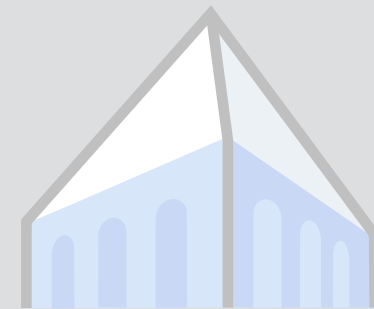
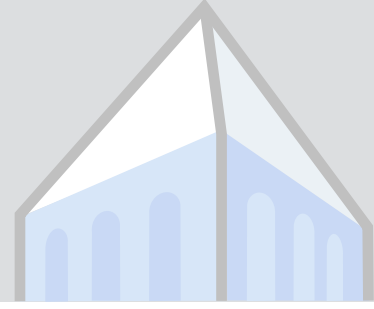


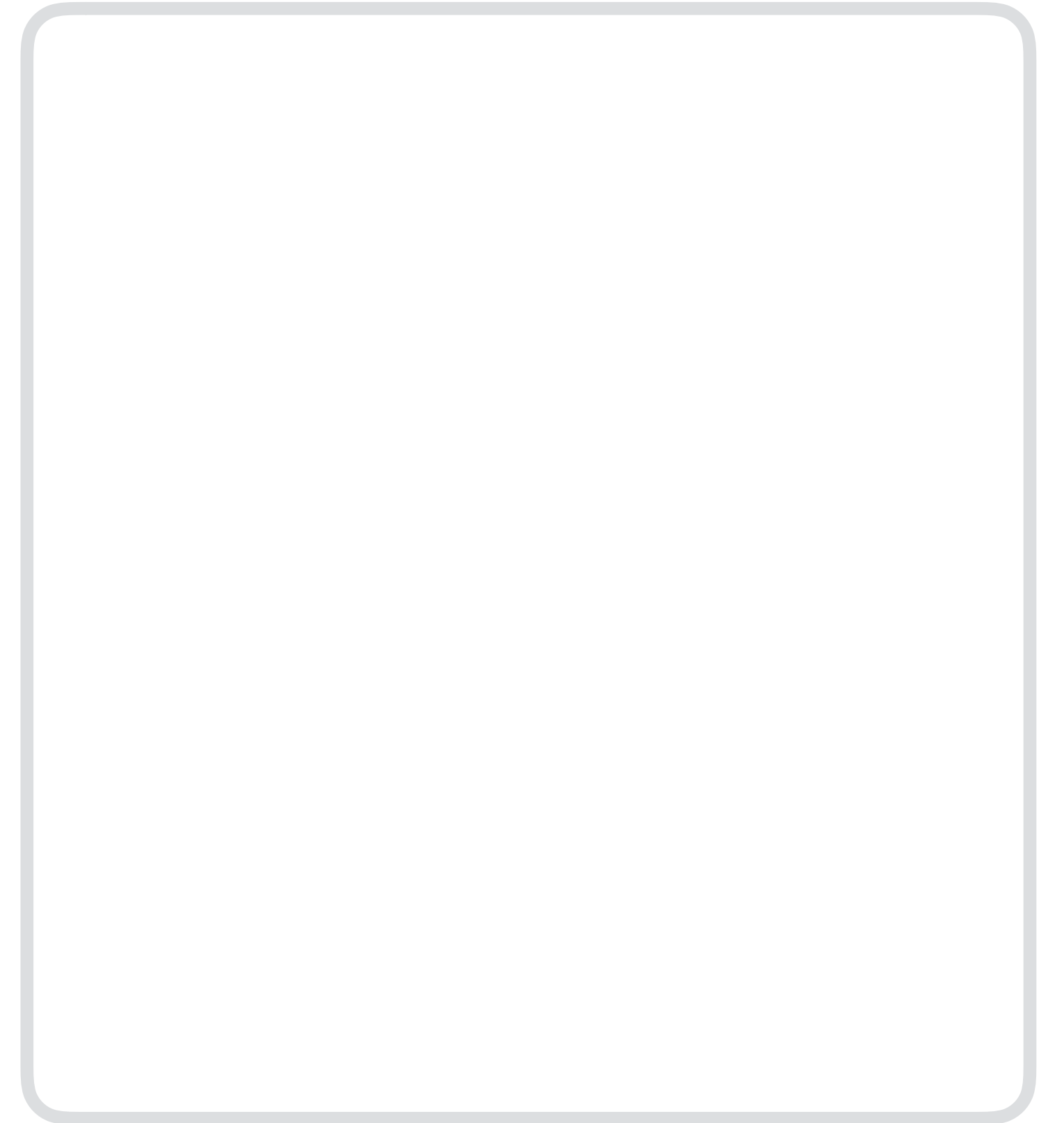
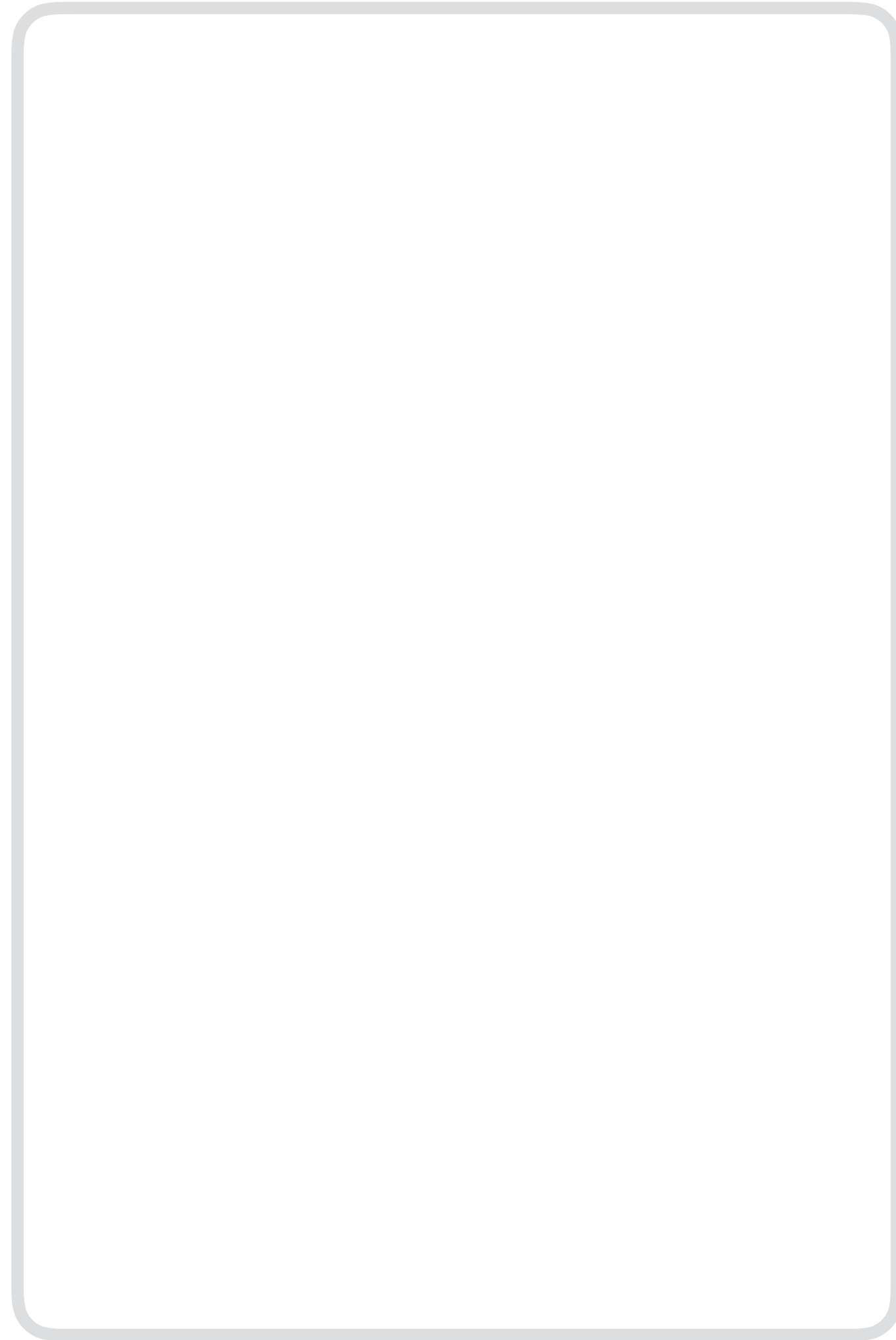
# Learning from Language



Jacob Andreas



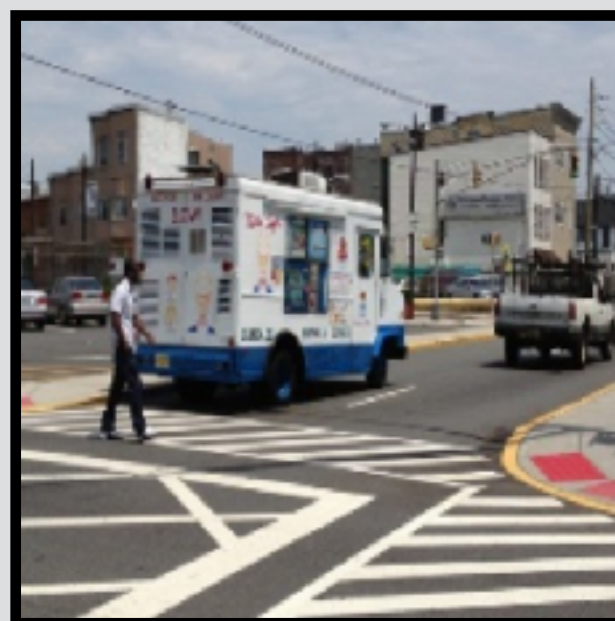
# Doing things with language





# Doing things with language

*Who is left of  
the truck?*



*A man with a  
white shirt and  
black pants.*



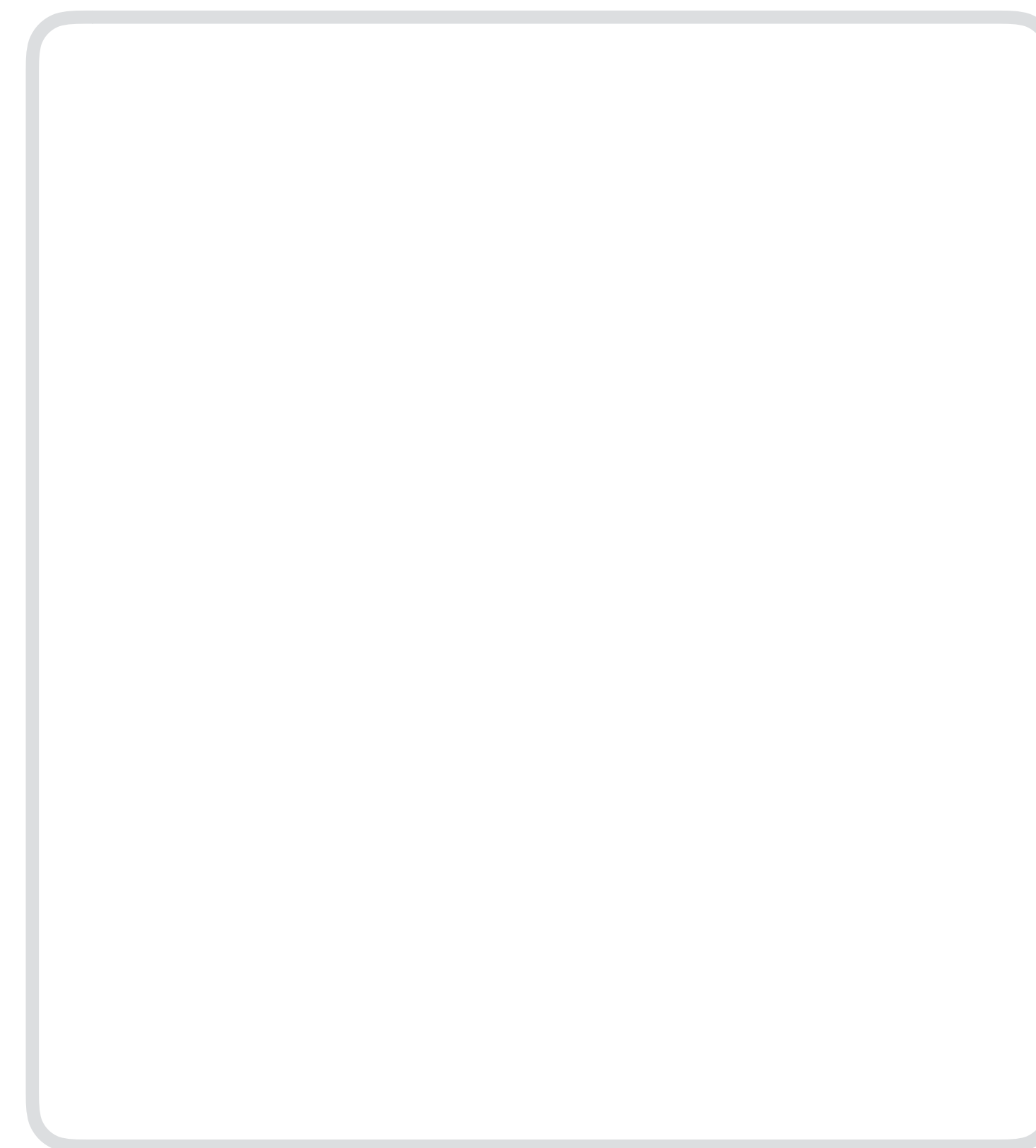
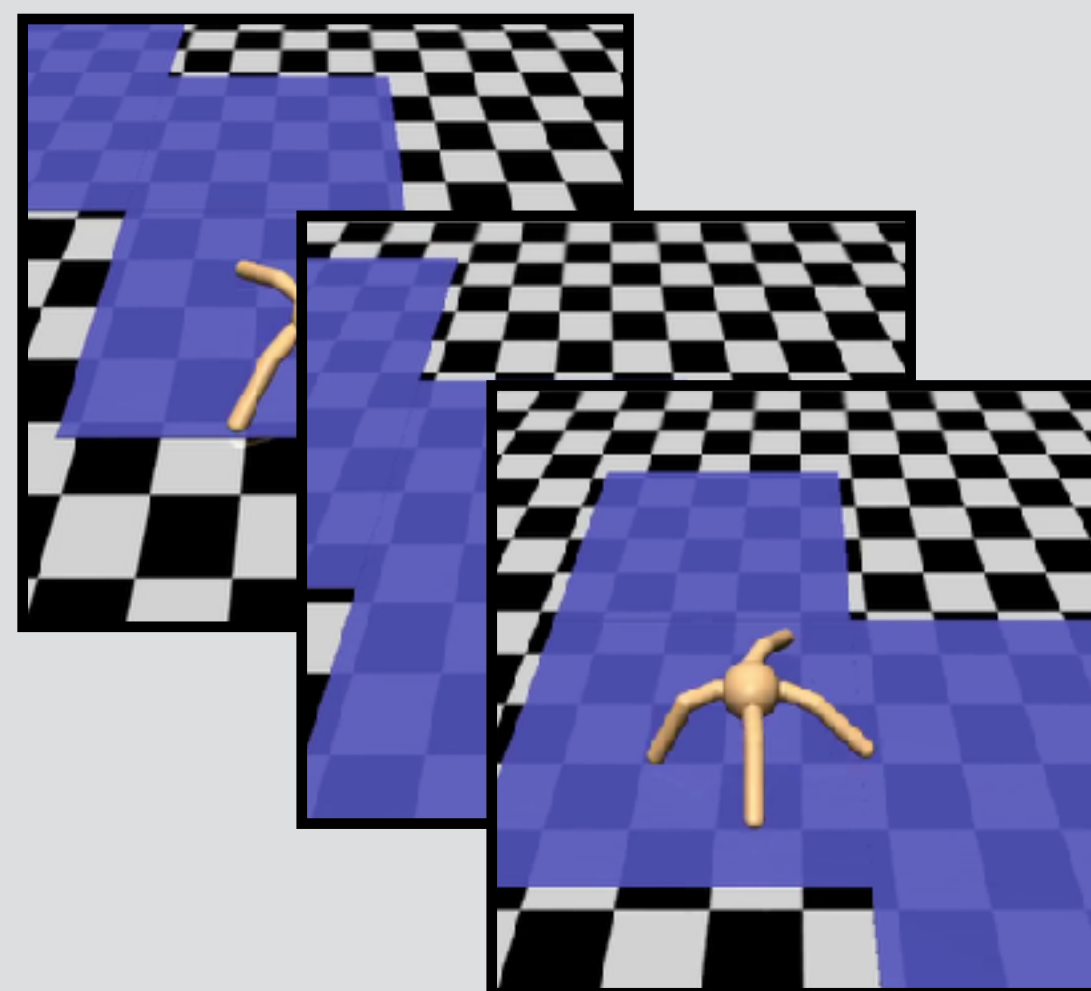
# Doing things with language

*Who is left of the truck?*



*A man with a white shirt and black pants.*

*Go up, then go left.*





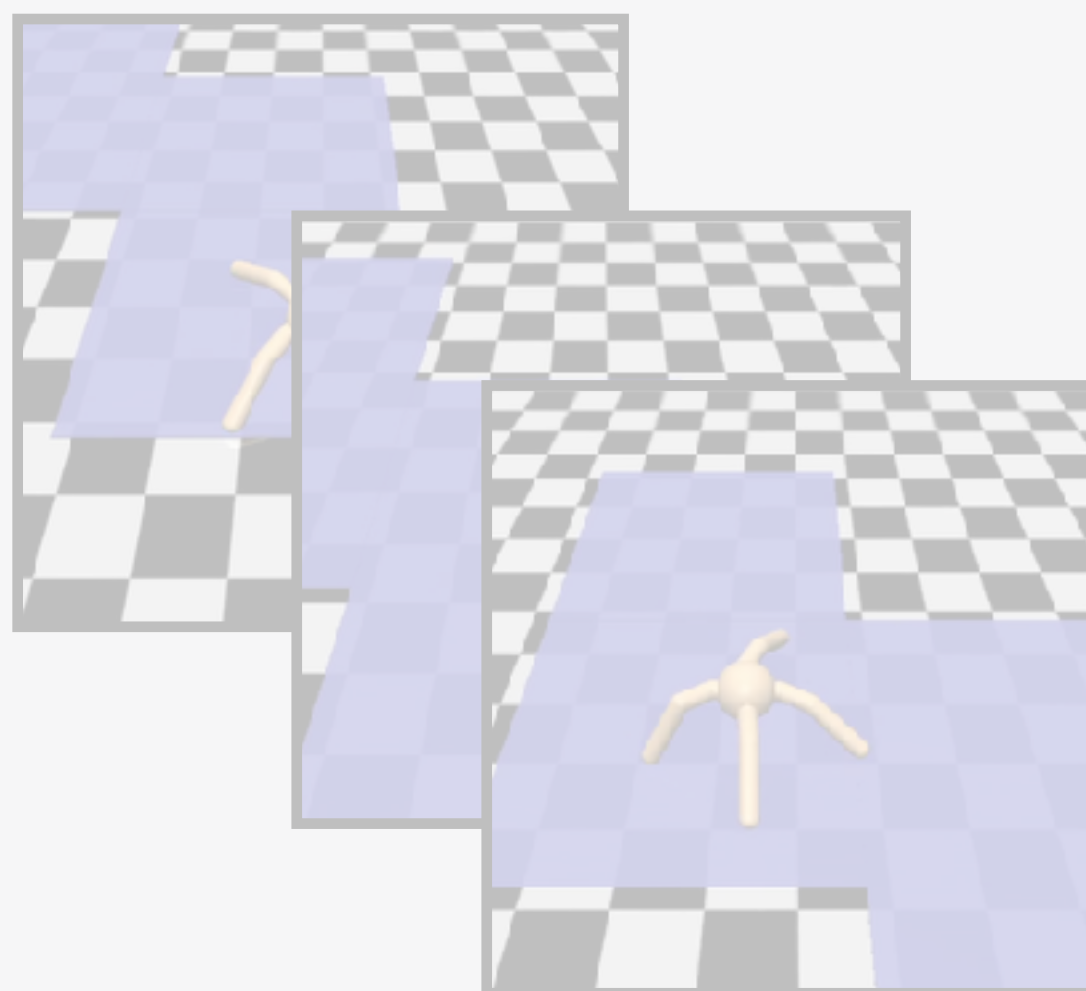
# Doing things with language

*Who is left of the truck?*



*A man with a white shirt and black pants.*

*Go up, then go left.*



*The hooded oriole is a large bird with black wings.*



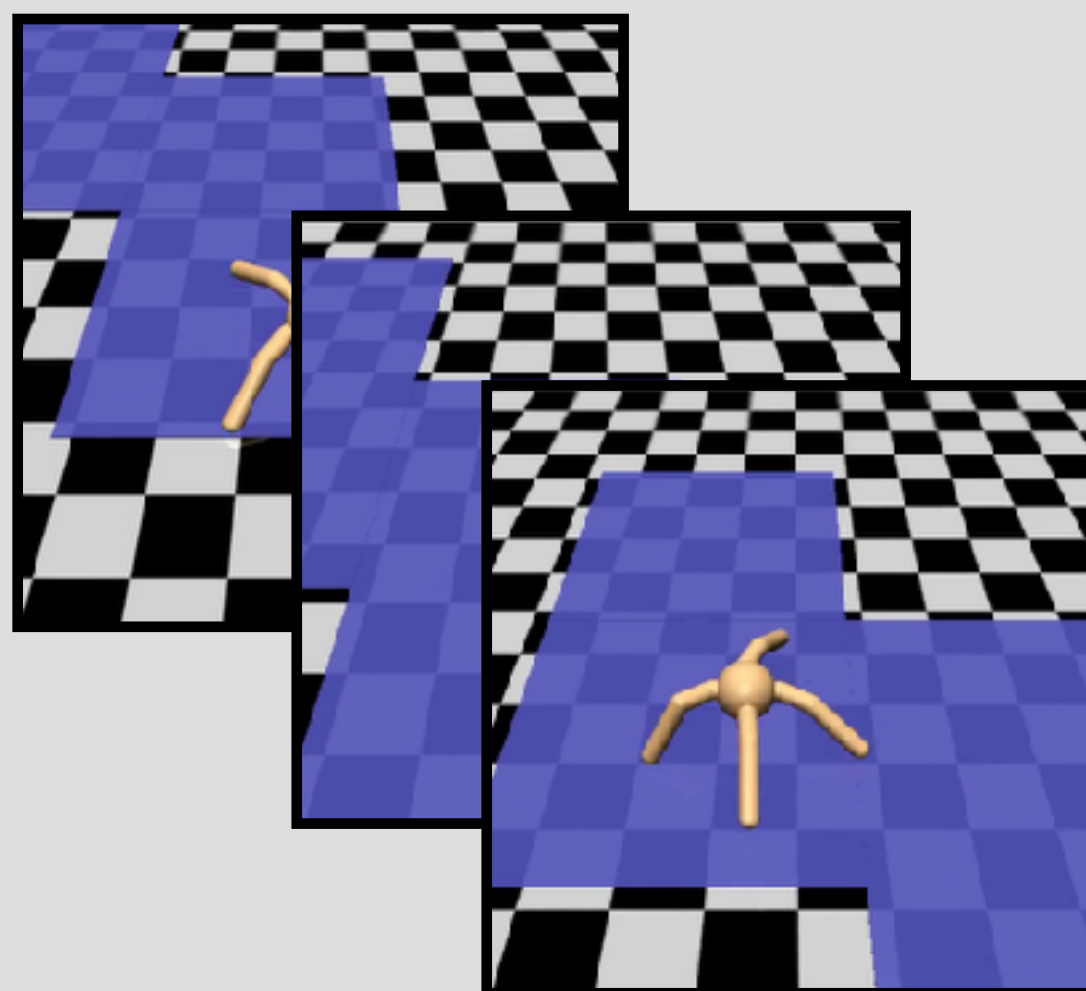
# Doing things with language

*Who is left of the truck?*



*A man with a white shirt and black pants.*

*Go up, then go left.*

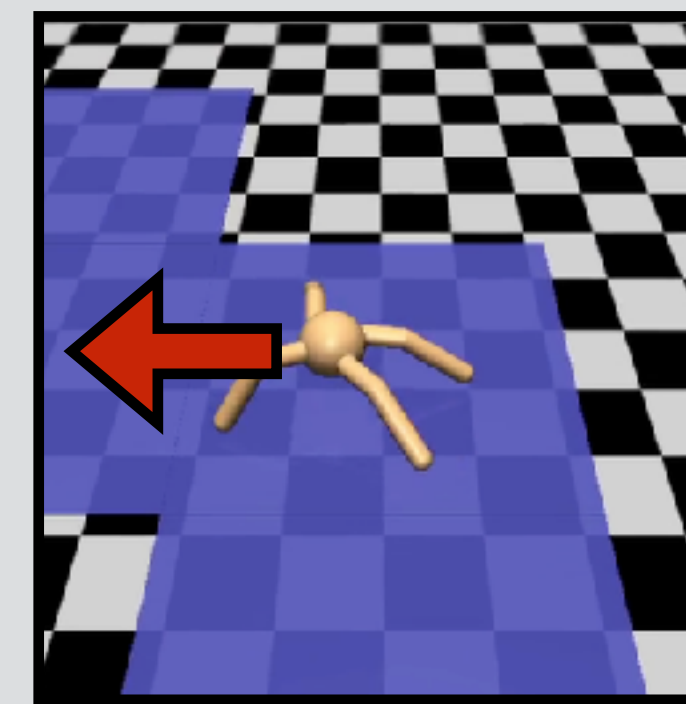
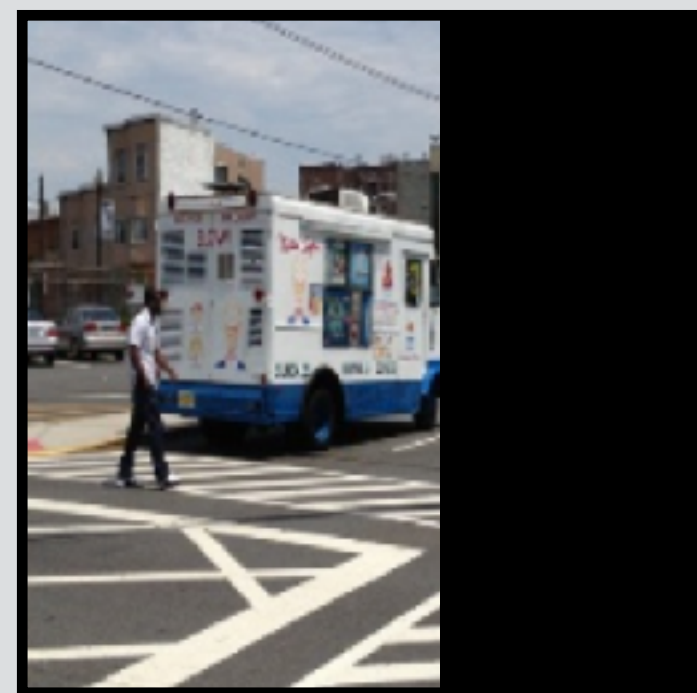
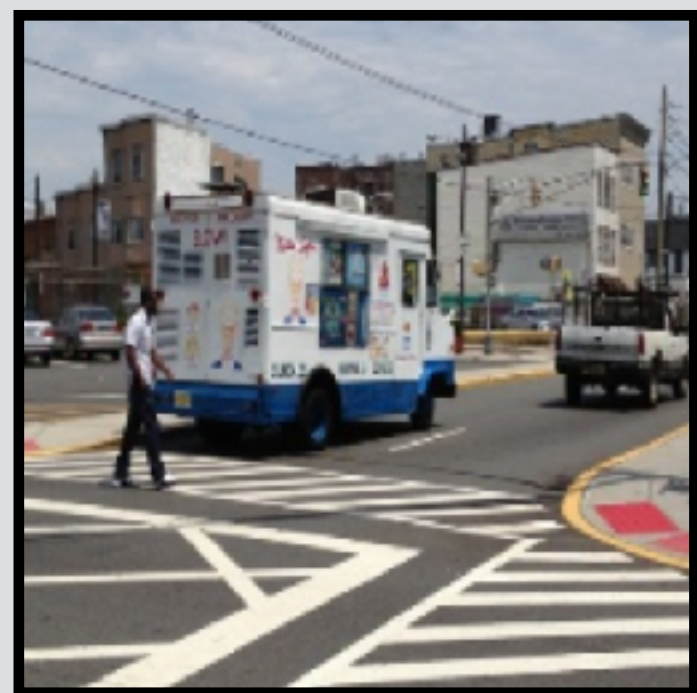


*The hooded oriole is a large bird with black wings.*

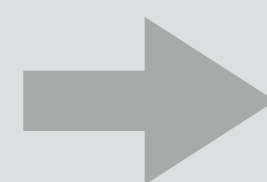
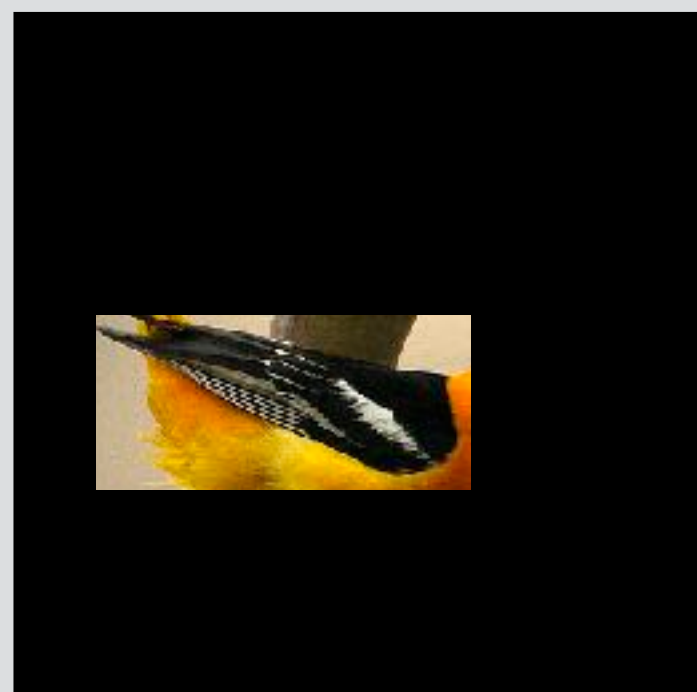


# Words and primitives

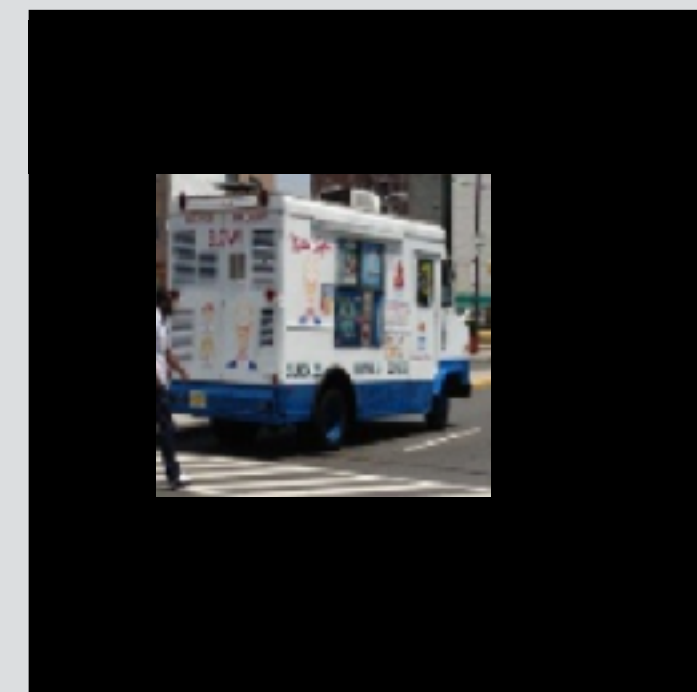
left



color



*black*

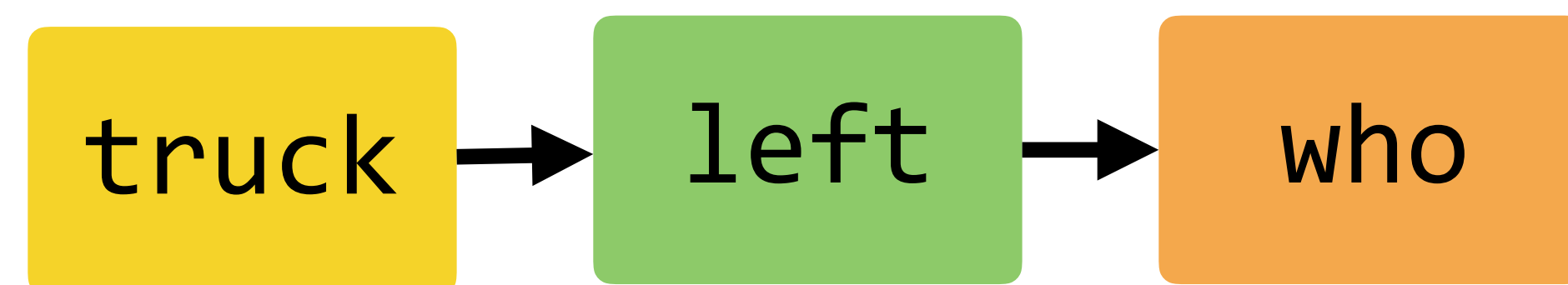


*white*

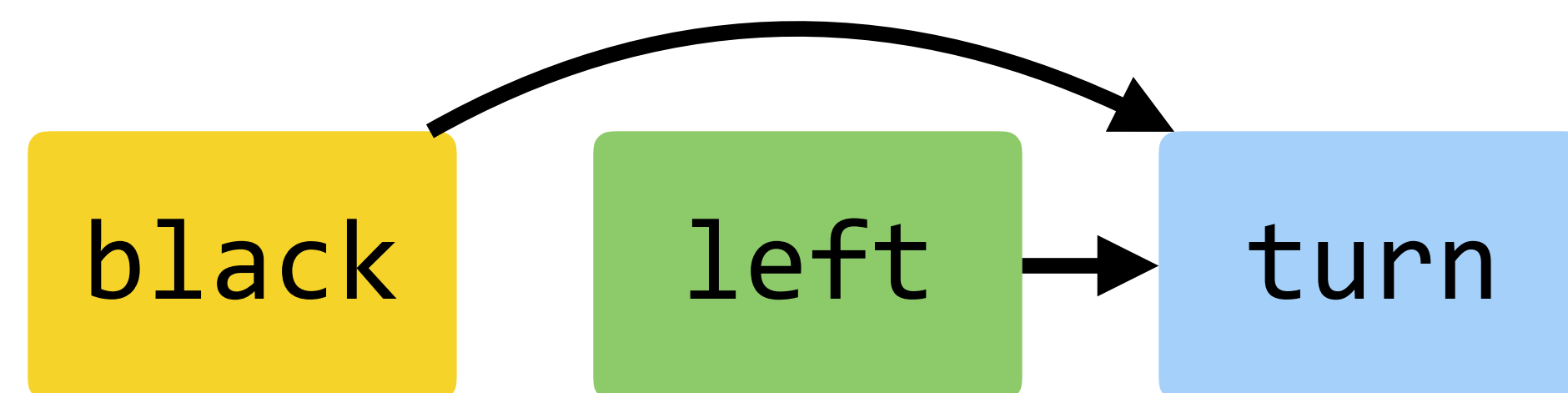


# Syntax and composition

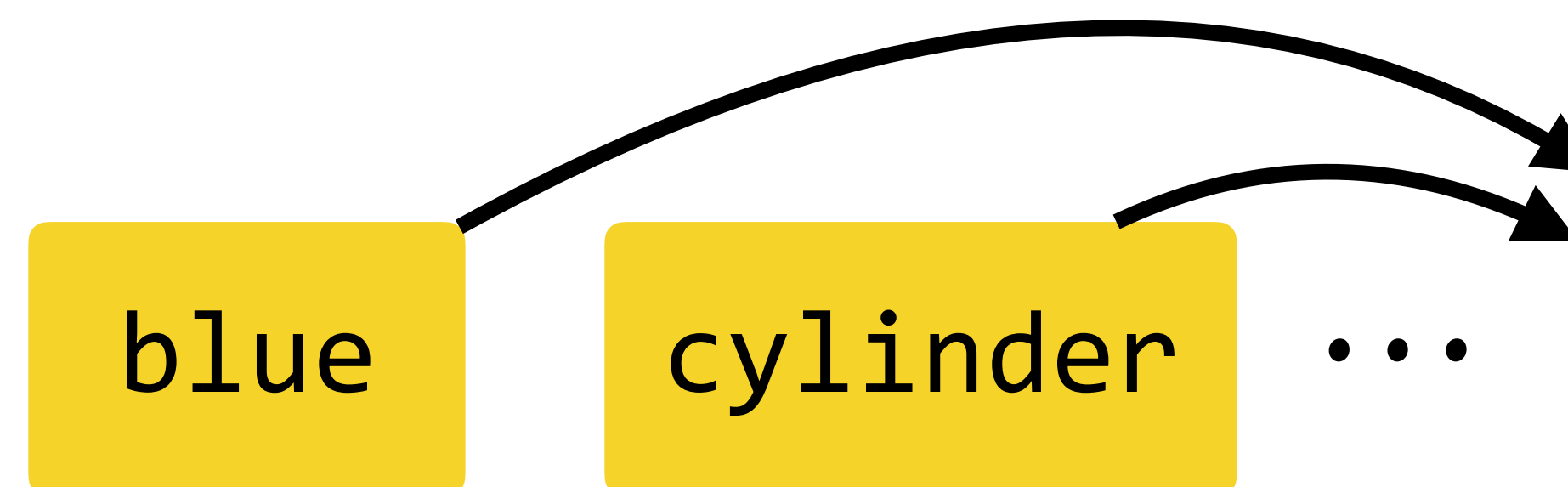
*[Who is [left of the truck]]?*



*[Turn [left] [at the black hallway]].*



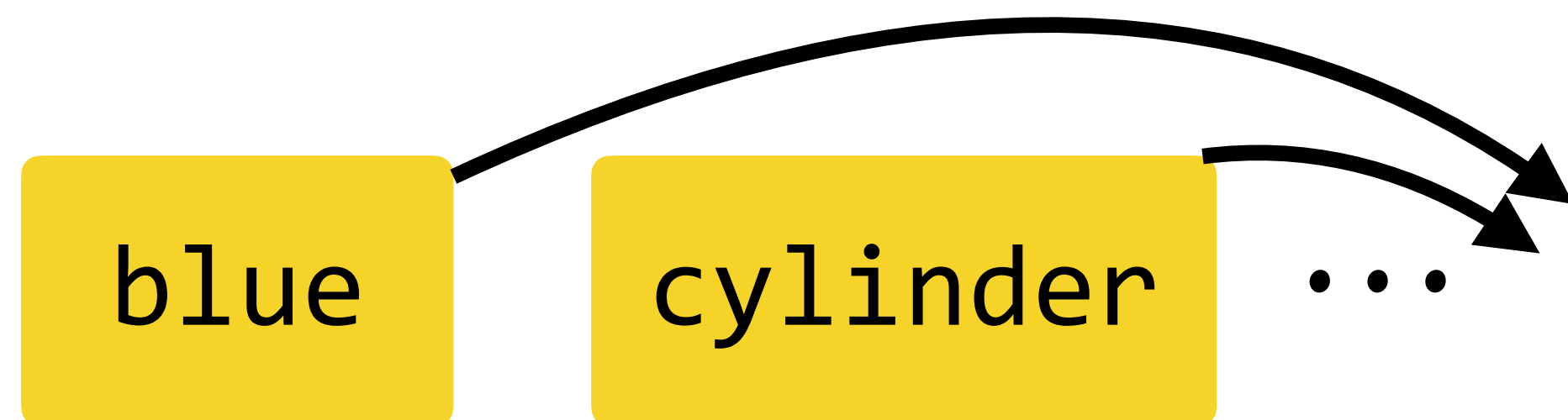
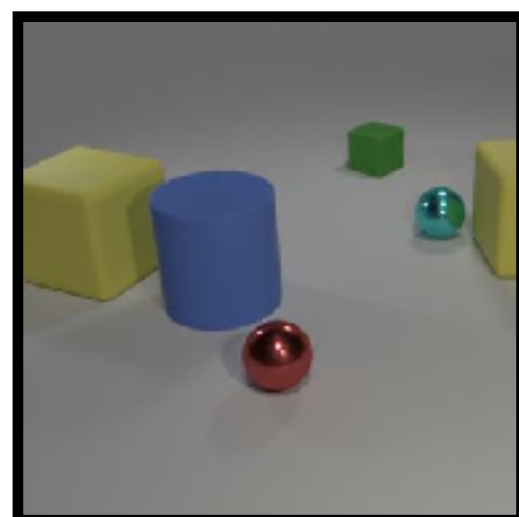
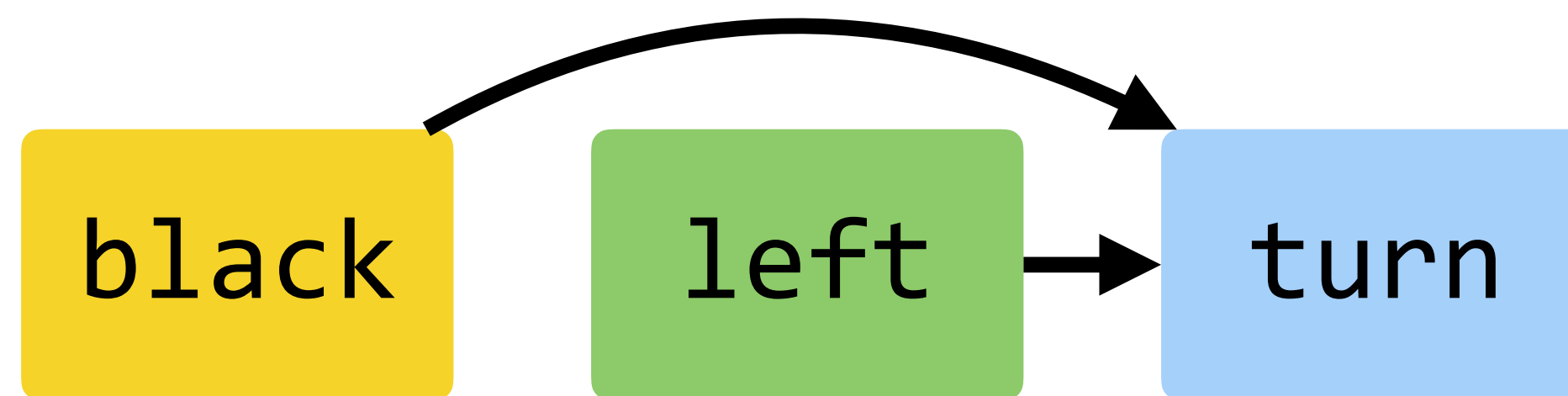
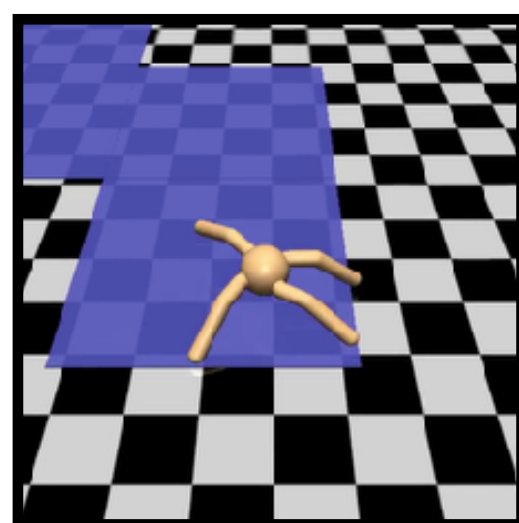
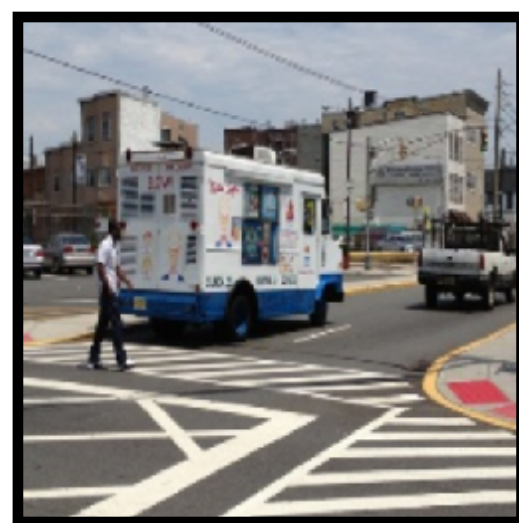
*[Does the [blue cylinder] have the [same material as the [big block [on the right side of [the red metallic thing]]]]?*





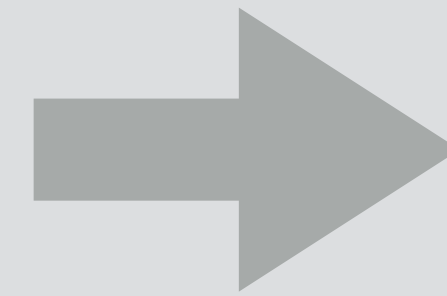
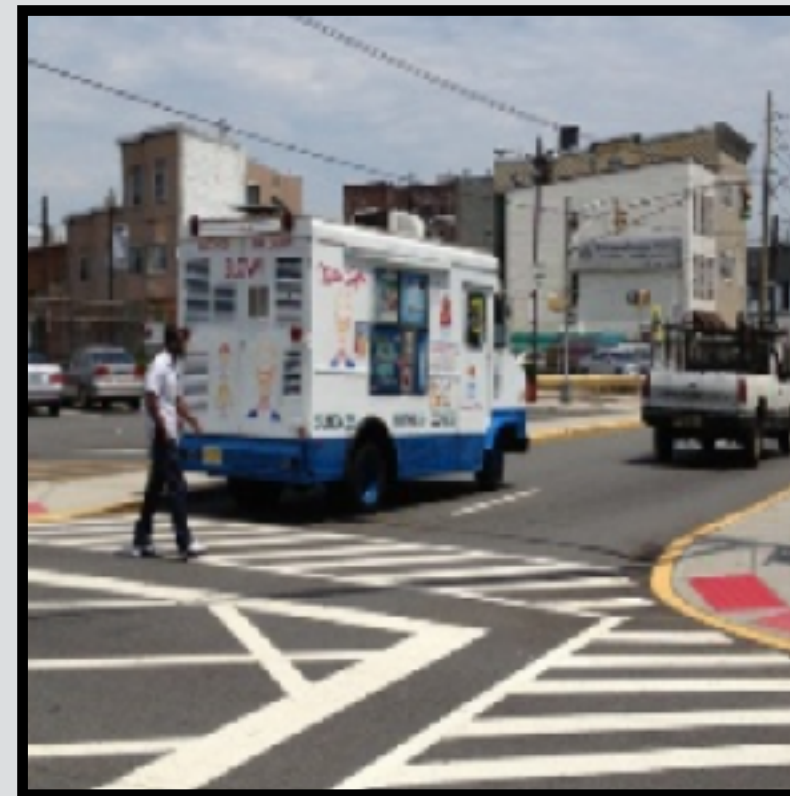


# Learning reusable abstractions



# LANGUAGE & REASONING

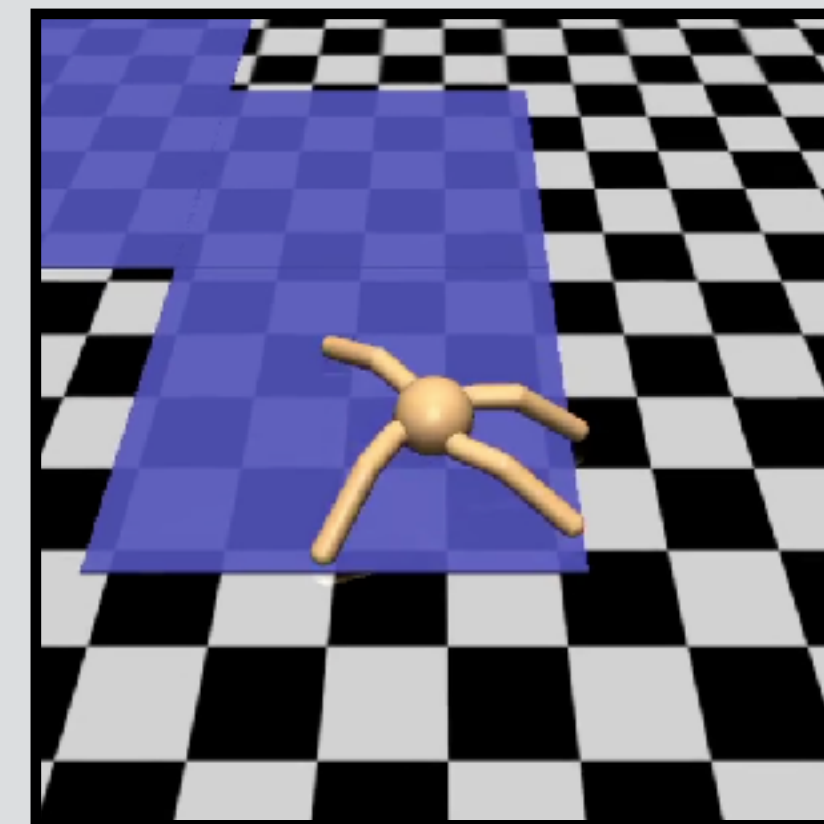
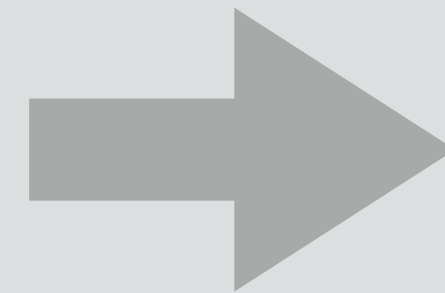
*What does the truck  
on the left sell?*



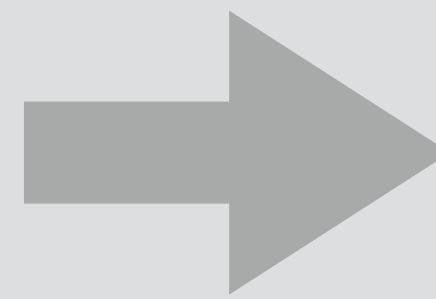
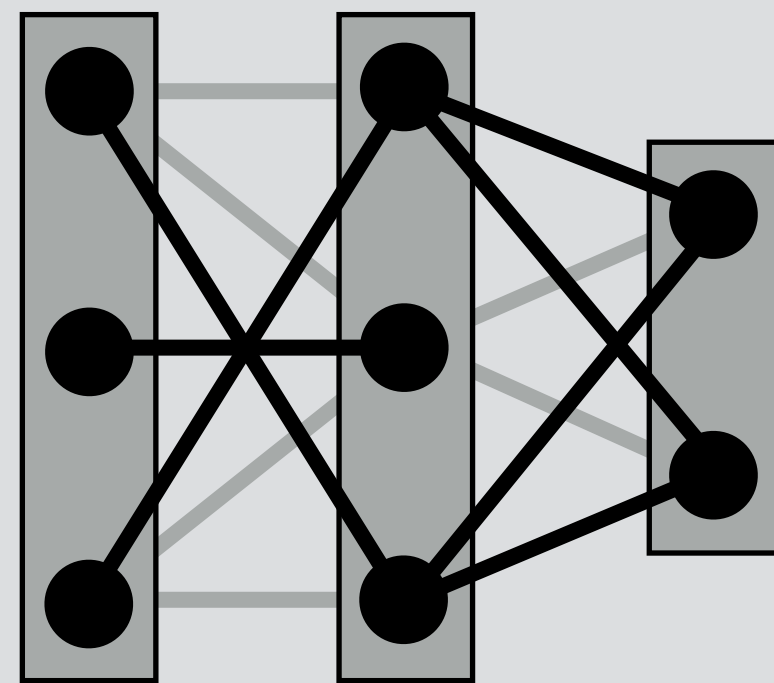
*ice cream*

# LANGUAGE & LEARNING

*Go up, then go left.*



# LANGUAGE & BELIEF



*large bird,  
black wings*

# LANGUAGE & REASONING

LEARNING

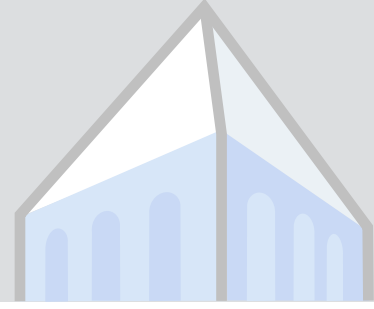
BELIEF

**A** et al. *Neural Module Networks*. CVPR 16.

**A** et al. *Learning to Compose Neural Networks for Question Answering*. NAACL 16.

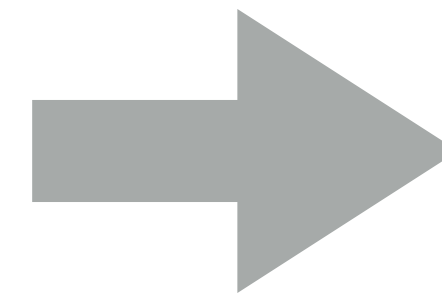
Hu, Rohrbach, **A** et al. *Modeling Relationships in Referential Expressions [...]*. CVPR 17.

Hu, **A** et al. *Learning to Reason: End-to-End Module Networks [...]*. ICCV 17.

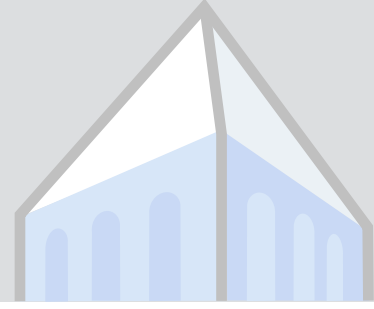


# Answering questions

*What color is  
the necktie?*



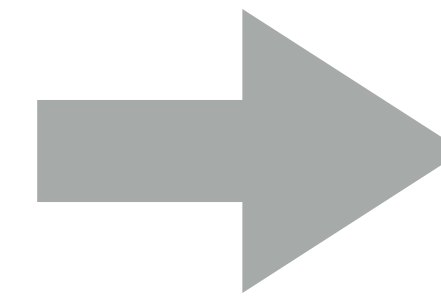
*yellow*



# Answering questions

*What rivers  
are in South  
Carolina?*

name	type	coastal
<i>Columbia</i>	city	no
<i>Cooper</i>	river	yes
<i>Charleston</i>	city	yes



*Cooper*



# Answering questions

*What rivers are in  
South Carolina?*

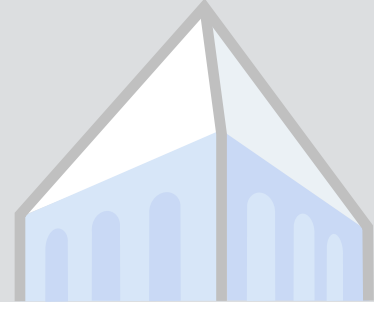
```
λx. river(x)  
  ∧ in(x, SC)
```

name	type	coastal
<i>Columbia</i>	city	no
<i>Cooper</i>	river	yes
<i>Charleston</i>	city	yes

**prolog**

*Cooper*





# Answering questions

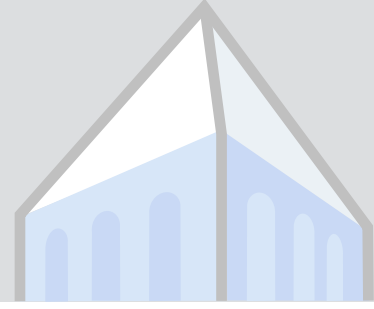
*What color is  
the necktie?*

$\lambda x. \exists y.$   
 $\text{color-of}(x, y)$   
 $\wedge \text{necktie}(y)$



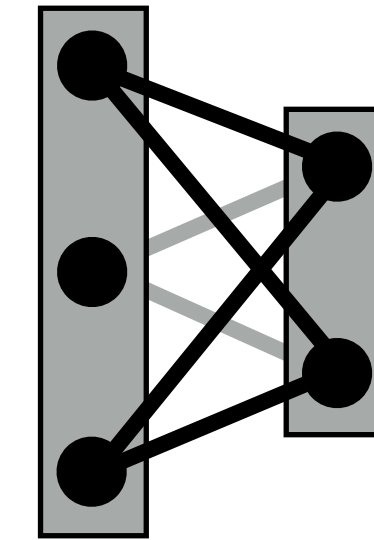
???

*yellow*

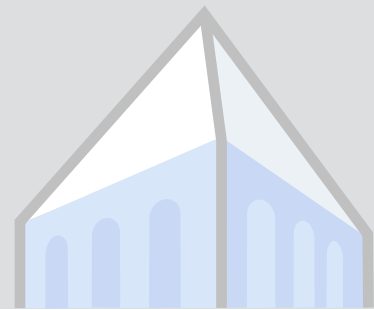


# Answering questions

*What color is  
the necktie?*



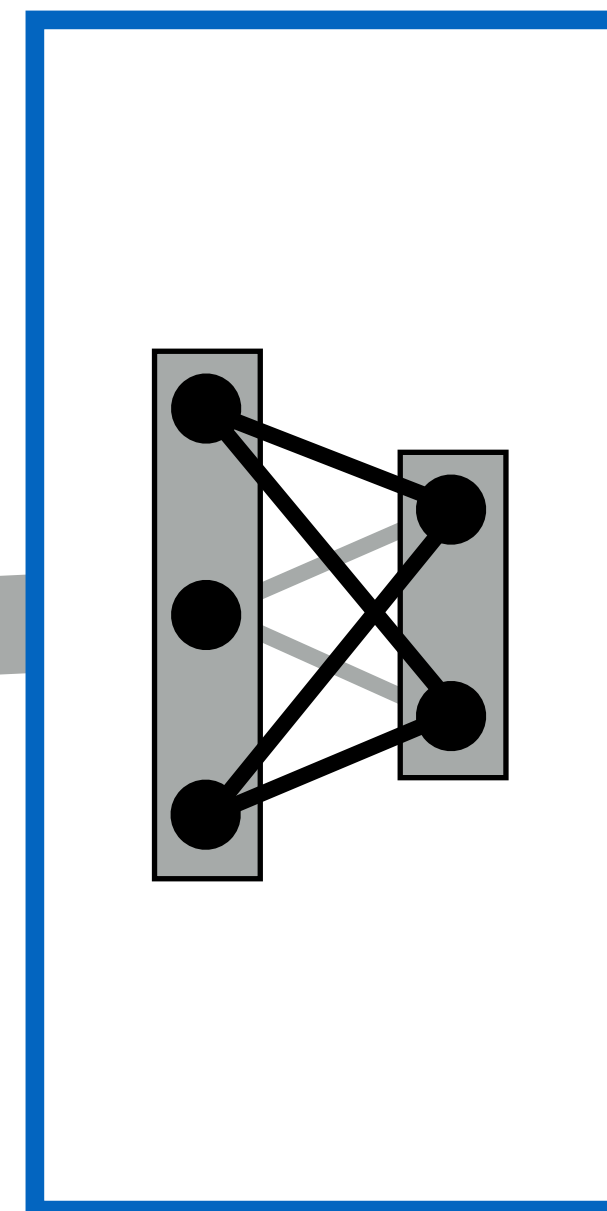
*yellow*



# Answering questions

*What color is  
the necktie?*

$\lambda x. \exists y.$   
 $\text{color-of}(x, y)$   
 $\wedge \text{necktie}(y)$

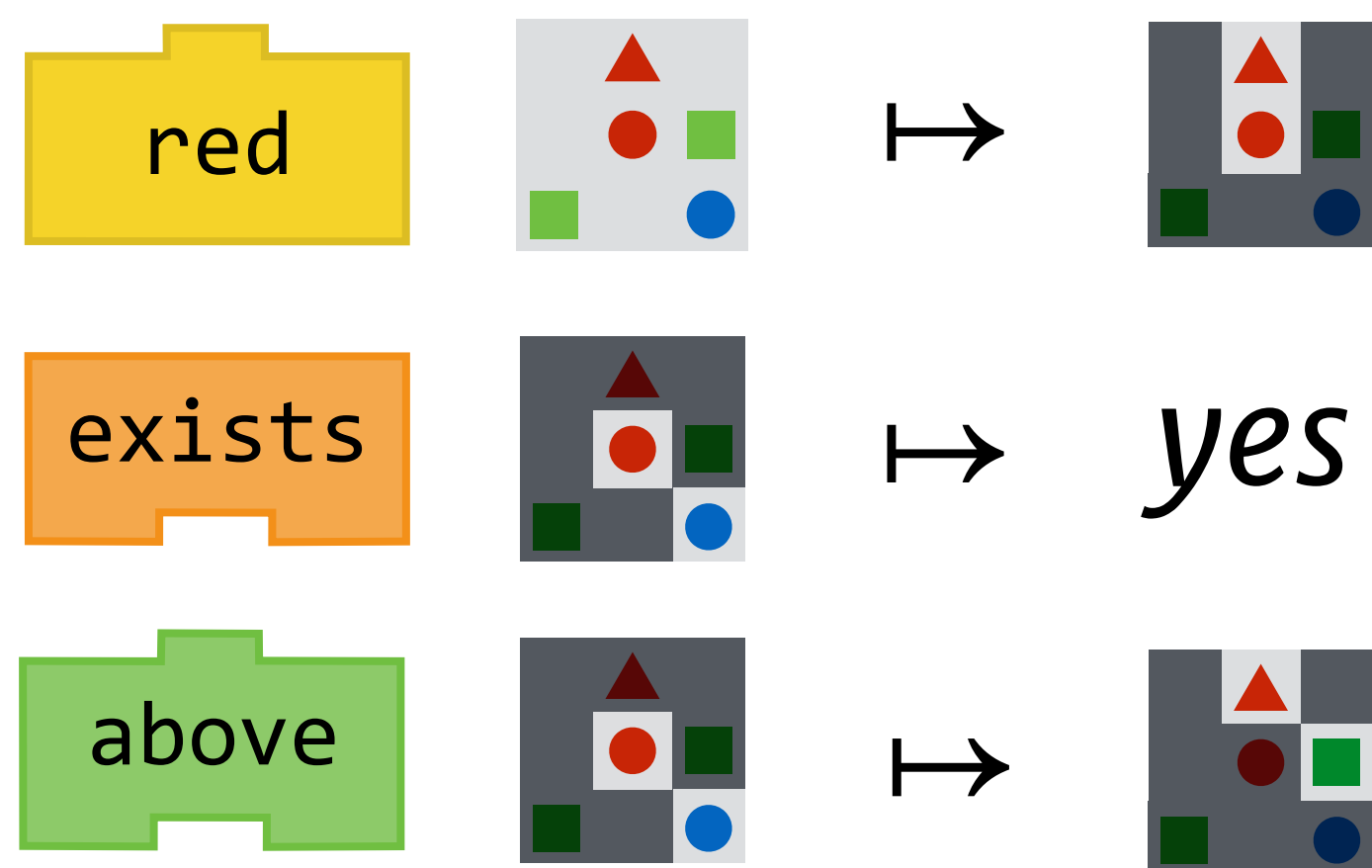


*yellow*



# Neural module networks

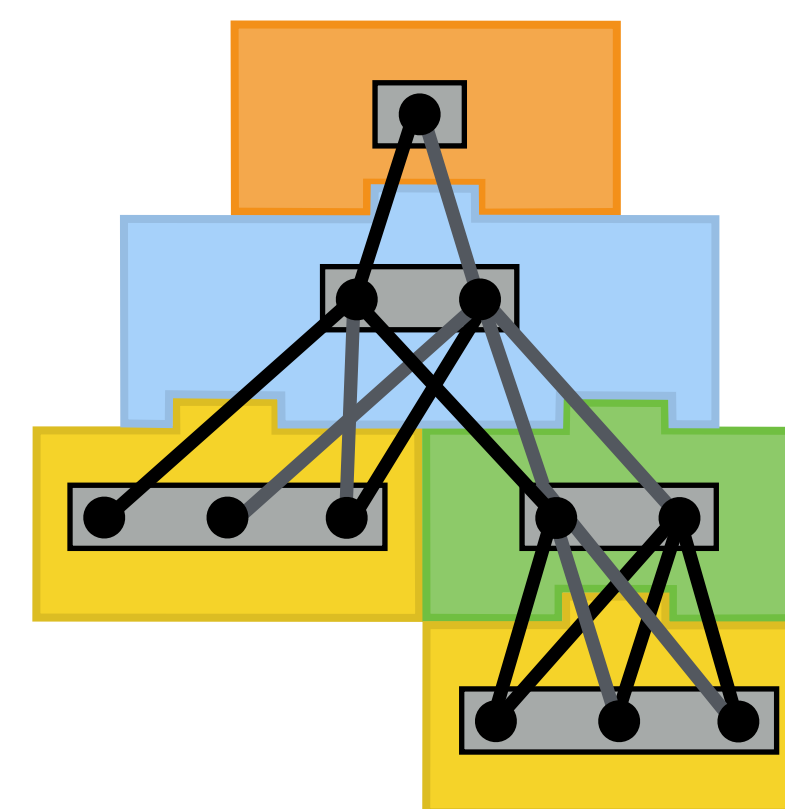
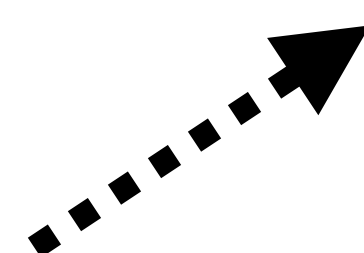
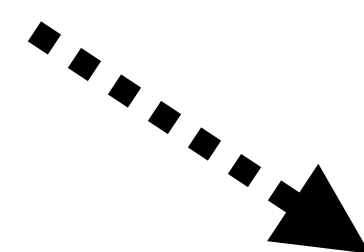
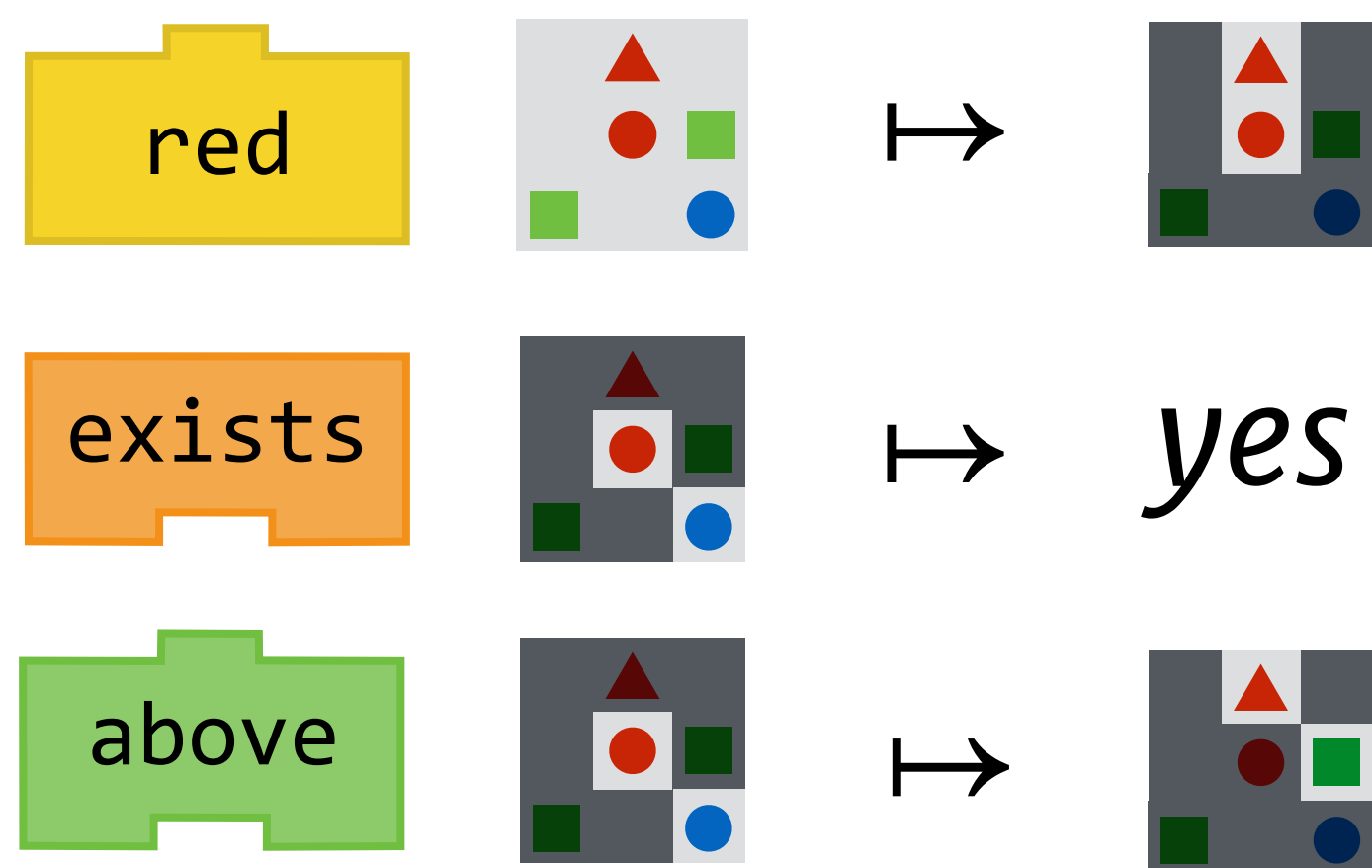
*Is there a red shape  
above a circle?*





