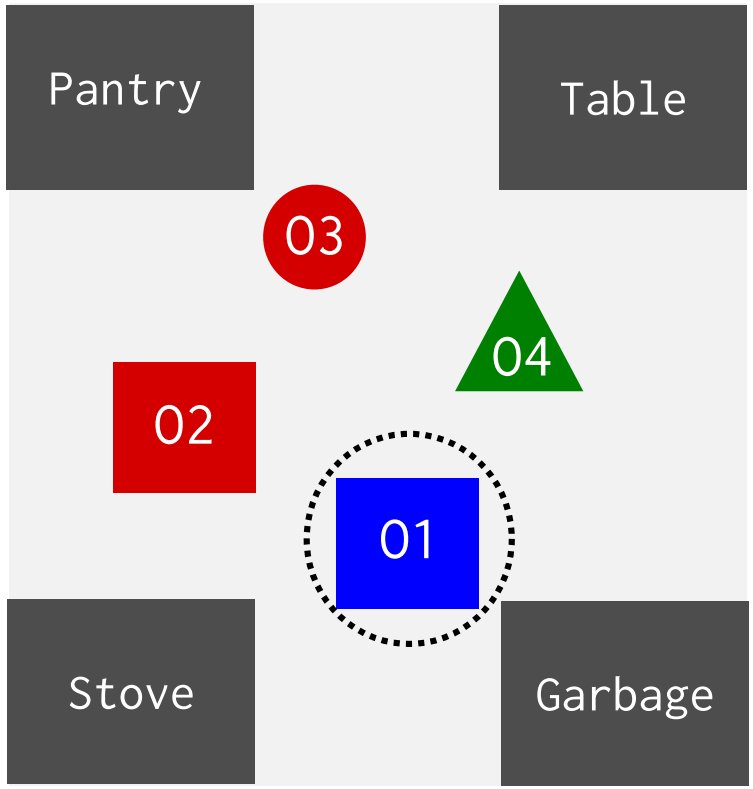
O_p = {O1,O2,O3,O4}
O_a = {O1}
O_f = {}

O_a: Soar's architectural activation; O_f: Dialog model based on Grosz and Sidner (1986)

RE Resolution in the Indexical Model

use GH heuristics to identify the candidates

Instructor: This is a blue rectangle.



Cognitive Status

O_p = {O1,O2,O3,O4}

O_a = {O1}

O_f = {}

GH Heuristic

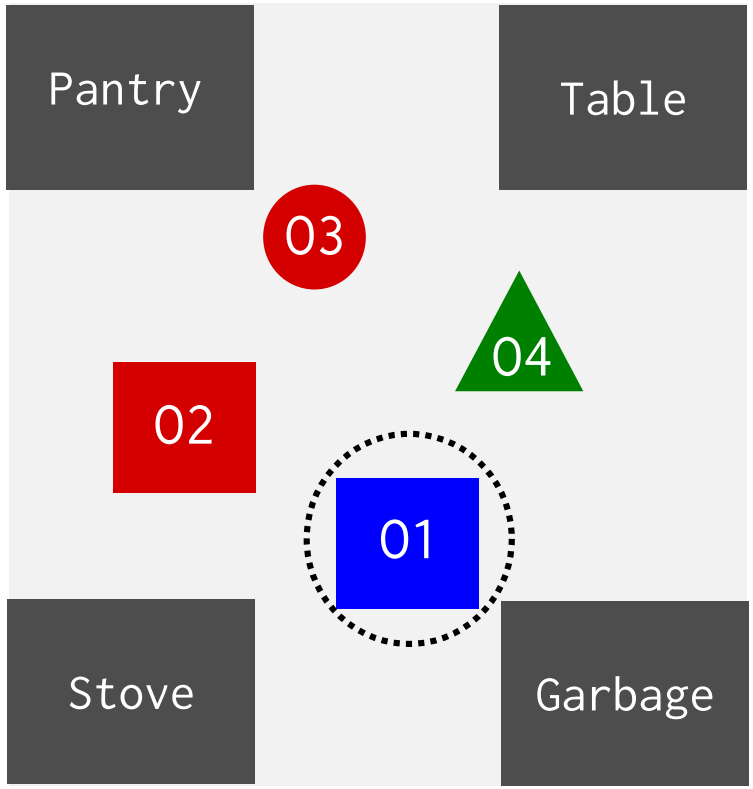
this/that N ⟶ active

O_a: Soar's architectural activation; O_f: Dialog model based on Grosz and Sidner (1986)

RE Resolution in the Indexical Model

use GH heuristics to identify the candidates

Instructor: This is a blue rectangle.



Cognitive Status

O_p = {O1,O2,O3,O4}

O_a = {O1}

O_f = {}

GH Heuristic

this/that N ⟶ active

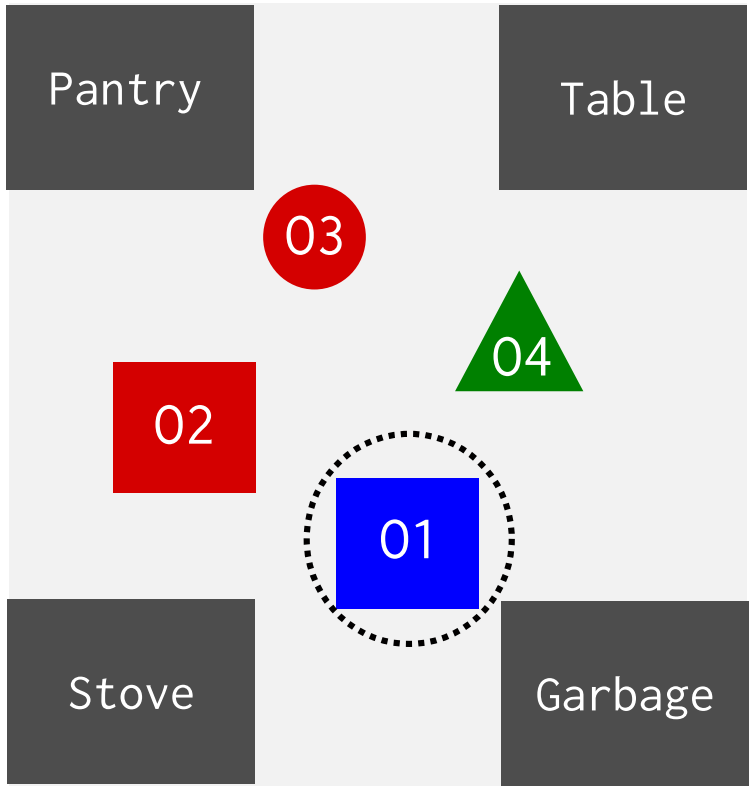
O_c = O_a

O_a: Soar's architectural activation; O_f: Dialog model based on Grosz and Sidner (1986)

RE Resolution in the Indexical Model

resolve

Instructor: This is a blue rectangle.



