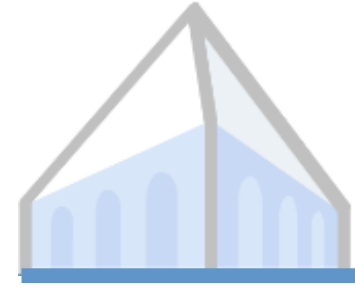


# Learning to compose neural networks for question answering



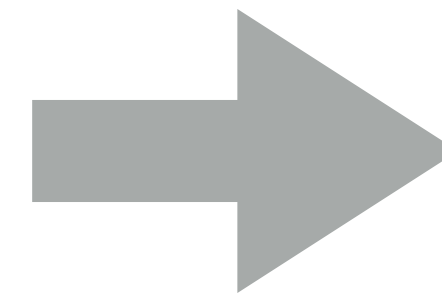
Jacob Andreas, Marcus Rohrbach,  
Trevor Darrell, Dan Klein



# Grounded question answering

---

*What color is  
the necktie?*



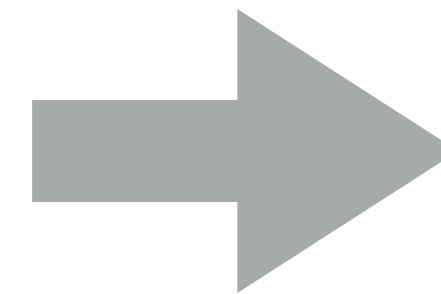
*yellow*



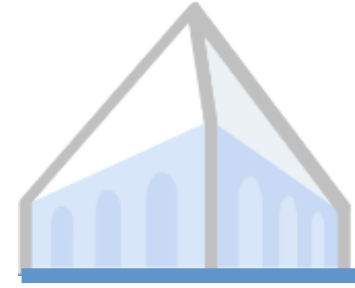
# Grounded question answering

*What rivers  
are in South  
Carolina?*

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



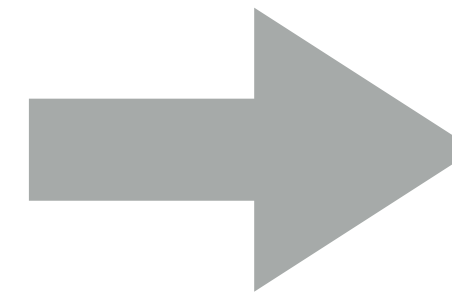
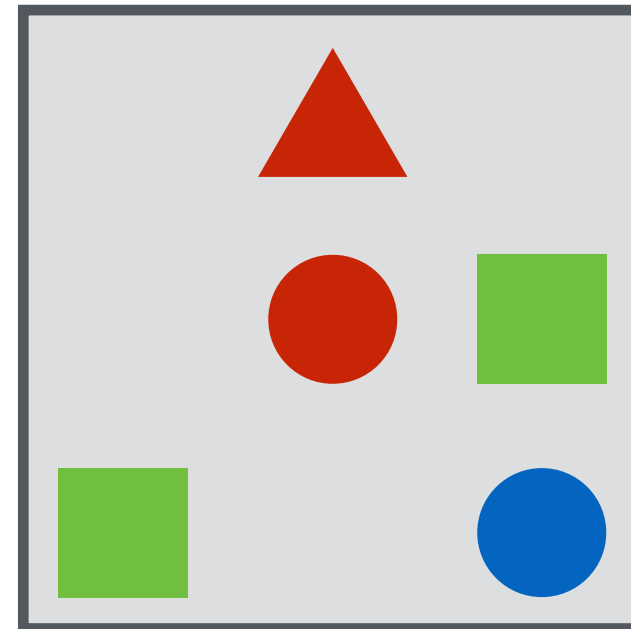
*Cooper*



# Grounded question answering

---

*Is there a red  
shape above  
a circle?*

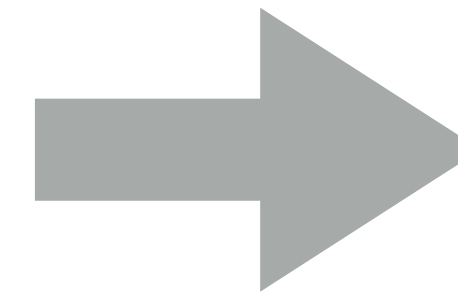
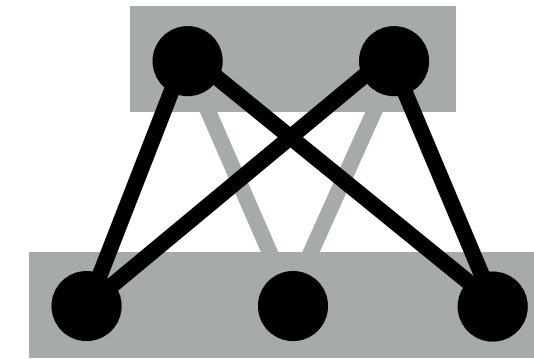
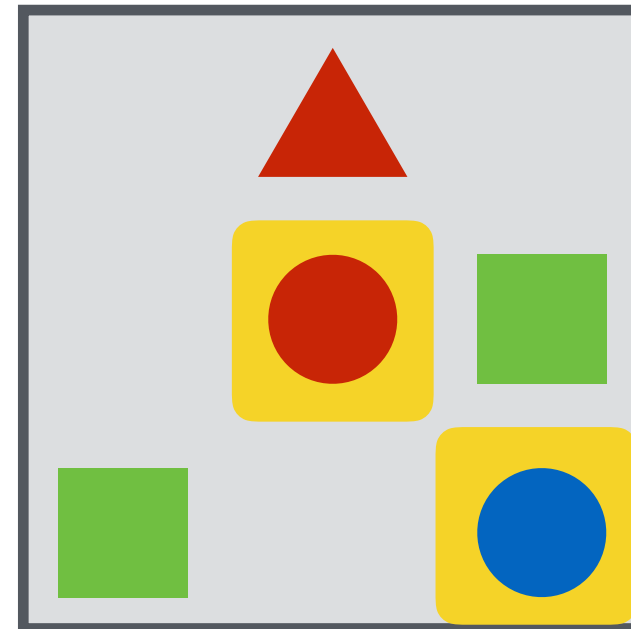


*yes*



# Neural nets learn lexical groundings

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



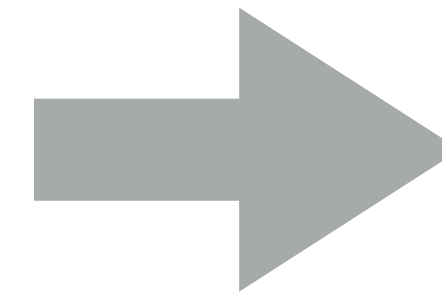
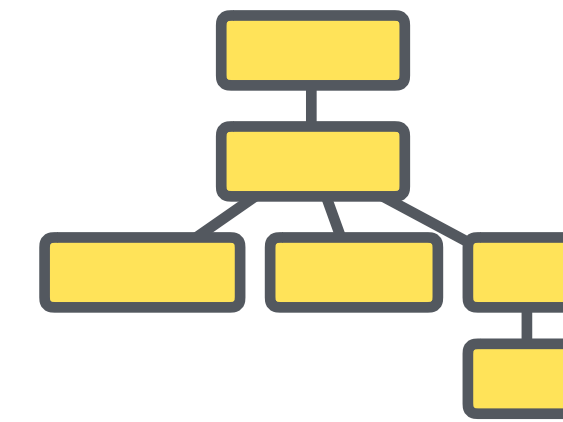
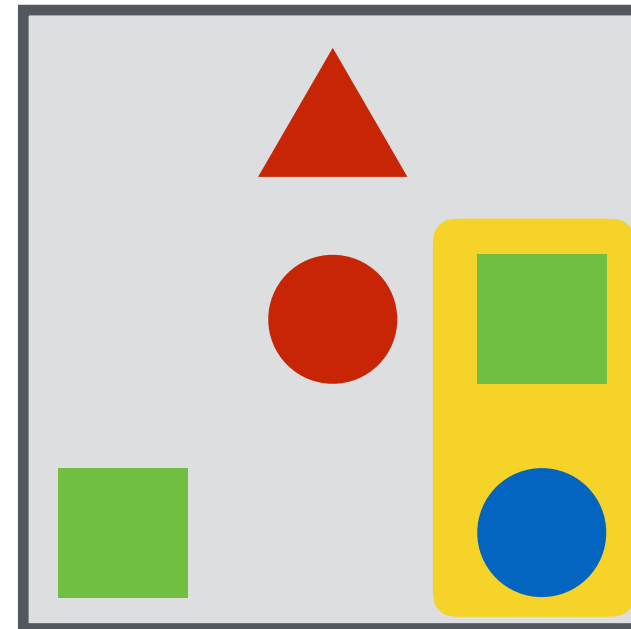
*yes*

[Iyyer et al. 2014, Bordes et al. 2014,  
Yang et al. 2015, Malinowski et al., 2015]



# Semantic parsers learn composition

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



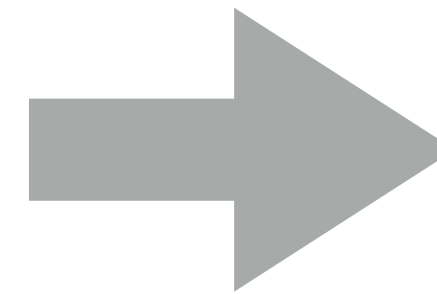
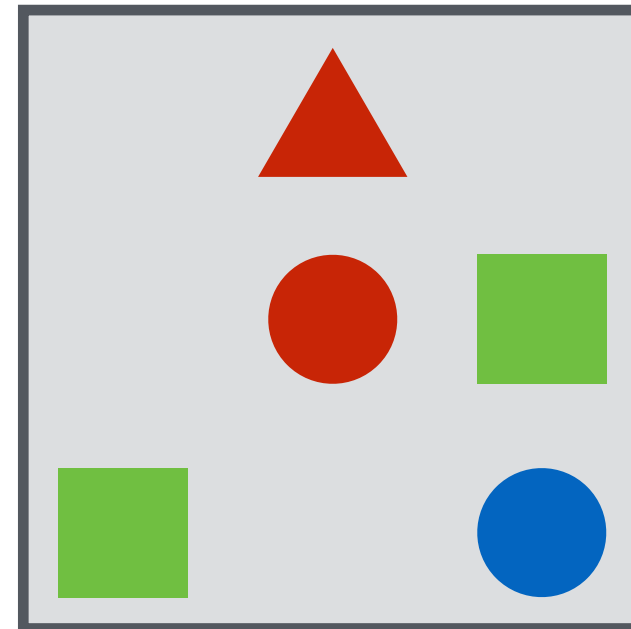
*yes*

[Wong & Mooney 2007, Kwiatkowski et al. 2010,  
Liang et al. 2011, A et al. 2013]

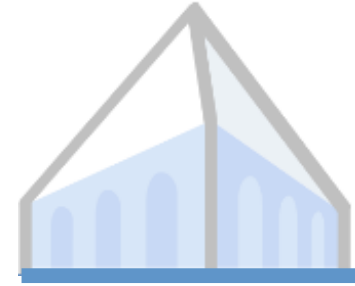


# Neural module networks learn both!

*Is there a red shape above a circle?*

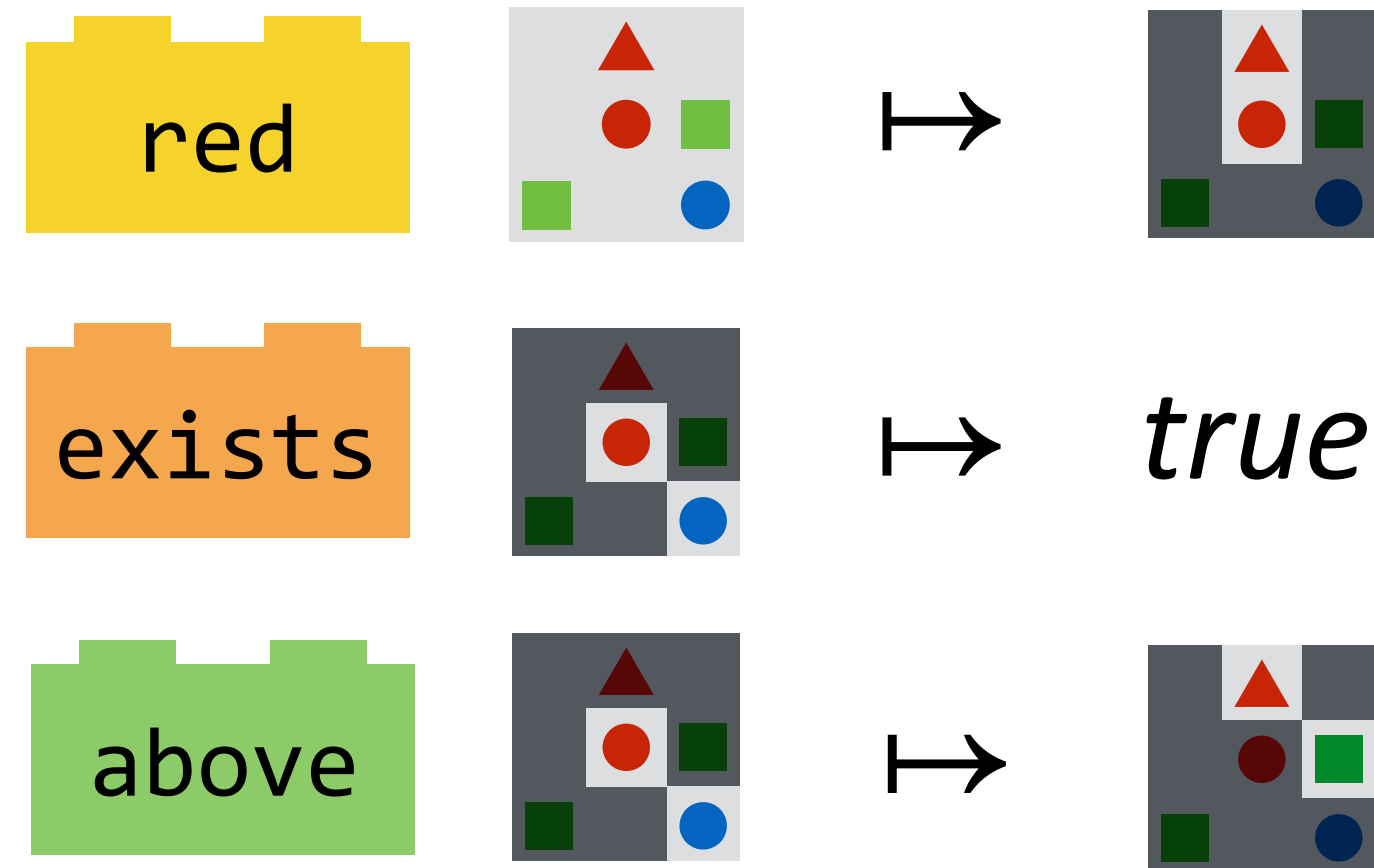


*yes*



# Neural module networks

*Is there a red shape  
above a circle?*

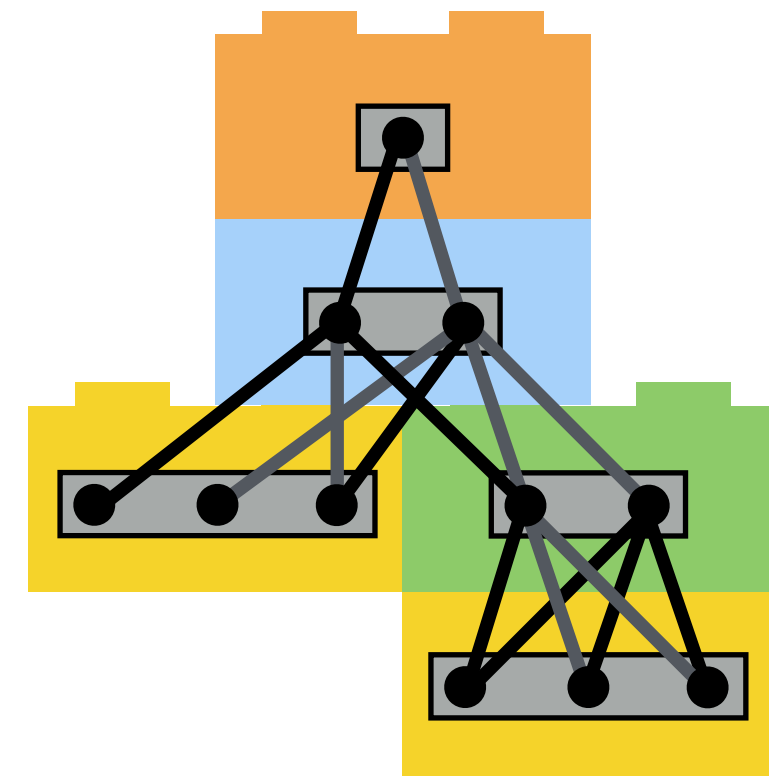
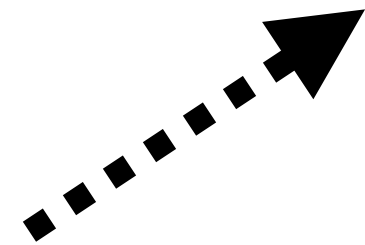
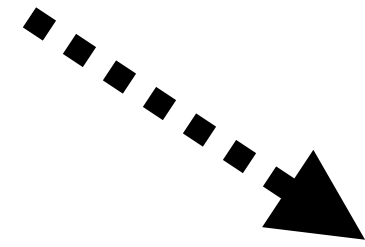
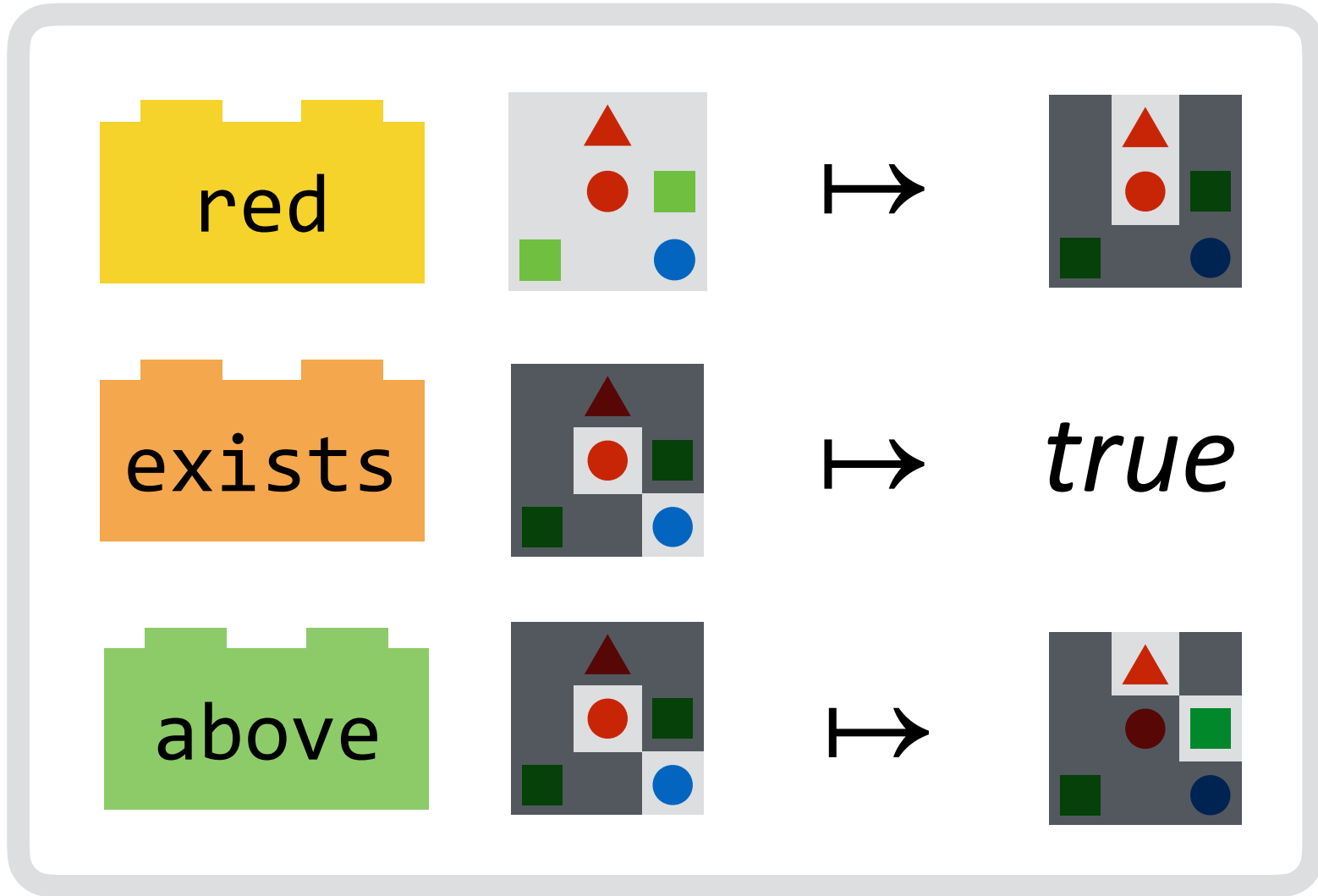






# Neural module networks

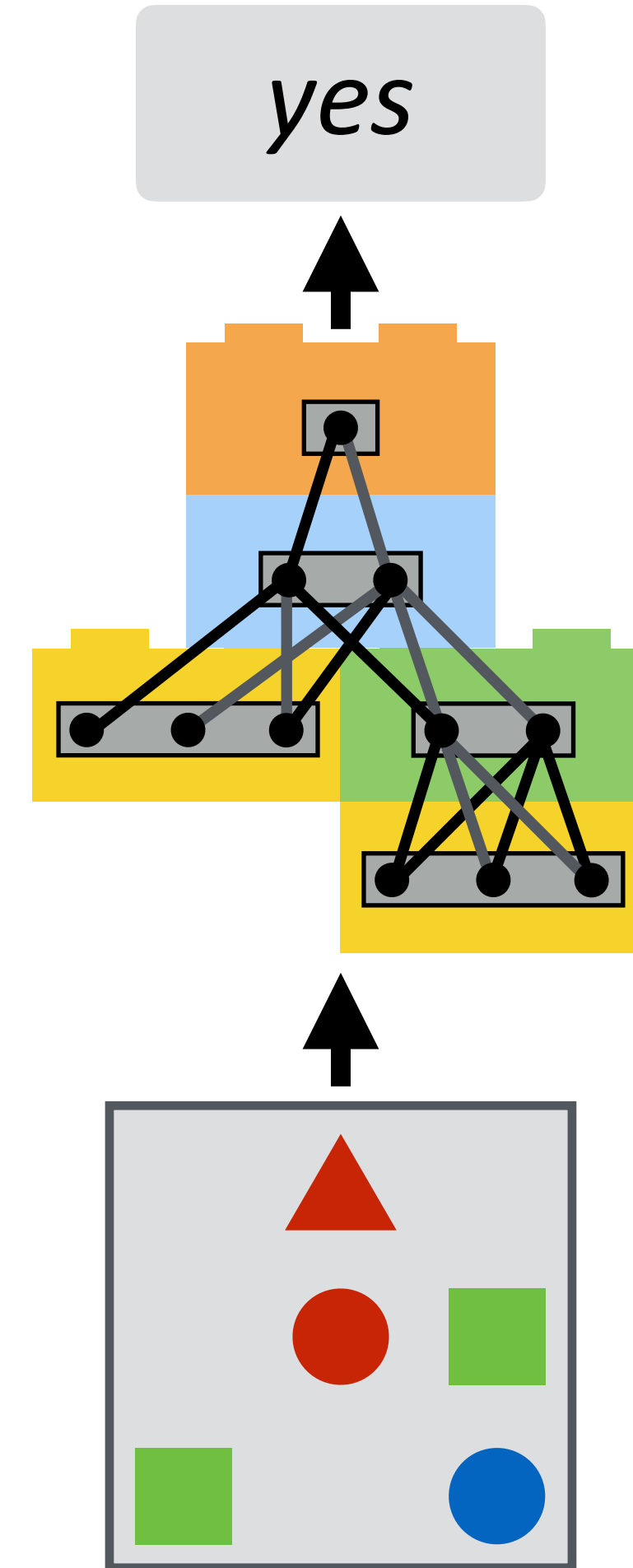
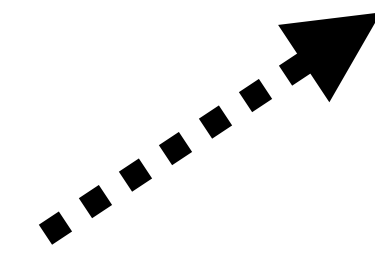
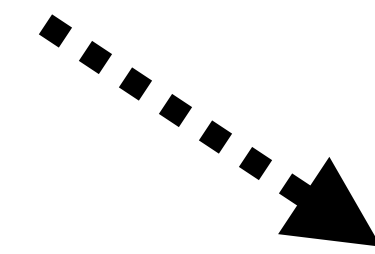
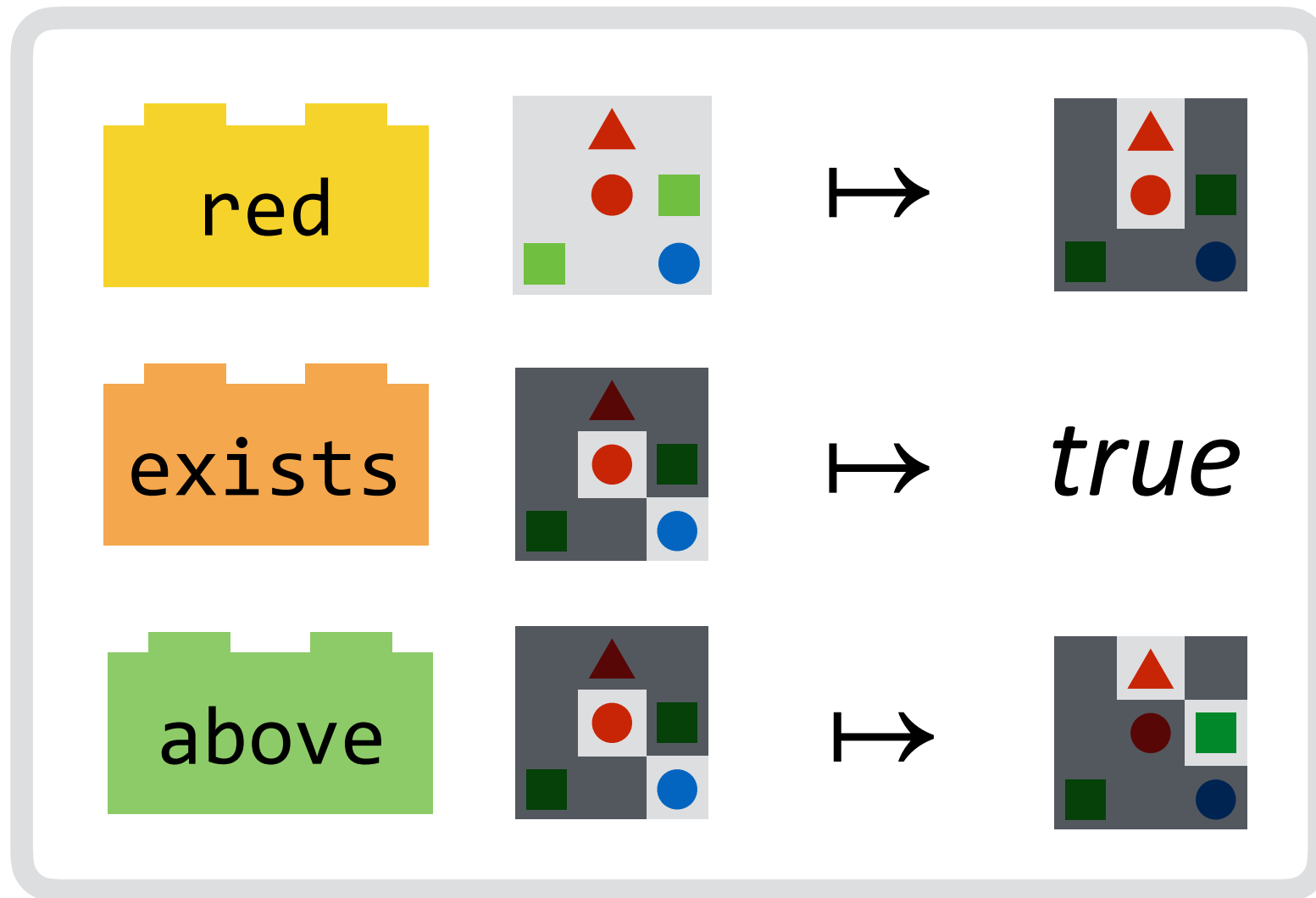
*Is there a red shape above a circle?*





# Neural module networks

*Is there a red shape above a circle?*





# Nearest neighbors

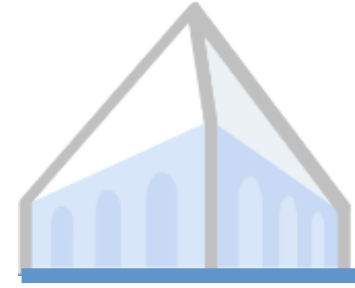
---

## **Structured neural models**

- [Socher et al. 2011, Bottou et al. 1997, Mnih et al. 2014]

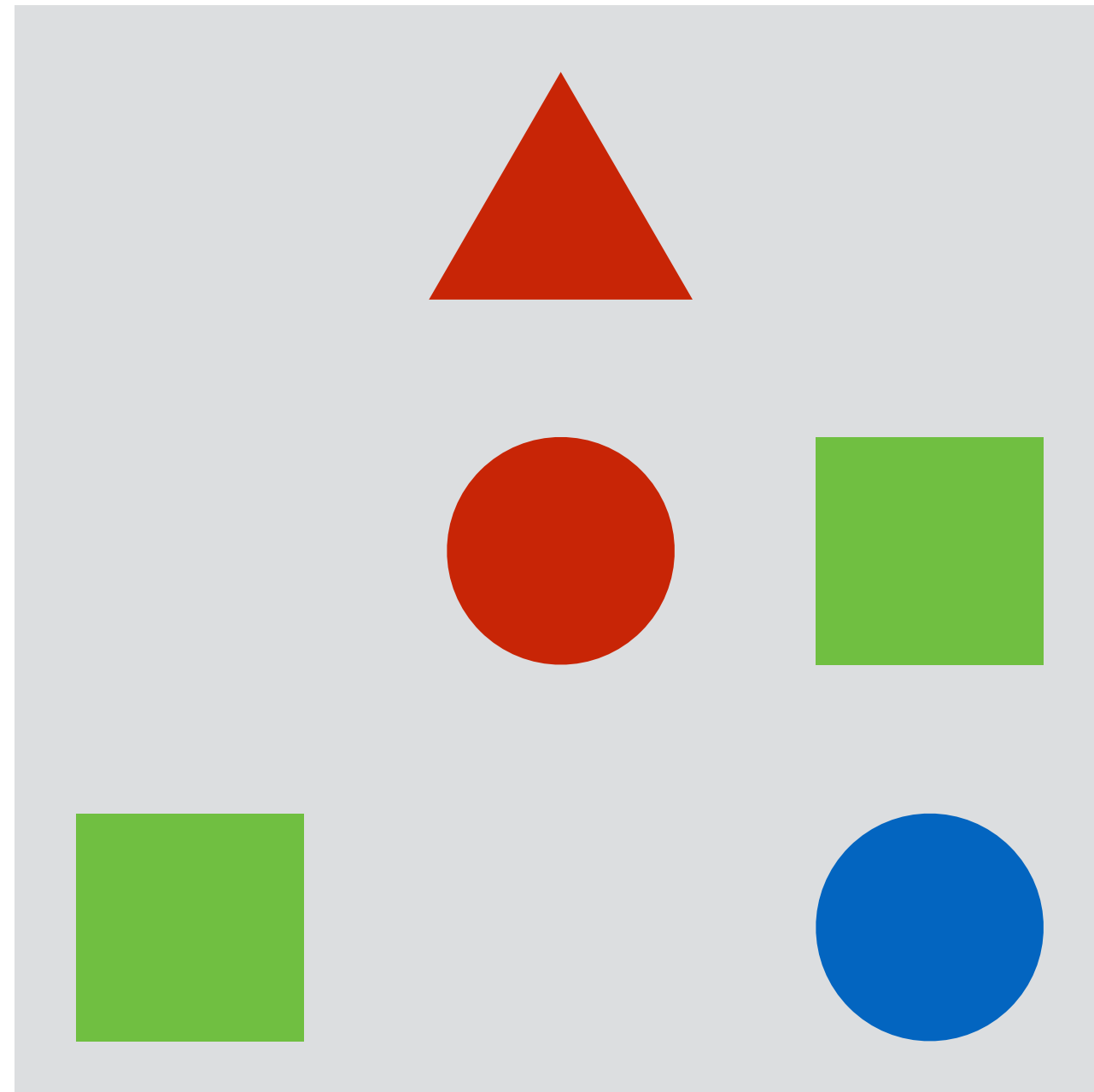
## **Probabilistic formal semantics / predicate learning**

- [Beltagy et al. 2013, Lewis & Steedman 2013, Malinowski & Fritz 2014]