# Neural module networks

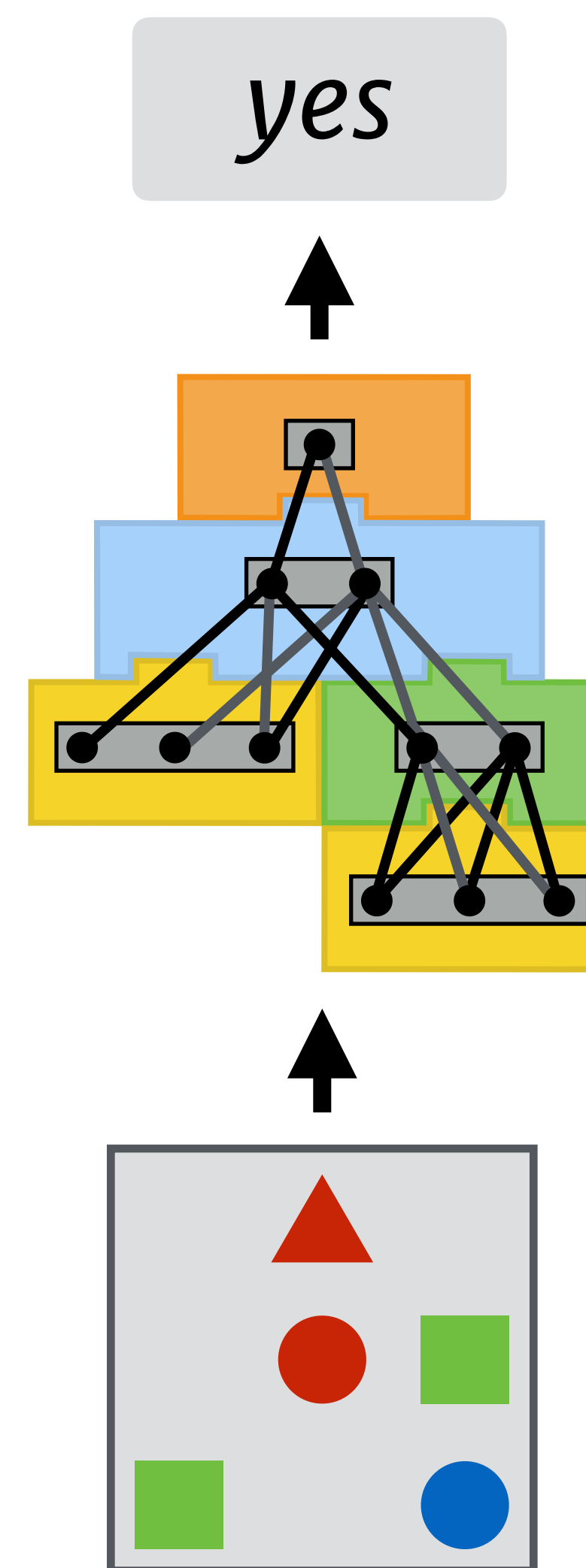
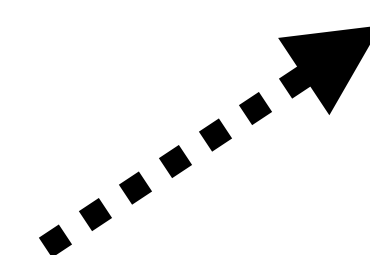
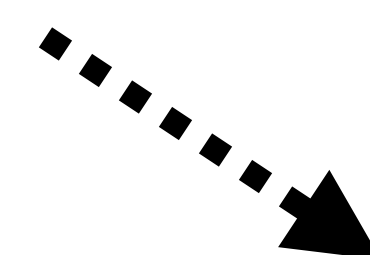
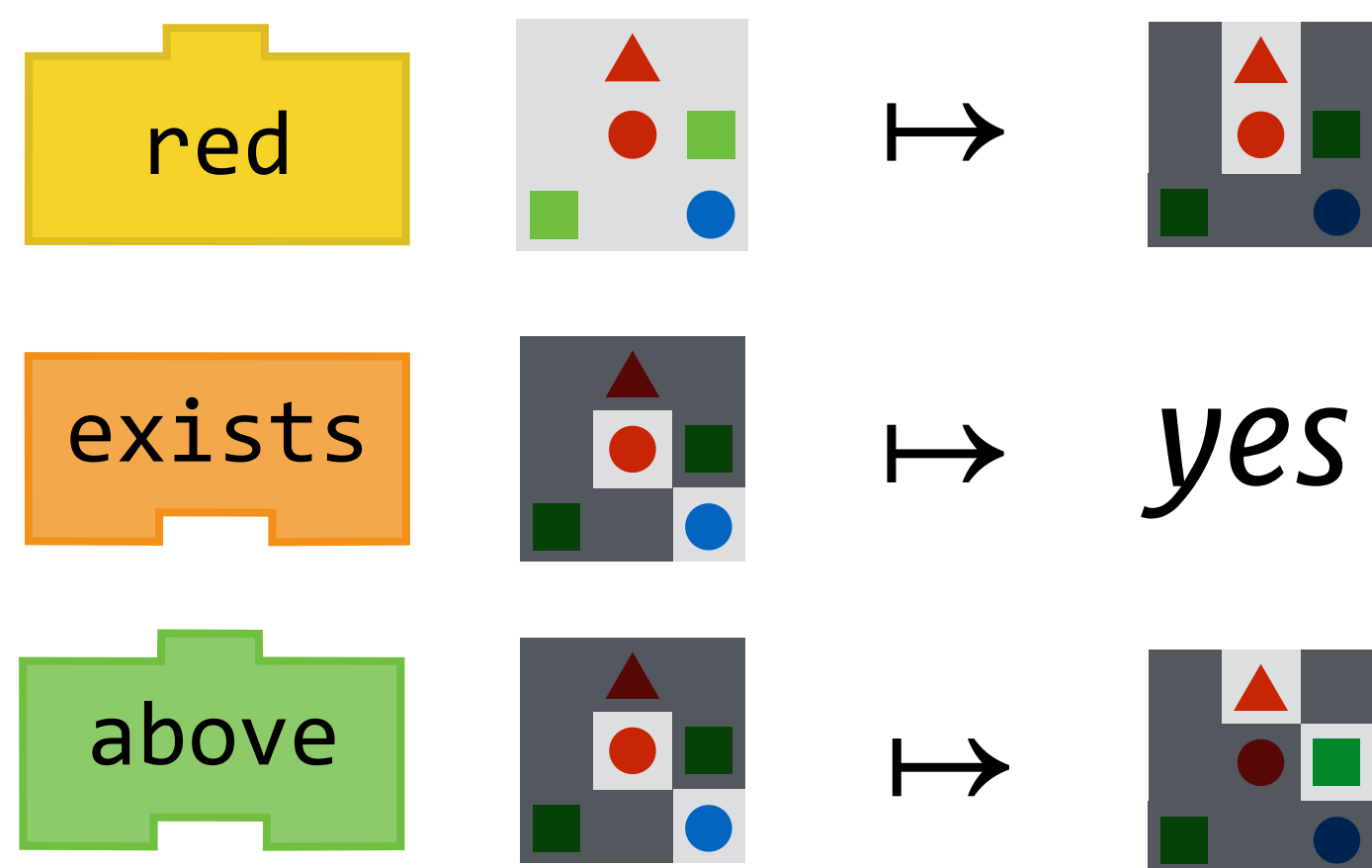
*Is there a red shape  
above a circle?*

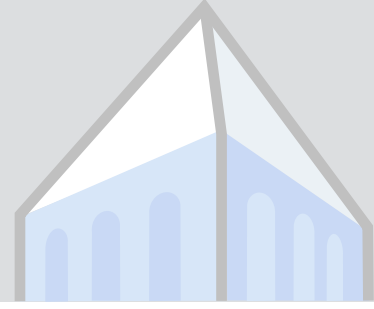




# Neural module networks

*Is there a red shape  
above a circle?*





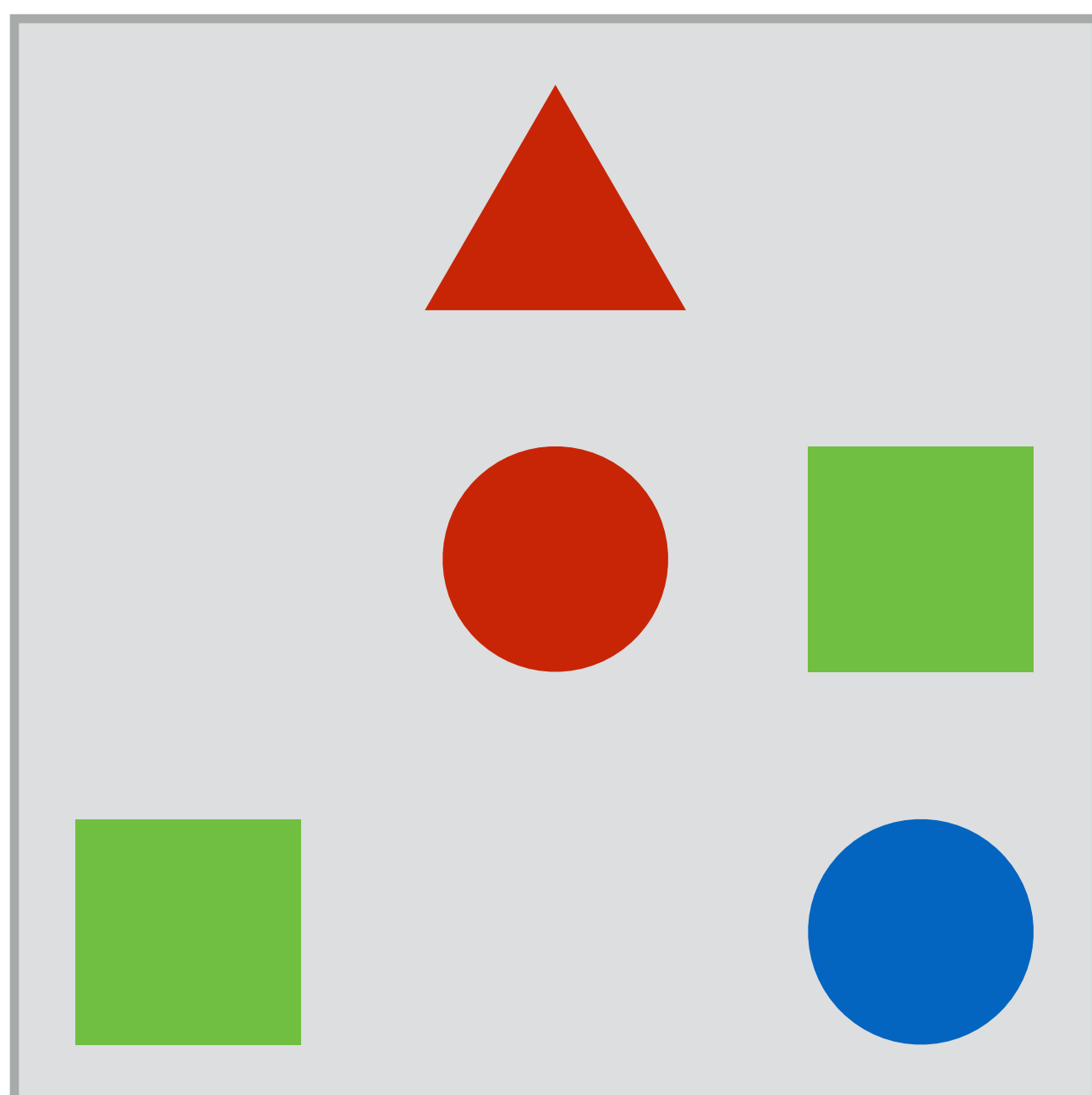
# Perceptual primitives



*What color is the necktie?*



# Perceptual primitives

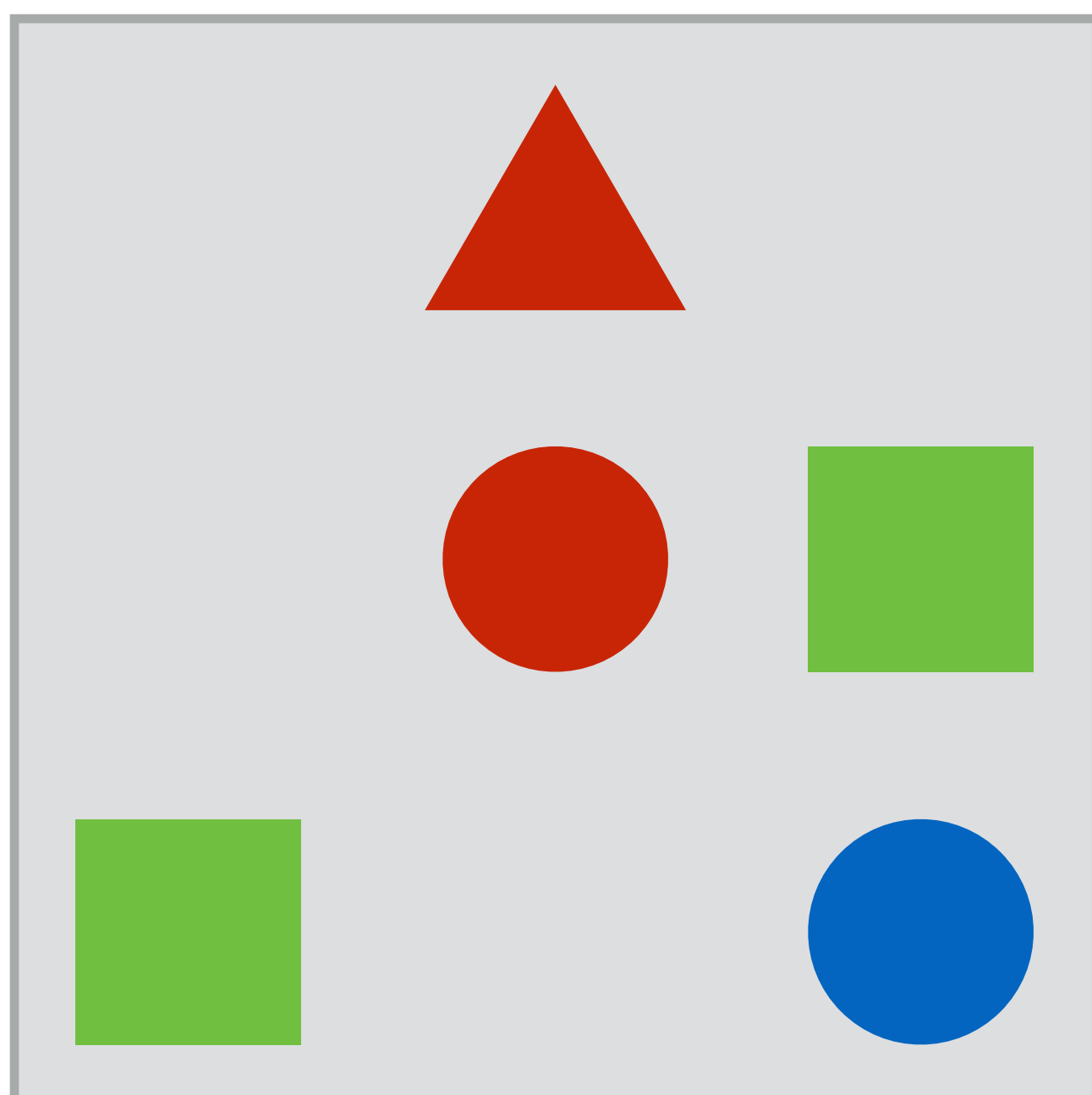


*Is there a red shape above a circle?*





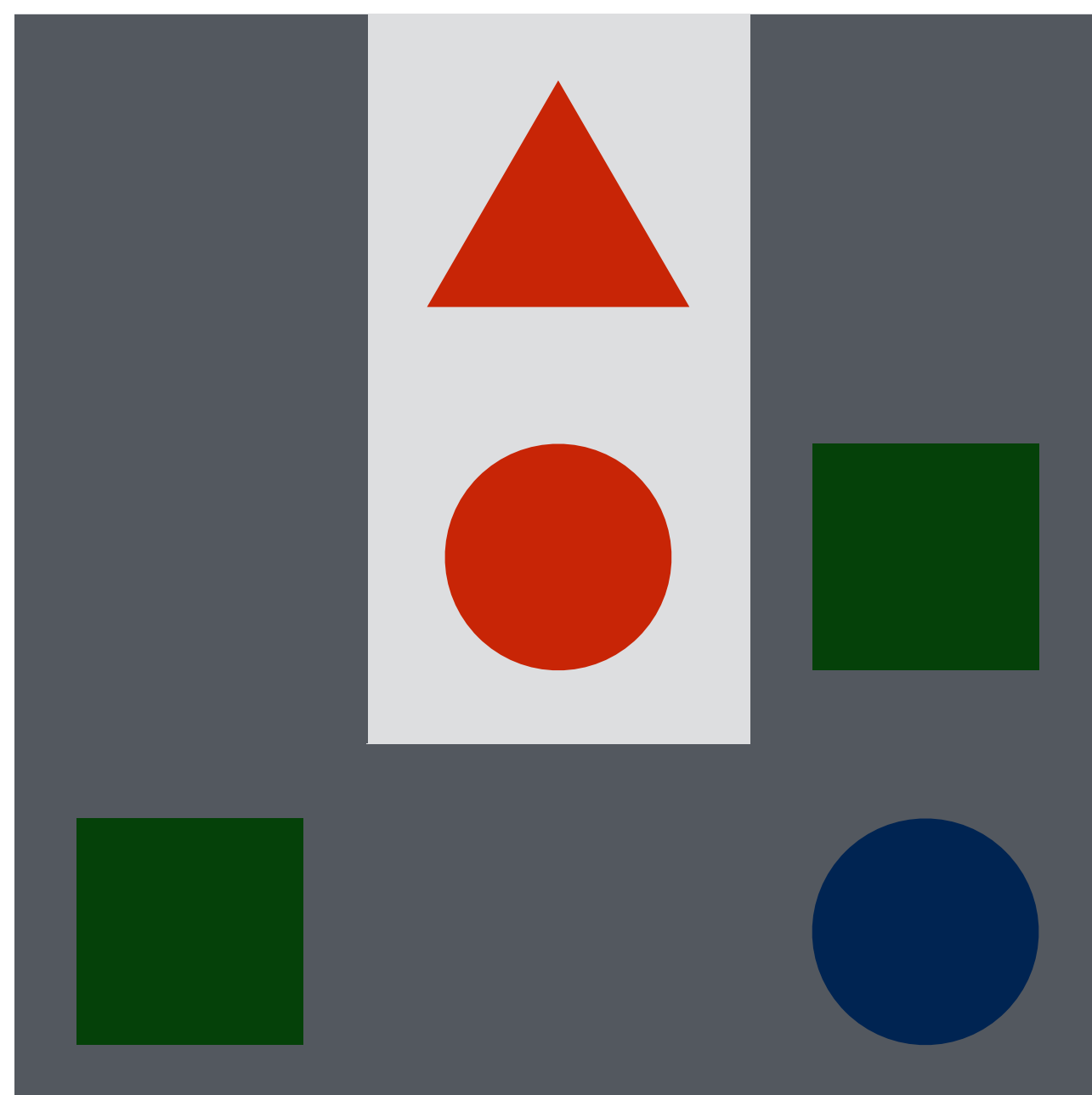
# Perceptual primitives



*Is there a **red** shape above a circle?*



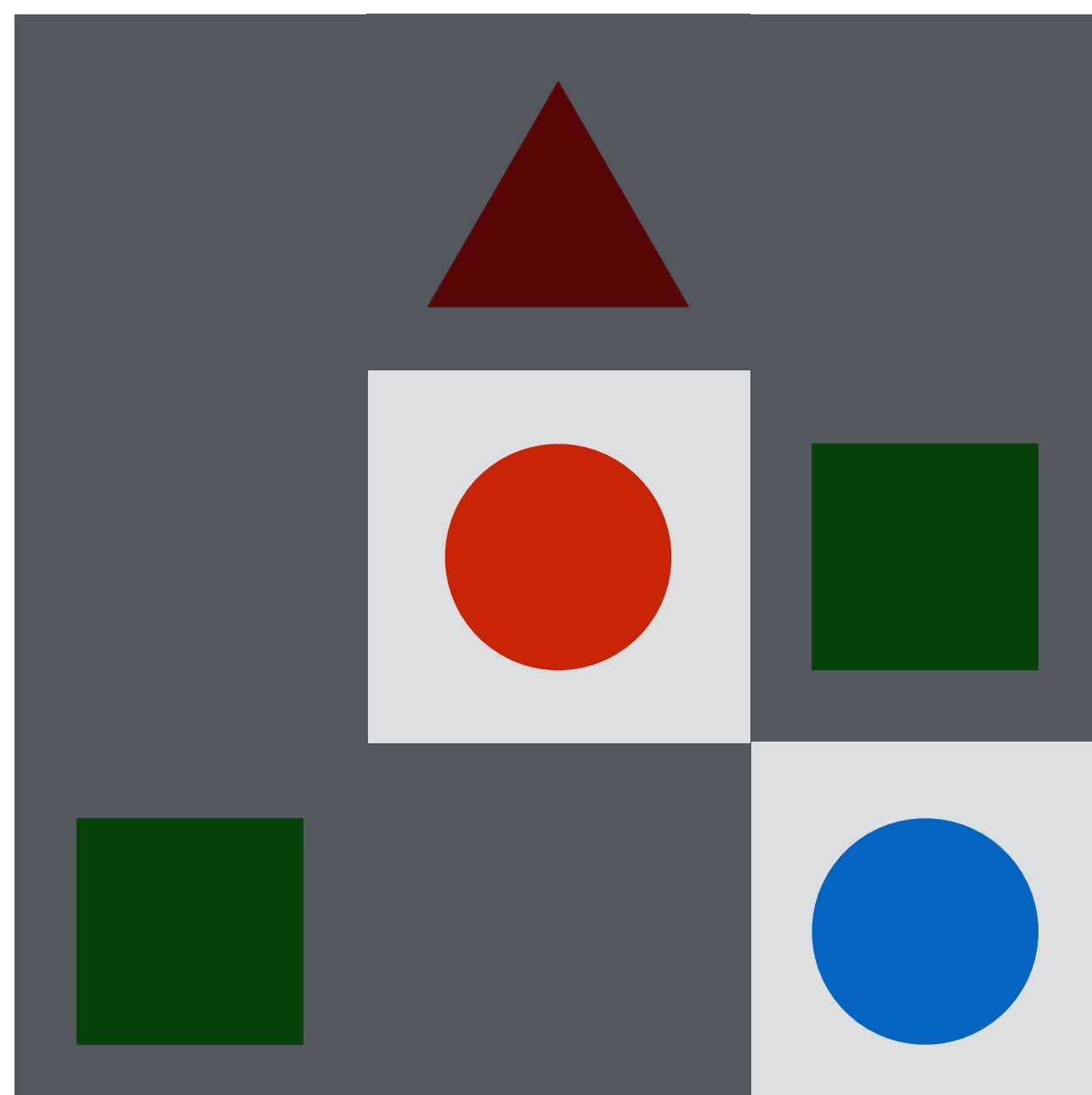
# Perceptual primitives



*Is there a **red** shape above a circle?*



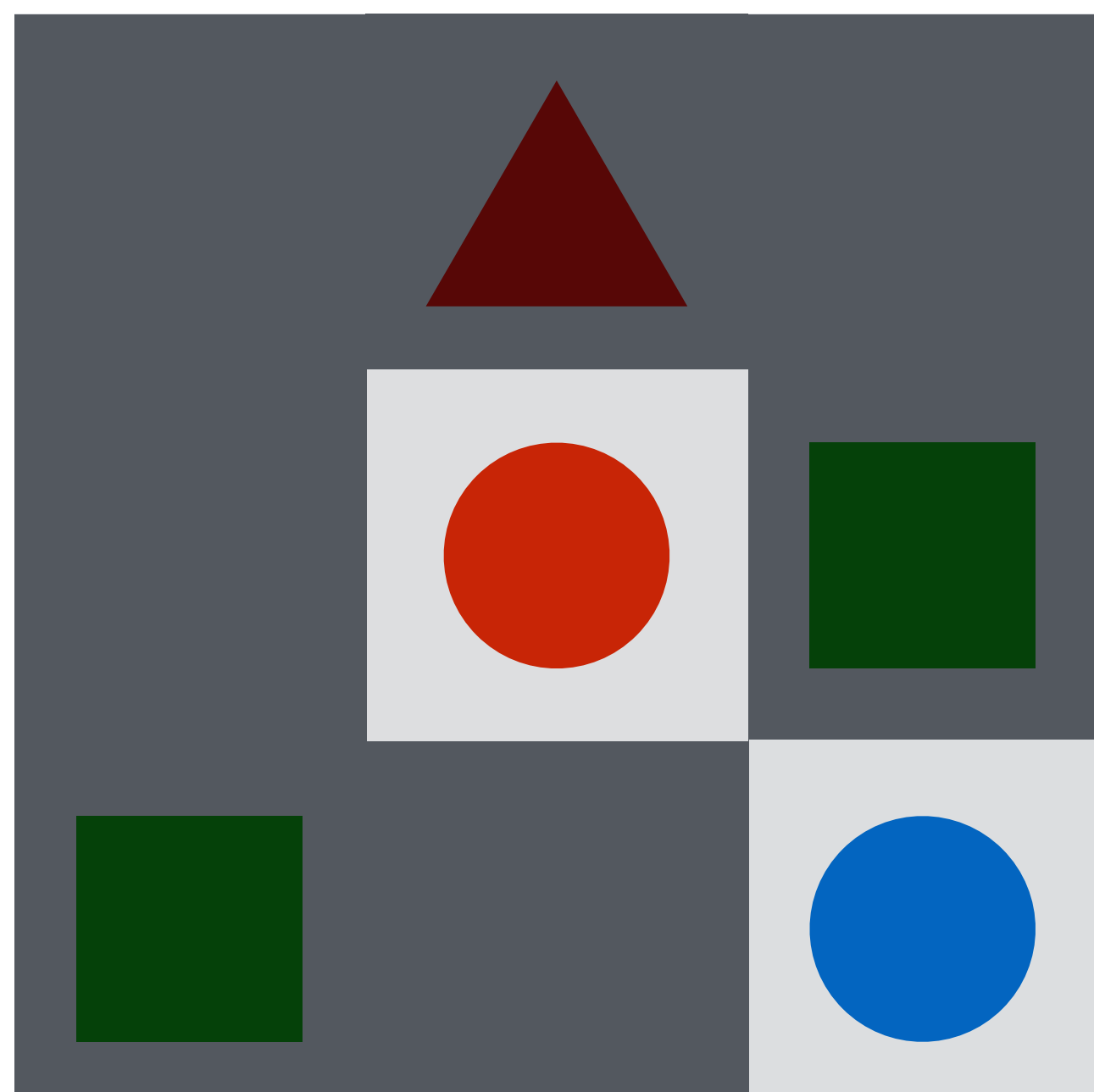
# Perceptual primitives



*Is there a red shape above a **circle**?*



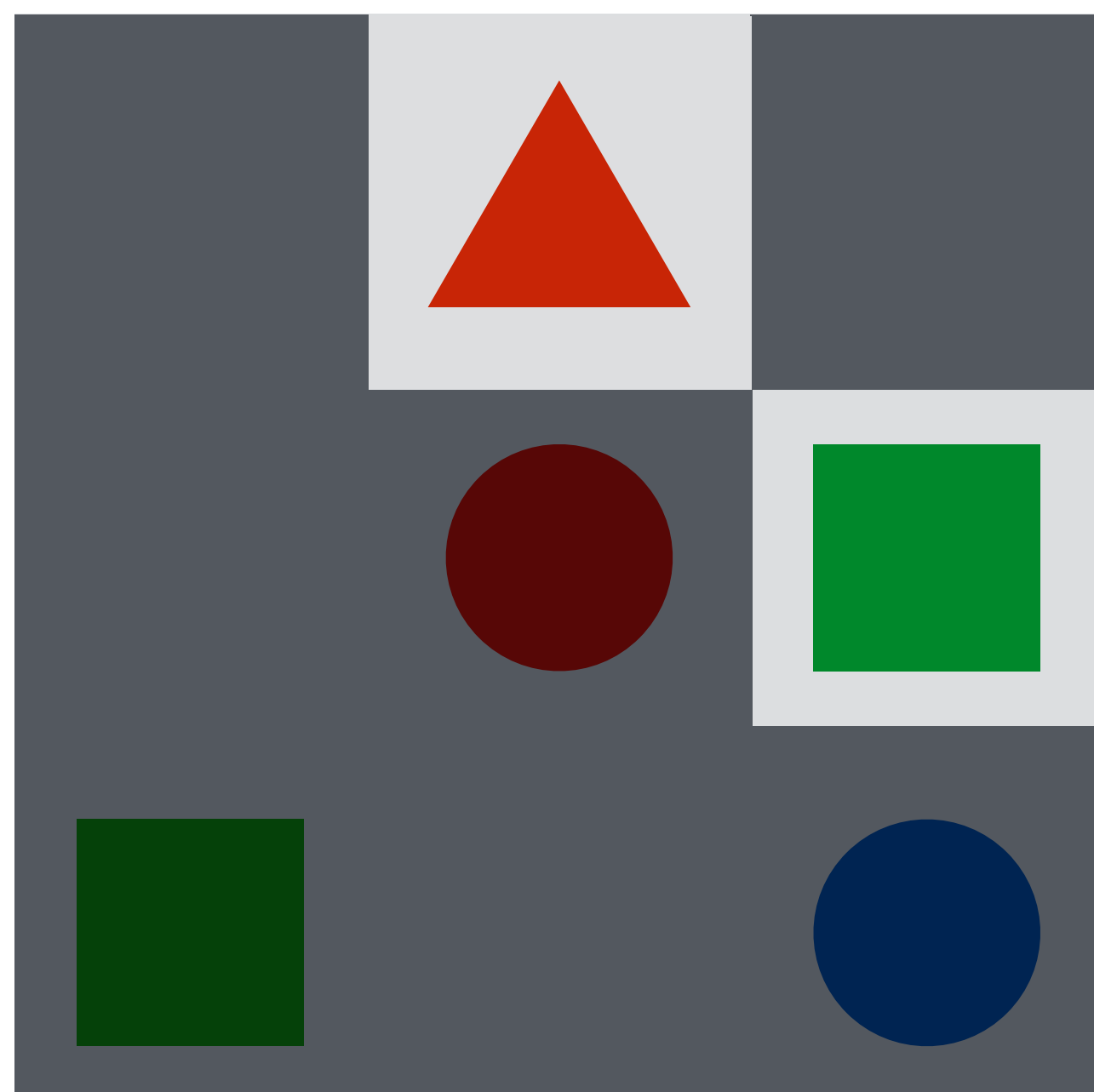
# Perceptual primitives



*Is there a red shape **above a circle?***



# Perceptual primitives

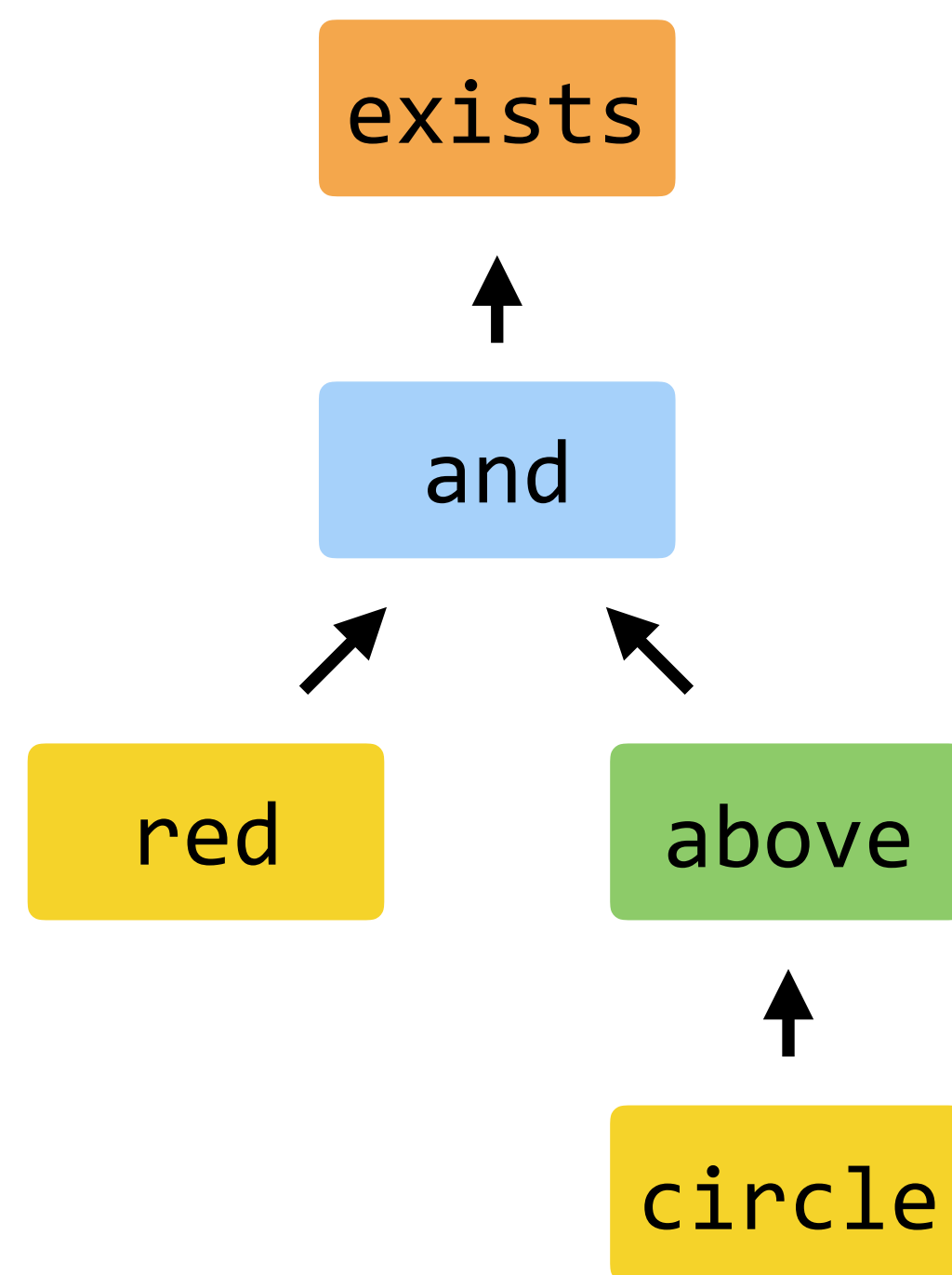
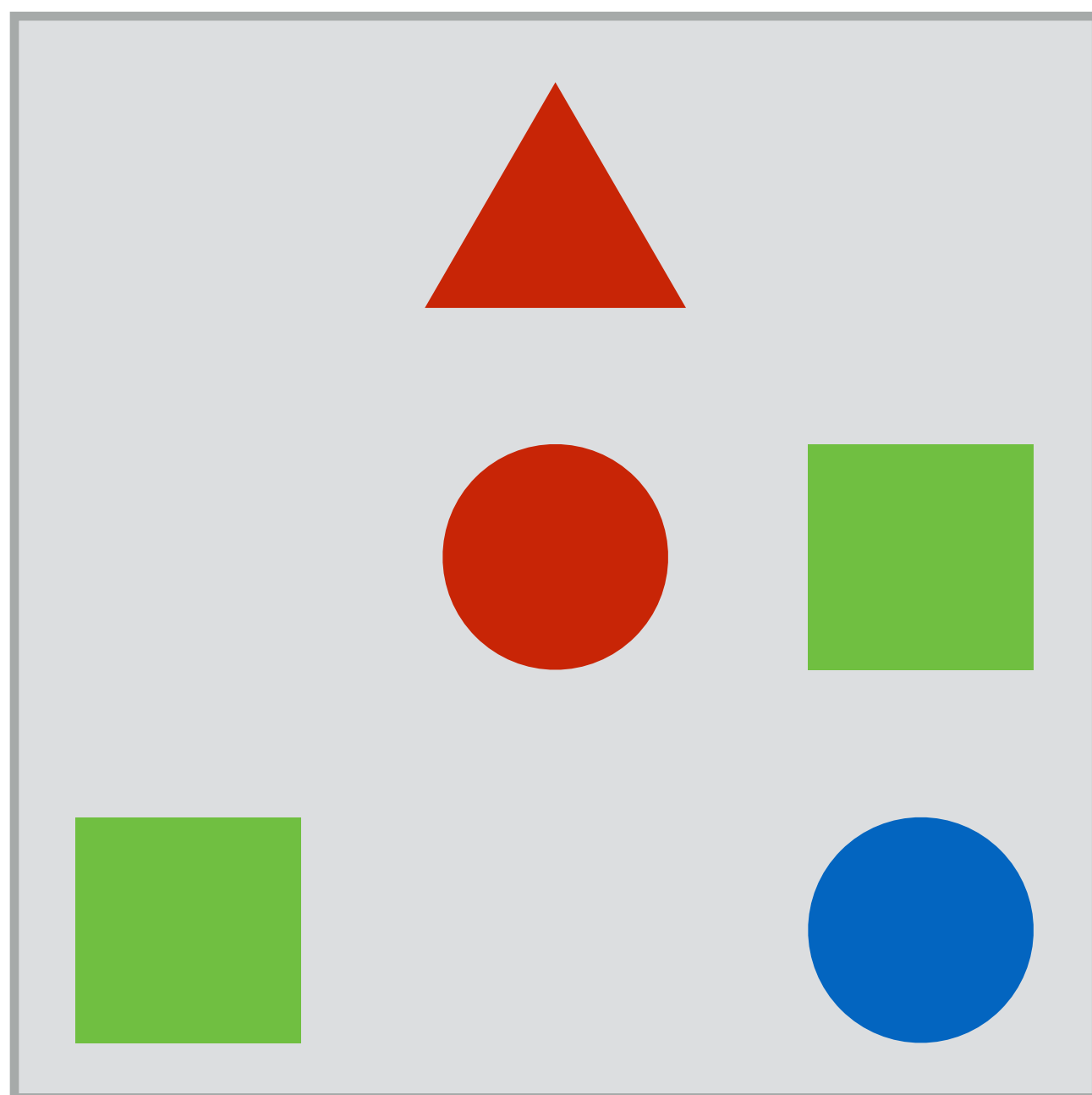


*Is there a red shape **above a circle**?*



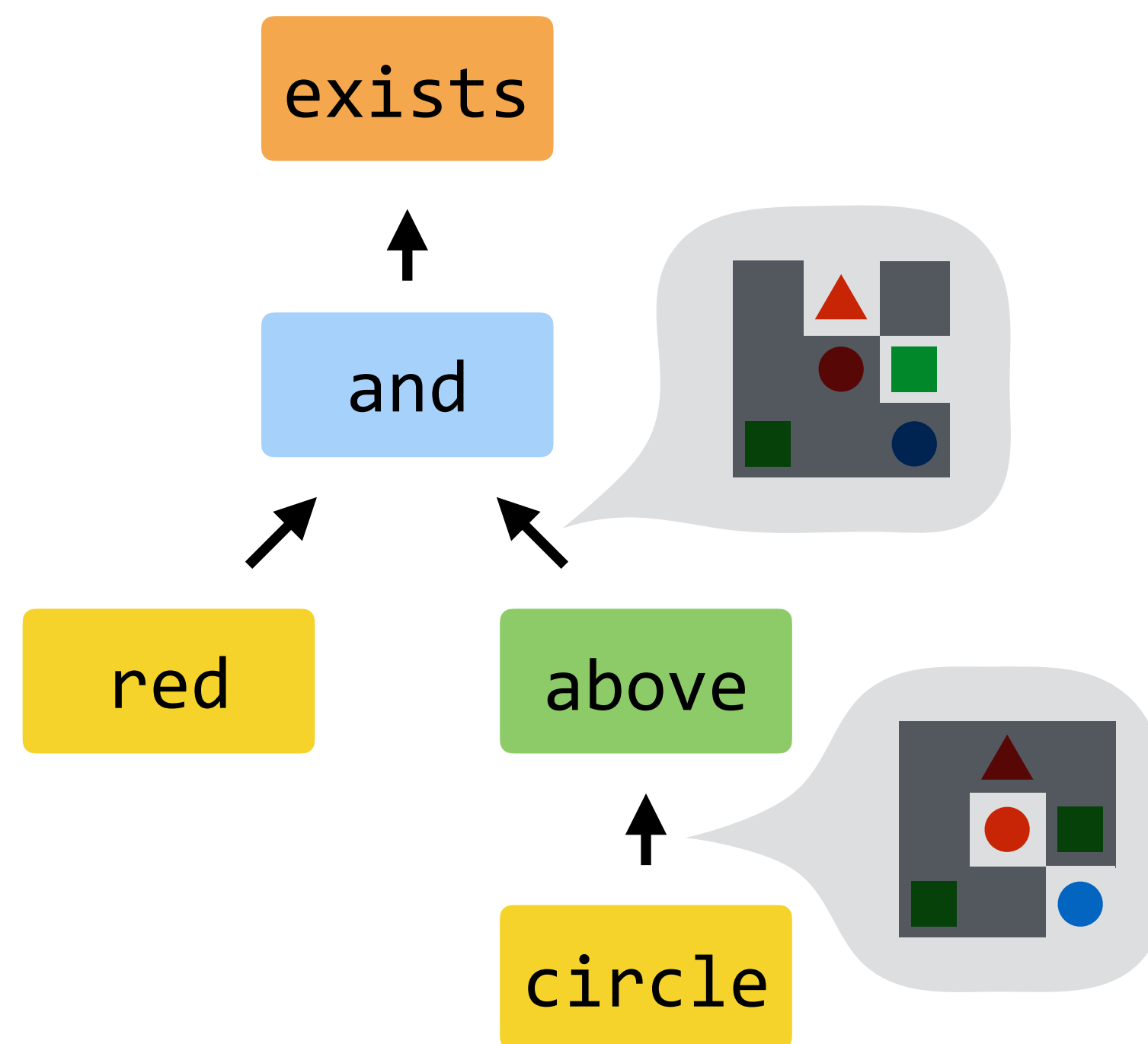
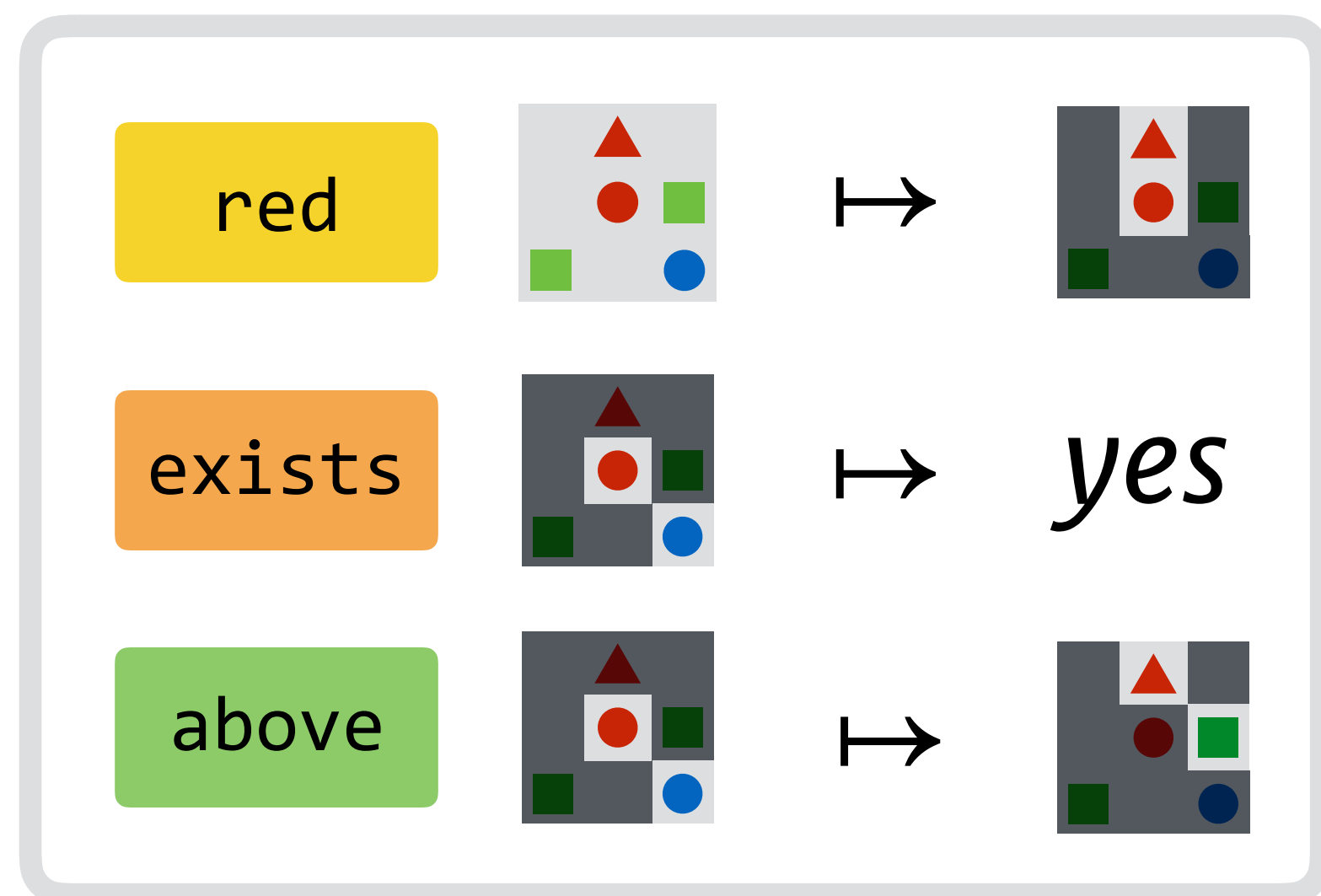
# Meanings are computations

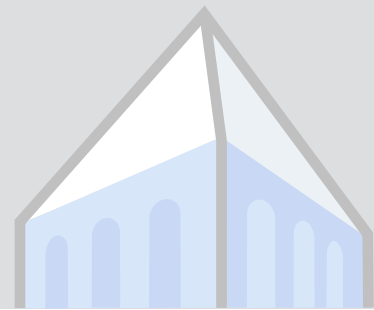
*Is there a red shape above a circle?*





# Meanings are computations



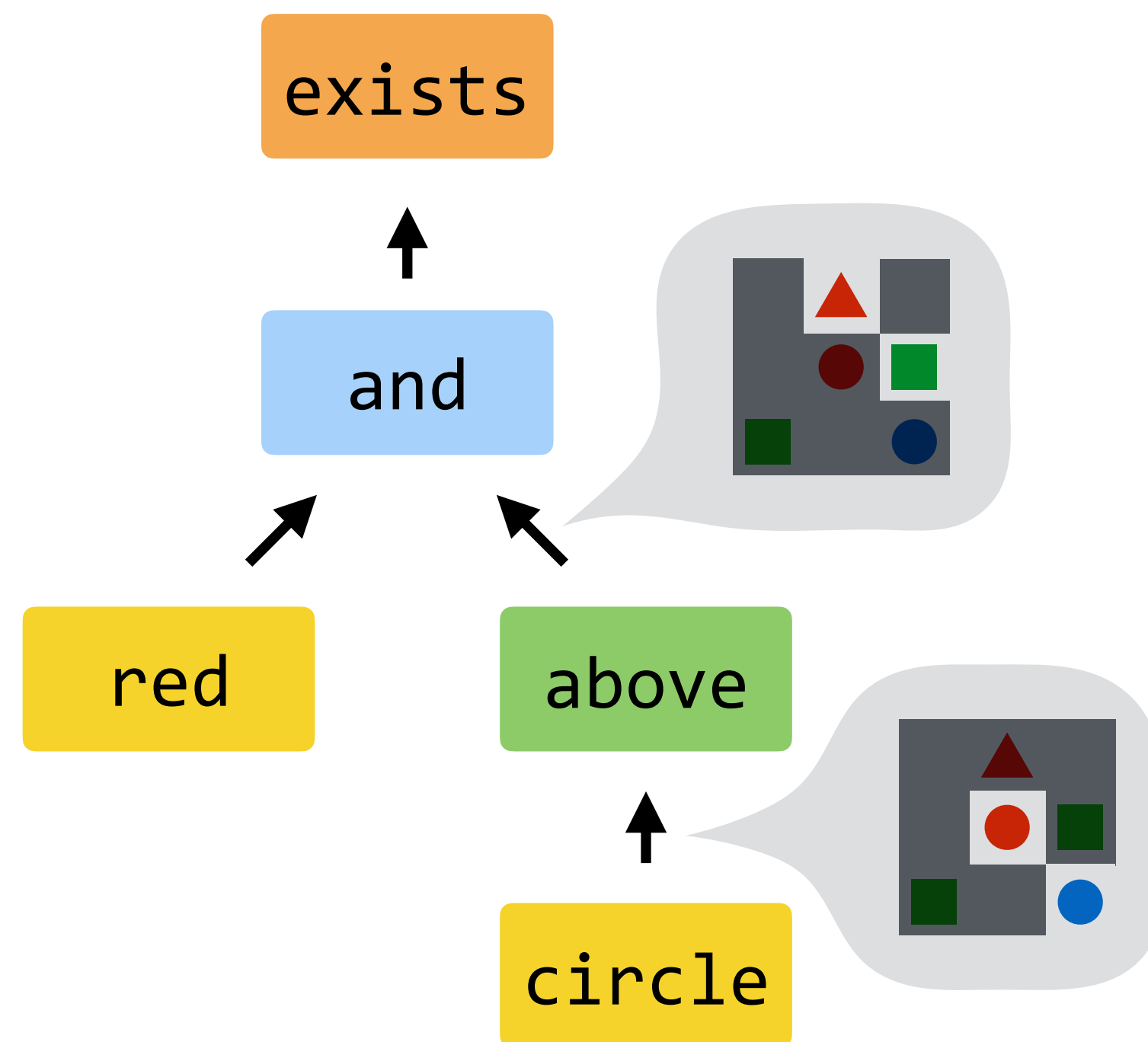


# Learning compositional operators

```
shapes.where(_.color == "red")
```

```
d => d.nonEmpty ? true : false
```

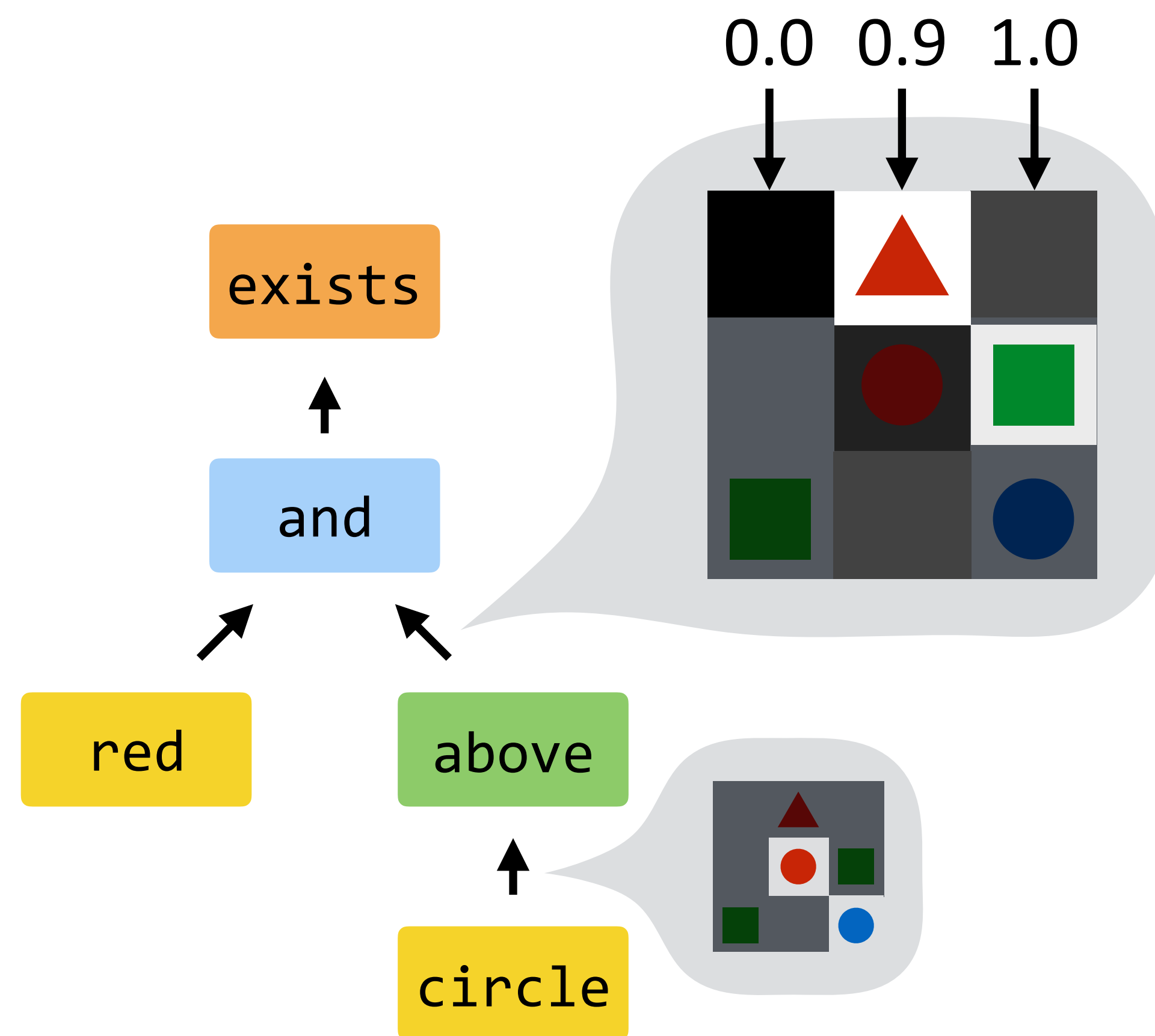
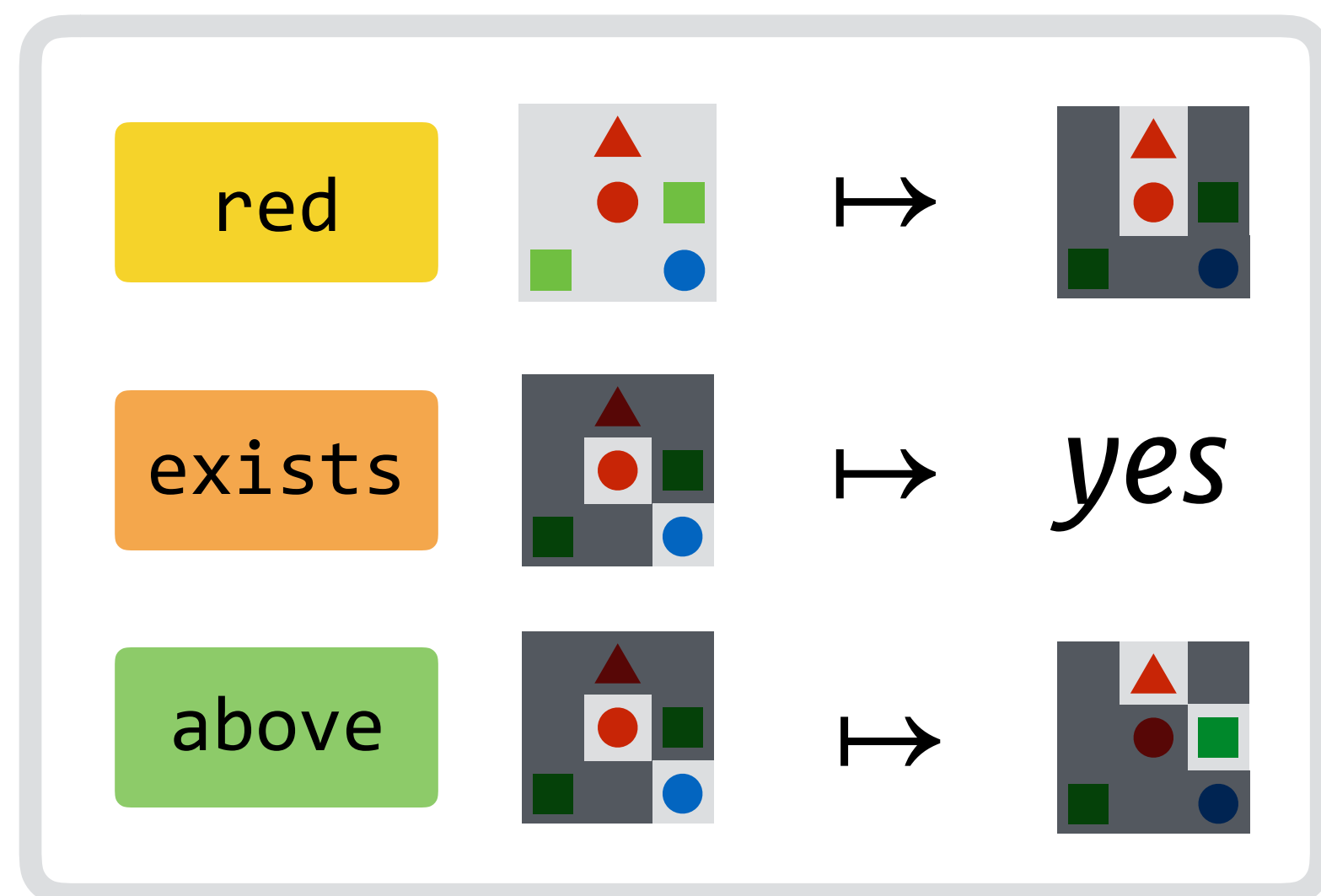
```
d => d.map(_.neighborAbove)
```





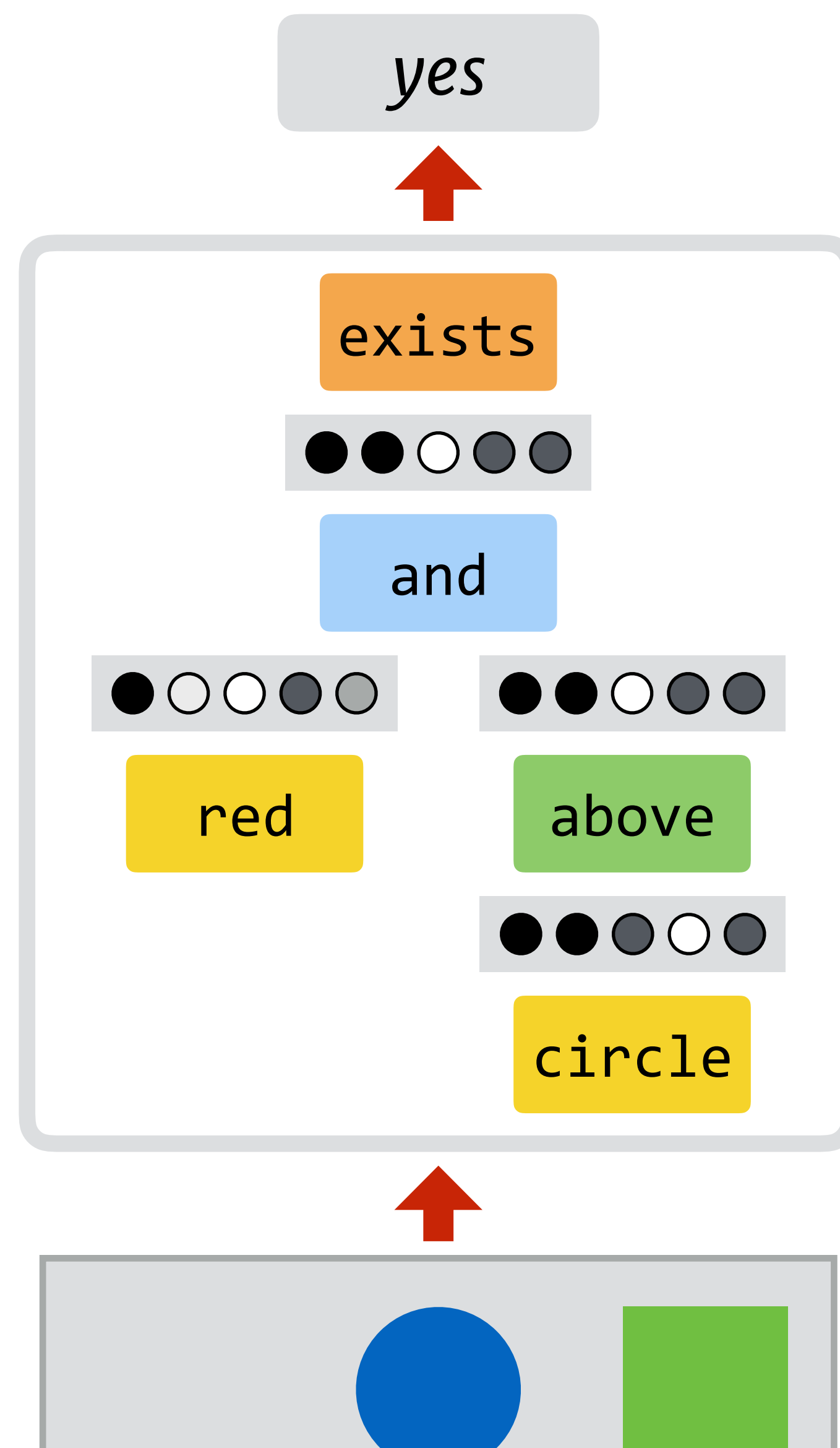
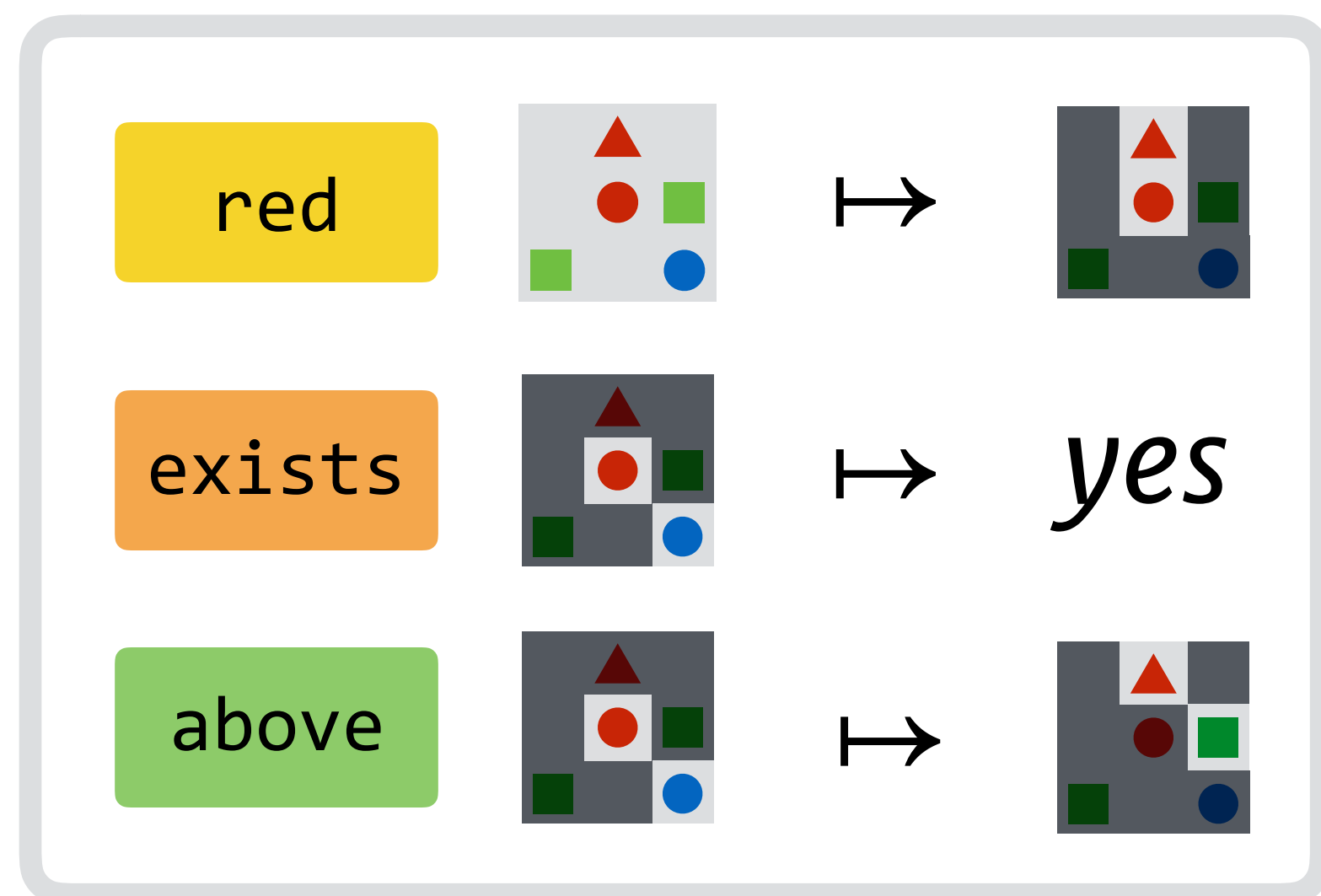


# Learning compositional operators



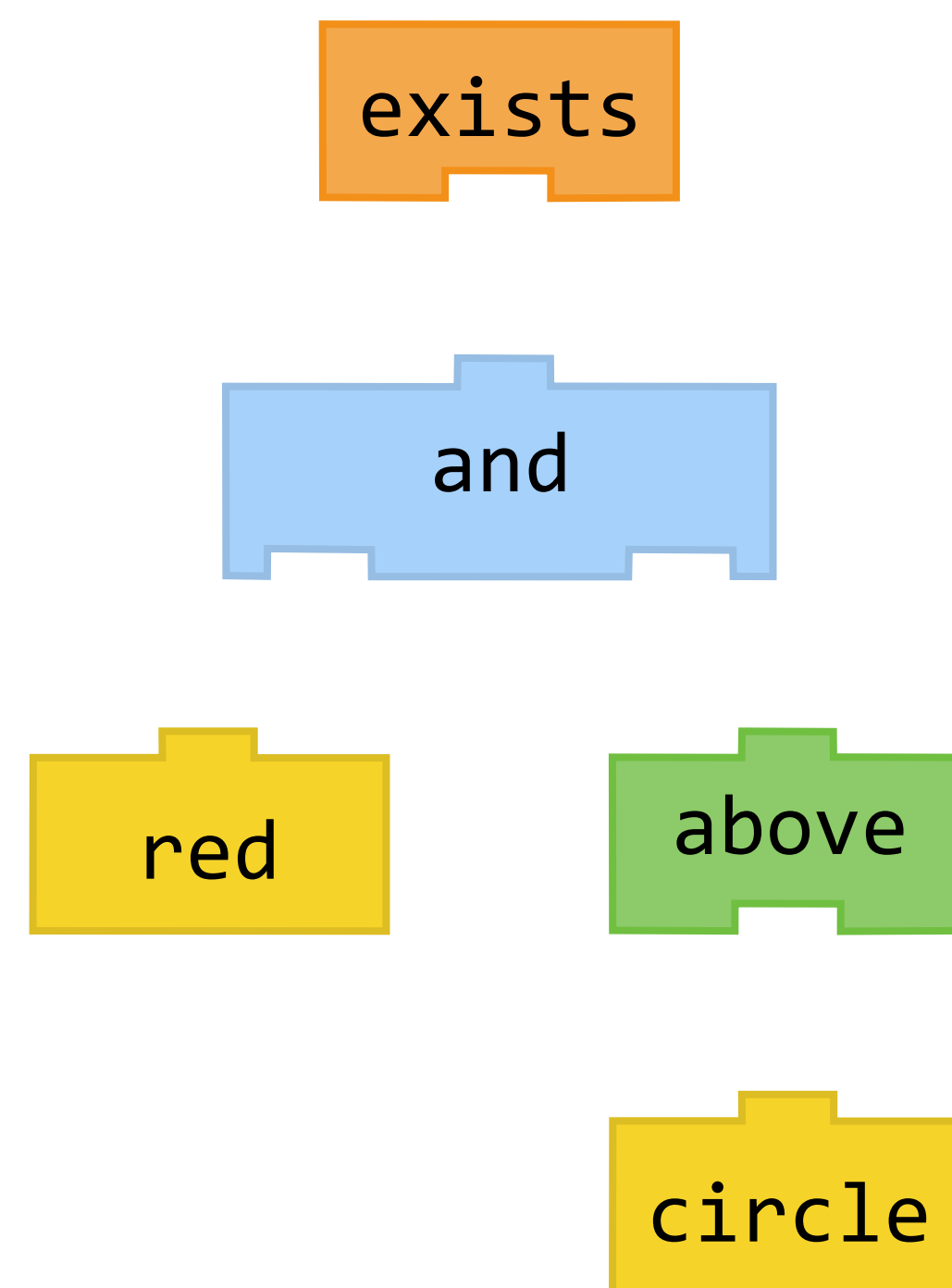
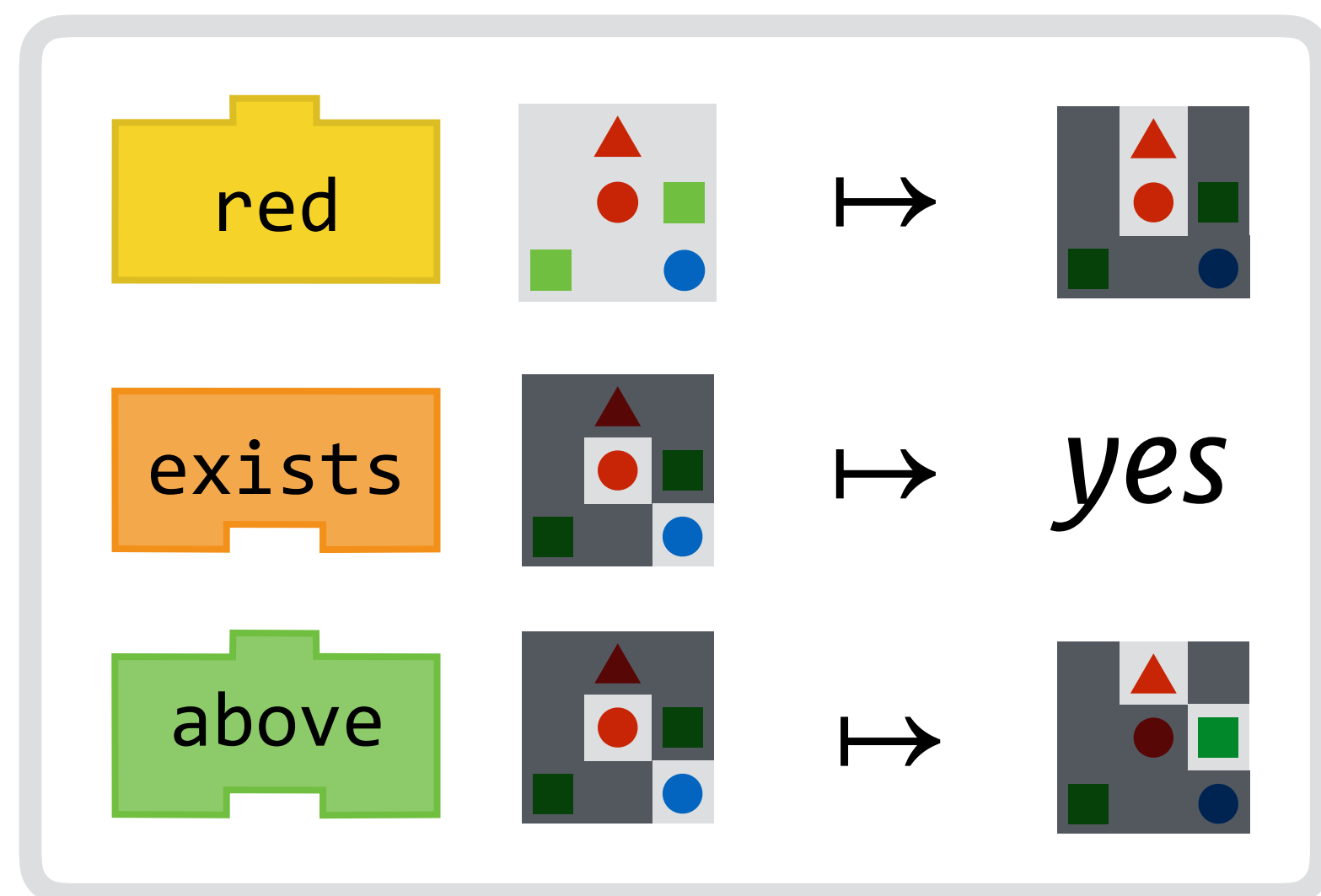