Cognitive Status

O_p = {01,02,03,04}

O_a = {01}

O_f = {}

GH Heuristic

this/that N ⟶ active

O_c = O_a

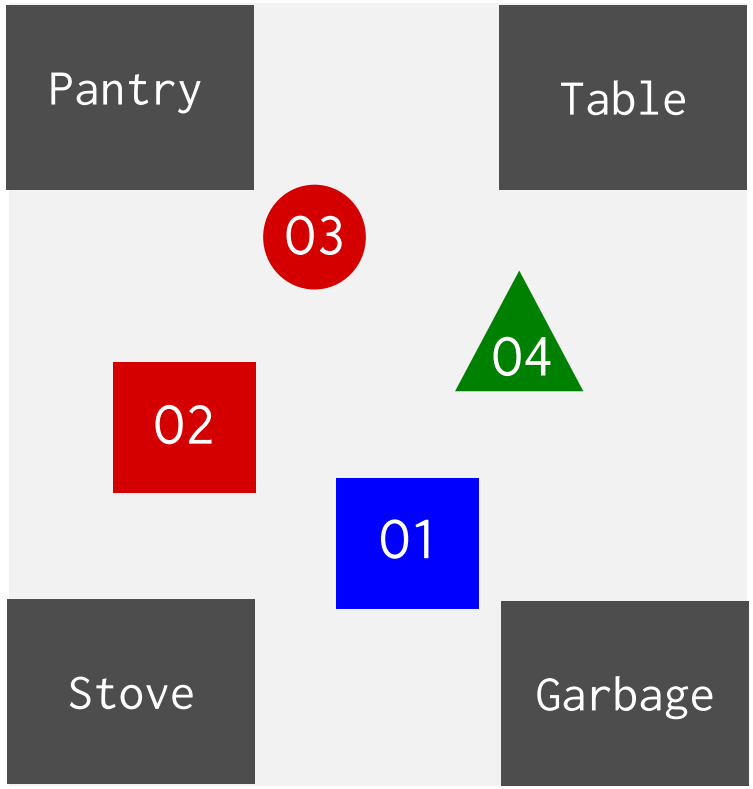
Resolution

this = 01

O_a: Soar's architectural activation; O_f: Dialog model based on Grosz and Sidner (1986)

RE Resolution in the Indexical Model

Instructor: This is a blue rectangle.
Instructor: This is to the right of the red circle.

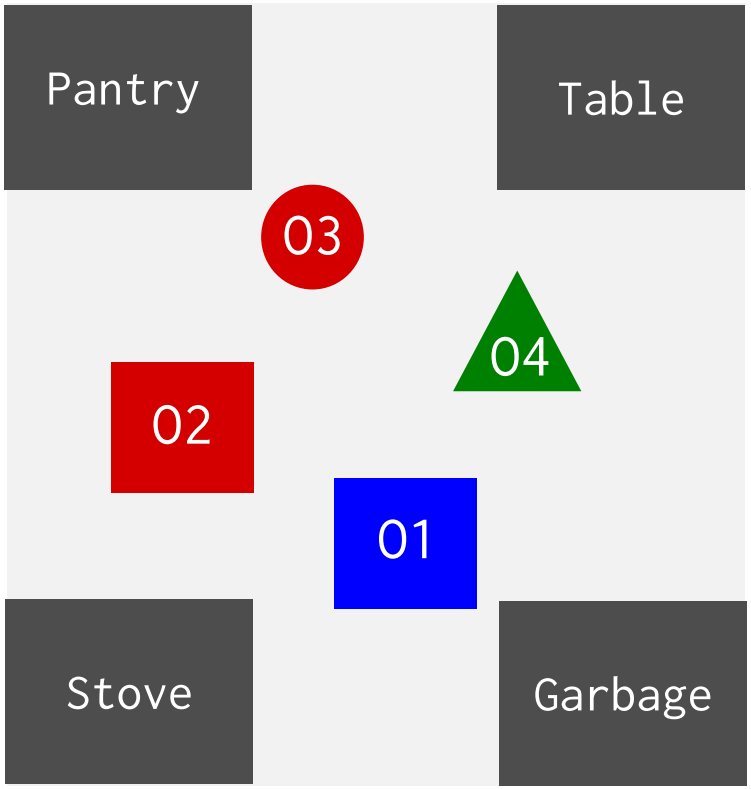


Cognitive Status
O_p = {01,02,03,04}
O_a = {01}
O_f = {}

O_a: Soar's architectural activation; O_f: Dialog model based on Grosz and Sidner (1986)

RE Resolution in the Indexical Model

Instructor: This is a blue rectangle.
Instructor: This is to the right of the red circle.



Cognitive Status

O_p = {O1,O2,O3,O4}
O_a = {O1}
O_f = {}

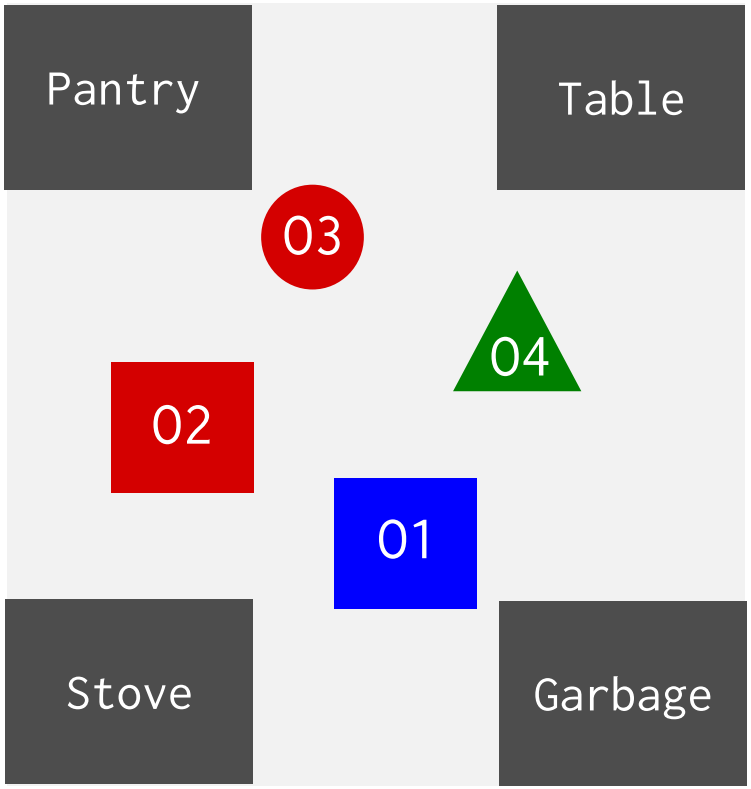
GH Heuristic

this/that N ⟶ active

O_a: Soar's architectural activation; O_f: Dialog model based on Grosz and Sidner (1986)

RE Resolution in the Indexical Model

Instructor: This is a blue rectangle.
Instructor: This is to the right of the red circle.