# Representing meaning

---

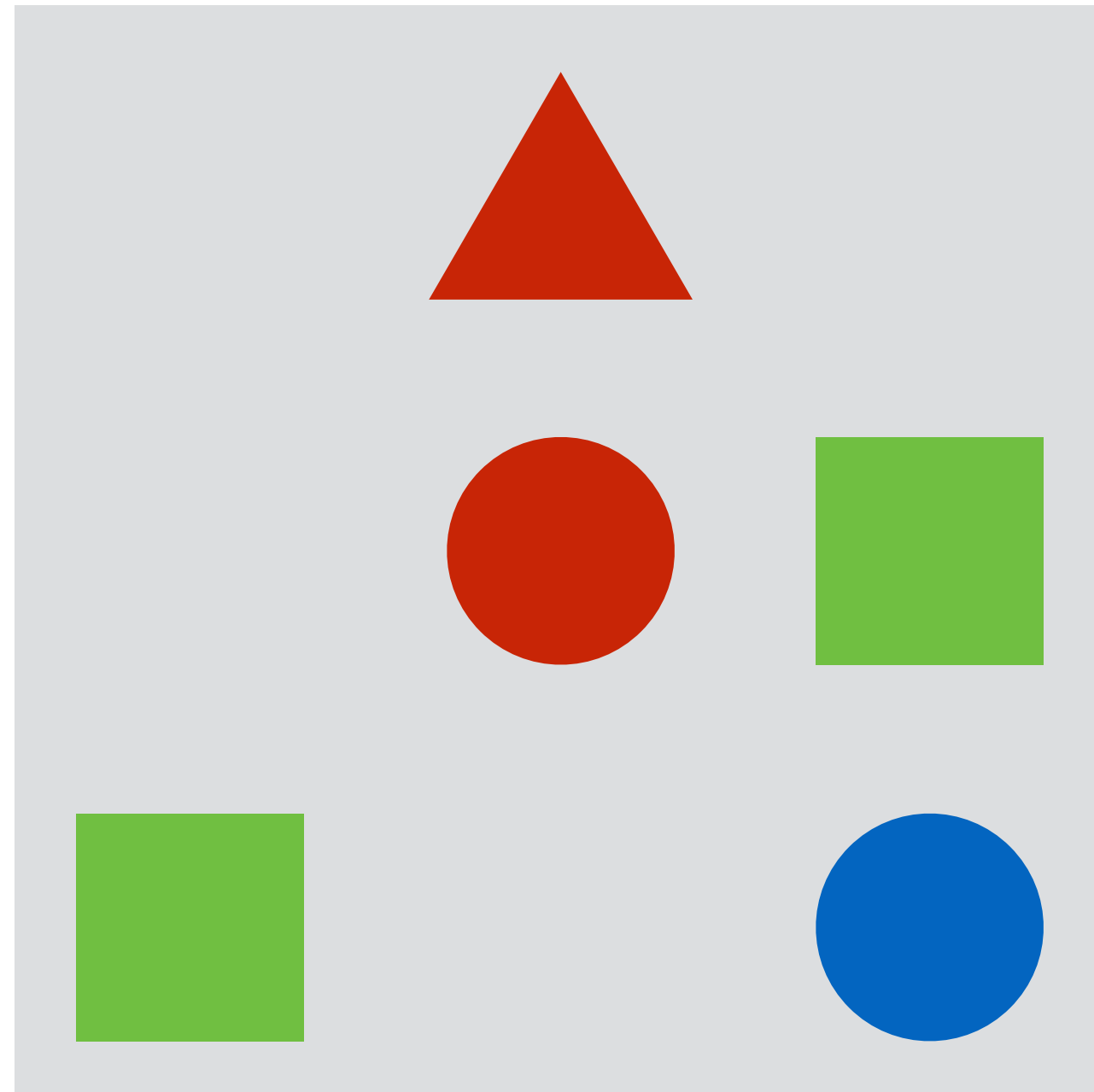


*Is there a red shape above a circle?*



# Representing meaning

---

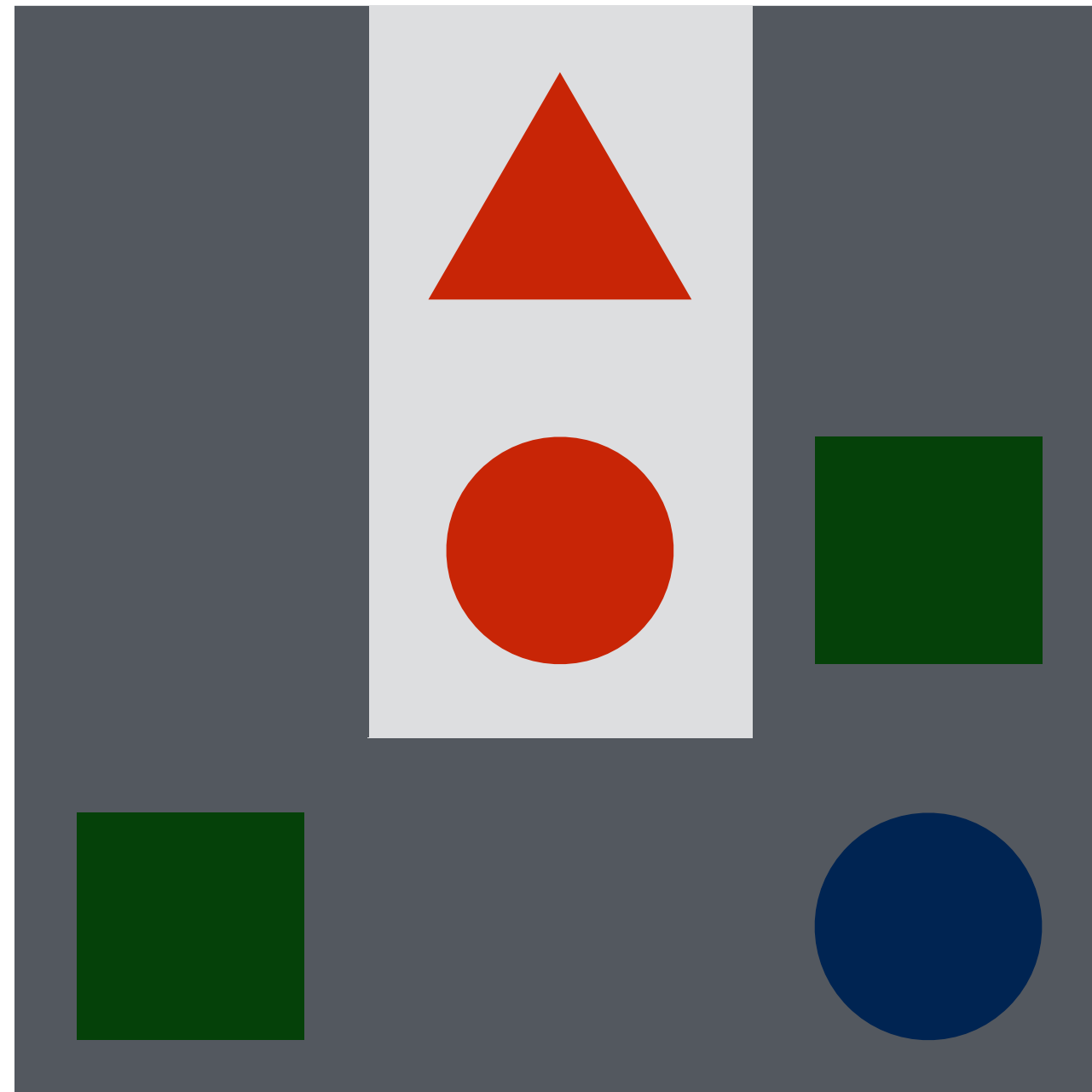


*Is there a **red** shape above a circle?*



# Sets encode meaning

---

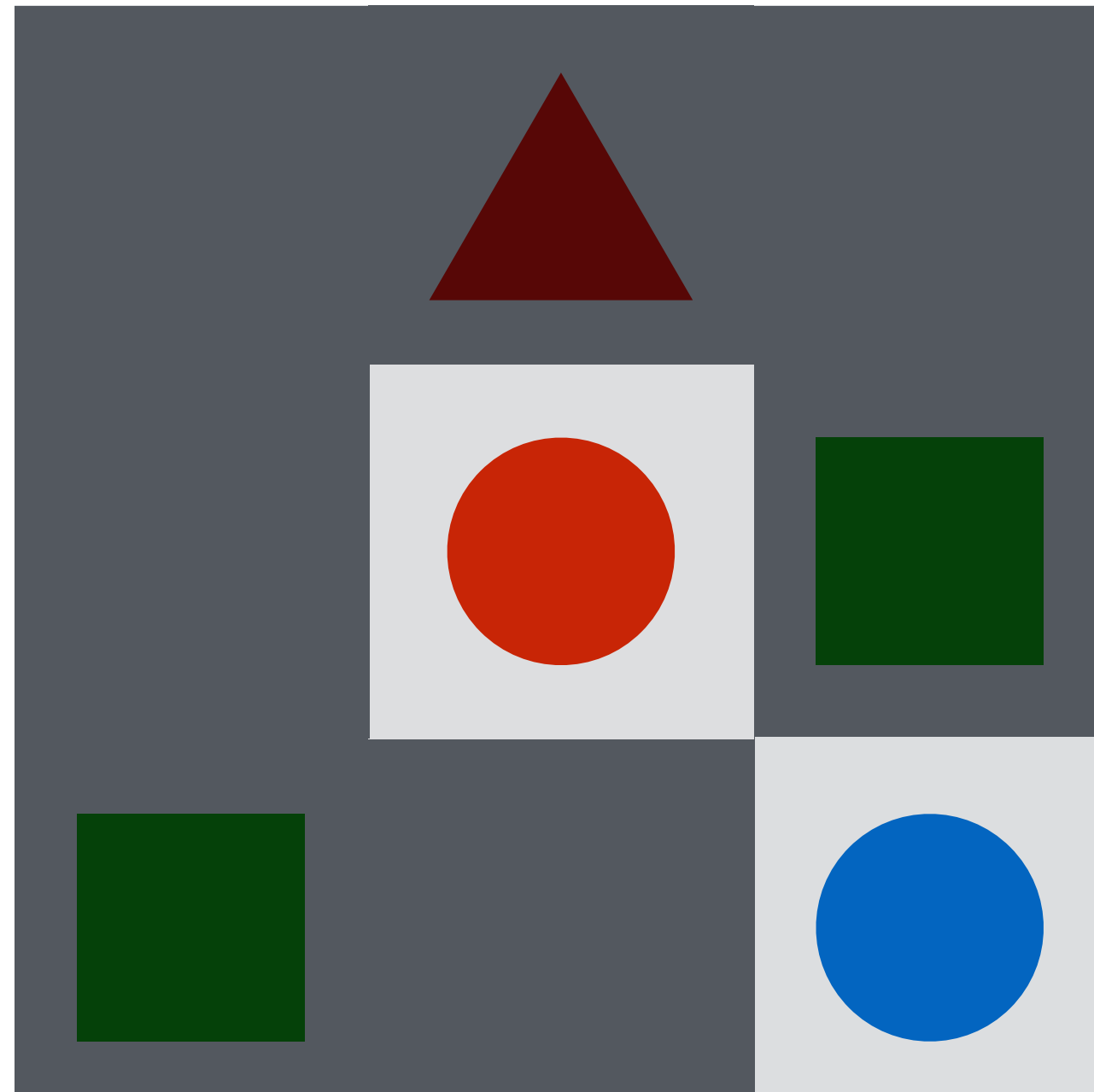


*Is there a **red** shape above a circle?*



# Sets encode meaning

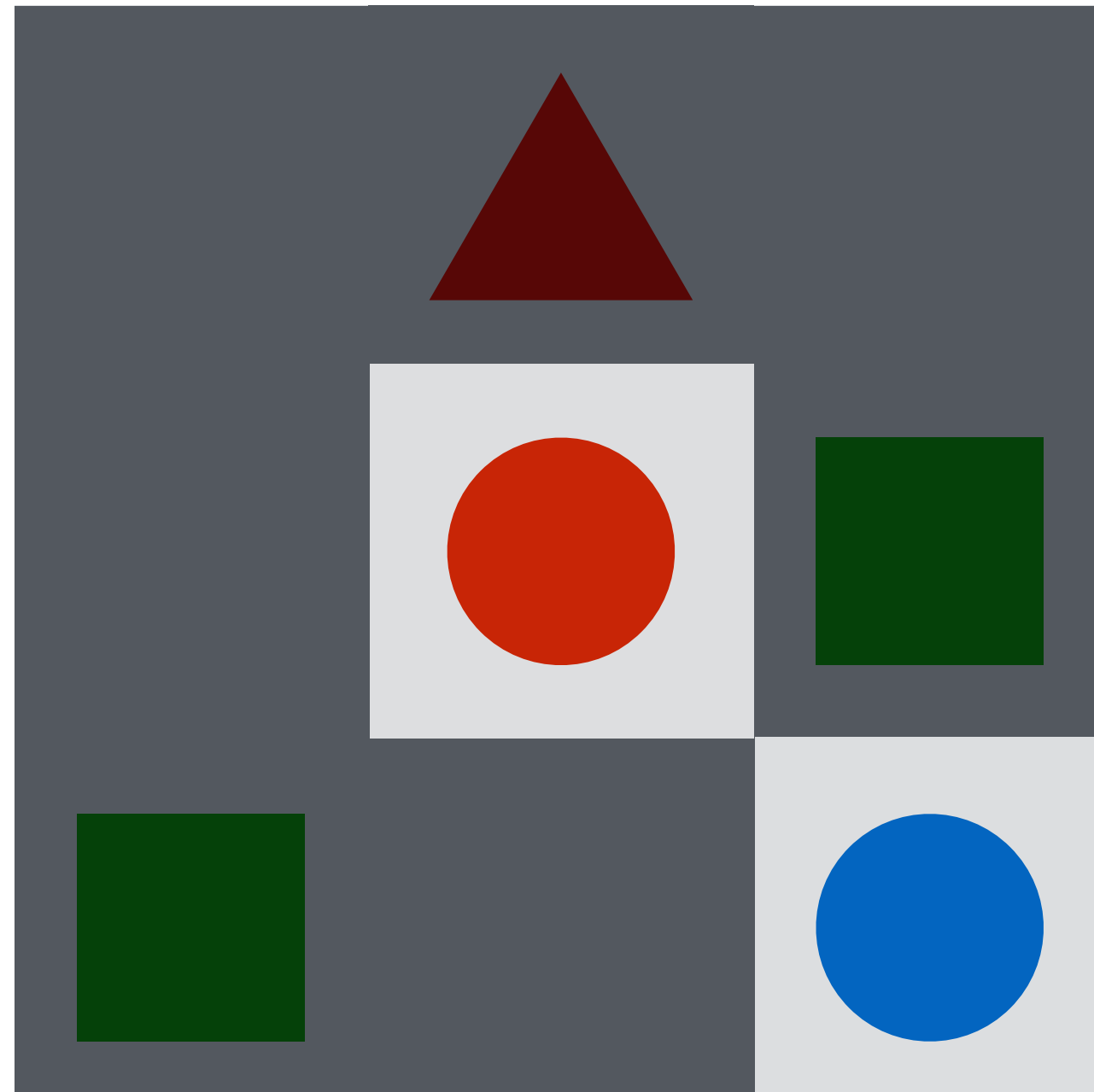
---



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



# Set transformations encode meaning

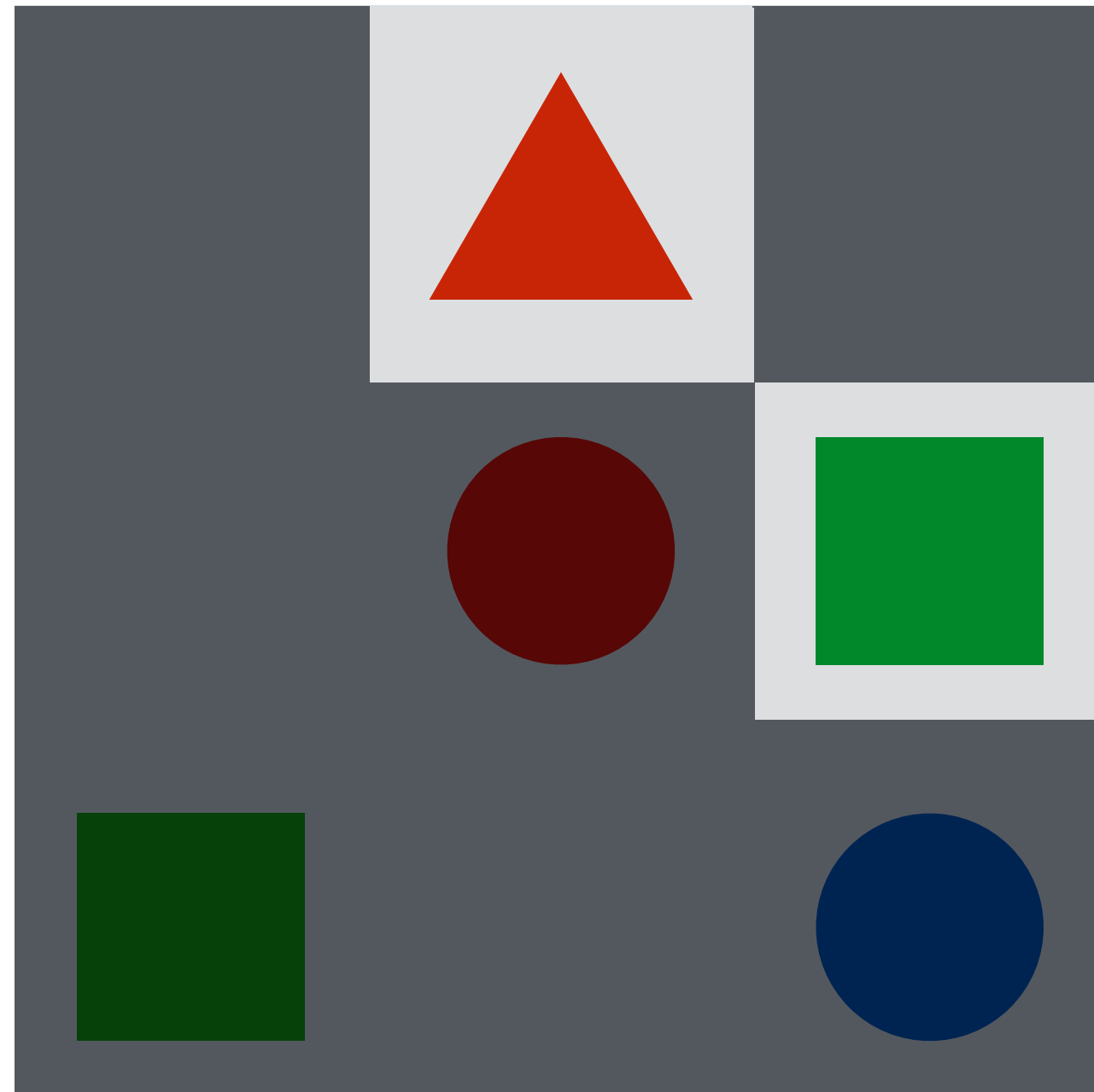


*Is there a red shape **above a circle**?*

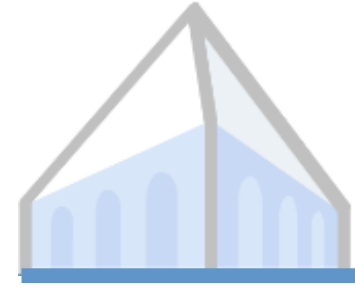




# Set transformations encode meaning

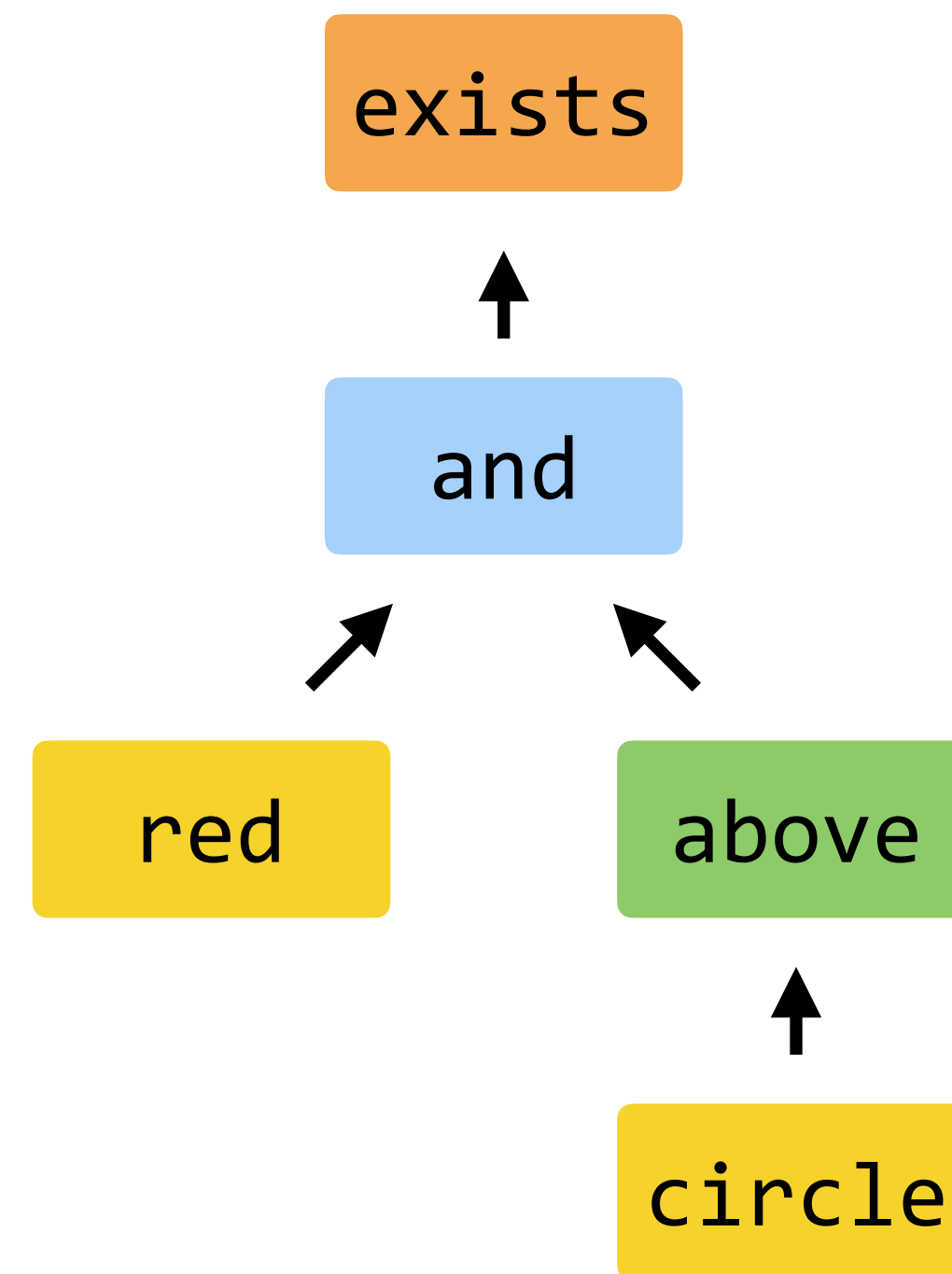
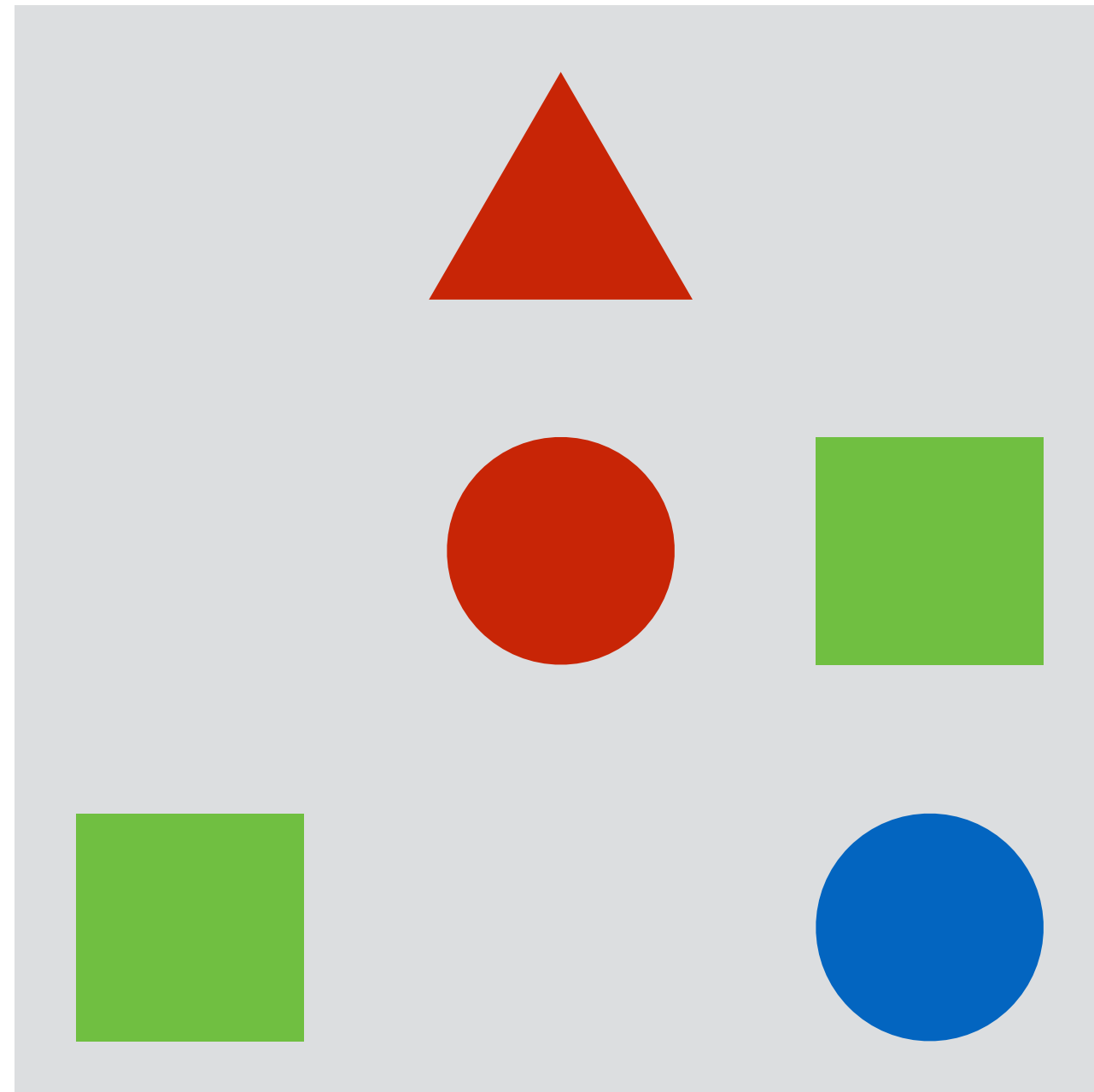


*Is there a red shape **above a circle**?*



# Sentence meanings are computations

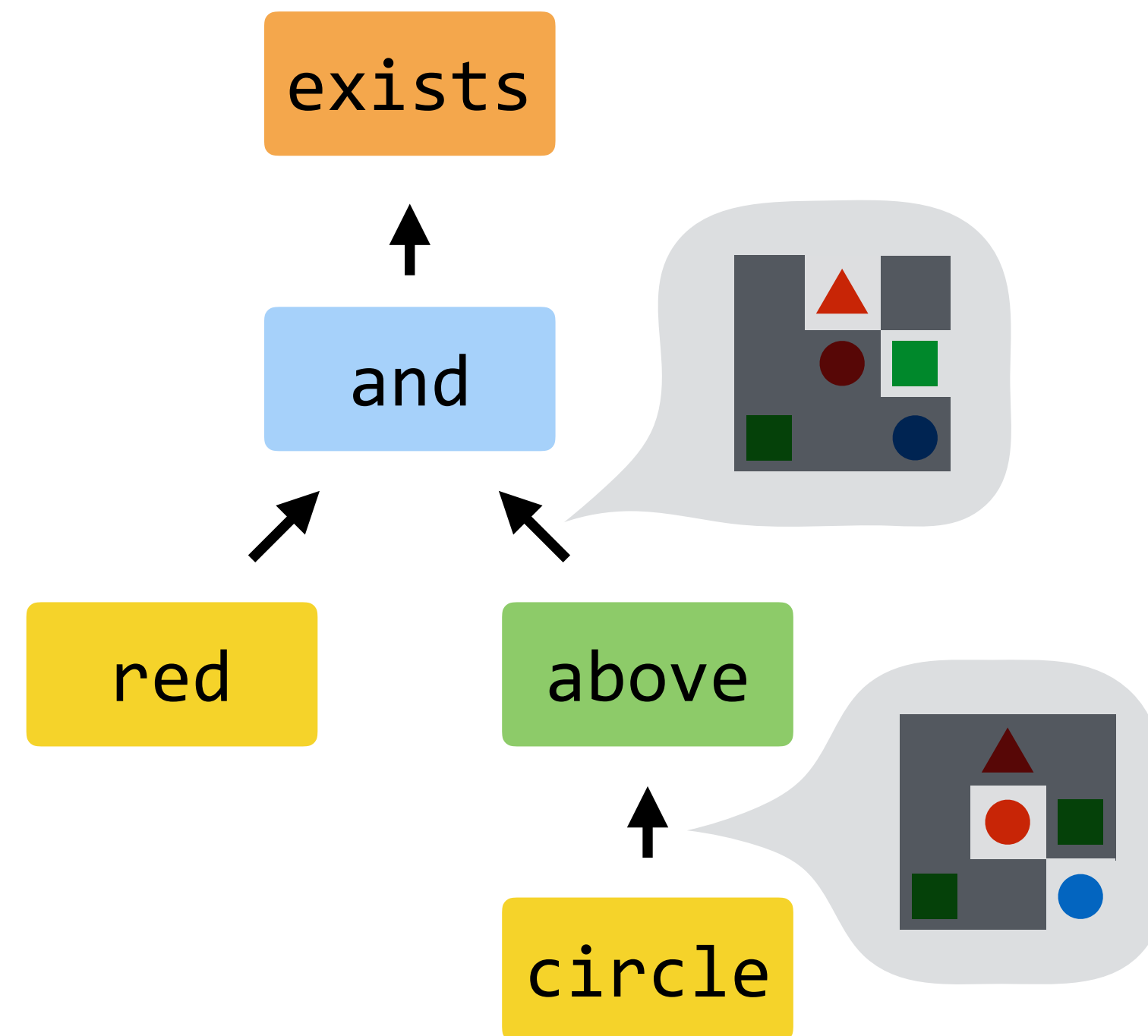
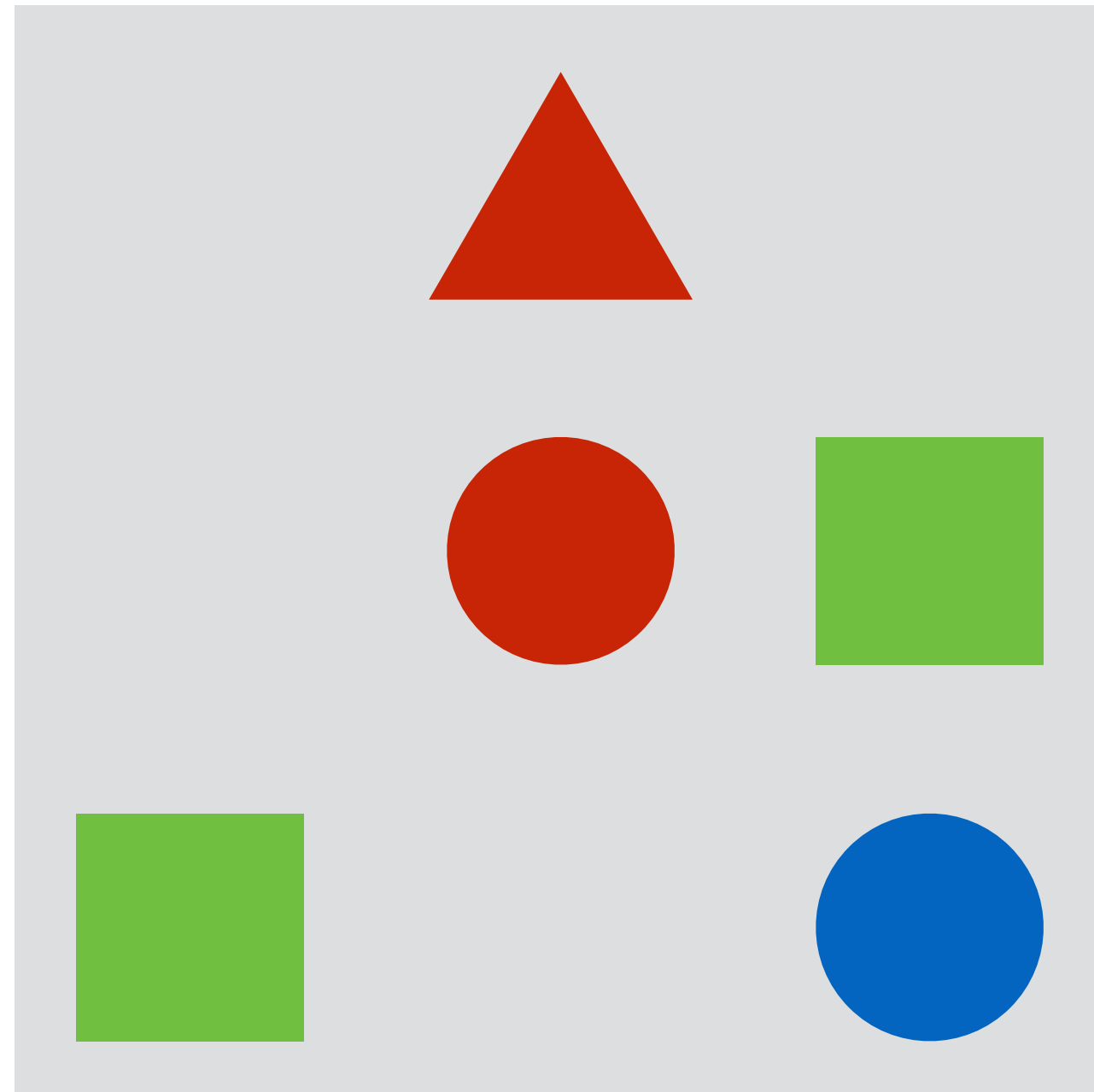
*Is there a red shape above a circle?*





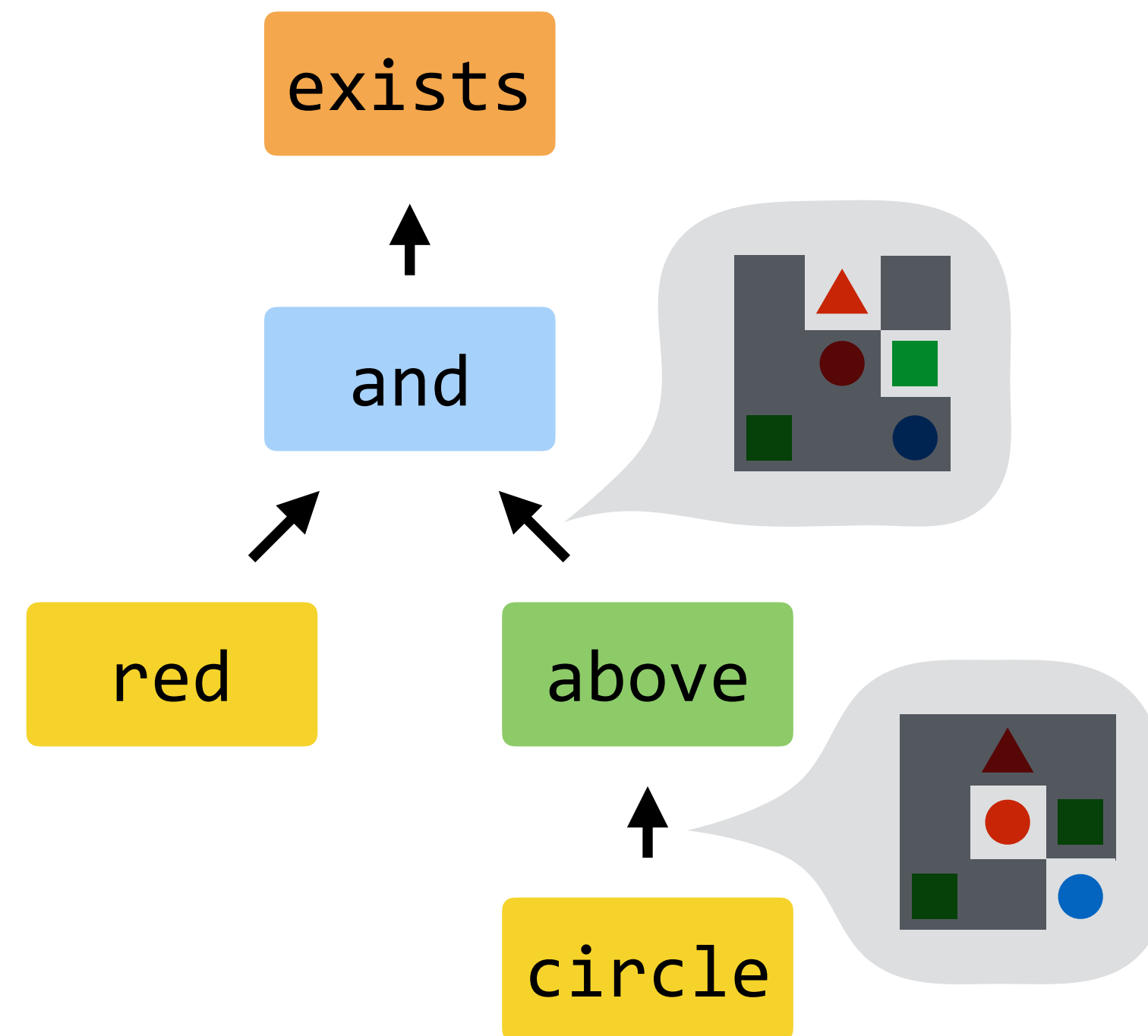
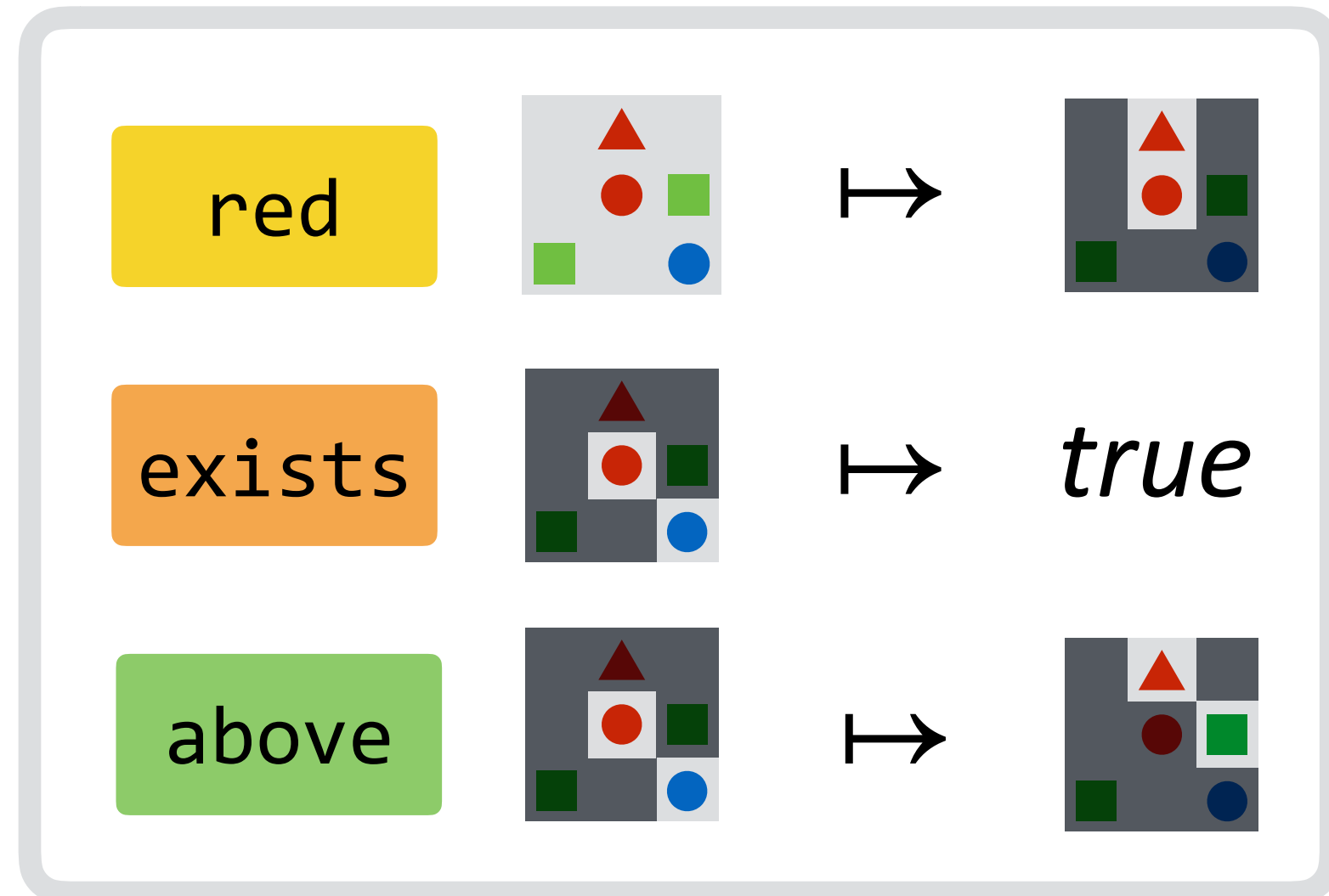
# Sentence meanings are computations

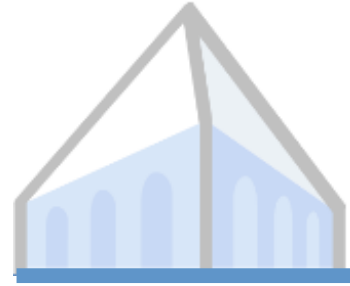
*Is there a red shape above a circle?*





# Computations are built from set functions



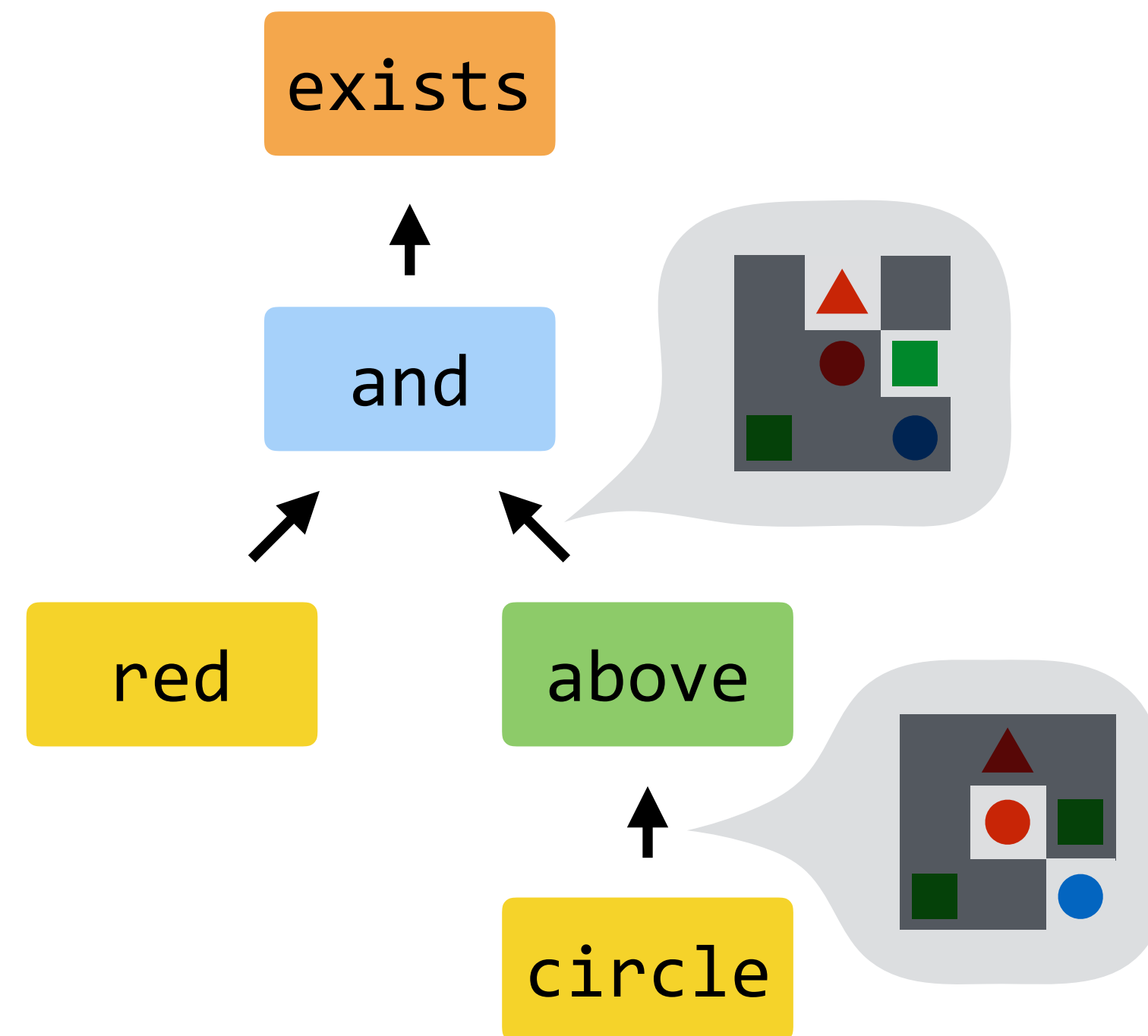


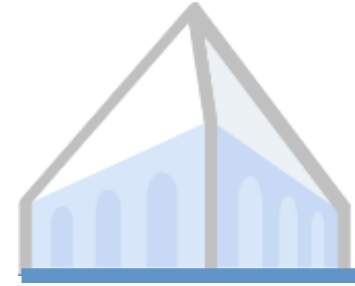
# Modules can be manually specified...