# Learning compositional operators



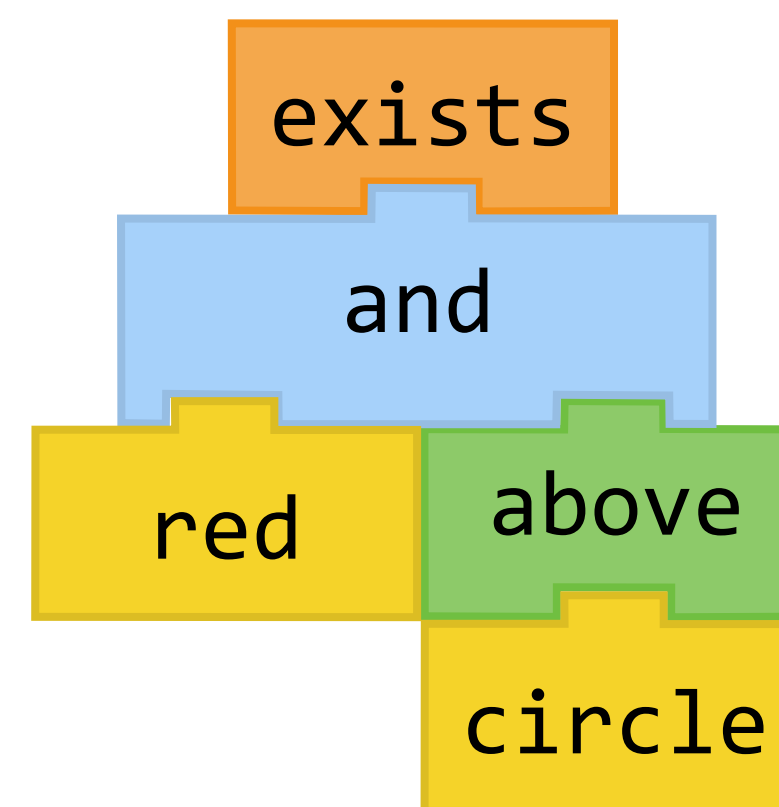
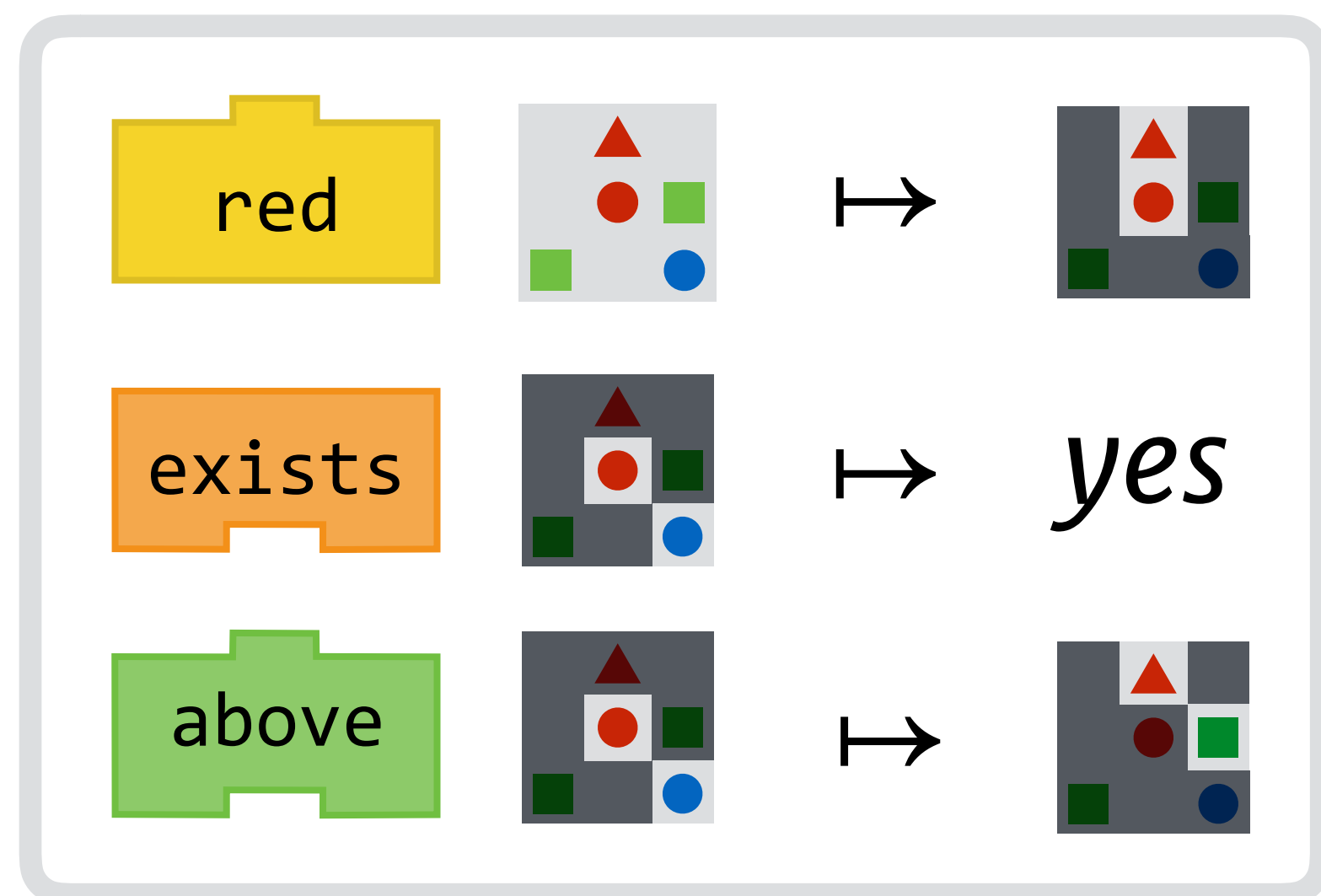


# Composing neural networks



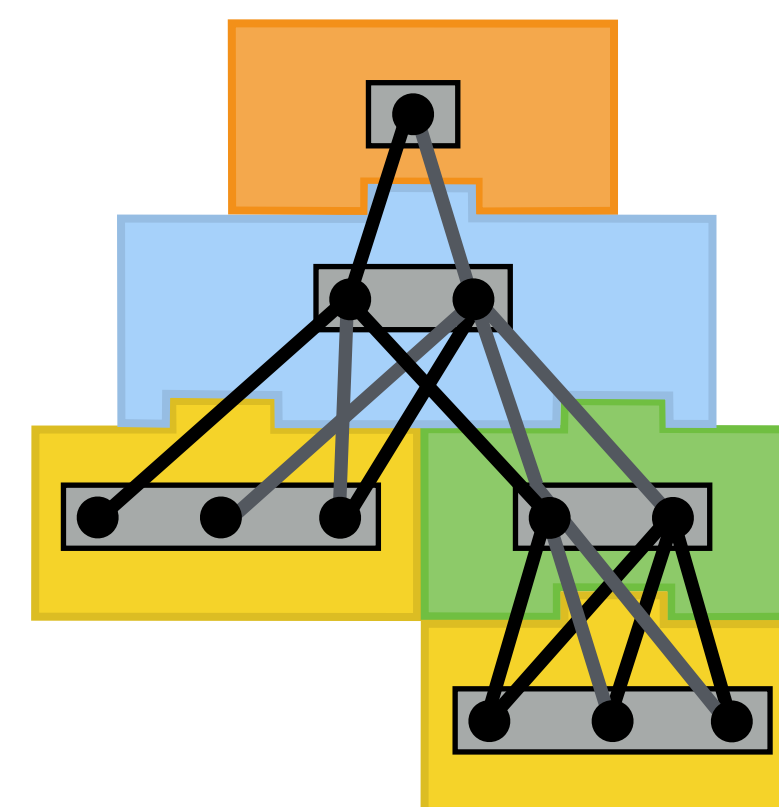
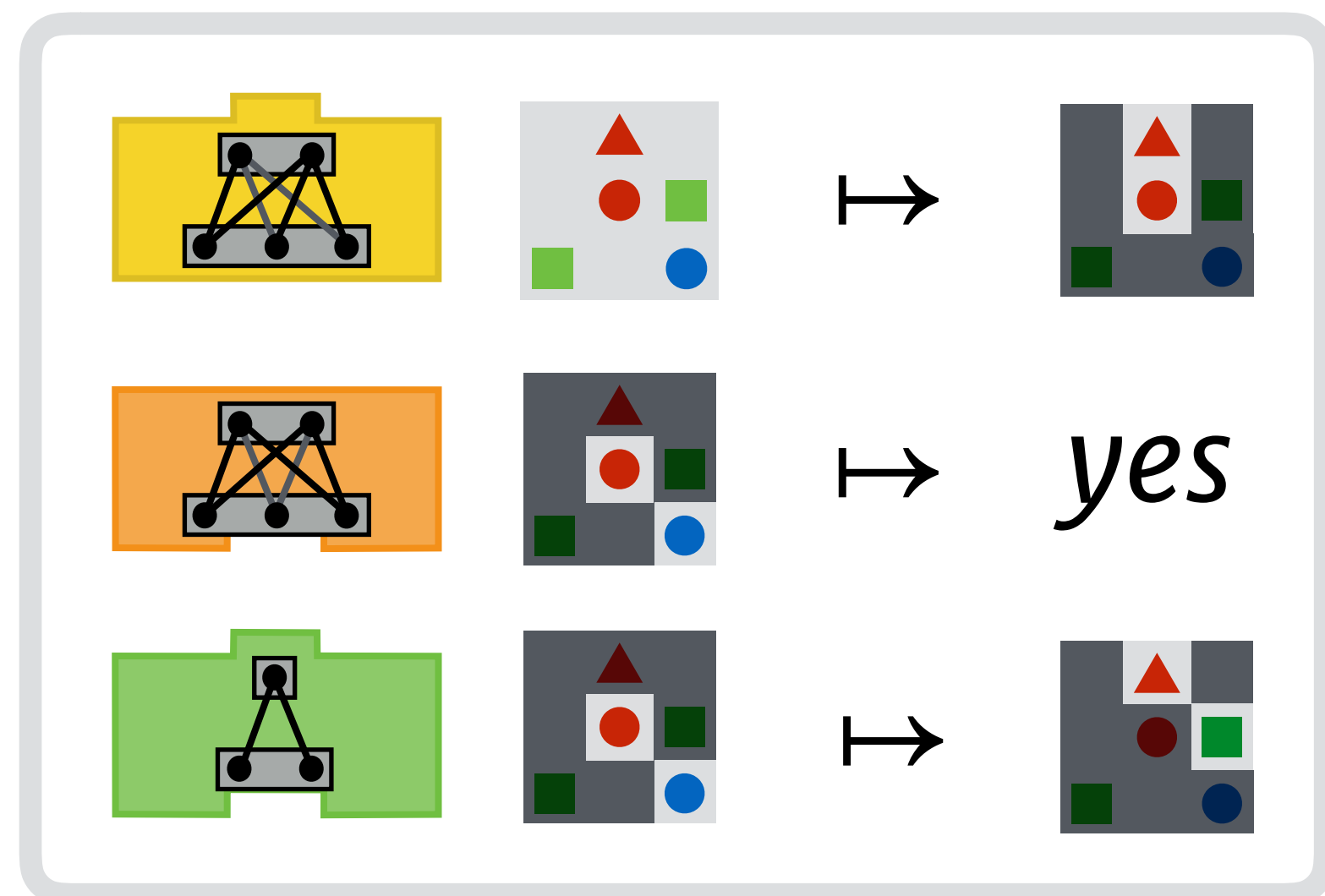


# Composing neural networks



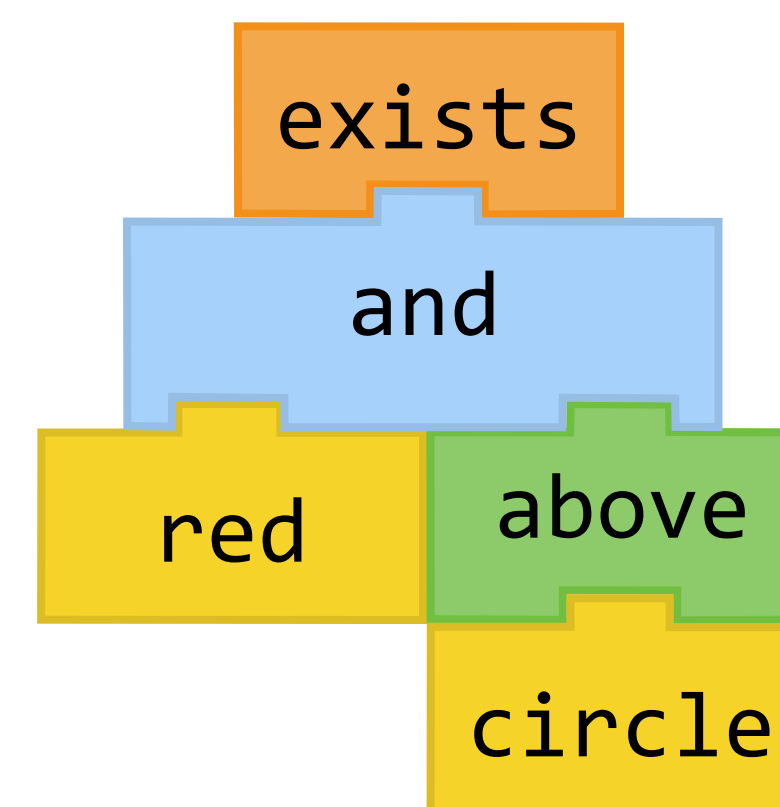
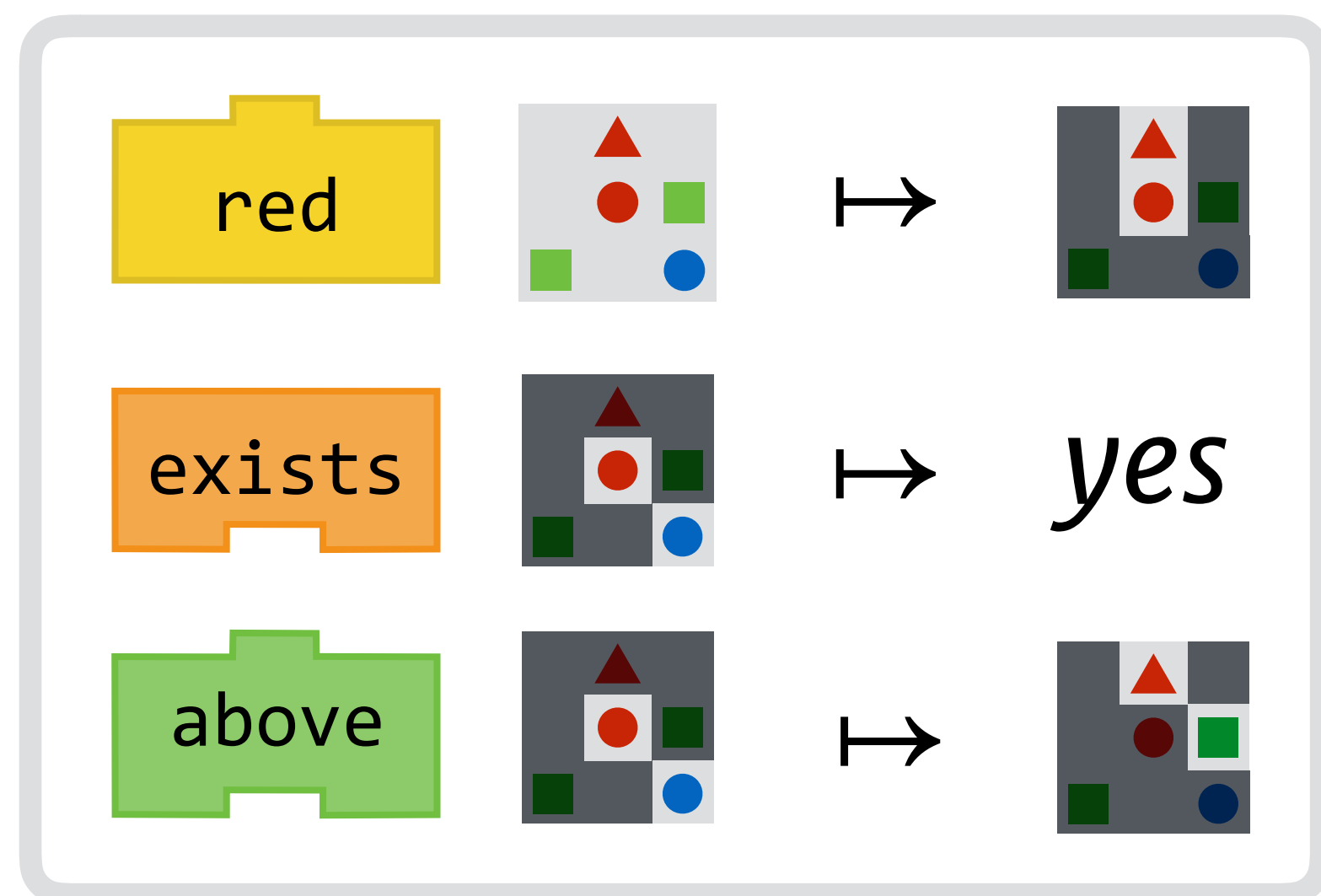


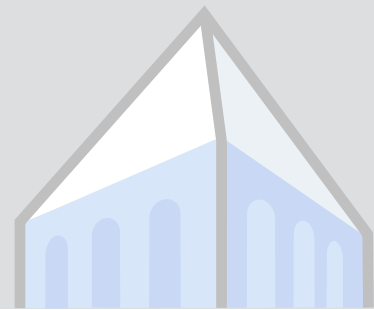
# Composing neural networks



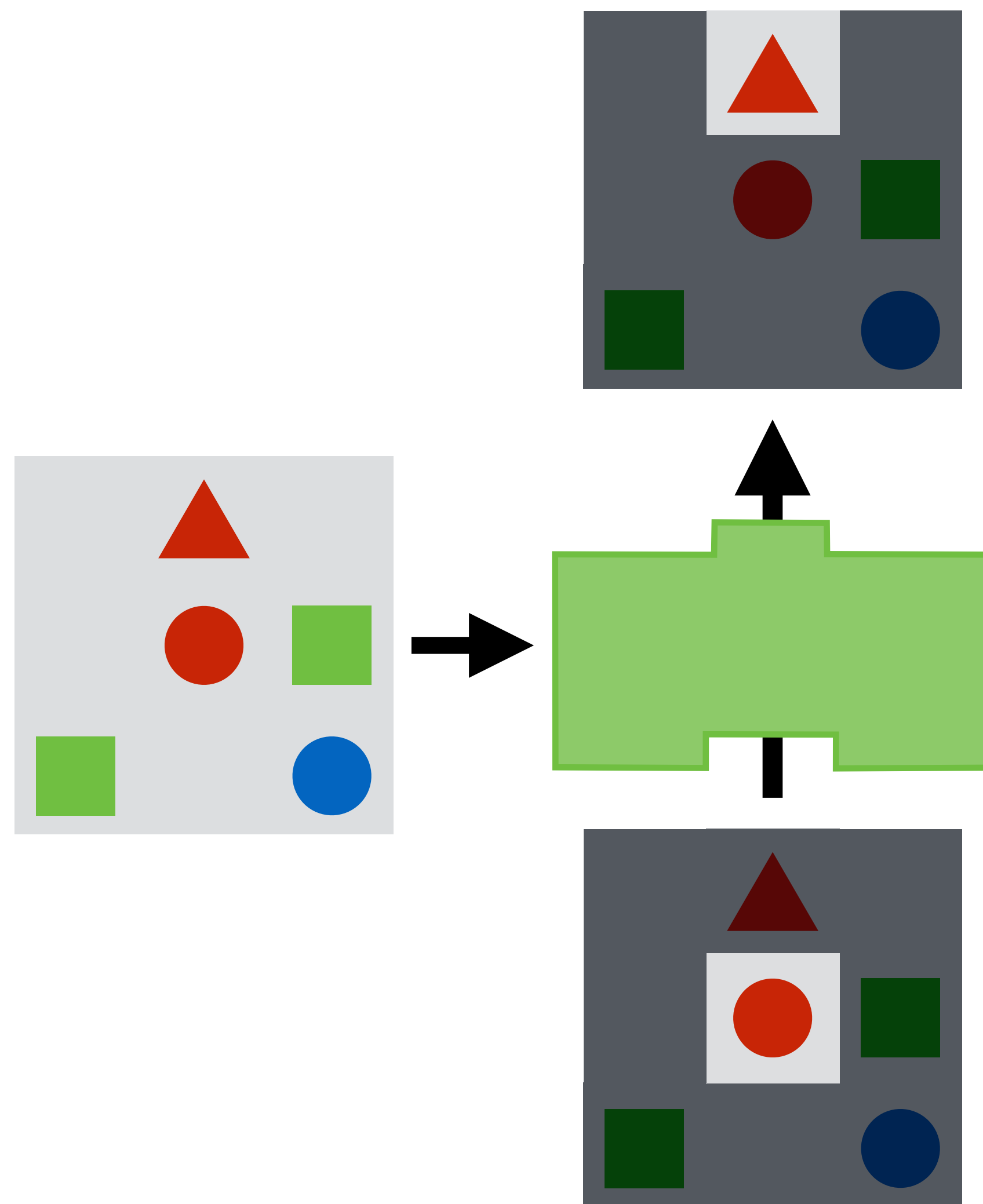


# Composing neural networks



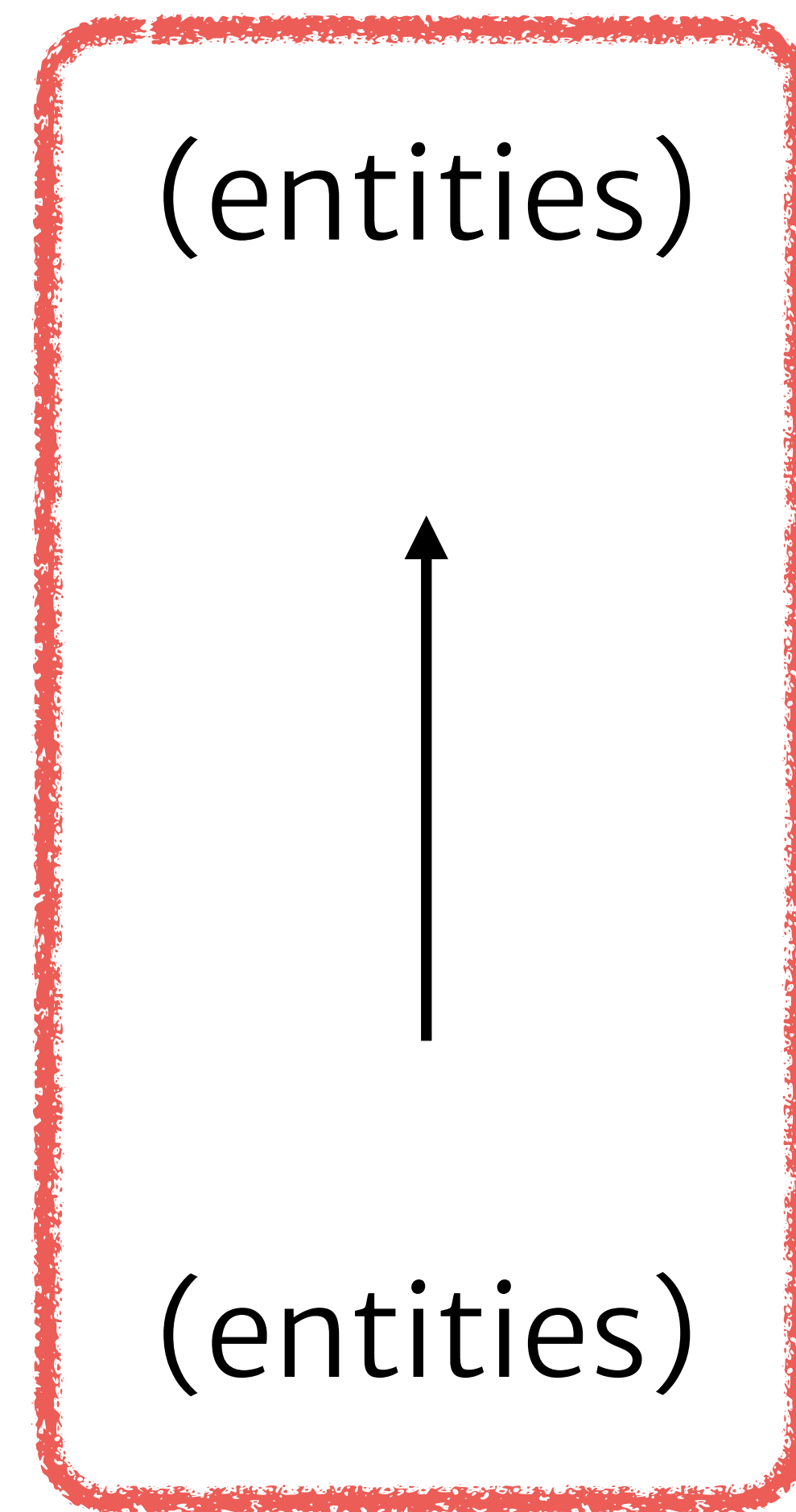
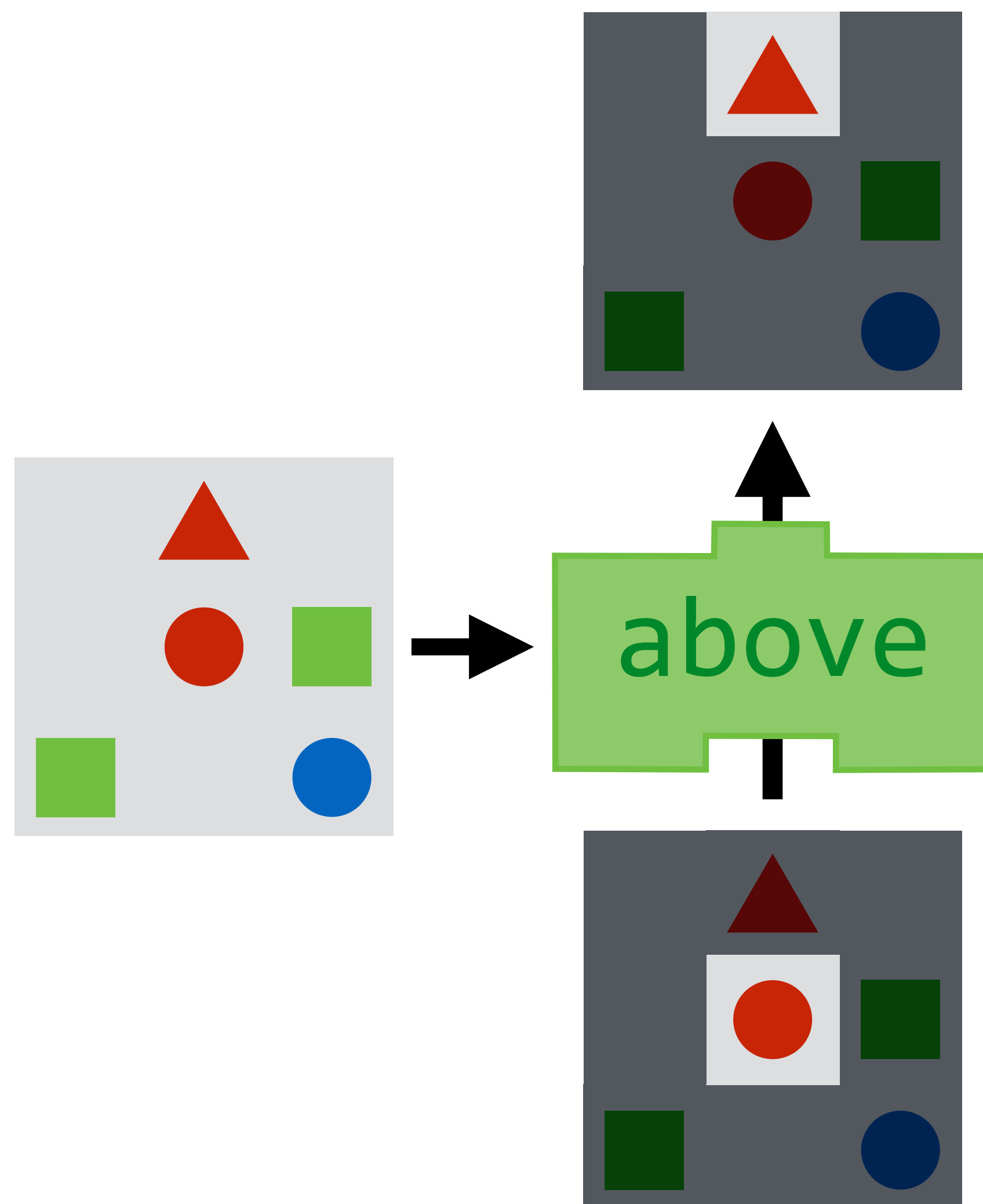


# Anatomy of a module: **Types**

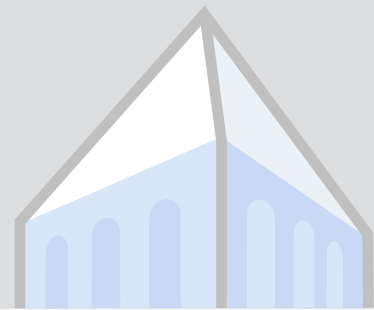




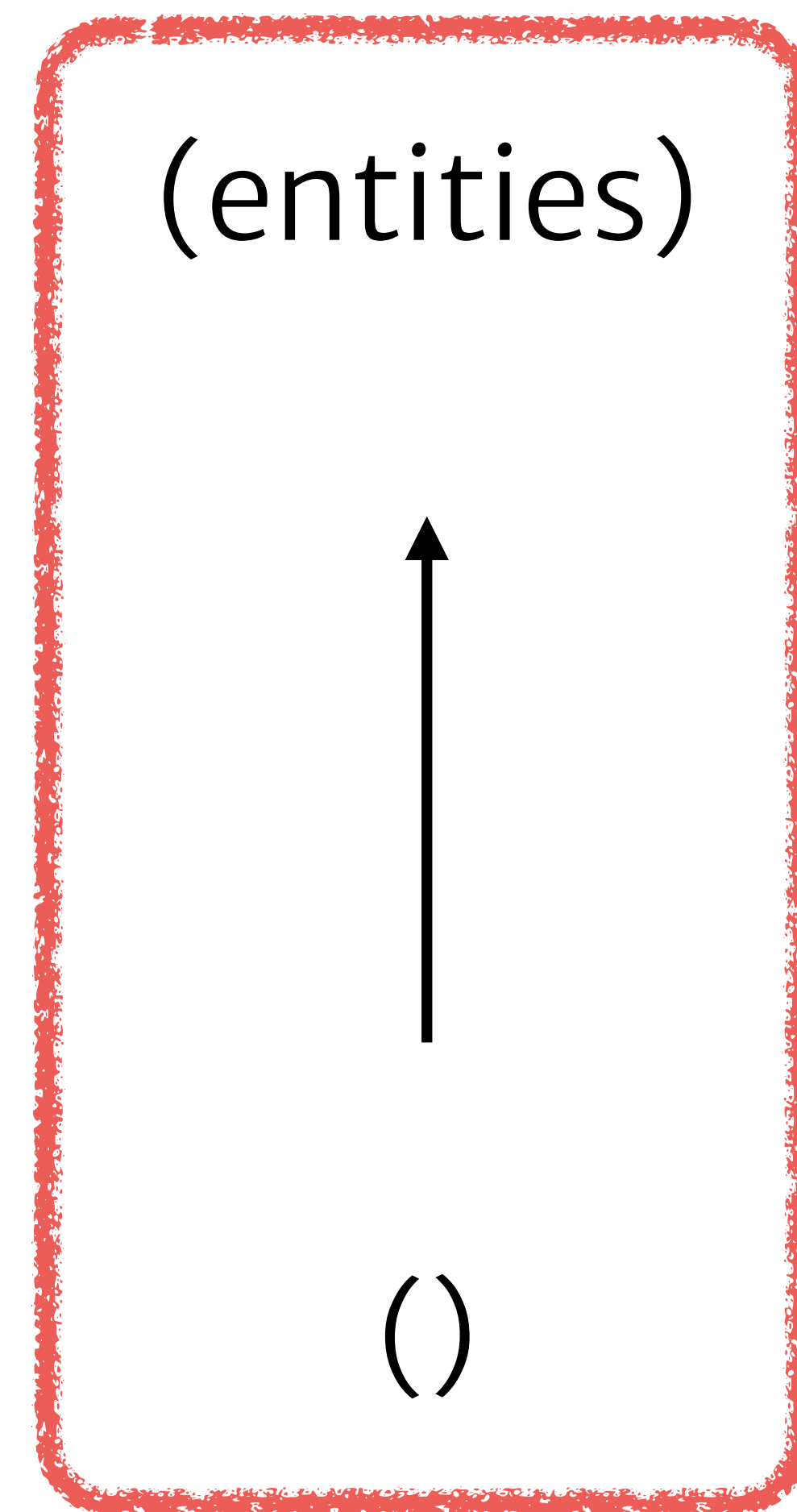
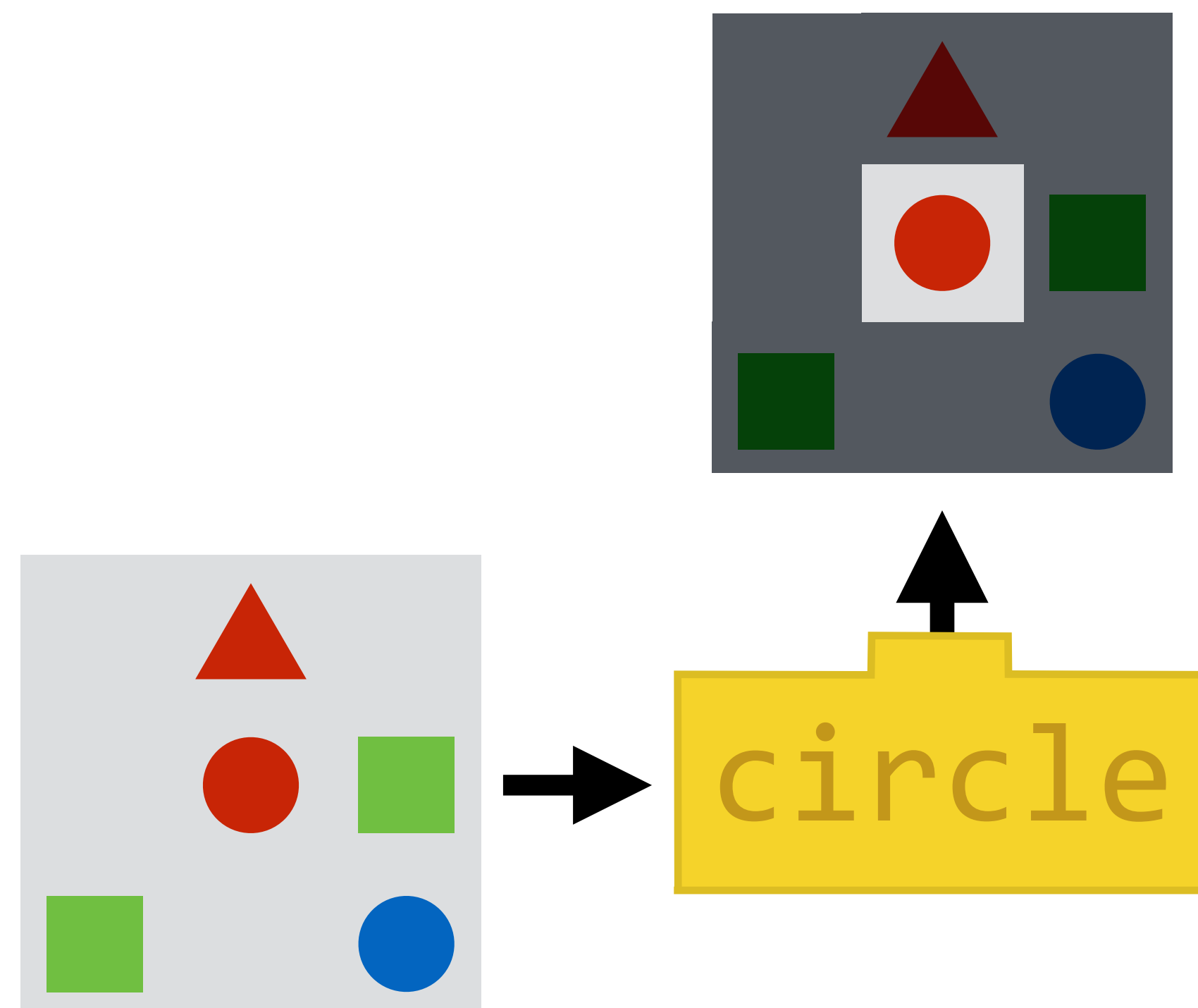
# Anatomy of a module: **Types**





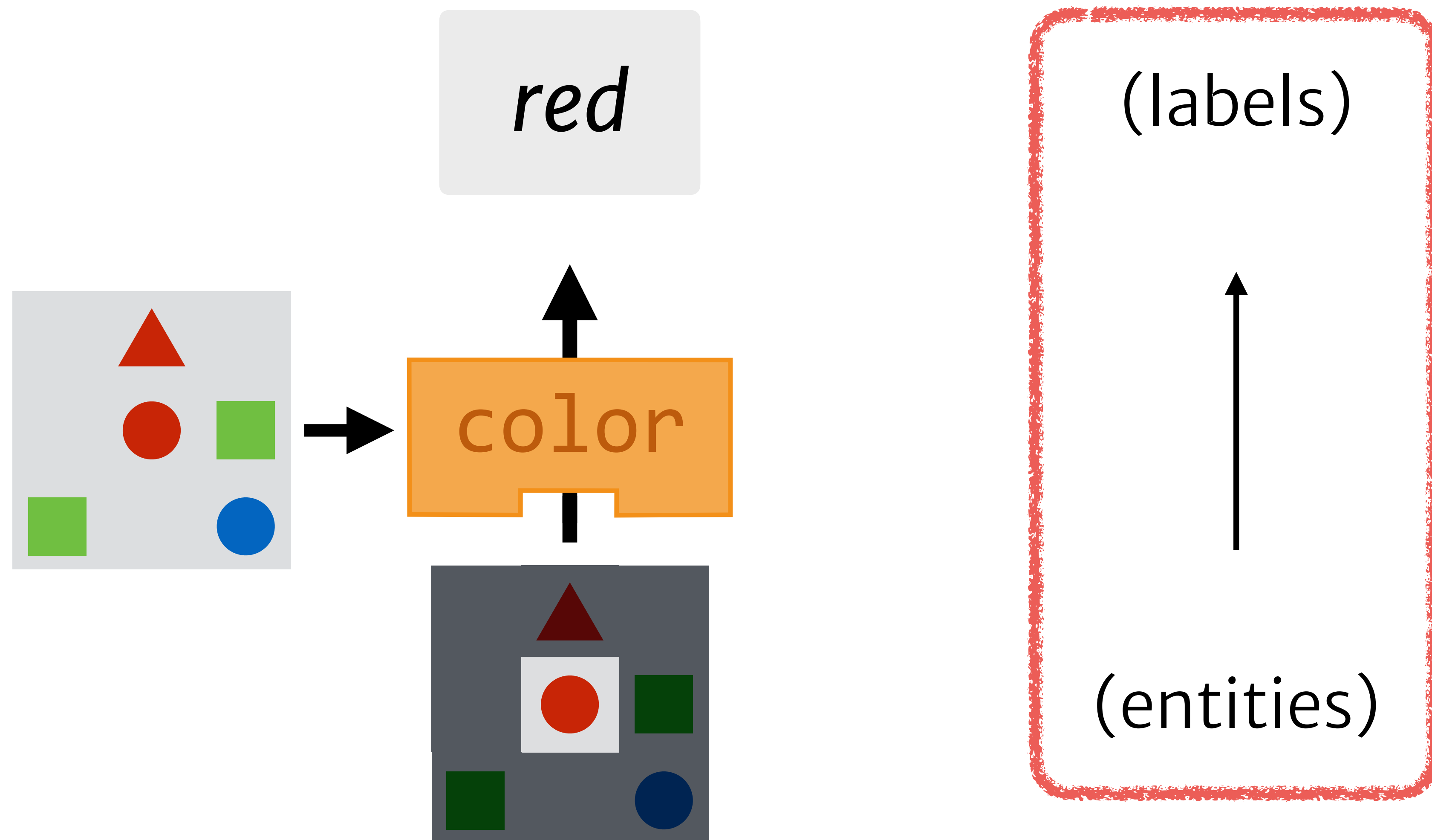


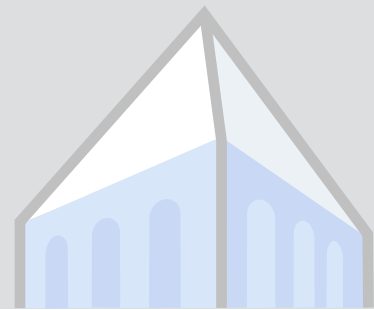
# Anatomy of a module: **Types**



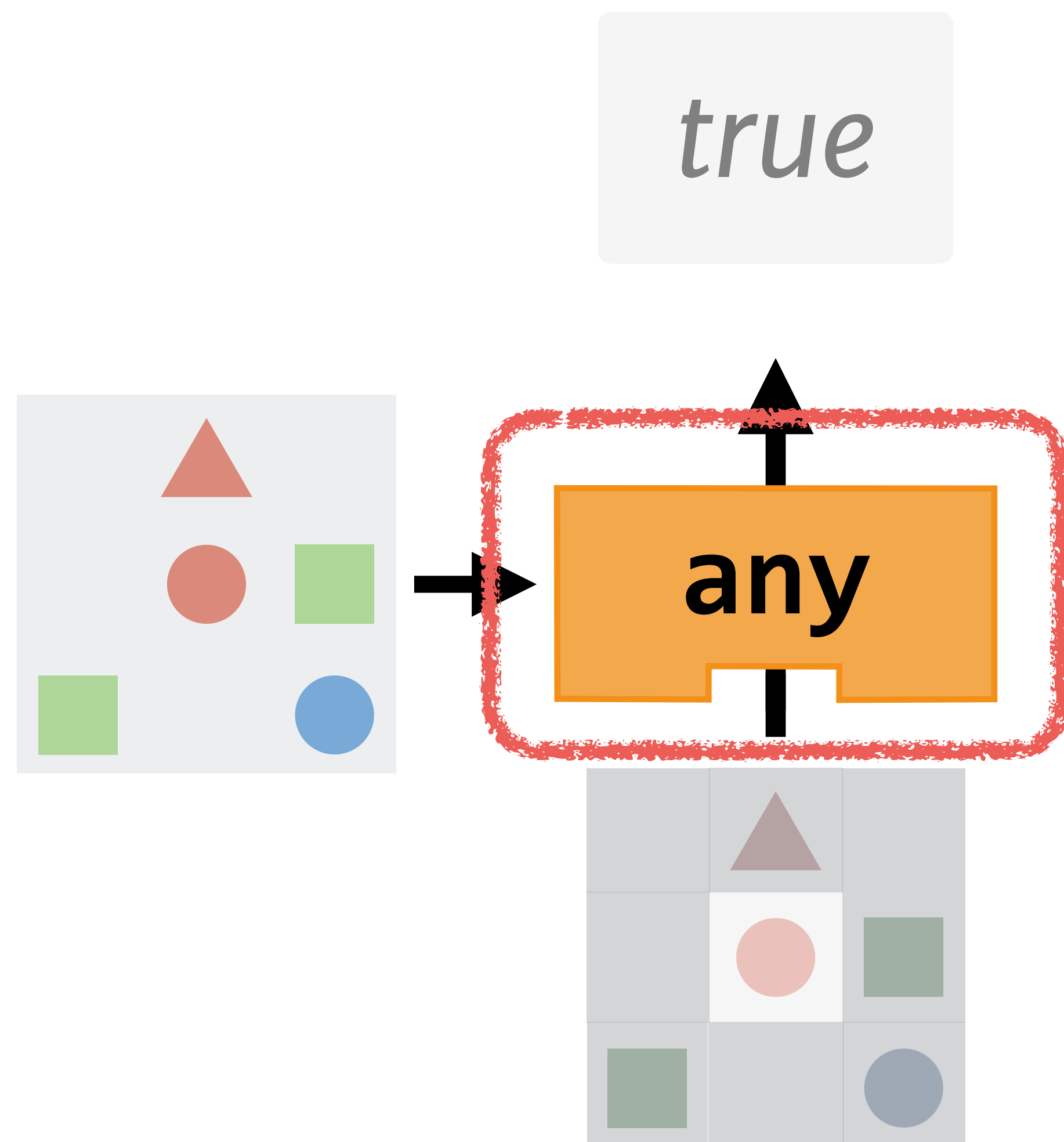


# Anatomy of a module: **Types**



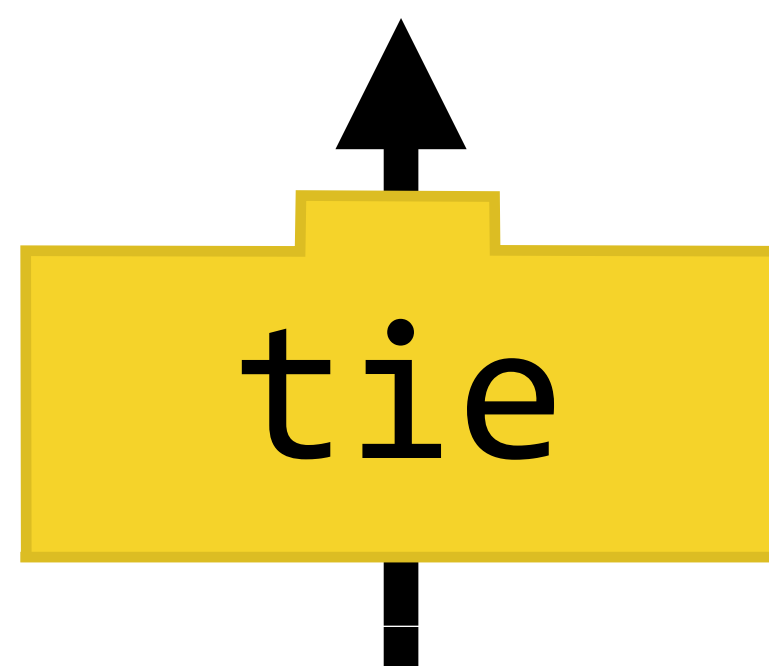
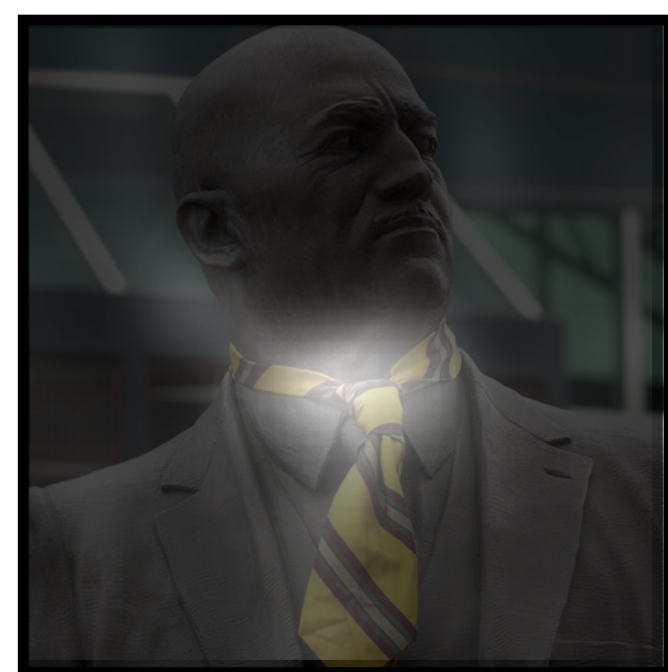


# Anatomy of a module: **Parameters**





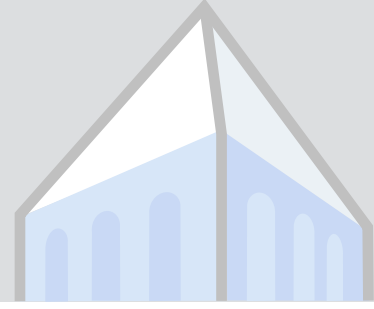
# Anatomy of a module



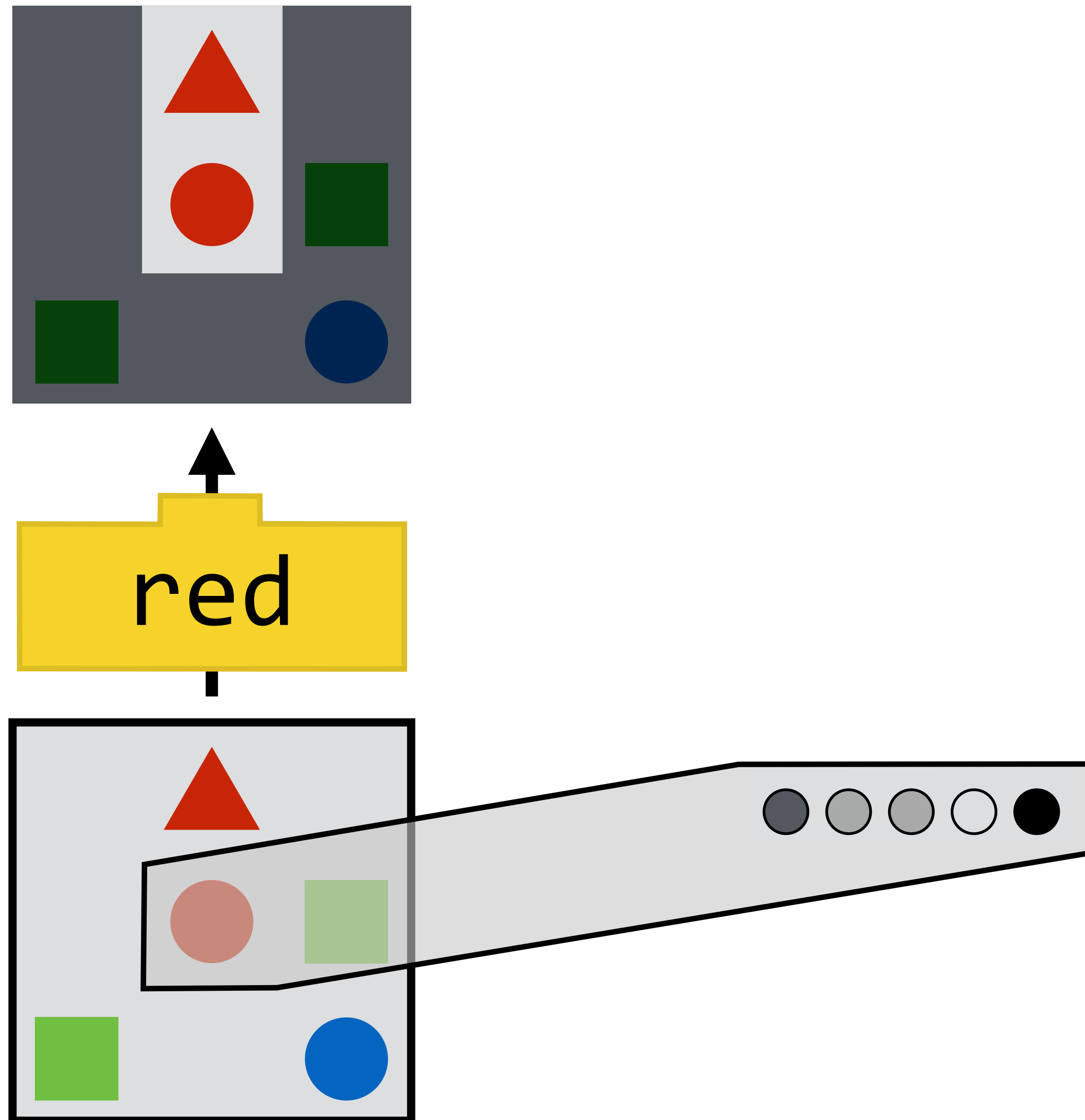
```
0.9 Columbia
0.1 Cooper
0.8 Myrtle Beach
```



name	type	coastal
<i>Columbia</i>	city	no
<i>Cooper</i>	river	yes
<i>Myrtle Beach</i>	city	yes

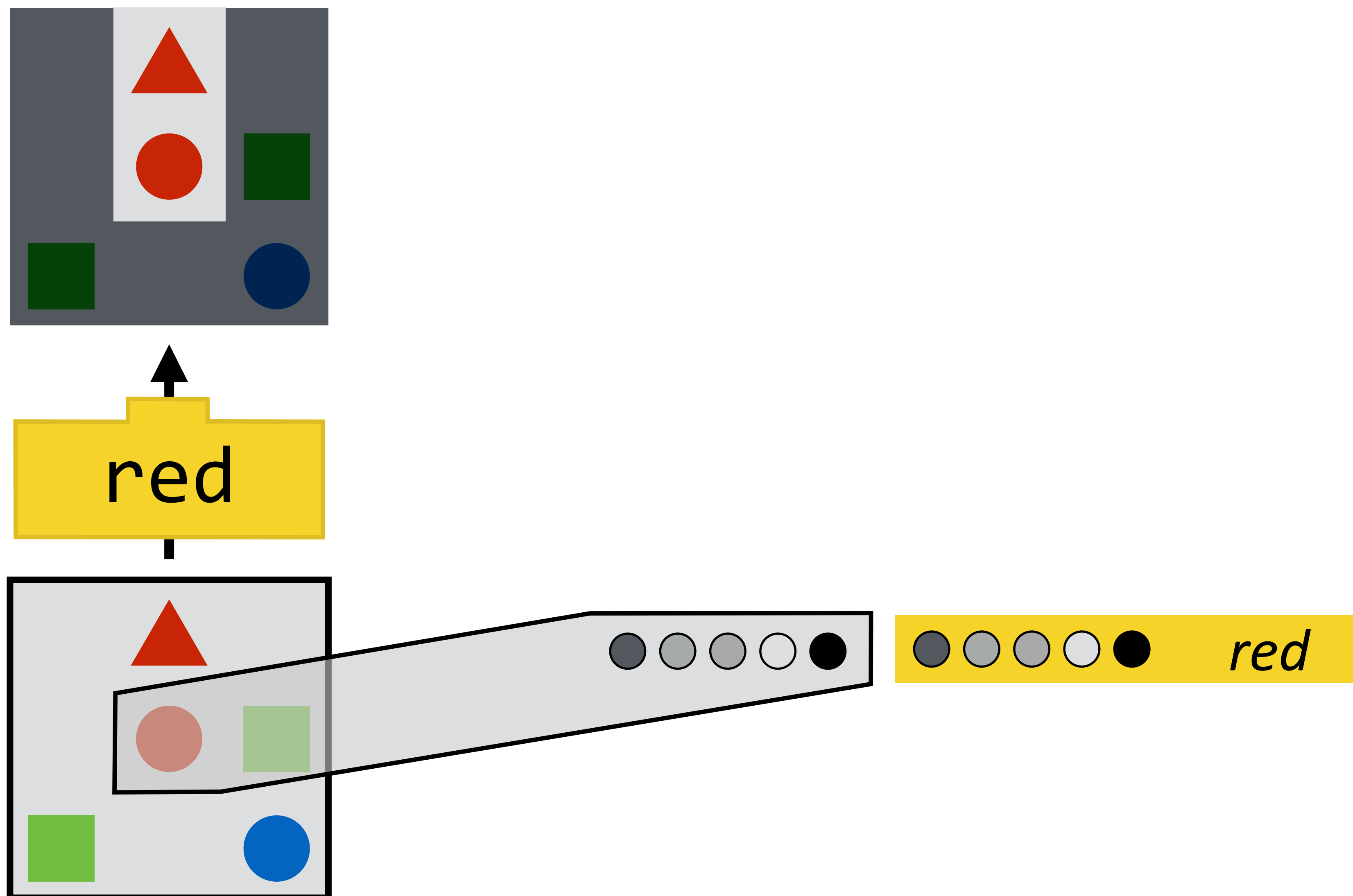


# Simple predicates



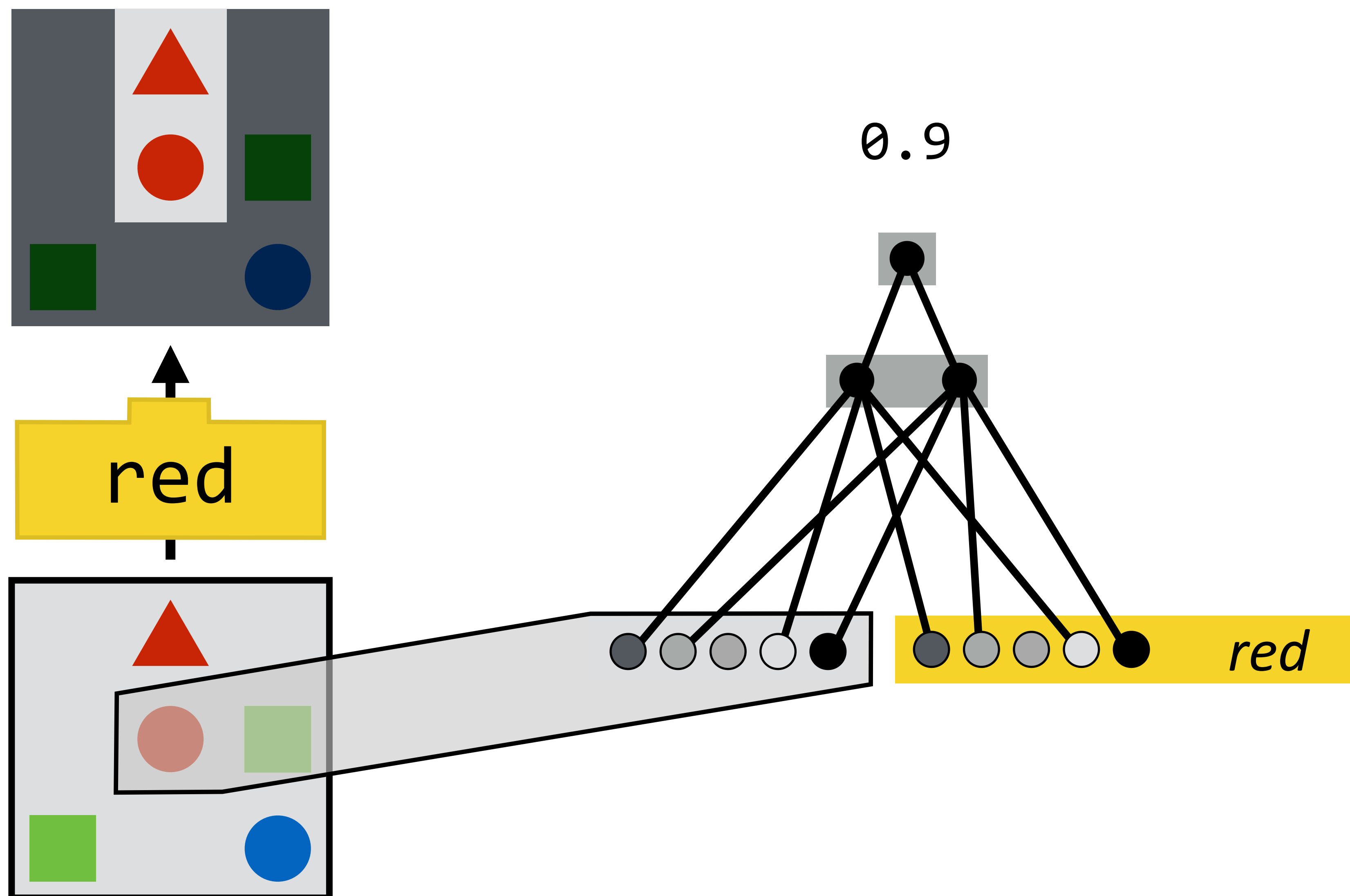


# Simple predicates



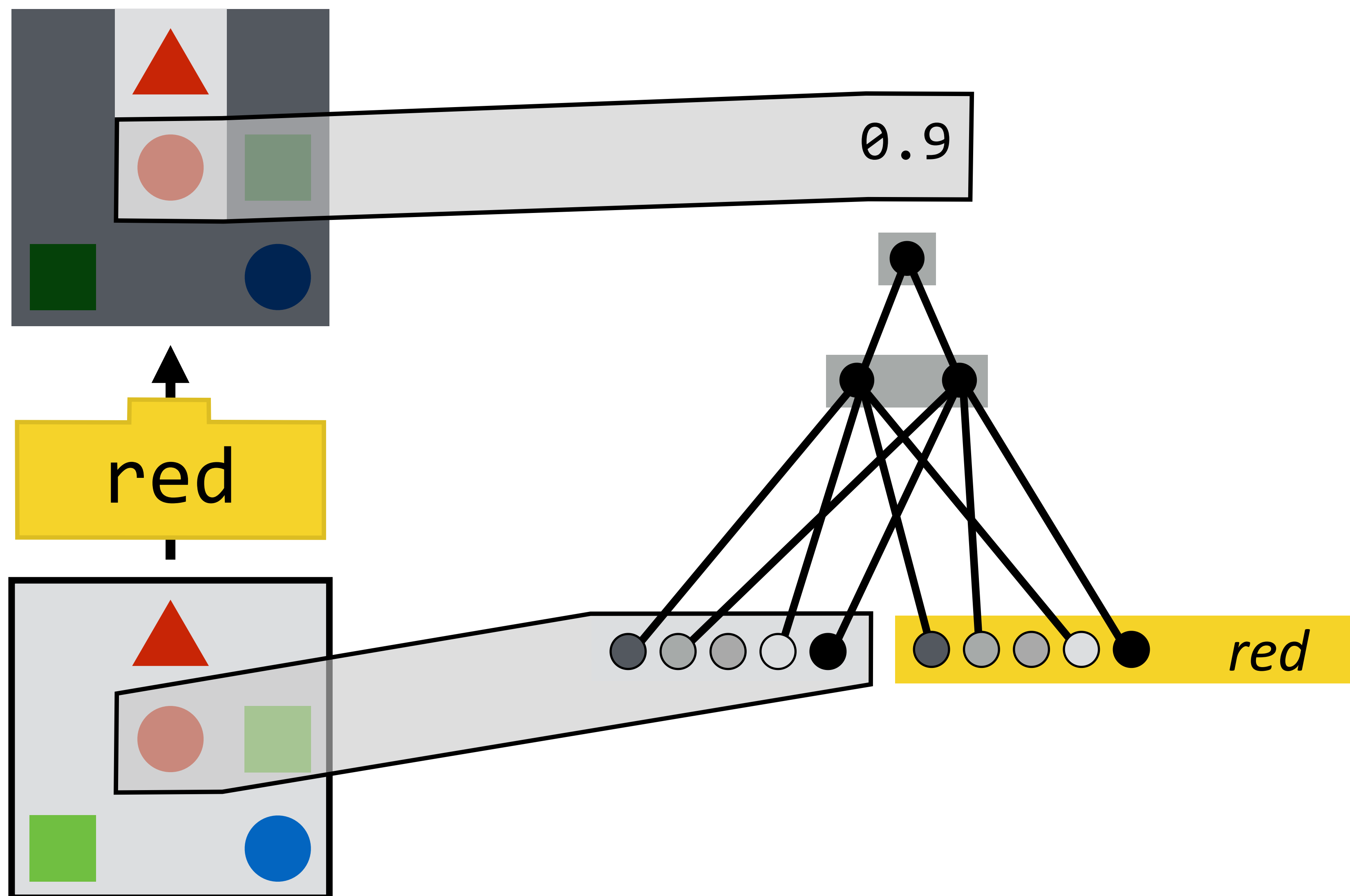


# Simple predicates





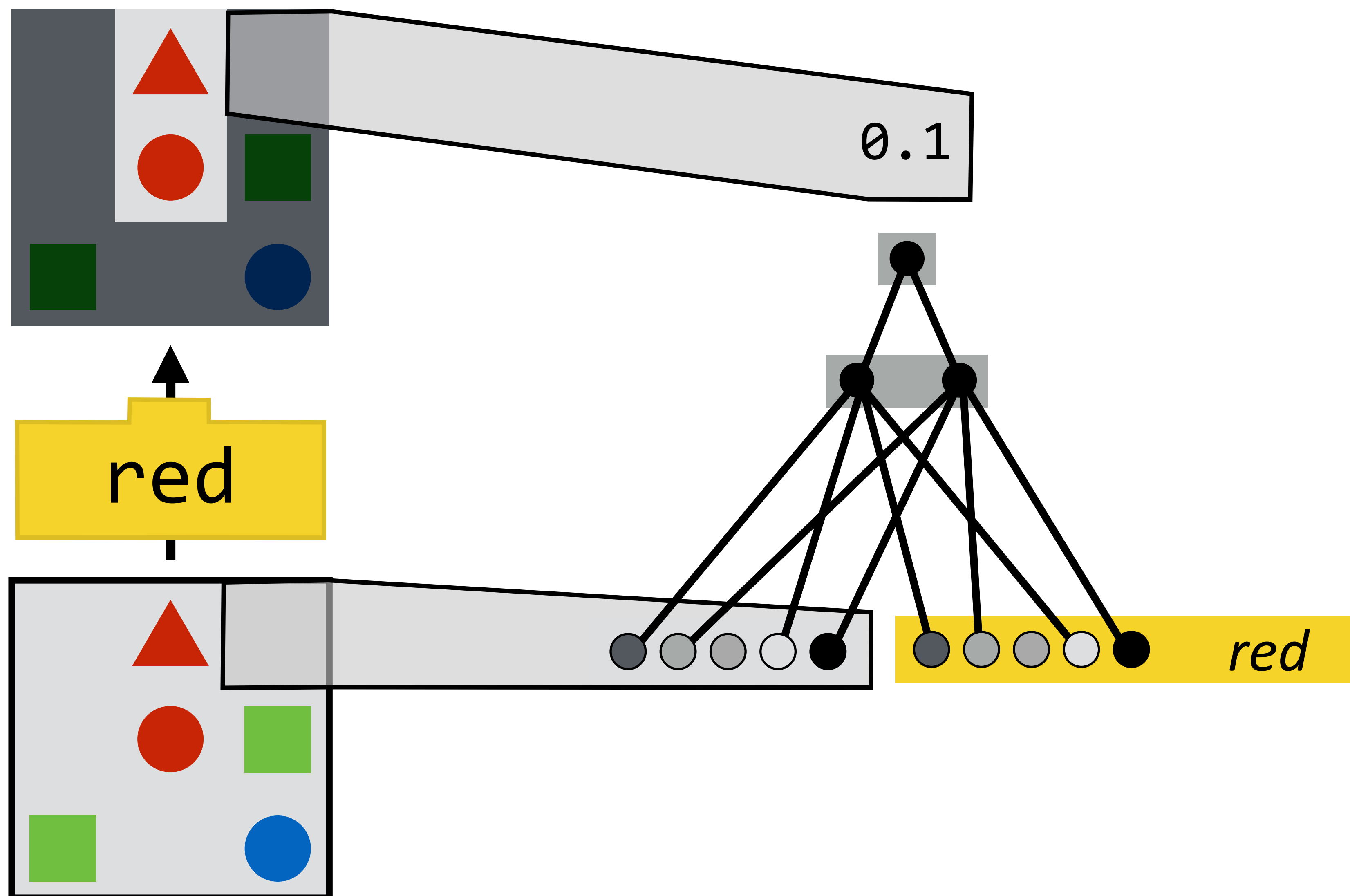
# Simple predicates







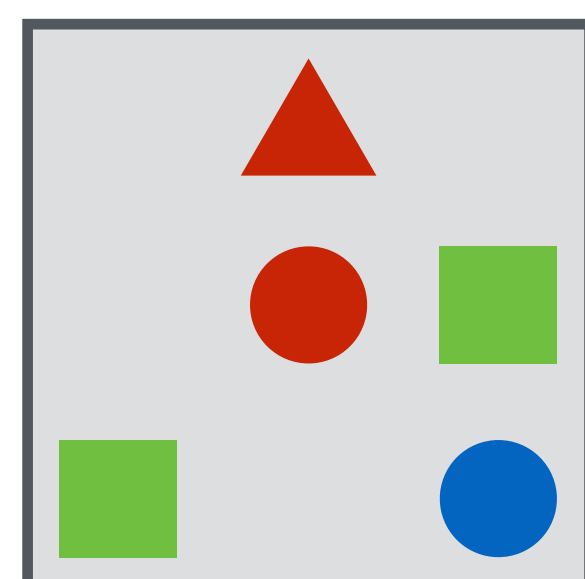
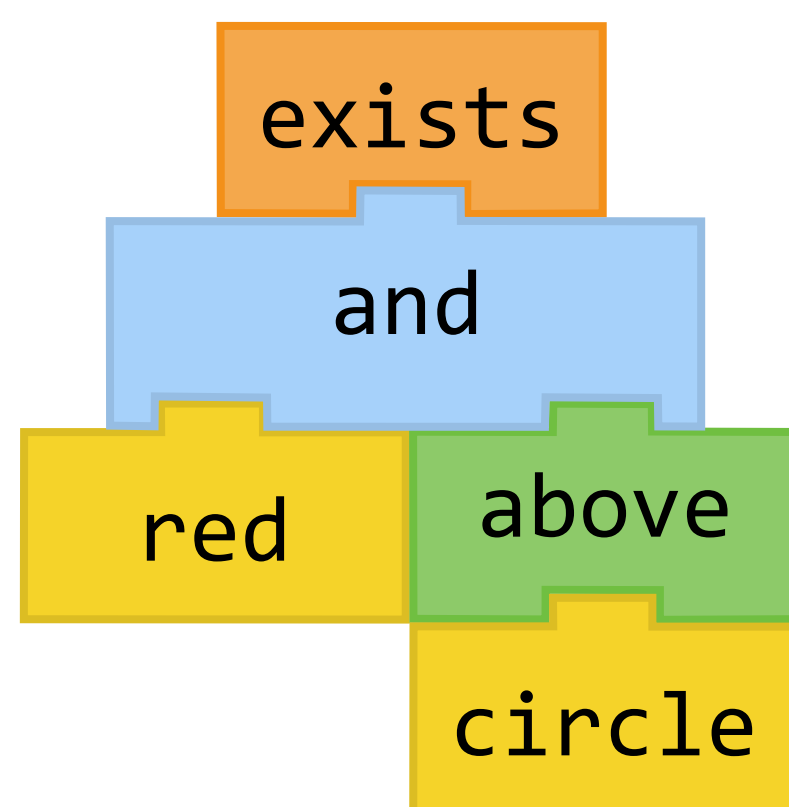
# Simple predicates