Cognitive Status

O_p = {O1,O2,O3,O4}
O_a = {O1}
O_f = {}

GH Heuristic

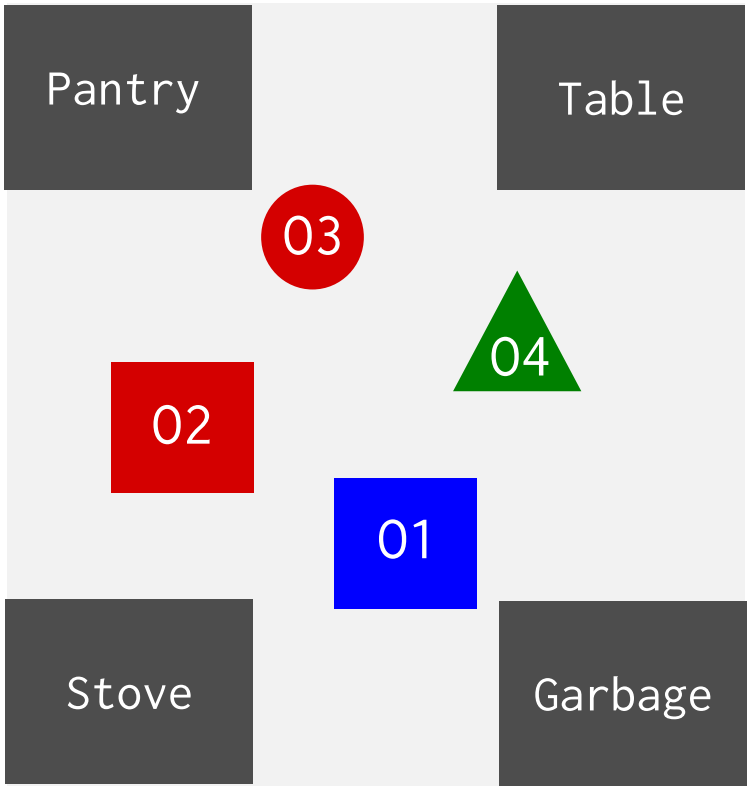
this/that N ⟶ active

O_c = O_a

O_a: Soar's architectural activation; O_f: Dialog model based on Grosz and Sidner (1986)

RE Resolution in the Indexical Model

Instructor: This is a blue rectangle.
Instructor: This is to the right of the red circle.



Cognitive Status

O_p = {O1,O2,O3,O4}
O_a = {O1}
O_f = {}

GH Heuristic

this/that N ⟶ active

O_c = O_a

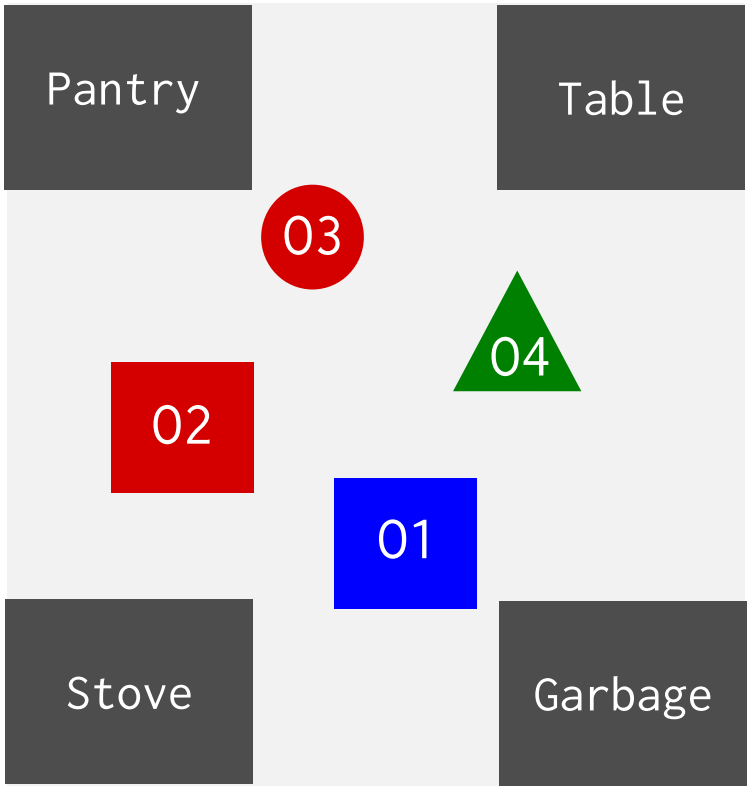
Resolution

this = O1

O_a: Soar's architectural activation; O_f: Dialog model based on Grosz and Sidner (1986)

RE Resolution in the Indexical Model

Instructor: This is a blue rectangle.
Instructor: This is to the right of the red circle.



Cognitive Status

O_p = {O1,O2,O3,O4}
O_a = {O1}
O_f = {}

GH Heuristic

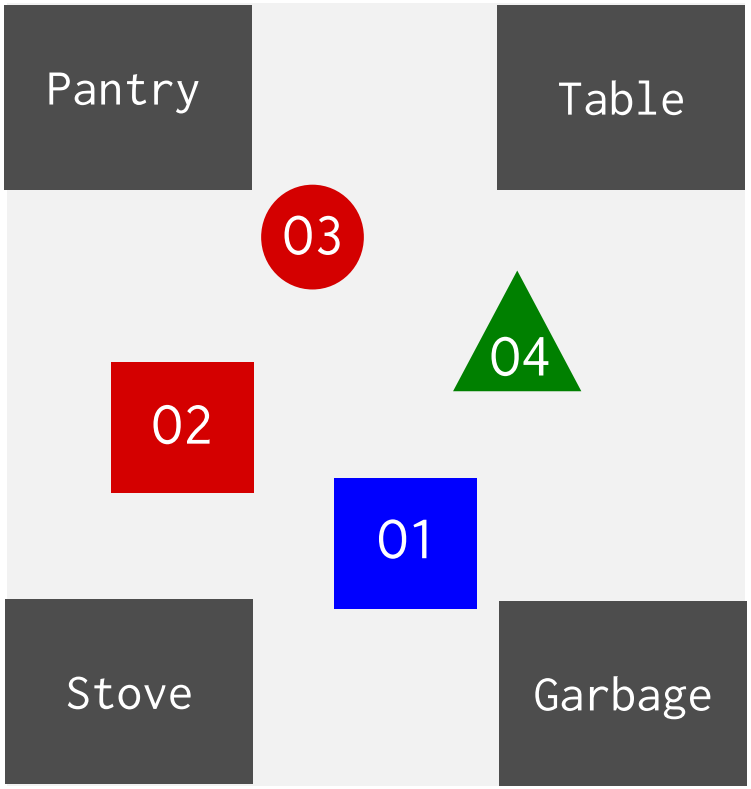
the N ⟶ percept

O_c = O_p

O_a: Soar's architectural activation; O_f: Dialog model based on Grosz and Sidner (1986)

RE Resolution in the Indexical Model

Instructor: This is a blue rectangle.
Instructor: This is to the right of the red circle.



Cognitive Status

O_p = {01,02,03,04}
O_a = {01}
O_f = {}

GH Heuristic

the N ⟶ percept

O_c = O_p

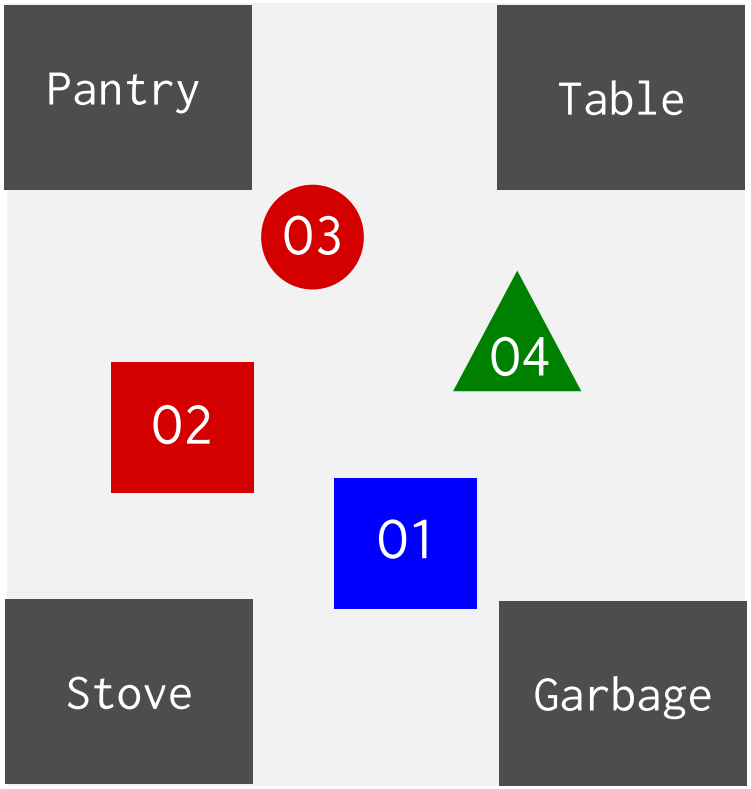
Filter(red)

O_c = {02, 03}

O_a: Soar's architectural activation; O_f: Dialog model based on Grosz and Sidner (1986)

RE Resolution in the Indexical Model

Instructor: This is a blue rectangle.
Instructor: This is to the right of the red circle.