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

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

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



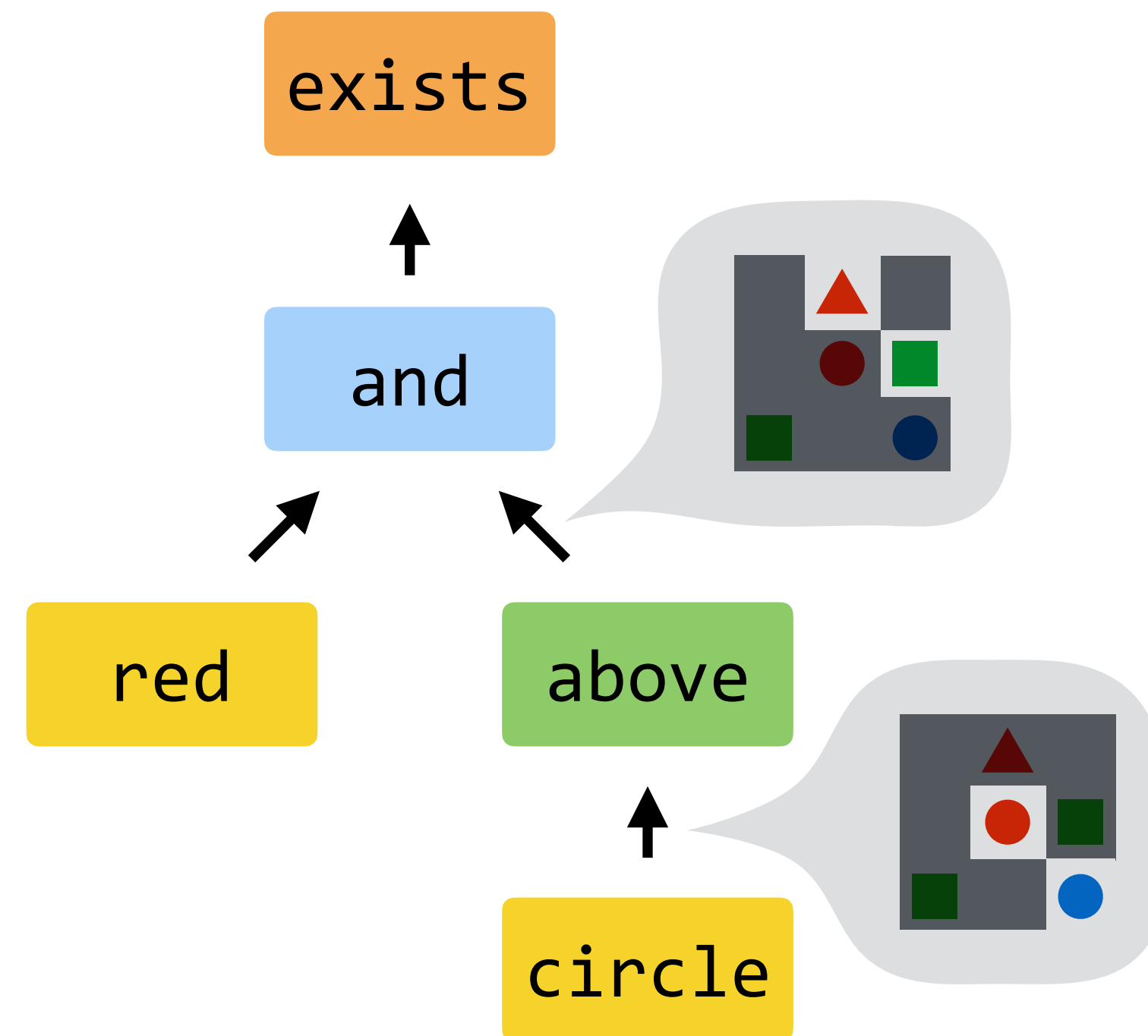


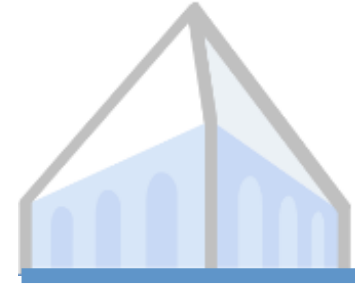
# ...or learned as classifiers...

$$x \mapsto \text{sign}(a^\top \phi(x))$$

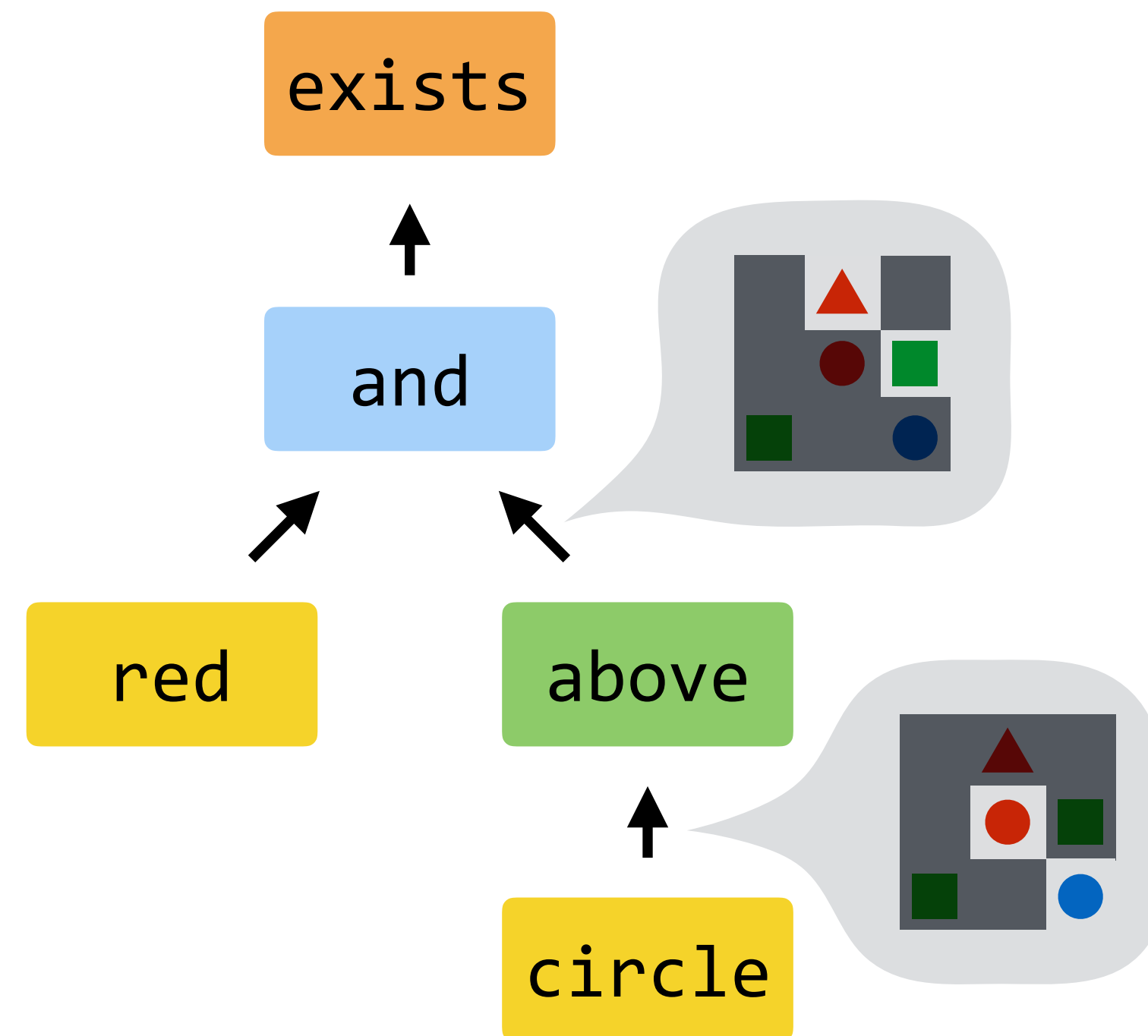
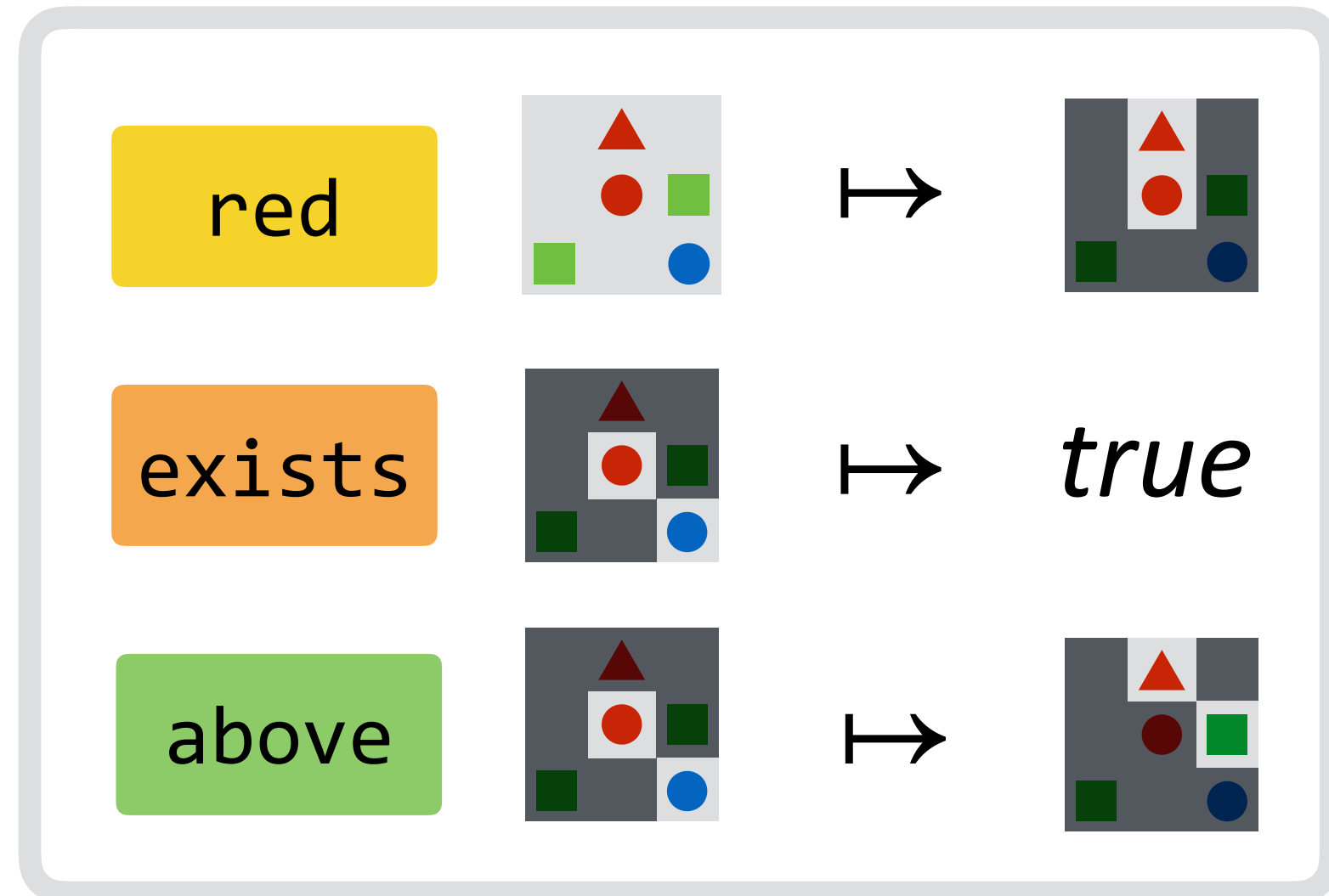
$$x \mapsto \text{sign}(b^\top \phi(x))$$

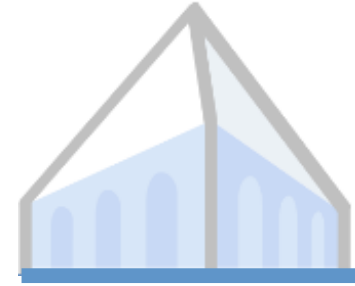
$$x \mapsto \text{sign}(c^\top \phi(x))$$



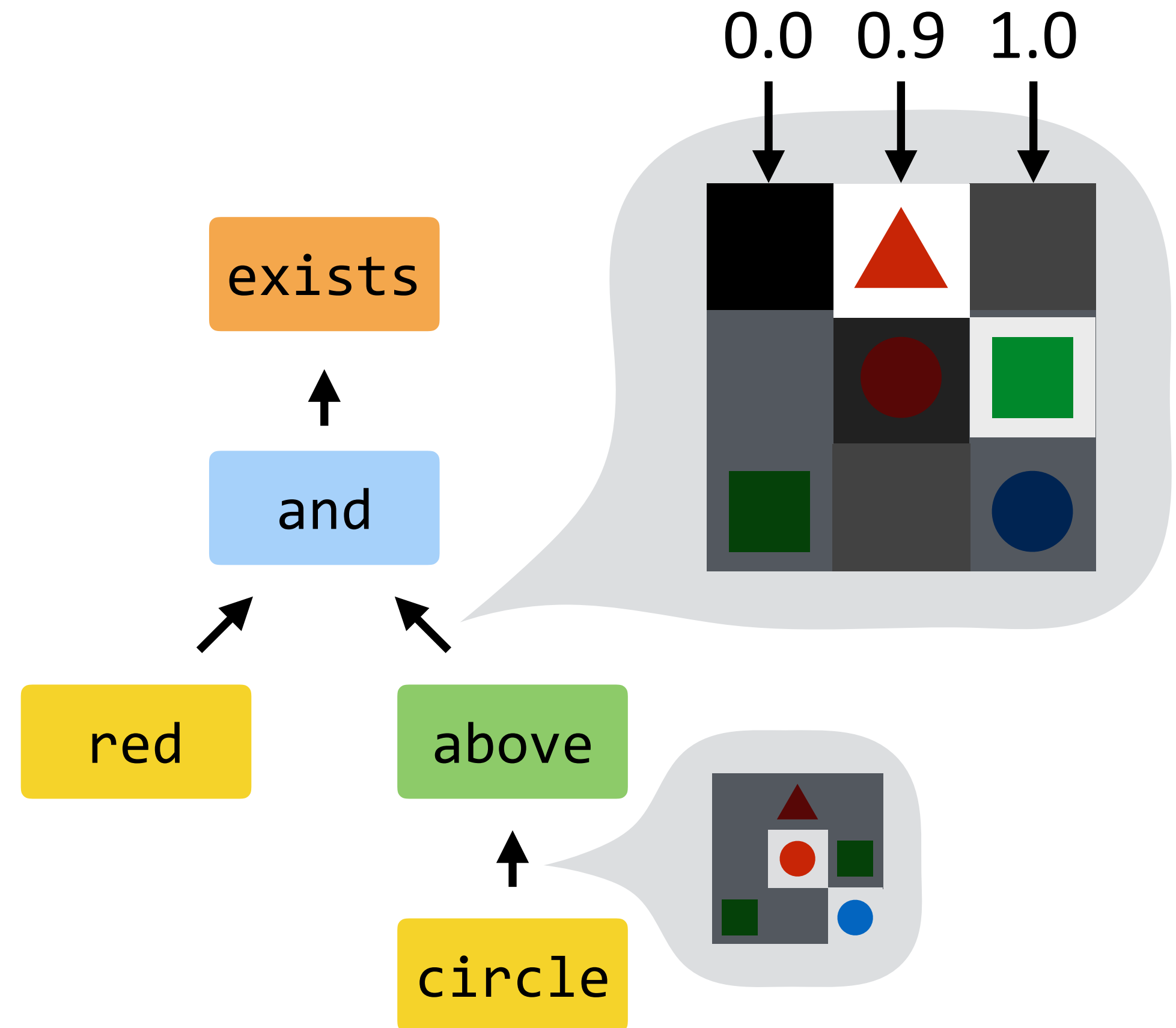
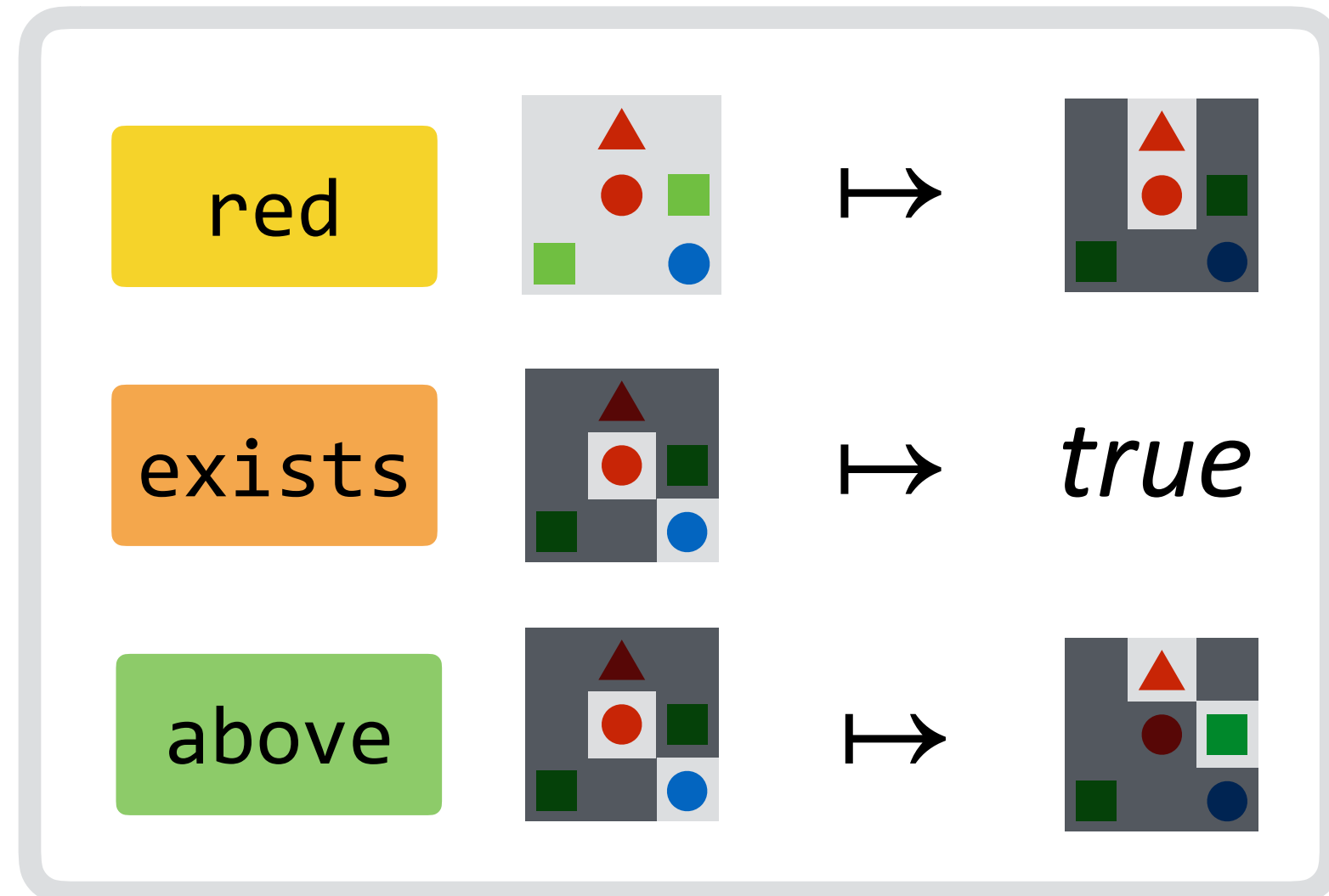


# ...or relaxed to real-valued vectors





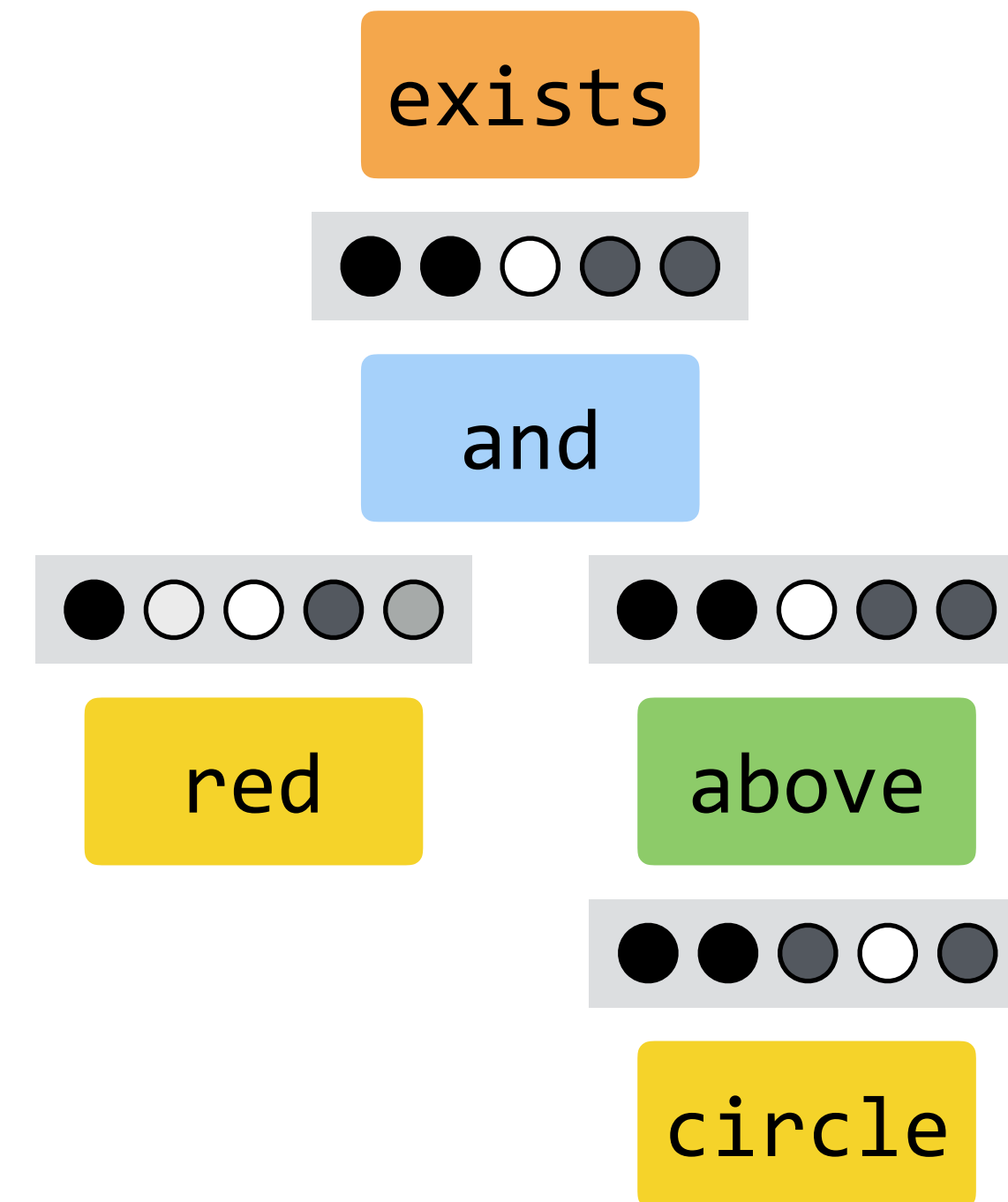
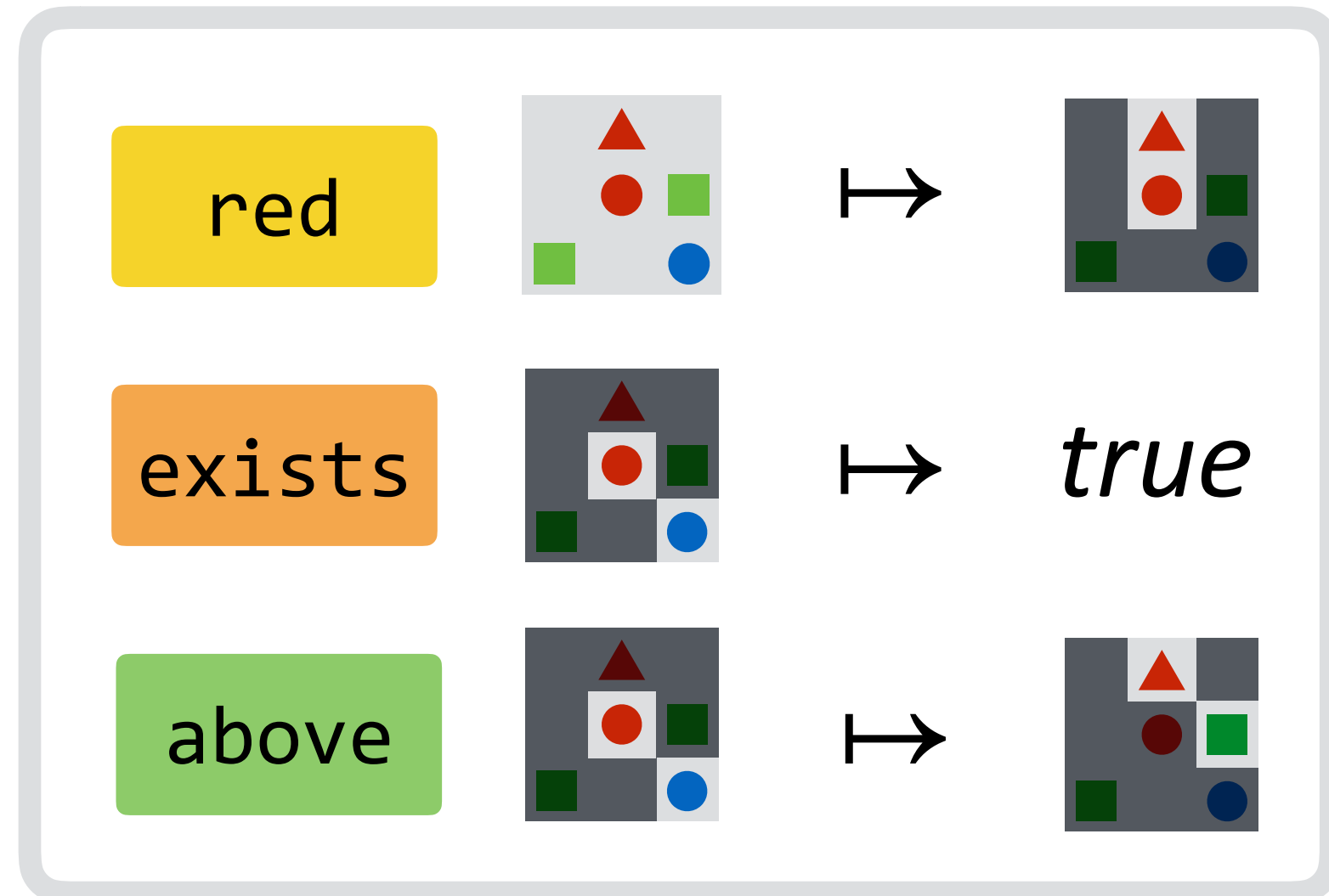
# ...or relaxed to real-valued vectors

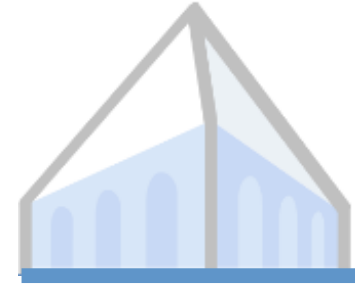




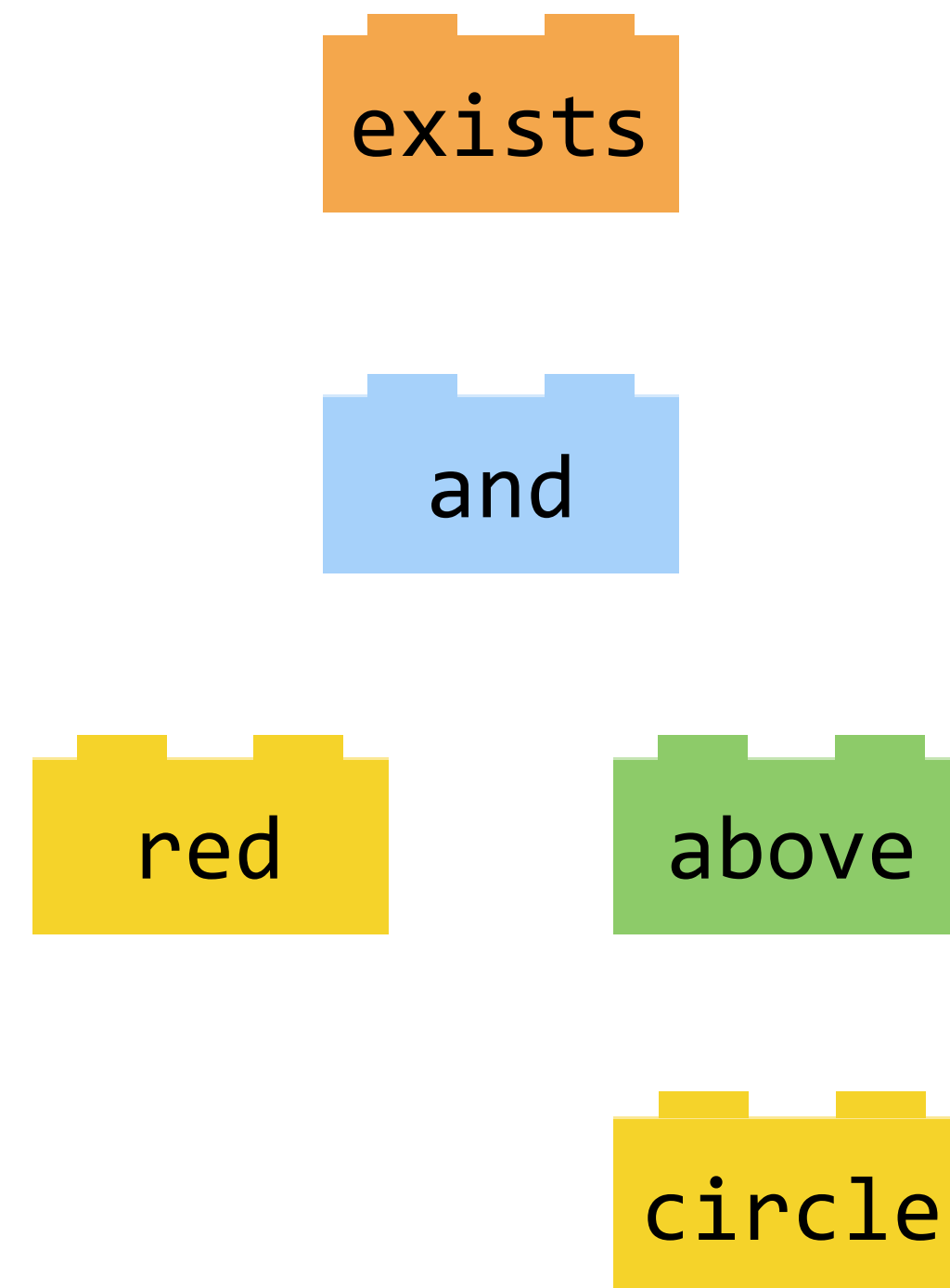
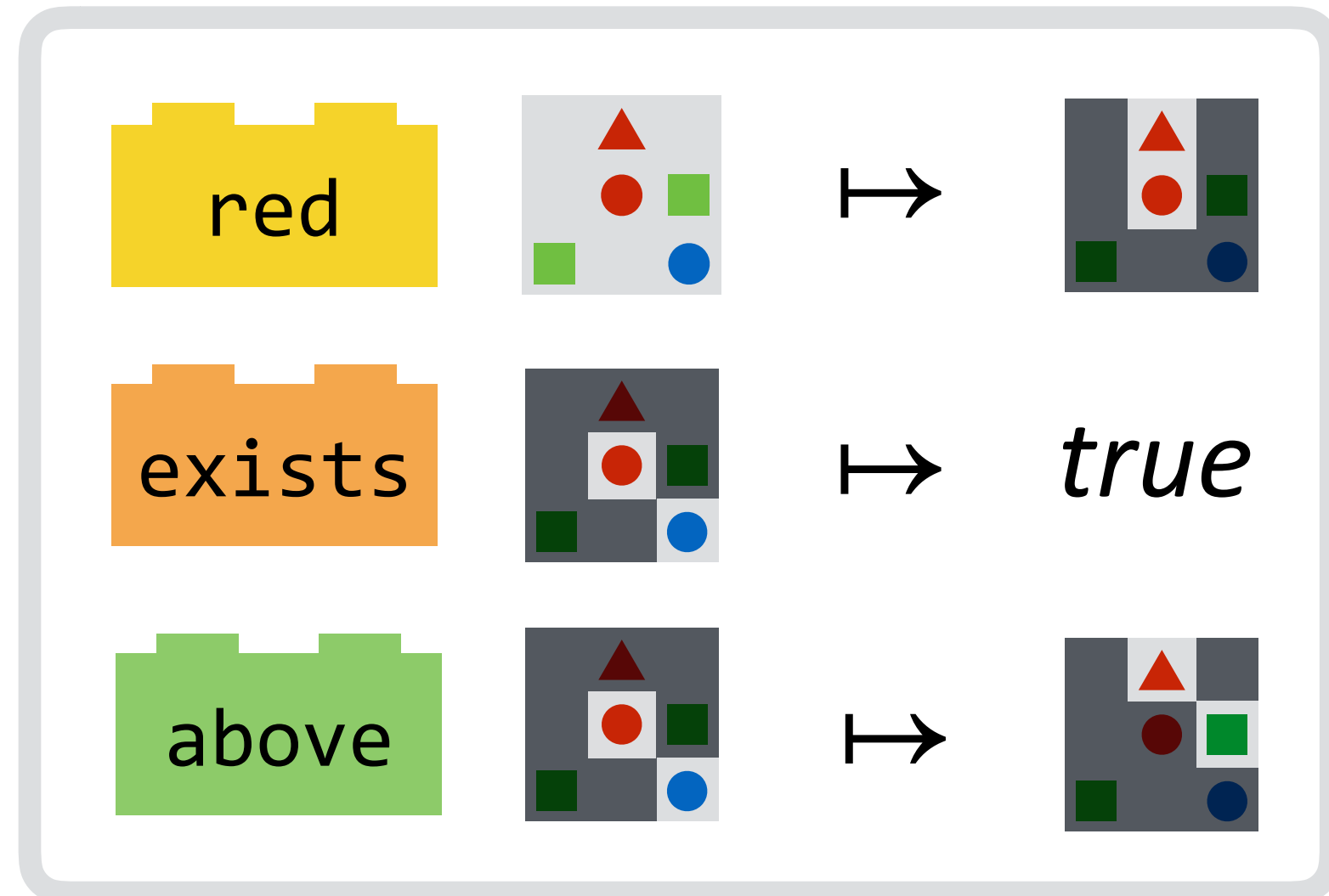


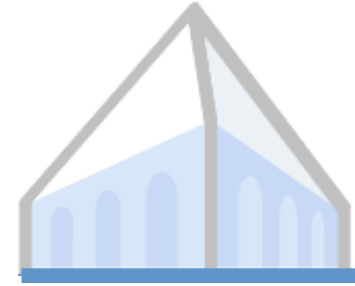
# Composing vector functions



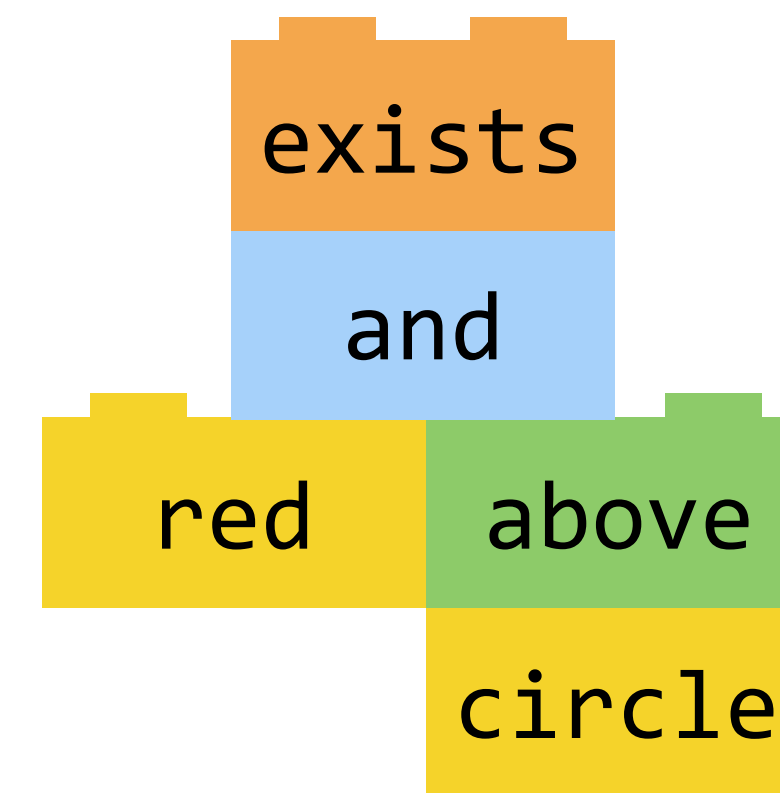
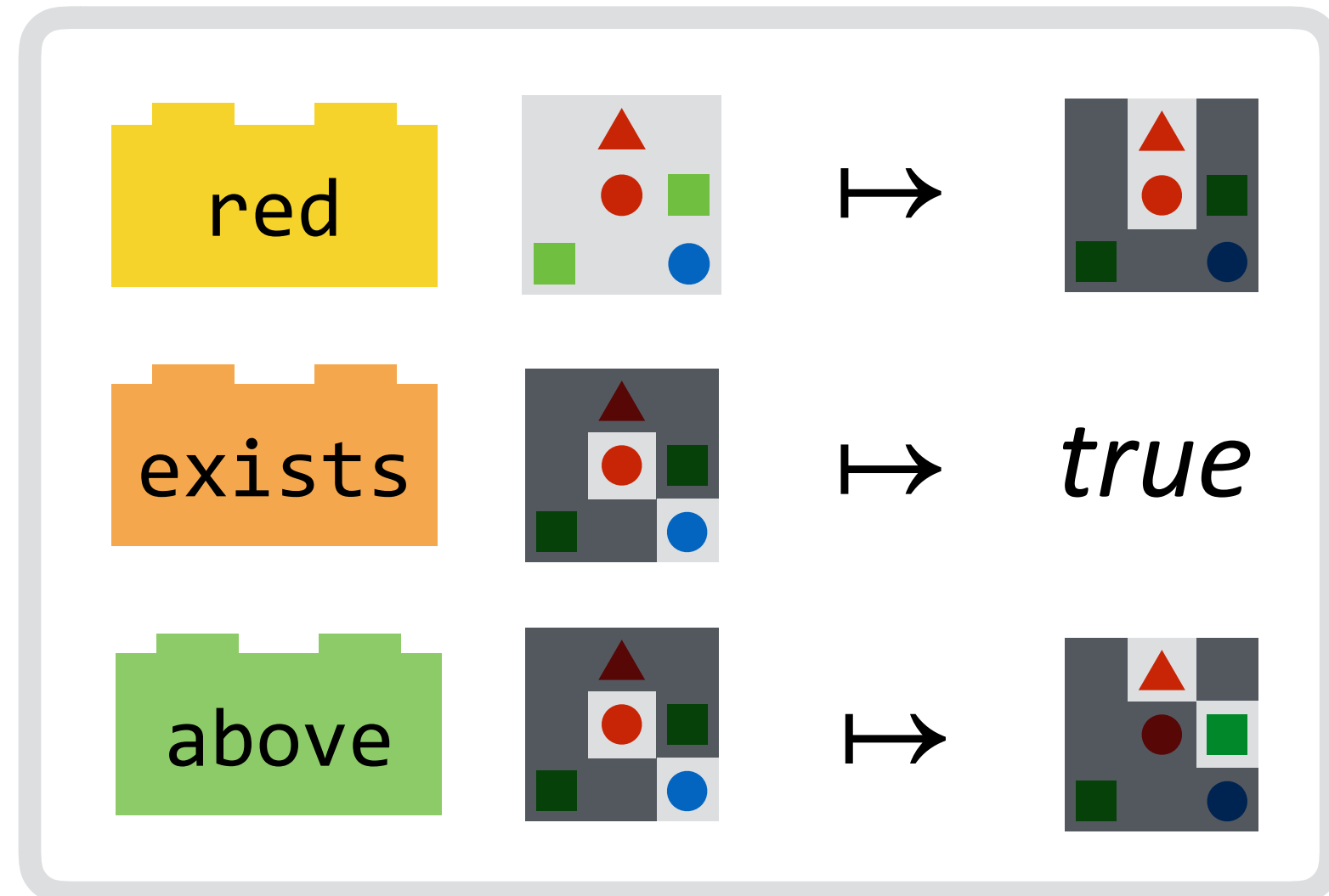