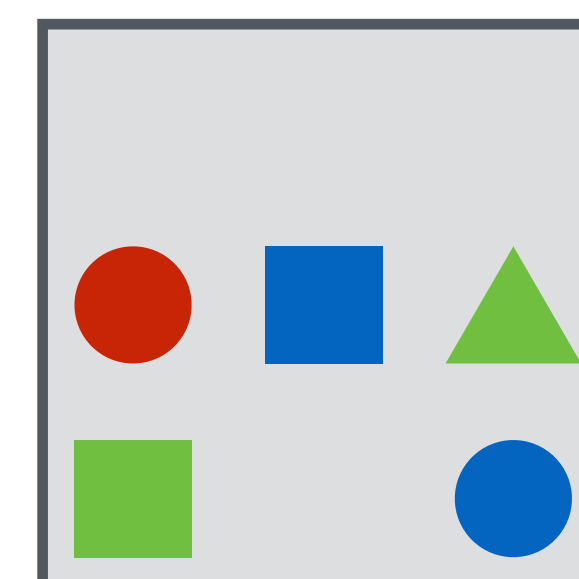
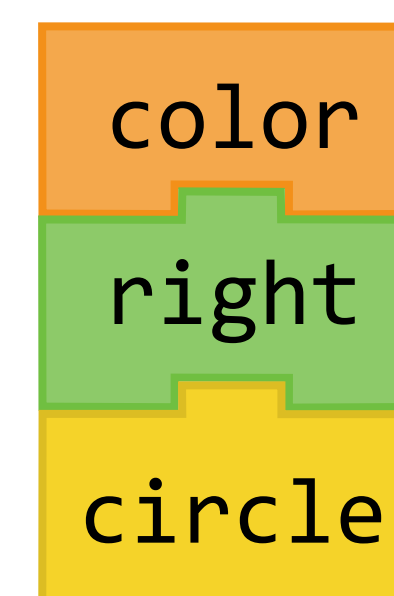
# Learning

*yes*



*Is there a red shape above a circle?*

*blue*

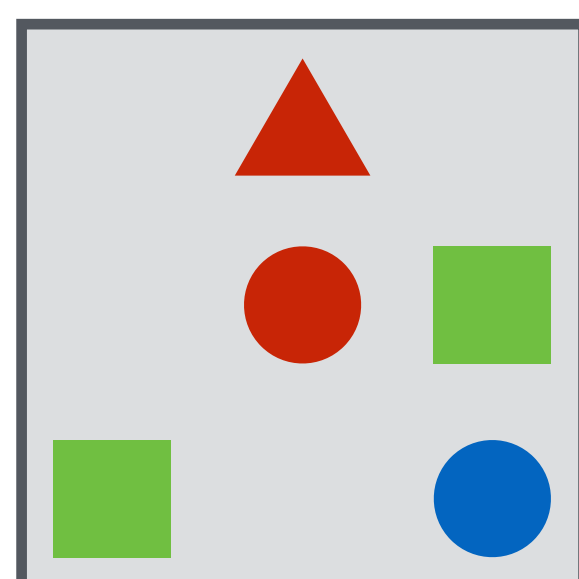
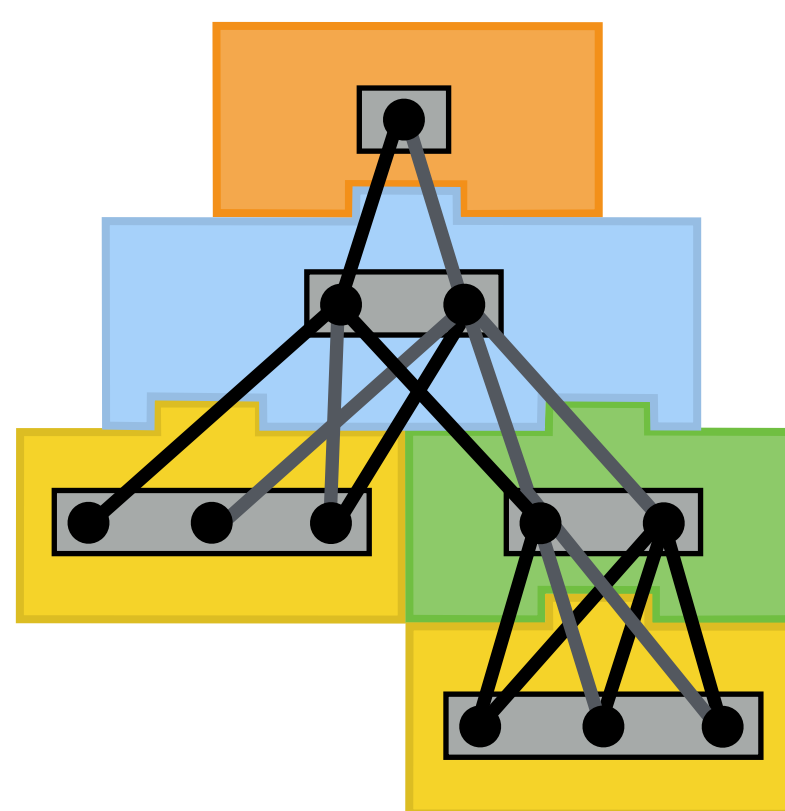


*What color is the shape right of a circle?*



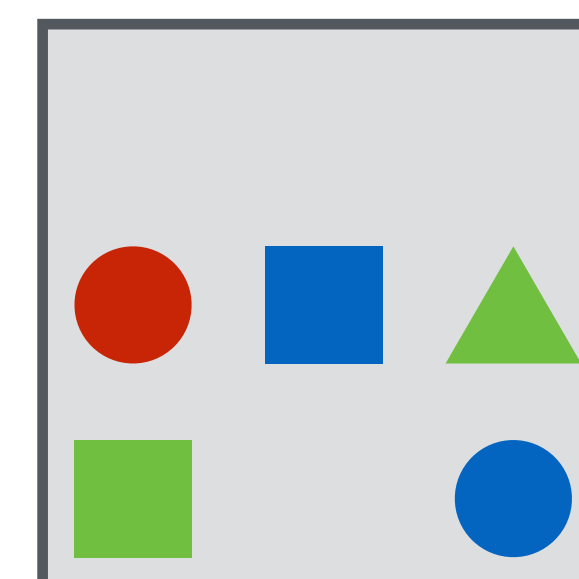
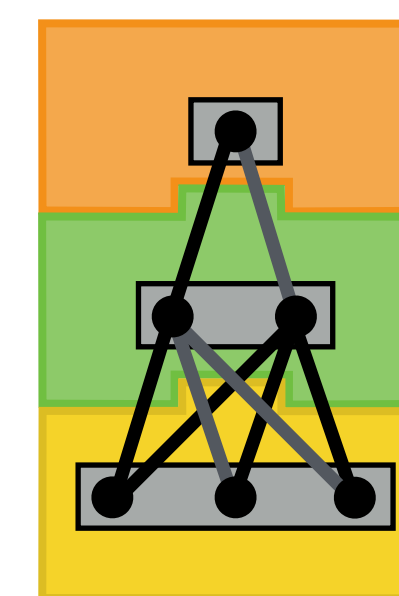
# Learning

*yes*



*Is there a red shape above a circle?*

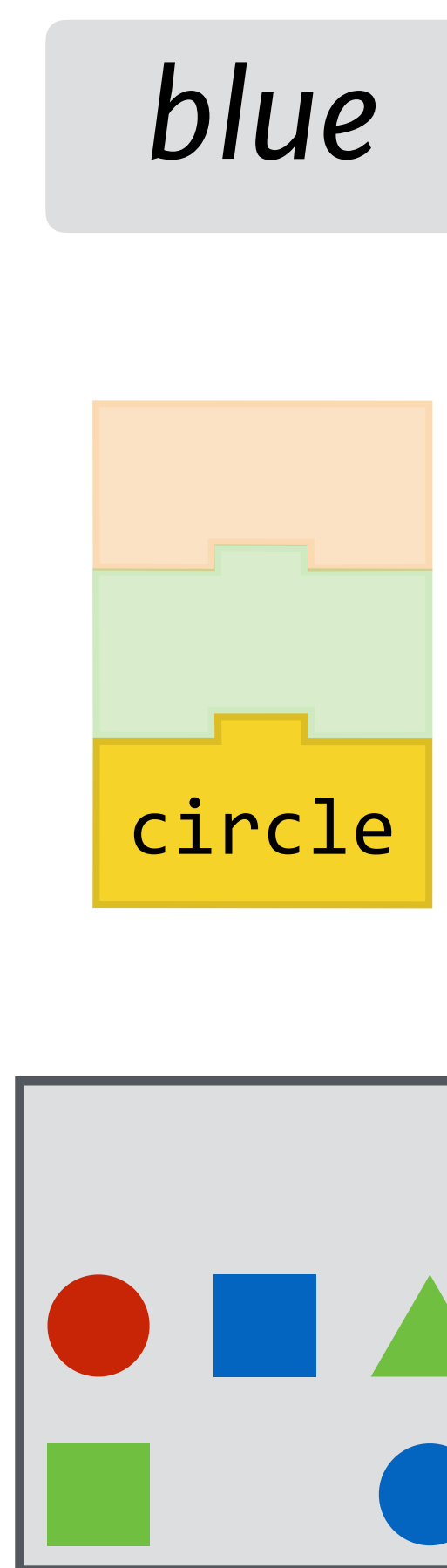
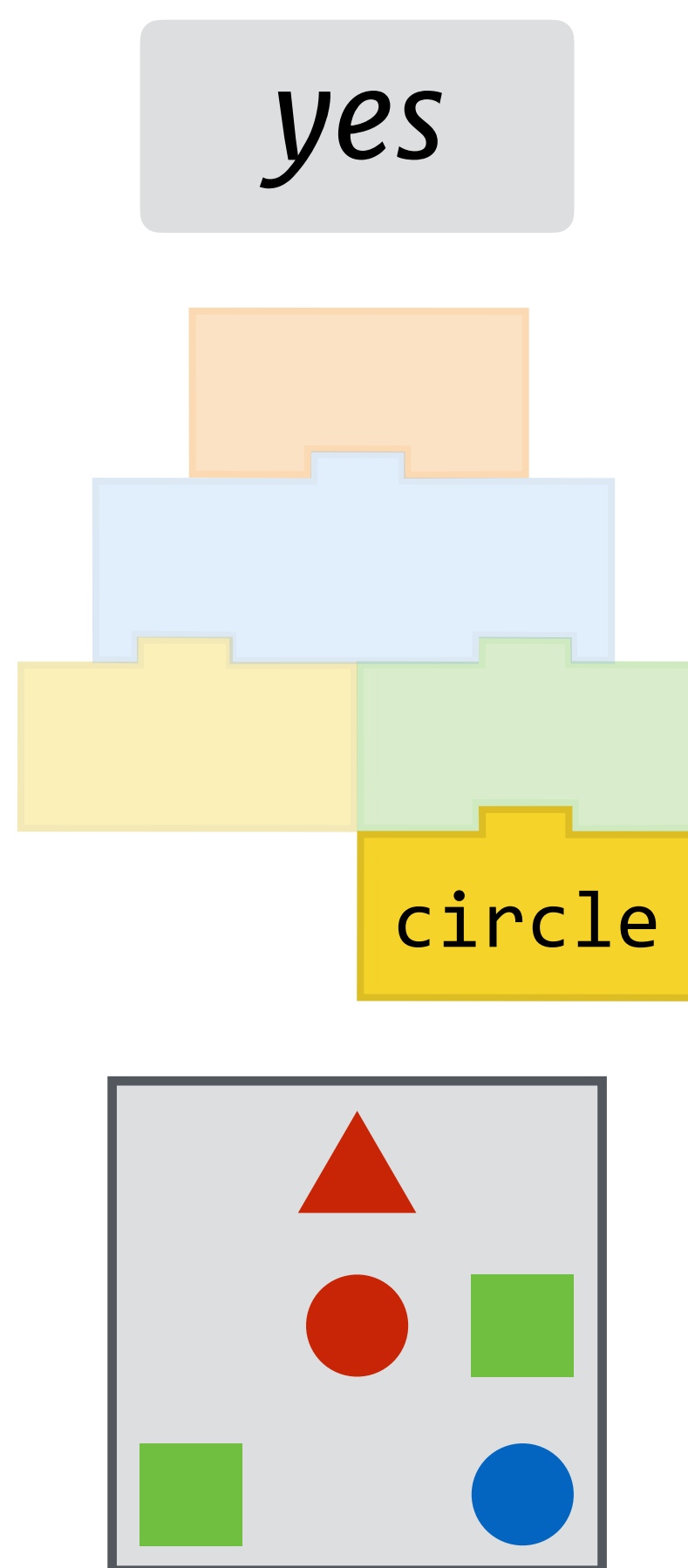
*blue*



*What color is the shape right of a circle?*



# Parameter tying

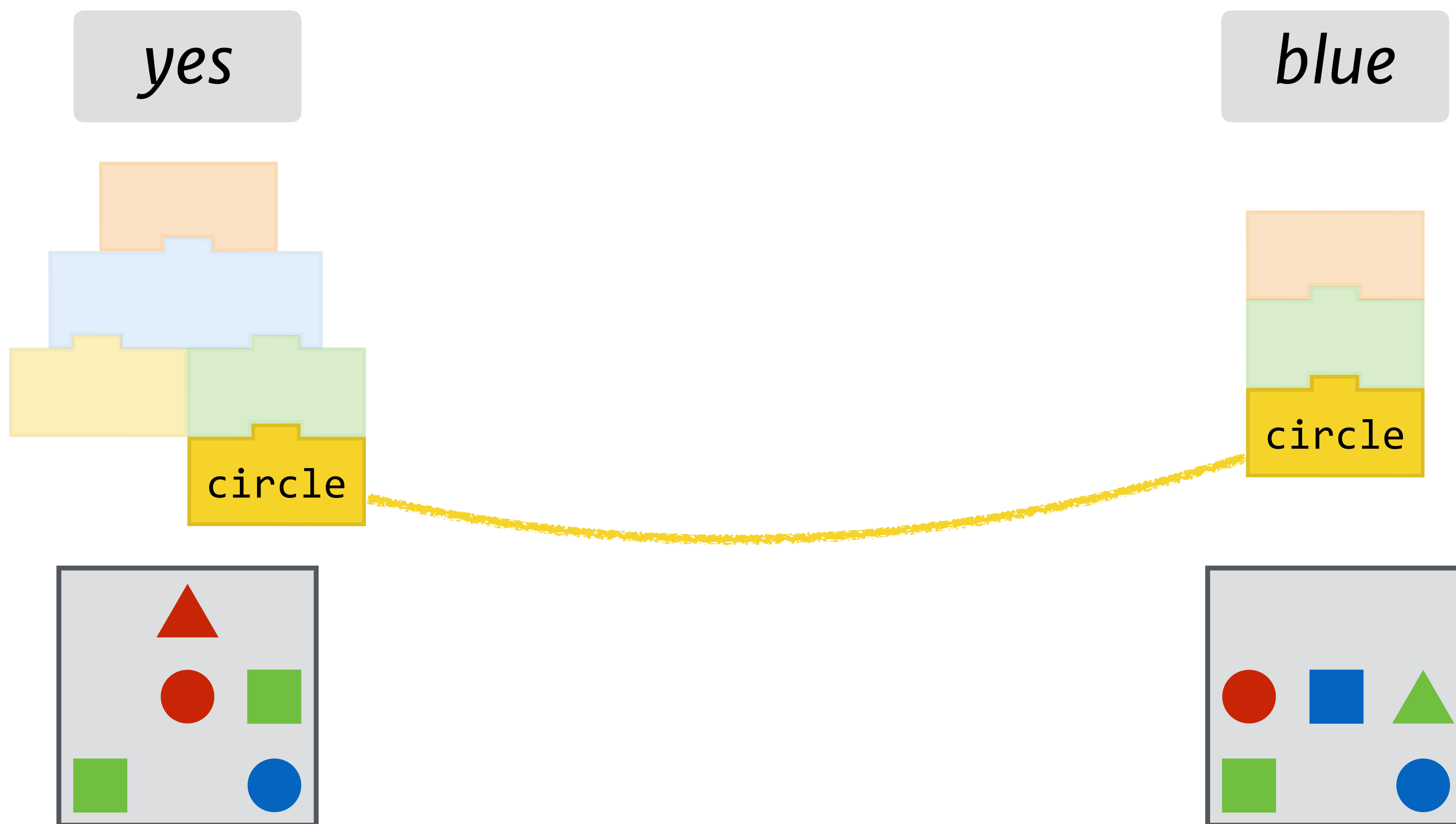


*Is there a red shape above a circle?*

*What color is the shape right of a circle?*



# Parameter tying

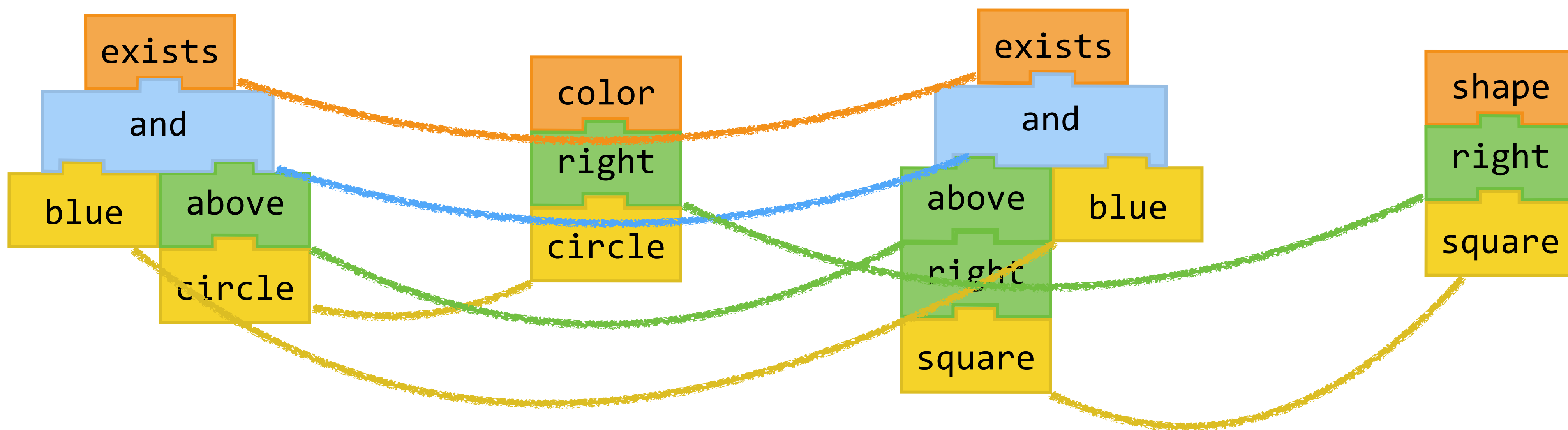


*Is there a red shape above a circle?*

*What color is the shape right of a circle?*

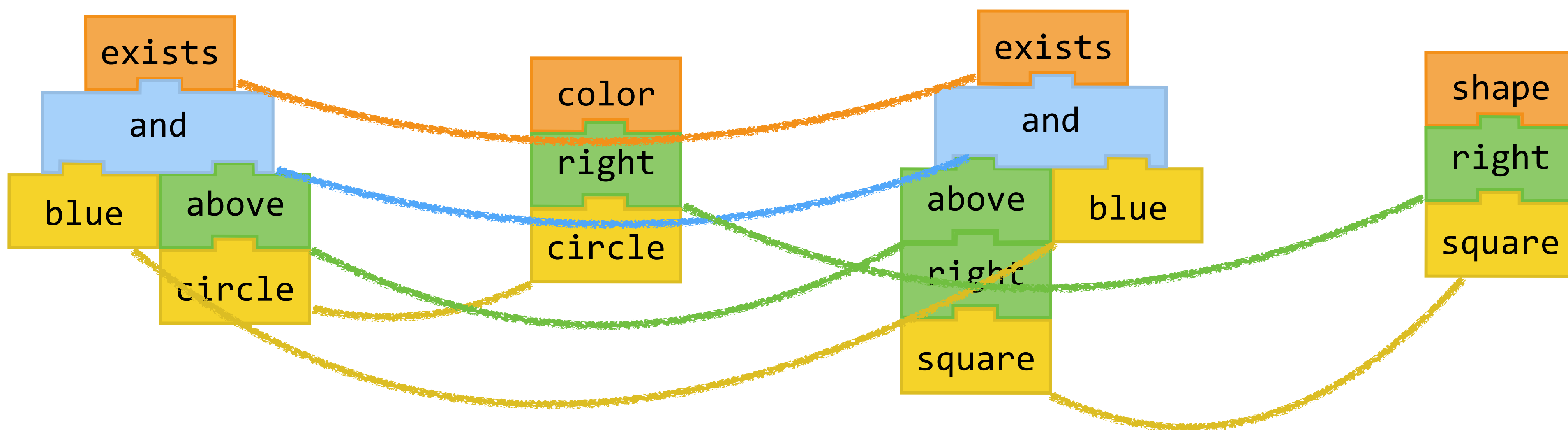


# EXTREME parameter tying

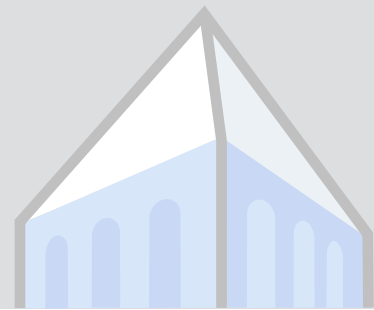




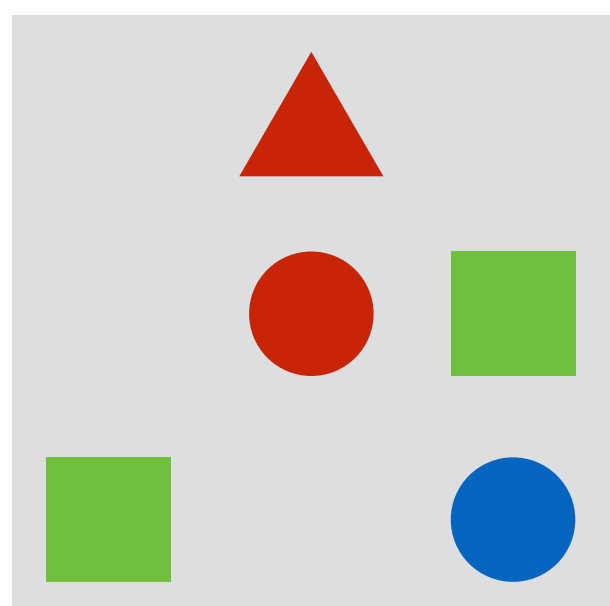
# EXTREME parameter tying



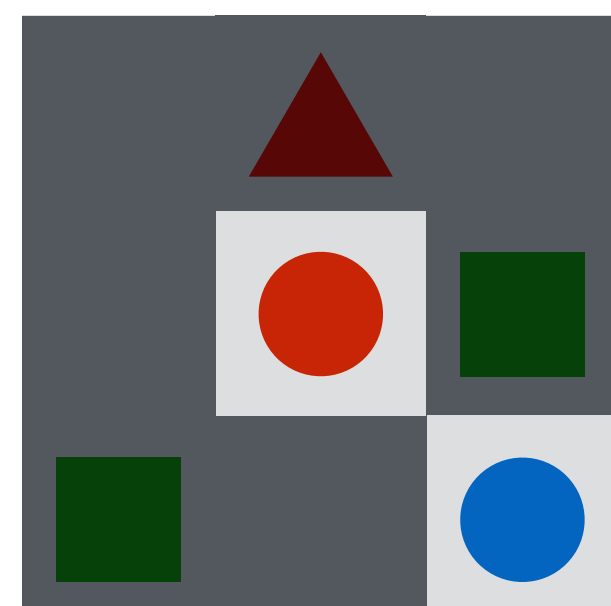
$$\arg \max_W \sum \log p(\text{yes} \mid \begin{array}{|c|} \hline \triangle \\ \hline \bullet \\ \hline \square \\ \hline \square \\ \hline \bullet \\ \hline \end{array}, \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \square \\ \hline \square \\ \hline \square \\ \hline \square \\ \hline \end{array}; W)$$



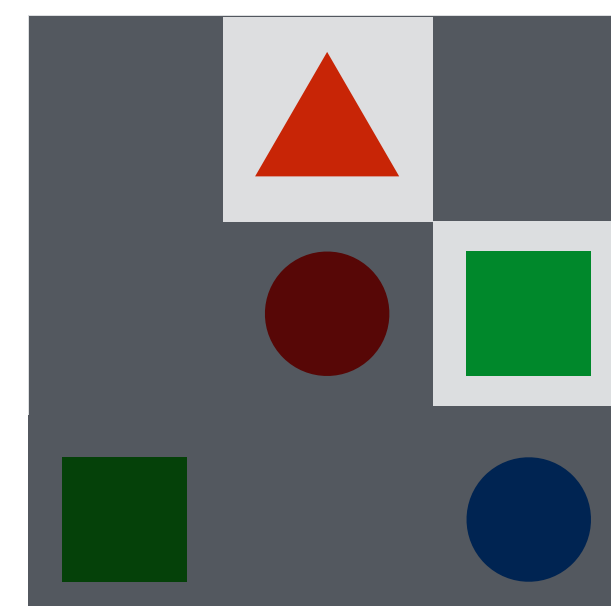
# Learning with fixed layouts is easy!



circle



above

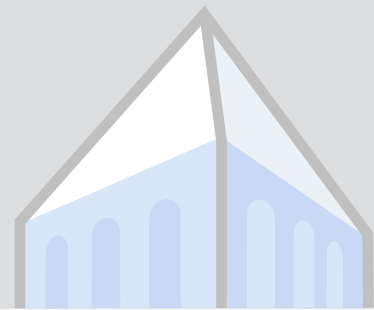


any

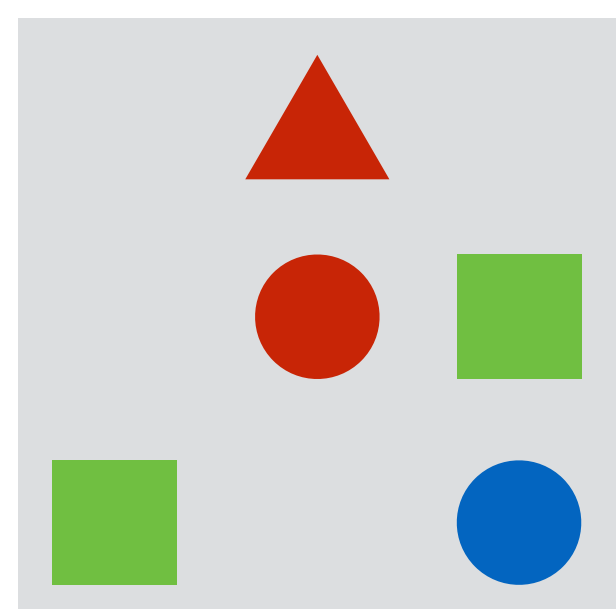
*true*

Module specialization is driven entirely by context!

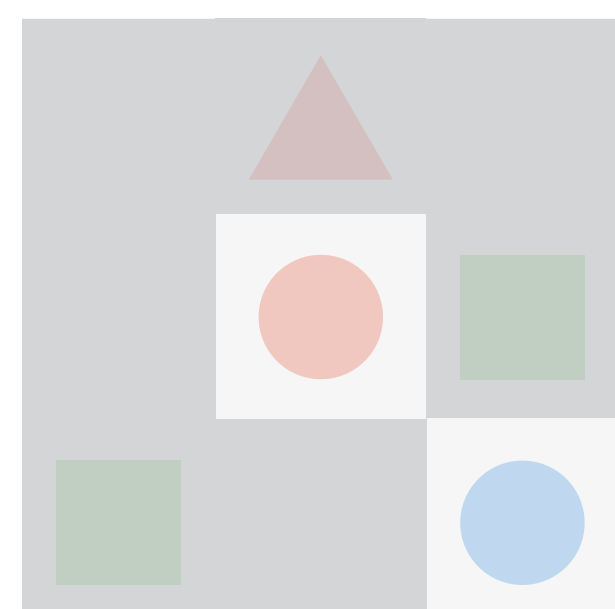




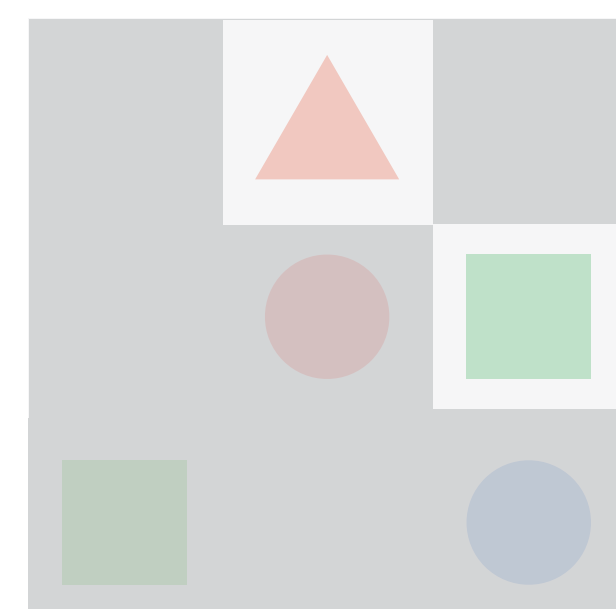
# Learning with fixed layouts is easy!



circle



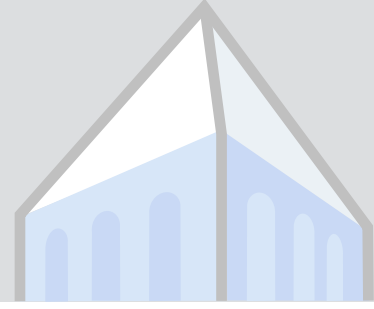
above



any

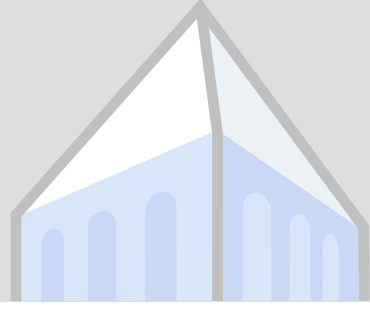
*true*

Module specialization is driven entirely by context!

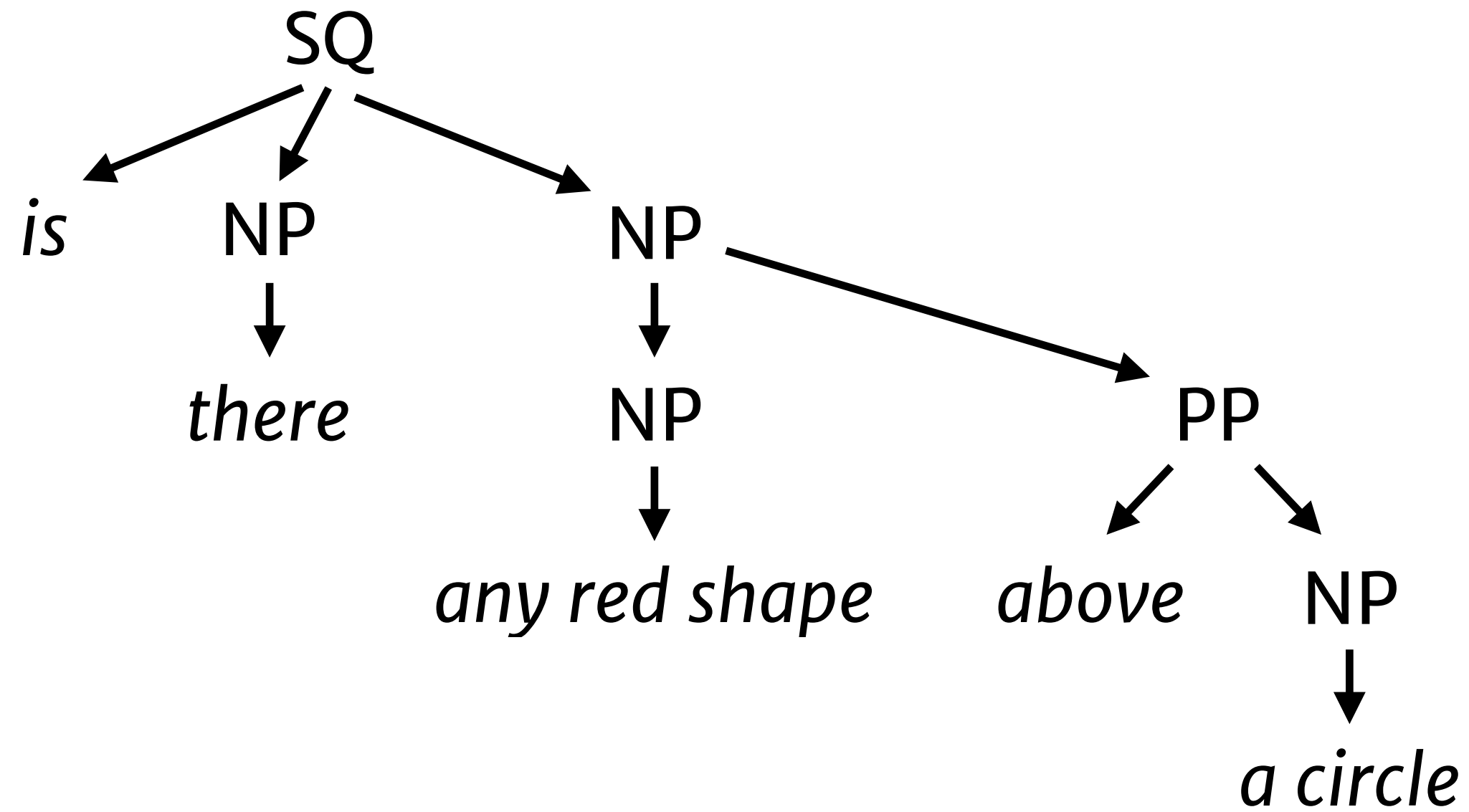


# Where do network structures come from?

*Is there any red shape above a circle?*



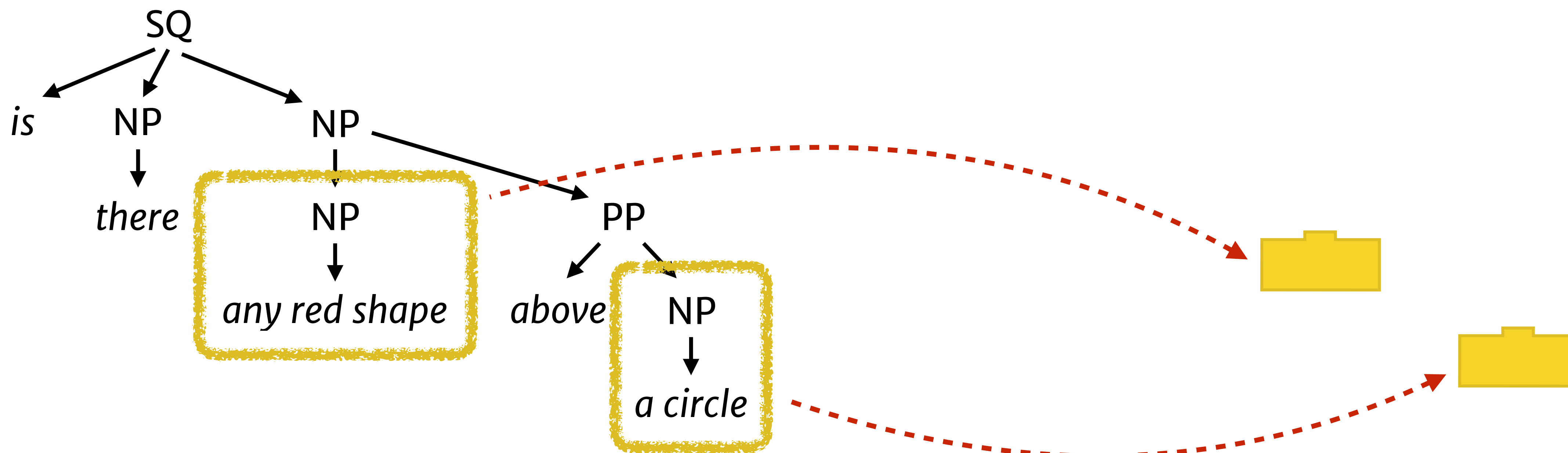
# Where do network structures come from?



*Is there any red shape above a circle?*



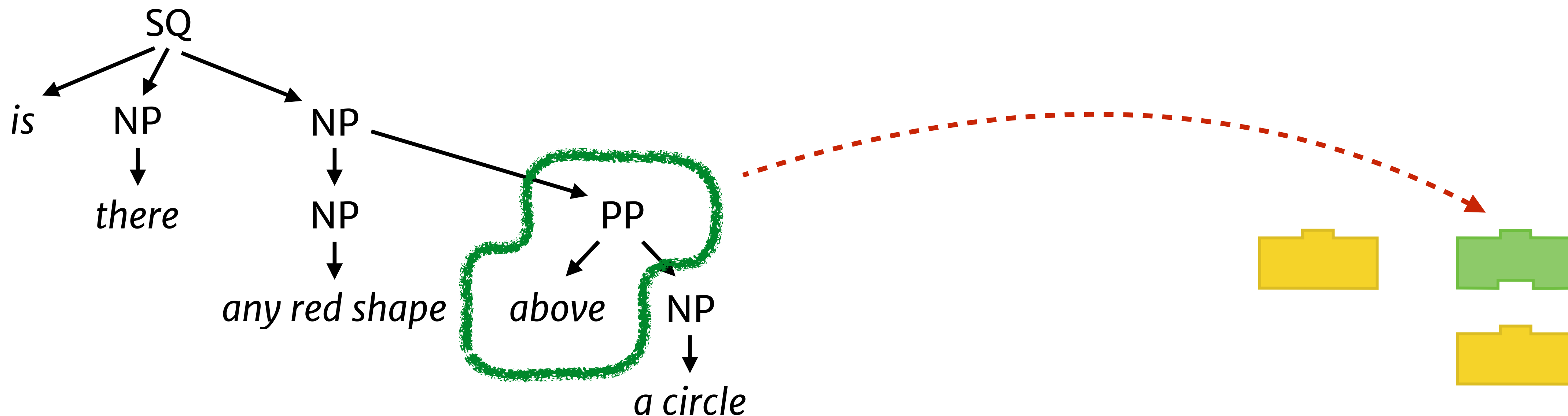
# Where do network structures come from?



*Is there any red shape above a circle?*



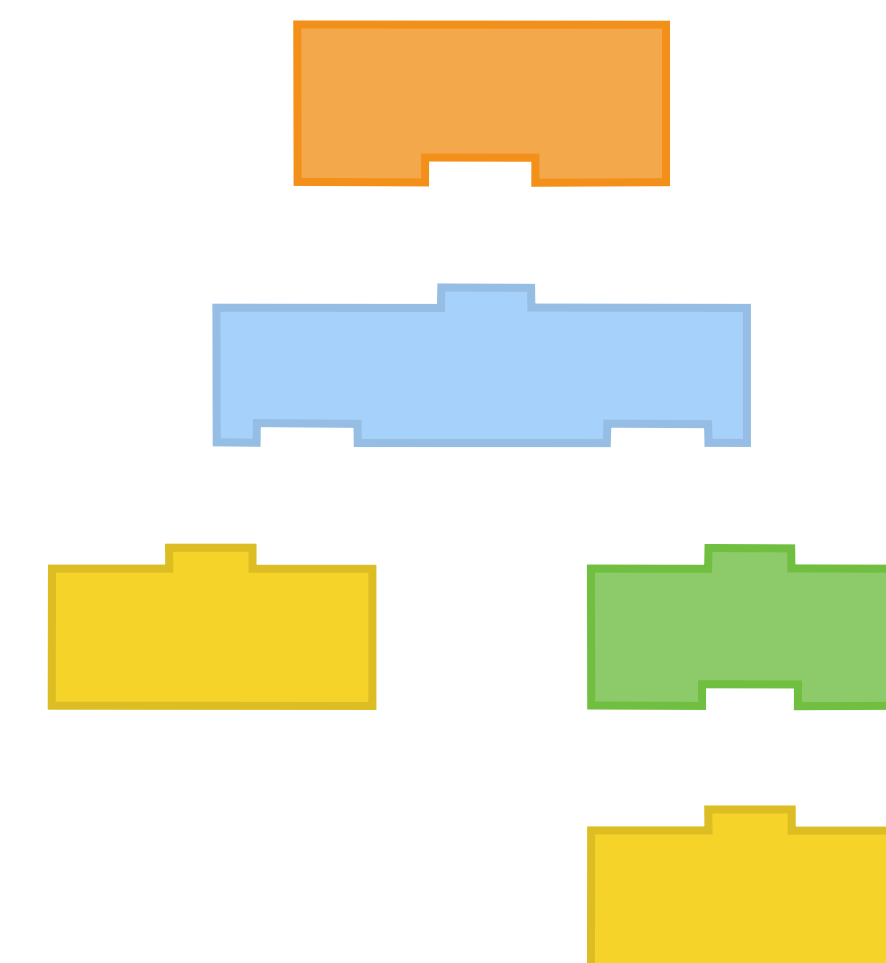
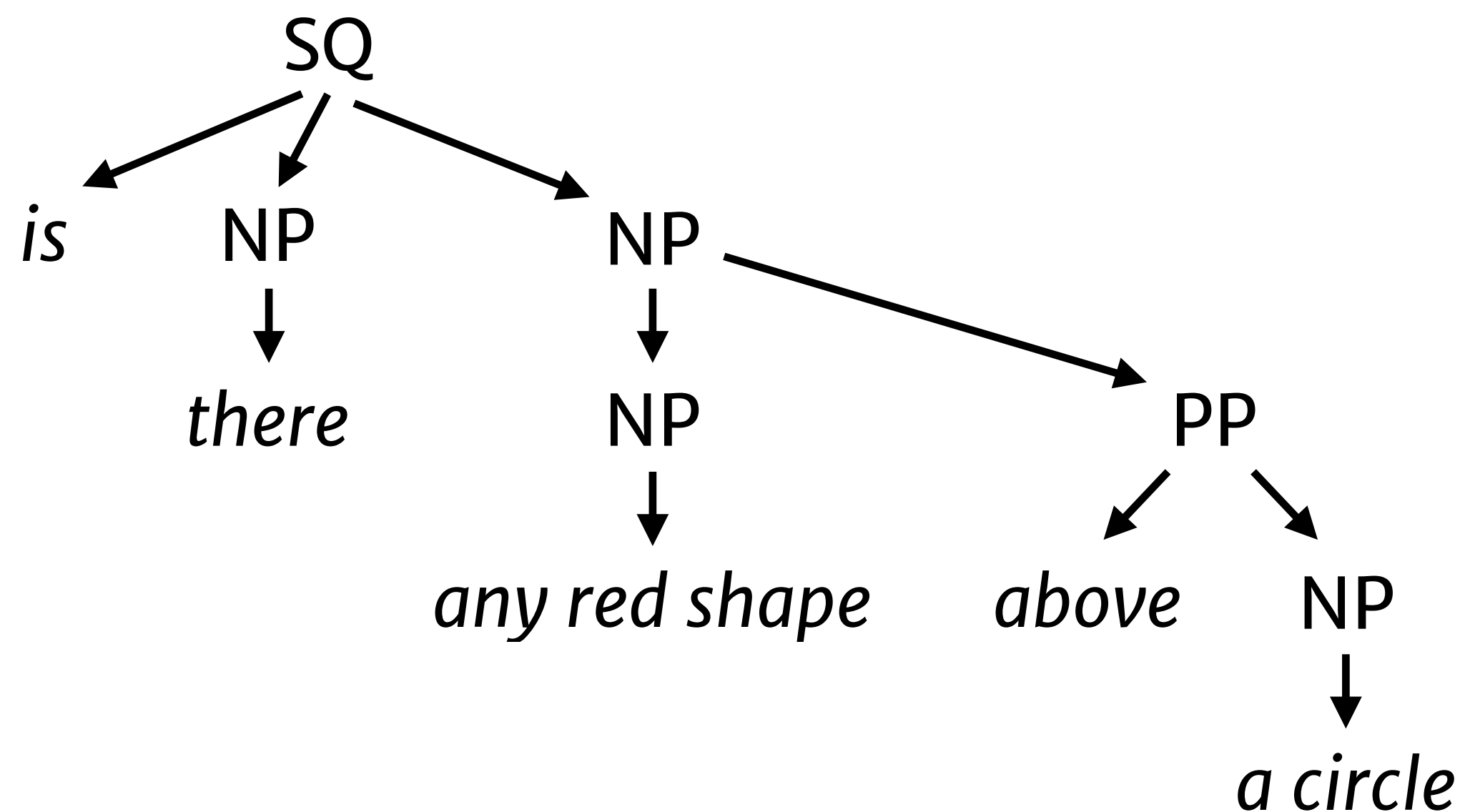
# Where do network structures come from?



*Is there any red shape above a circle?*



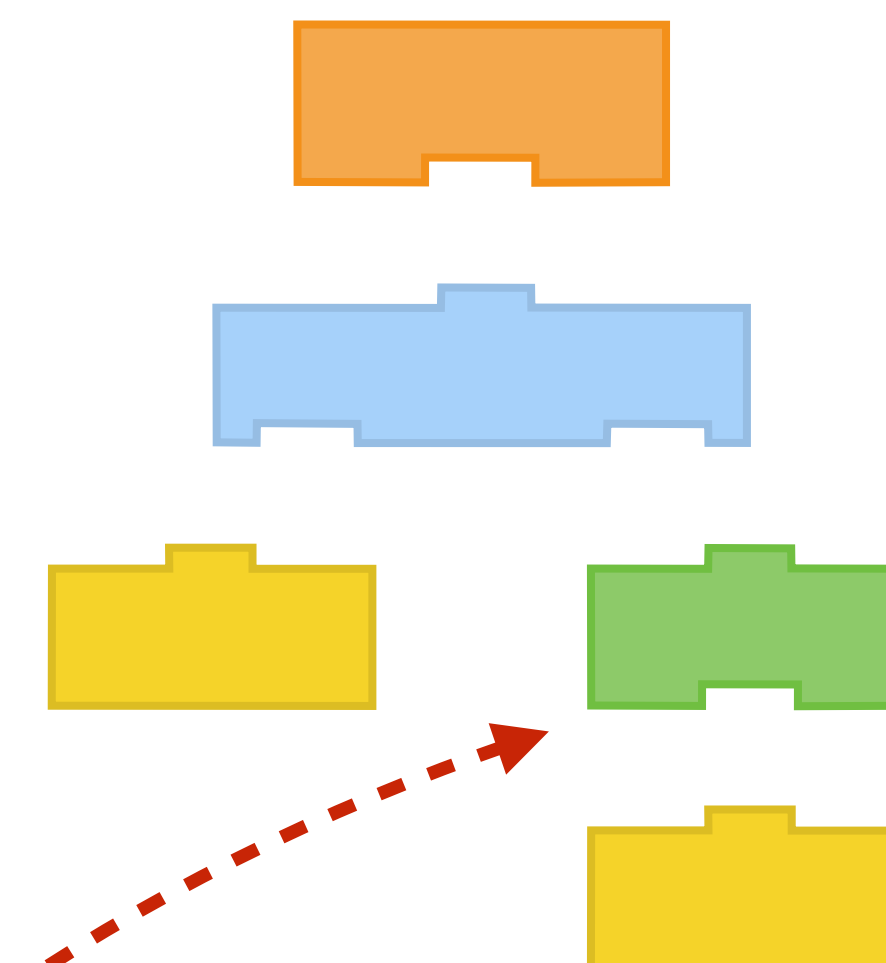
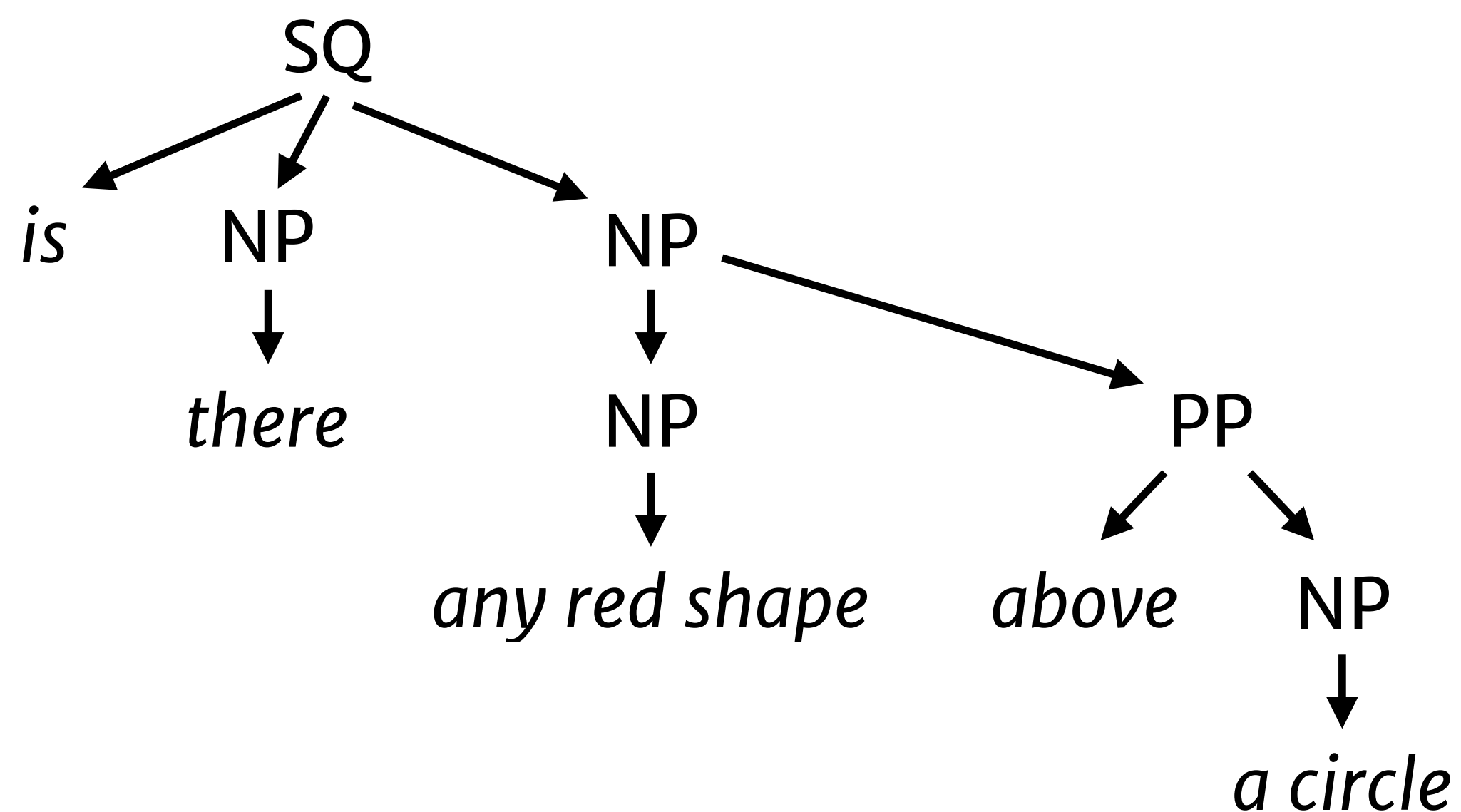
# Where do network structures come from?



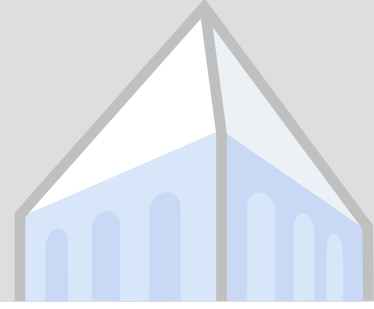
*Is there any red shape above a circle?*



# Where do network structures come from?



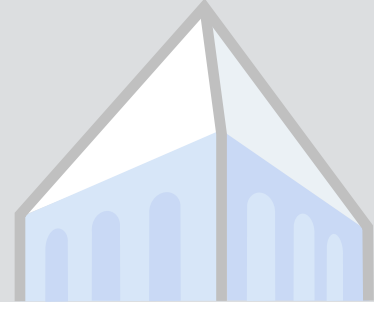
Is there any red *shape* *above* a circle?



# Where do network structures come from?

hard choice of layouts and answers





# Where do network structures come from?

hard choice of layouts and answers

$$\arg \max_{\theta} \mathbb{E} \log p(\text{ans} \mid \text{ques}, \text{img})$$



# Where do network structures come from?

hard choice of layouts and answers

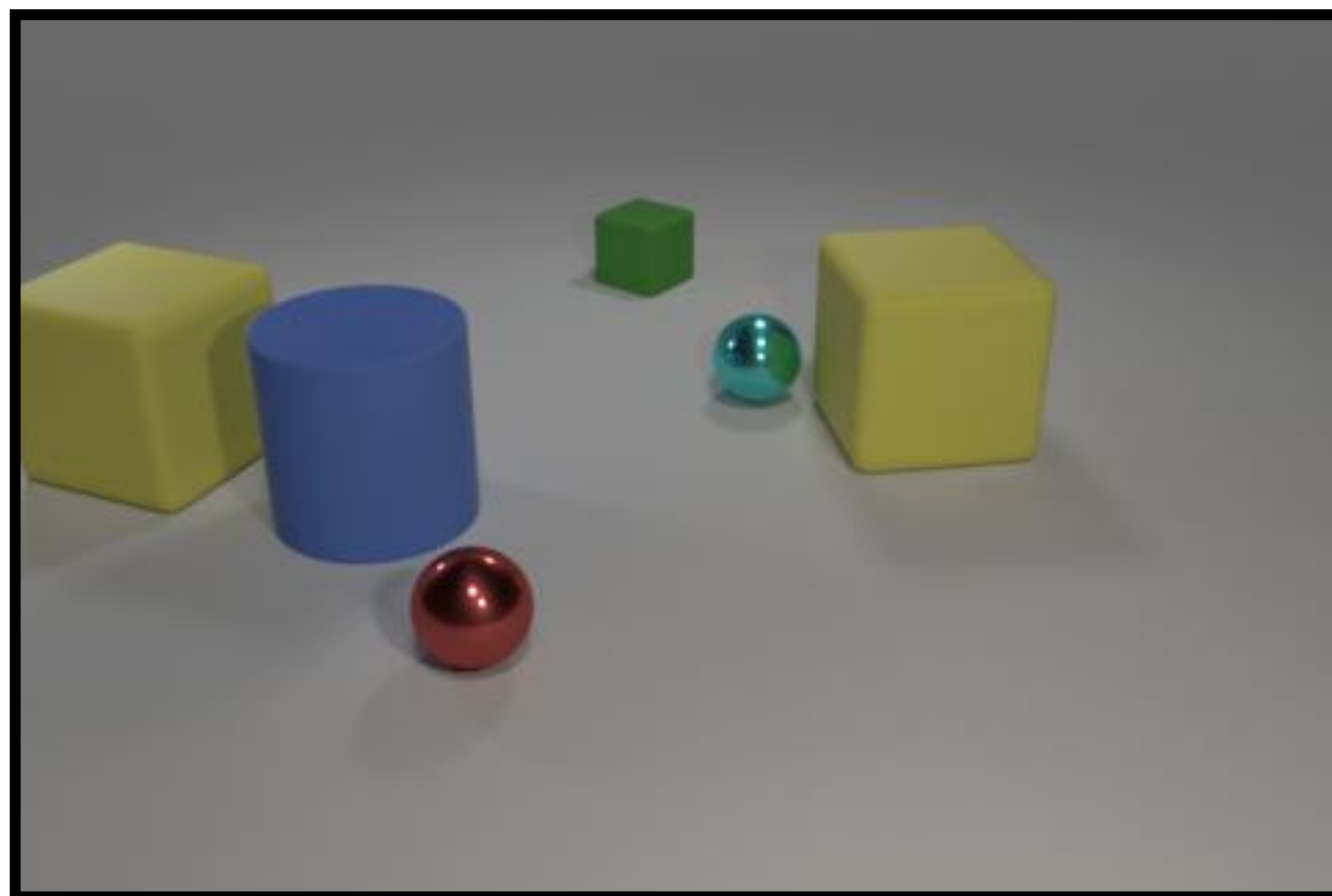
$$\arg \max_{\theta} \mathbb{E} \log p(\text{ans} \mid \text{ques}, \text{img})$$

hybrid supervised / policy gradient

$$\nabla \mathbb{E} r \approx \nabla \log p(\text{ans} \mid \text{img}, \text{img}) + \lambda (\nabla \log p(\text{img} \mid \text{ques})) \cdot \log p(\text{ans} \mid \text{img}, \text{img})$$



# Experimental evaluation



*Does the blue cylinder have the same material as the big block on the right side of the red metallic thing?*



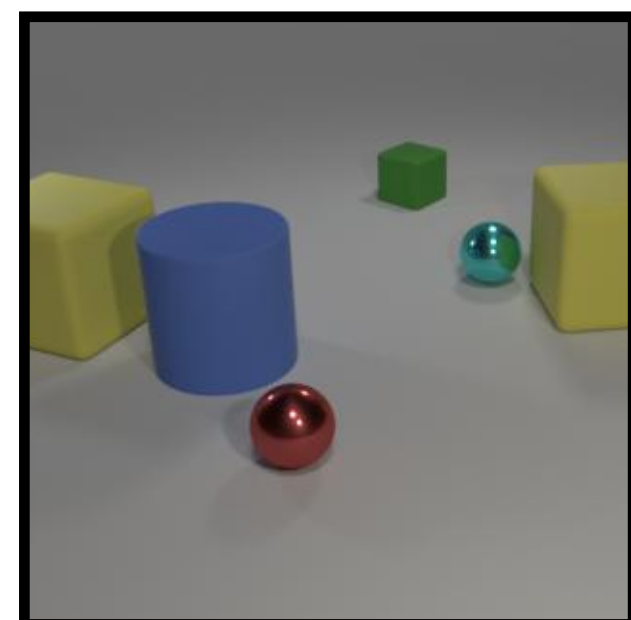
# Experimental evaluation



*What is in the sheep's ear?*



# Experimental evaluation [ARDK16a, HARDS17]



*How many other things are the same size as the cylinder?*

68.5

CNN +  
RNN

83.7  
NMN

96.9  
NMN  
[JH+17]



*What color is she wearing?*

62.5

CNN +  
RNN

64.7

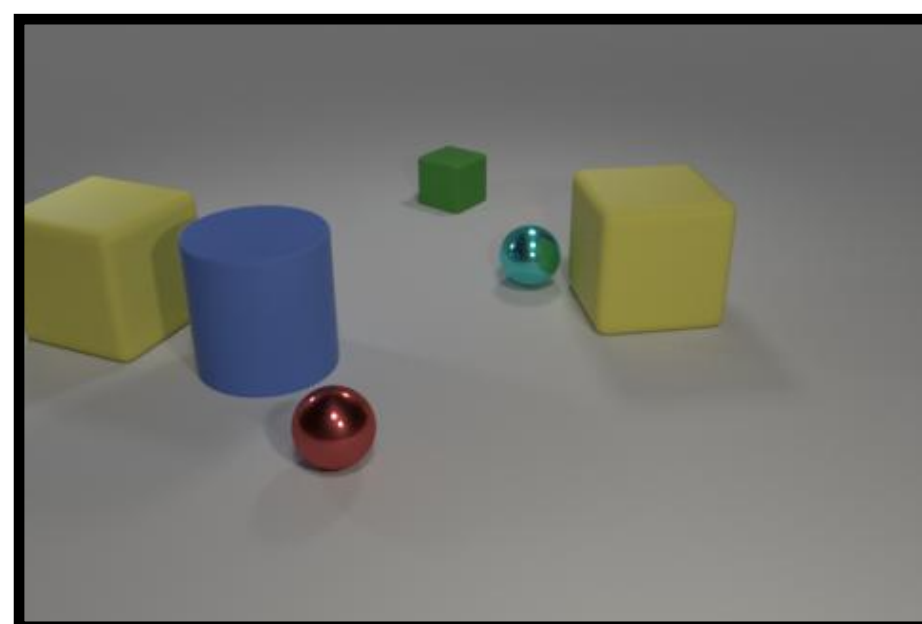
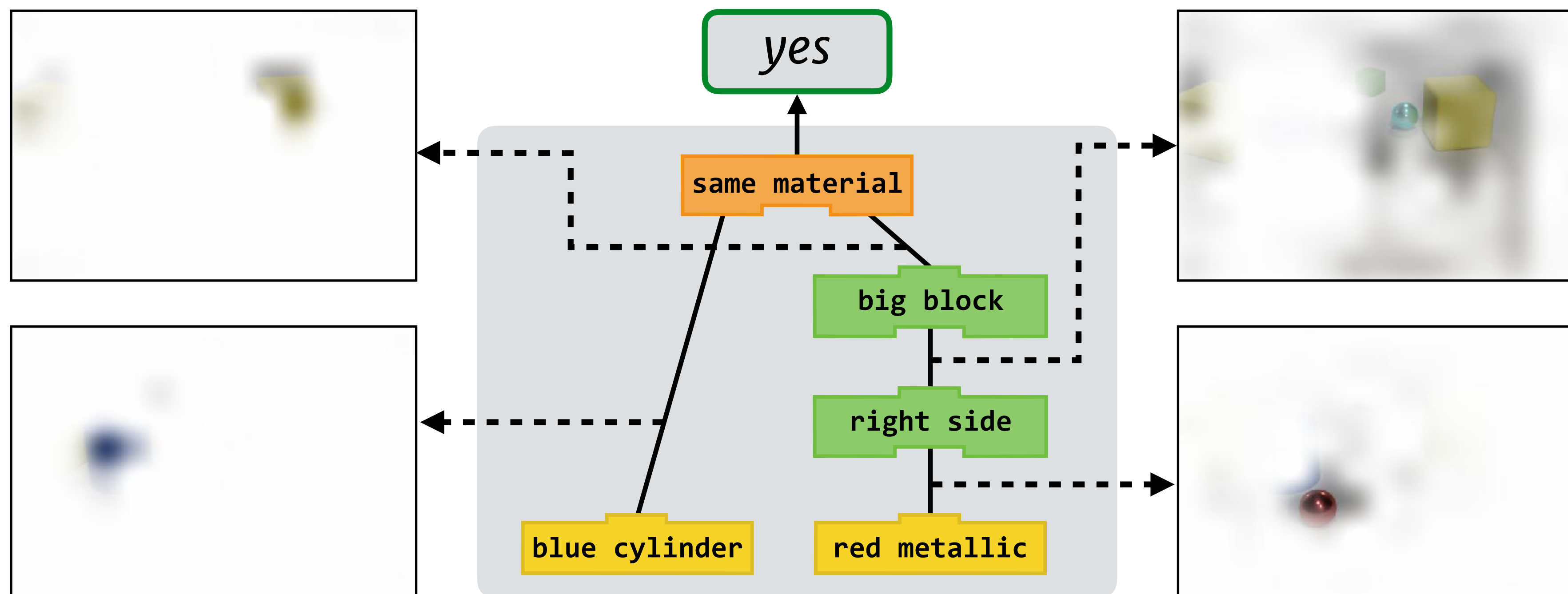
MCB