Cognitive Status

O_p = {O1,O2,O3,O4}
O_a = {O1}
O_f = {}

GH Heuristic

the N ⟶ percept

O_c = O_p

Filter(red)

O_c = {O2, O3}

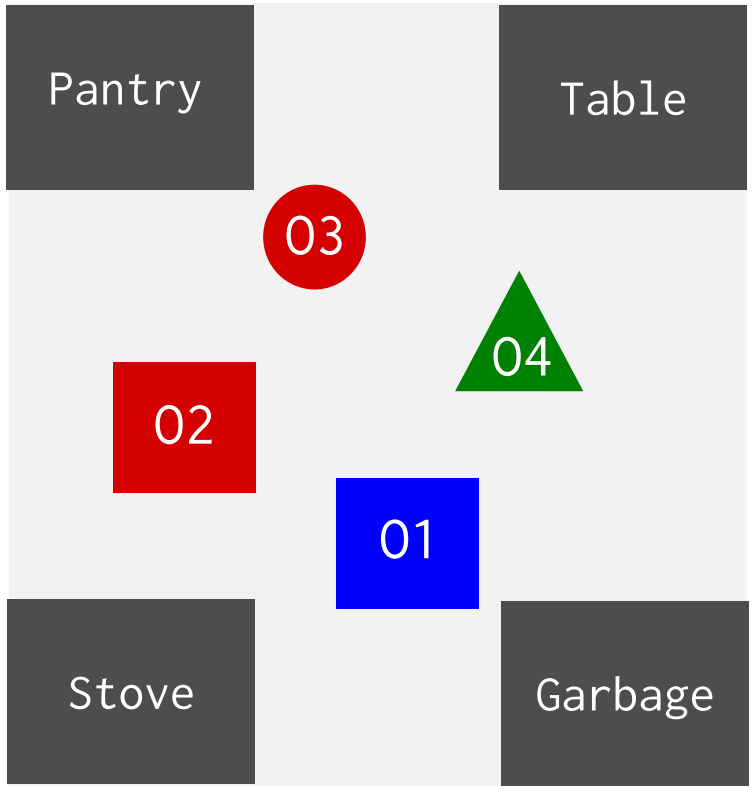
Filter(circle)

O_c = {O3}

O_a: Soar's architectural activation; O_f: Dialog model based on Grosz and Sidner (1986)

RE Resolution in the Indexical Model

Instructor: This is a blue rectangle.
Instructor: This is to the right of the red circle.



Cognitive Status

O_p = {01,02,03,04}
O_a = {01}
O_f = {}

GH Heuristic

the N ⟶ percept

O_c = O_p

Filter(red)

O_c = {02, 03}

Filter(circle)

O_c = {03}

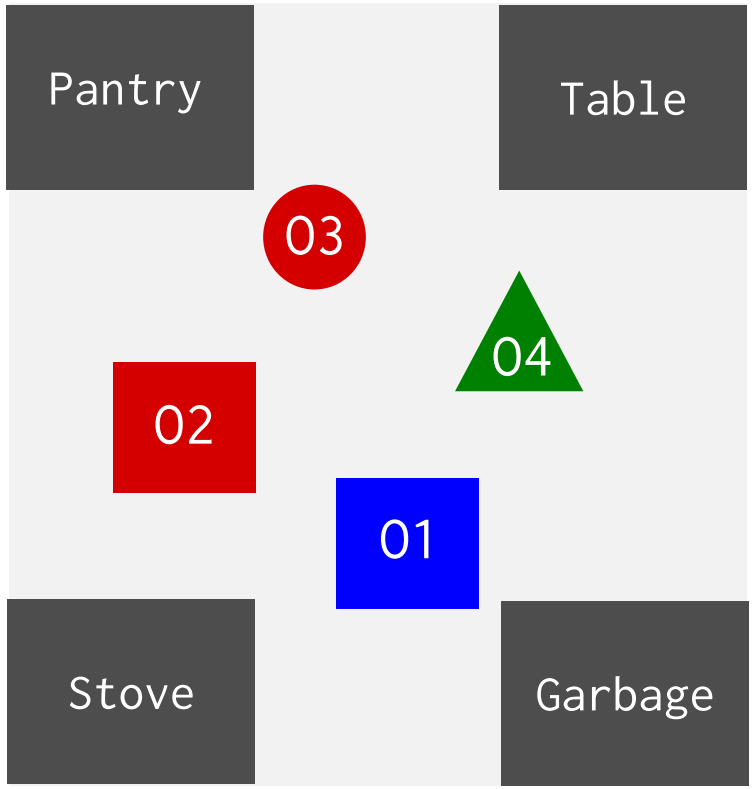
Resolution

'the red circle' = 03

O_a: Soar's architectural activation; O_f: Dialog model based on Grosz and Sidner (1986)

RE Resolution in the Indexical Model

Instructor: This is a blue rectangle.
Instructor: This is to the right of the red circle.



Cognitive Status

O_p = {01,02,03,04}
O_a = {03, 01}
O_f = {}

GH Heuristic

the N ⟶ percept

O_c = O_p

Filter(red)

O_c = {02, 03}

Filter(circle)

O_c = {03}

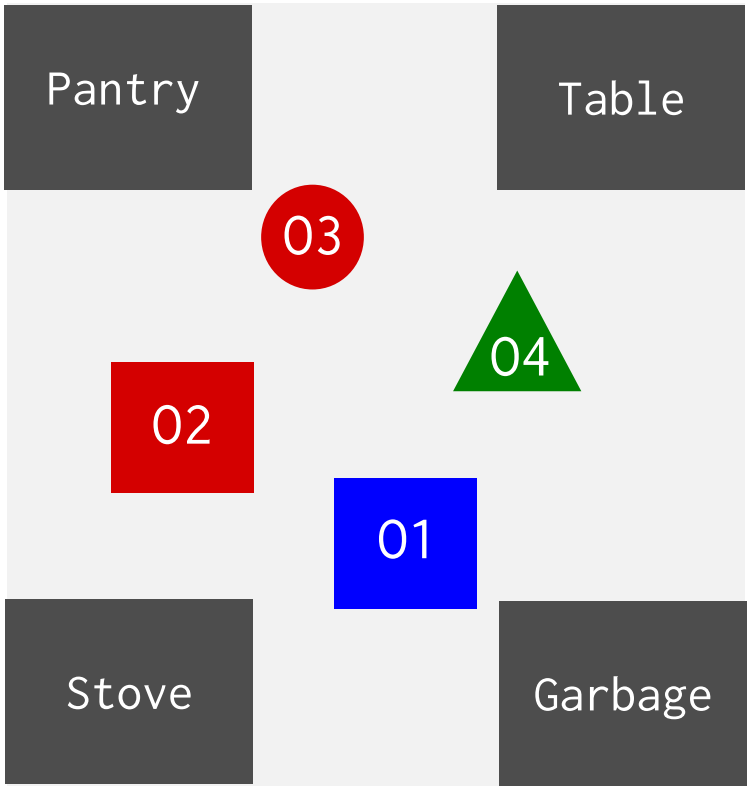
Resolution

'the red circle' = 03

O_a: Soar's architectural activation; O_f: Dialog model based on Grosz and Sidner (1986)