# Composing vector functions



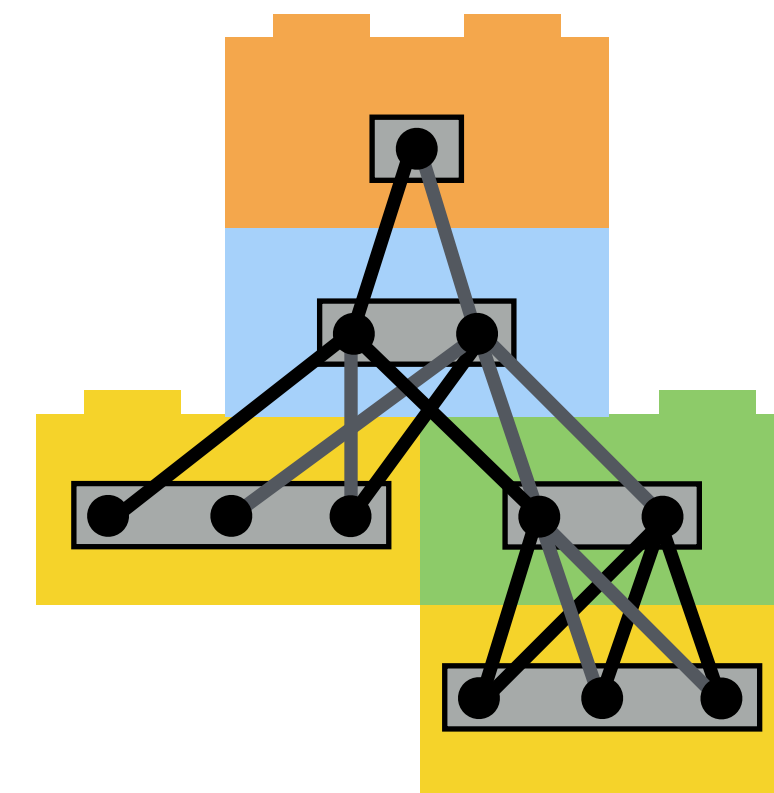
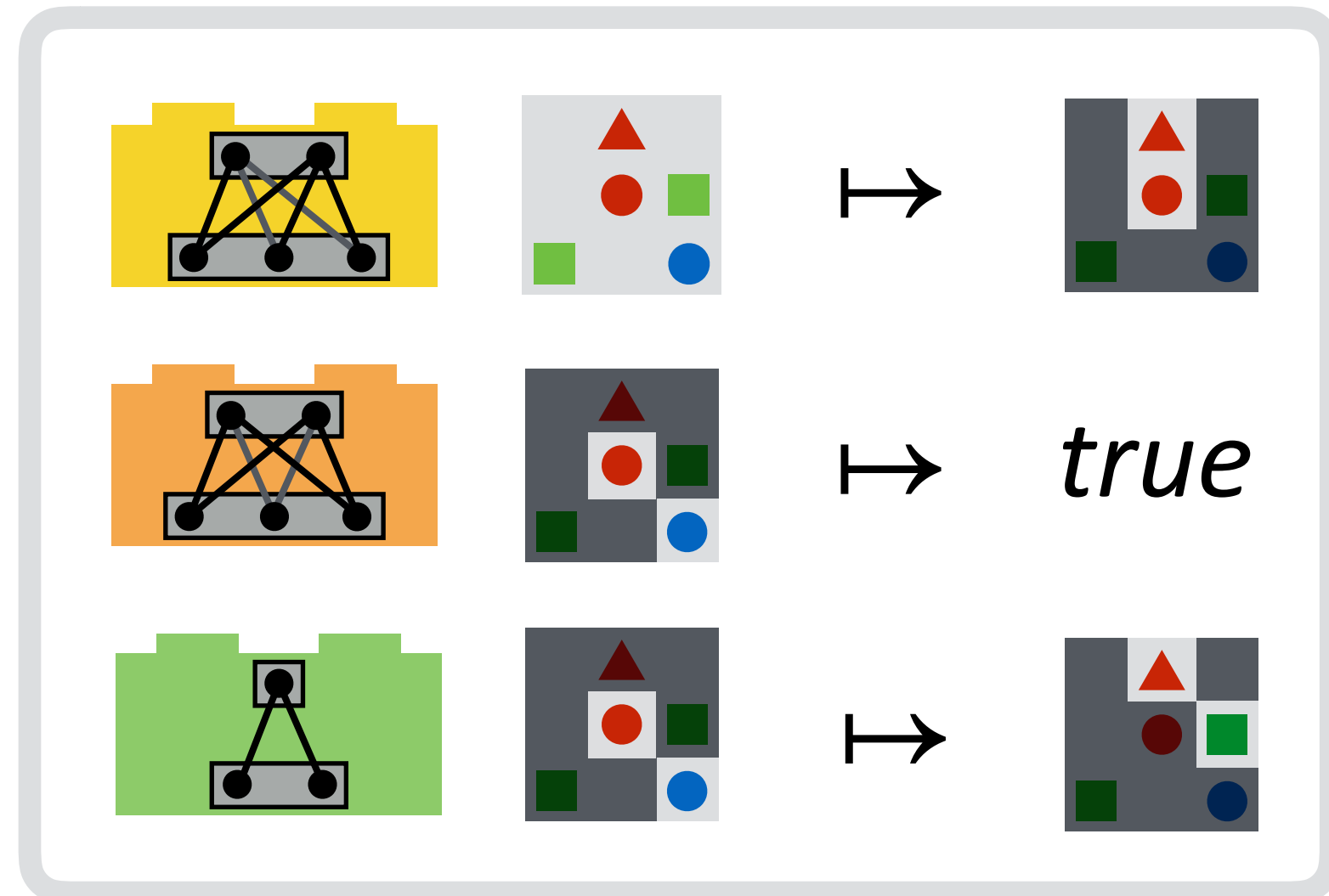


# Composing vector functions



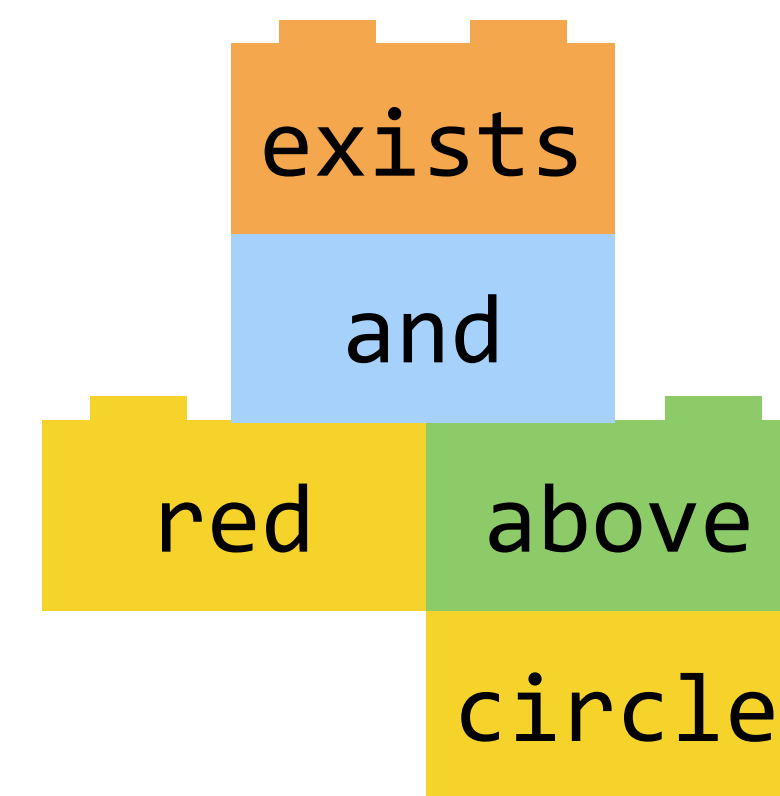
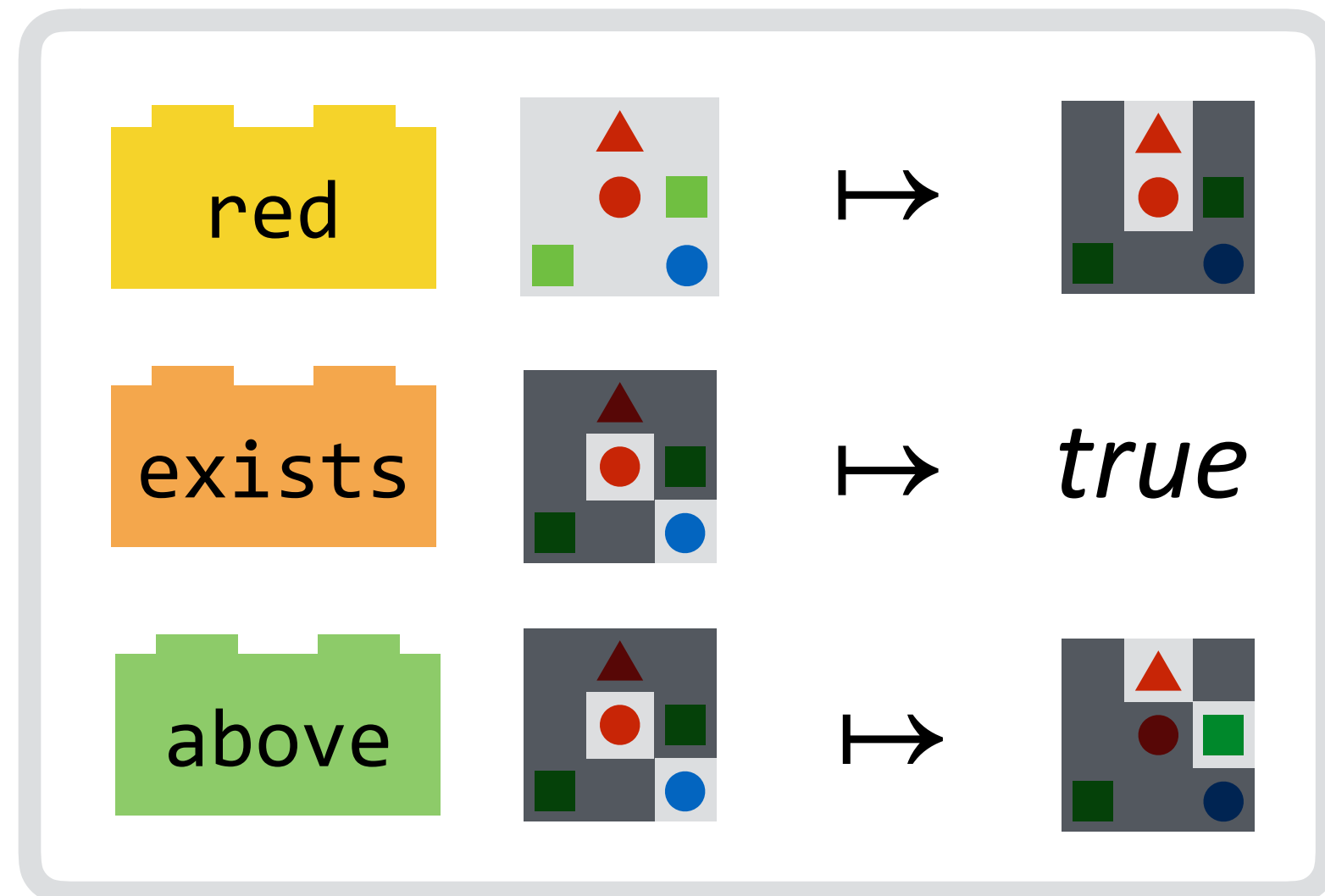


# Compositions of vector functions are neural nets





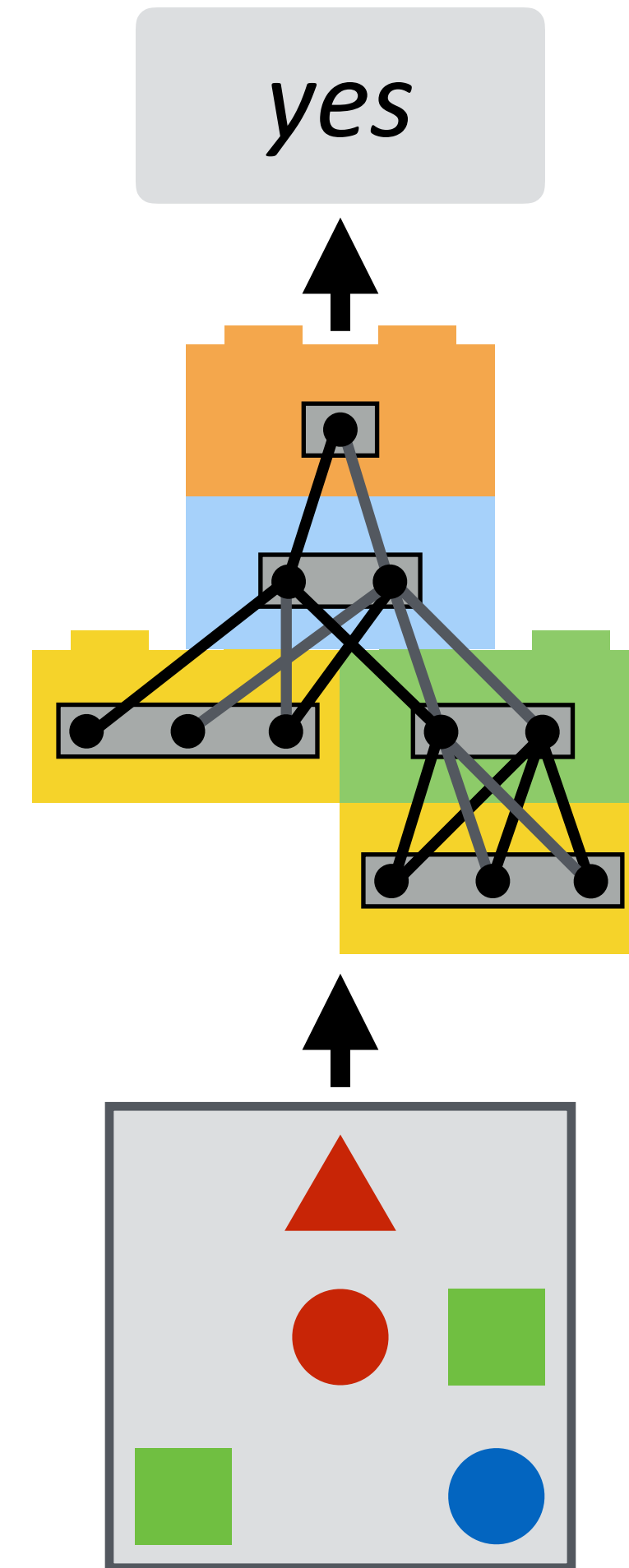
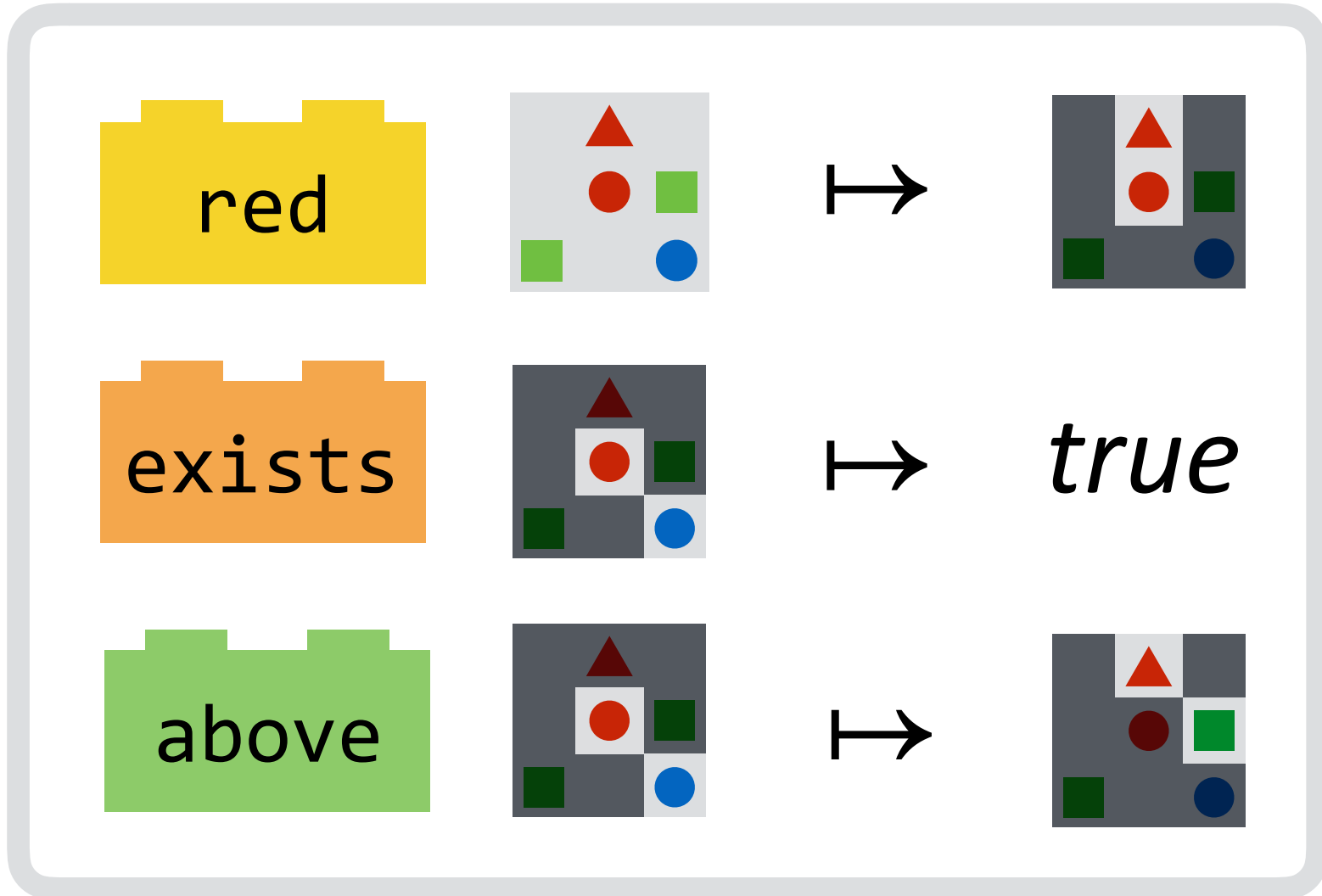
# Compositions of vector functions are neural nets





# Outline

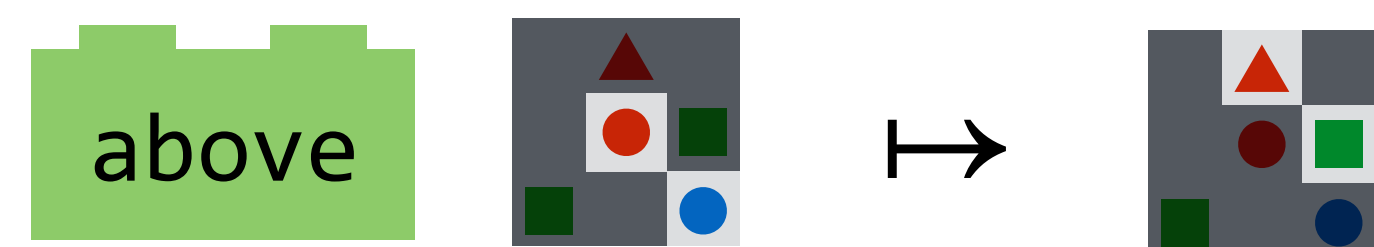
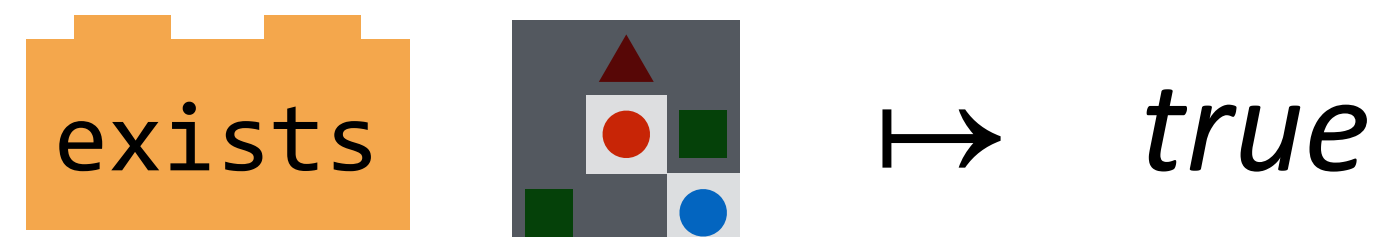
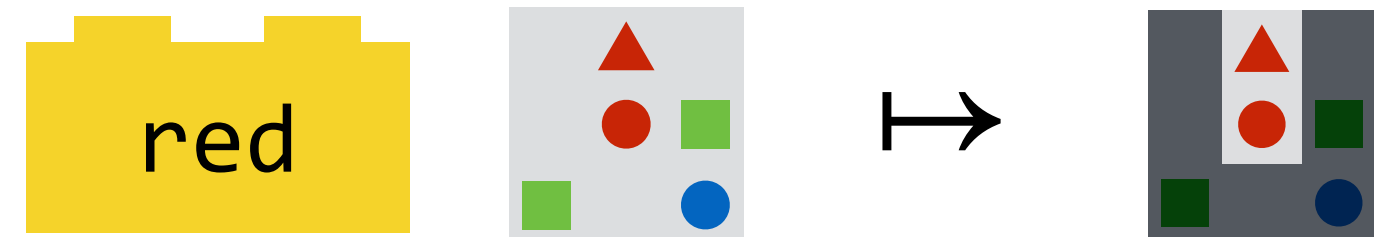
*Is there a red shape above a circle?*



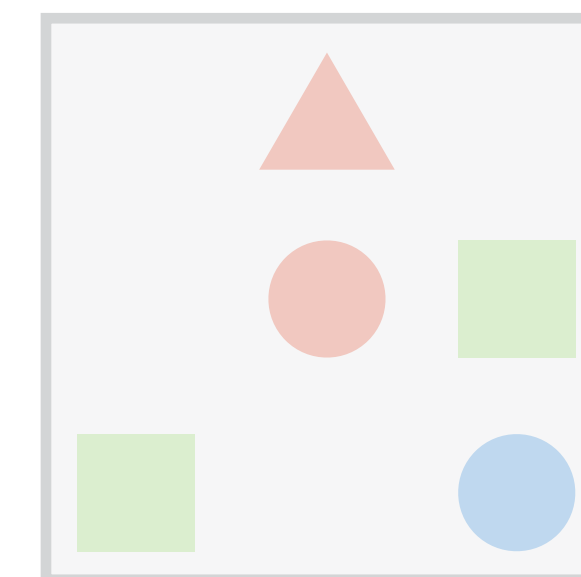


# Outline

*Is there a red shape  
above a circle?*



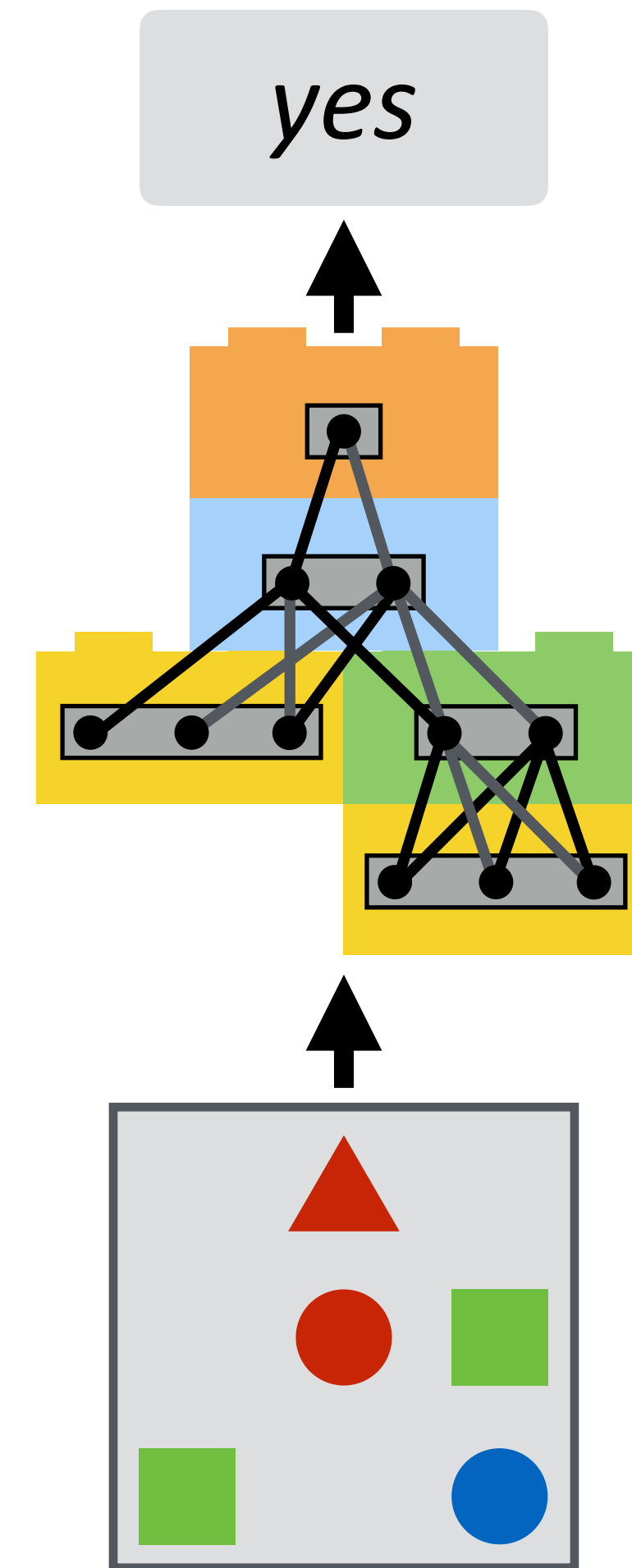
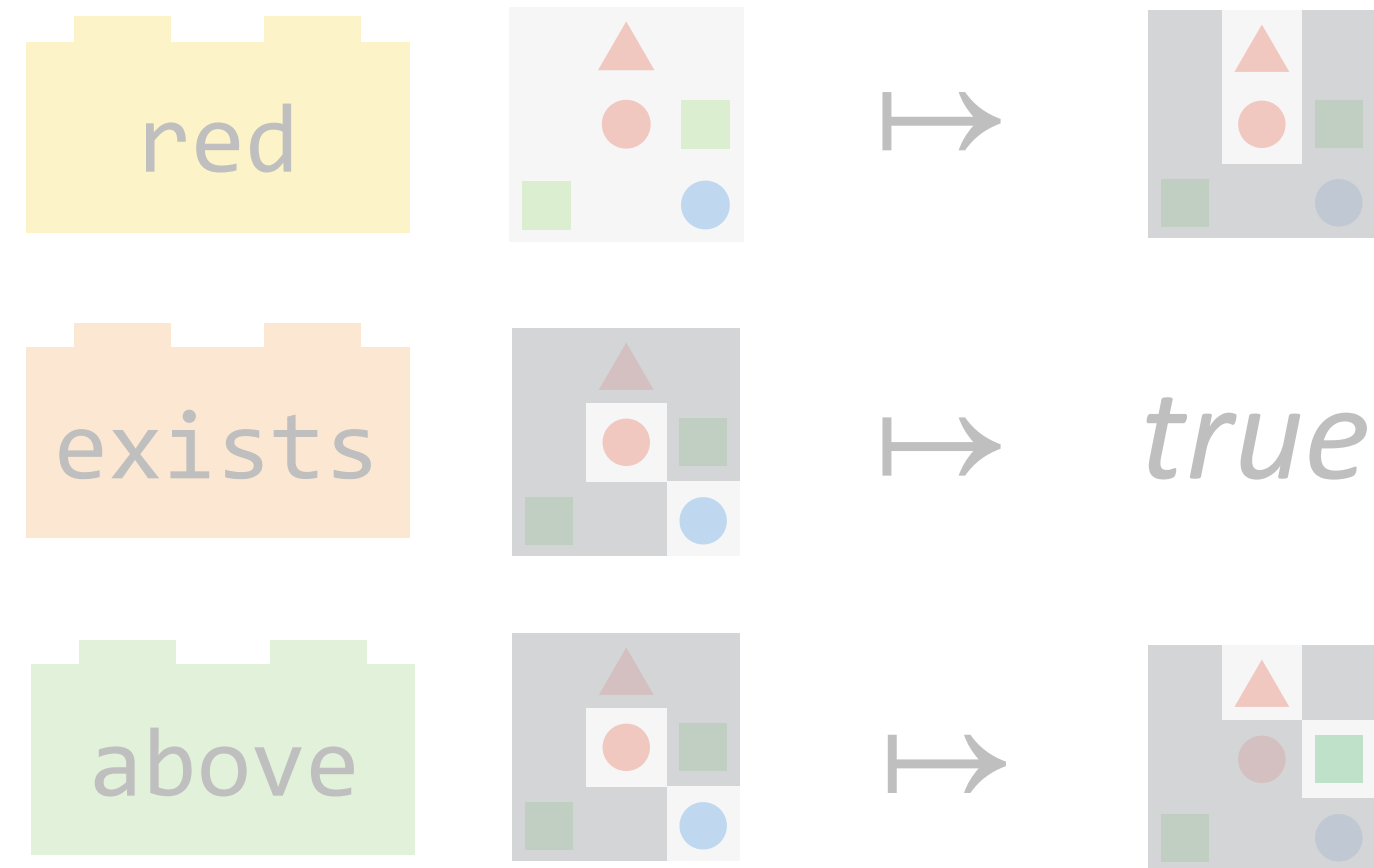
*yes*





# Outline

*Is there a red shape  
above a circle?*



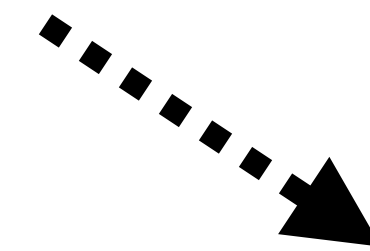




# Outline

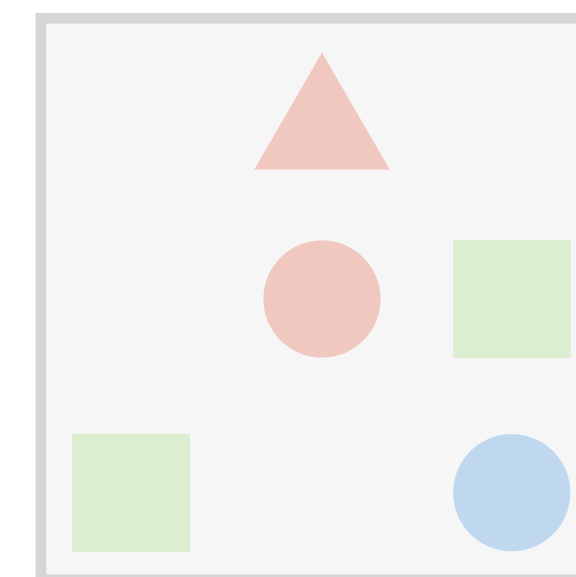
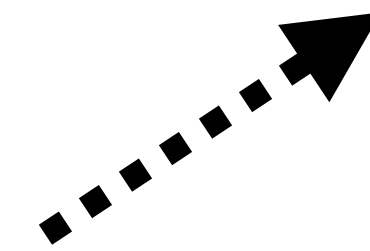
*Is there a red shape  
above a circle?*

red		→	
exists		→	<i>true</i>
above		→	



*yes*

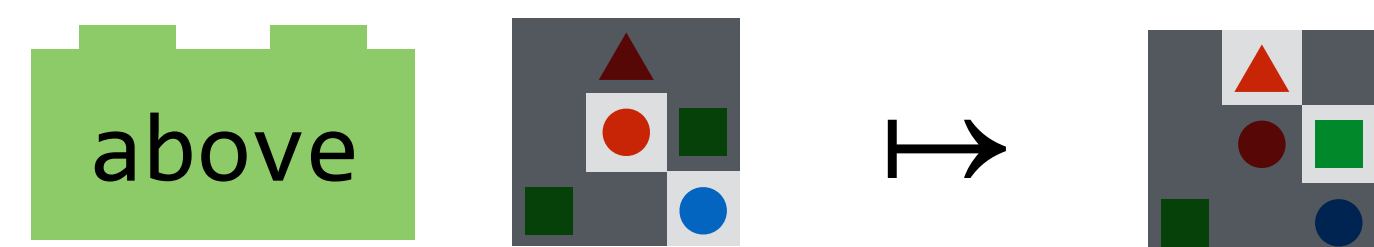
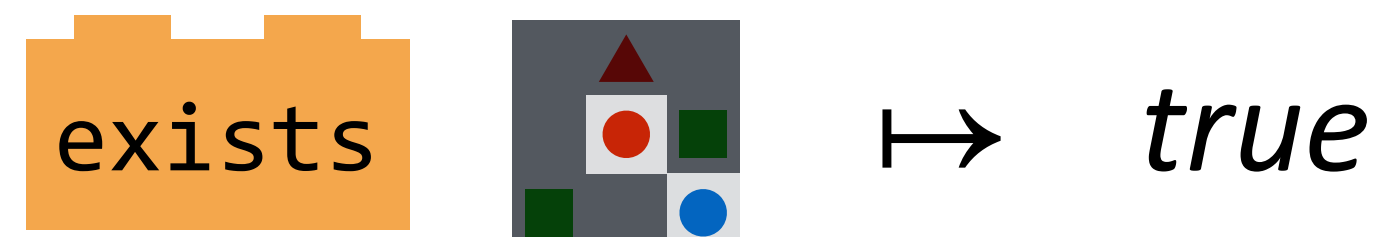
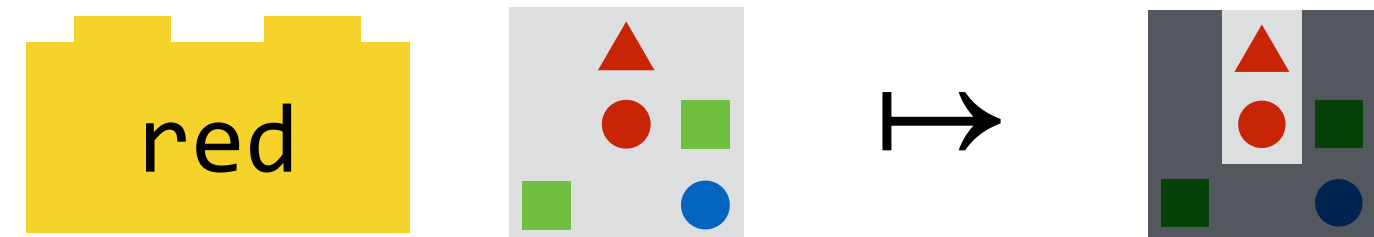
exists	
and	
red	above
	circle