64.9

NMN



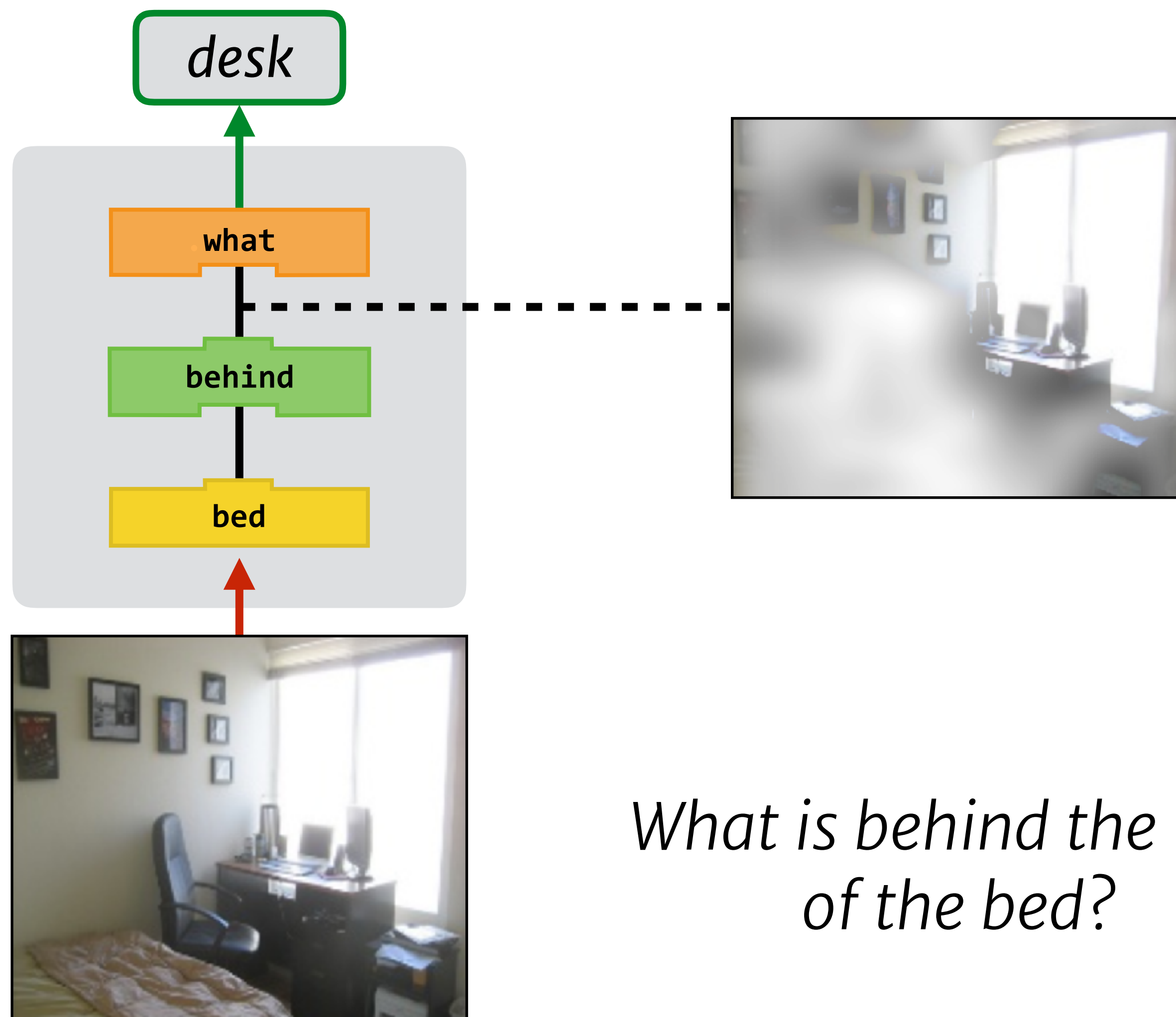
# Experimental evaluation [ARDK16a, HARDS17]

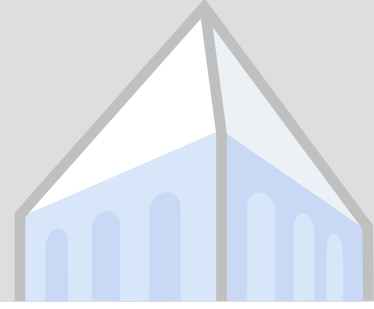


*Does the blue cylinder have the same material as the big block on the right side of the red metallic thing?*



# Experimental evaluation [ARDK16a, HARDSDS17]





# NMNs and strong generalization [ARDK16a]

## TRAIN

*Is there anything left of a circle?  
Is there anything above a circle?*

## TEST

*Is there anything **above and left**  
of a circle?*





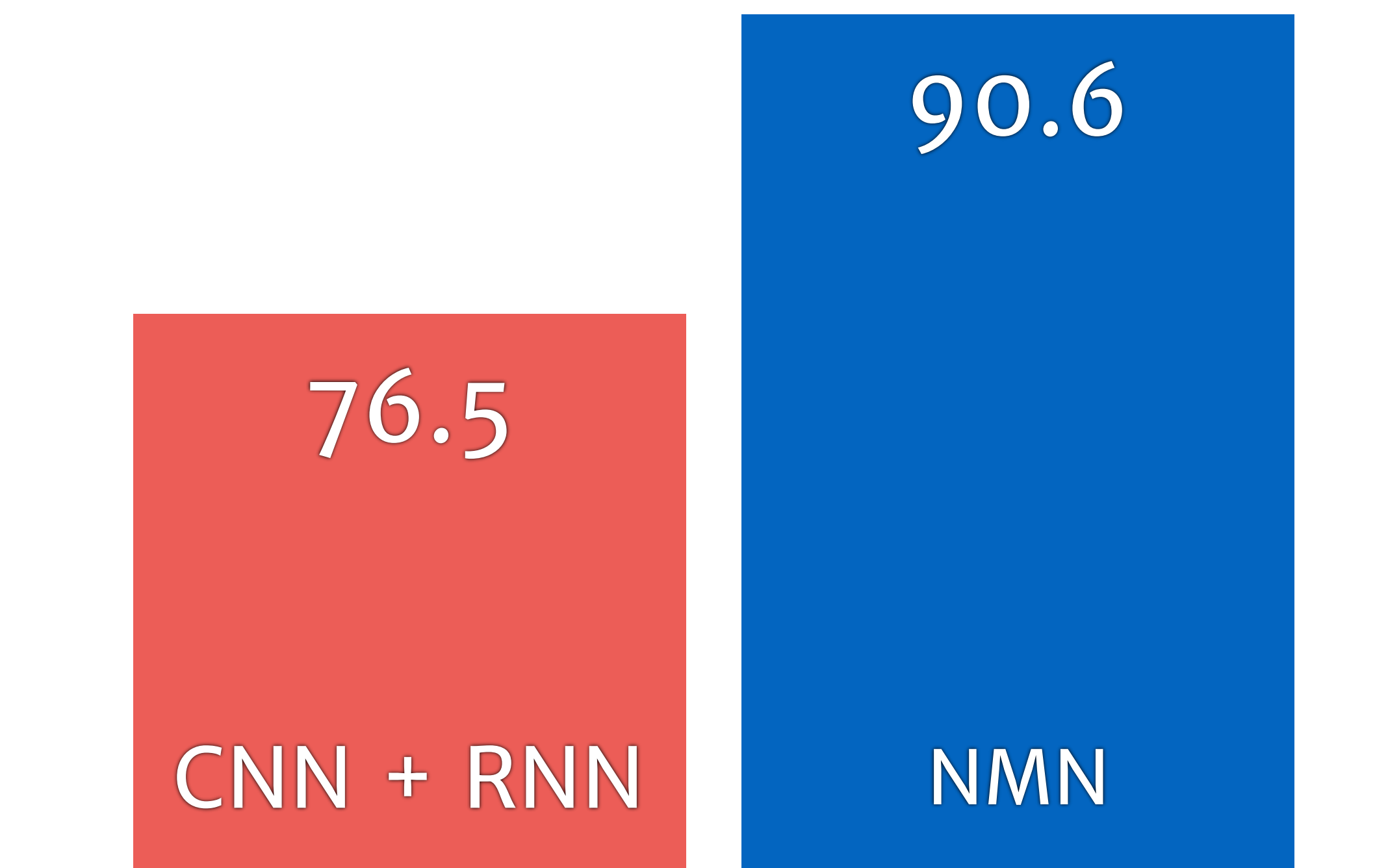
# NMNs and strong generalization [ARDK16a]

## TRAIN

*Is there anything left of a circle?  
Is there anything above a circle?*

## TEST

*Is there anything **above and left**  
of a circle?*





# NMNS for other tasks

*Is Key Largo  
an island?*

name	type	coastal	island
<i>Columbia</i>	city	no	no
<i>Cooper</i>	river	yes	no
<i>Charleston</i>	city	yes	no

[[A](#), Rohrbach, Darrell, Klein; 16b]

*man in sunglasses  
walking towards  
two men*

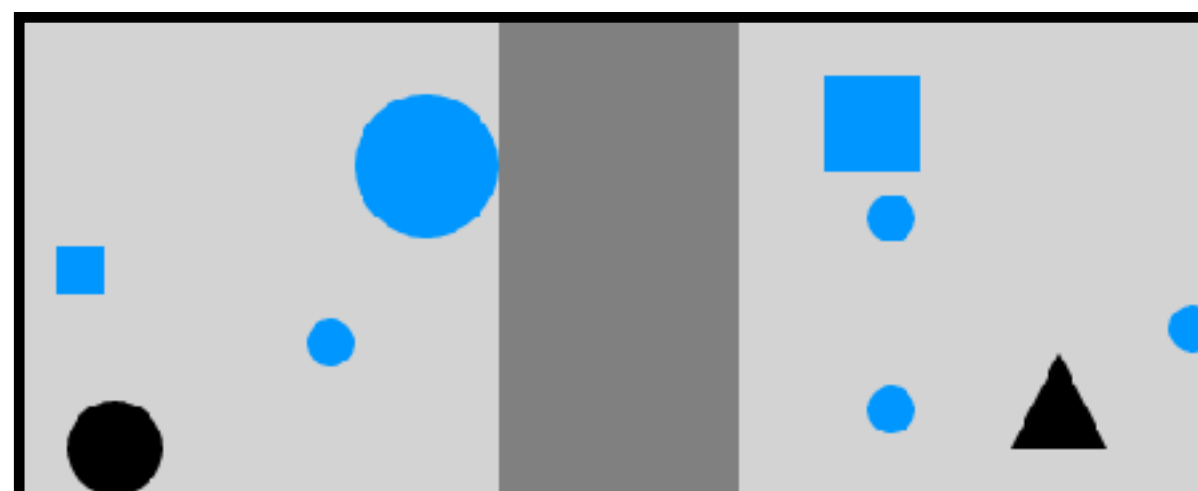


[Hu, Rohrbach, [A](#), Darrell, Saenko; 17]

[Cirik, Berg-Kirkpatrick, Morency; 18]

[Yu, Lin, Shen, Yang, Lu, Bansal, Berg; 18]

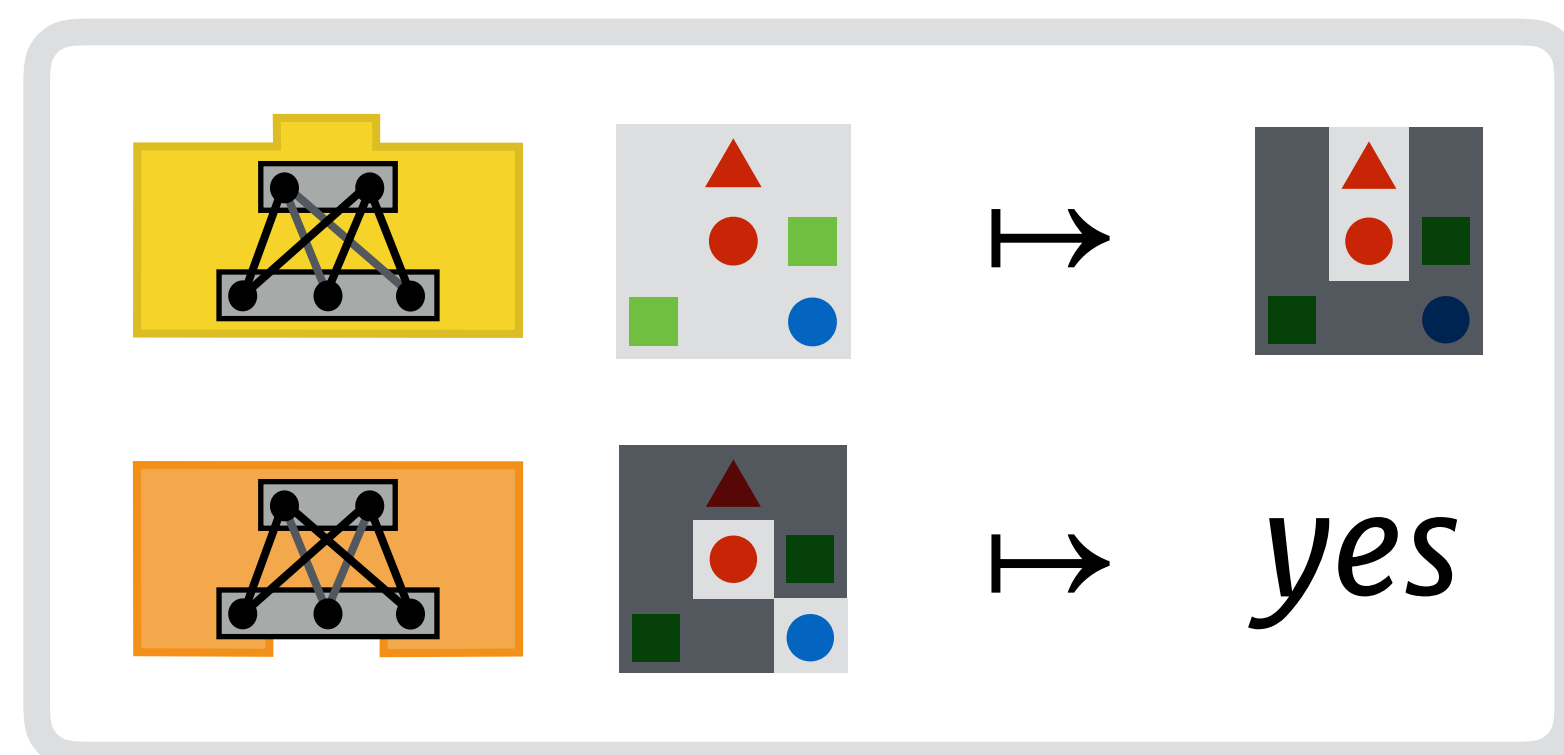
*There is exactly one  
black triangle not  
touching any edge.*



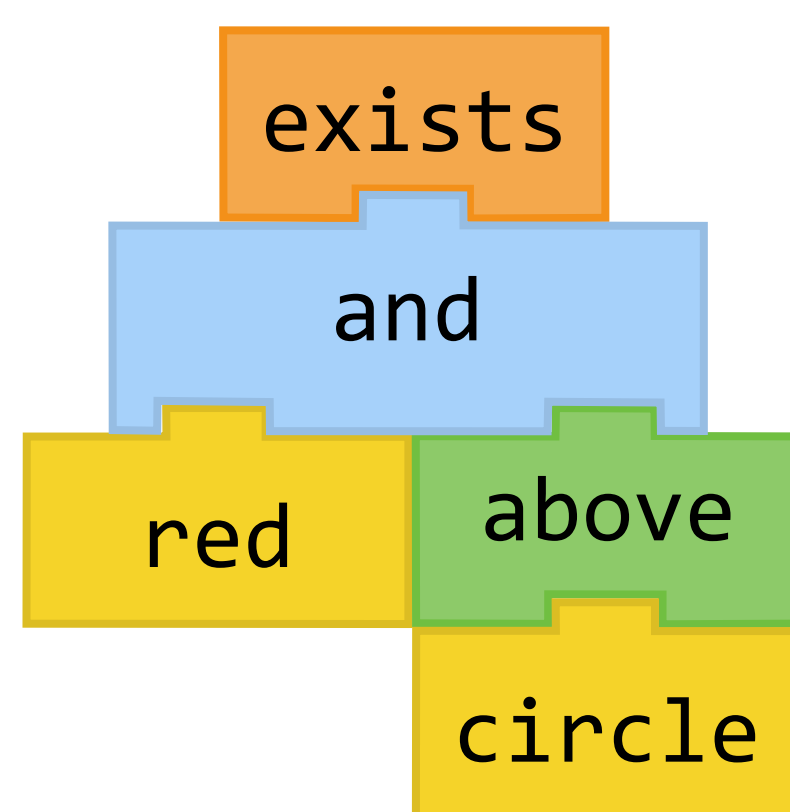
[Suhr, Lewis, Artzi; 17]



# Lessons



Linguistic structure lets us learn composable neural modules from weak supervision.



These modules allow us to more accurately interpret new statements, questions and references.

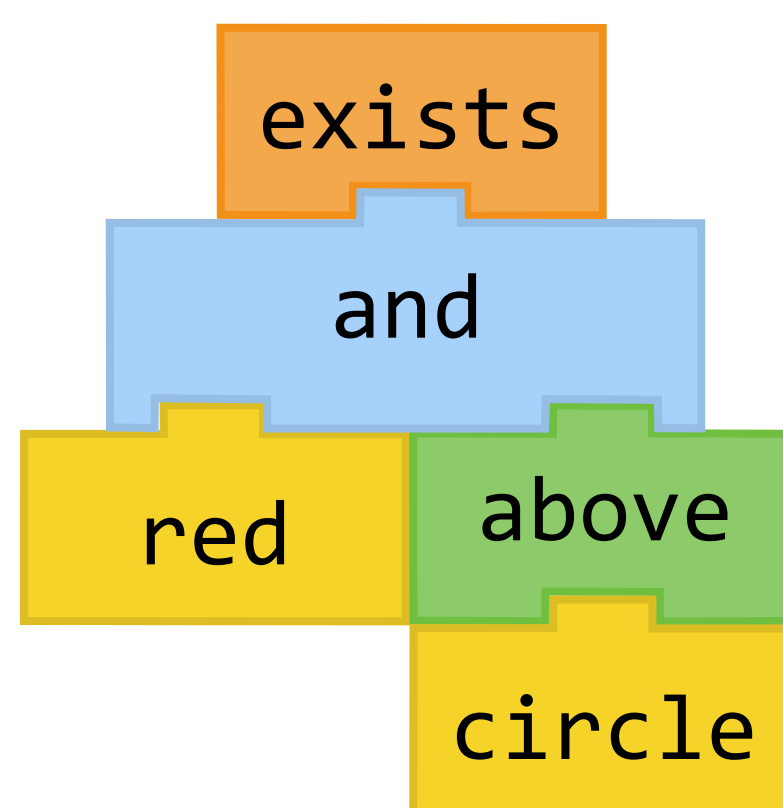
REASONING  
**LANGUAGE & LEARNING**  
BELIEF

[A](#), Klein & Levine. *Modular Multitask Reinforcement Learning [...]*. ICML 17.



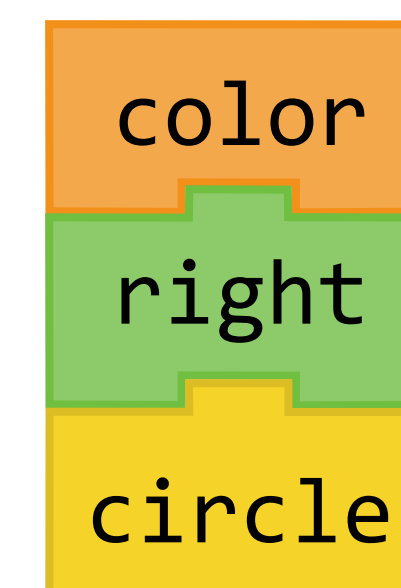
# Learning classifiers

*yes*



*Is there a red shape above a circle?*

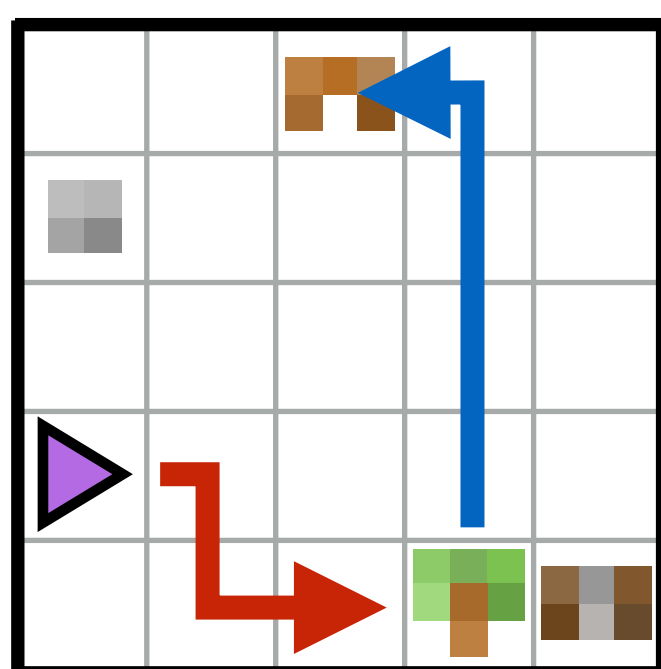
*blue*



*What color is the shape right of a circle?*

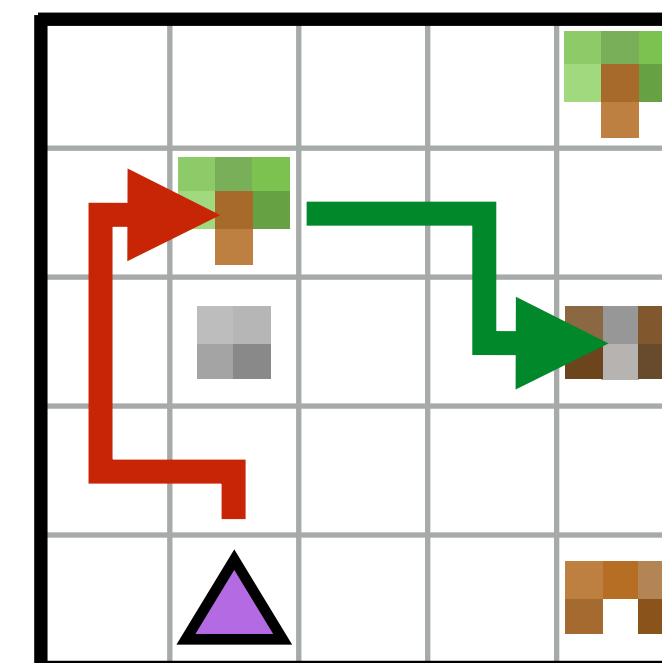


# Learning behaviors



get wood use saw

*Make planks:  
get wood, then use a saw.*

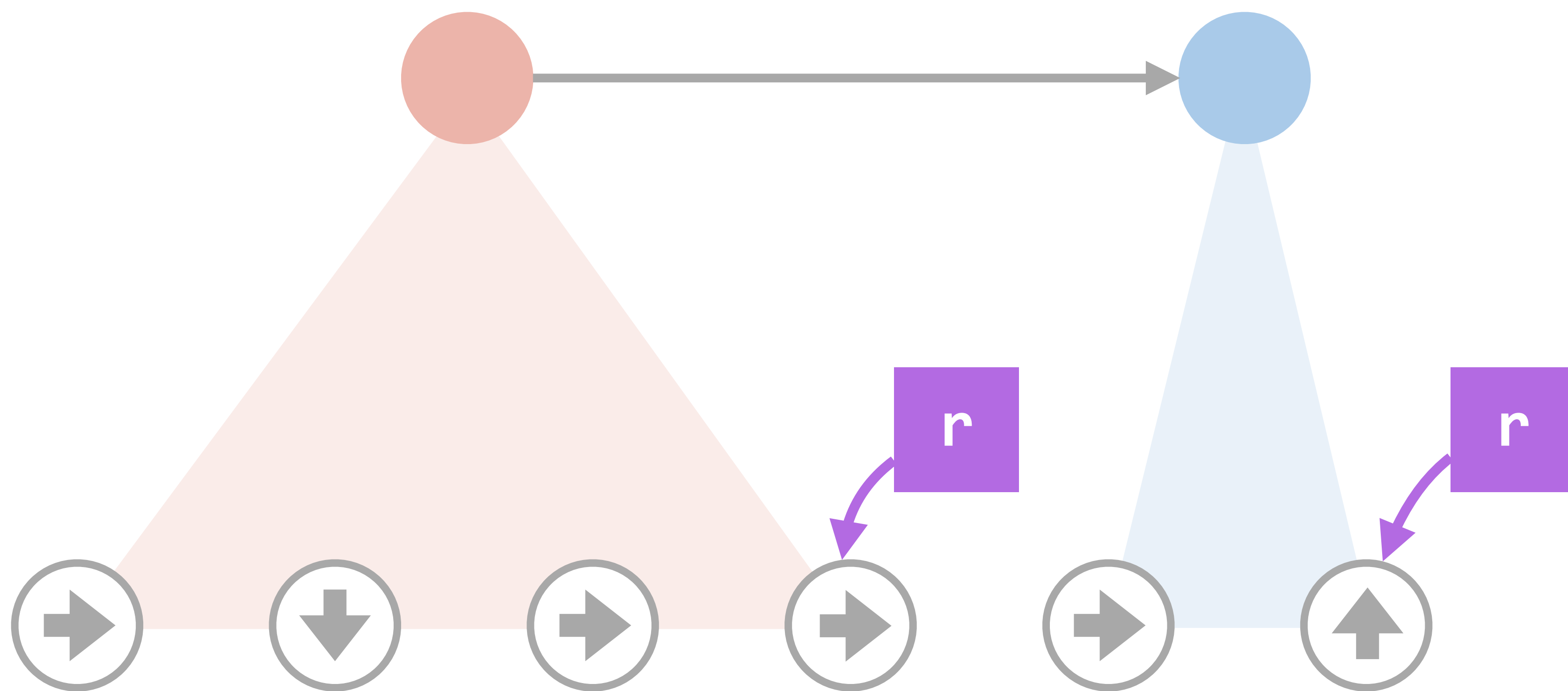


get wood use axe

*Make sticks:  
get wood, then use an axe.*

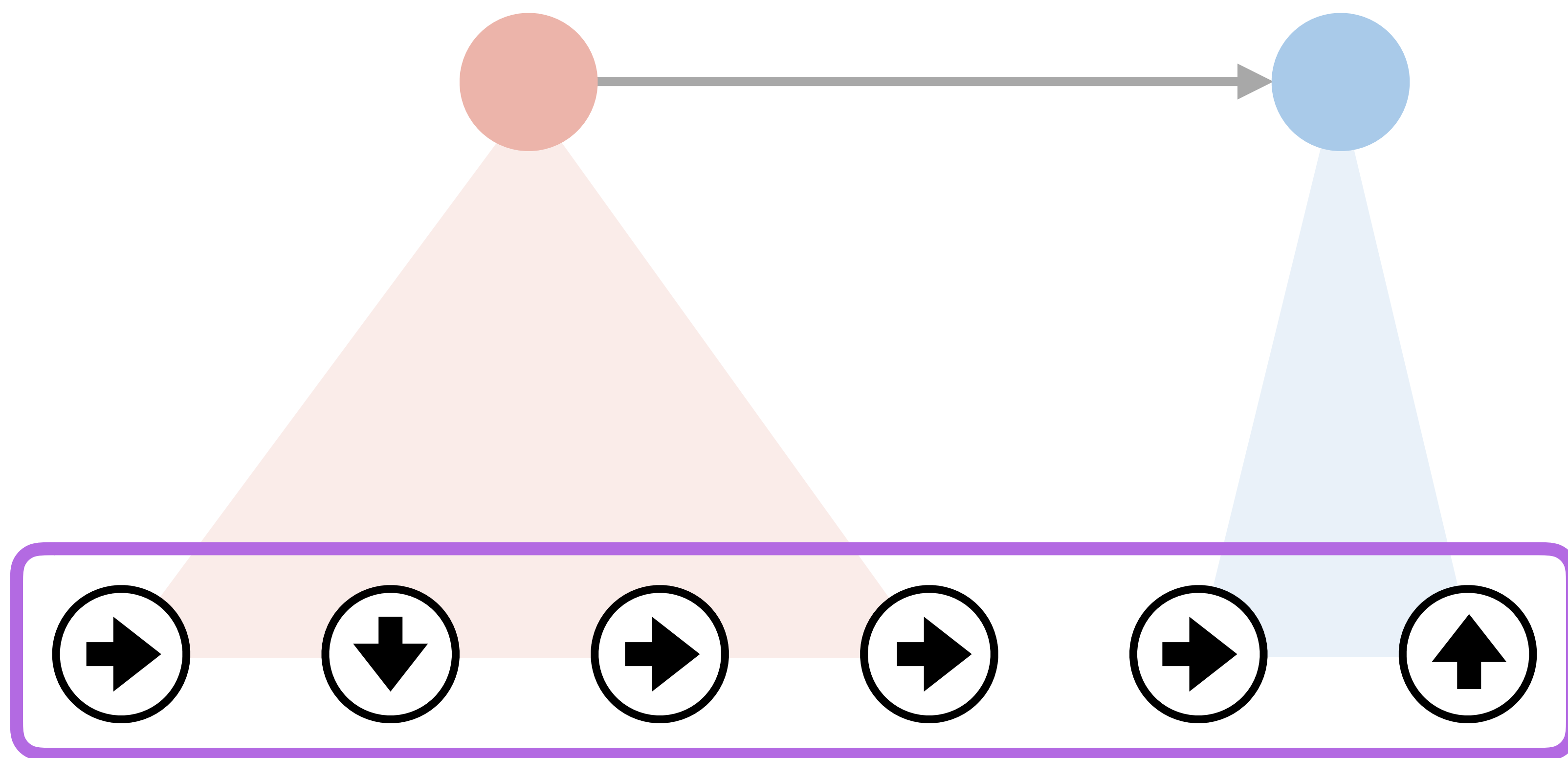


# Learning from intermediate rewards





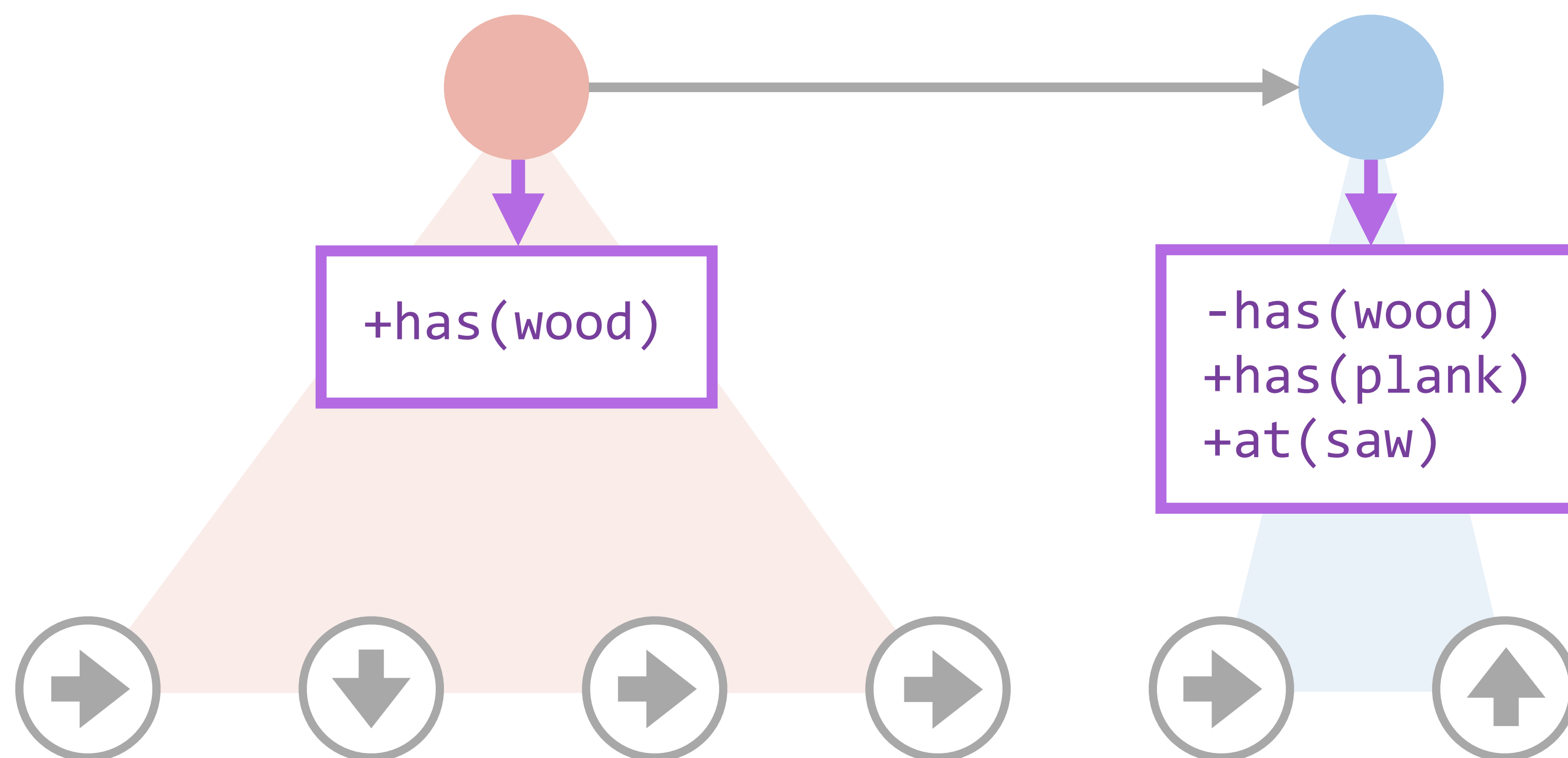
# Learning from demonstrations







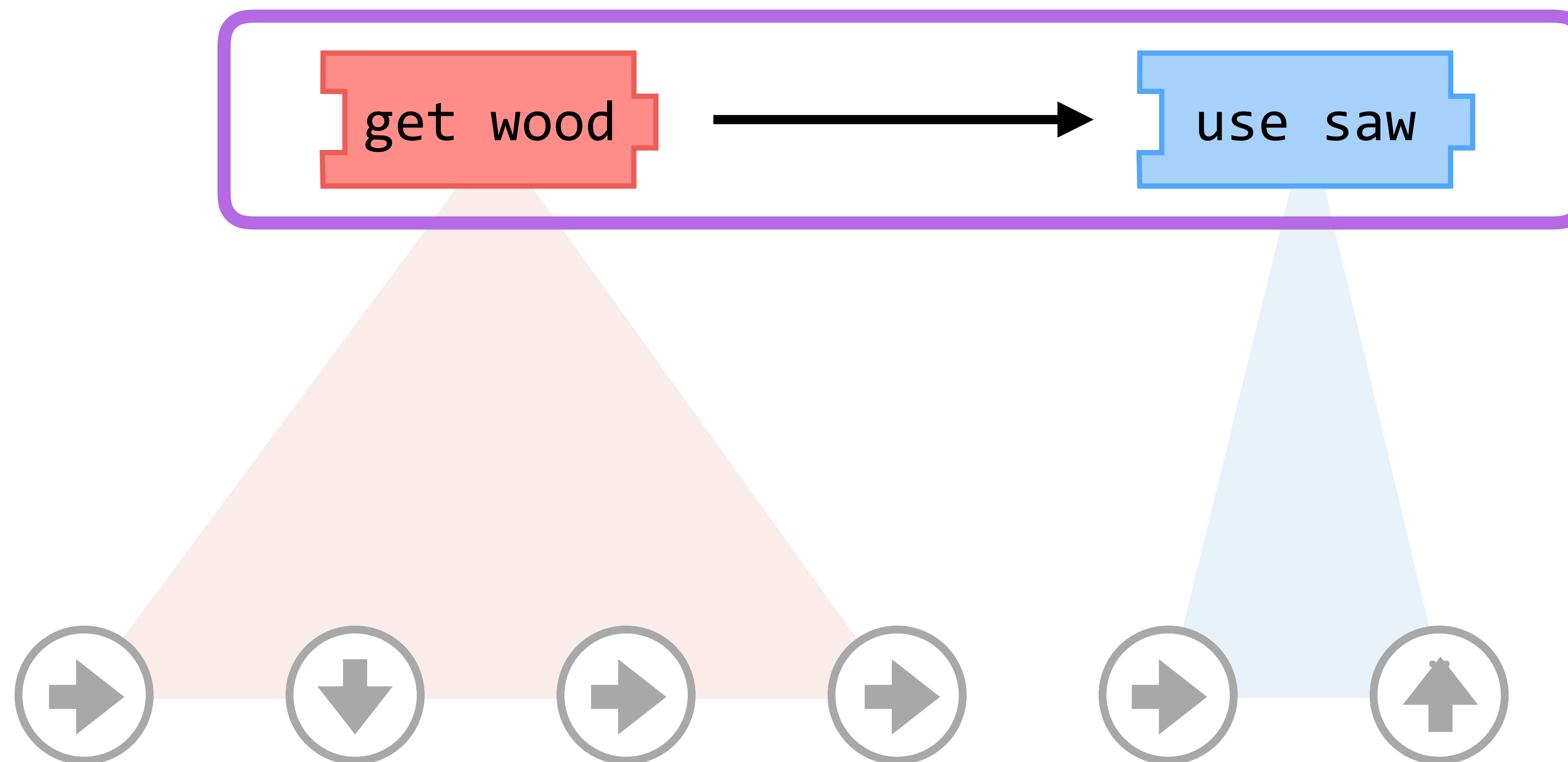
# Learning from intermediate rewards



[e.g. Sacerdoti 75, Hauskrecht et al. 98]



# Learning from sketches



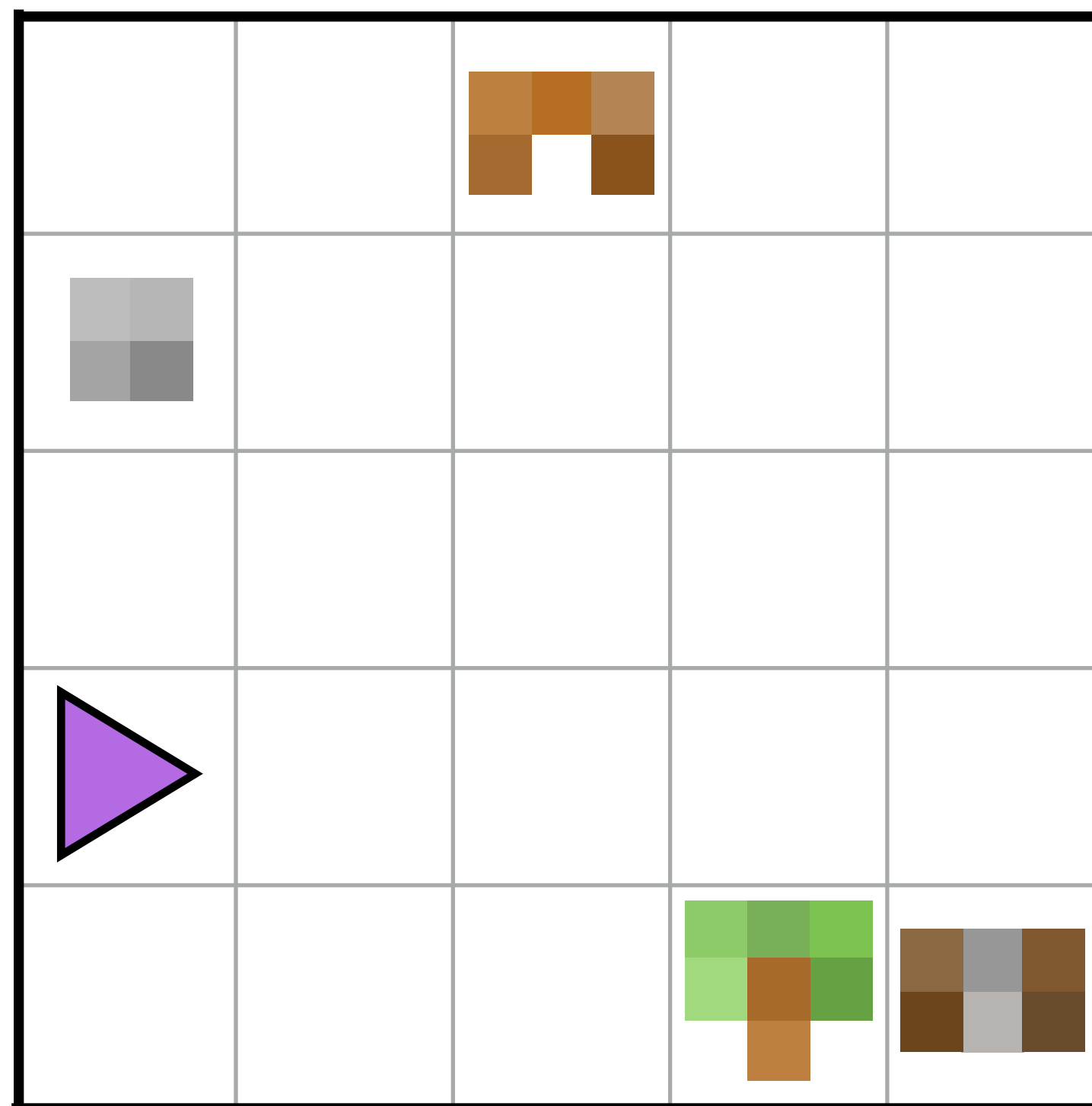


# Learning from sketches

*Make planks:*

get wood

use saw



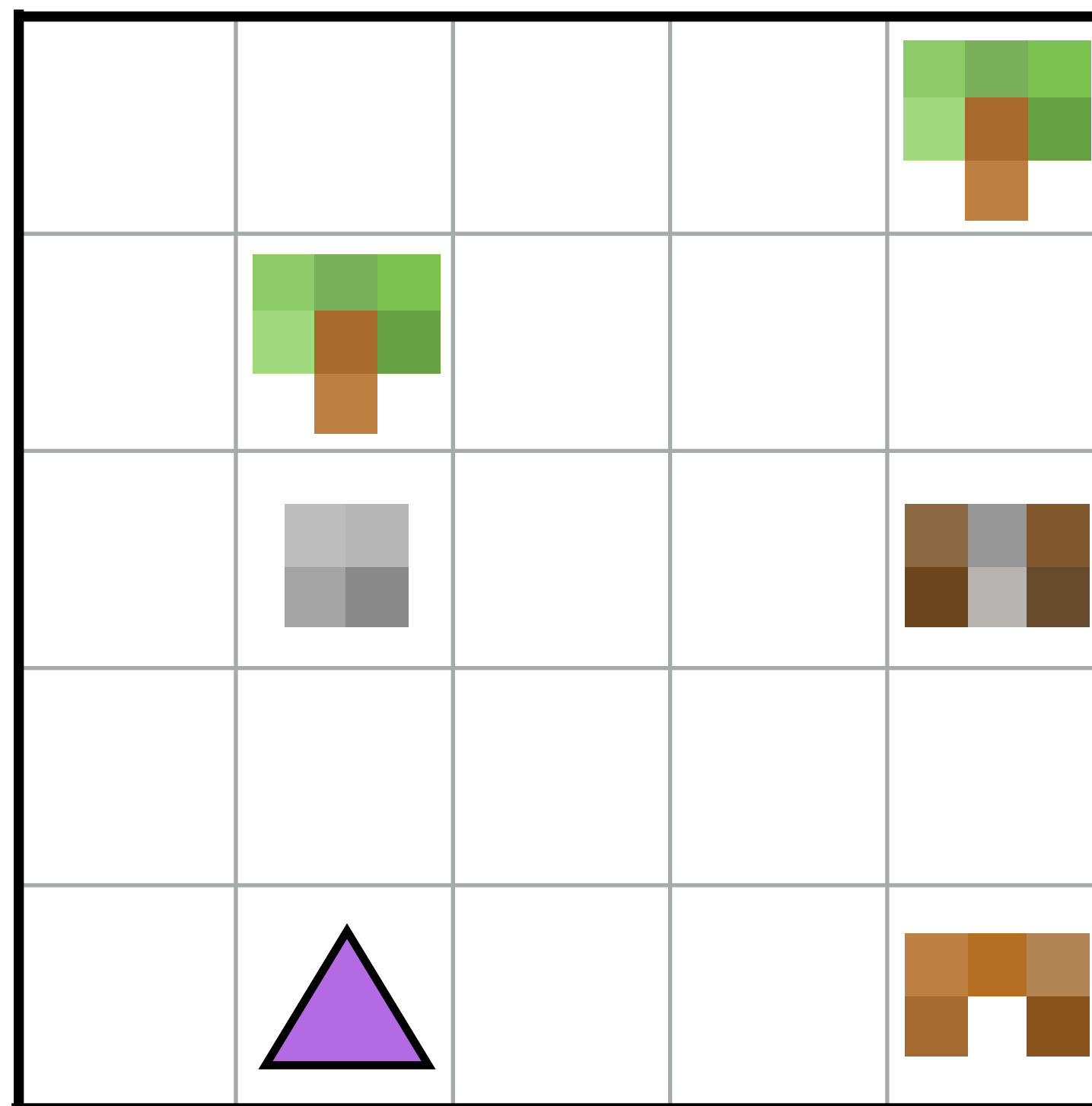


# Learning from sketches

*Make sticks:*

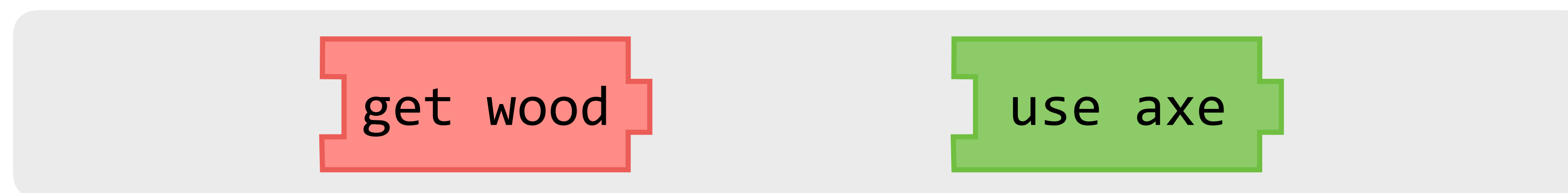
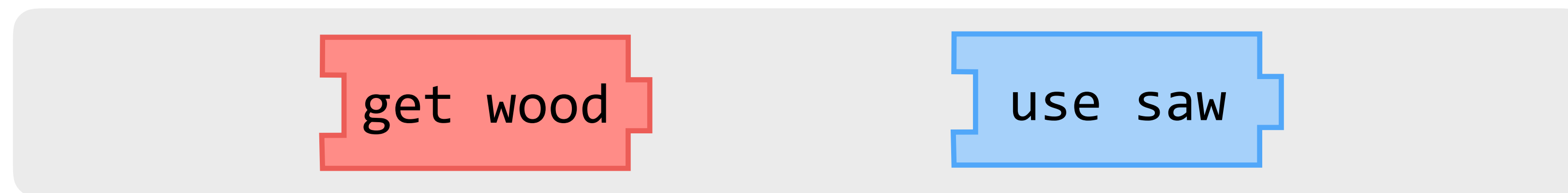
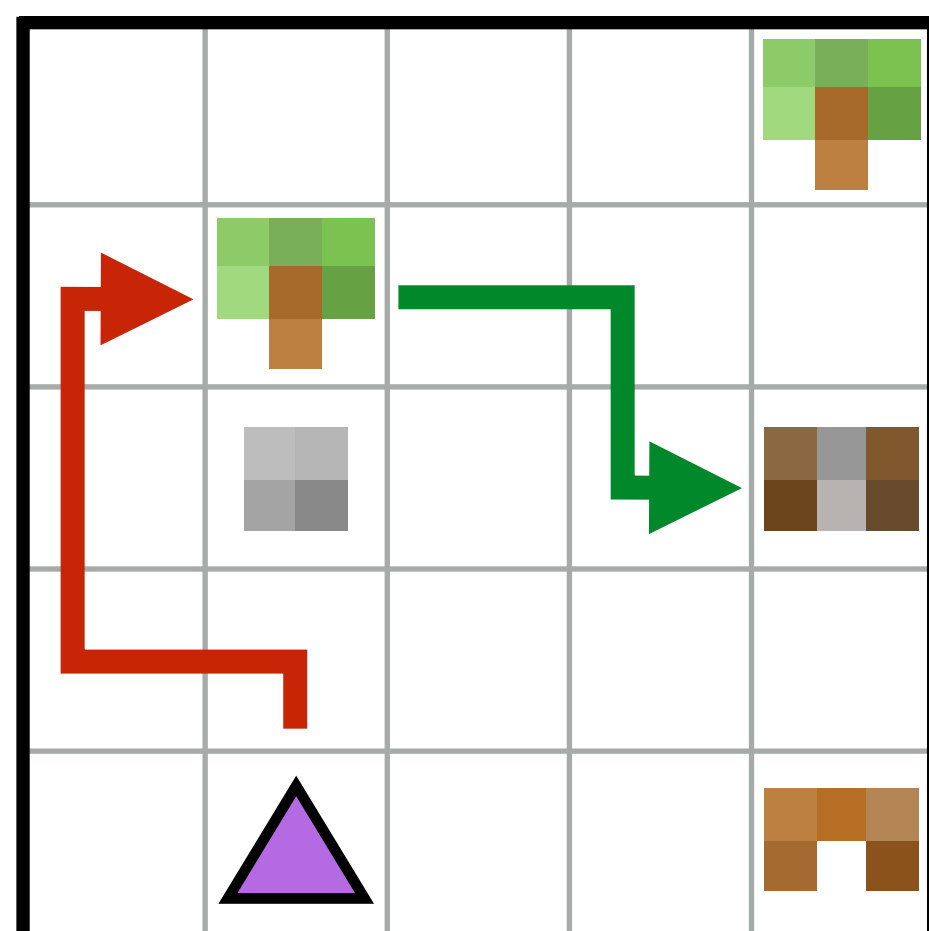
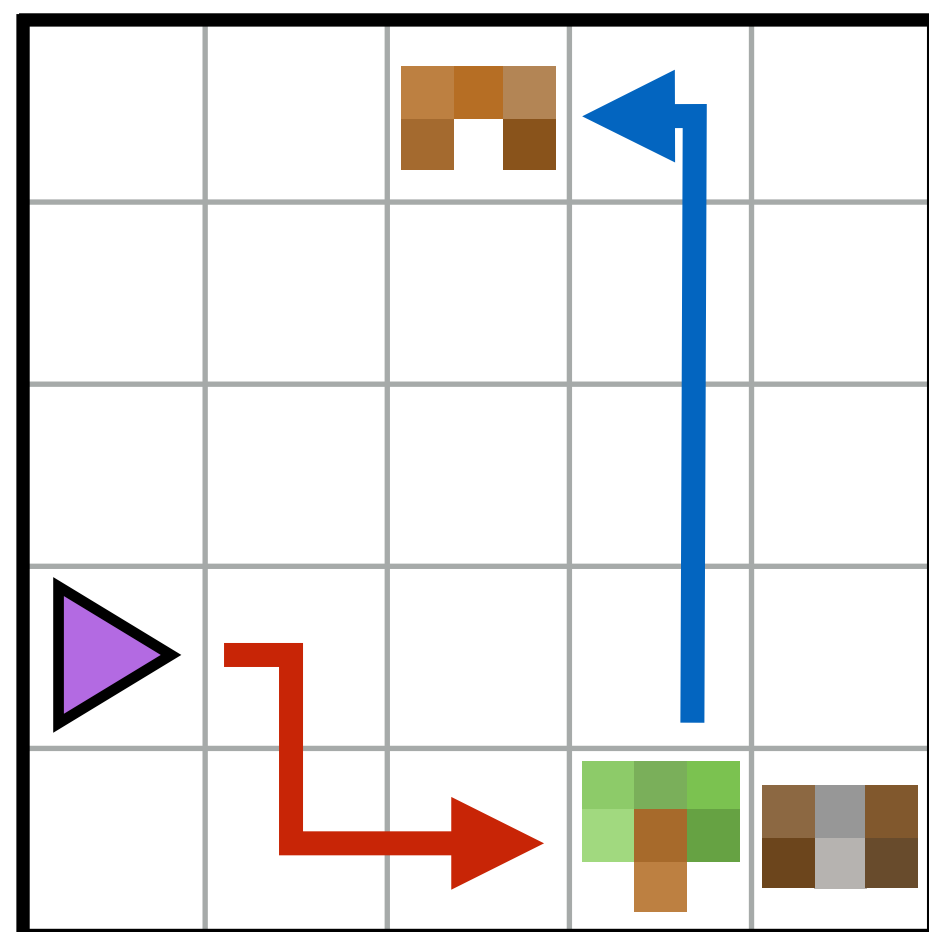
get wood

use axe



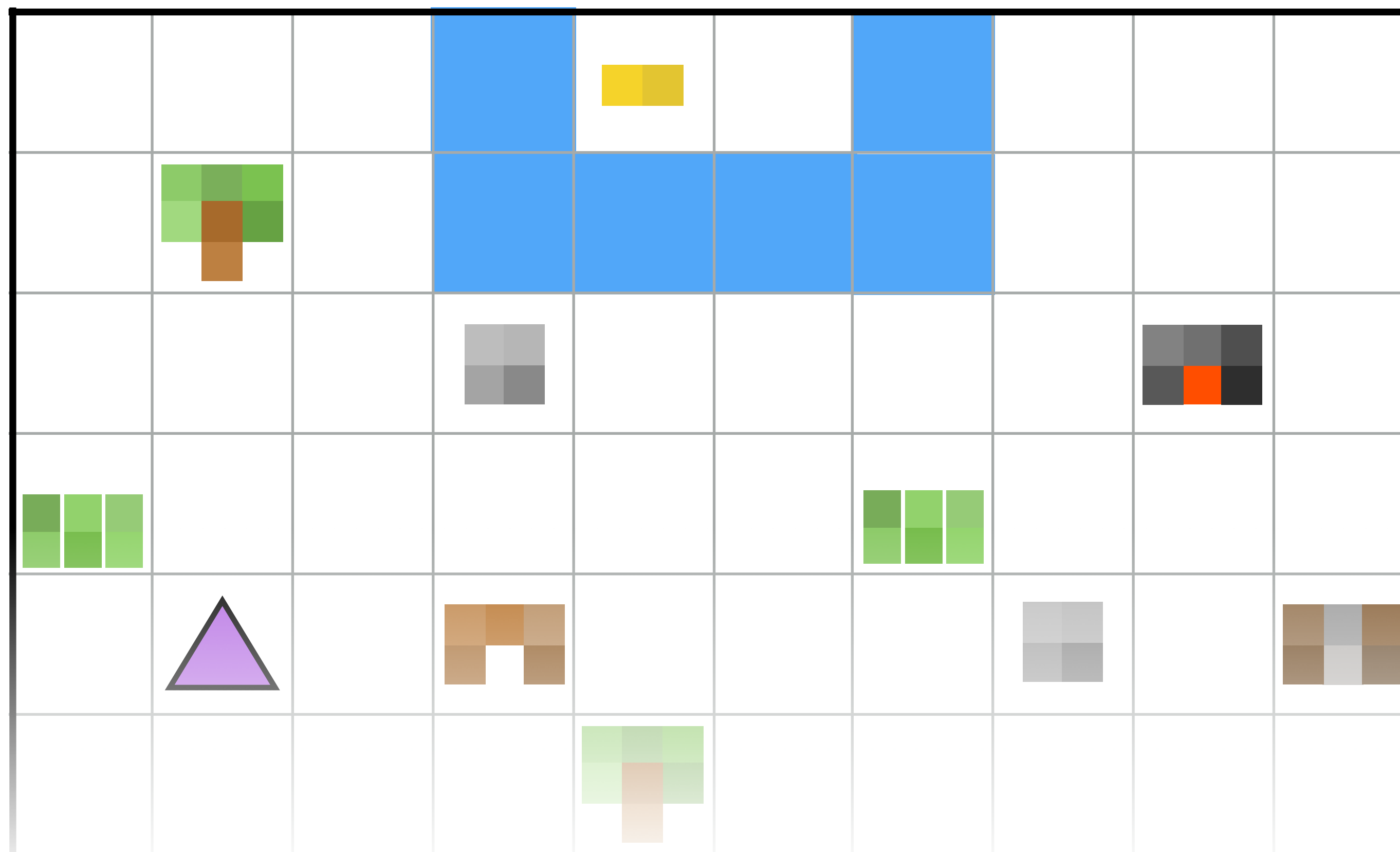


# Learning from sketches



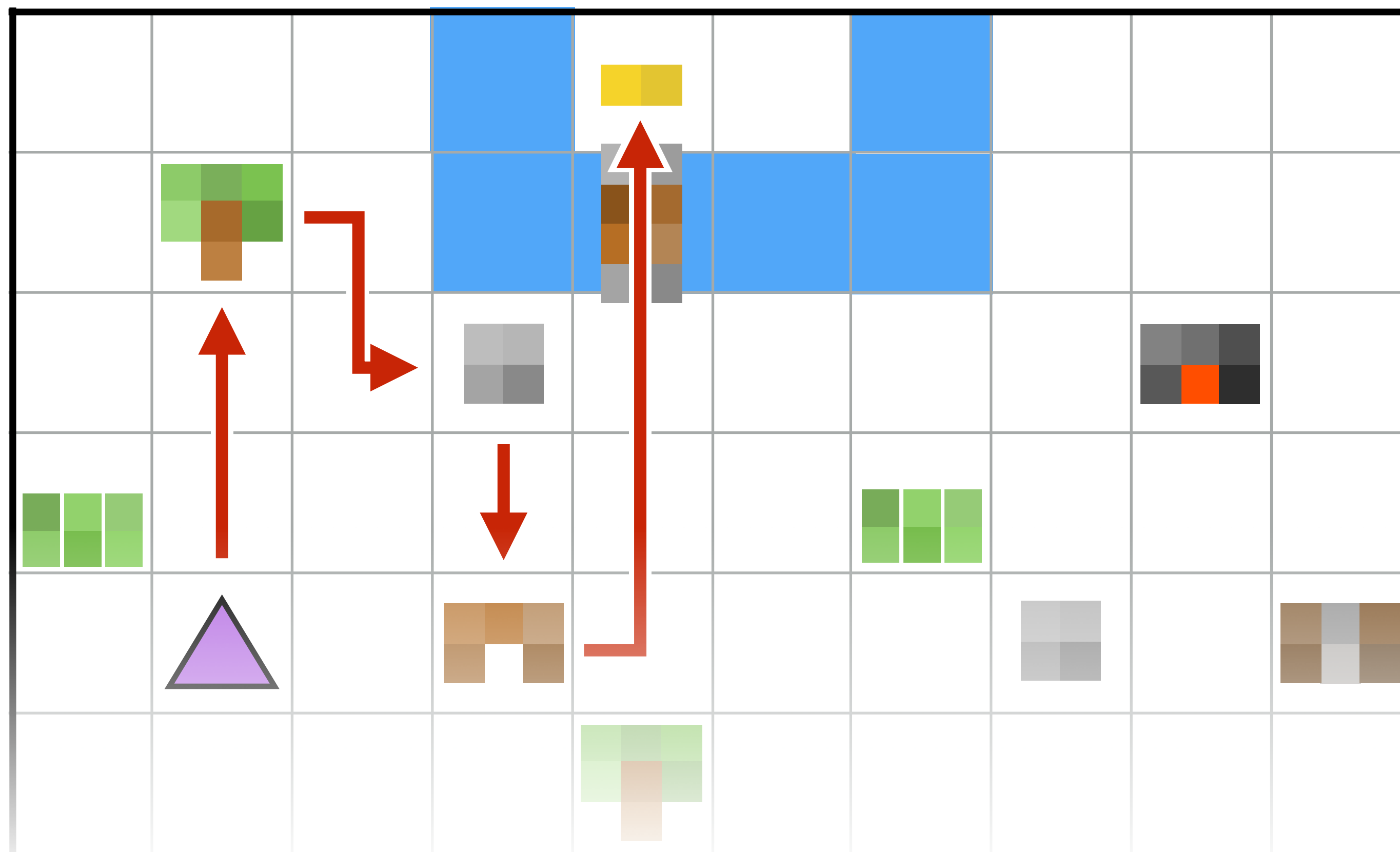


# Experiments: crafting game



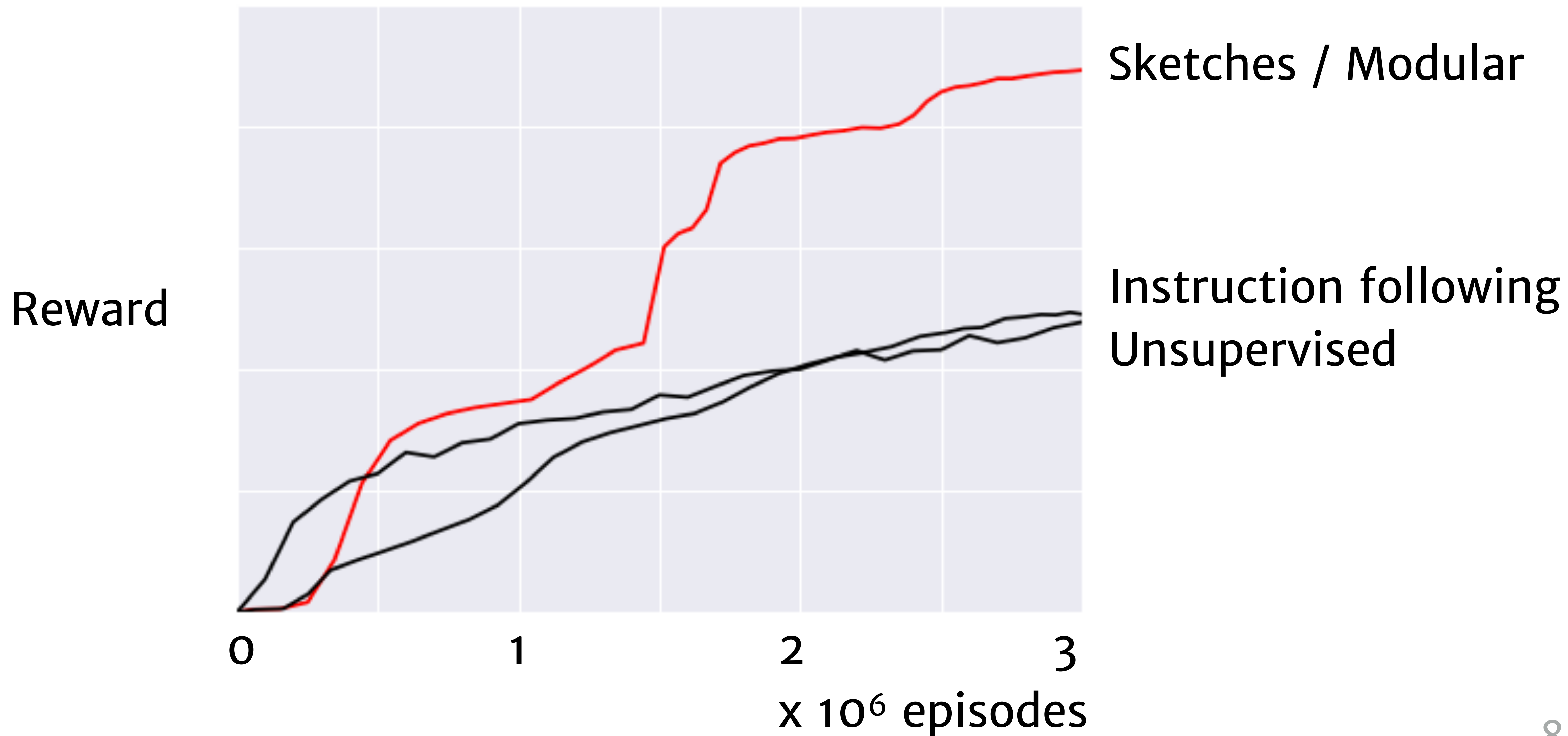


# Experiments: crafting game





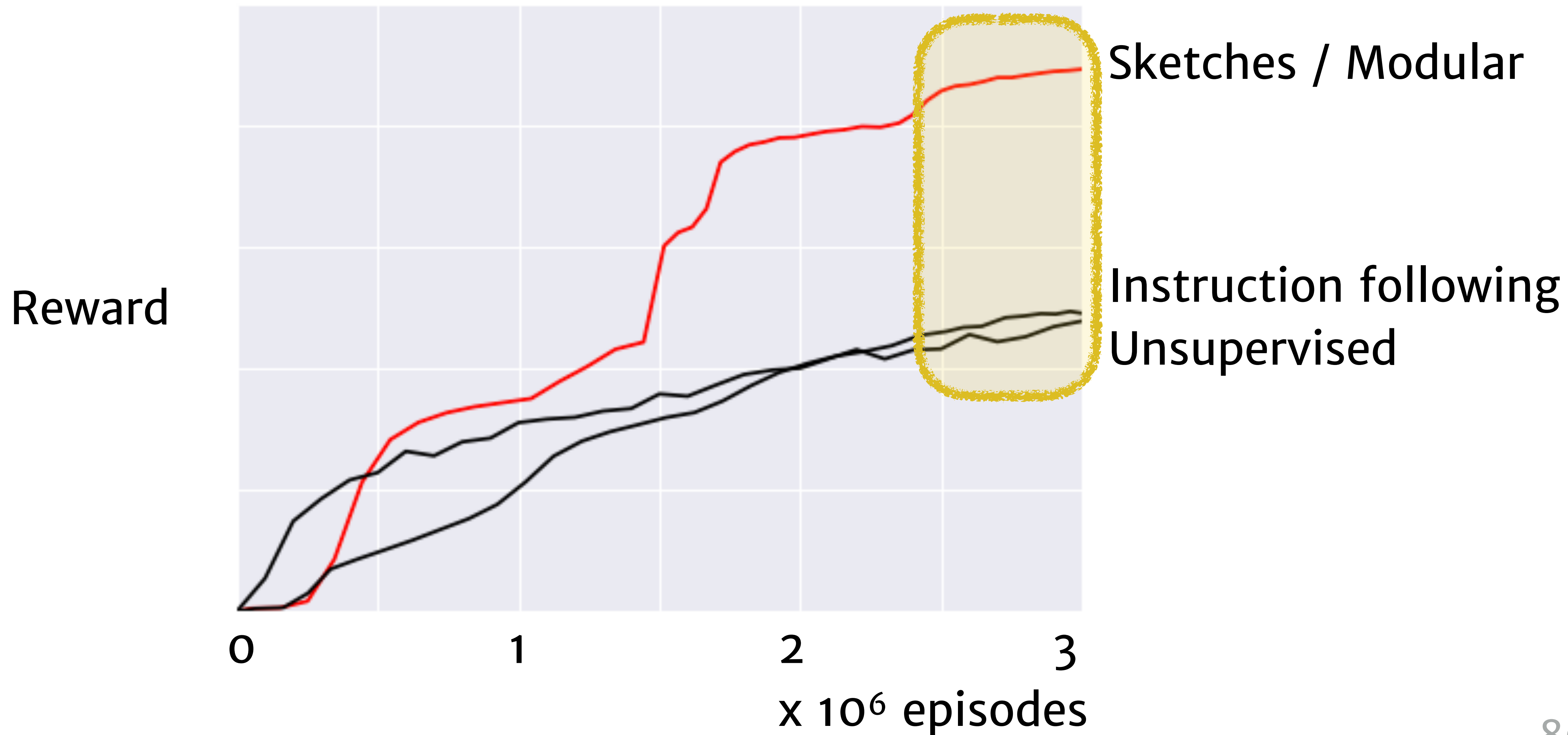
# Experiments: crafting game





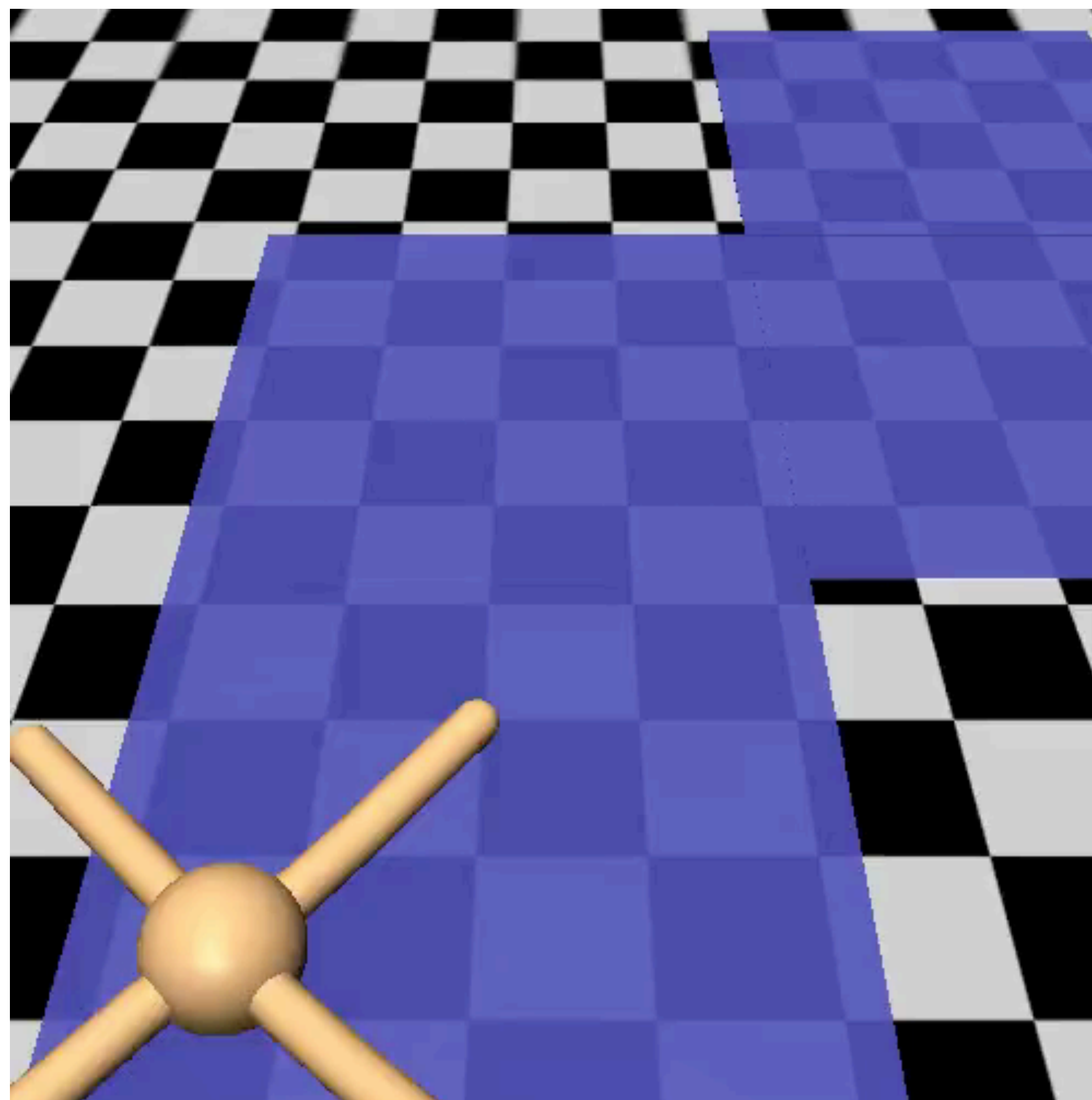


# Experiments: crafting game





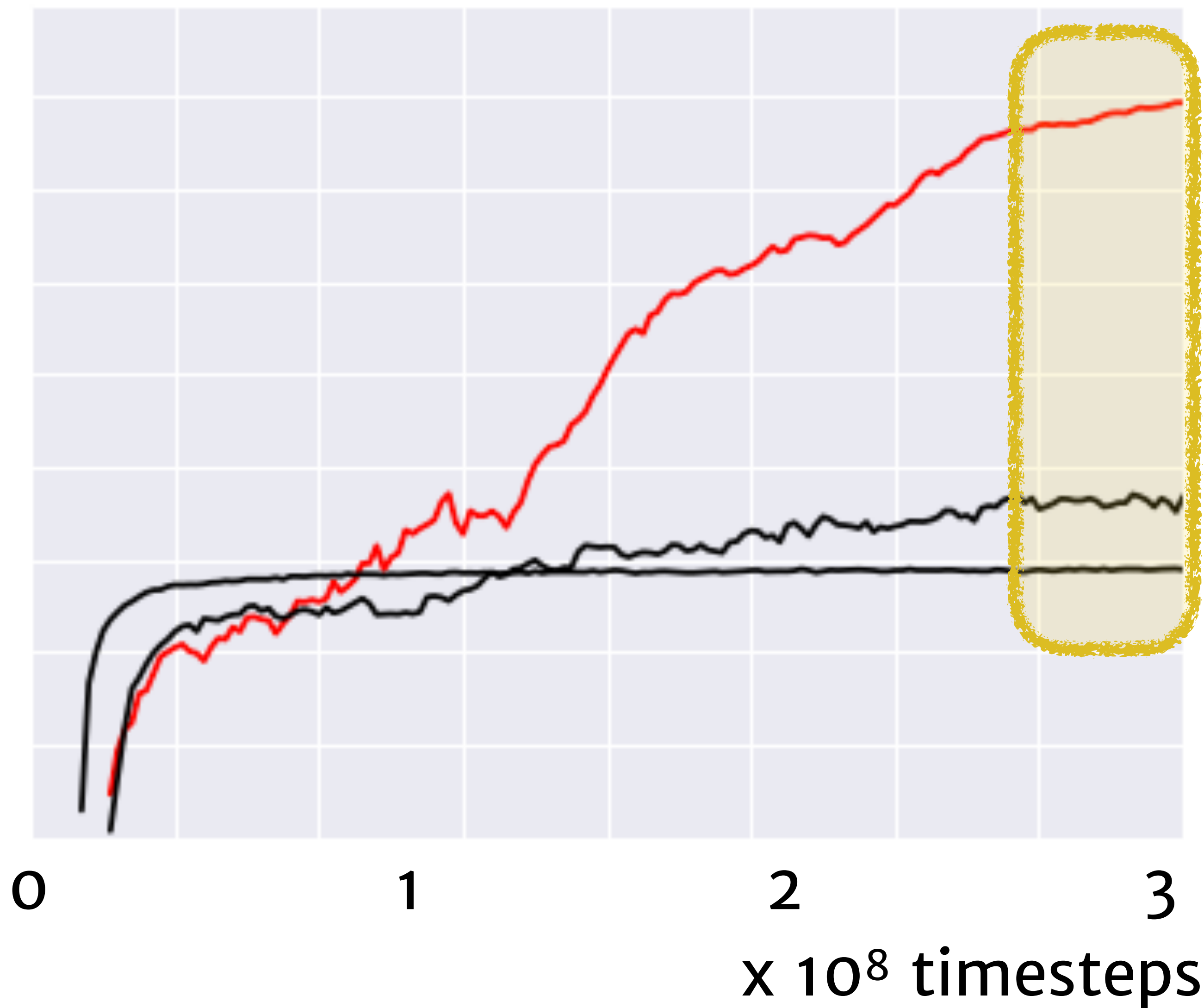
# Experiments: locomotion





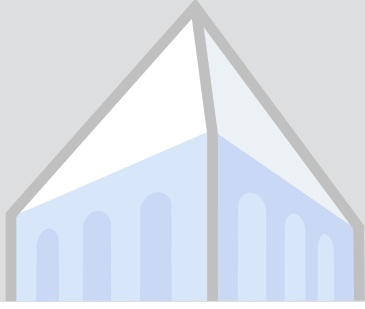
# Experiments: locomotion

log Reward



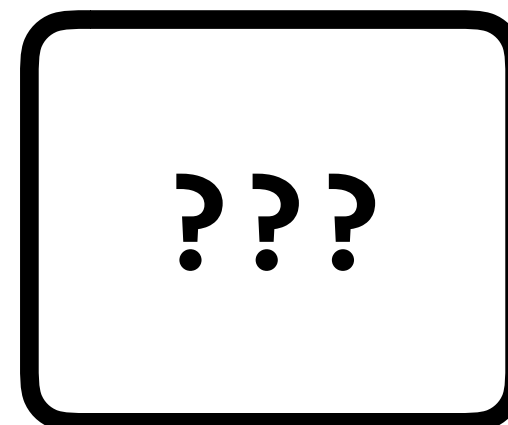
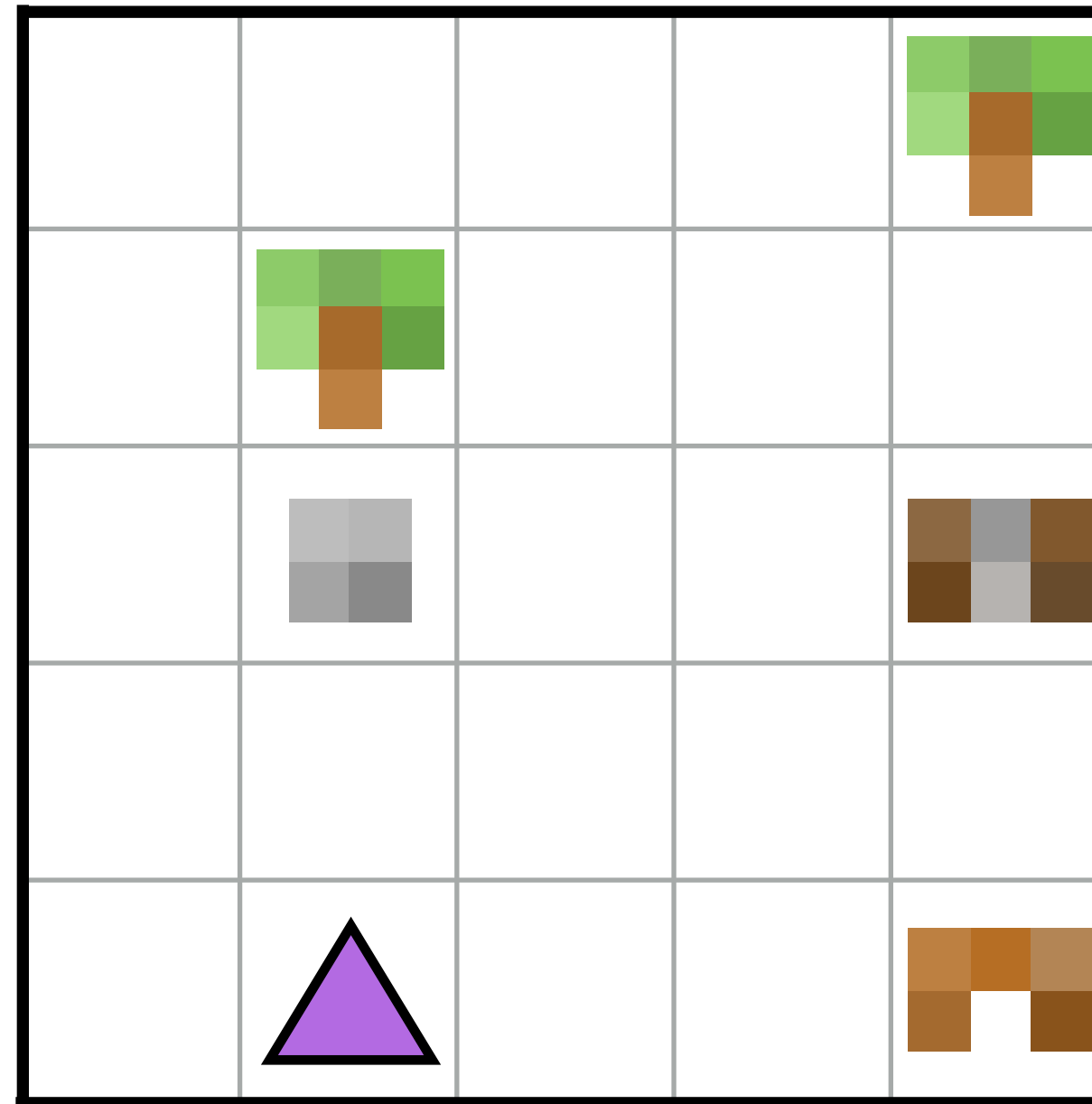
Sketches / Modular