RE Resolution in the Indexical Model

Instructor: This is a blue rectangle.
Instructor: This is to the right of the red circle.
Instructor: Move the rectangle to the pantry.



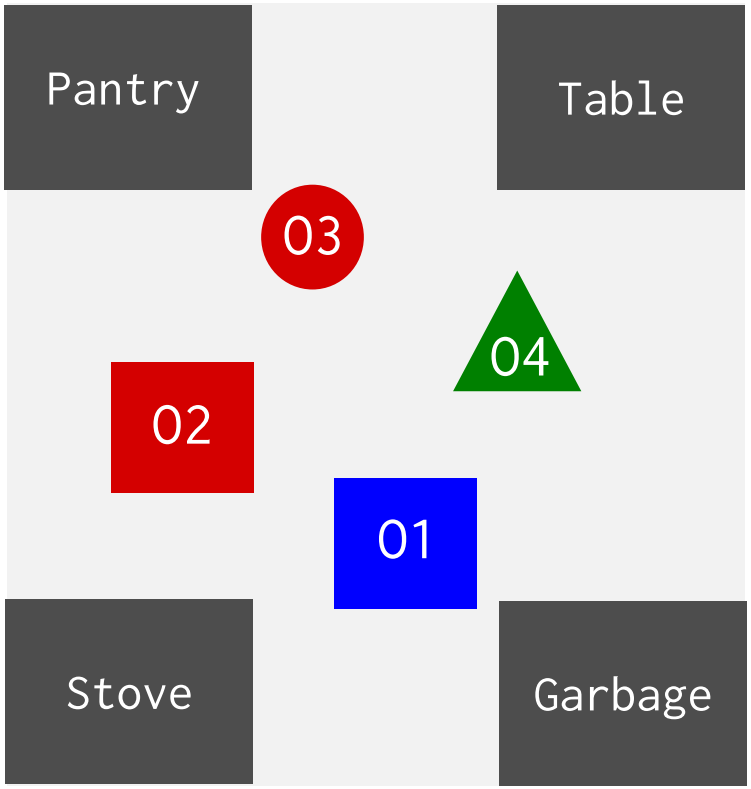
Cognitive Status
O_p = {01,02,03,04}
O_a = {03, 01}
O_f = {}

GH Heuristic
the N ⟶ percept
O_c = O_p
Filter(rectangle)
O_c = {01, 02}

O_a: Soar's architectural activation; O_f: Dialog model based on Grosz and Sidner (1986)

RE Resolution in the Indexical Model

Instructor: This is a blue rectangle.
Instructor: This is to the right of the red circle.
Instructor: Move the rectangle to the pantry.



Cognitive Status
O_p = {01,02,03,04}
O_a = {03, 01}
O_f = {}

GH Heuristic
the N ⟶ percept
O_c = O_p

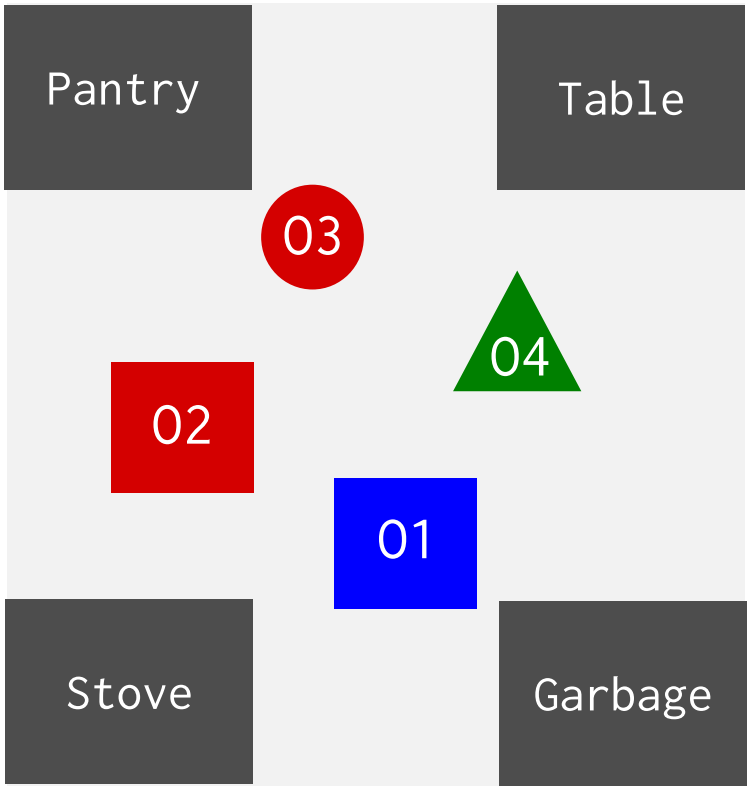
Filter(rectangle)
O_c = {01, 02}

Resolution
'the rectangle' = 01

O_a: Soar's architectural activation; O_f: Dialog model based on Grosz and Sidner (1986)

RE Resolution in the Indexical Model

Instructor: This is a blue rectangle.
Instructor: This is to the right of the red circle.
Instructor: Move the rectangle to the pantry.

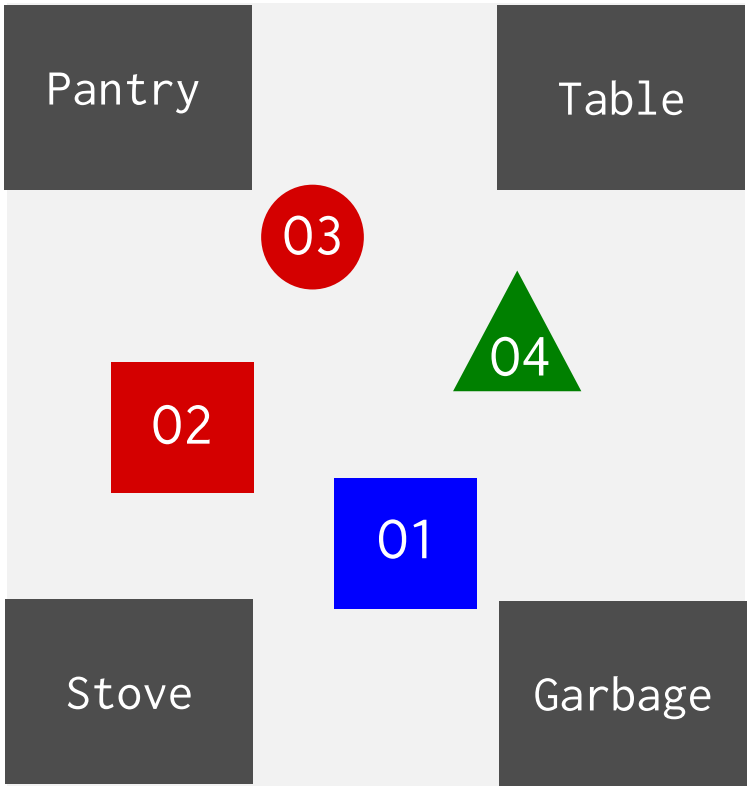


Cognitive Status
O_p = {O1,O2,O3,O4}
O_a = {O3, O1}
O_f = {O1, pantry}

O_a: Soar's architectural activation; O_f: Dialog model based on Grosz and Sidner (1986)

RE Resolution in the Indexical Model

Instructor: This is a blue rectangle.
Instructor: This is to the right of the red circle.
Instructor: Move the rectangle to the pantry.
Instructor: Pick it.