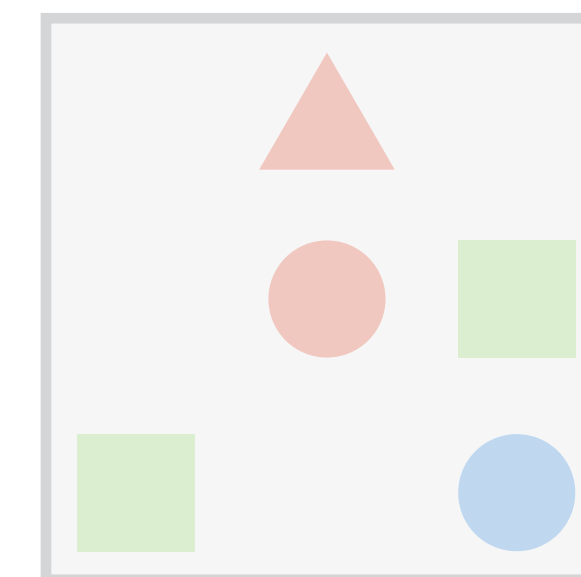


# Outline

*Is there a red shape  
above a circle?*



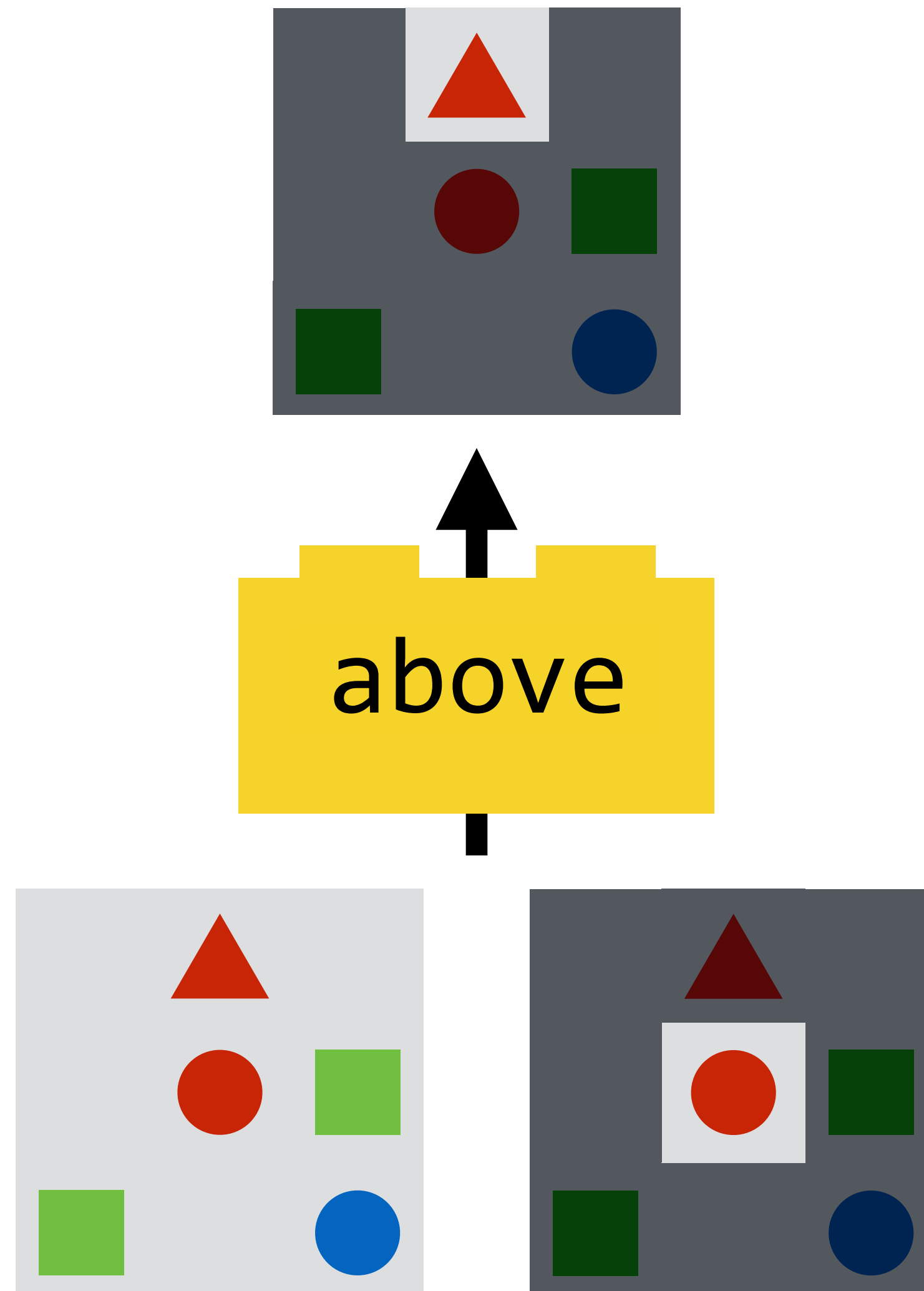
*yes*

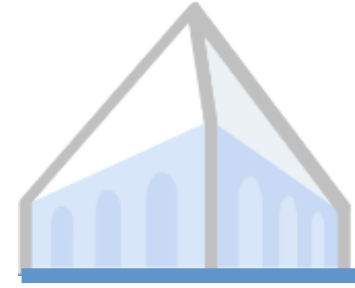




# Anatomy of a module

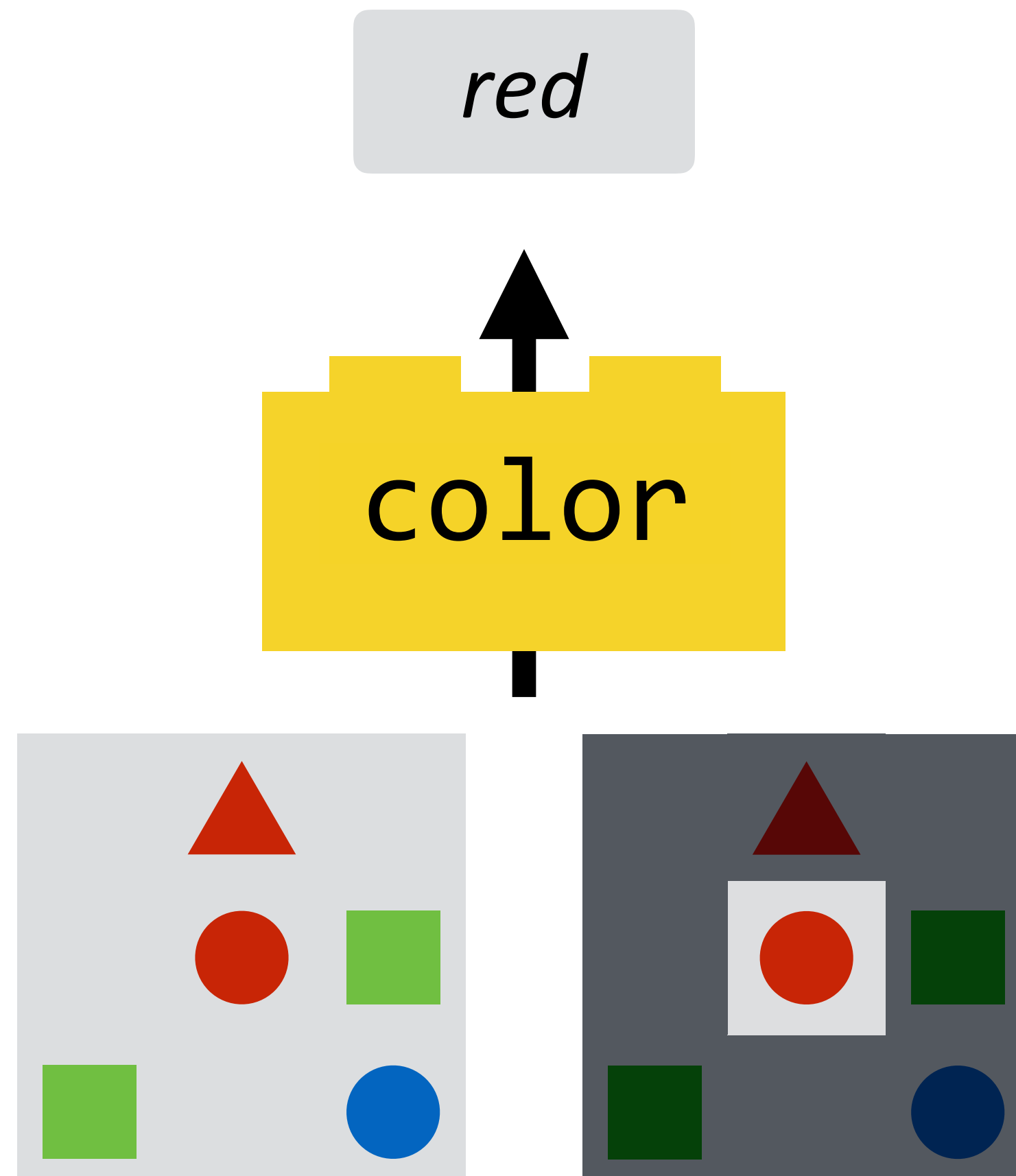
---





# Anatomy of a module

---





# What modules do we need?

---

*Is there a red shape above a circle?*

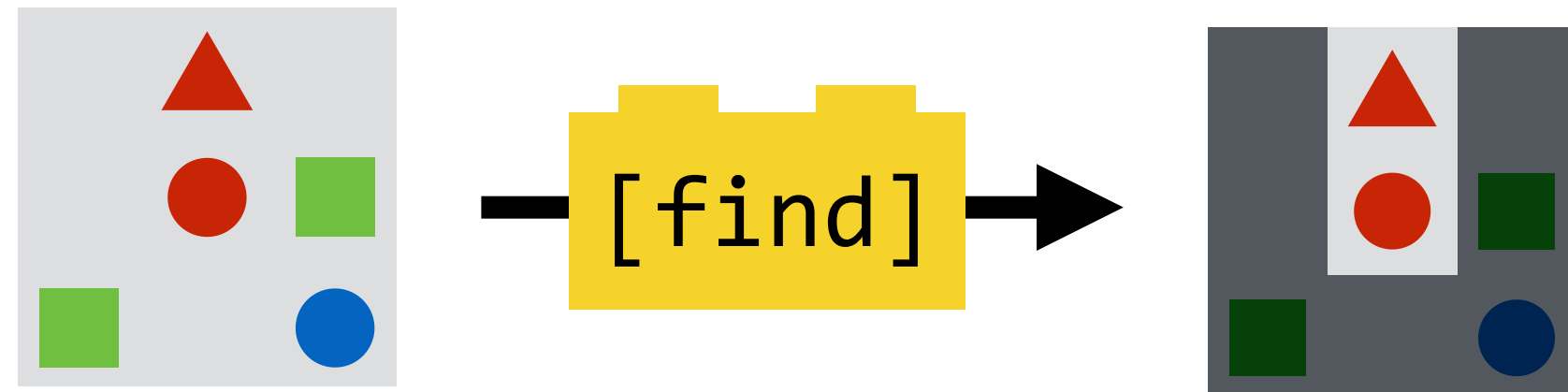
*What color is the triangle?*

*Who is running in the grass?*

*What cities are south of San Diego?*



# A module for predicates

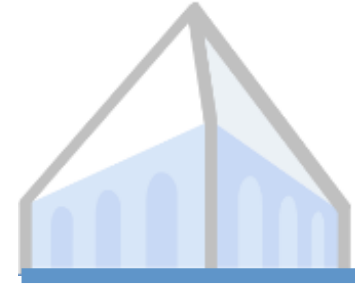


*Is there a **red** shape above a **circle**?*

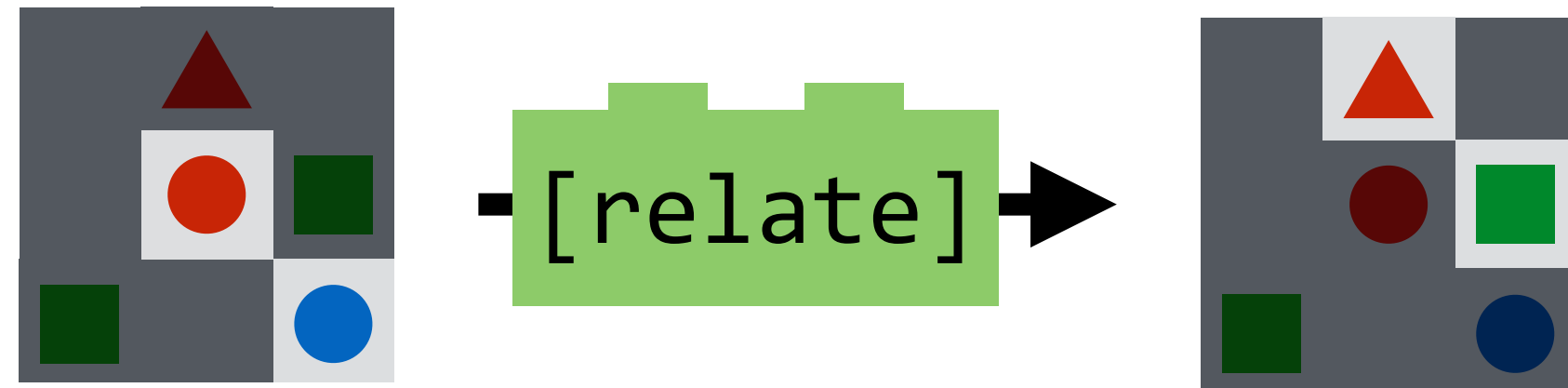
*What color is the **triangle**?*

*Who is **running** in the **grass**?*

*What **cities** are south of **San Diego**?*



# A module for relations



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

*What color is the triangle?*

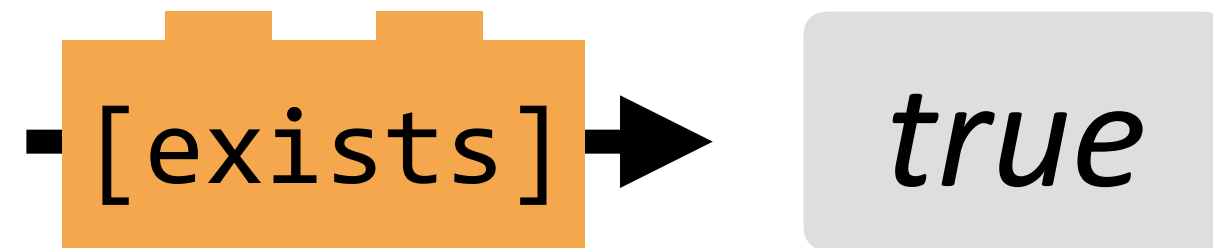
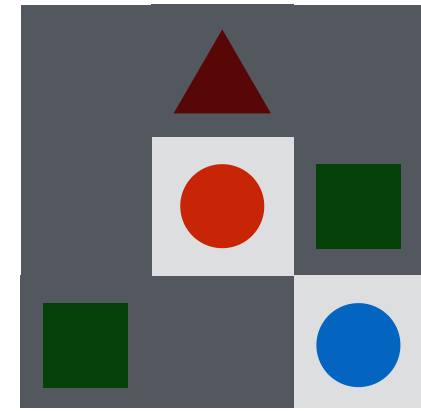
*Who is running **in** the grass?*

**[find]**

*What cities are **south of** San Diego?*



# A module for quantifiers



*Is there a red shape above a circle?*

*What color is the triangle?*

*Who is running in the grass?*

[find]

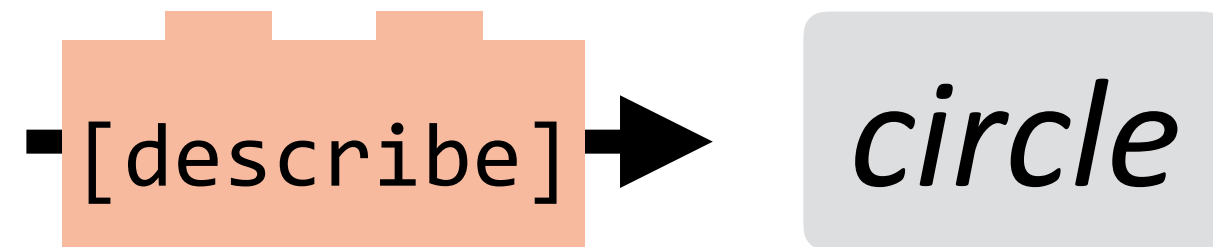
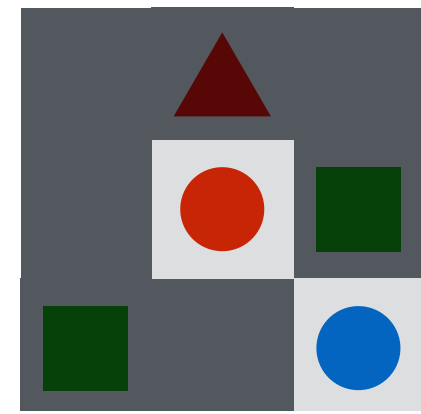
[relate]

*What cities are south of San Diego?*





# A module for attributes



*Is there a red shape above a circle?*

*What color is the triangle?*

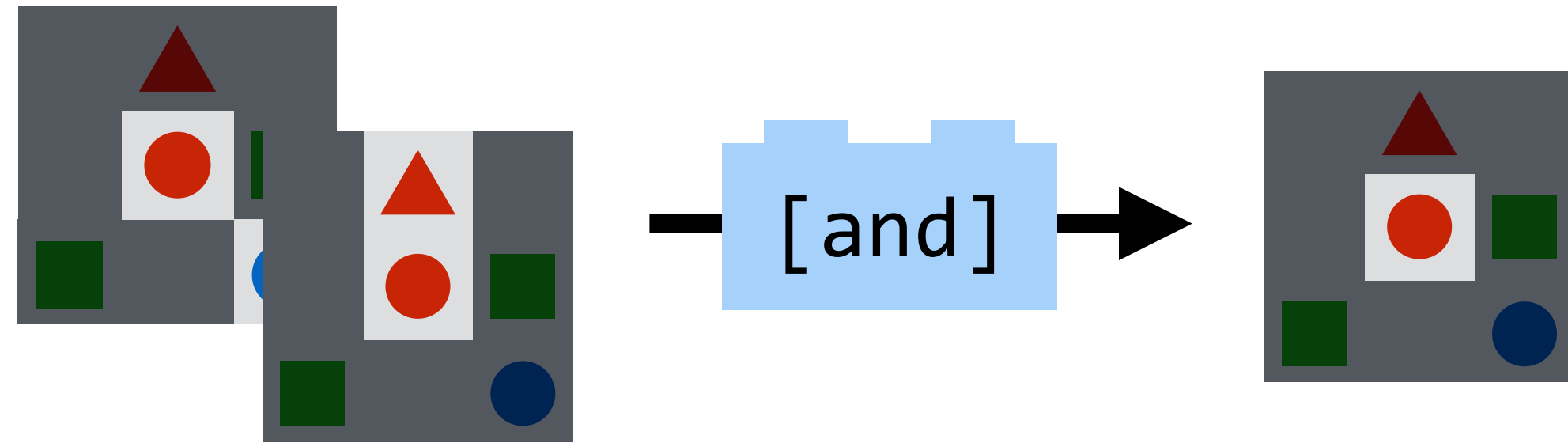
*Who is running in the grass?*



*What cities are south of San Diego?*



# A module for logic



*Is there a red shape above a circle?*

*What color is the triangle?*

*Who is running in the grass?*

*What cities are south of San Diego?*

[describe]

[find]

[relate]

[exists]



# Module inventory



*Is there a red shape above a circle?*

*What color is the triangle?*

*Who is running in the grass?*

*What cities are south of San Diego?*

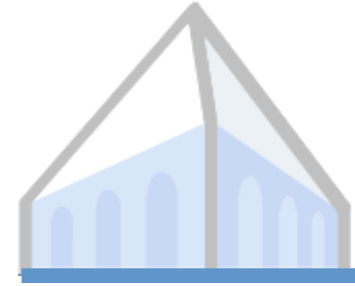
[describe]

[and]

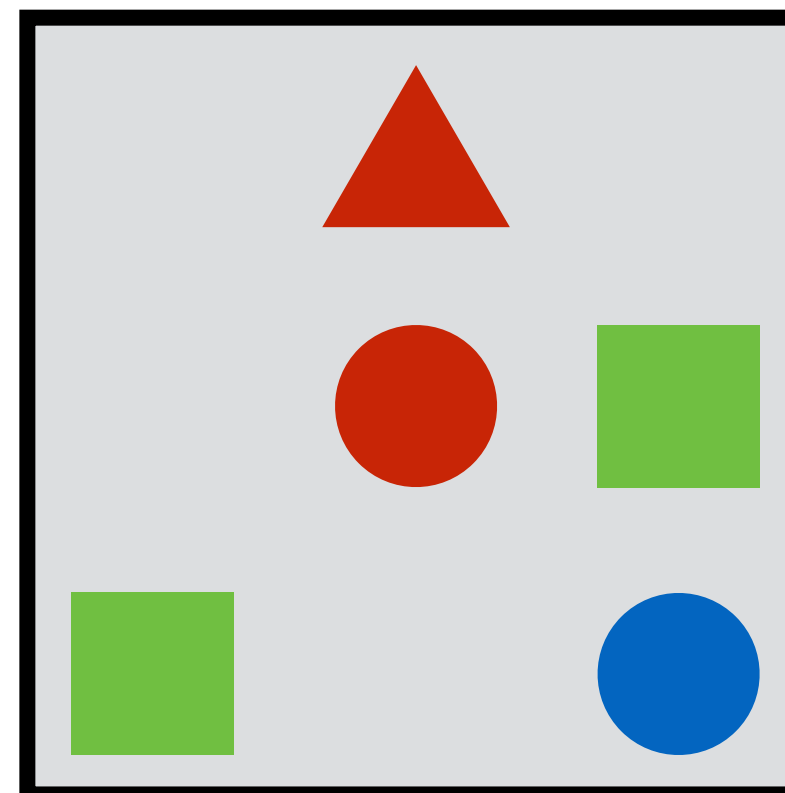
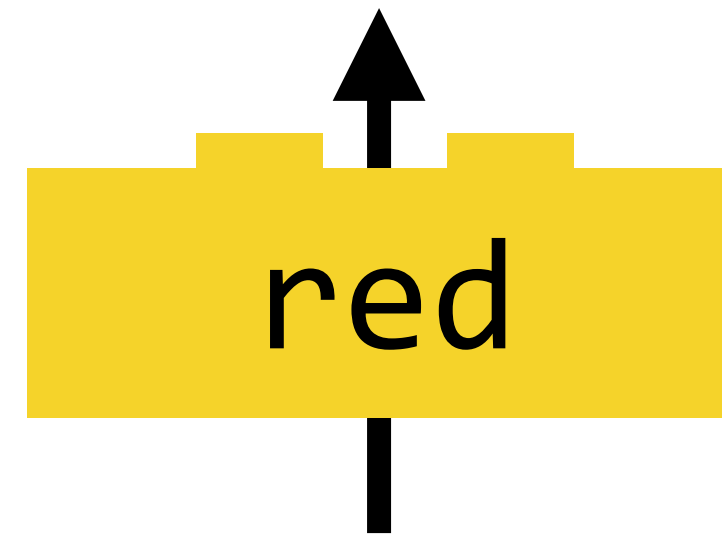
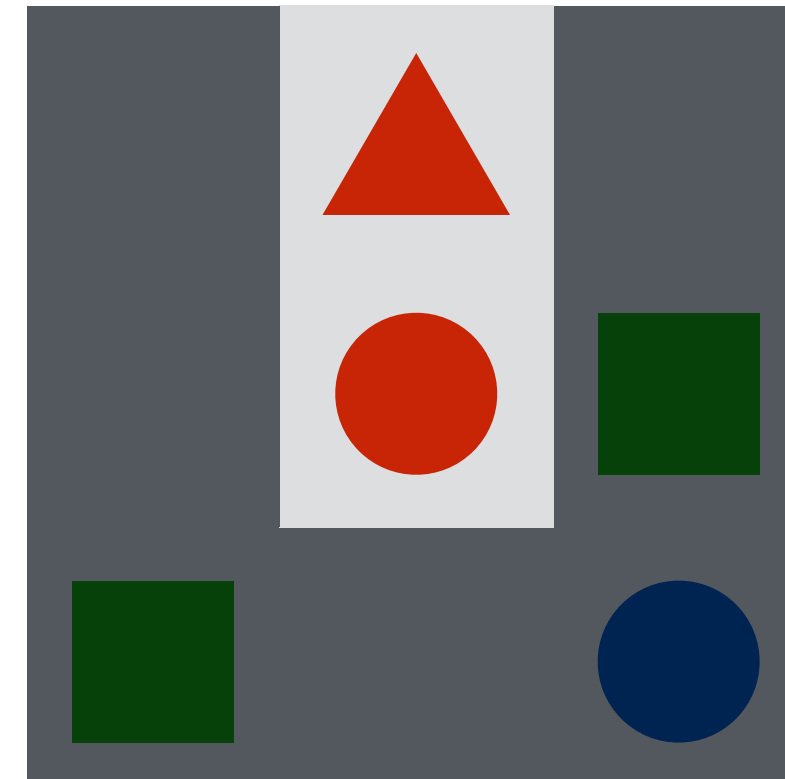
[find]

[relate]

[exists]



# The [find] module





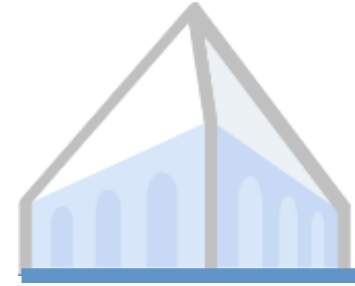
# The [find] module

---



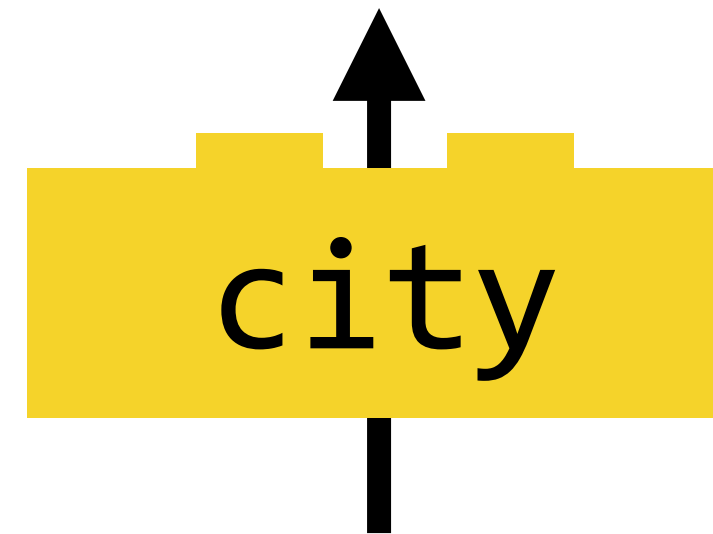
↑  
necktie  
↓



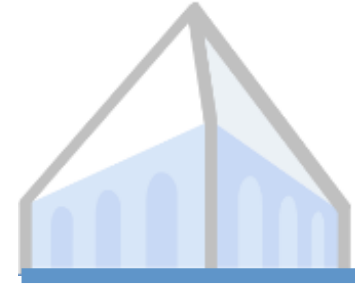


# The [find] module

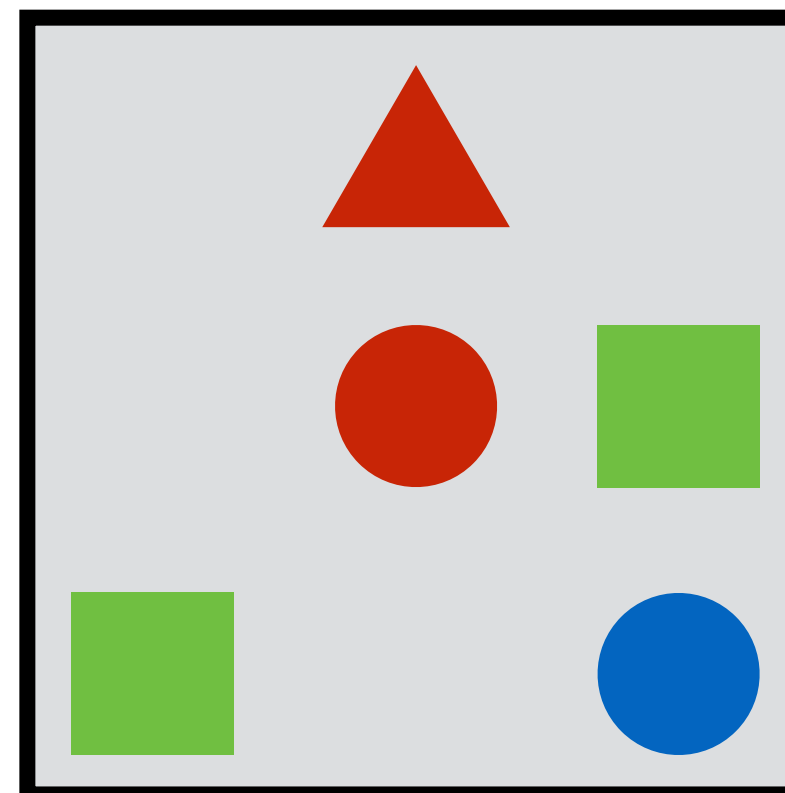
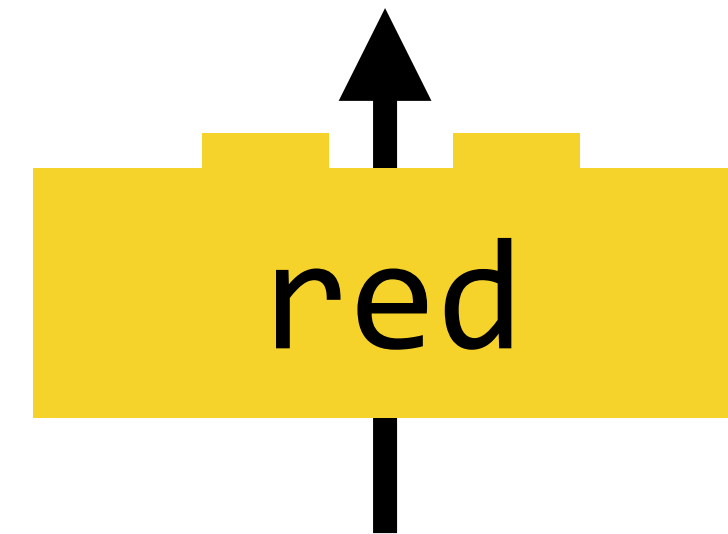
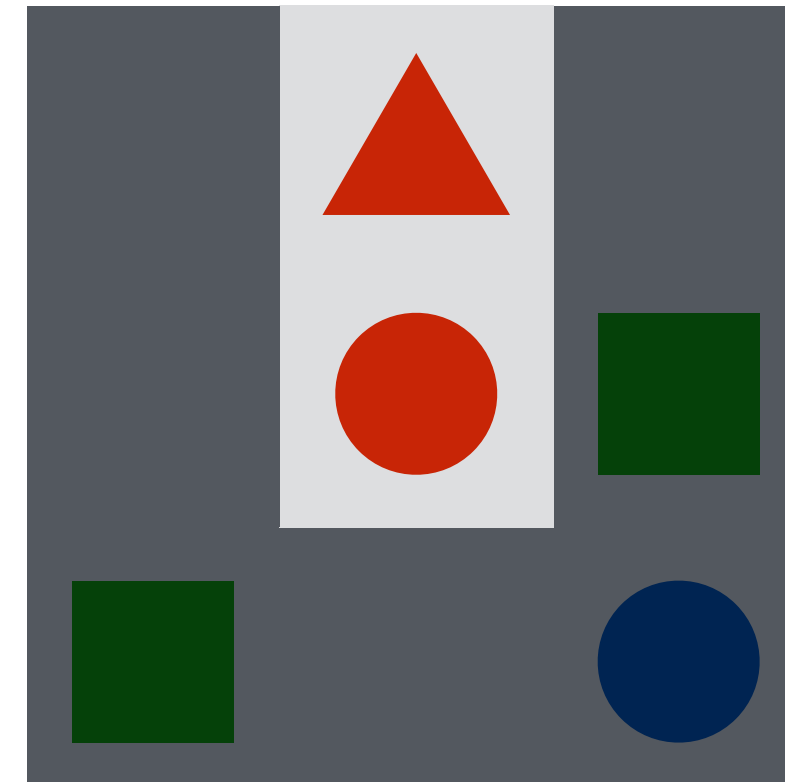
```
0.9 Columbia  
0.1 Cooper  
0.8 Myrtle Beach
```