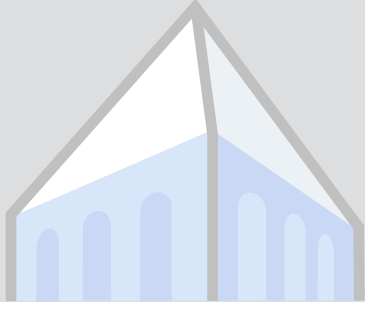
Instruction following  
Unsupervised



# Fast adaptation

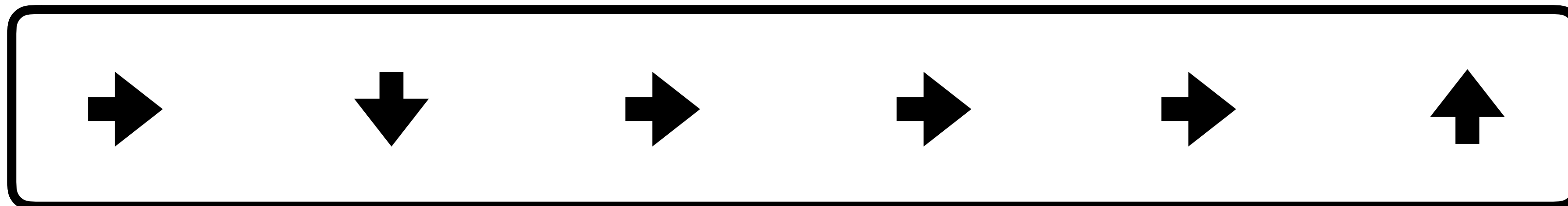
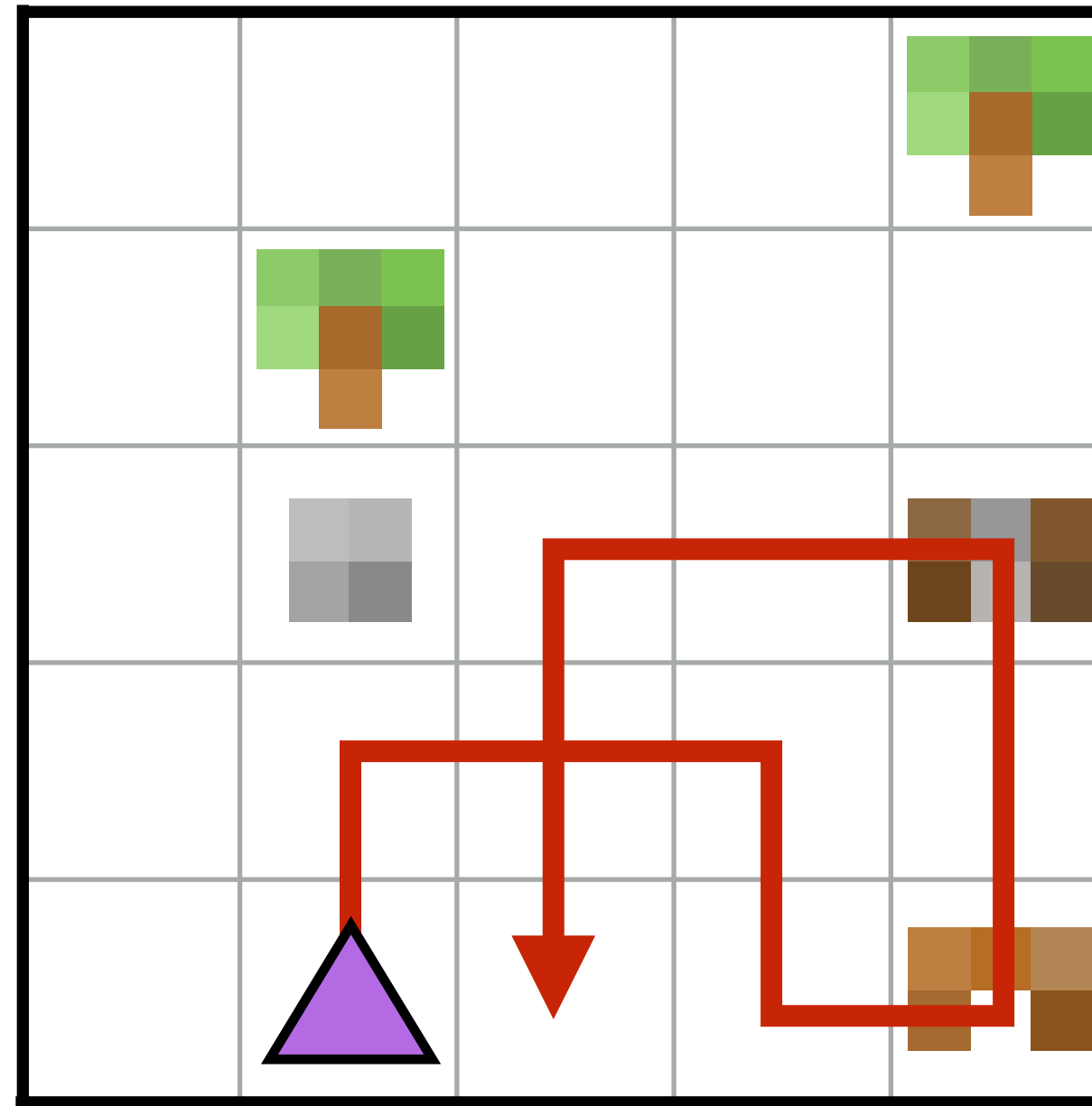
What if I don't get a sketch at test time?

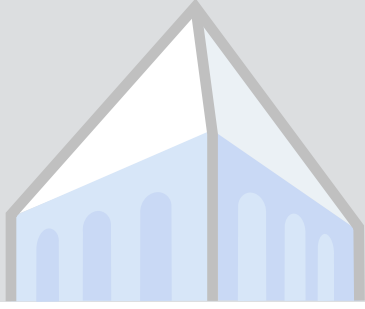




# Fast adaptation

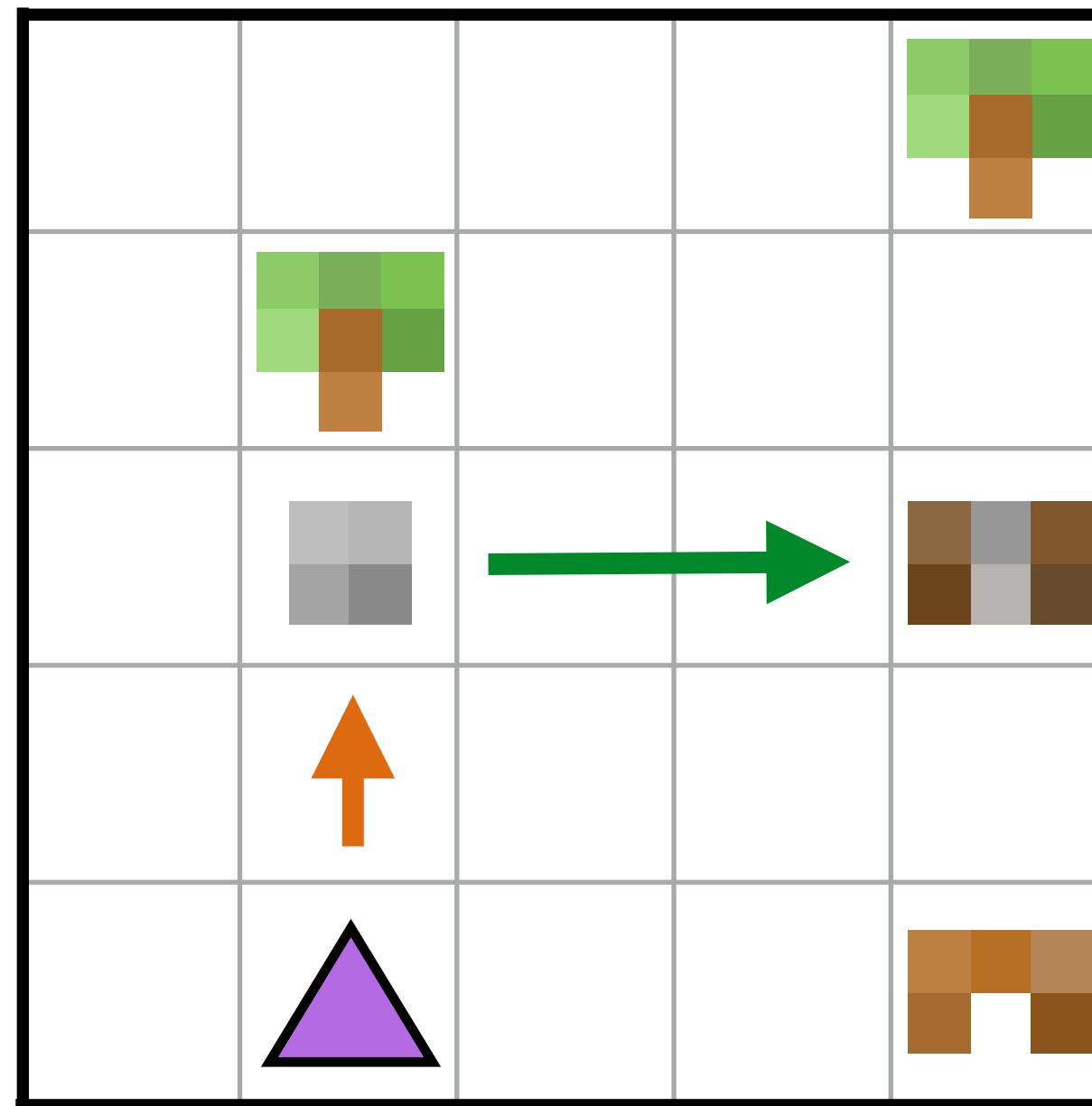
What if I don't get a sketch at test time?



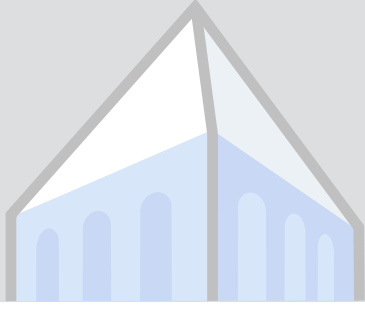


# Fast adaptation

What if I don't get a sketch at test time?

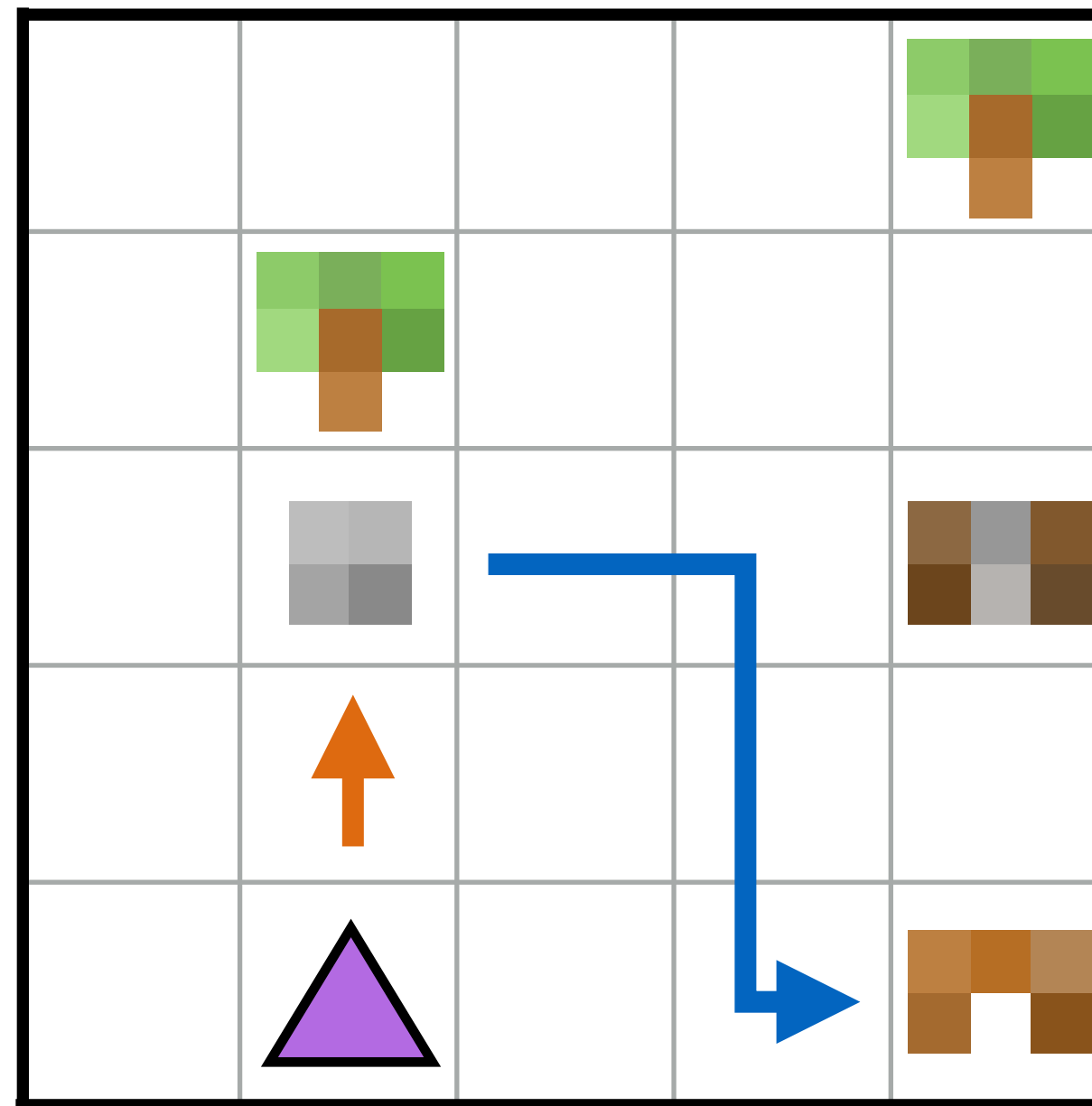


get iron use axe



# Fast adaptation

What if I don't get a sketch at test time?

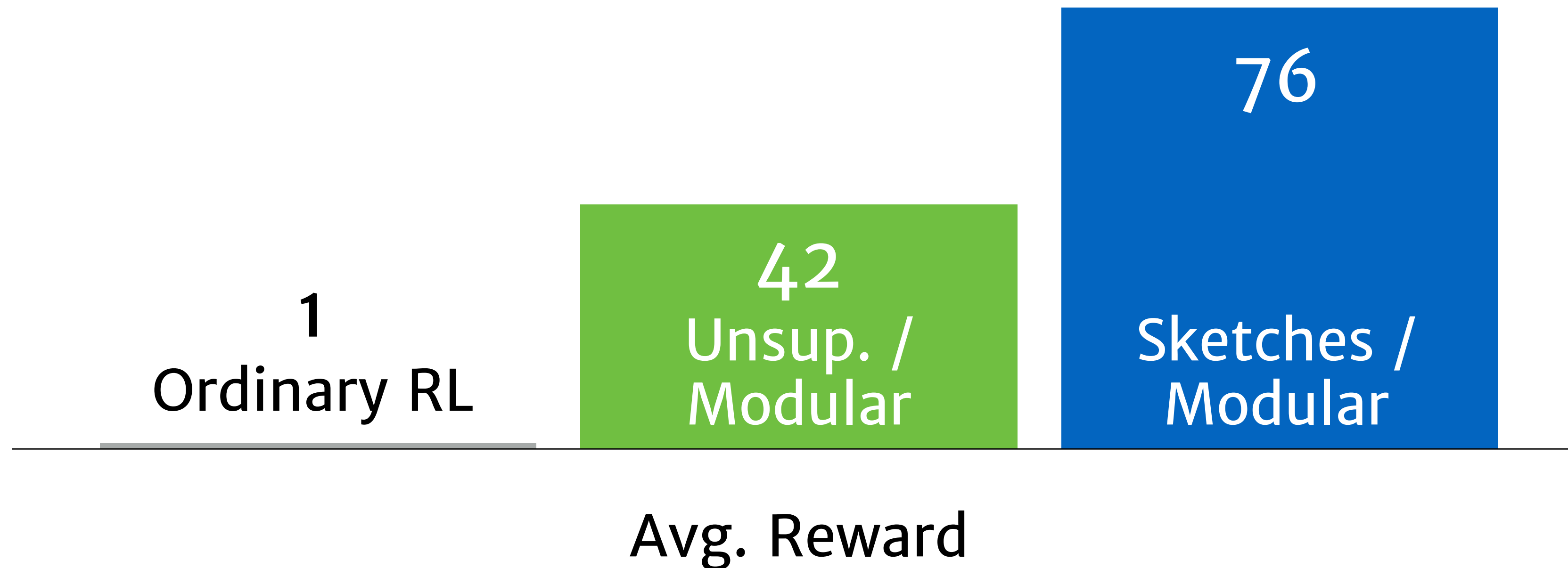


get iron use saw

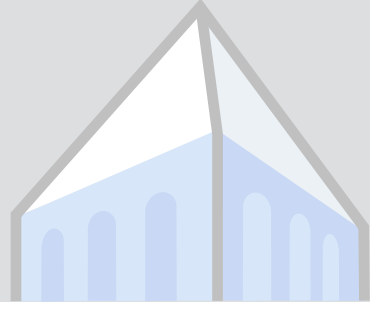


# Fast adaptation

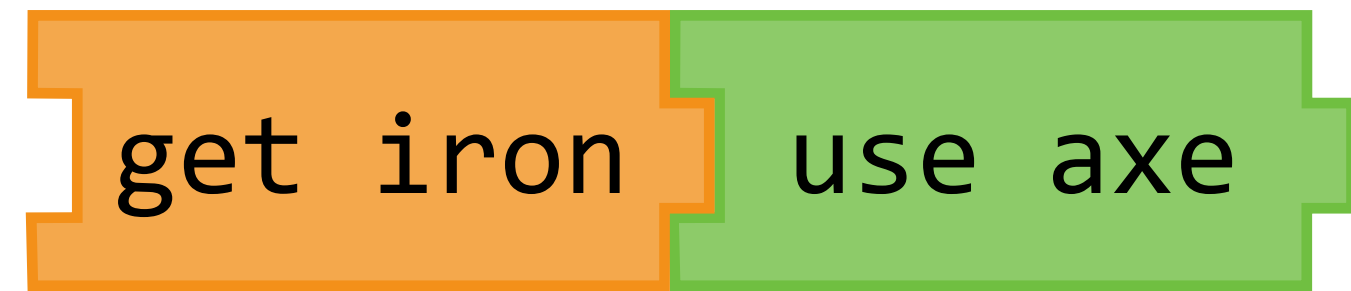
What if I don't get a sketch at test time?



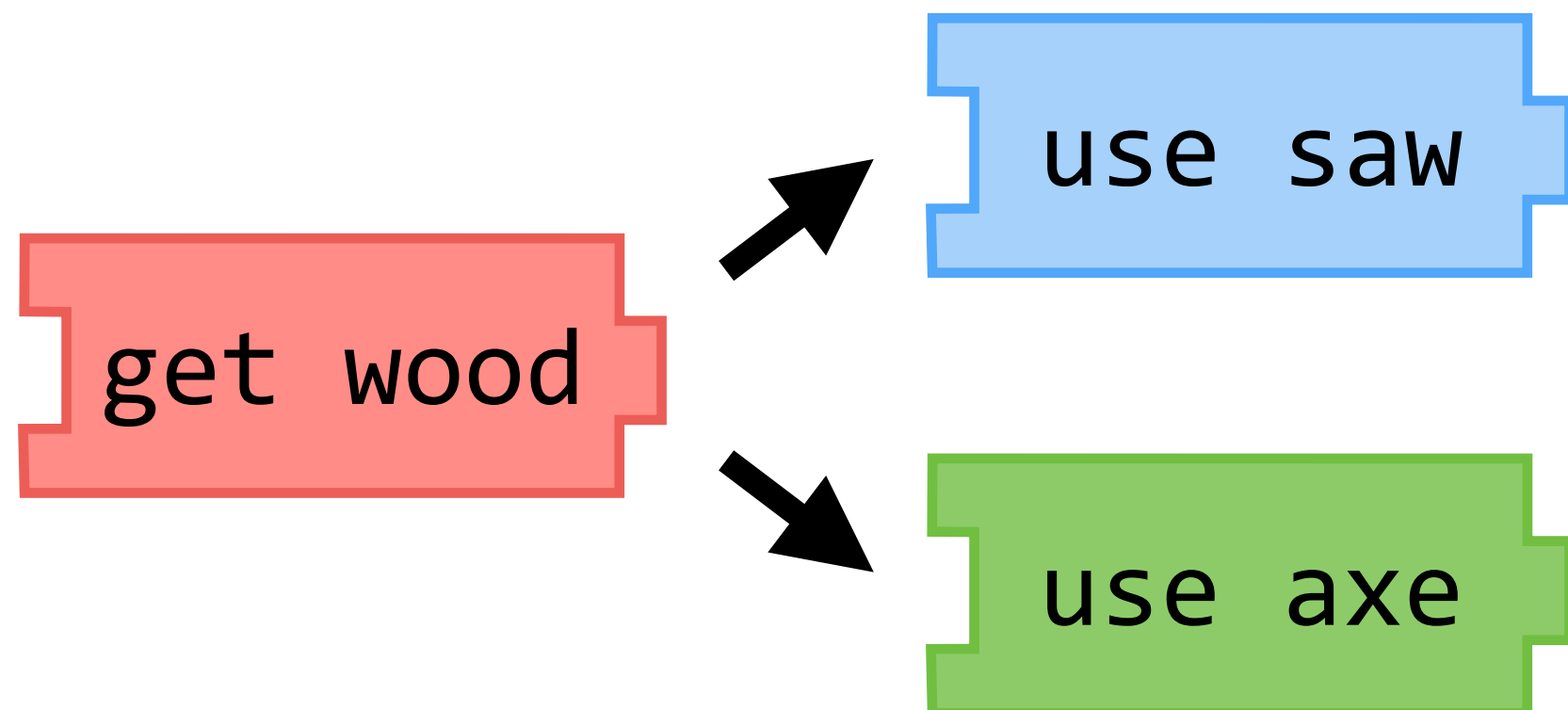




# Lessons



We can also learn modular behaviors from ungrounded “sketches” of abstract plans.



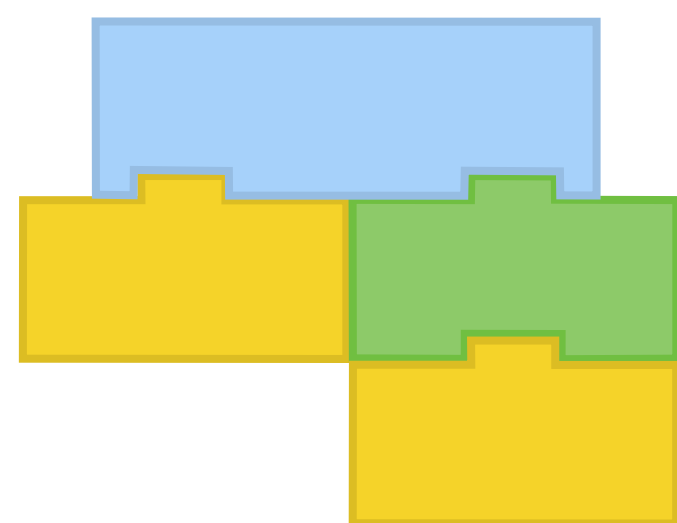
We can use these modules to help reinforcement learning even when sketches are not available.



# Beyond “tasks”

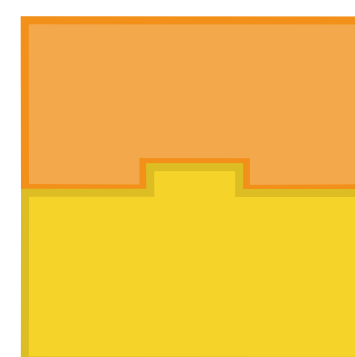
## LOCALIZATION

*Man in glasses near two men.*



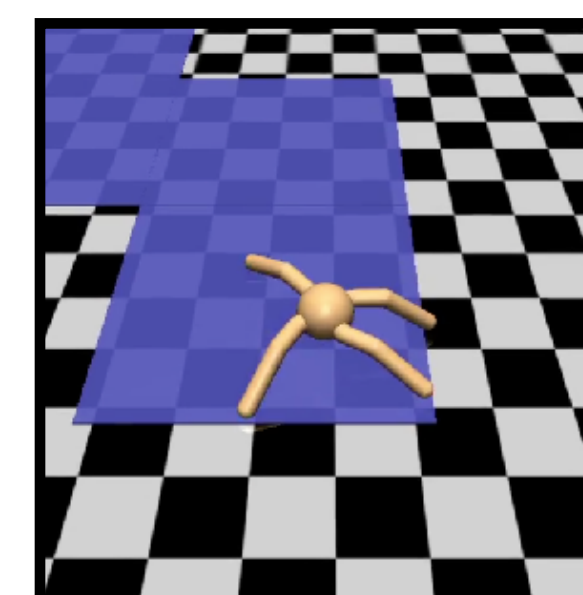
## Q&A

*How many men?*



## POLICY SEARCH

*go near the corner*





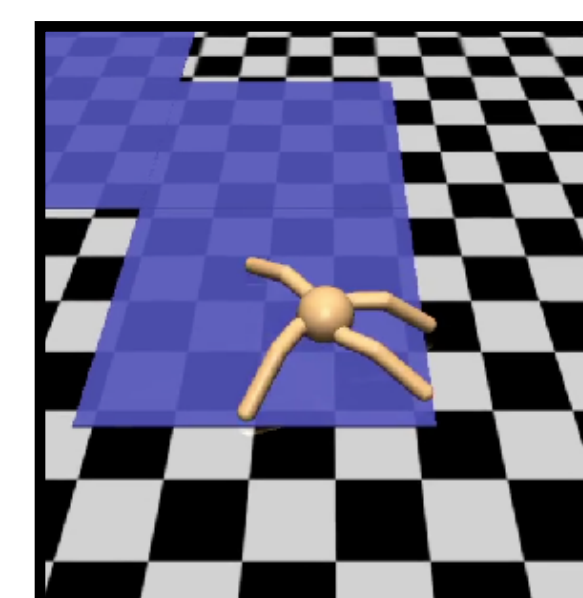
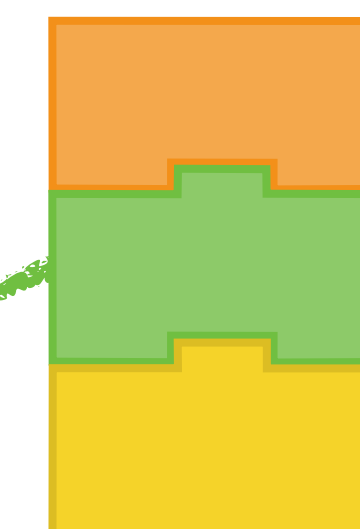
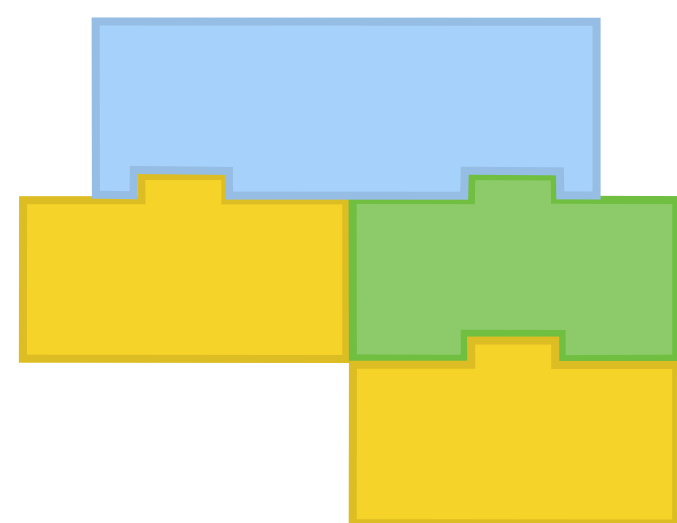
# Toward a model of everything

## LANGUAGE LEARNING

*Man in glasses  
near two men.*

*How many  
men?*

*go near the  
corner*



REASONING  
LEARNING

# LANGUAGE & BELIEF

**A** & Klein. *Reasoning about Pragmatics with Neural Listeners and Speakers*. EMNLP 16.

**A**, Drăgan & Klein. *Translating Neuralese*. ACL 17.

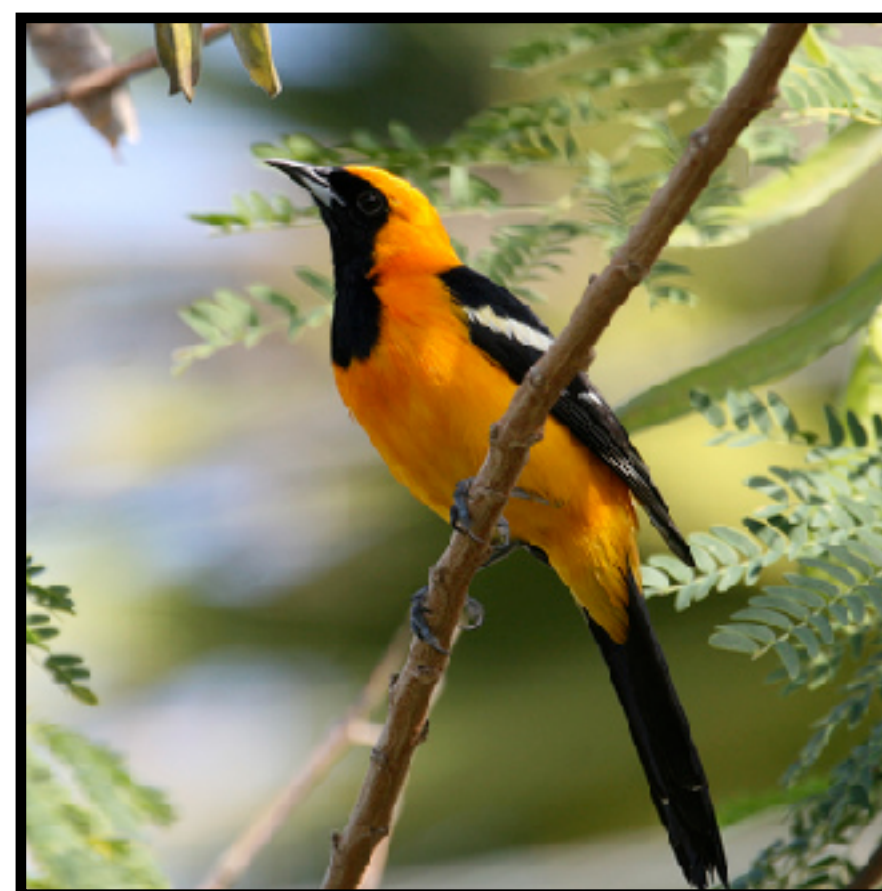
**A** & Klein. *Analogs of Linguistic Structure in Deep Representations*. EMNLP 17.

Fried, **A** & Klein. *Unified Pragmatic Models for Generating and Following [...]*. NAACL 18.

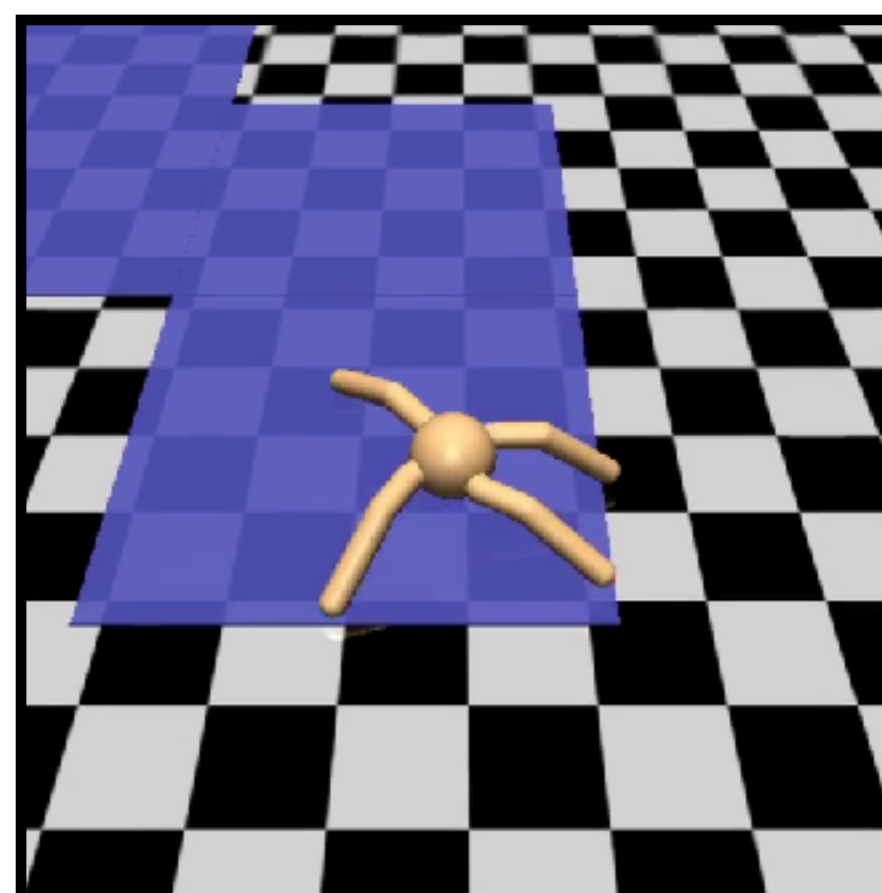


# Interpreting language

*What kind of bird is this?*



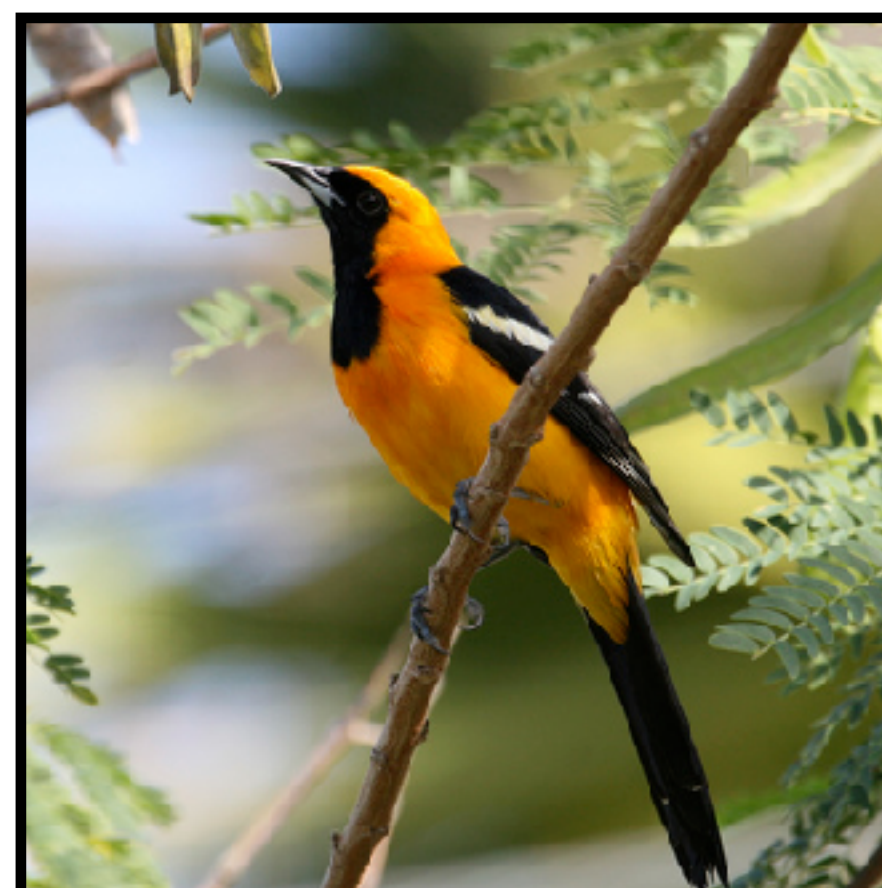
*What are you going to do?*





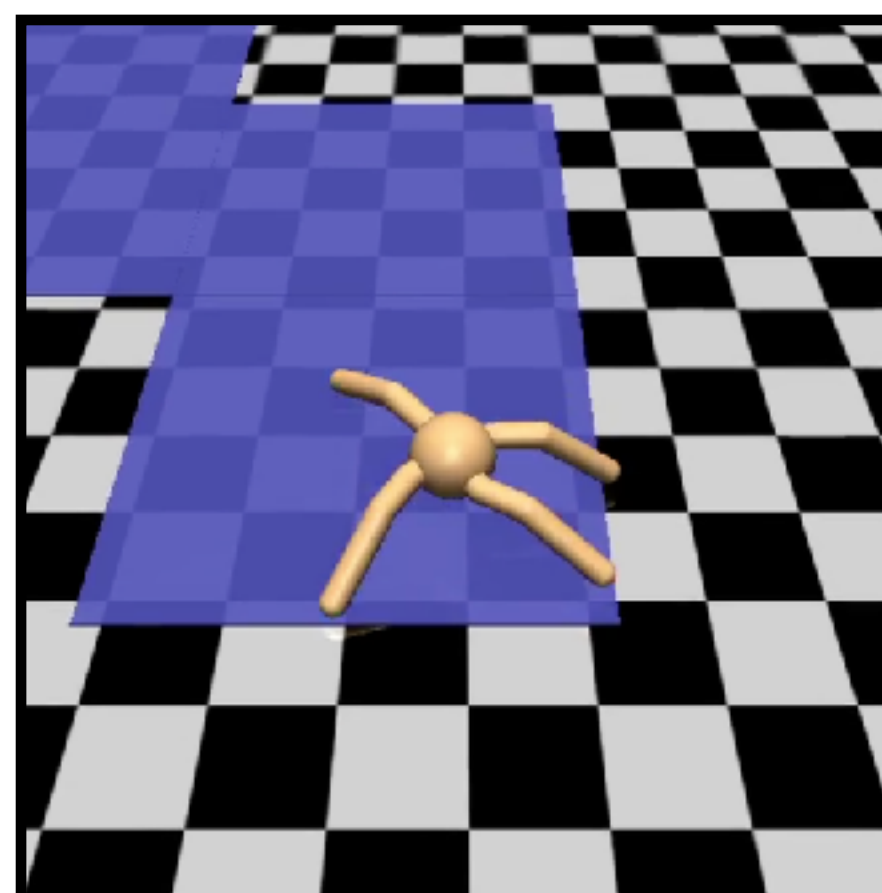
# Generating “language”

*What kind of  
bird is this?*



oriole

*What are you  
going to do?*

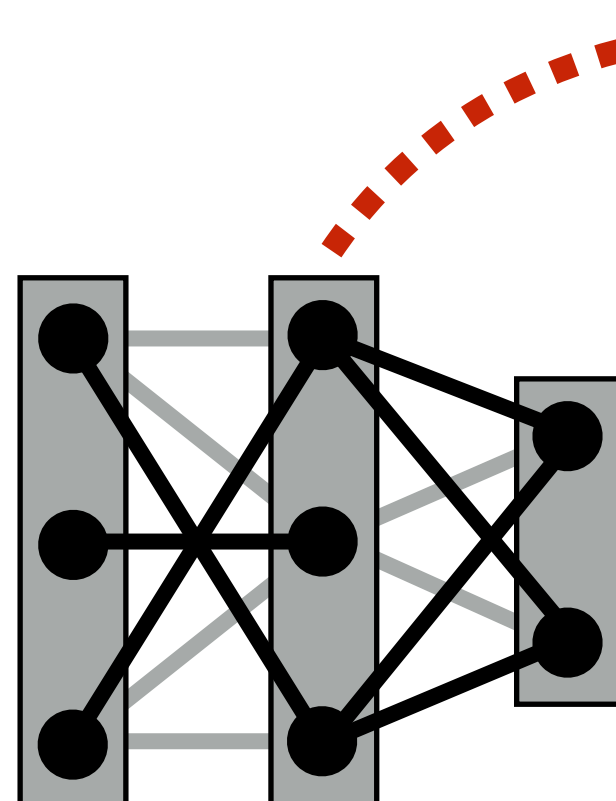


[GO NORTH, GO WEST]



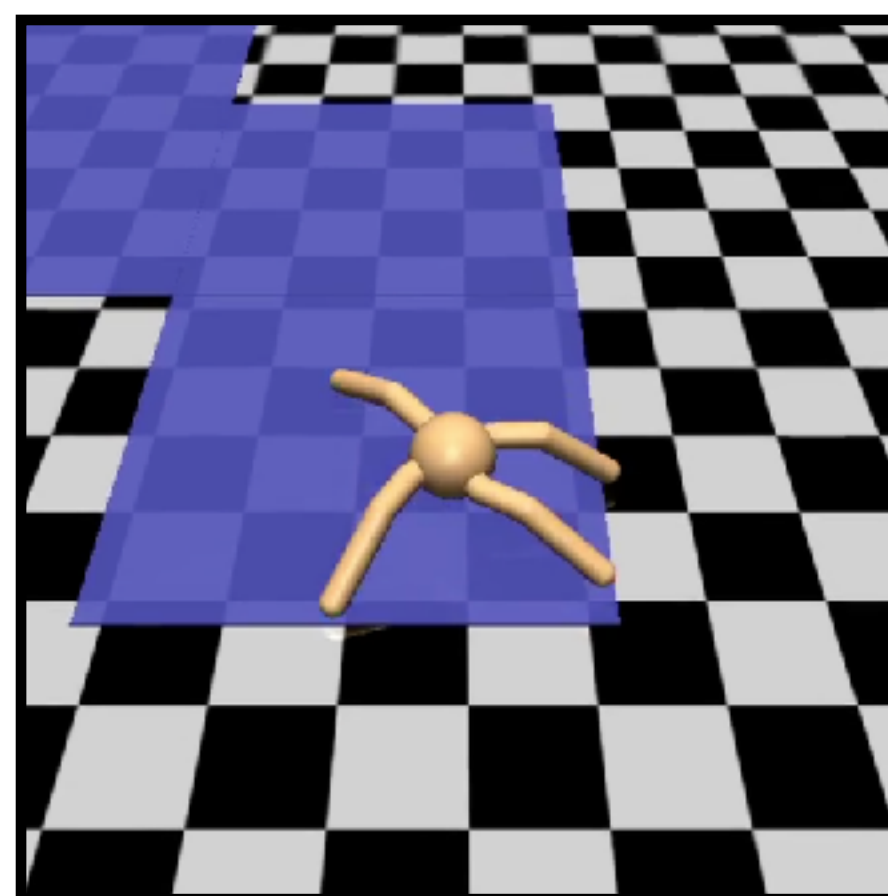
# Generating “language”

*What kind of  
bird is this?*



???

*What are you  
going to do?*

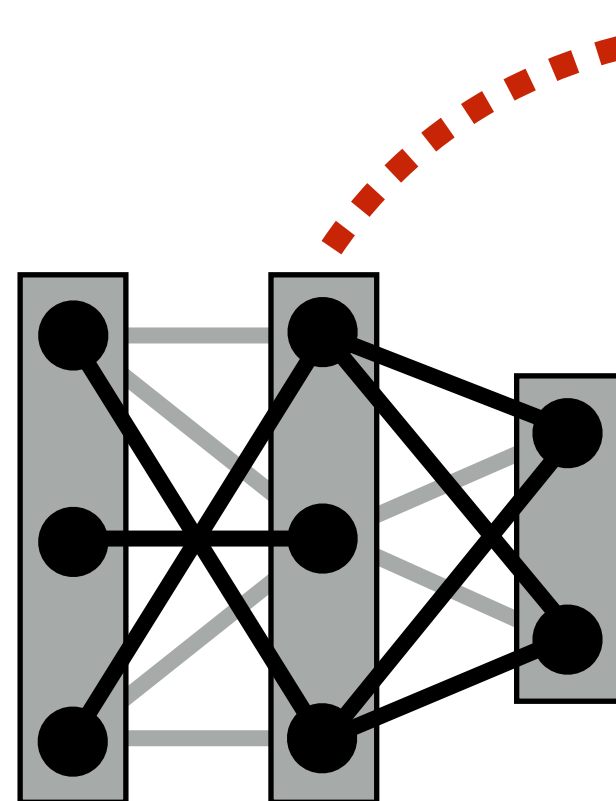


[GO NORTH, GO WEST]



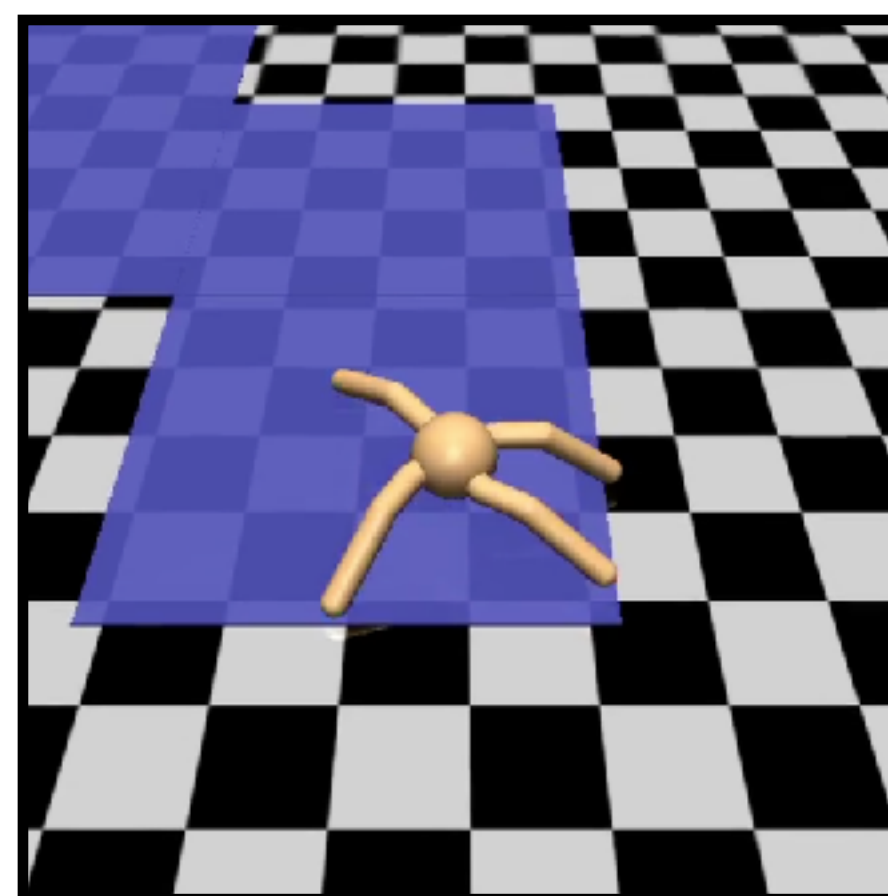
# Generating informative language

*What kind of bird is this?*



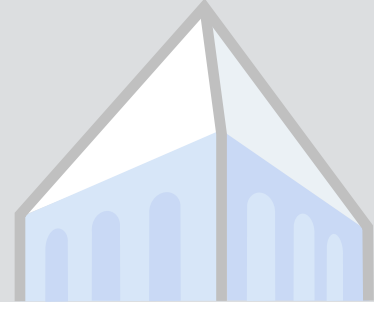
*large bird, black wings*

*What are you going to do?*

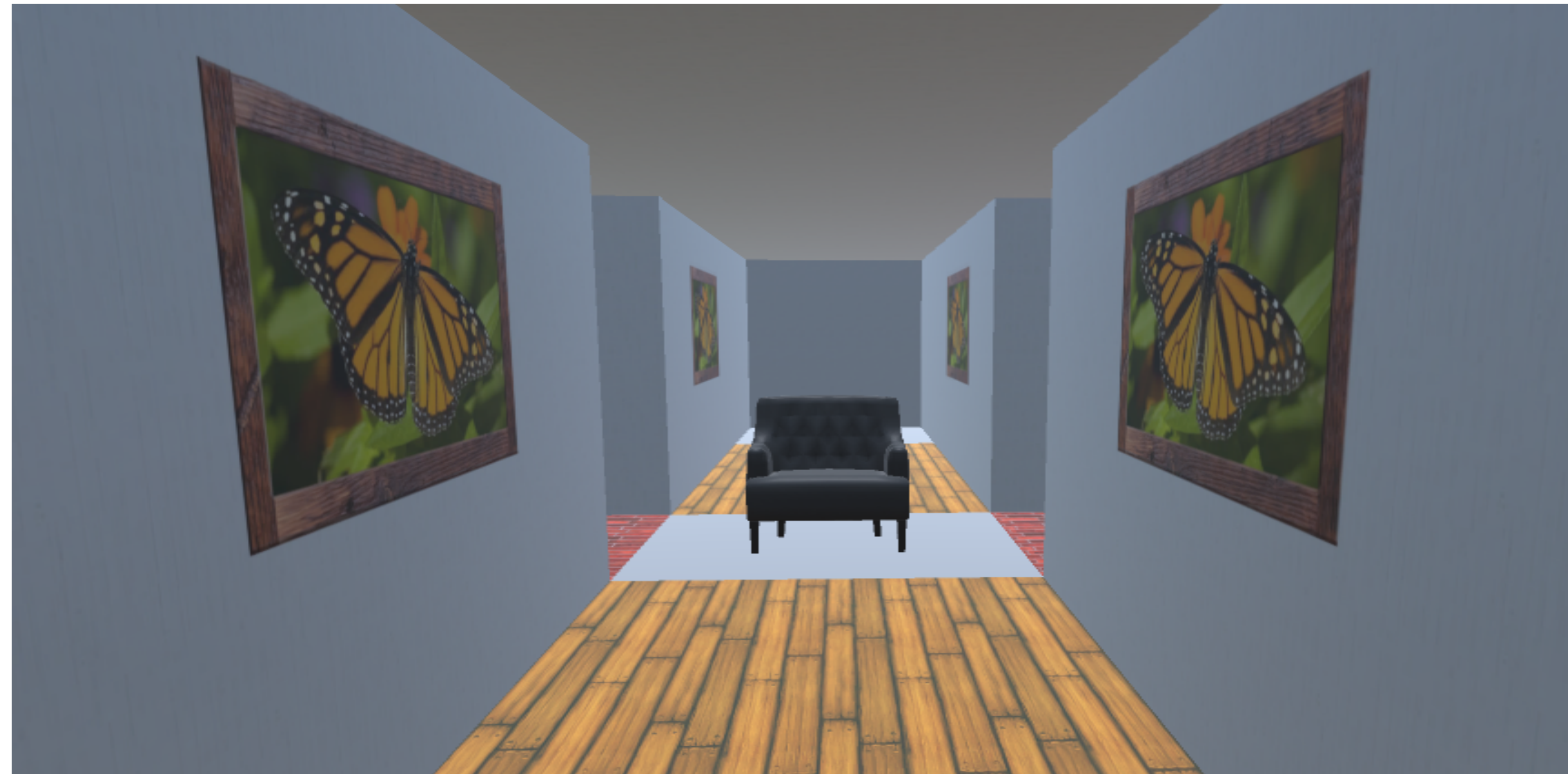


*Reach the end of the blue path.*



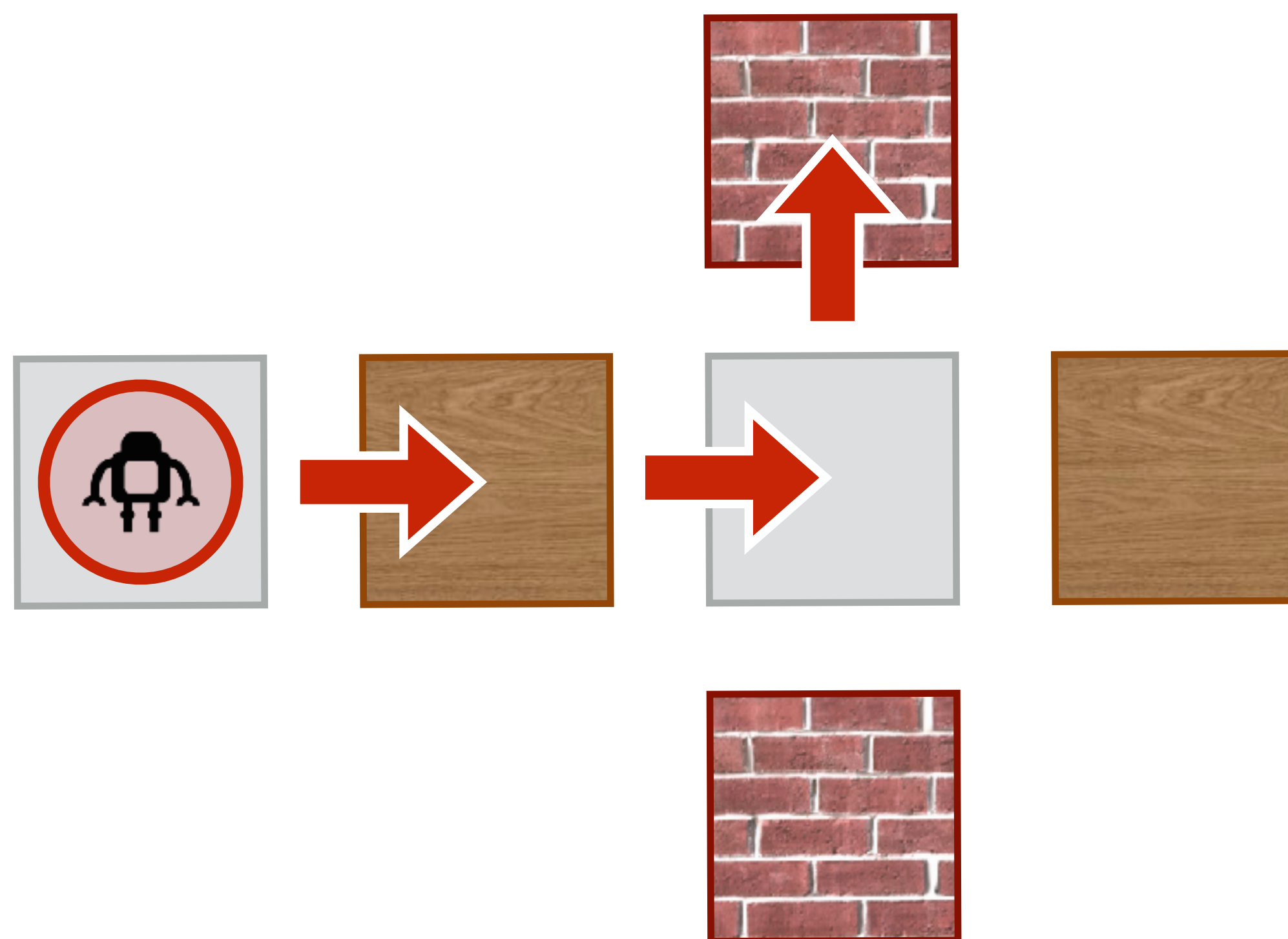


# Explaining behaviors



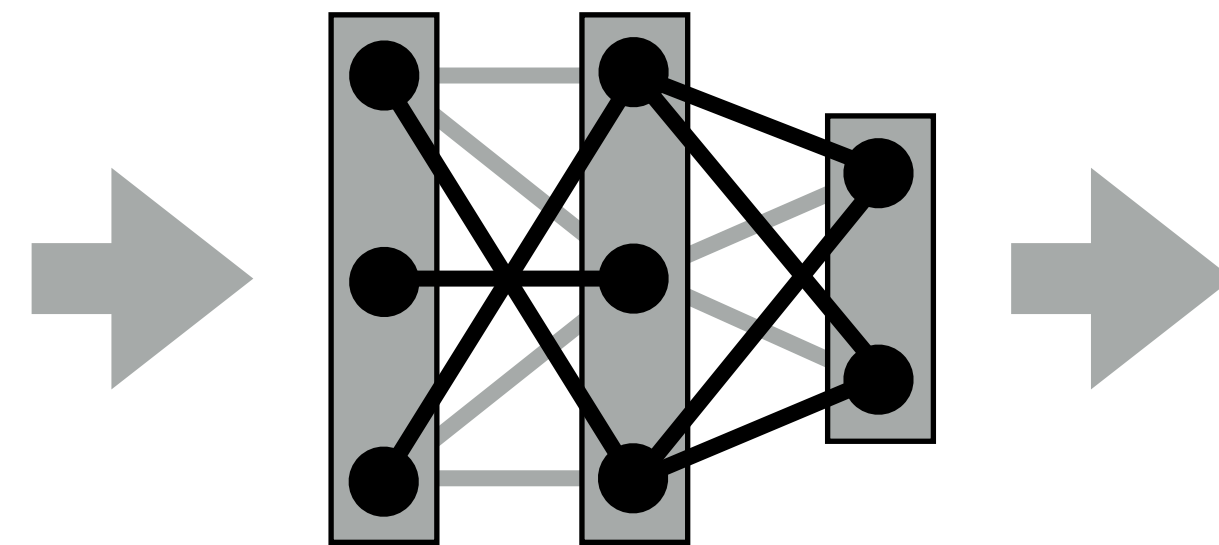


# Explaining behaviors





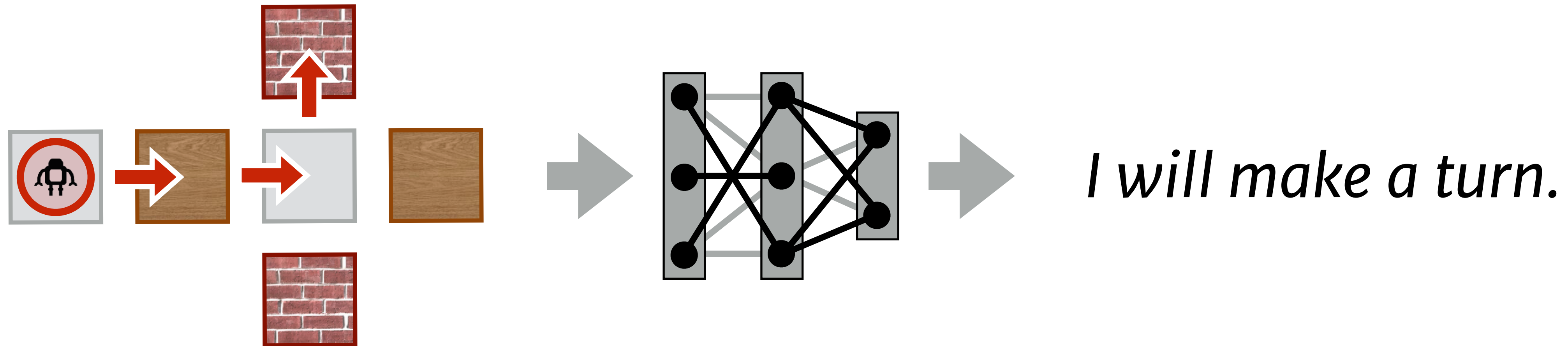
# Image captioning



*A group of young men  
playing a game of soccer.*

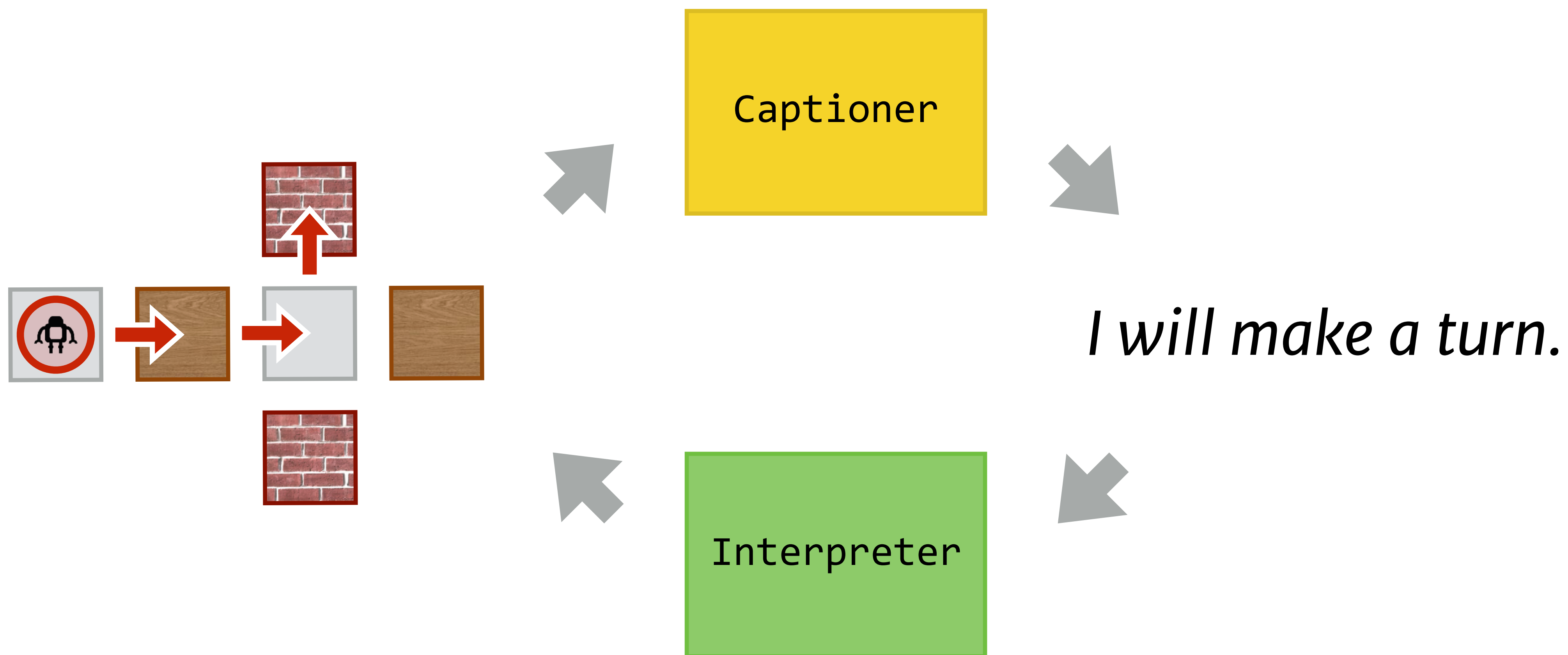


# Image captioning



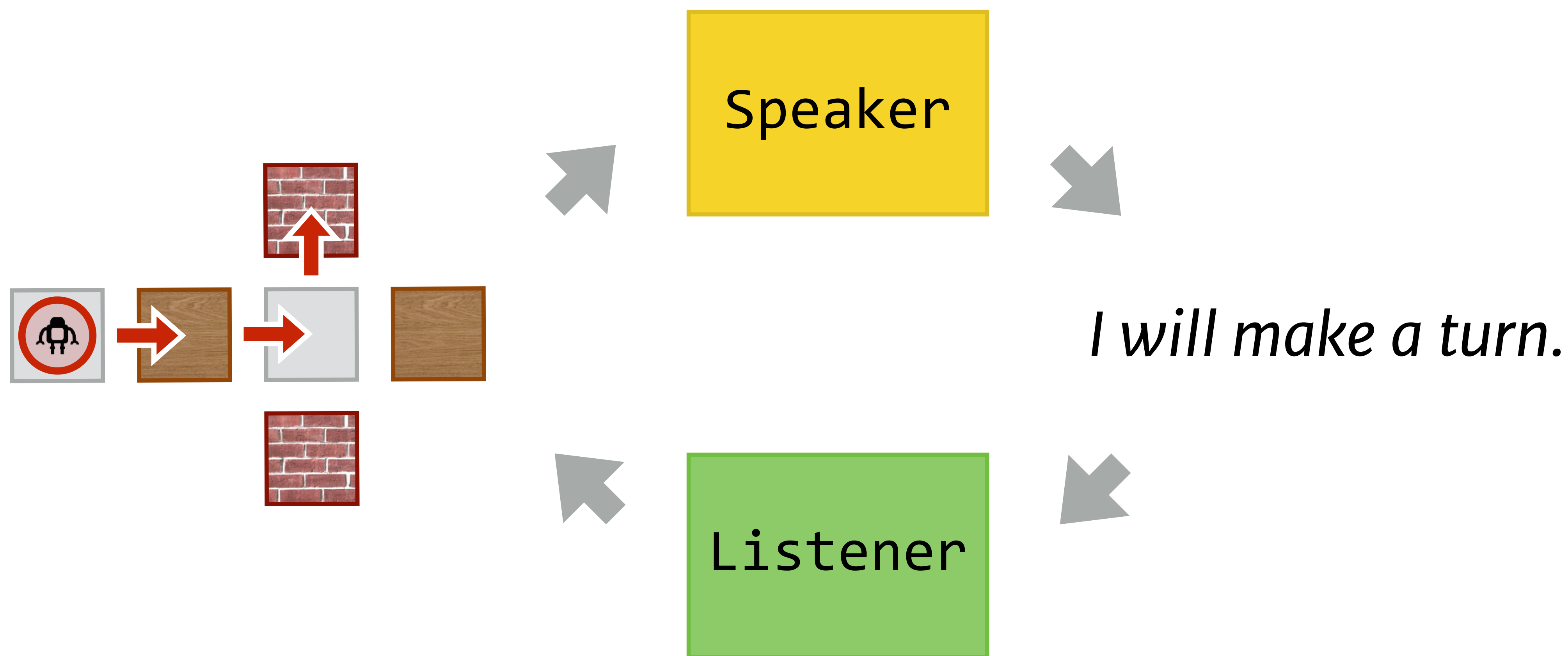


# Reasoning about outcomes



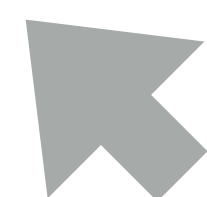
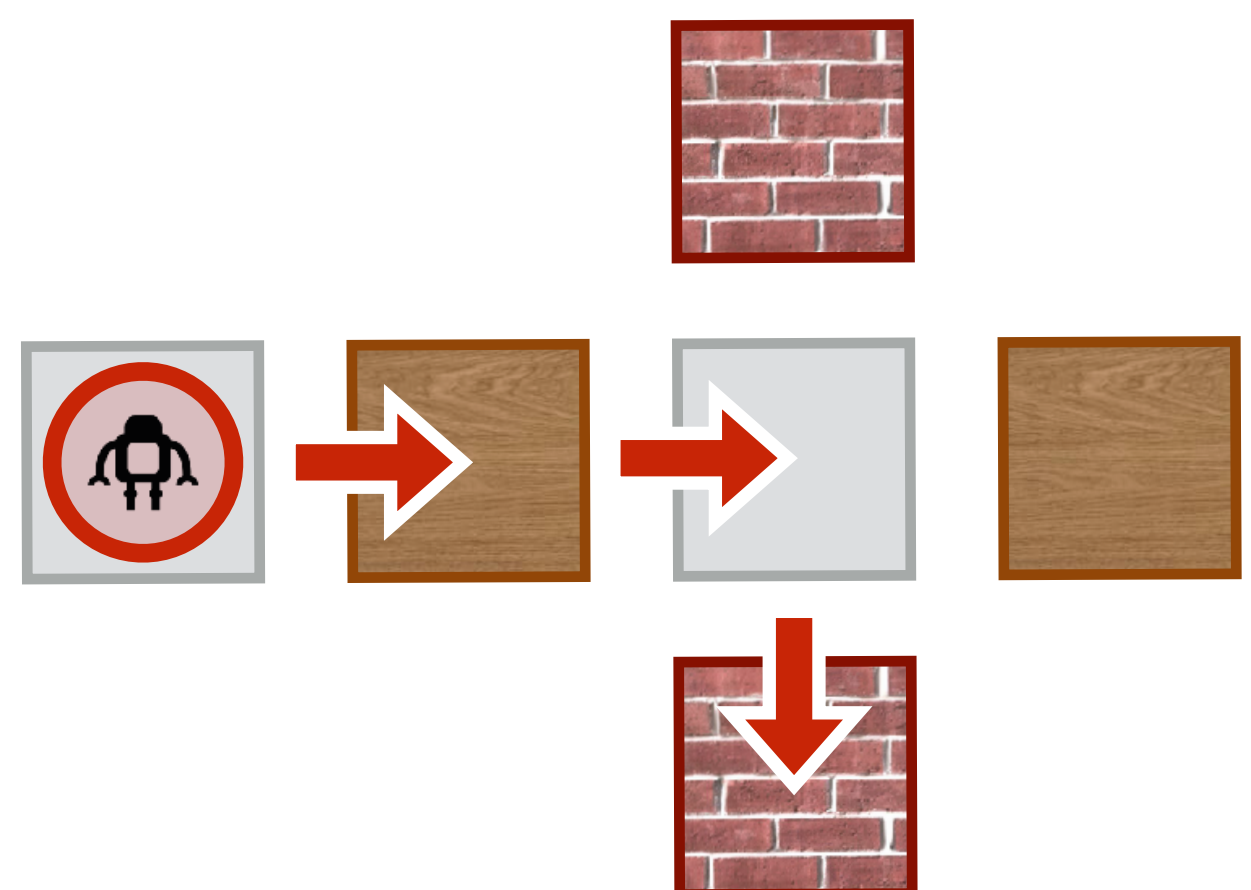