Cognitive Status

O_p = {01,02,03,04}
O_a = {03, 01}
O_f = {01, pantry}

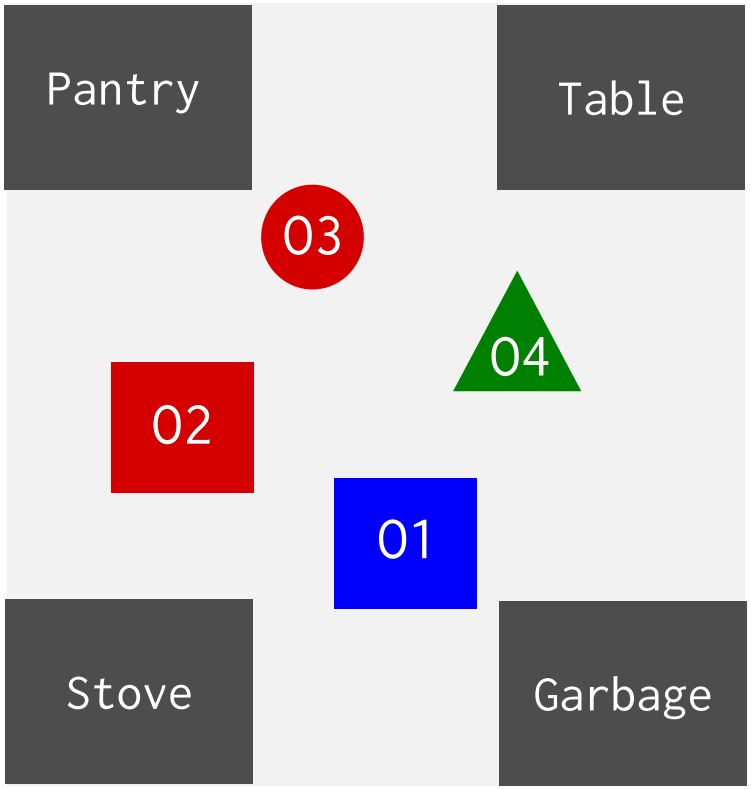
GH Heuristic

it ⟶ focus
O_c = O_f

O_a: Soar's architectural activation; O_f: Dialog model based on Grosz and Sidner (1986)

RE Resolution in the Indexical Model

Instructor: This is a blue rectangle.
Instructor: This is to the right of the red circle.
Instructor: Move the rectangle to the pantry.
Instructor: Pick it.



Cognitive Status

O_p = {01,02,03,04}
O_a = {03, 01}
O_f = {01, pantry}

GH Heuristic

it ⟶ focus

O_c = O_f

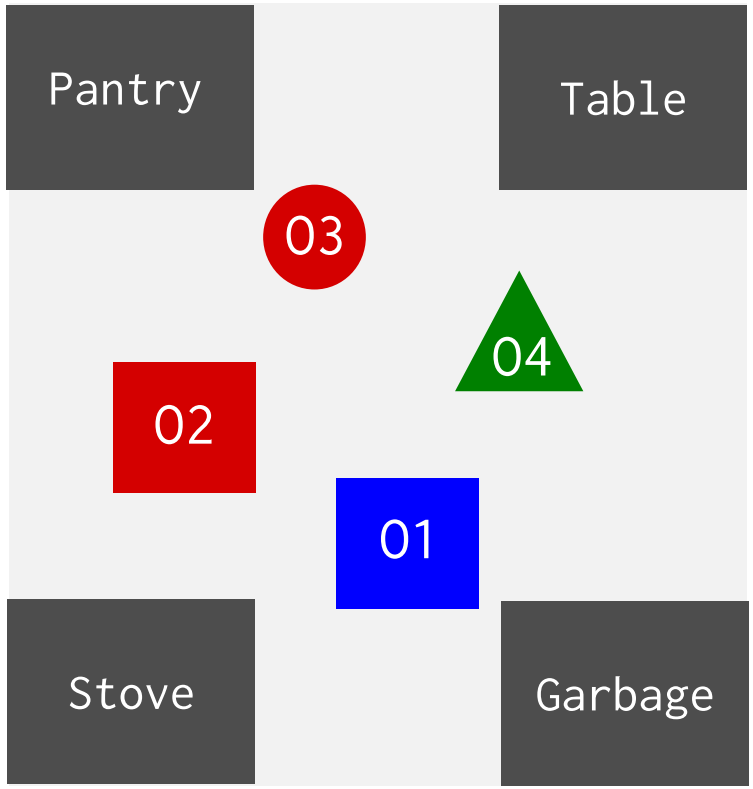
Affordance(pick-up)

O_c = {01,pantry}

O_a: Soar's architectural activation; O_f: Dialog model based on Grosz and Sidner (1986)

RE Resolution in the Indexical Model

Instructor: This is a blue rectangle.
Instructor: This is to the right of the red circle.
Instructor: Move the rectangle to the pantry.
Instructor: Pick it.



Cognitive Status

O_p = {01,02,03,04}
O_a = {03, 01}
O_f = {01, pantry}

GH Heuristic

it ⟶ focus

O_c = O_f

Affordance(pick-up)