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

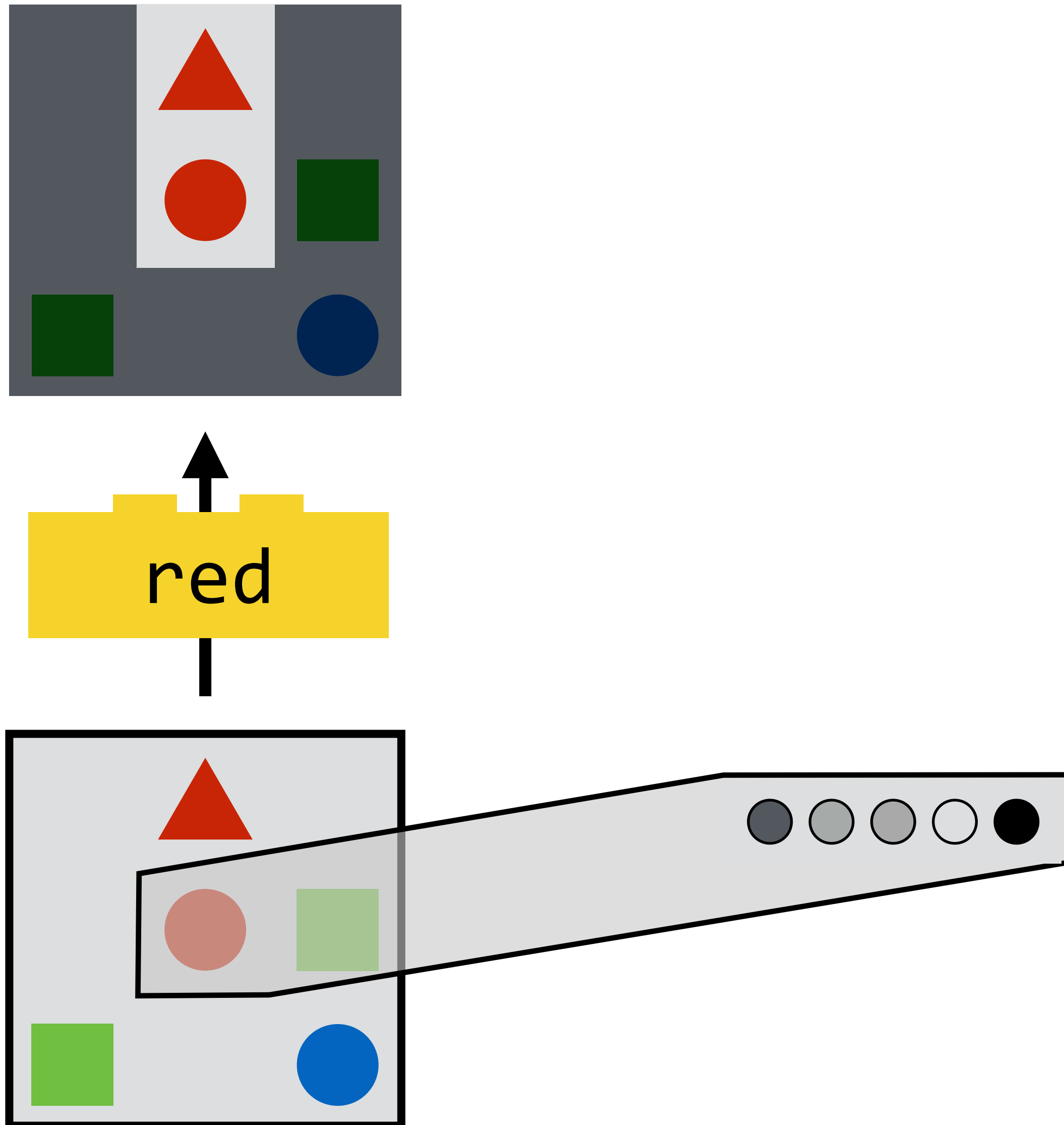


# The [find] module





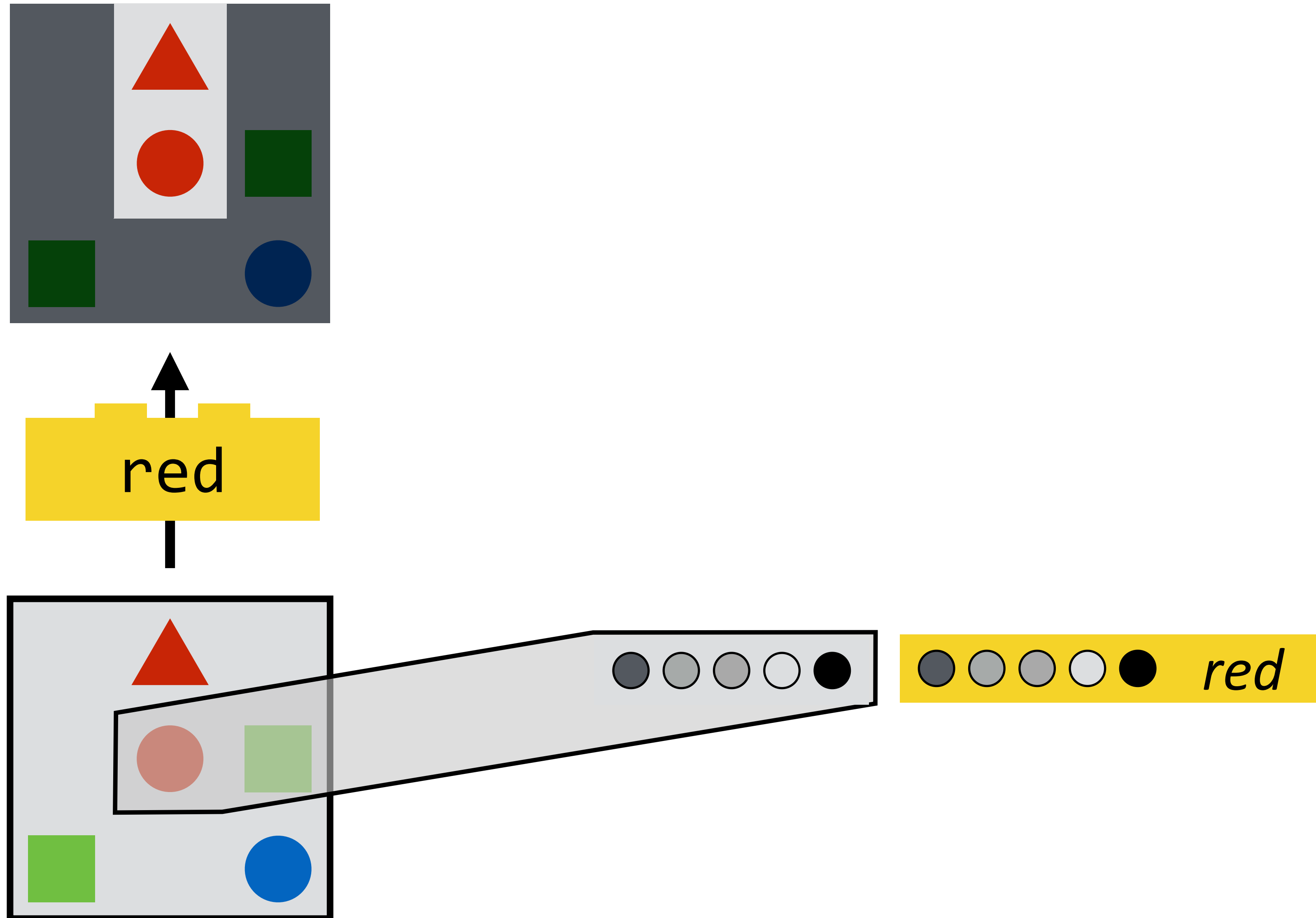
# The [find] module





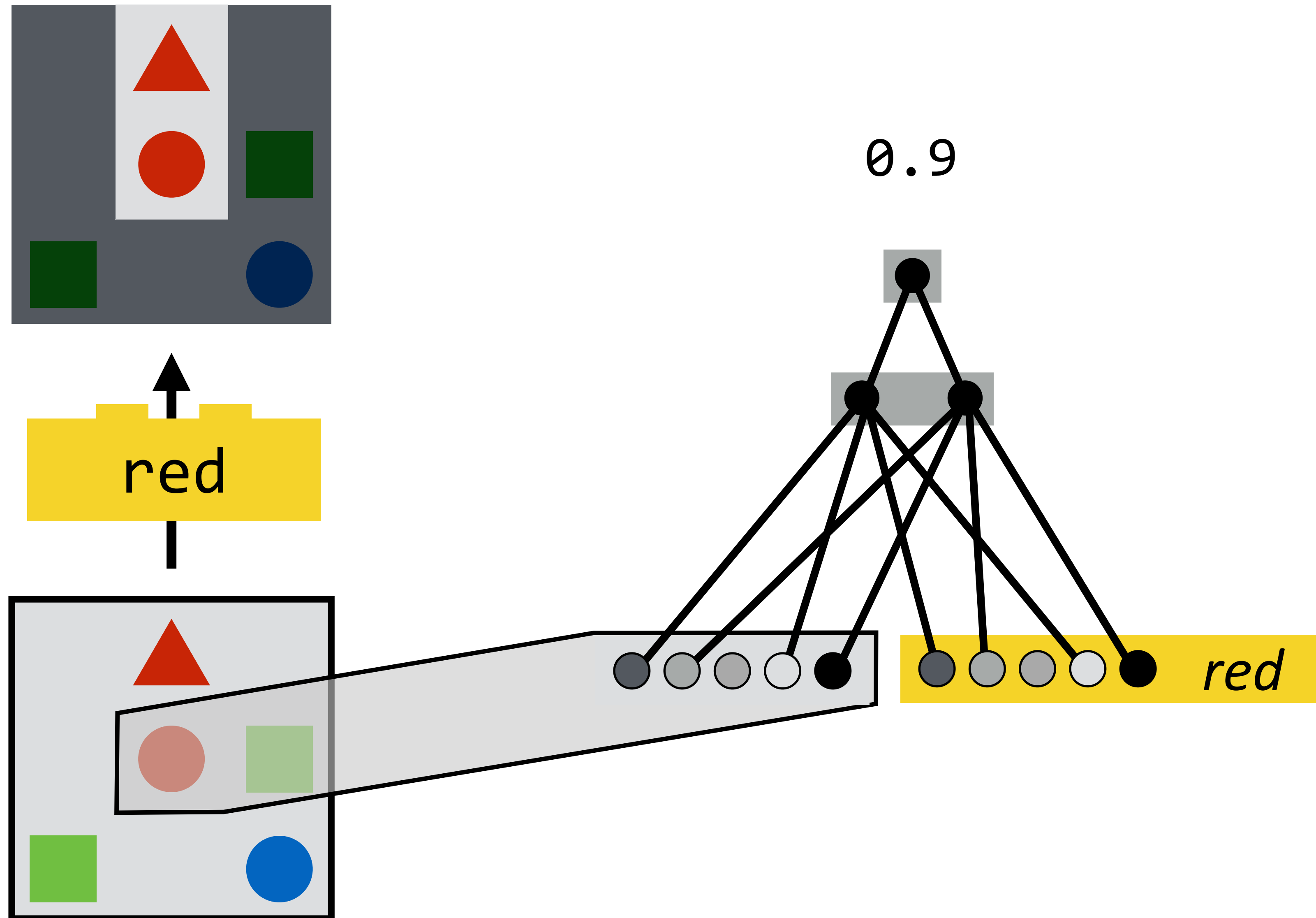


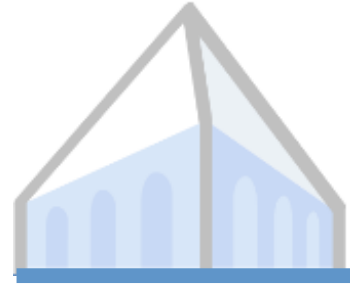
# The [find] module



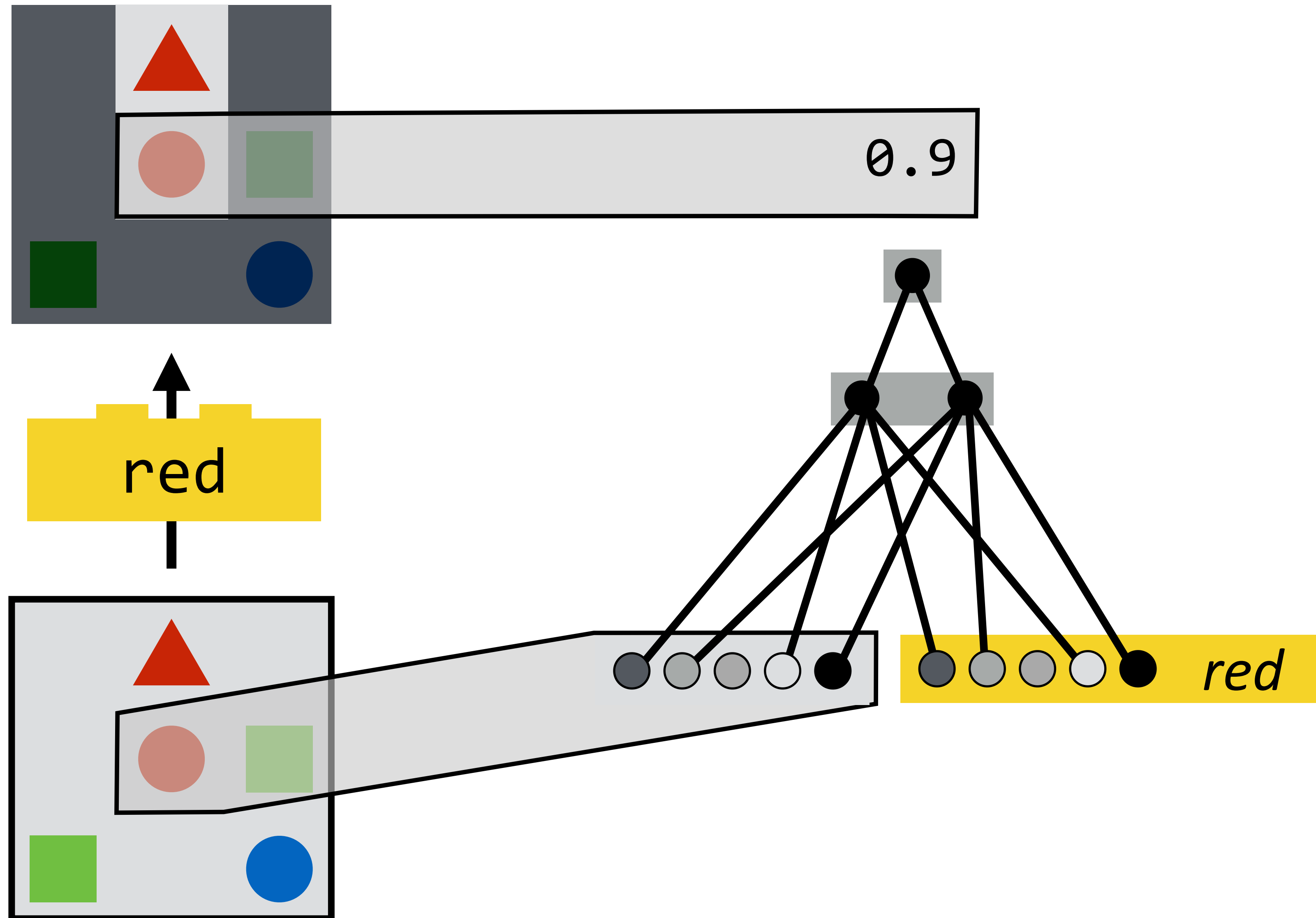


# The [find] module



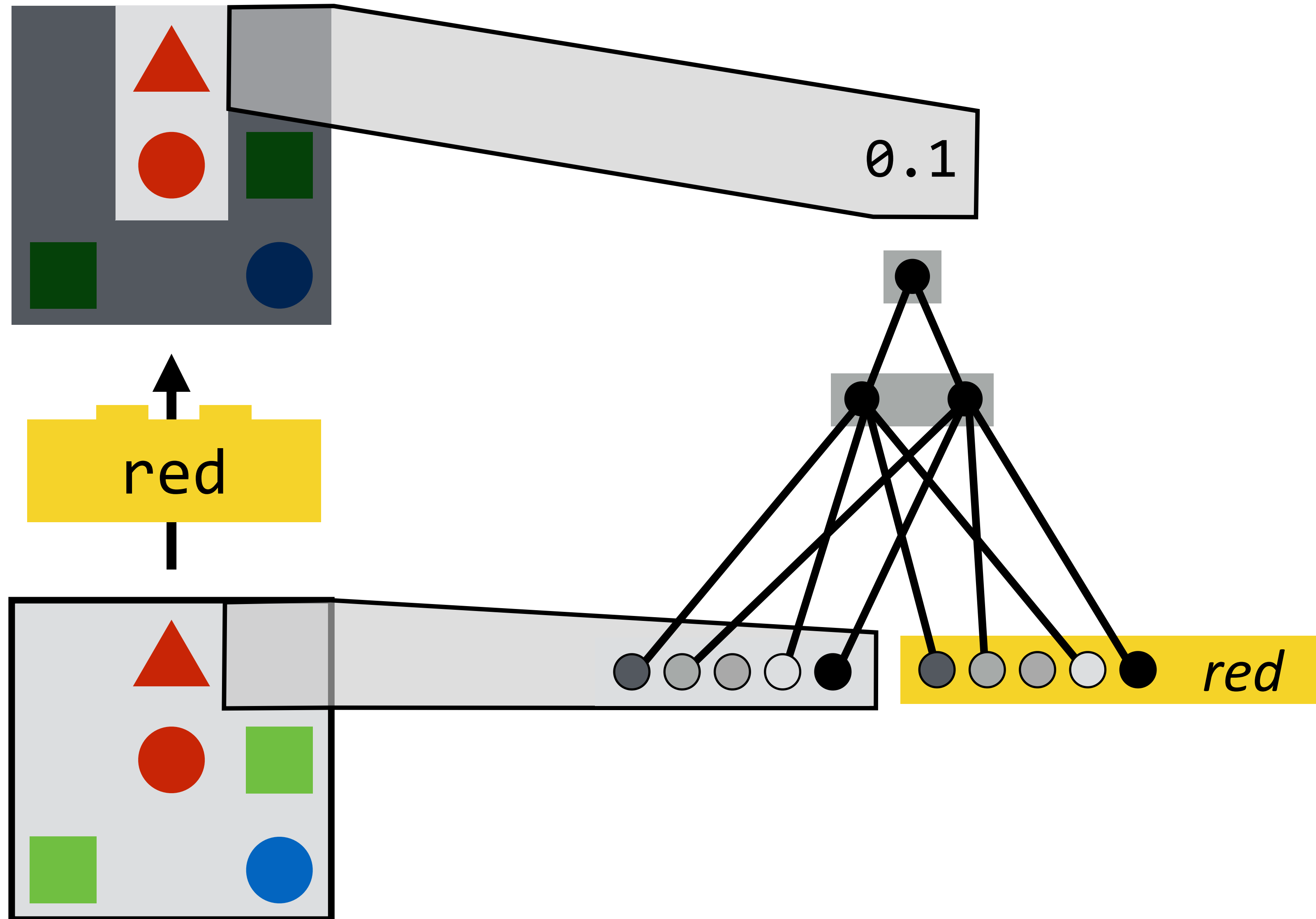


# The [find] module





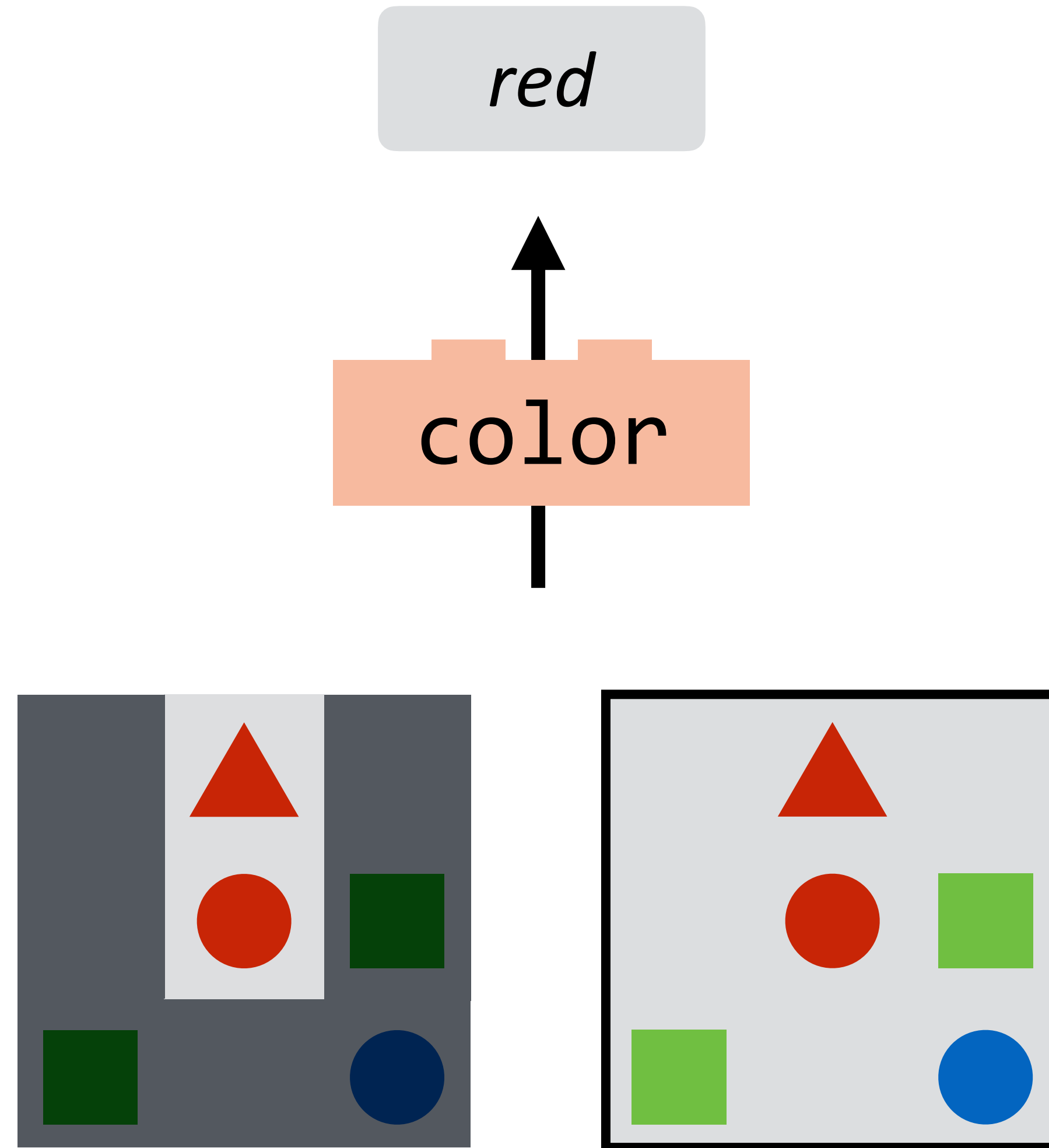
# The [find] module





# The [describe] module

---





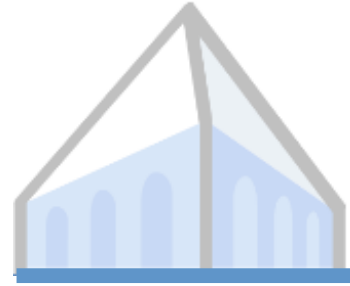
# The [describe] module

---

*necktie*

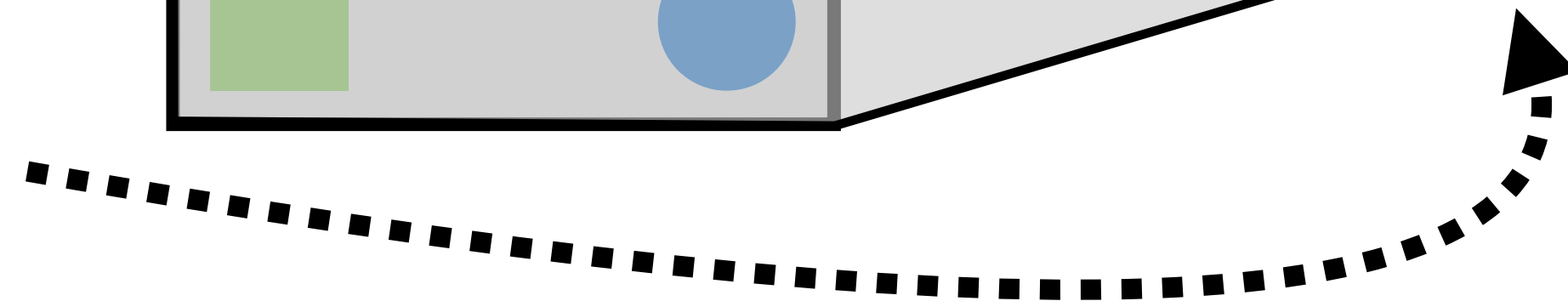
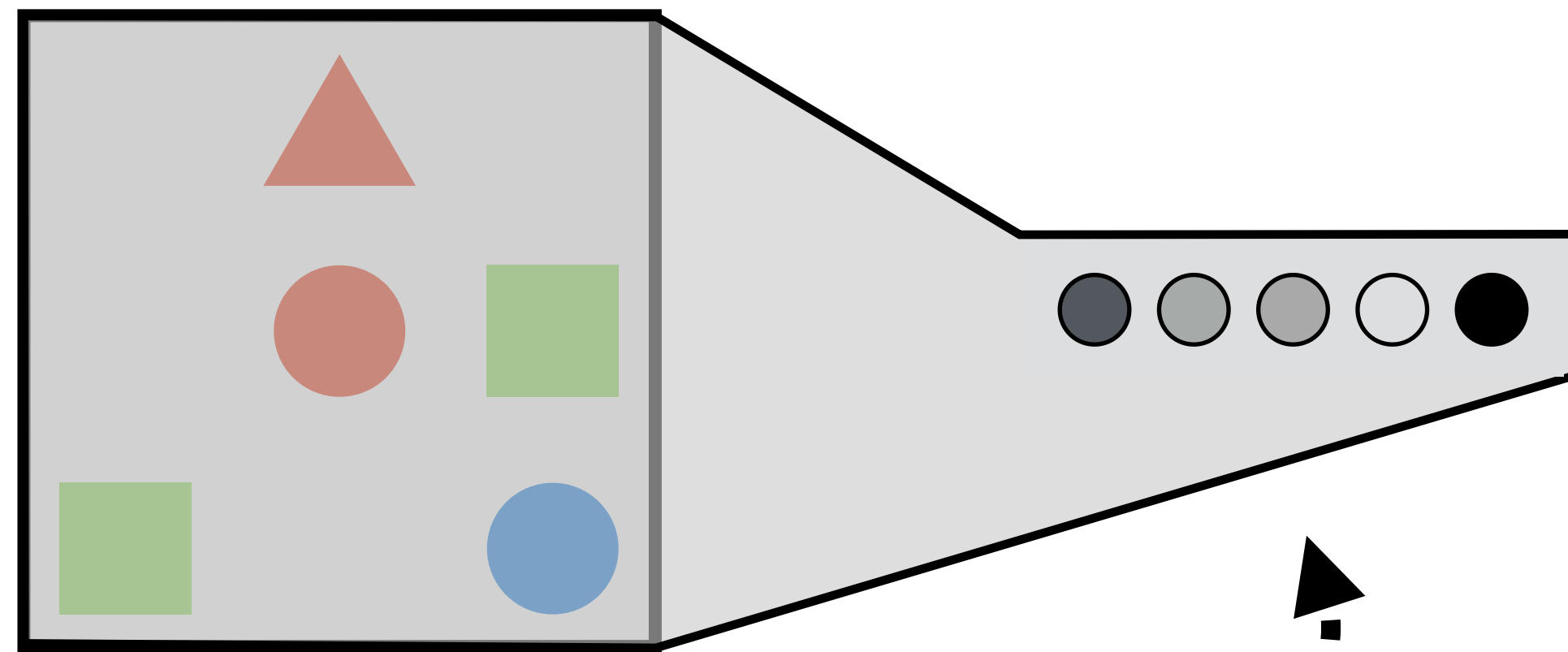
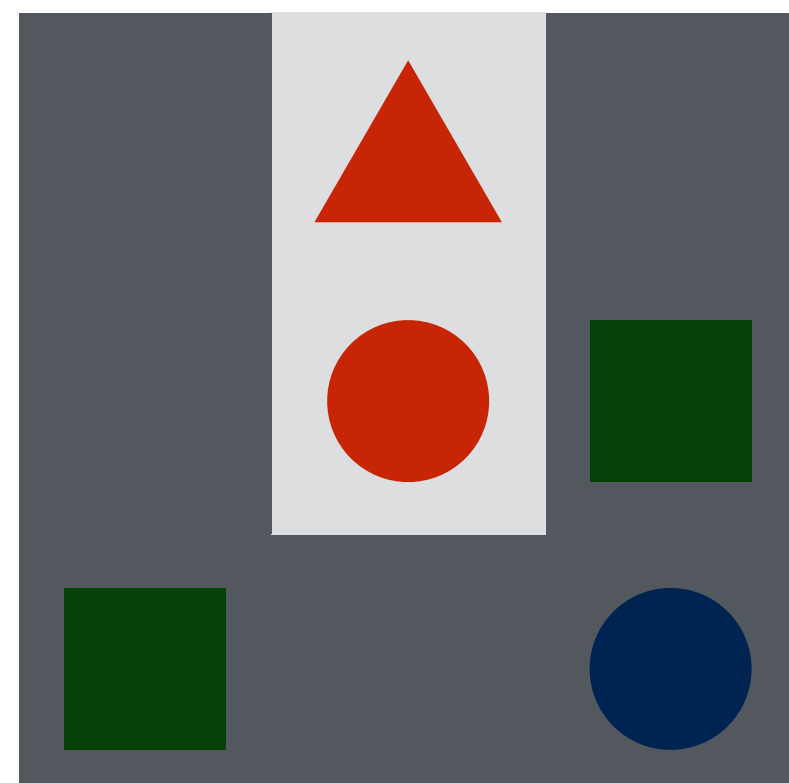
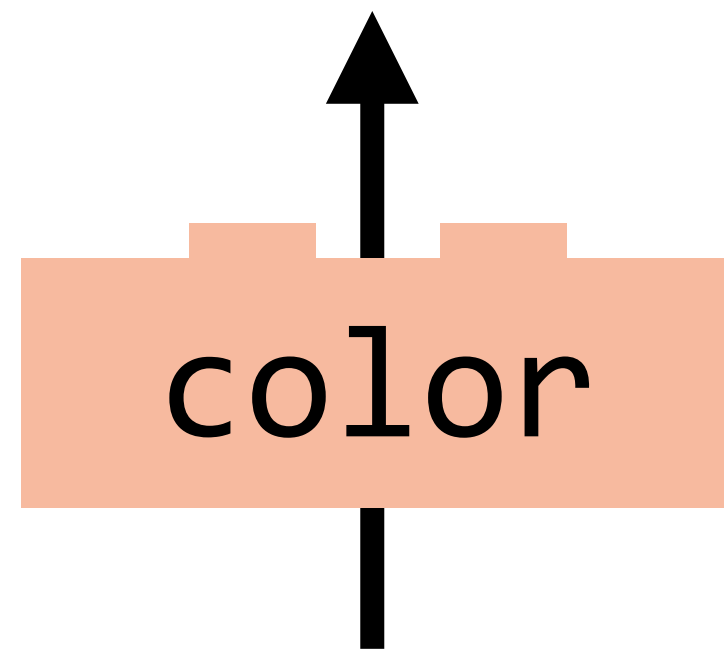
what





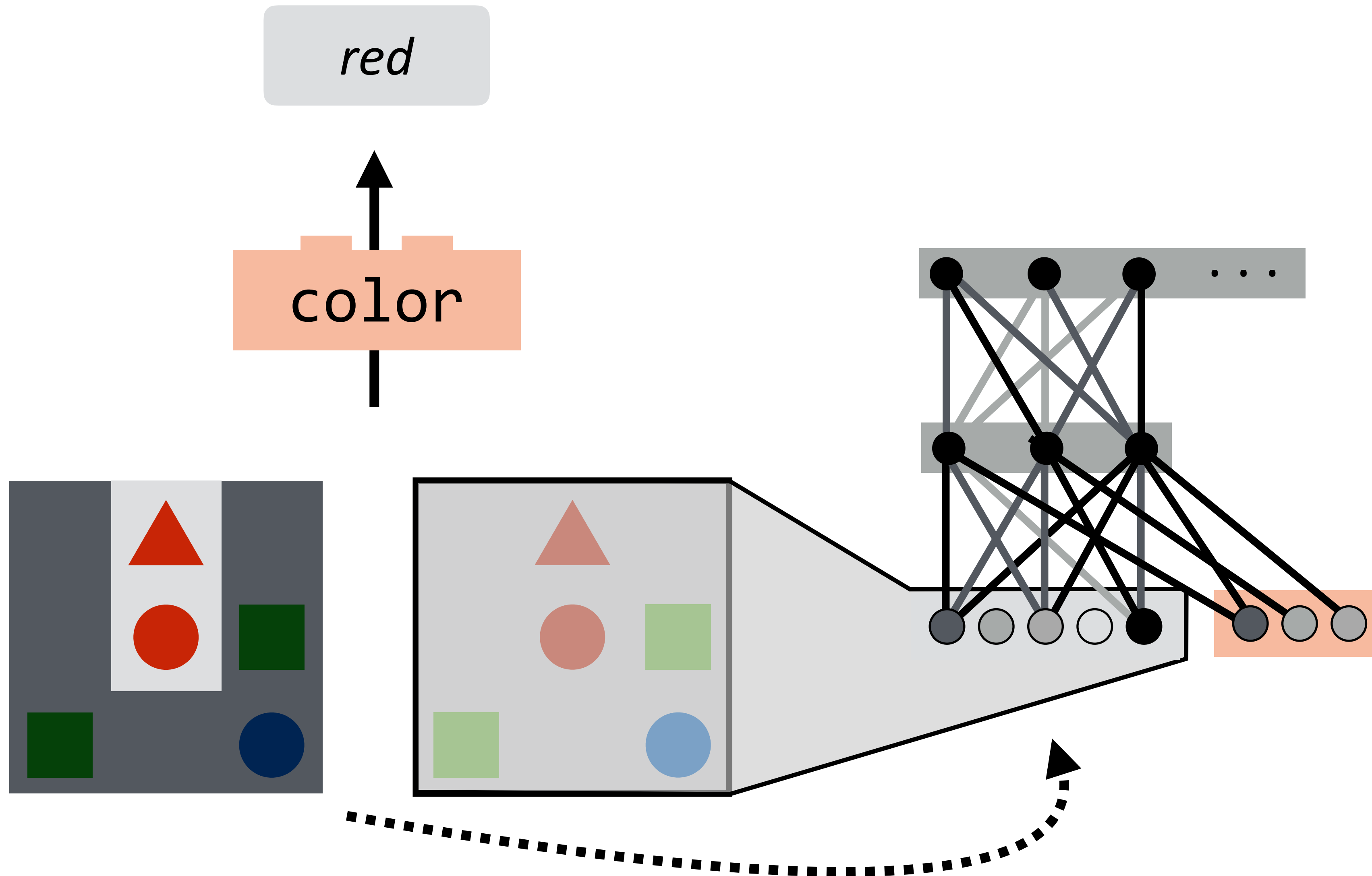
# The [describe] module

*red*





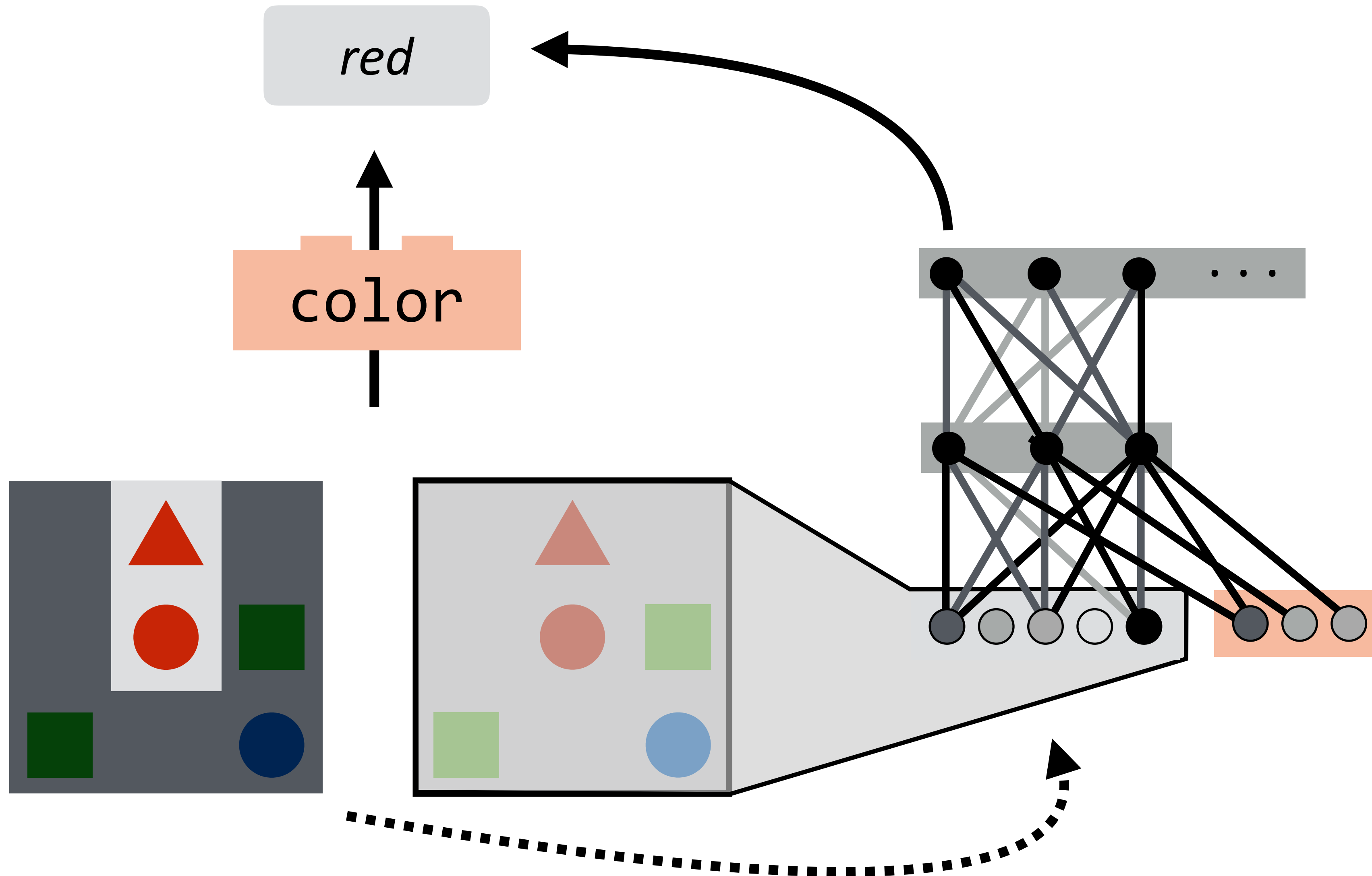
# The [describe] module







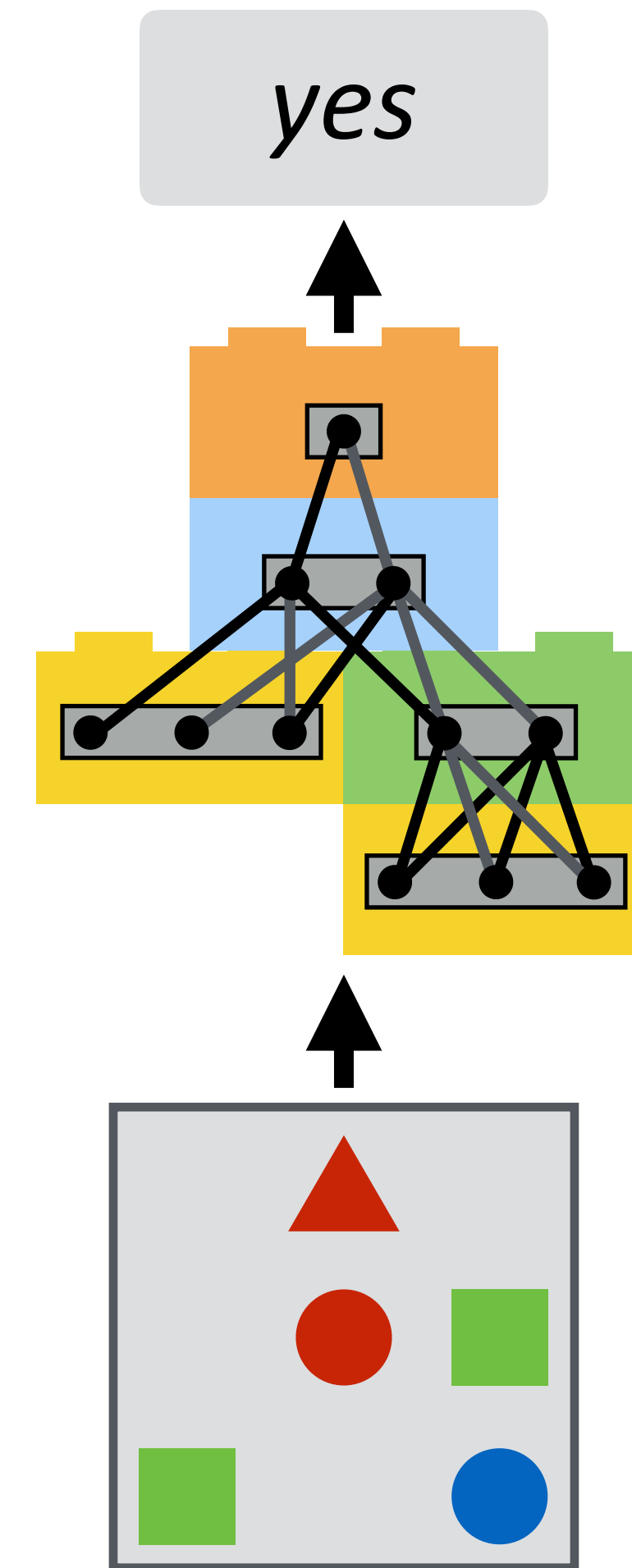
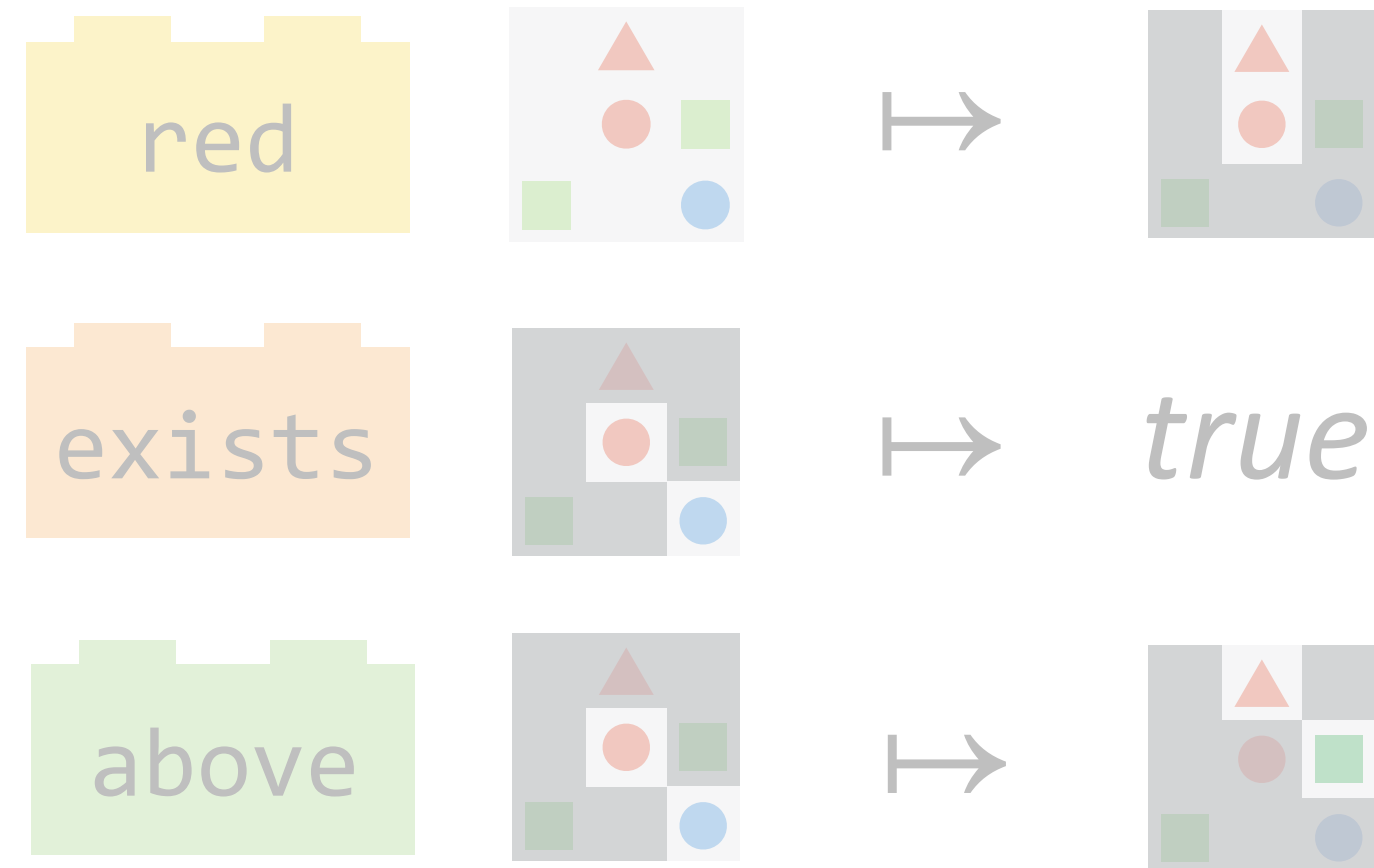
# The [describe] module

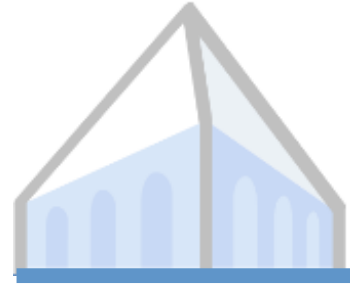




# Outline

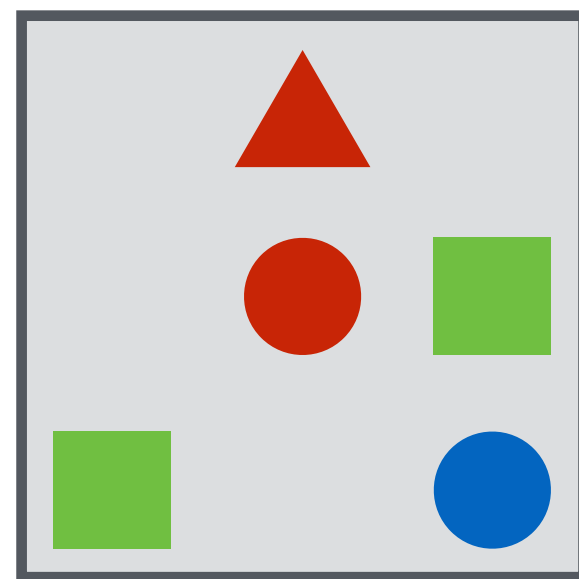
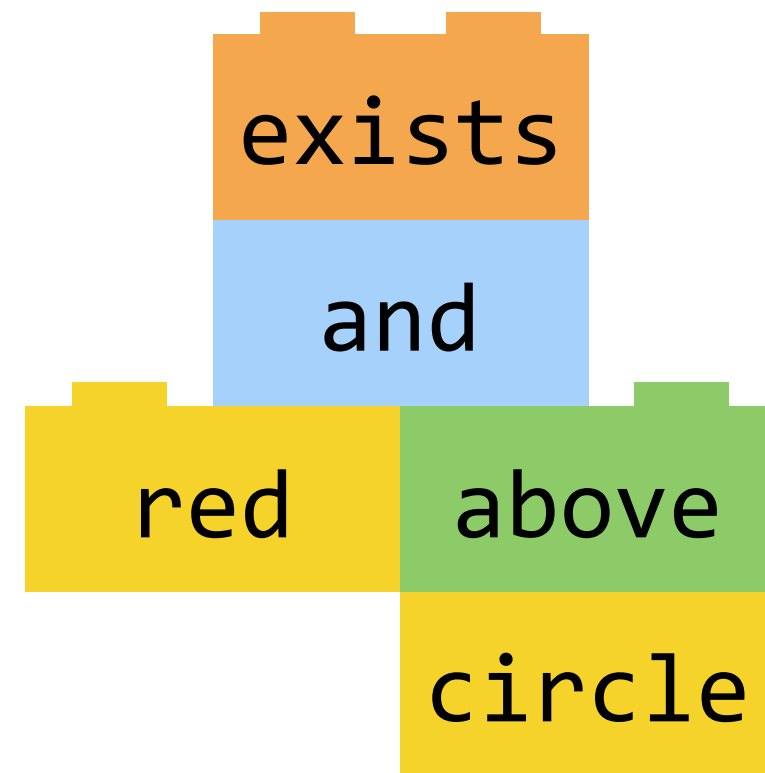
*Is there a red shape  
above a circle?*





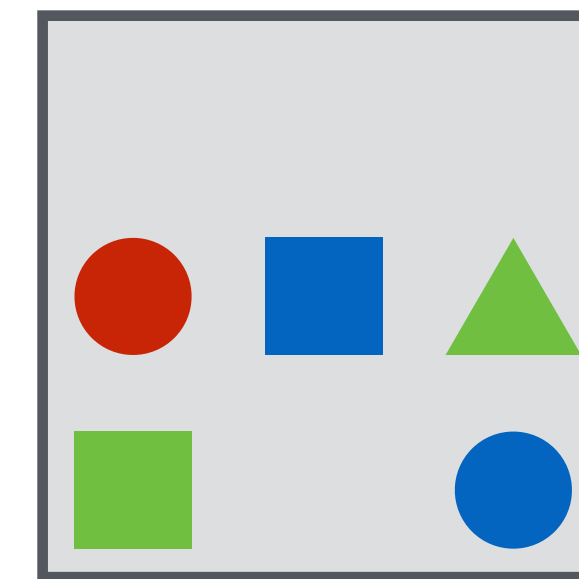
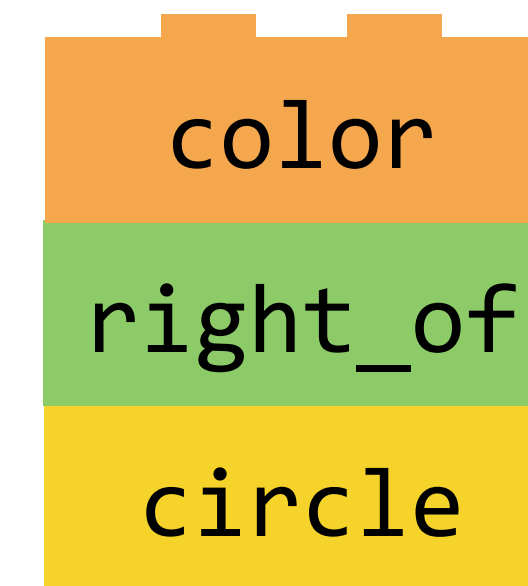
# Learning