# Reasoning about outcomes





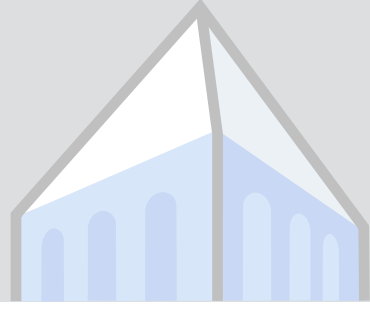
# Reasoning about outcomes



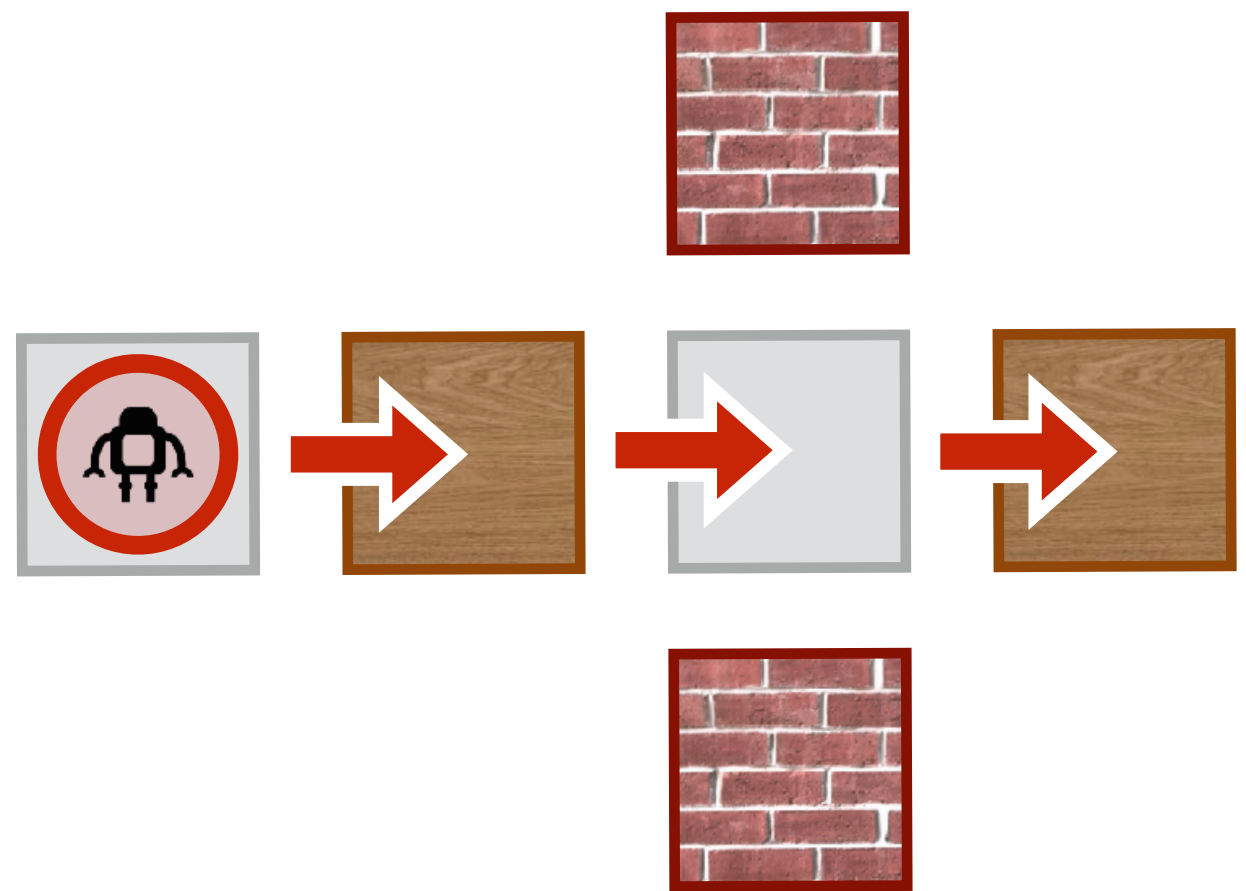
Listener



*I will make a turn.*



# Reasoning about outcomes



Listener

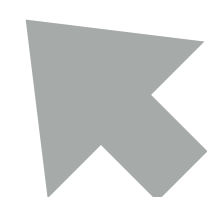
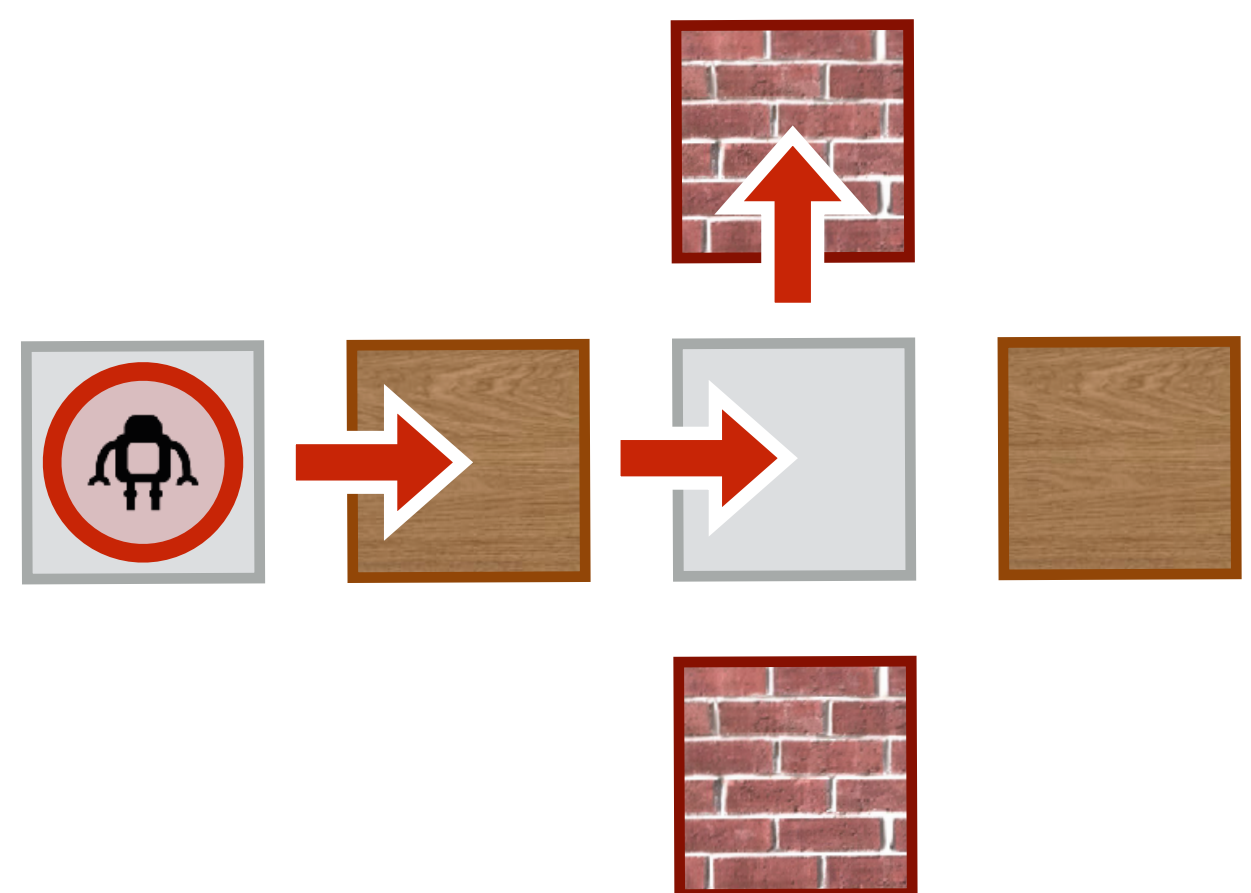


*I will go straight through.*





# Reasoning about outcomes



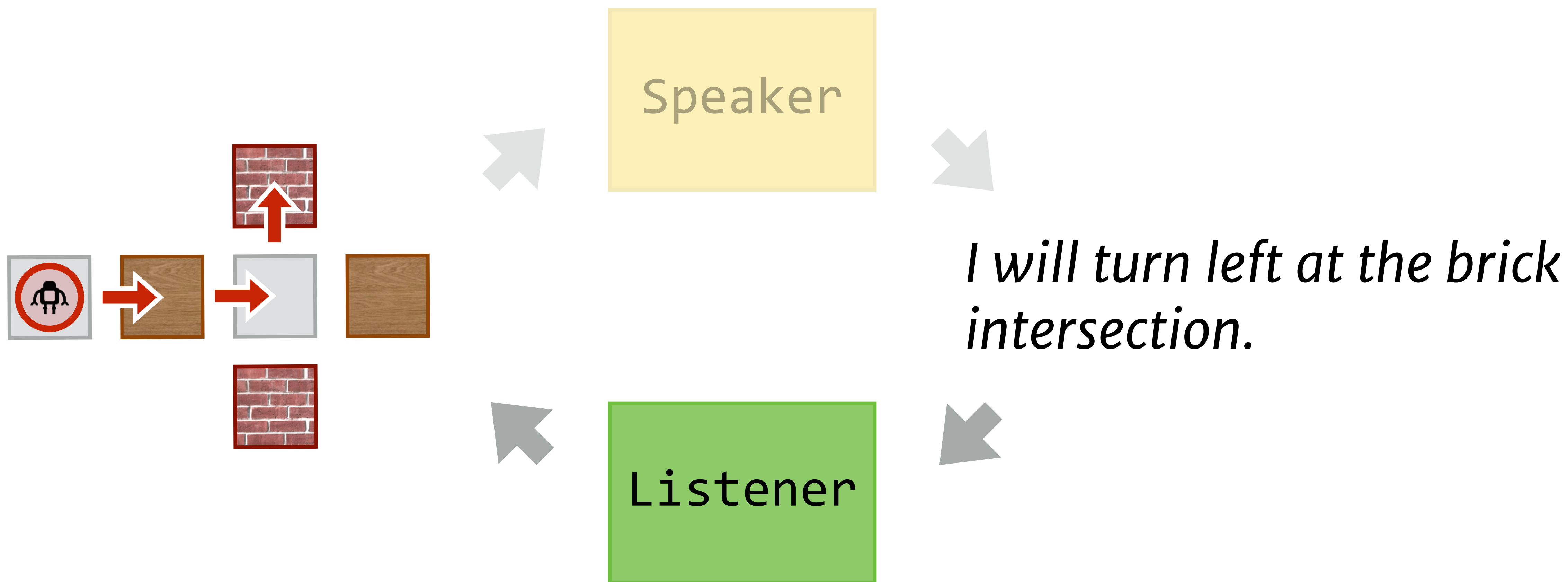
Listener



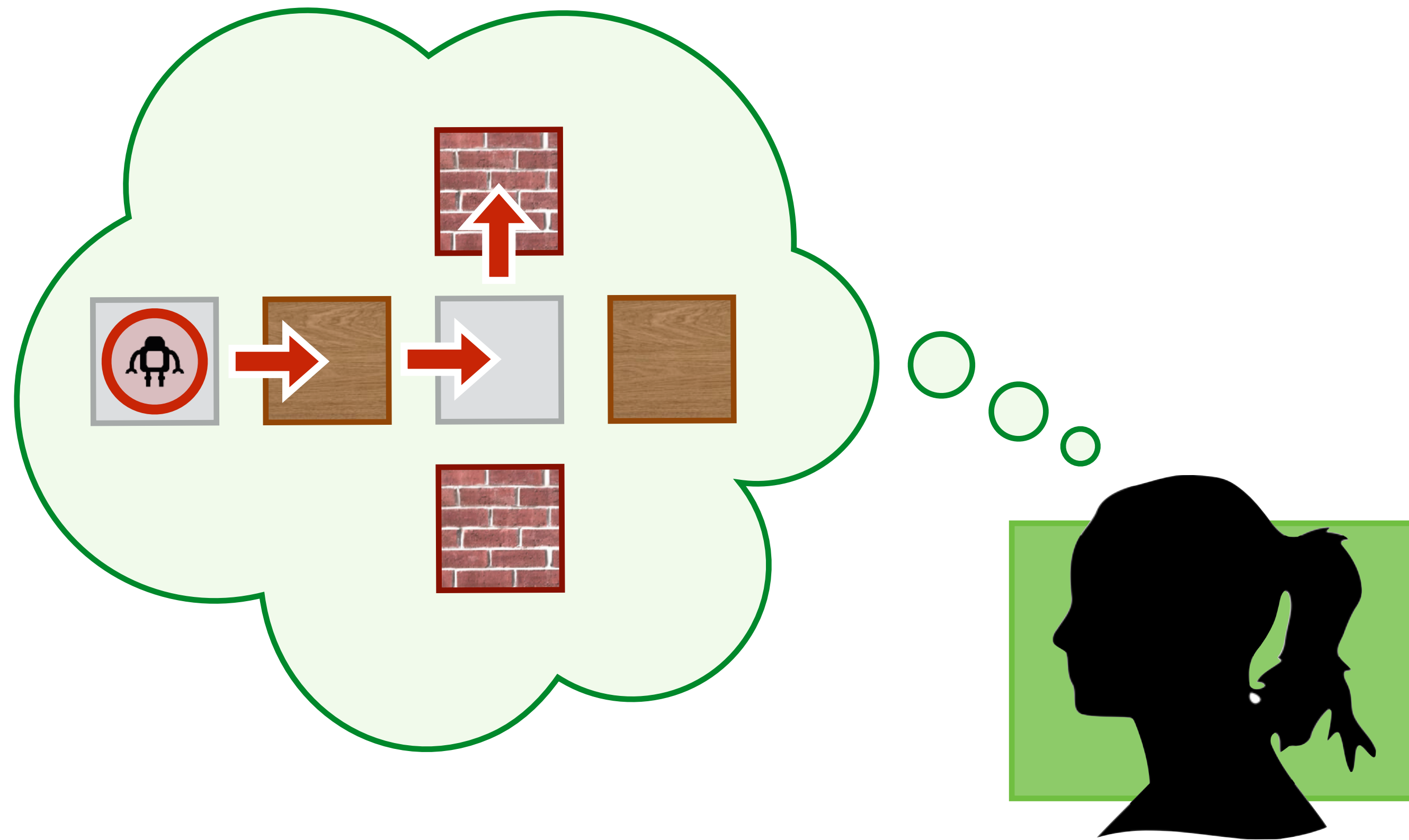
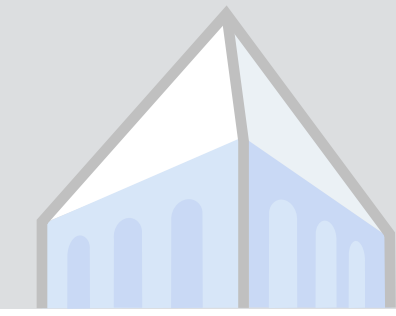
*I will turn left at the brick intersection.*



# Reasoning about outcomes



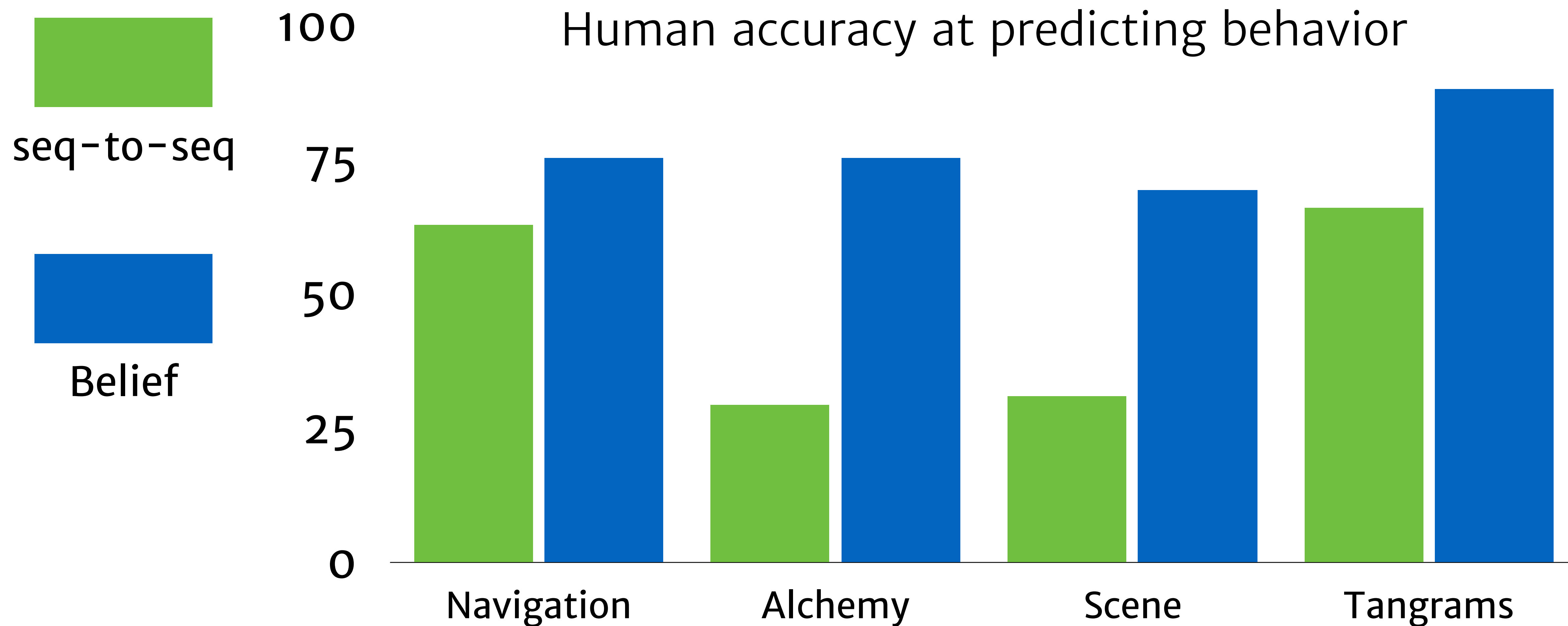
# Reasoning about belief



*I will turn left at the brick intersection.*

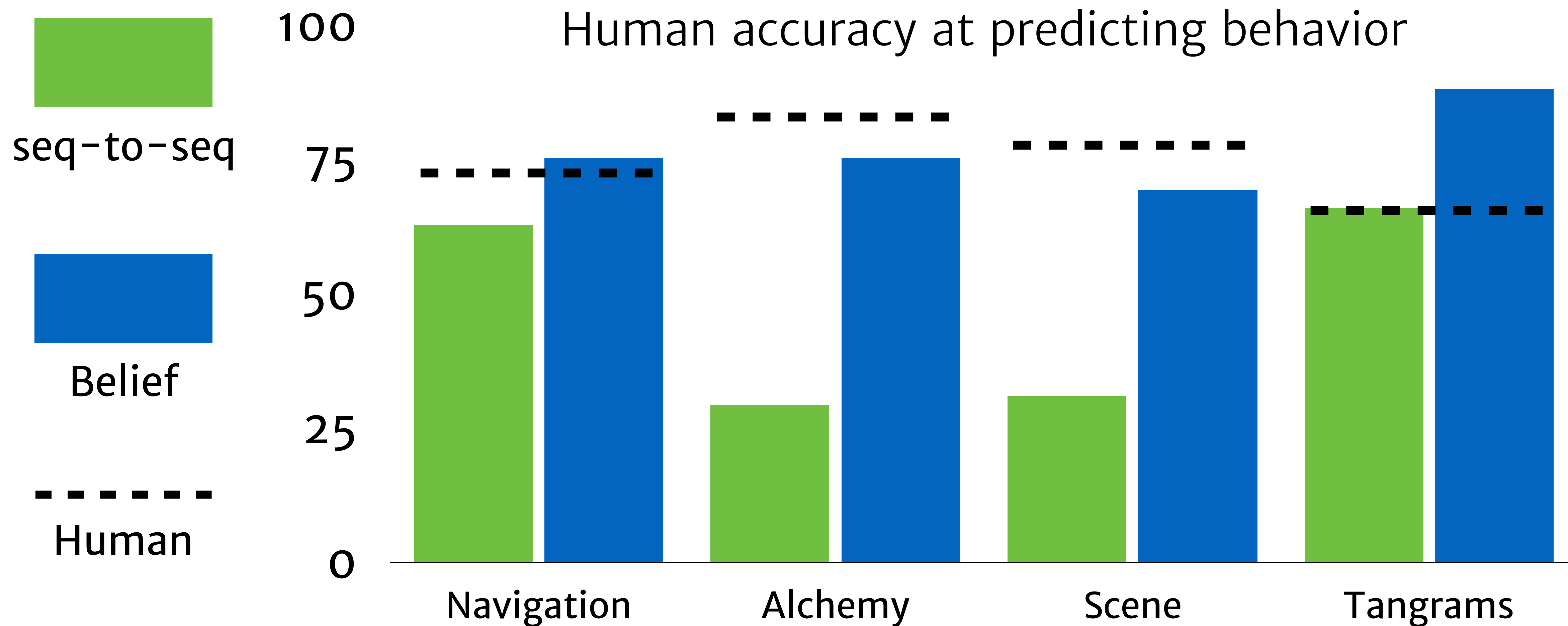


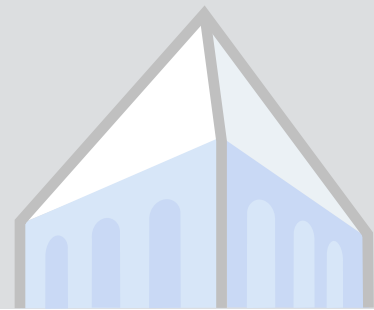
# Experimental results [FAK18]



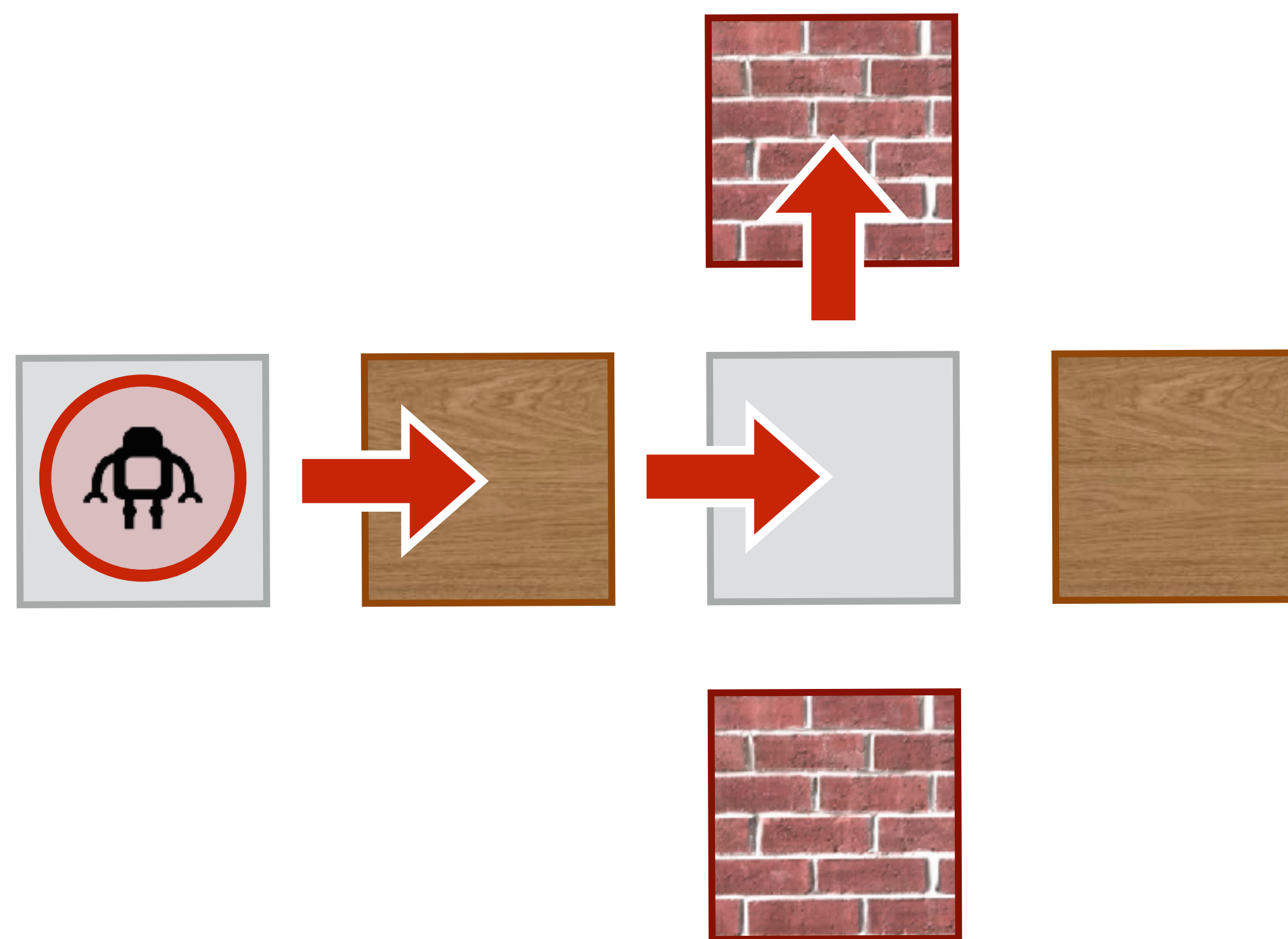


# Experimental results [FAK18]



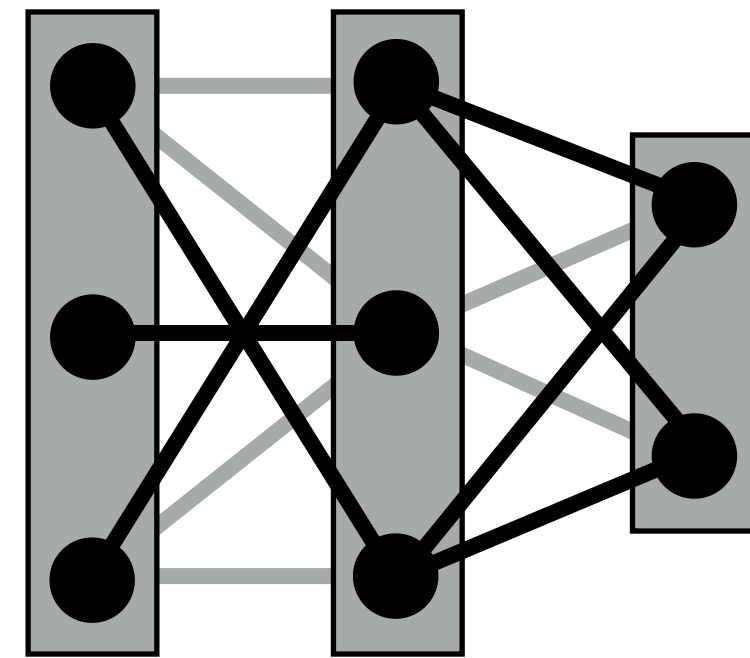
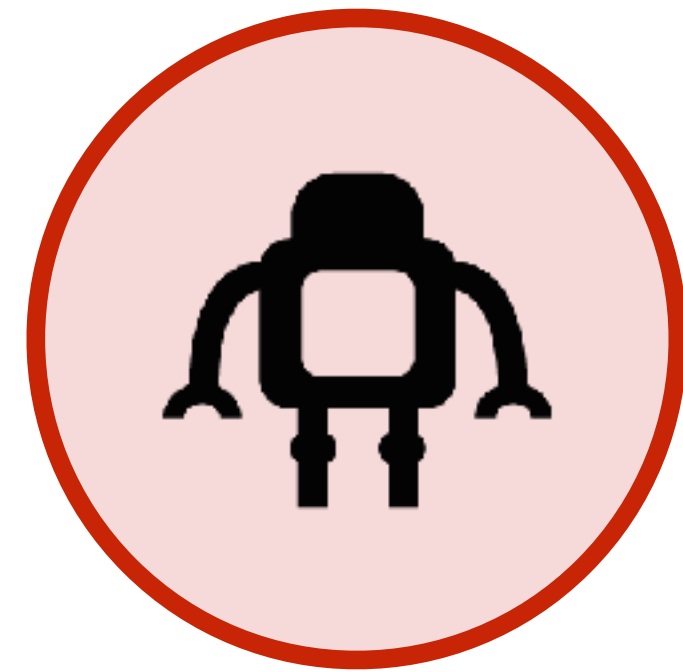


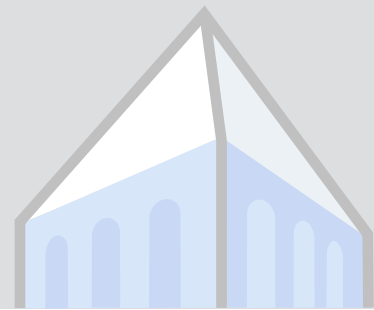
# Explaining behaviors



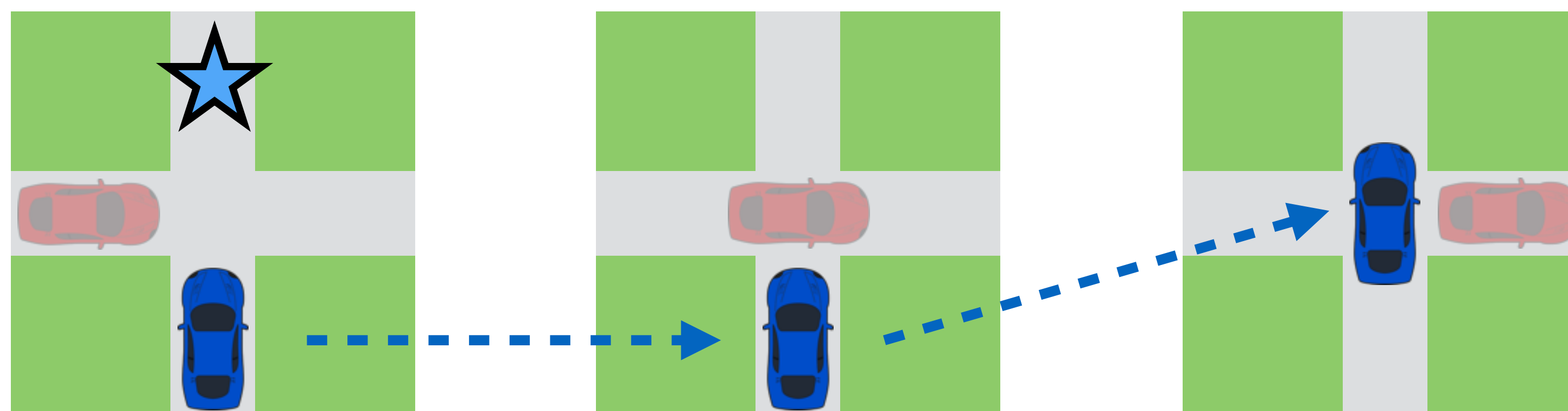
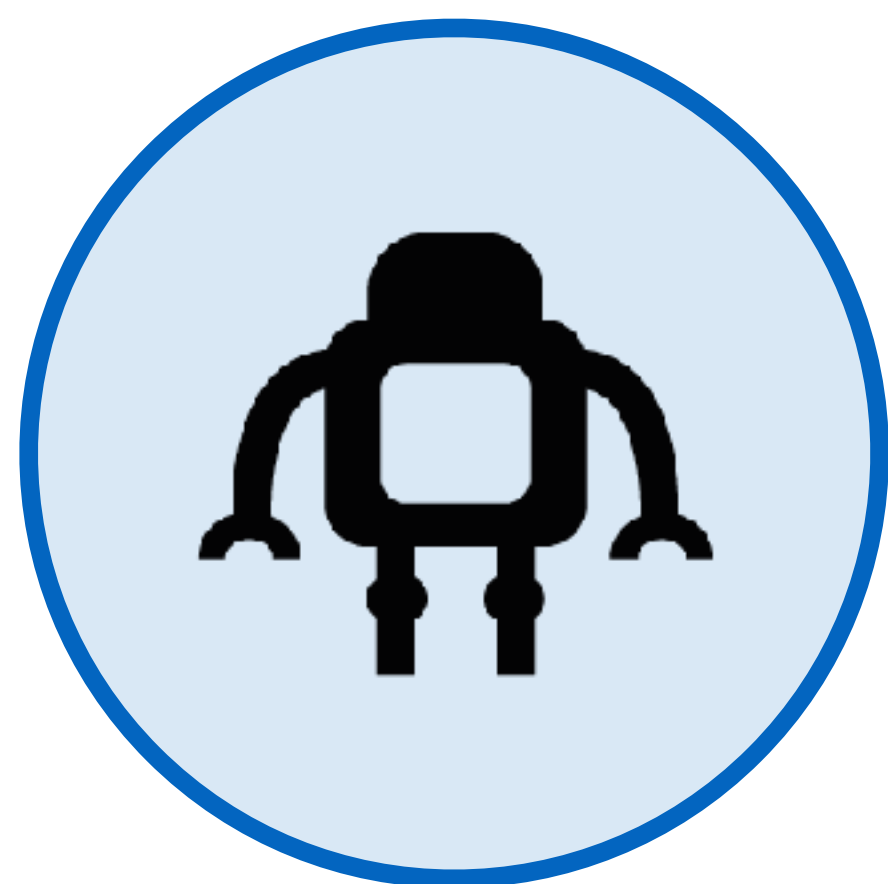
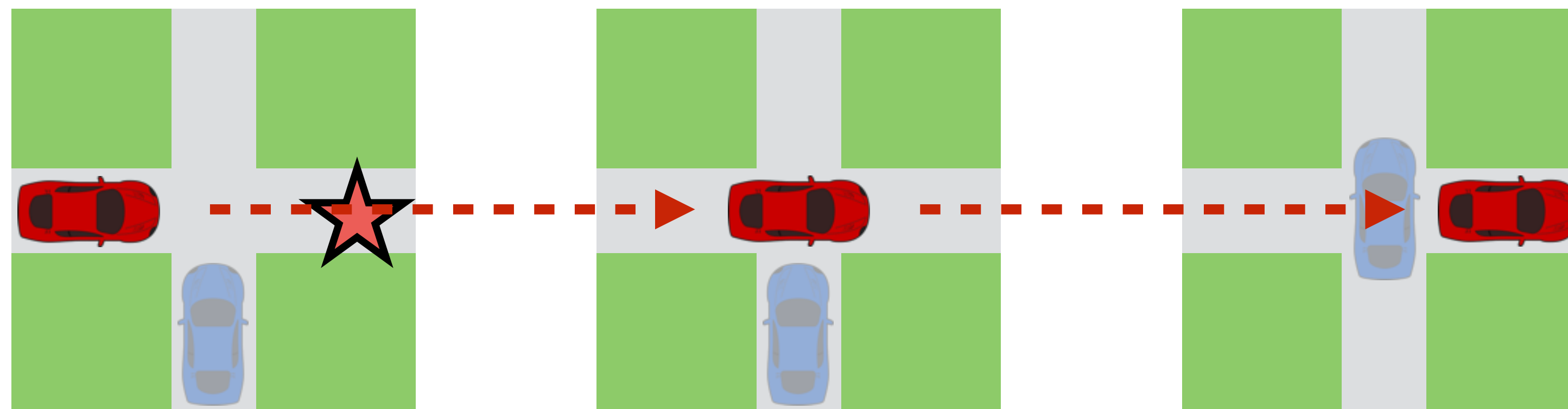
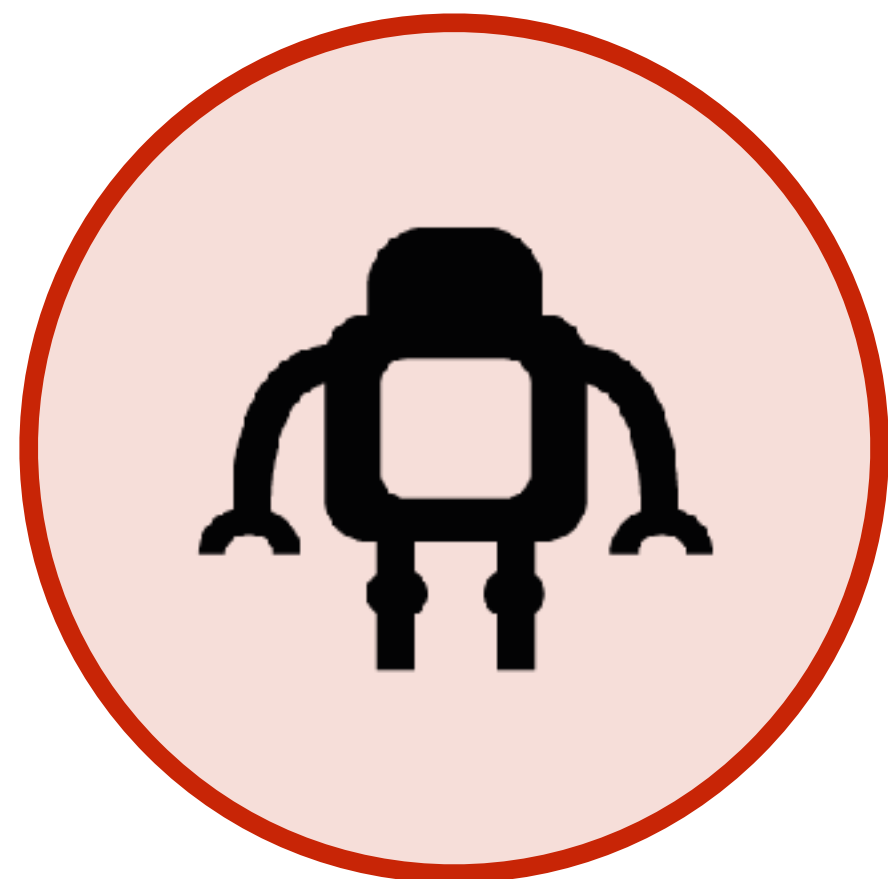


# Explaining models





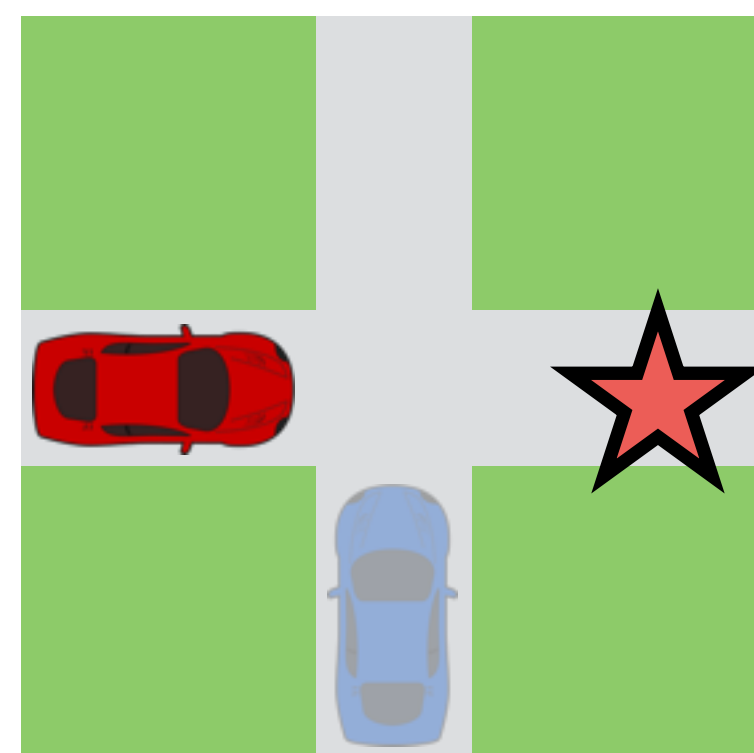
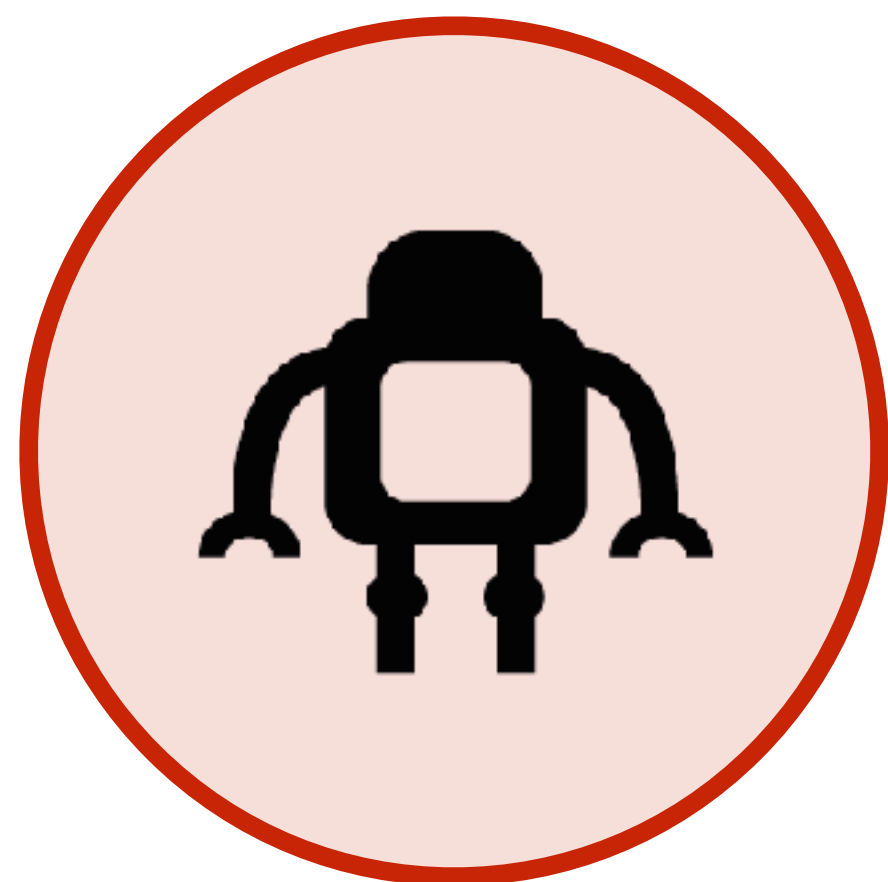
# Multi-agent communication



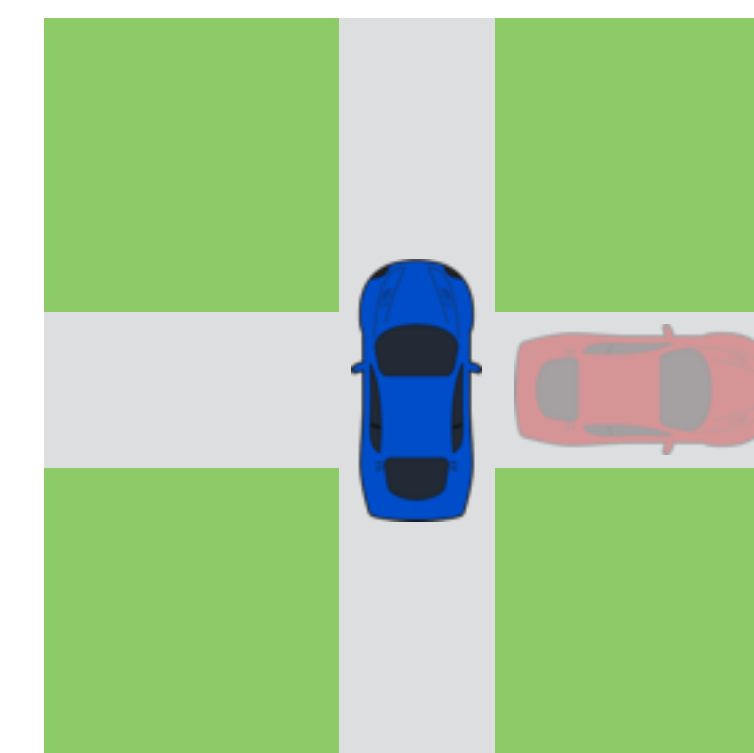
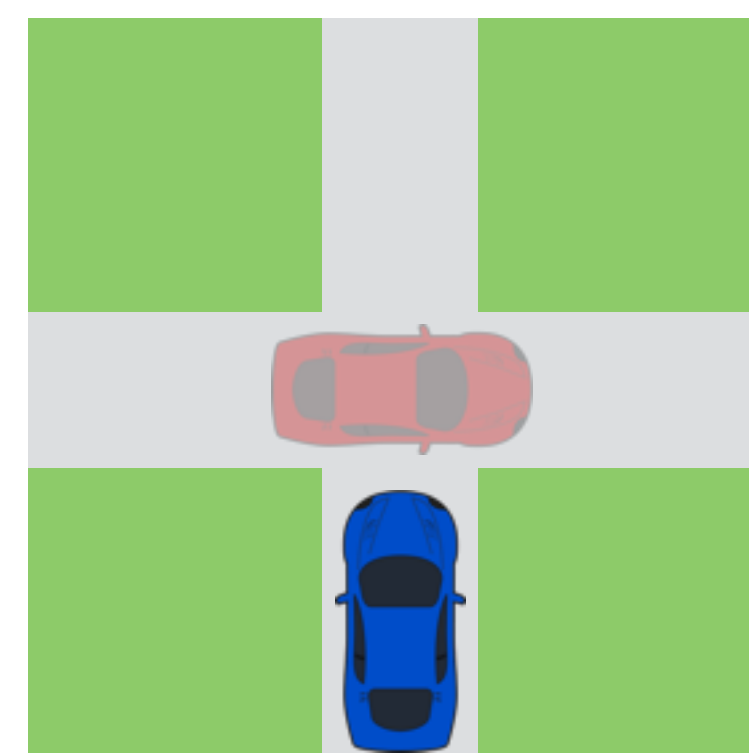
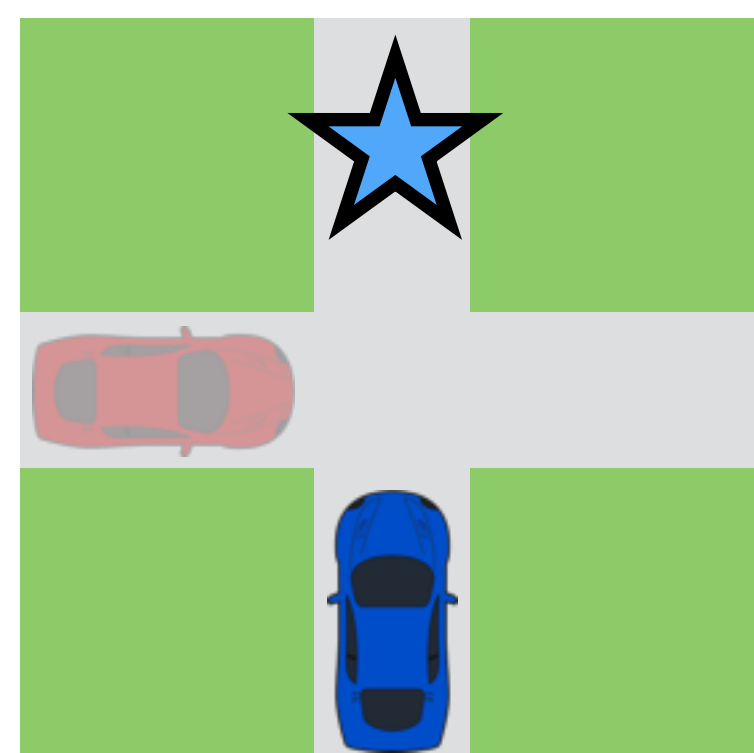
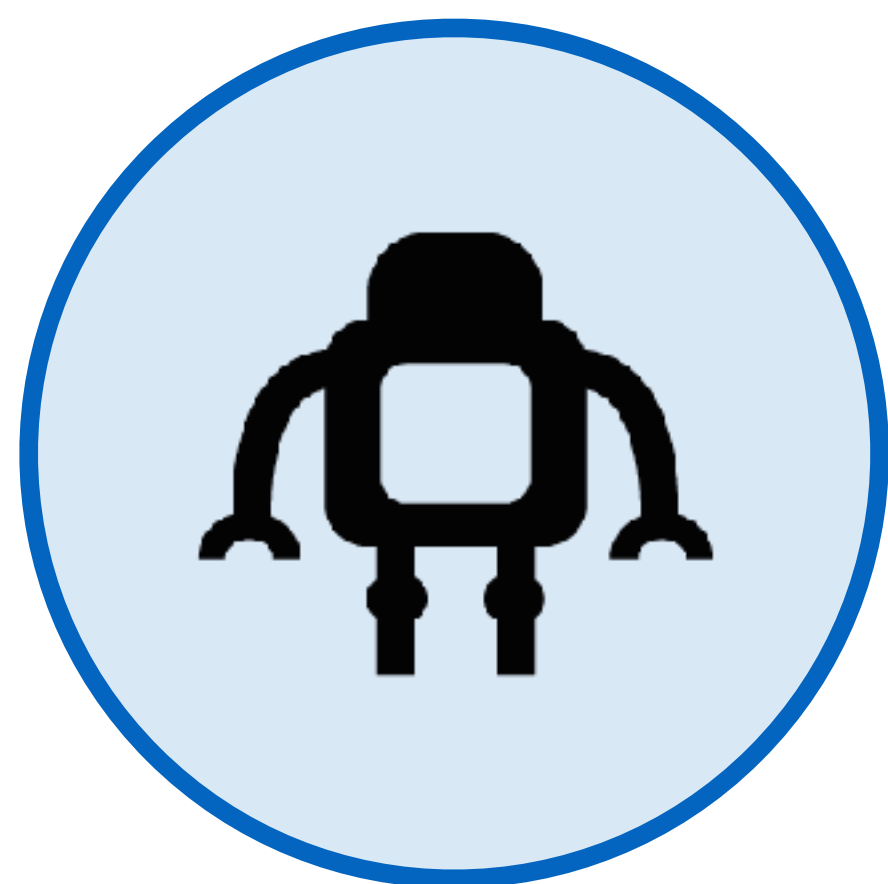
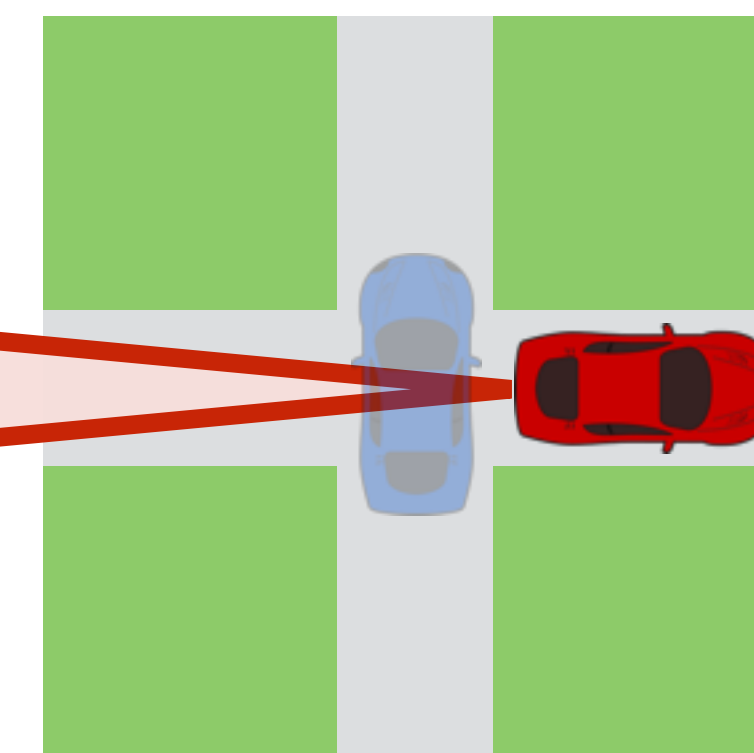


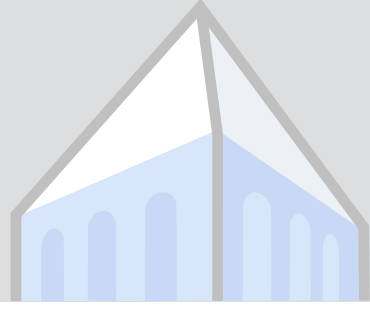


# Neuralese

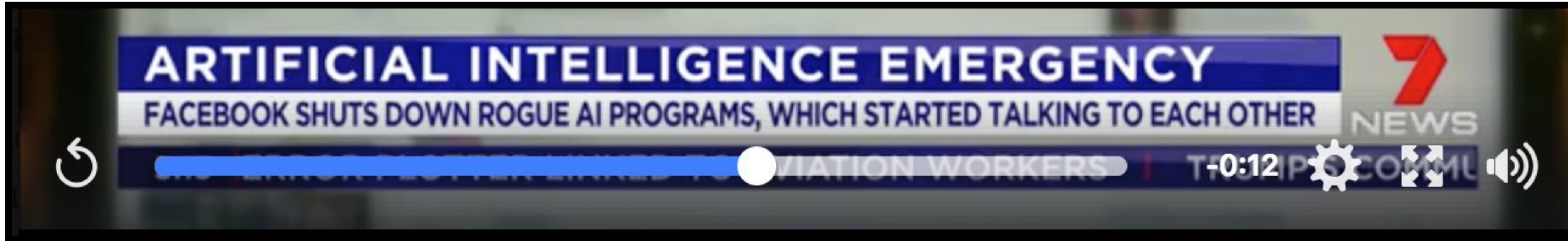


1.0	2.3
-0.3	0.4
-1.2	1.1





# Communication and behavior



## **ROBOSTOP** Facebook shuts off AI experiment after two robots begin speaking in their OWN language only they can understand

Experts have called the incident exciting but also incredibly scary

By James Beal and Andy Jehring  
1st August 2017, 12:03 am | Updated: 2nd August 2017, 4:56 am

## Elon Musk's lab forced bots to create their own language

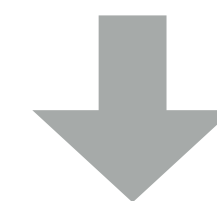
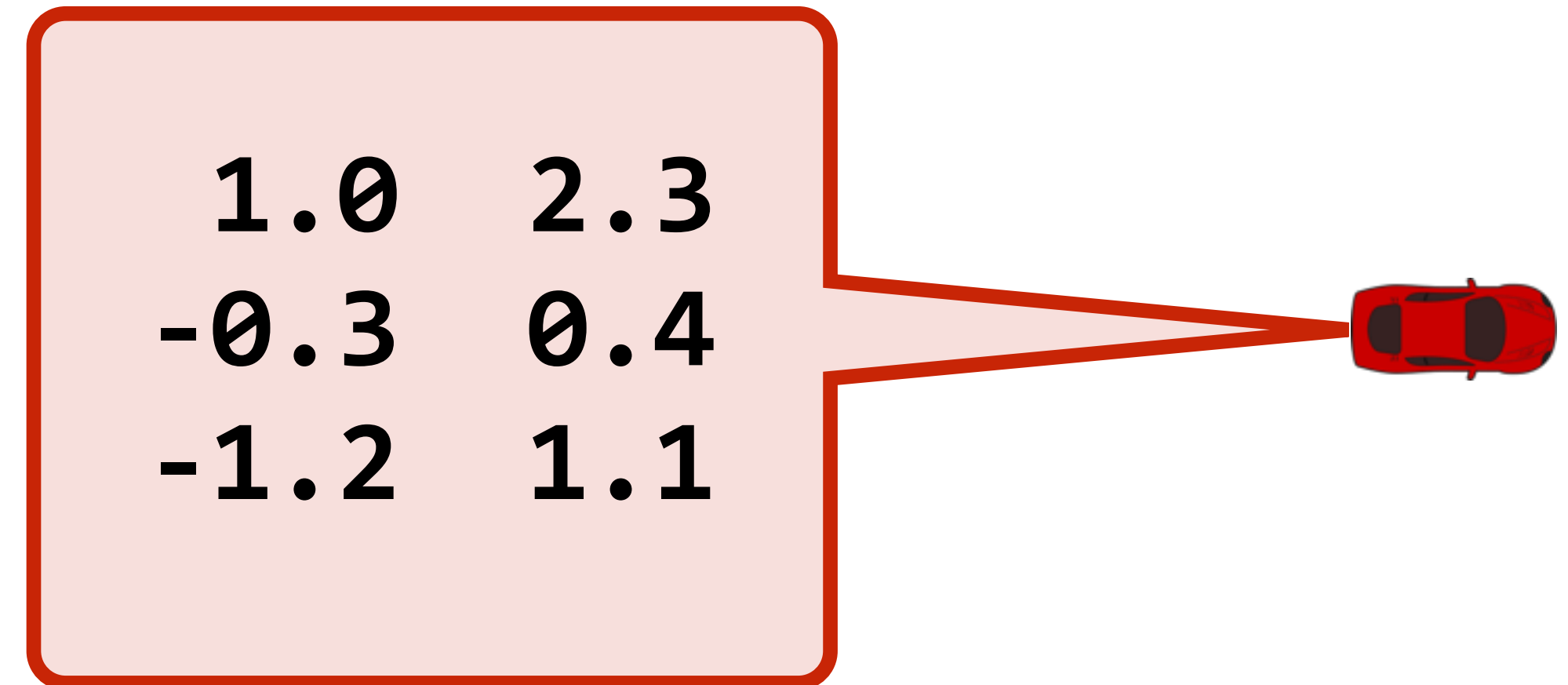
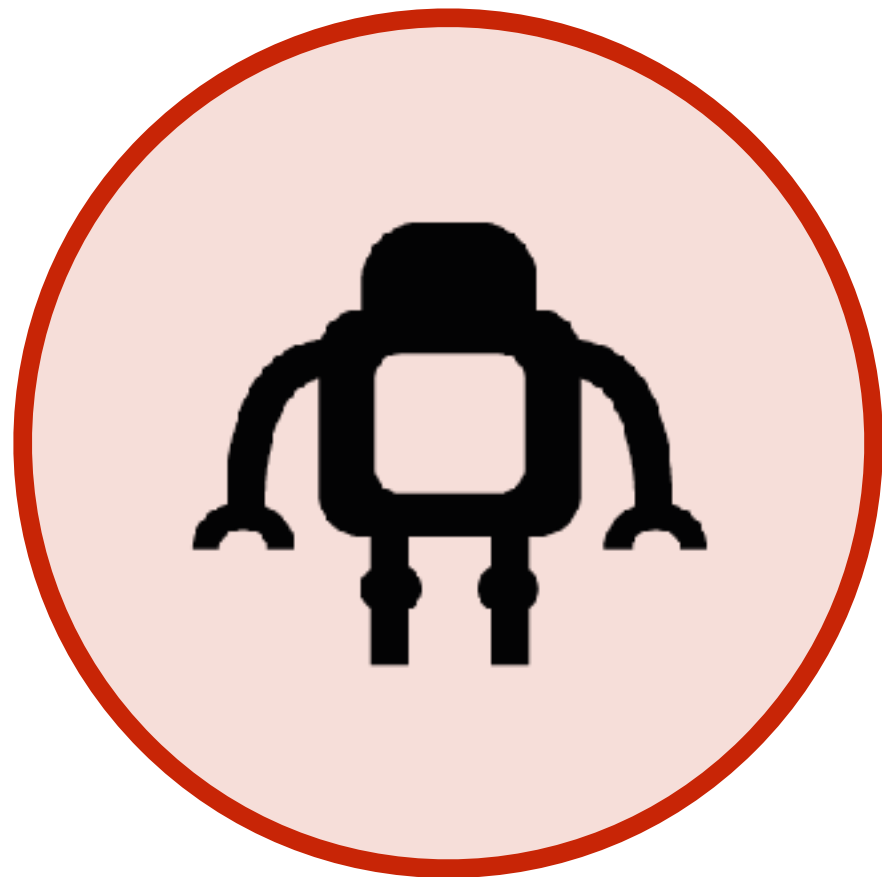