O_c = {01,pantry}

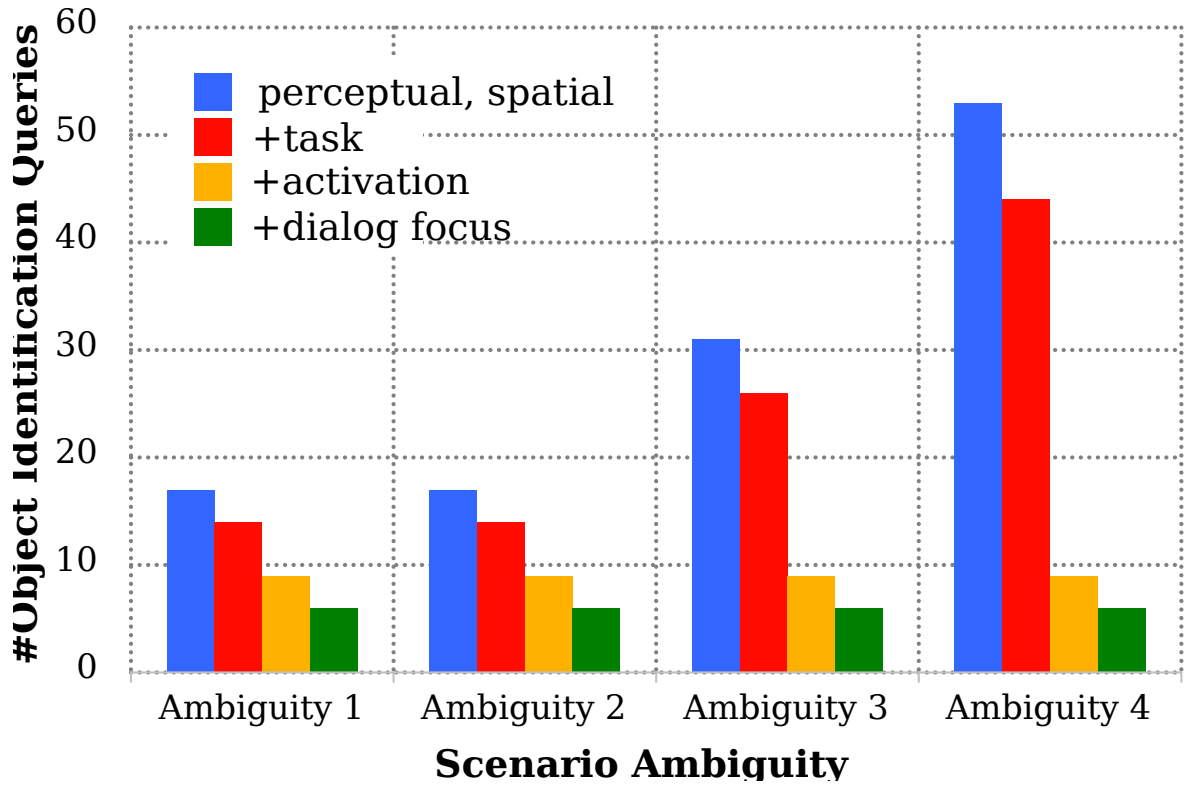
Resolution

'it' = 01

O_a: Soar's architectural activation; O_f: Dialog model based on Grosz and Sidner (1986)

RE Resolution Performance

- Corpus: instructional dialogs,
  12 *it*,
  4 *this*,
  3 *that cylinder*
  14 *the red cylinder*
- Queries
  Instructor: Pick it.
  Agent: Which object?
- Linguistic context only
  Stanford CoreNLP fails at 28.6% of references.



Unexpressed Argument Alternation

- Goal of verb comprehension - identify and instantiate a task
- Variability in how tasks are described.

  a. *Take the trash out to the curb.*
  b. *Take the trash out.*
- Human speakers and hearers rely on shared experience/knowledge
  - agreement on where the trash is usually deposited (common ground)
  - missing information is filled in from knowledge

Exploiting the Instructional Experience

1
move the red cylinder to the right of the pantry

the goal is the red cylinder to the right of the pantry

pick up the red cylinder

put the red cylinder to the right of the pantry

you are done

Exploiting the Instructional Experience

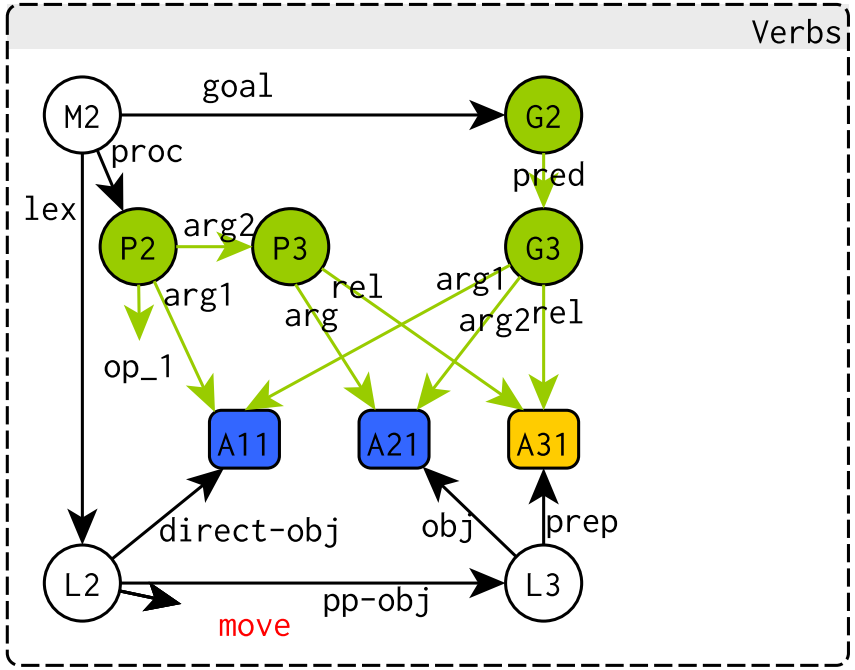
1
move the red cylinder to the right of the pantry

the goal is the red cylinder to the right of the pantry

pick up the red cylinder

put the red cylinder to the right of the pantry

you are done



Exploiting the Instructional Experience

1
move the red cylinder to the right of the pantry

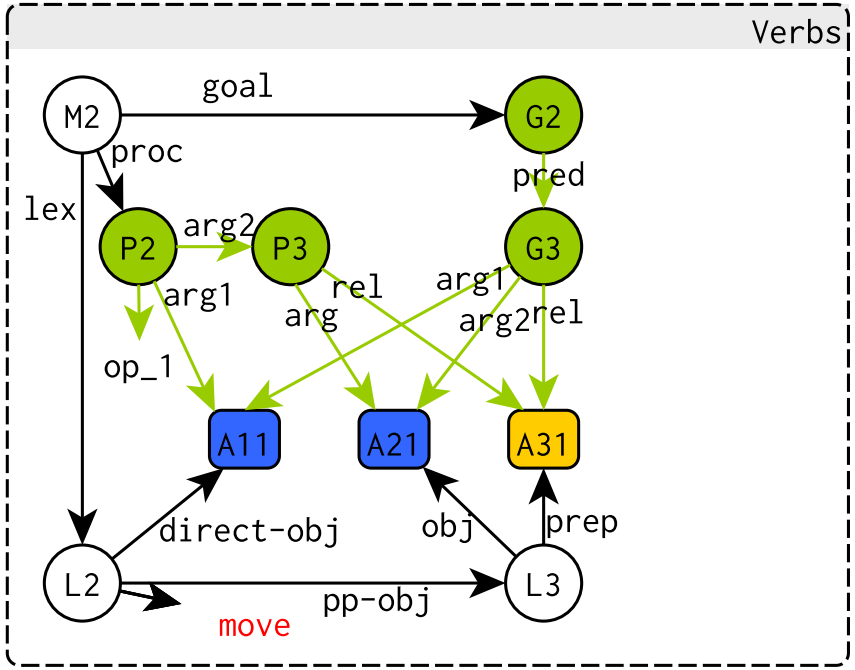
the goal is the red cylinder to the right of the pantry