*yes*



*Is there a red shape above a circle?*

*blue*

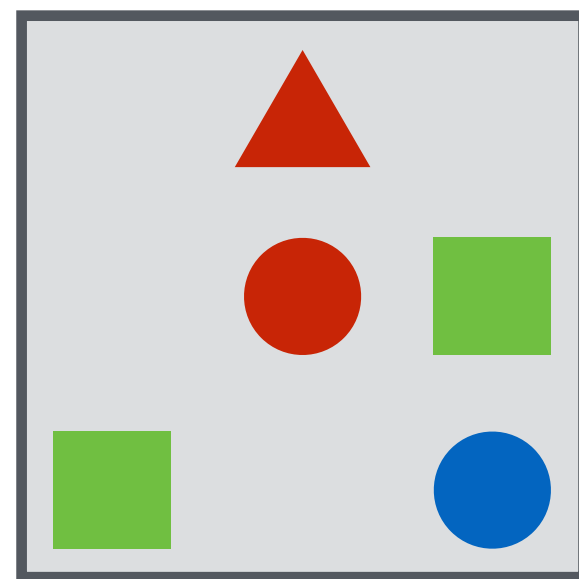
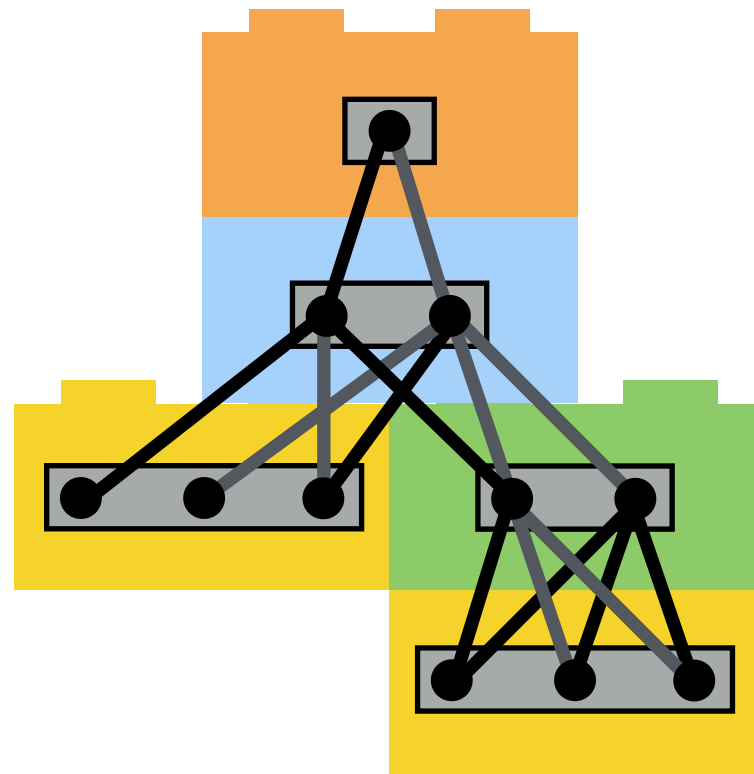


*What color is the shape right of a circle?*



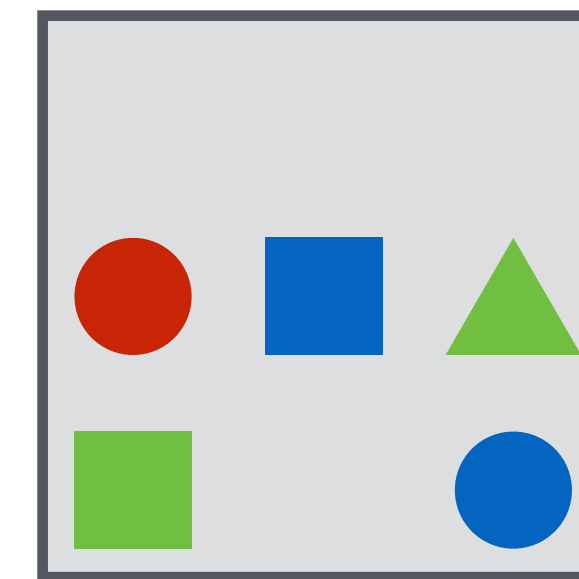
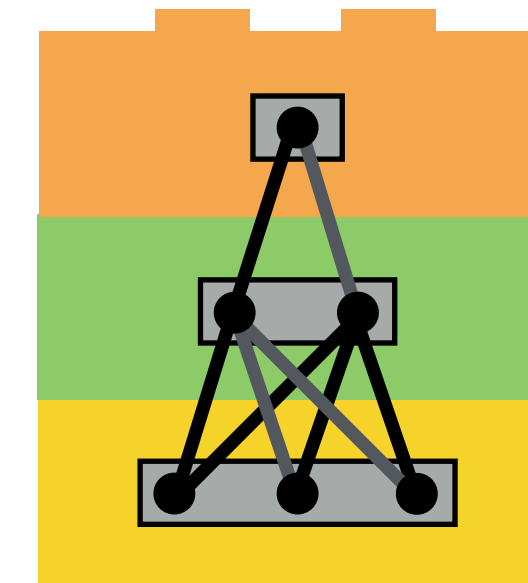
# Learning

*yes*



*Is there a red shape above a circle?*

*blue*

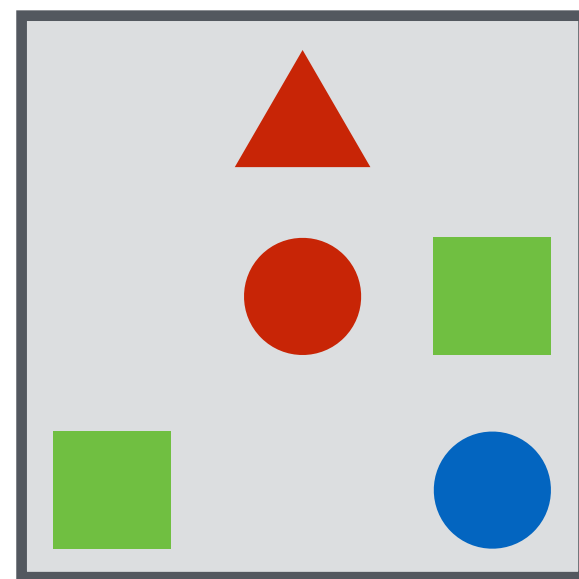
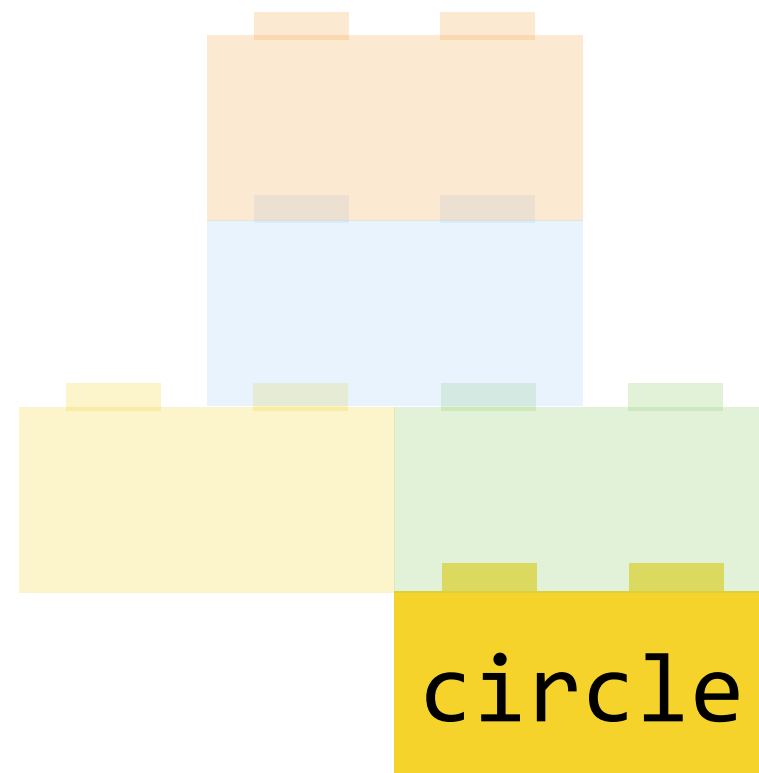


*What color is the shape right of a circle?*



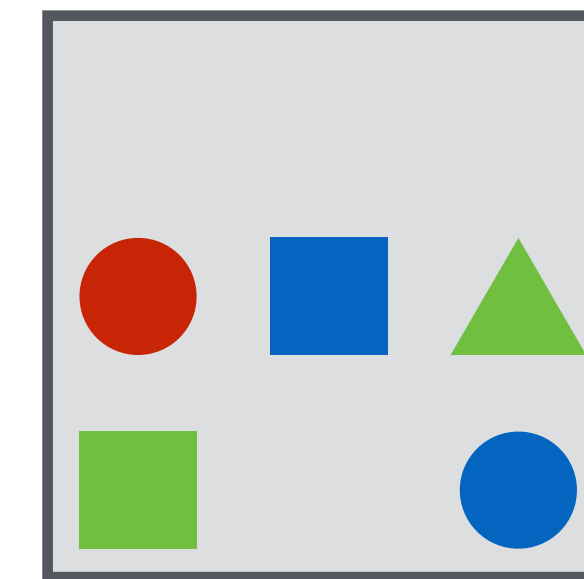
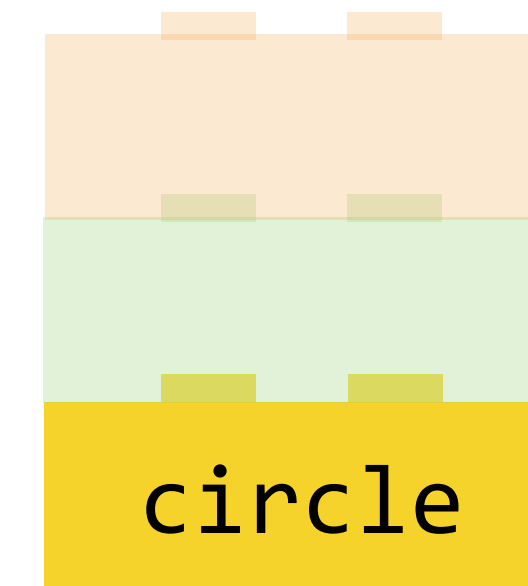
# Parameter tying

*yes*



*Is there a red shape above a circle?*

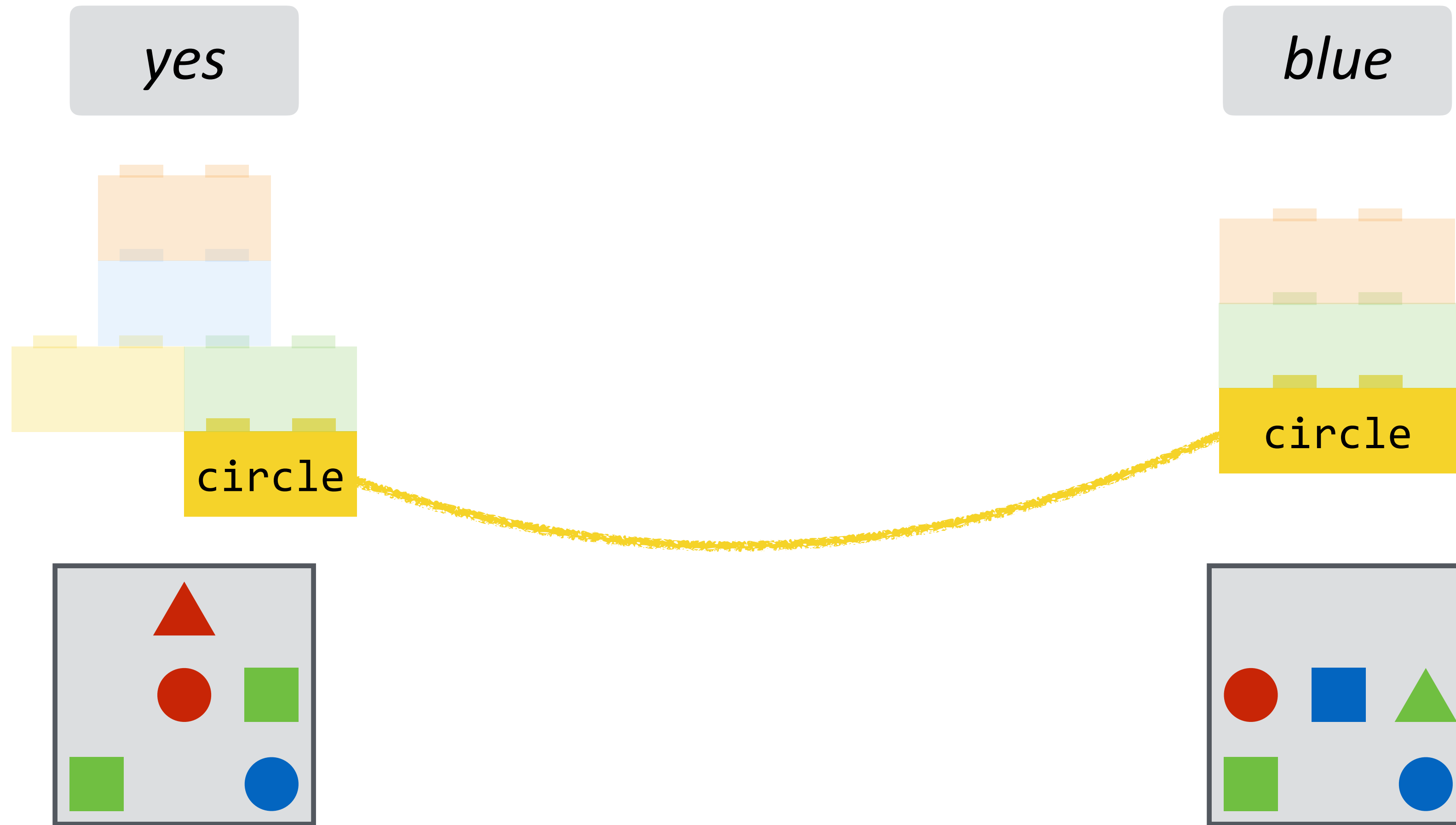
*blue*



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

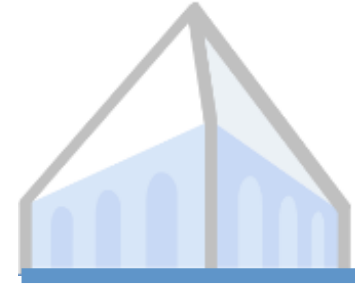


# Parameter tying

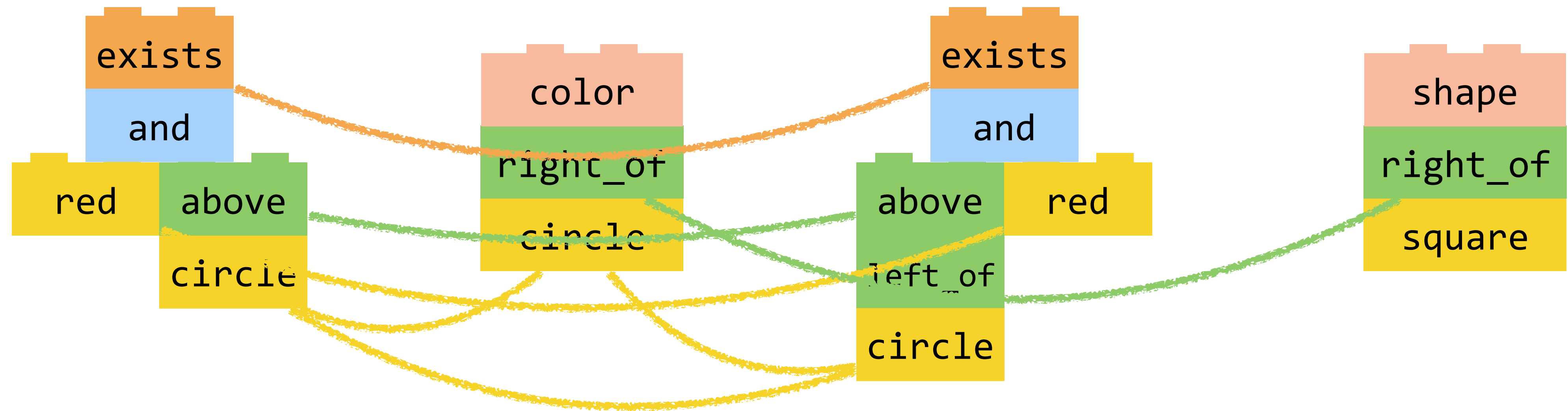


*Is there a red shape above a circle?*

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



# Extreme parameter tying





# Learning with fixed layouts is easy!

---

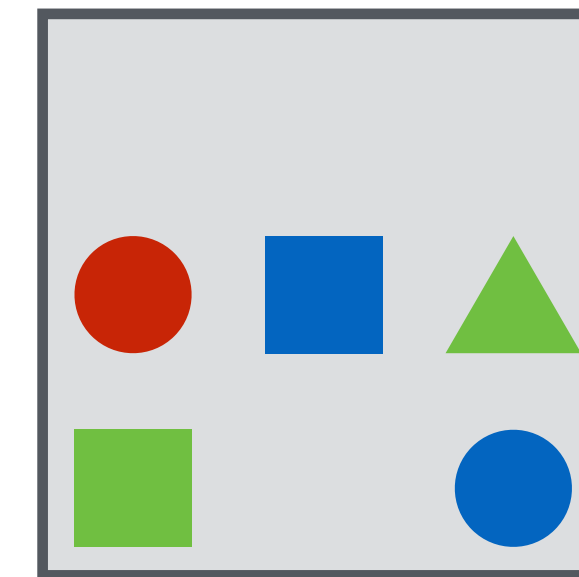
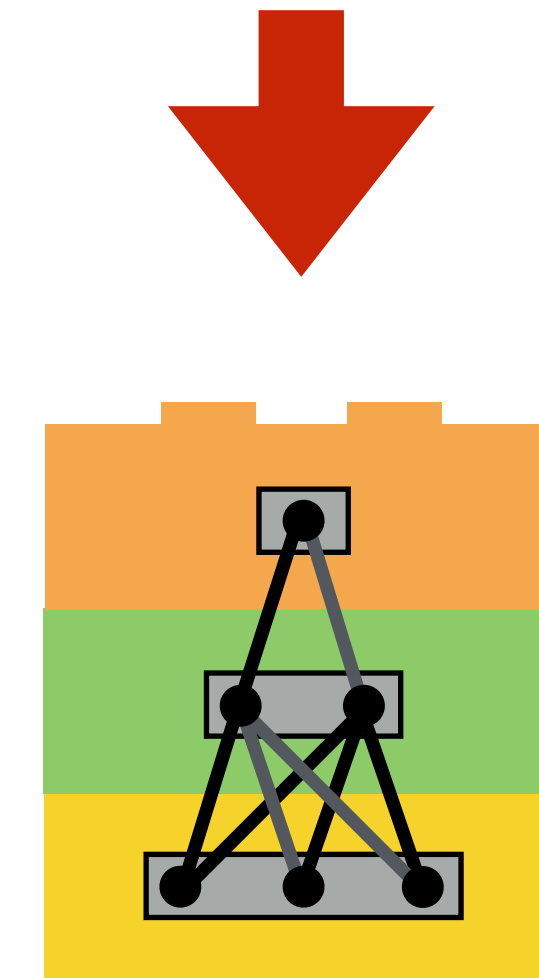
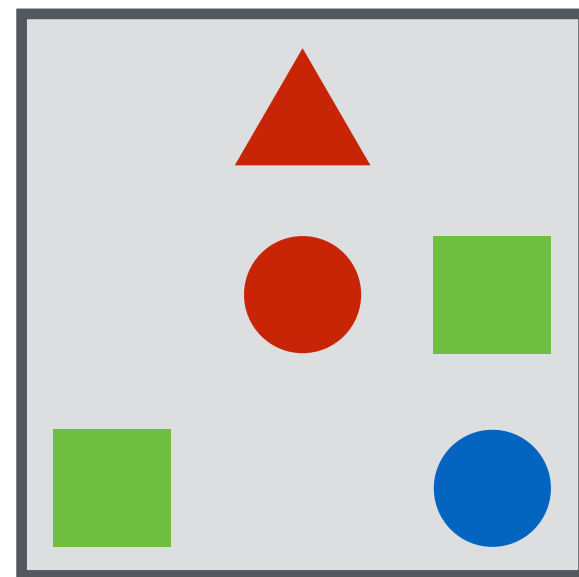
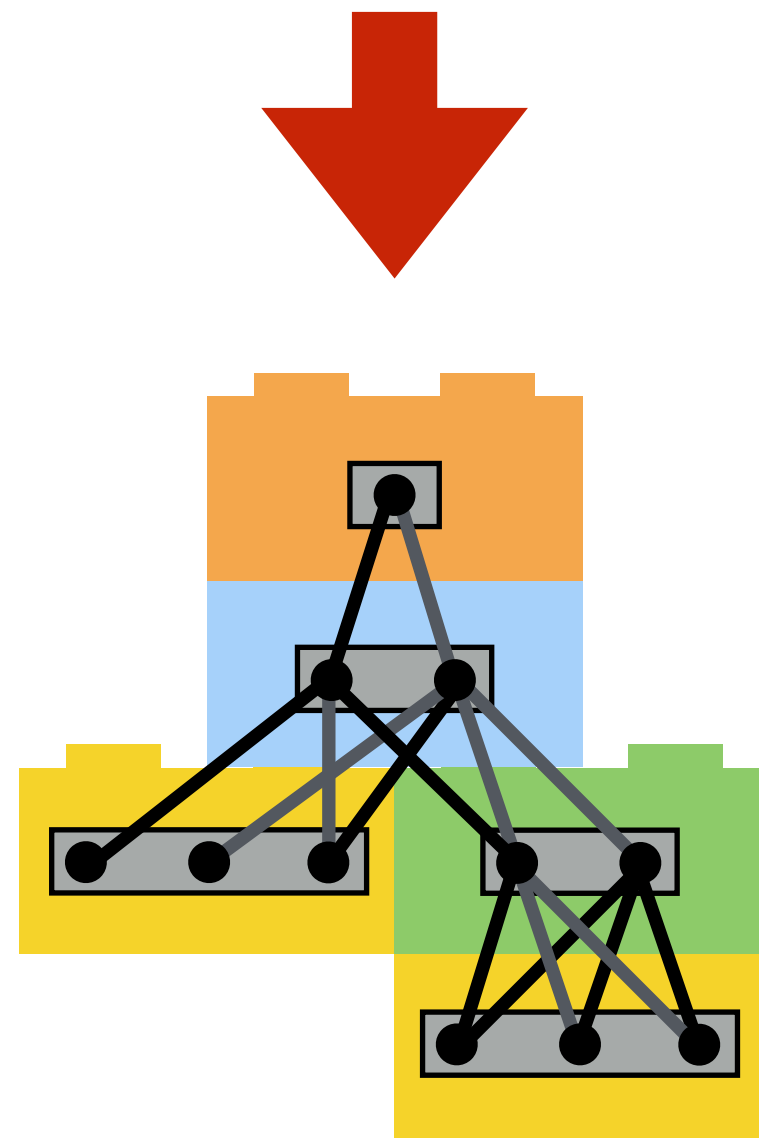
$$\arg \max_W \sum p(\text{yes} \mid \begin{array}{|c|} \hline \triangle \\ \hline \bullet \\ \hline \square \\ \hline \square \\ \hline \bullet \\ \hline \end{array}, \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \square \\ \hline \square \\ \hline \square \\ \hline \square \\ \hline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

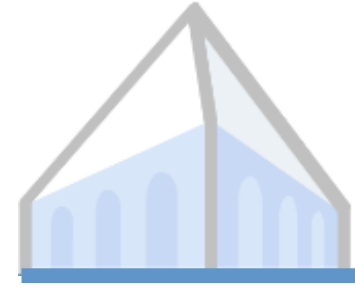
(where every root module outputs a distribution over answers  
and  $W$  is the set of all module parameters)





# Maximum likelihood estimation





# Properties of learning

---

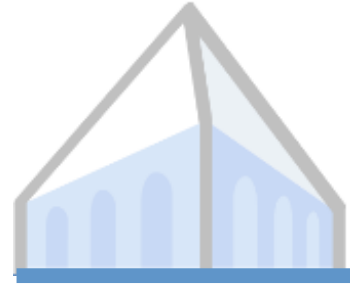
right\_of  
square

Module specialization is driven entirely by context

green

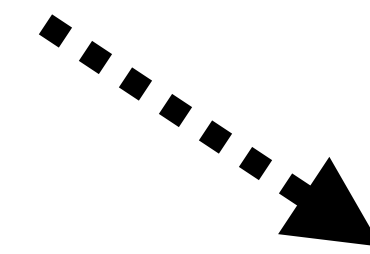
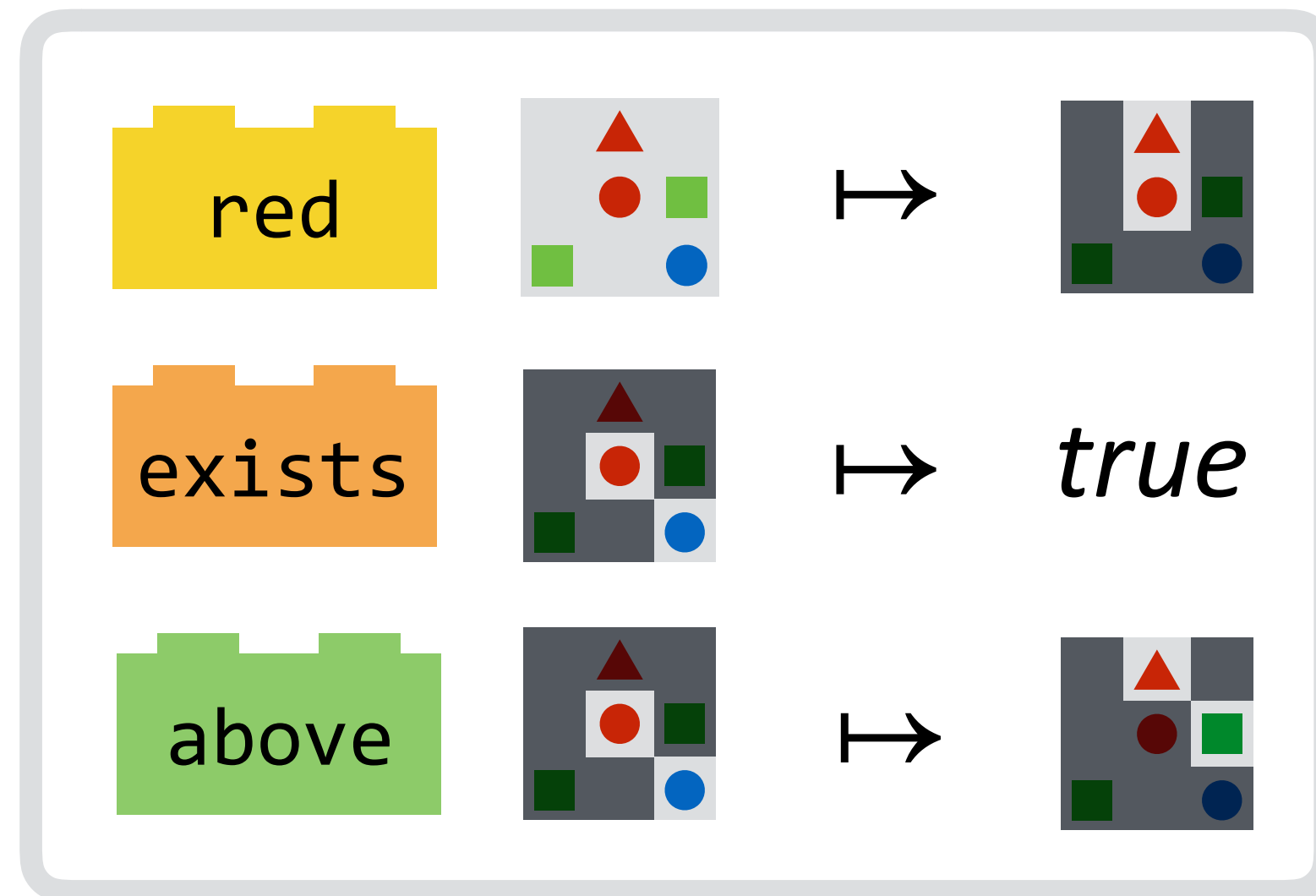
chartreuse

“Lexicon learning” is a continuous optimization problem

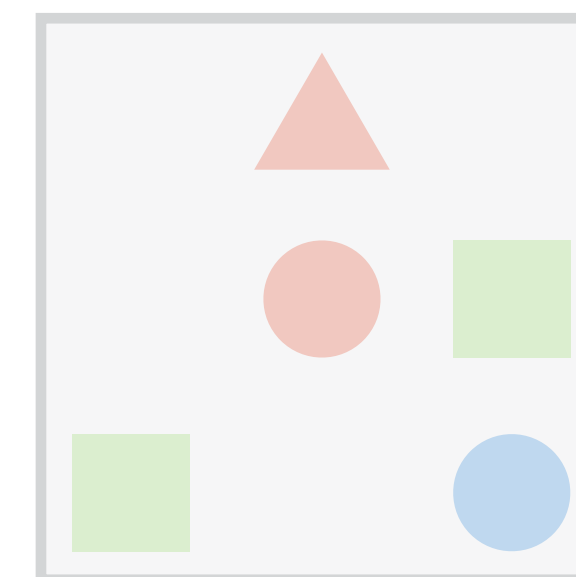
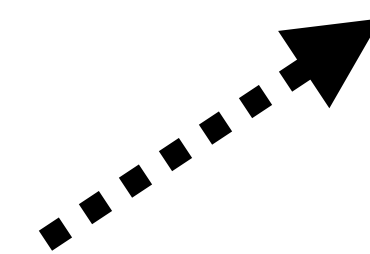
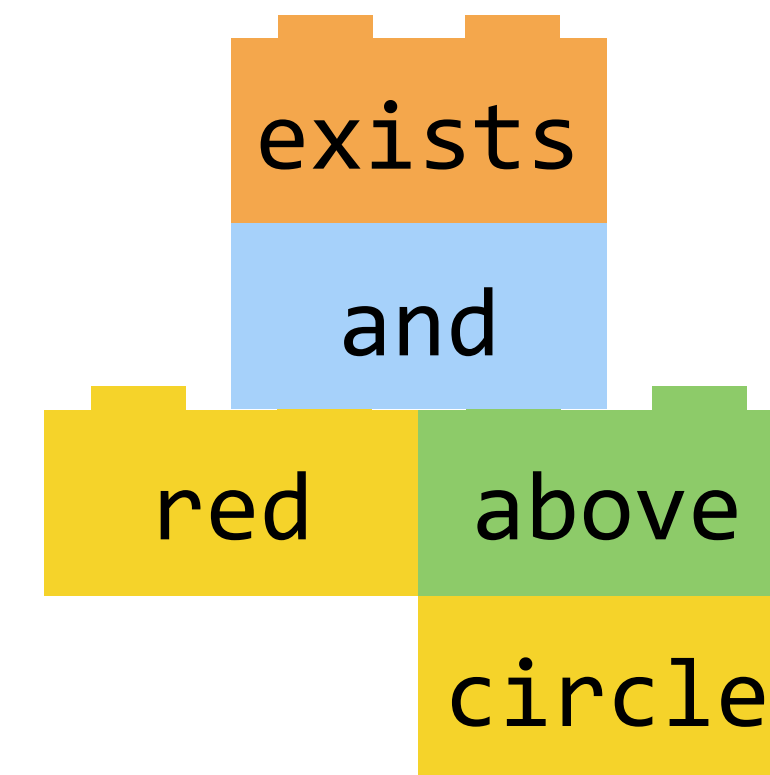