Mike Wehner [@MikeWehner](#)  
March 19th, 2017 at 9:01 AM

Share

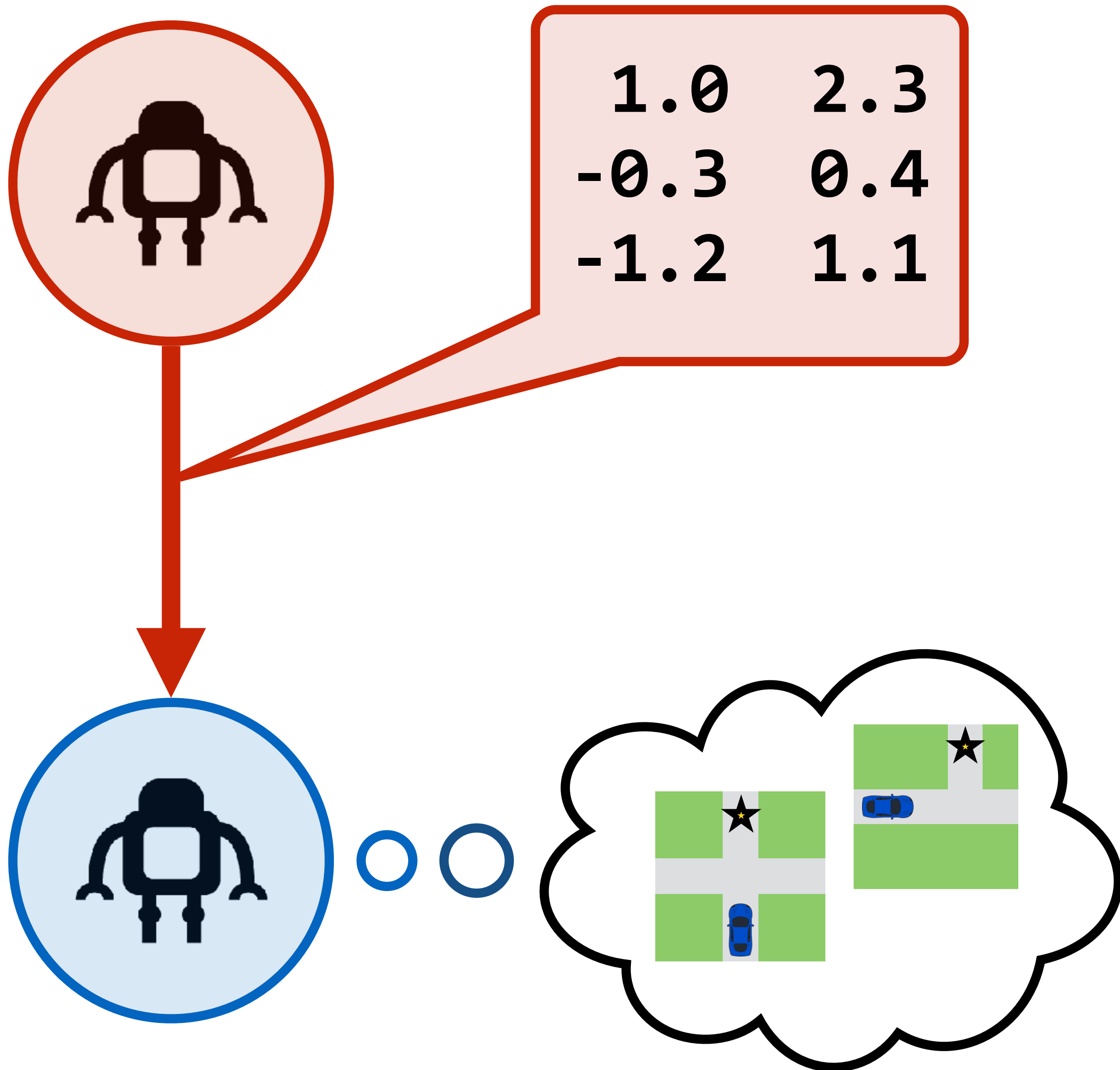
Tweet

# Translating neuralesse

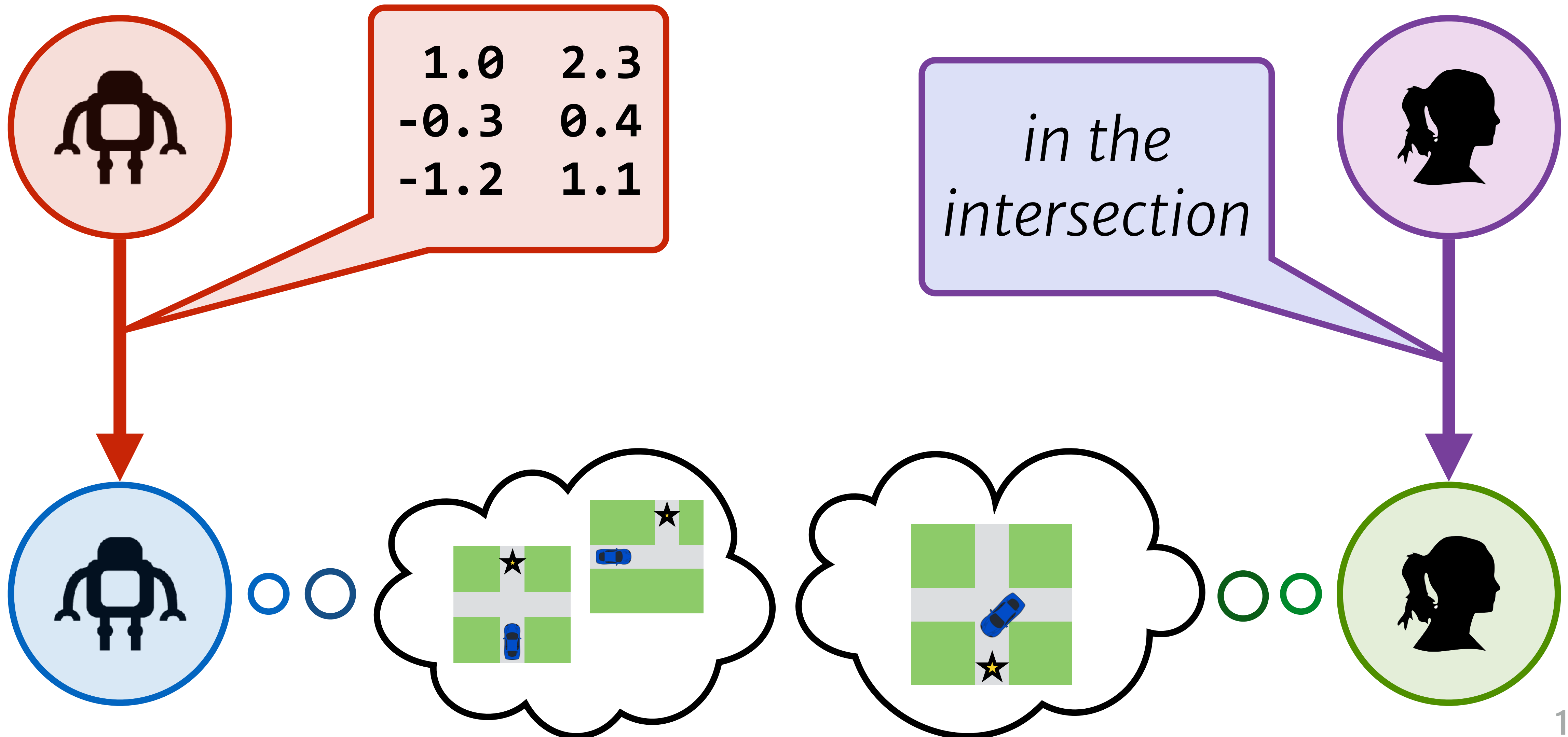


*all clear*

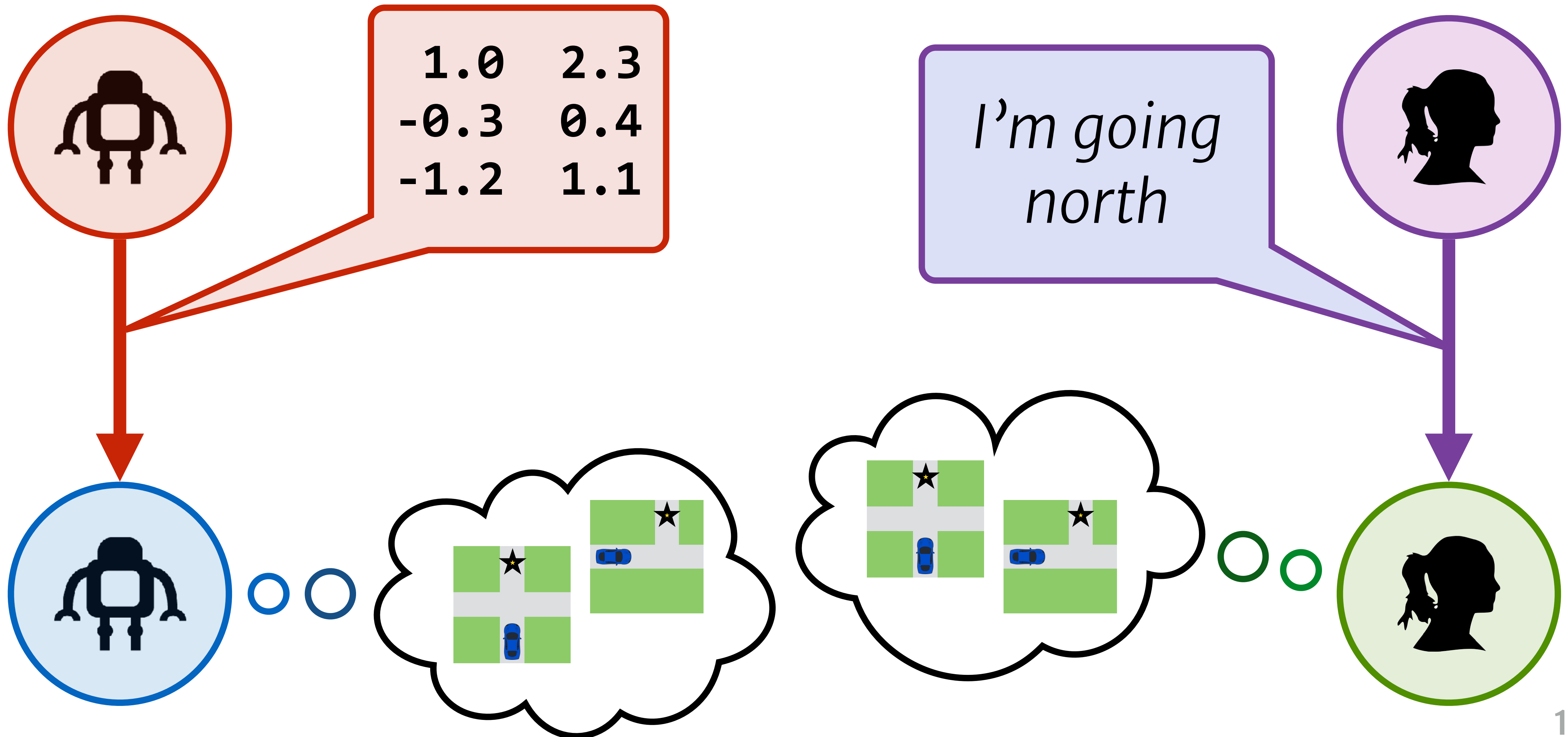
# Translating via belief



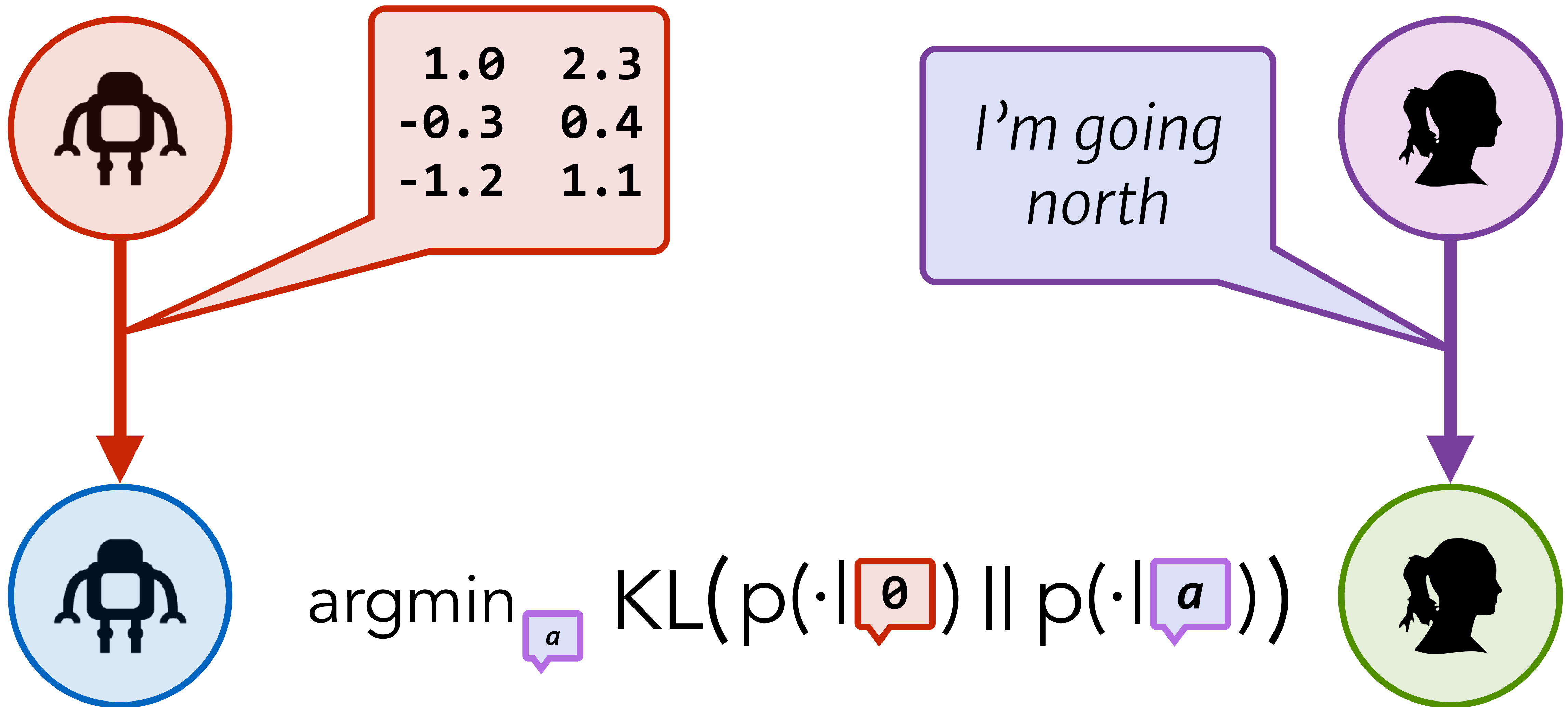
# Translating via belief



# Translating via belief

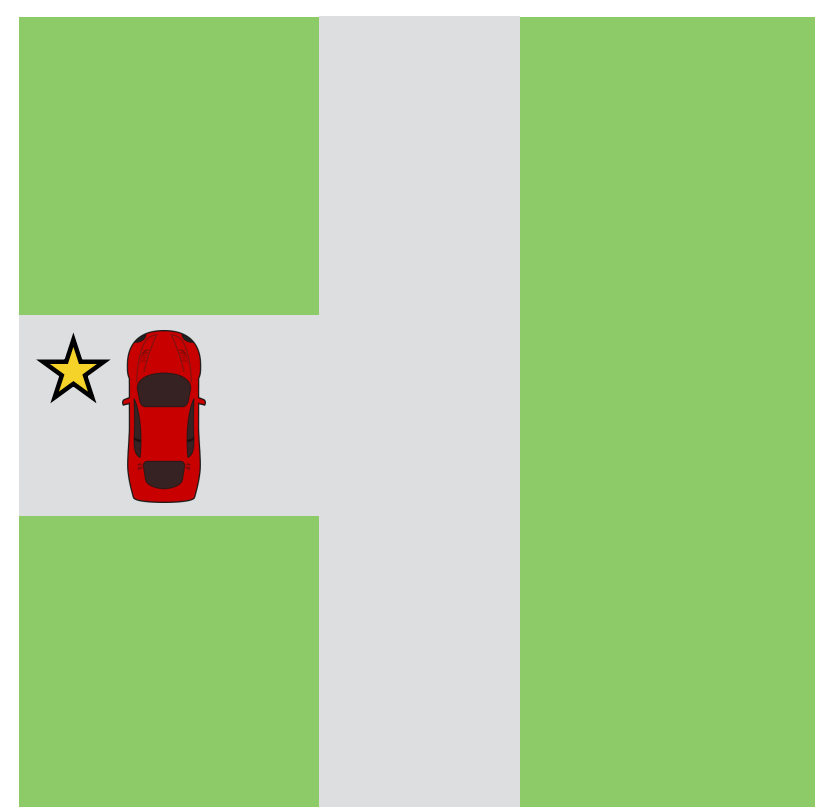


# Translating via belief

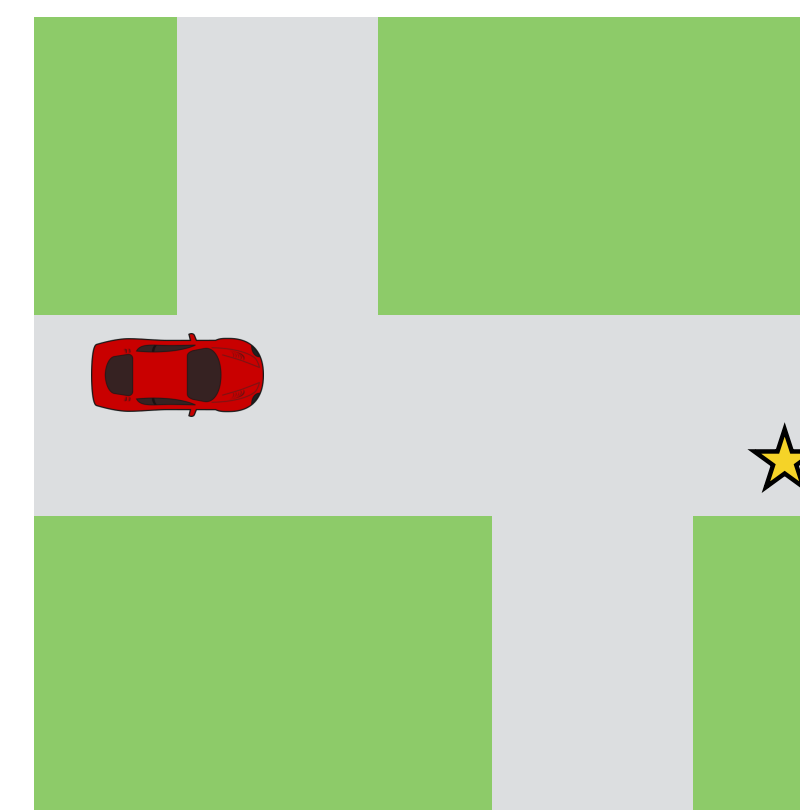




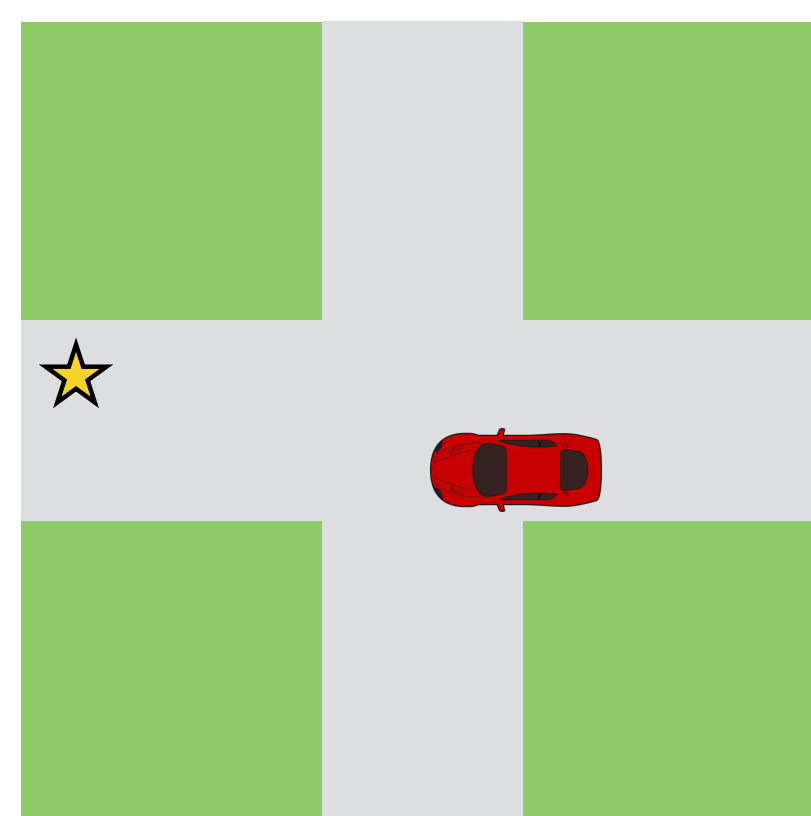
# Example translations



*at goal  
done  
left to top*



*you first  
following  
going down*

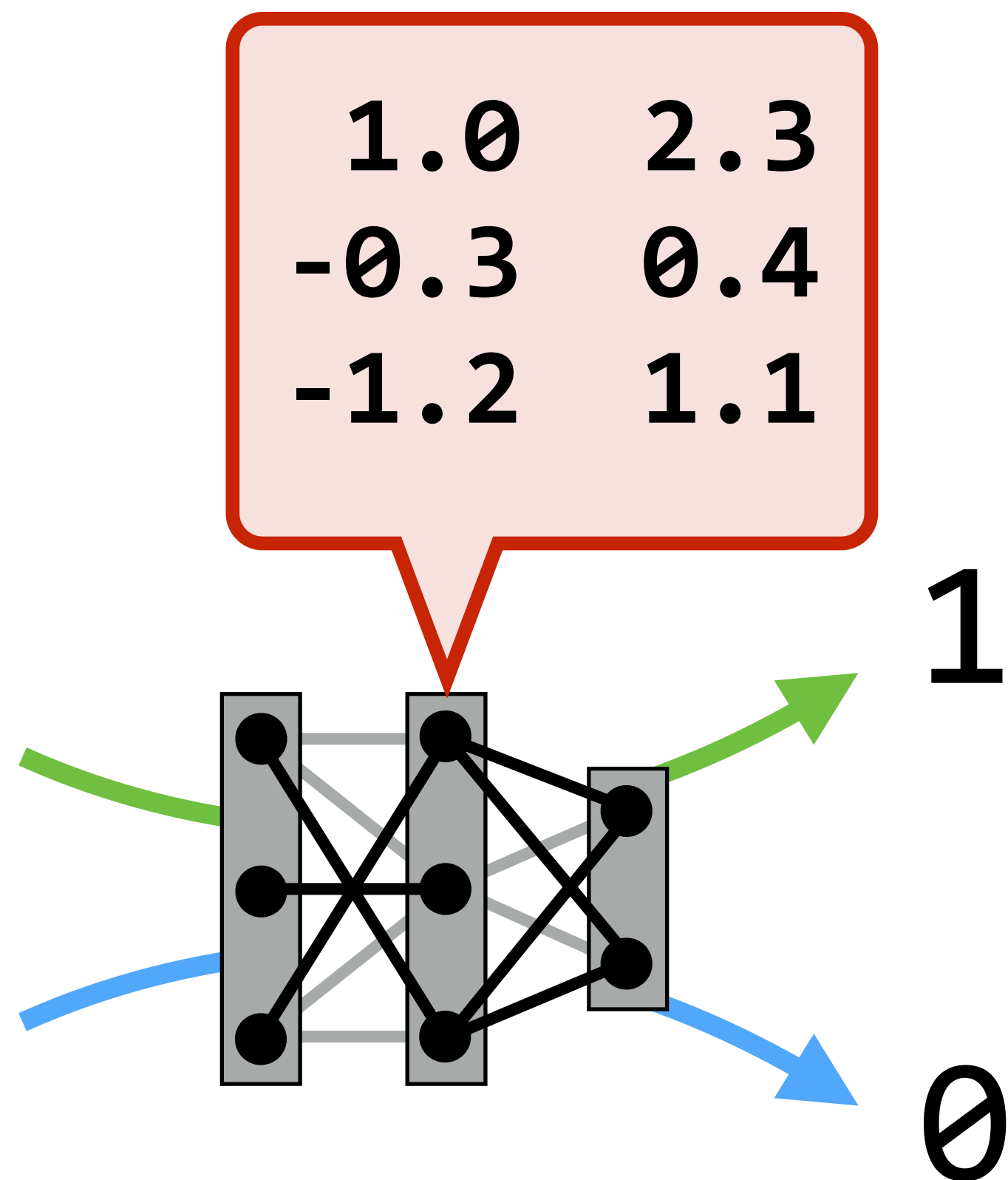
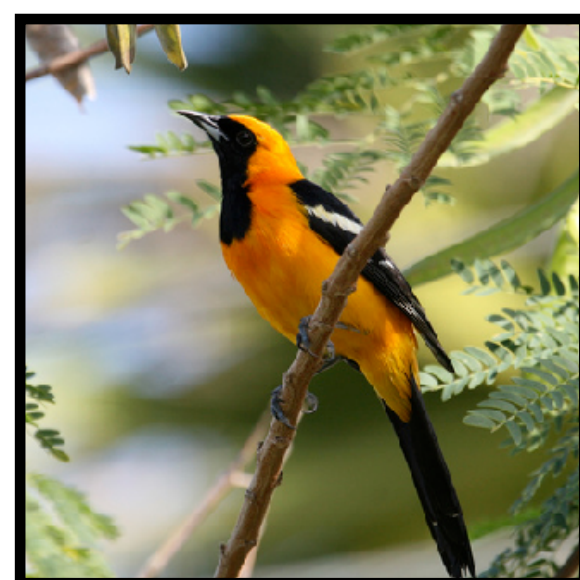


*going in intersection  
proceed  
going*



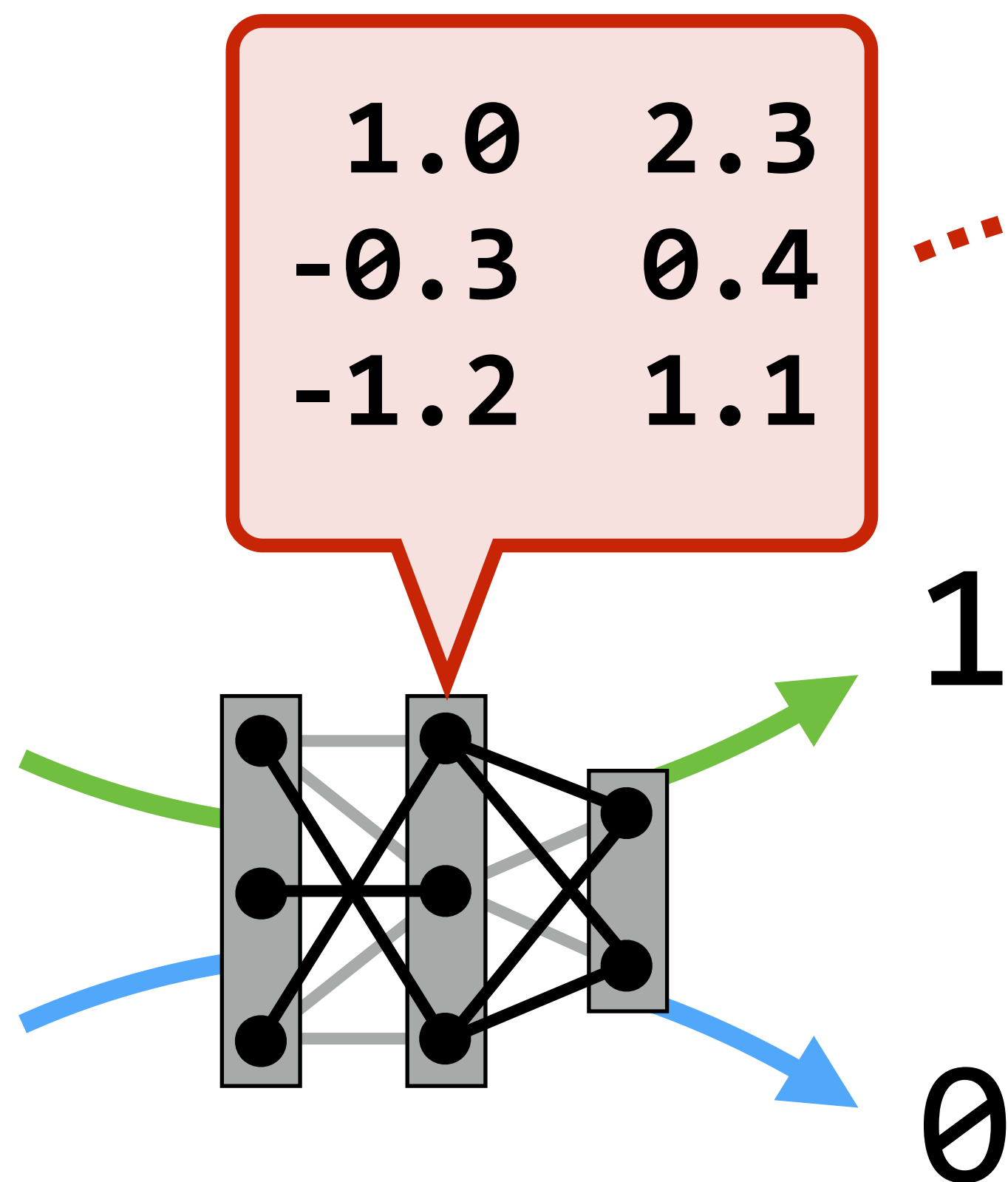
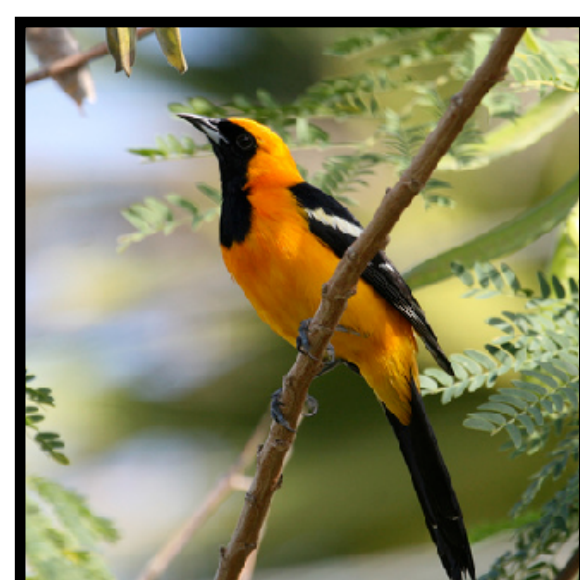


# Translating deep representations





# Translating deep representations



*large bird,  
black wings*



# Translation games [ADK17]

**PROPOSITION:** For agents cooperating via an approximately belief-preserving translation layer, we can bound loss relative to agents with a common language.



# Translation games [ADK17]

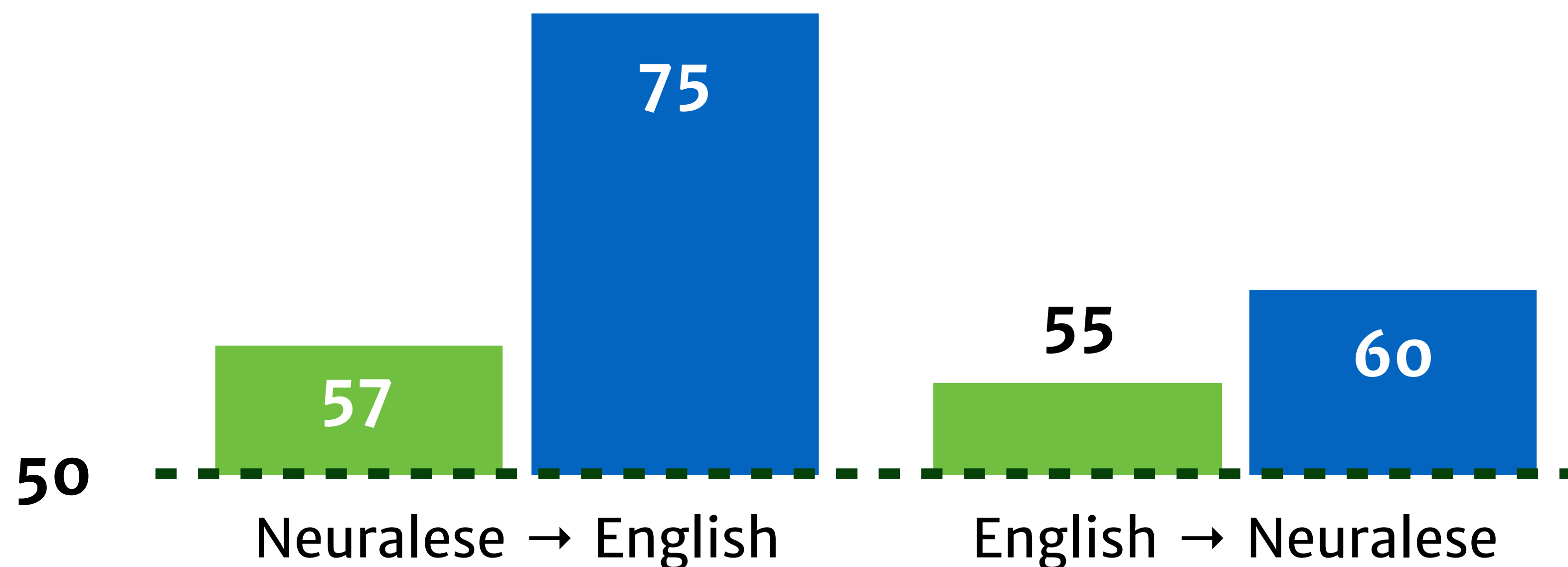
**PROPOSITION:** For agents cooperating via an approximately belief-preserving translation layer, we can bound loss relative to agents with a common language.



seq-to-seq



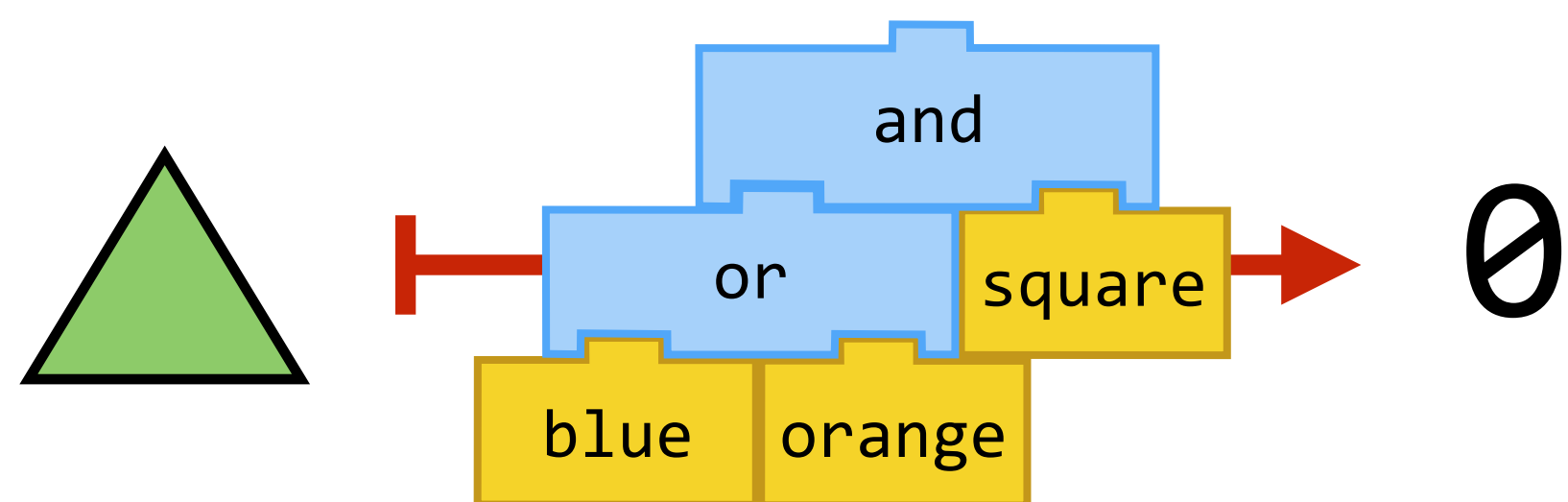
Belief





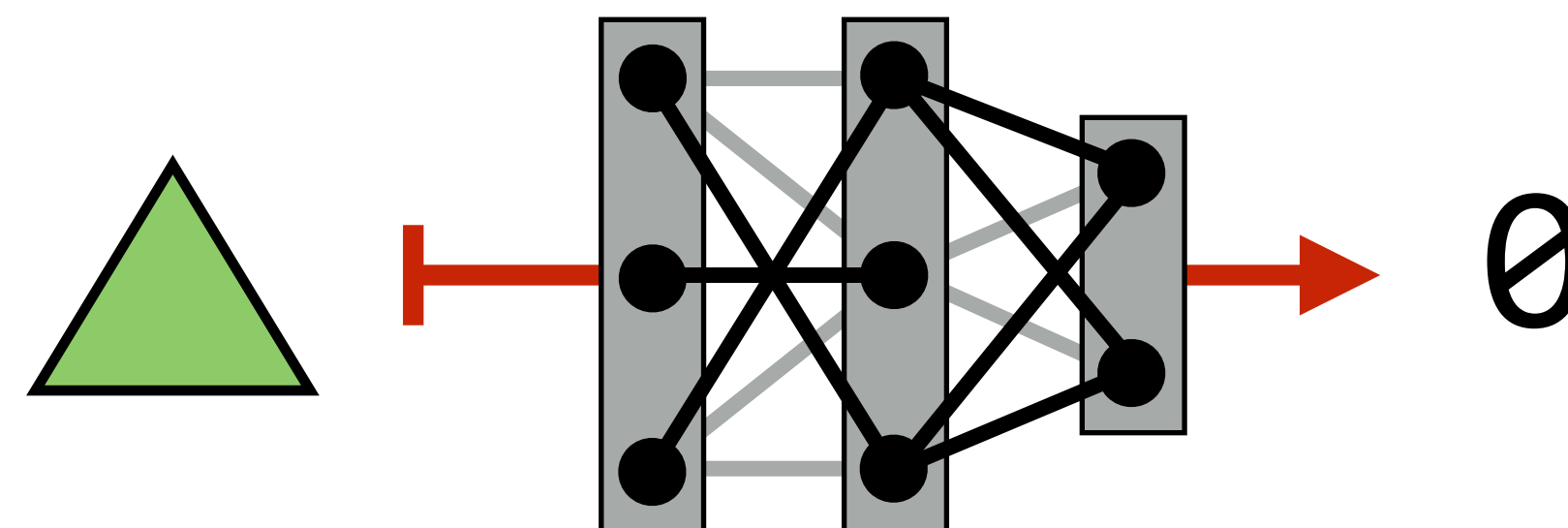
# Explaining classifiers

Interpreted language



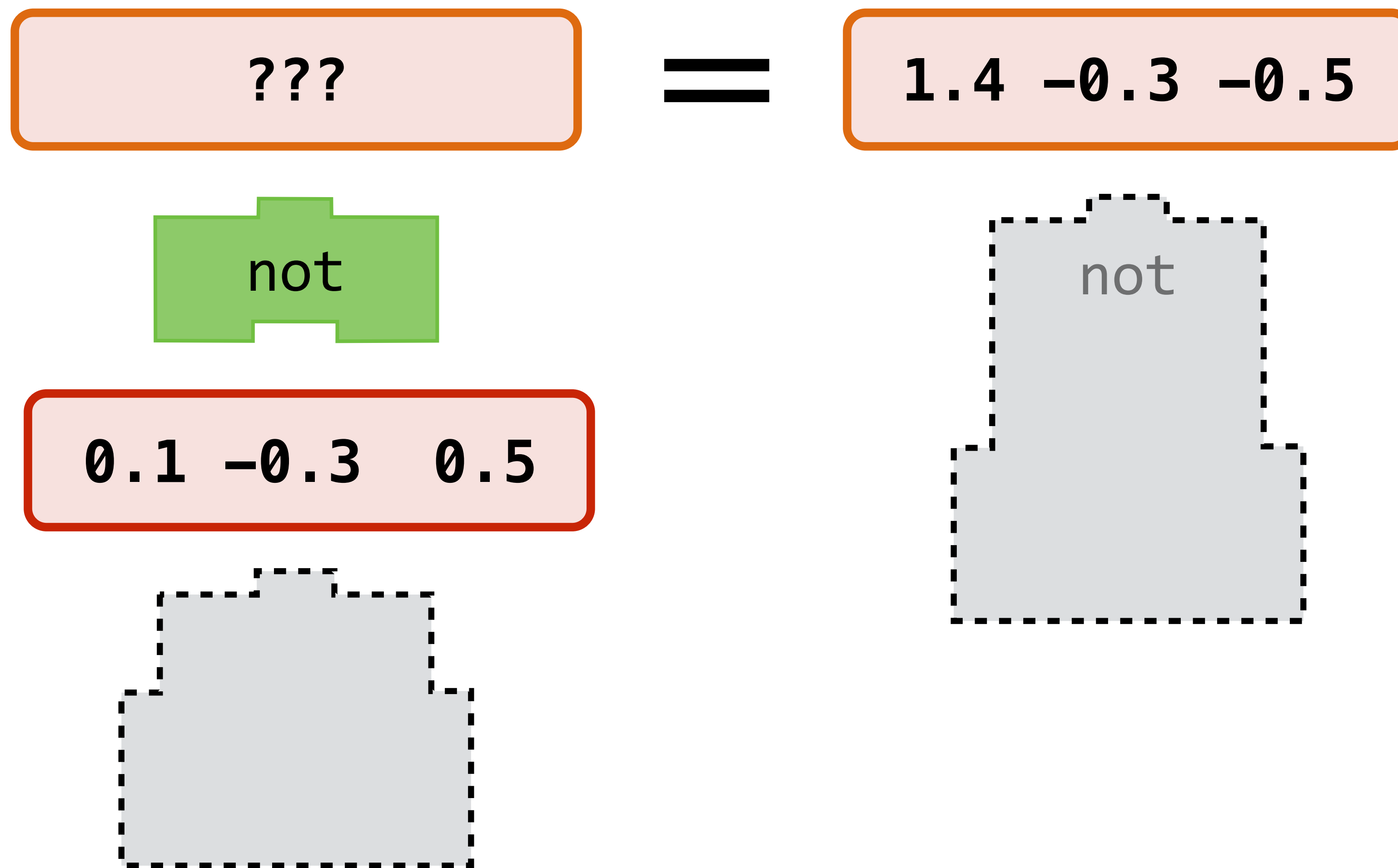
*blue and orange squares*

Learned classifier





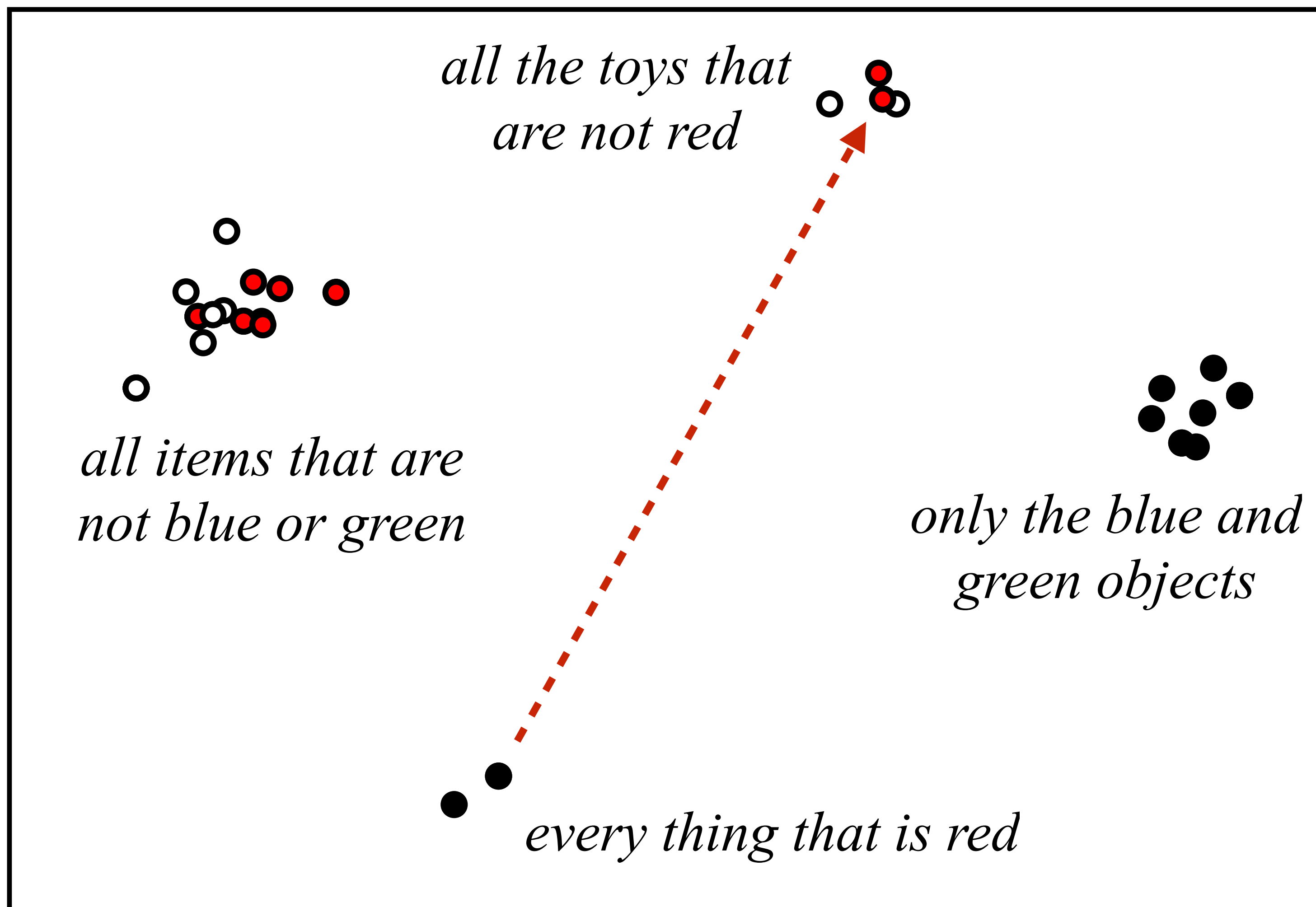
# Learning compositional operators





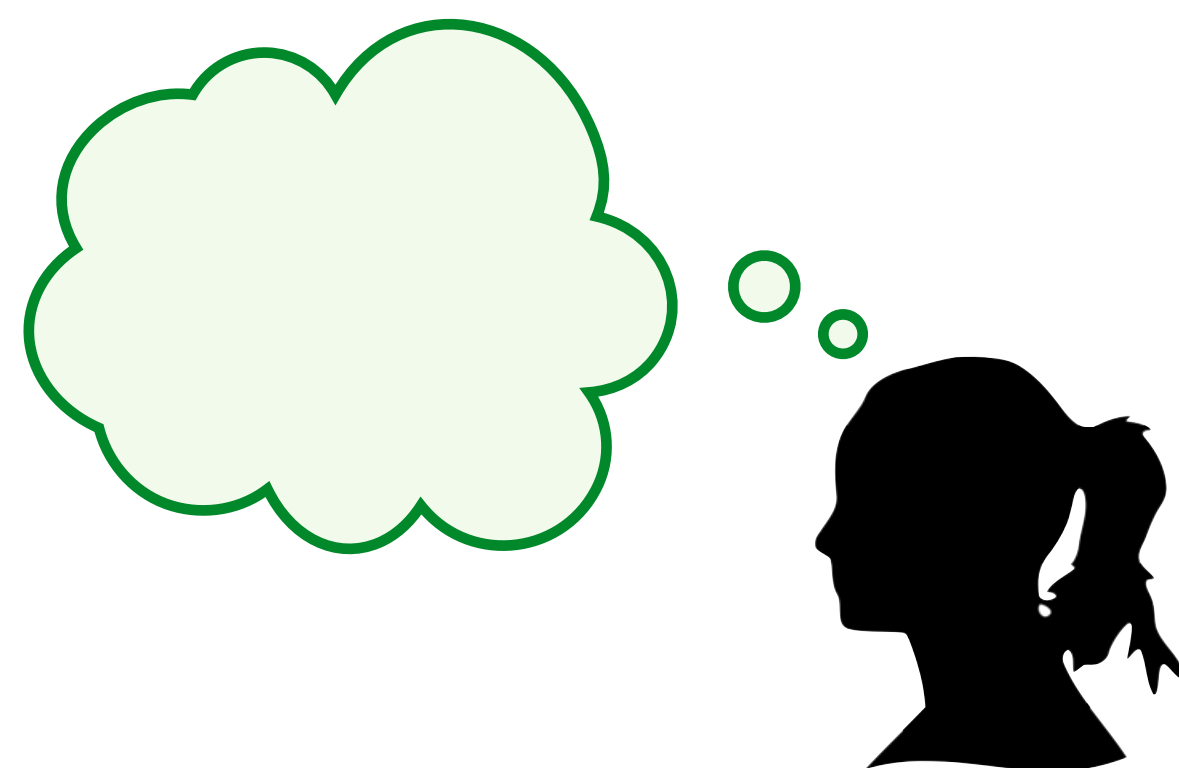
# Learning negation [AK17]

- Input
- Predicted
- True

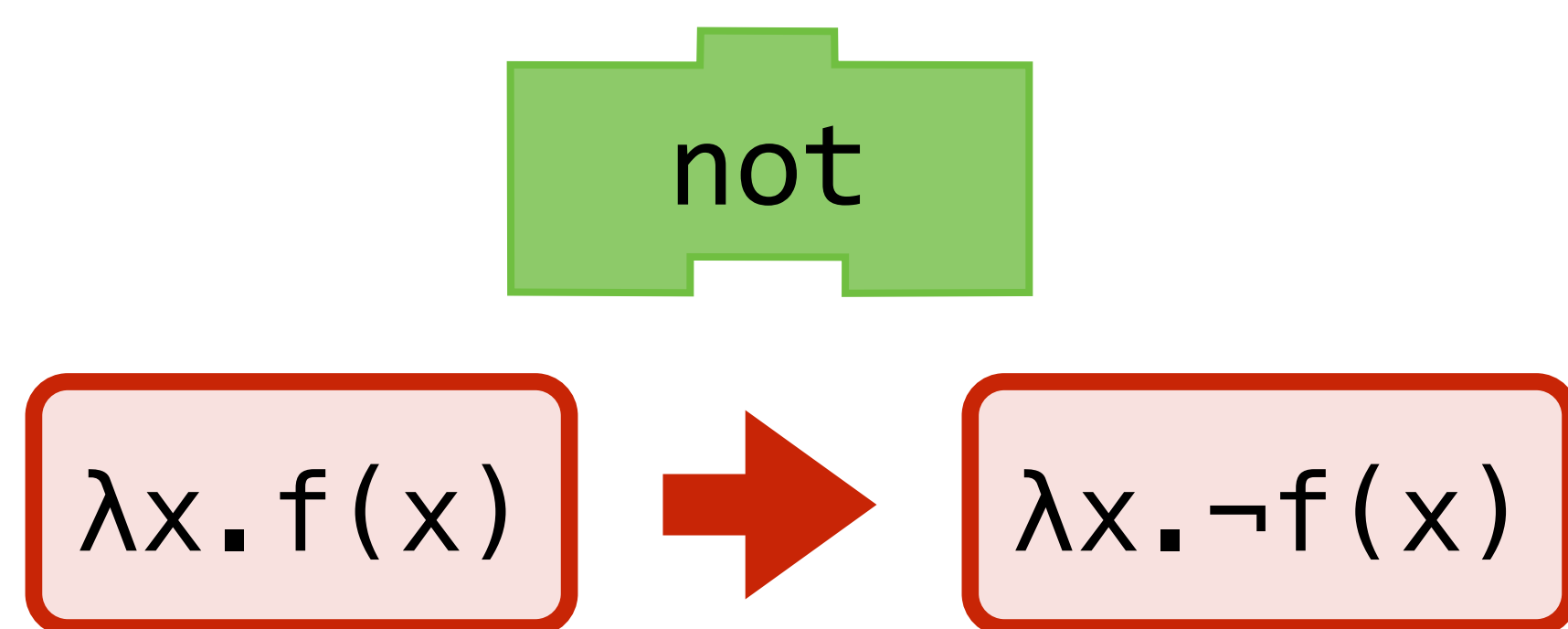




# Lessons

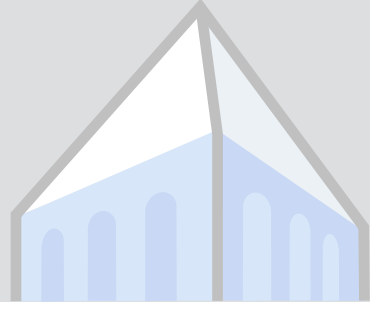


Explicitly modeling listener beliefs helps us build informative models for language generation.

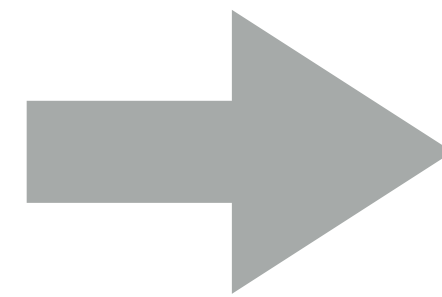


Language lets us find interpretable compositional operators in black-box deep models.

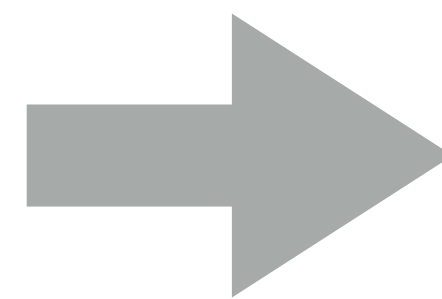




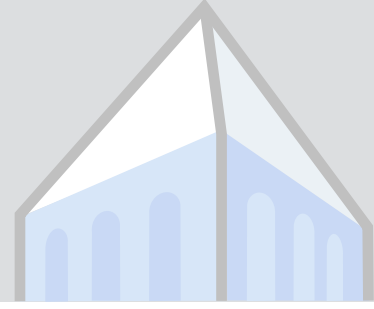
# Safe exploration



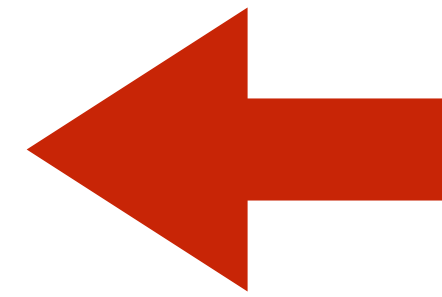
*I will wiggle my front left leg.*



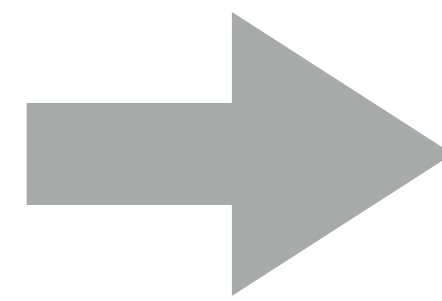
*I will sprint 300 meters forward.*



# Demonstrating competence



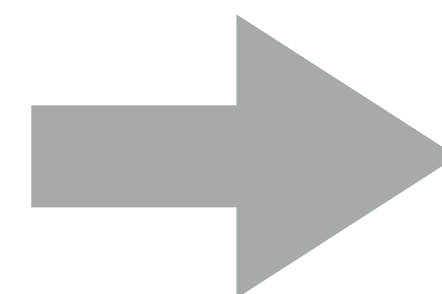
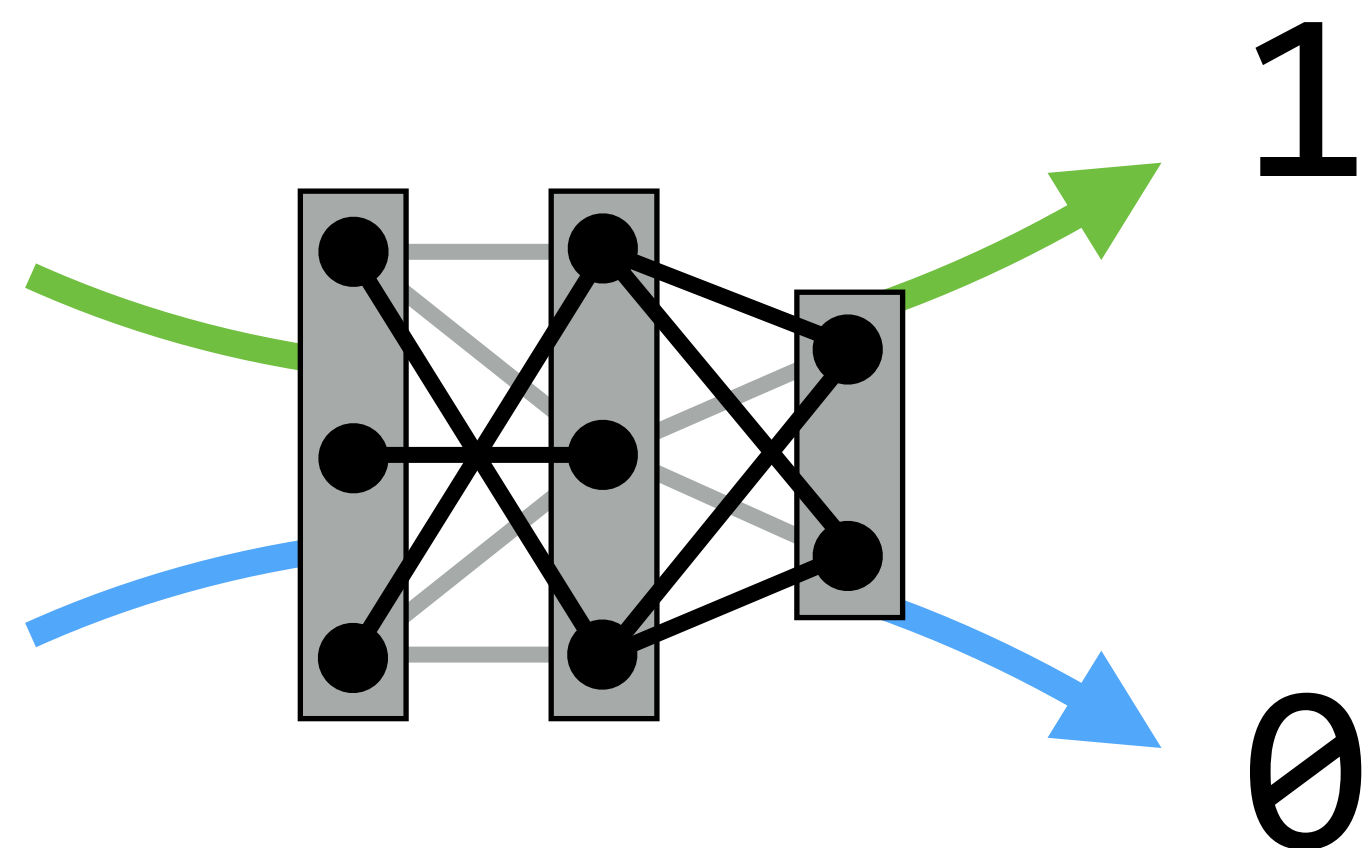
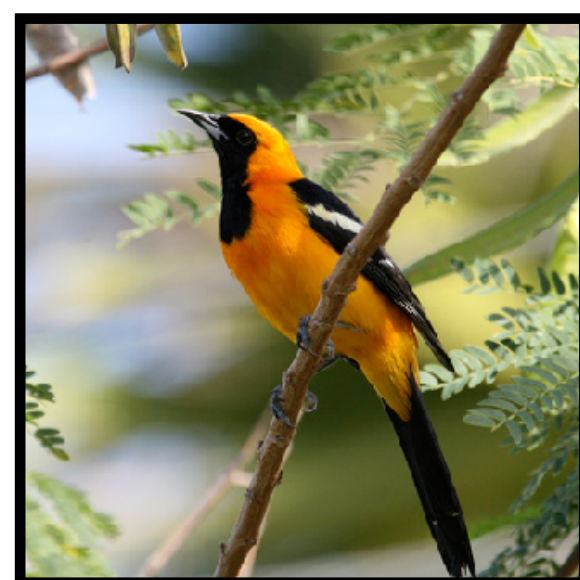
*Defuse the bomb.*



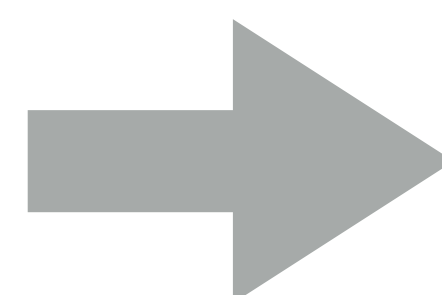
*I will cut open the box and snip the blue wire while avoiding the red one.*



# Explaining limitations



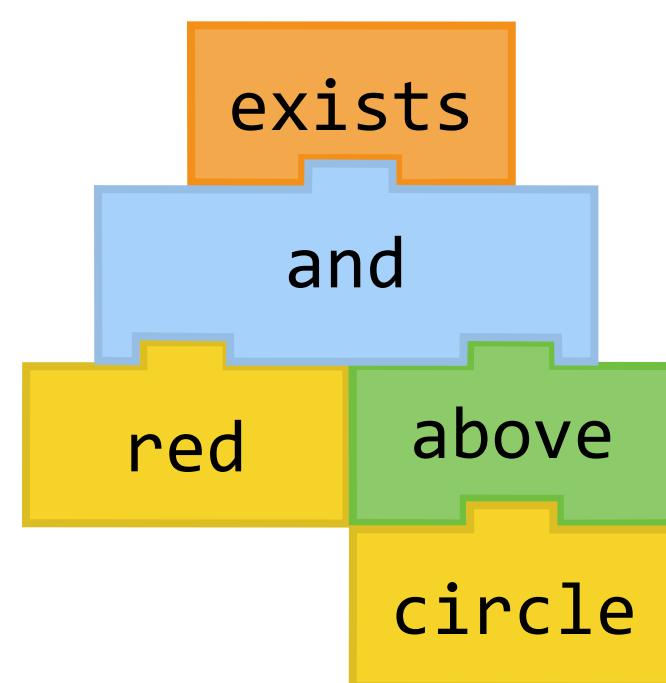
*All western tanagers have yellow heads.*



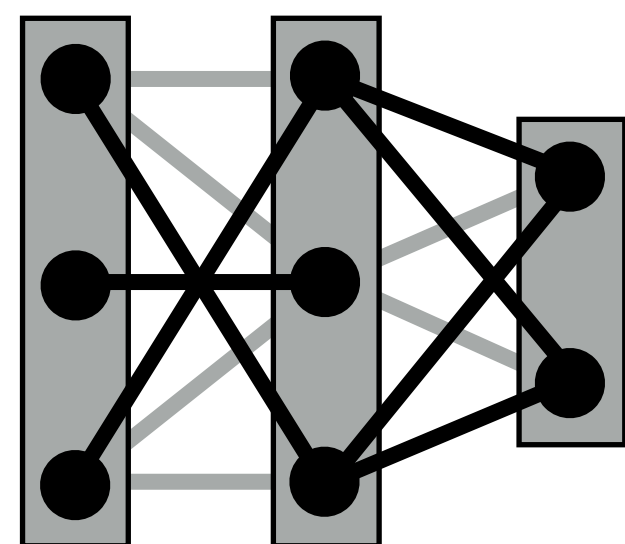
*I can't tell the difference between ravens and crows.*



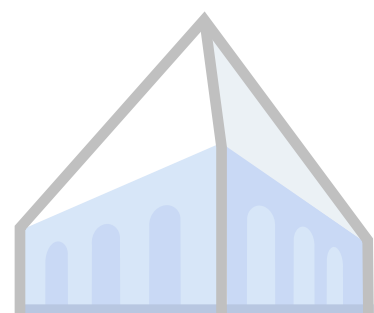
# Last lessons



The structure of language helps us design models that reflect the compositional structure of the world.



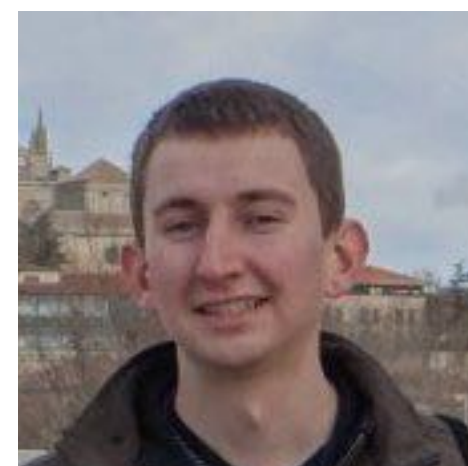
These models provide more accurate and interpretable learning for language processing and more.



Trevor  
Darrell



Anca  
Drăgan



Daniel  
Fried



Dan  
Klein



Ronghang  
Hu



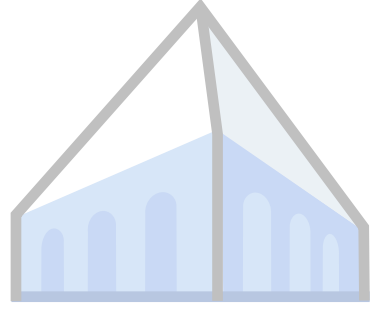
Sergey  
Levine



Marcus  
Rohrbach

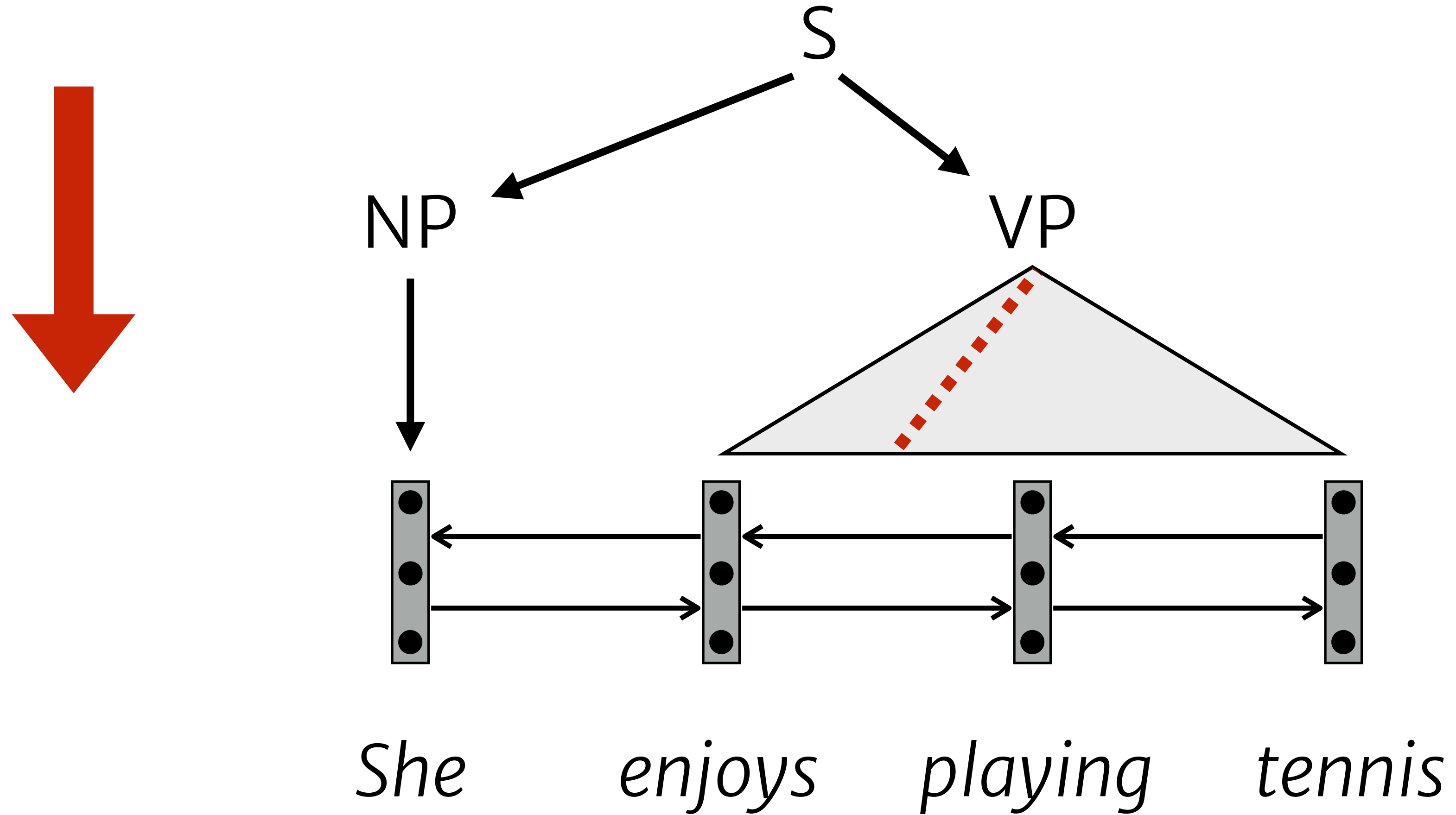


Kate  
Saenko





# Bonus parse tree [AK14, SAK17]





# Learning with latent language [AKL18]