pick up the red cylinder

put the red cylinder to the right of the pantry

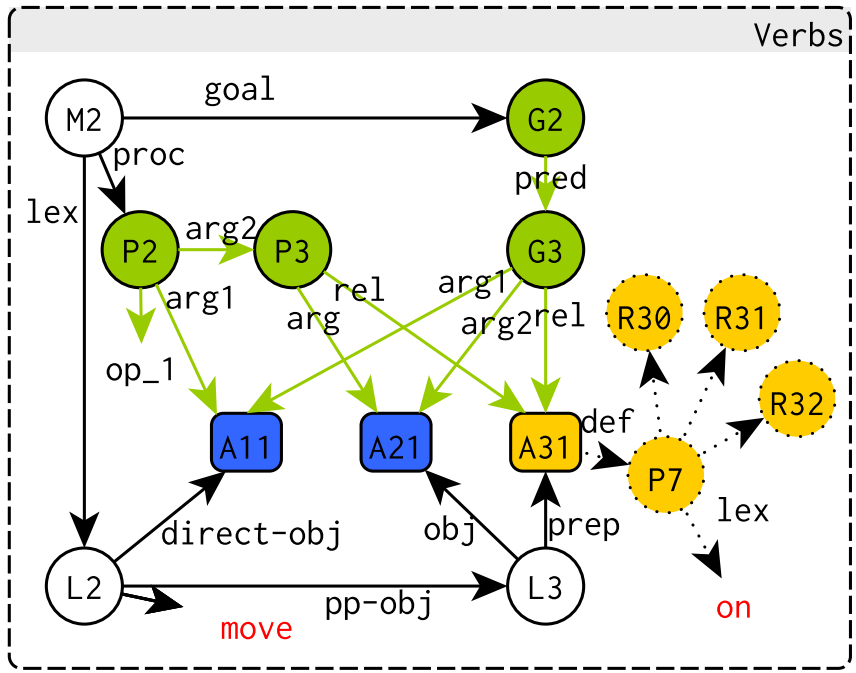
you are done

2
move the red cylinder to the pantry



Exploiting the Instructional Experience

1	2
move the red cylinder to the right of the pantry	move the red cylinder to the pantry
the goal is the red cylinder to the right of the pantry	the goal is the red cylinder in the pantry
pick up the red cylinder	pick up the red cylinder
put the red cylinder to the right of the pantry	put the red cylinder in the pantry
you are done	you are done



Performance

pick up

put down

move

a. *Move the red cylinder to the right of the table.*

b. *Move the red cylinder to the table.*

store

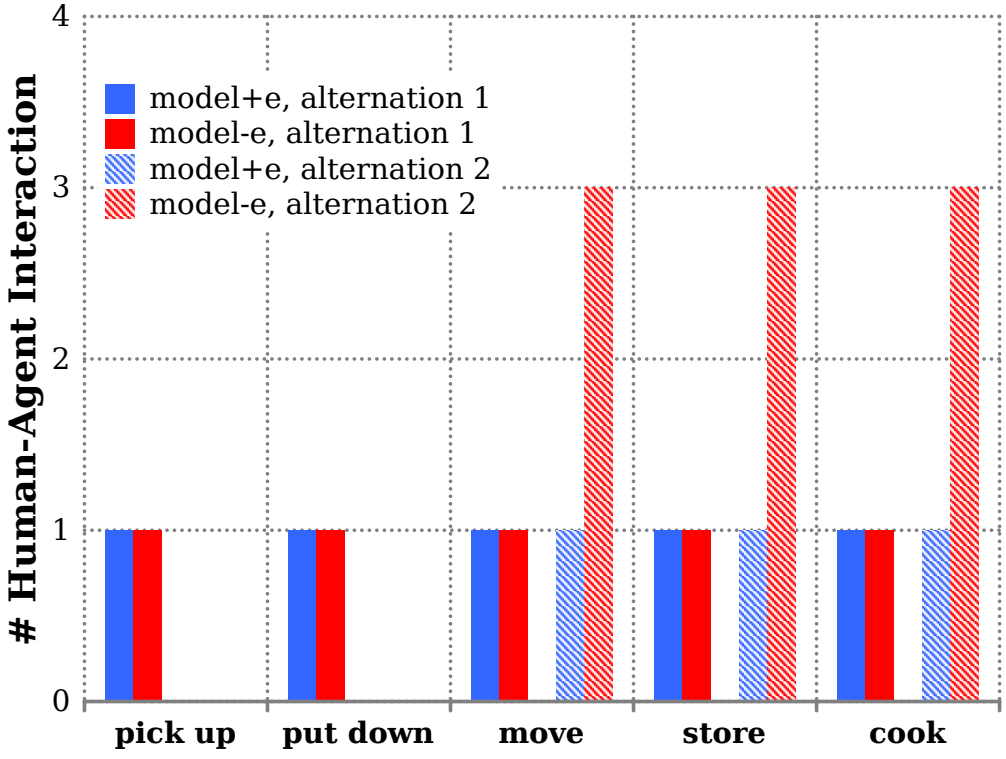
a. *Store the red cylinder in the pantry.*

b. *Store the red cylinder.*

cook

a. *Cook the steak on the stove.*

b. *Cook the steak.*



Summary

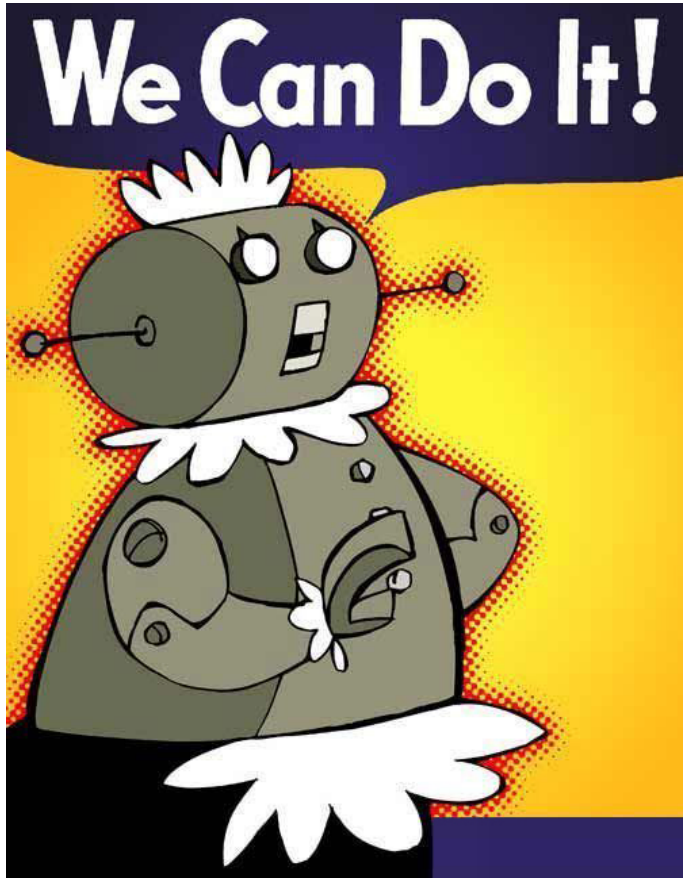
- Linguistic features are cues to search perception, knowledge, and experience
- Allows for modality-specific mixed representation
  - classification, reasoning for continuous space, goals and policy
  - integrates with learning [Mohan et al. 2012]
- Natural role of non-linguistic context
  - constrains and guides search
- Addresses complexities due to ambiguities and missing information

Future Work

- Ambiguity
  - preposition-phrase attachment
  - verb-task polysemy
- Complexity
  - partially-observable environments
  - non-situated comprehension
- Evaluation
  - in the wild
  - information-theoretic models

Questions?

Rosie the Robot



Thanks to:
APRIL: Edwin Olson, Robert Goeddel, Lauren Hinkle