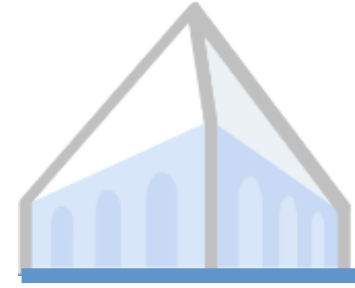
# Outline

*Is there a red shape  
above a circle?*



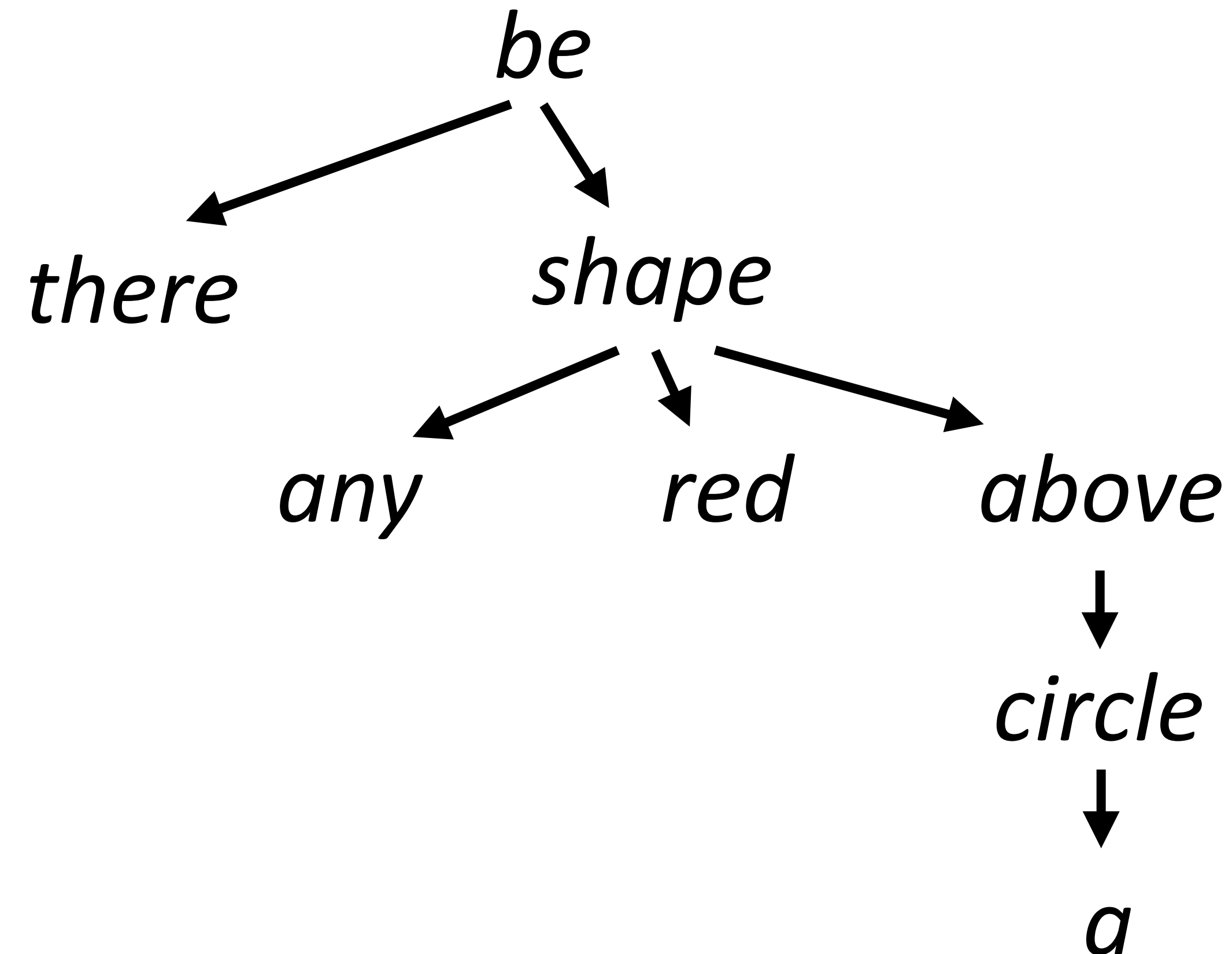
*yes*





# Where do layouts come from?

*Is there a red shape above a circle?*

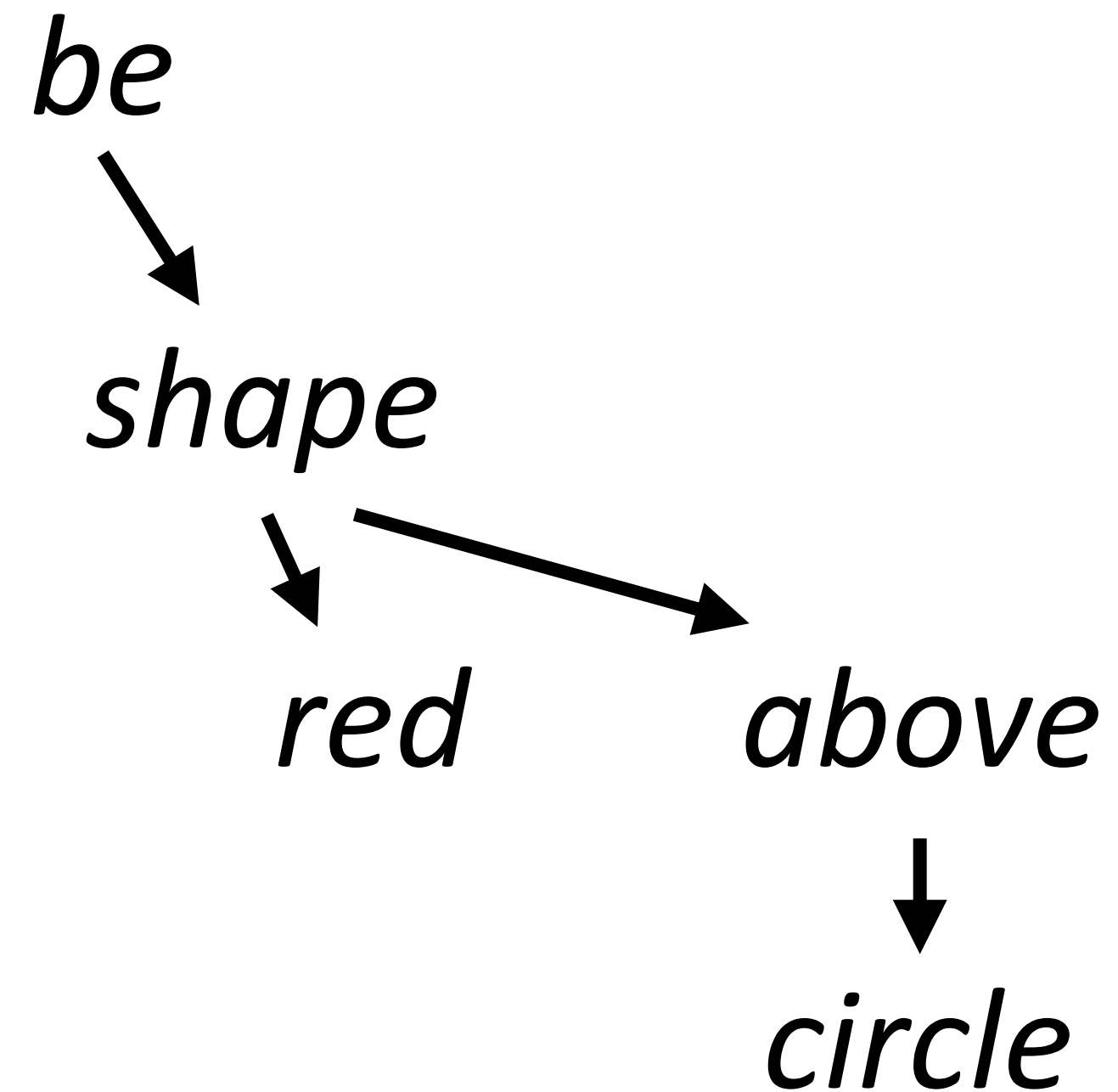


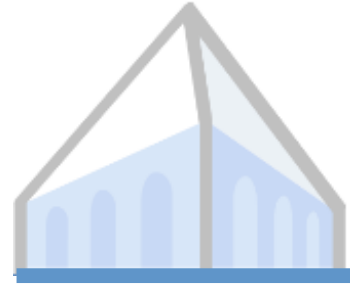


# Where do layouts come from?

---

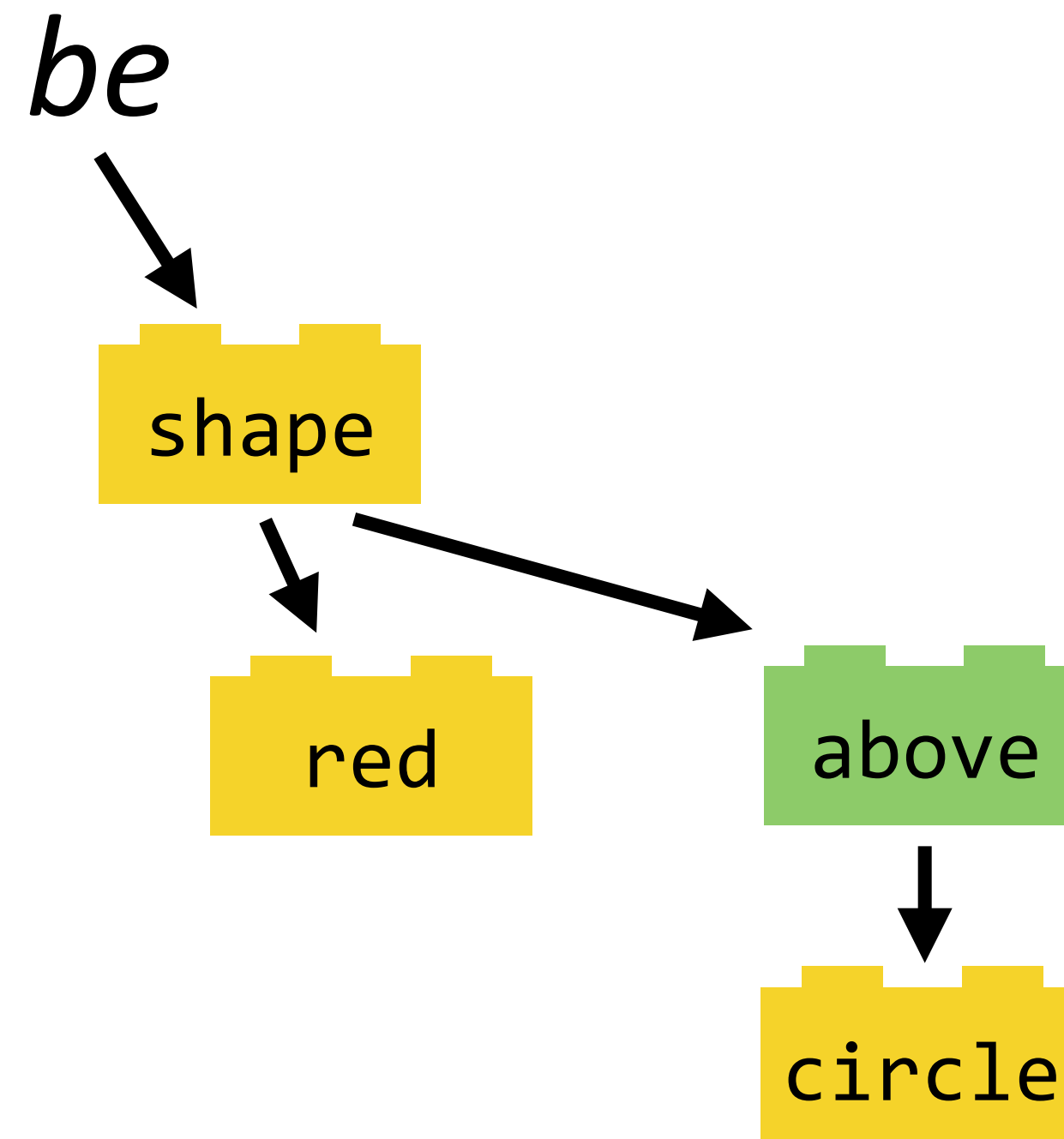
*Is there a red shape above a circle?*





# Where do layouts come from?

*Is there a red shape above a circle?*





# Where do layouts come from?

---

*Is there a red shape above a circle?*

shape

red

above

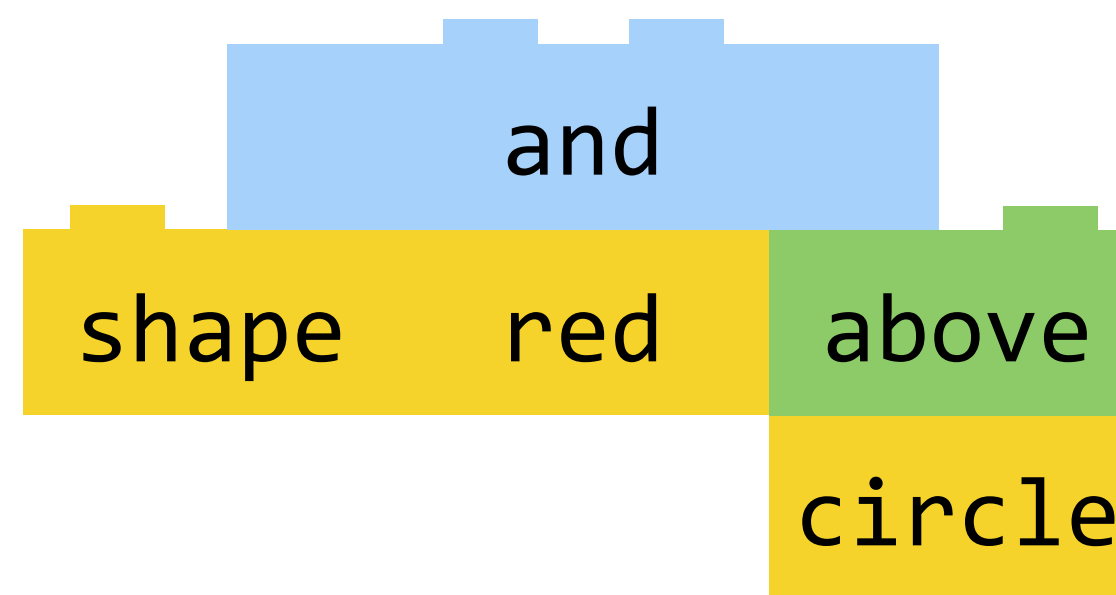
circle



# Where do layouts come from?

---

*Is there a red shape above a circle?*

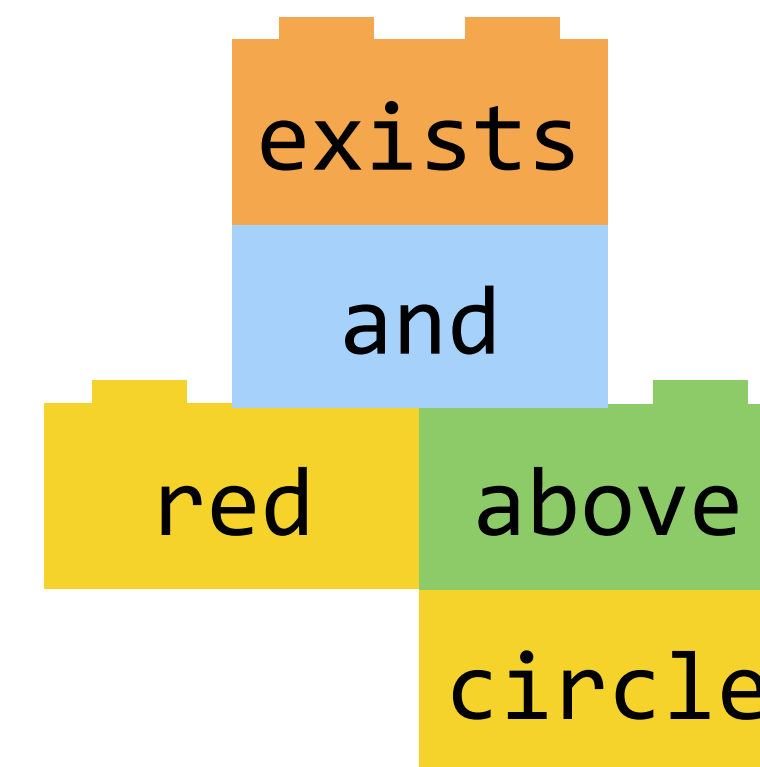
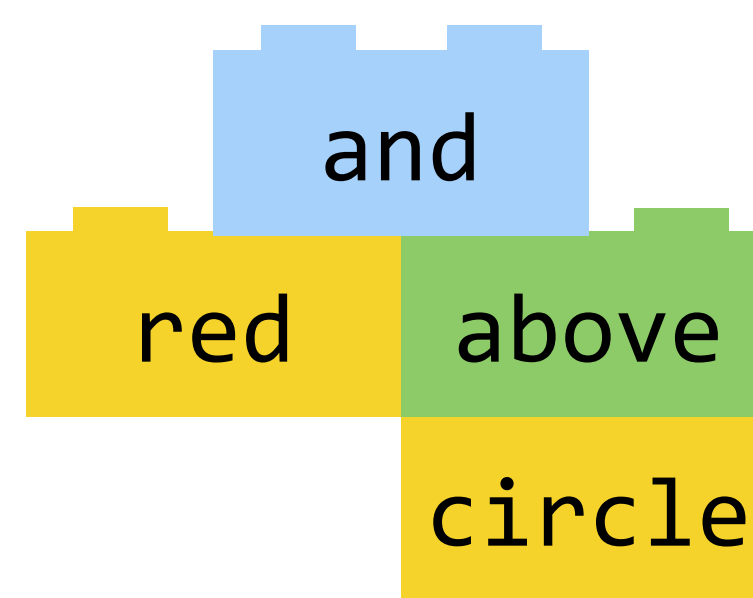
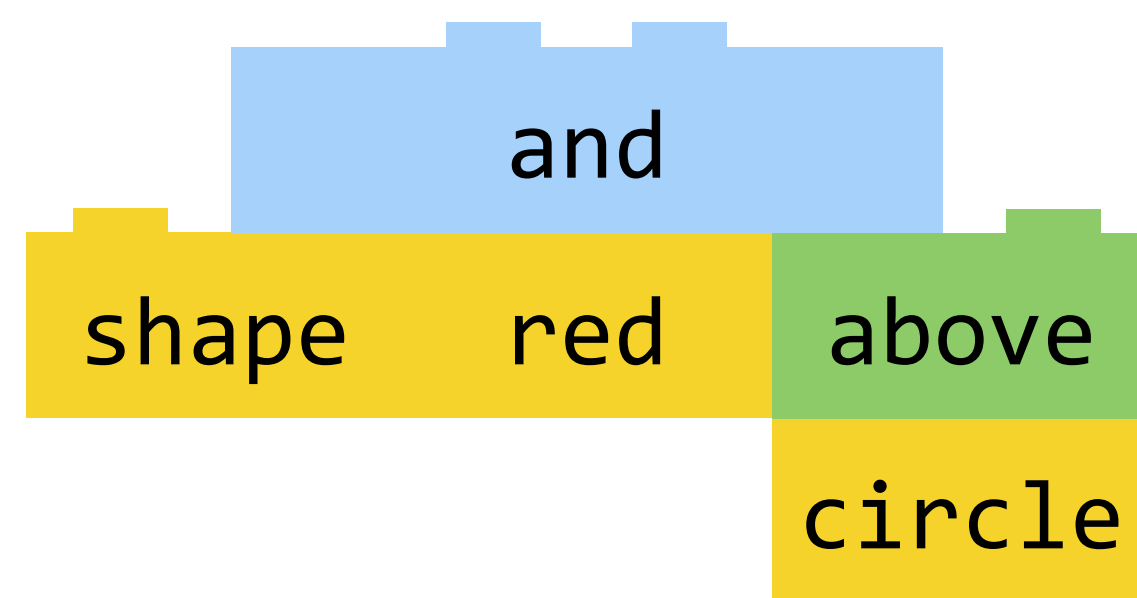




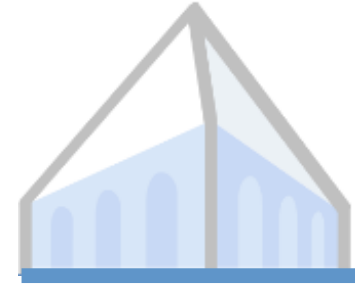


# Where do layouts come from?

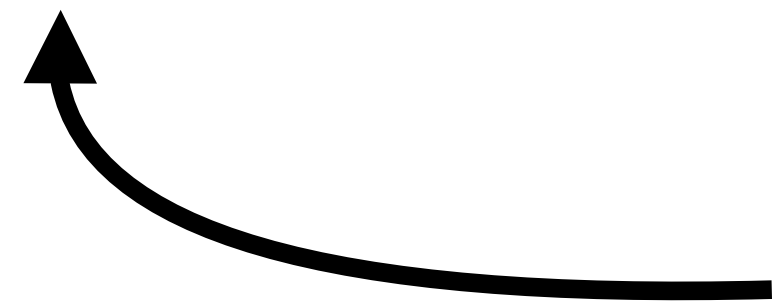
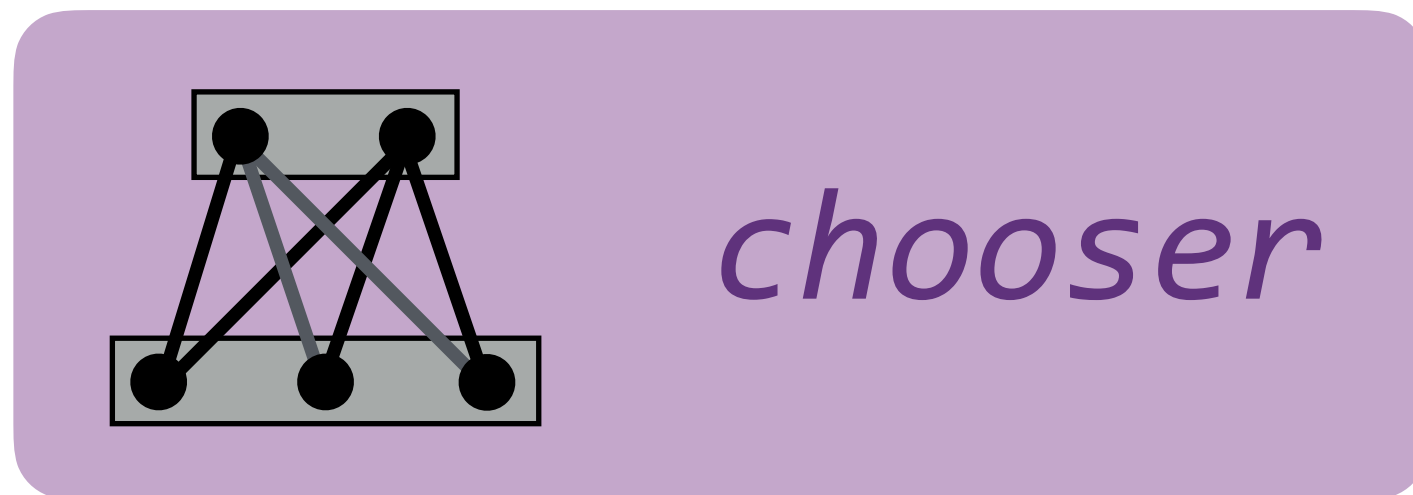
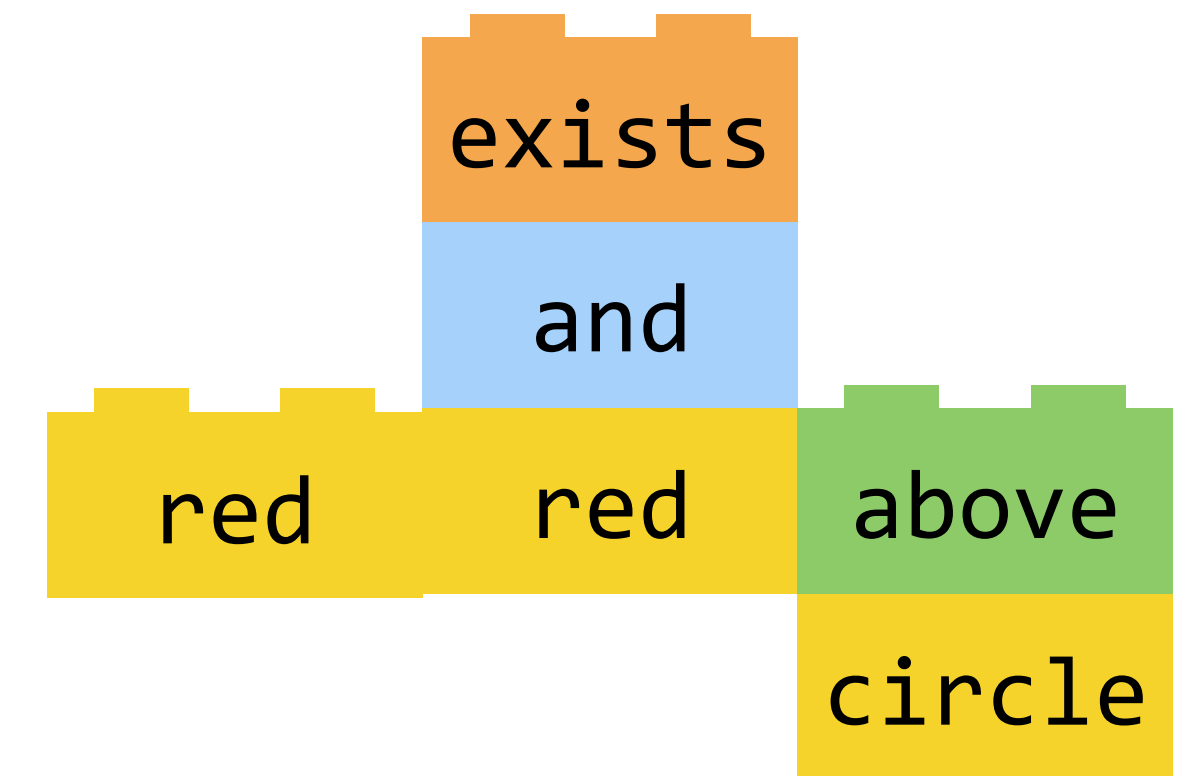
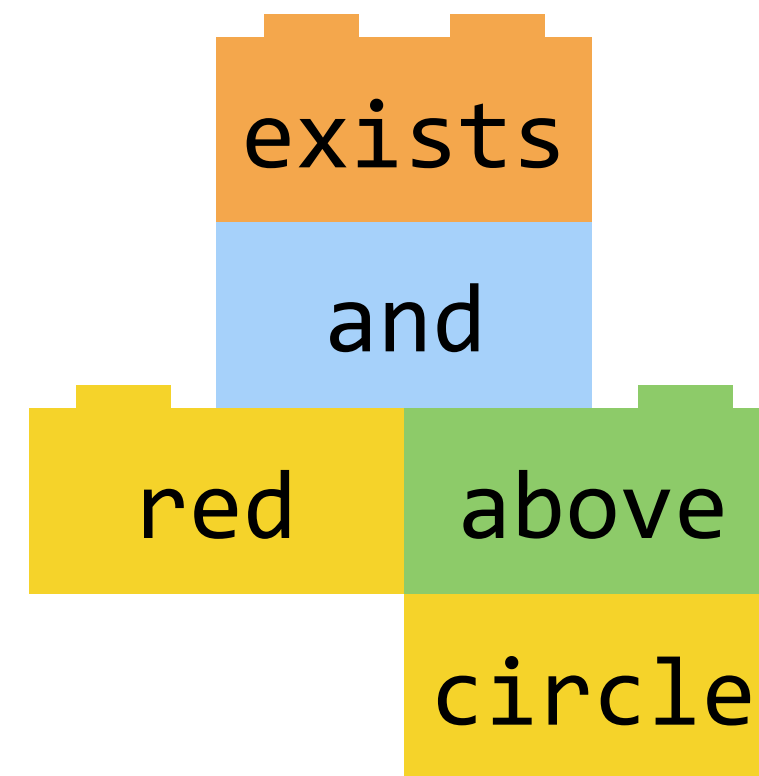
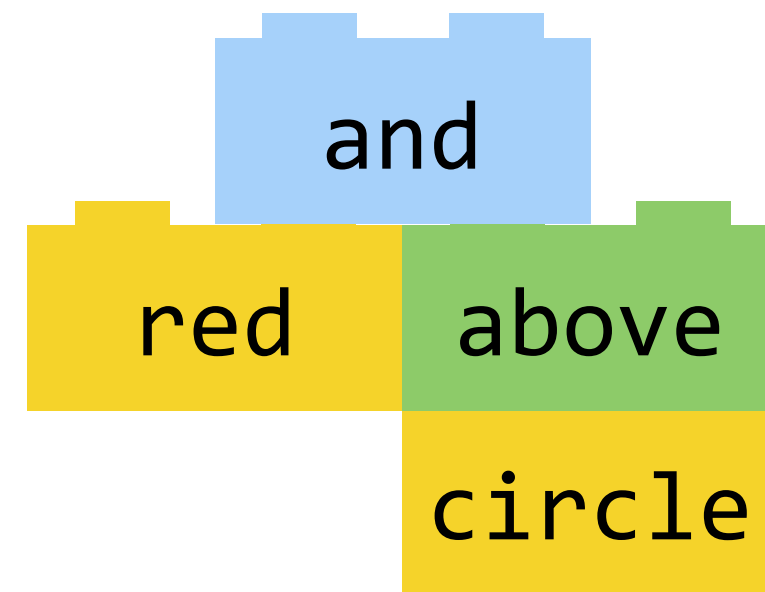
*Is there a red shape above a circle?*



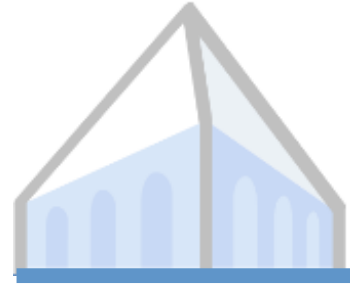
...



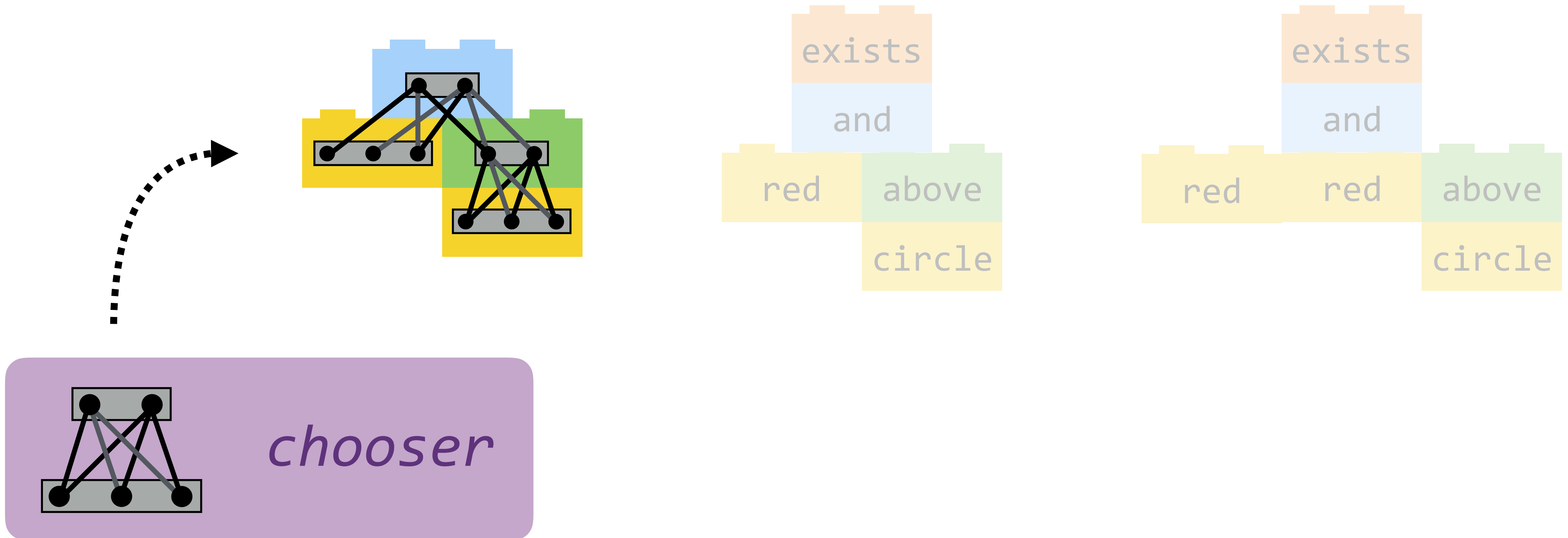
# Choosing among layouts



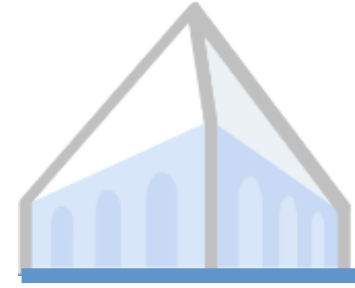
*Is there a red shape above a circle?*



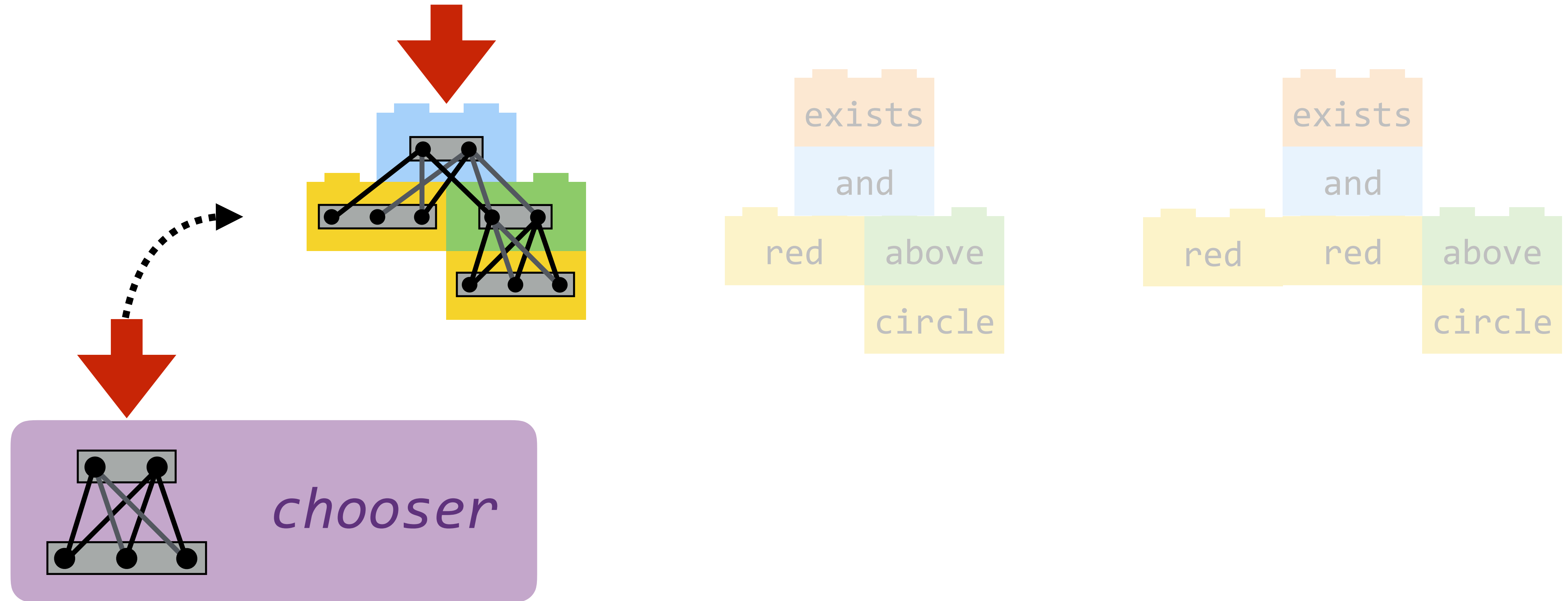
# Learning to choose layouts



*Is there a red shape above a circle?*



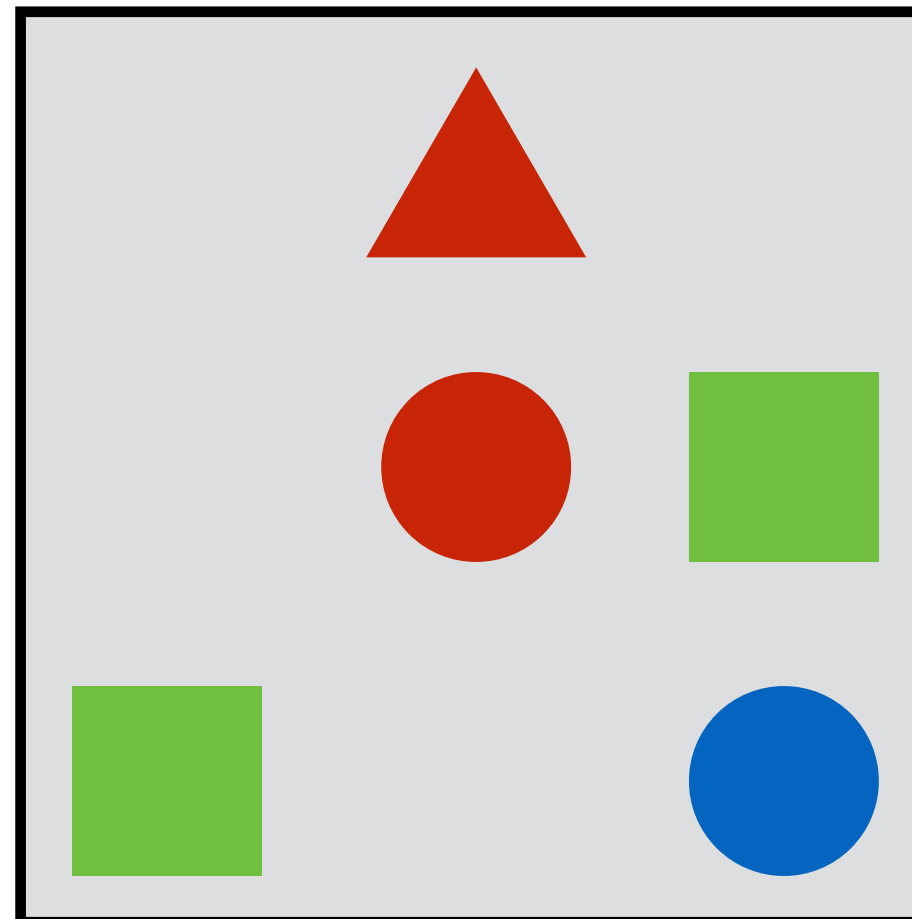
# Learning with unknown layouts uses RL



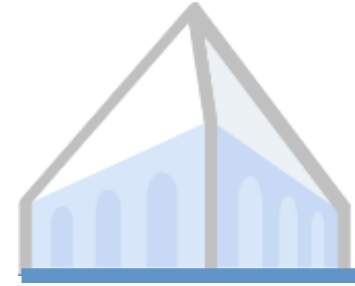
*Is there a red shape above a circle?*



# Experiments

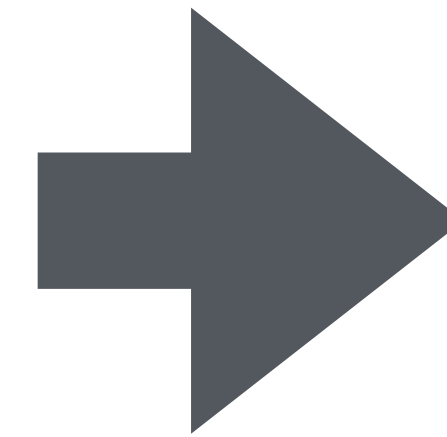


name	type	coastal
<i>Columbia</i>	city	no
<i>Cooper</i>	river	yes
<i>Charleston</i>	city	yes



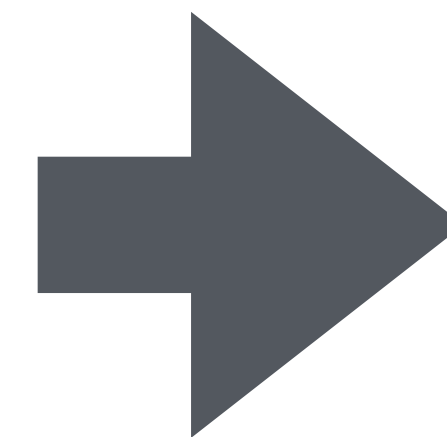
# Experiments: VQA dataset

*What color  
is the necktie?*



*yellow*

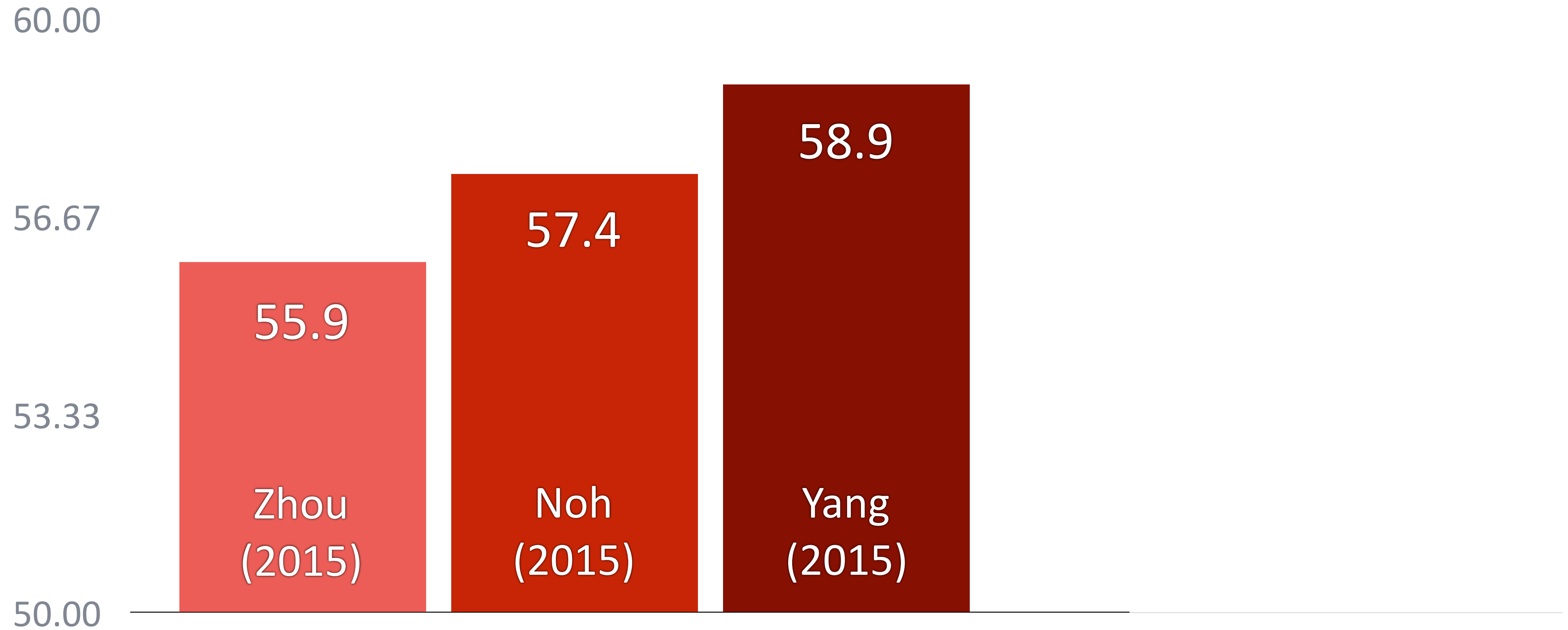
*What is in the  
sheep's ear?*



*tag*

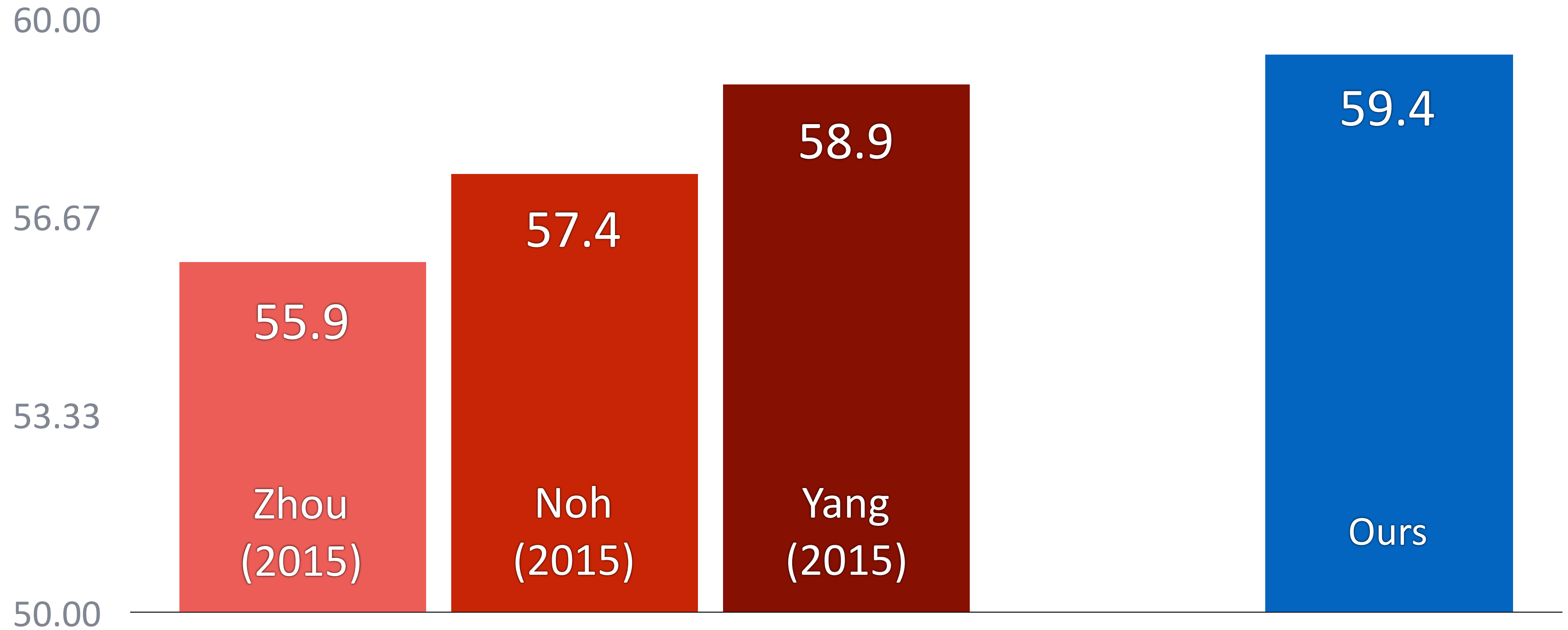


# Experiments: VQA dataset





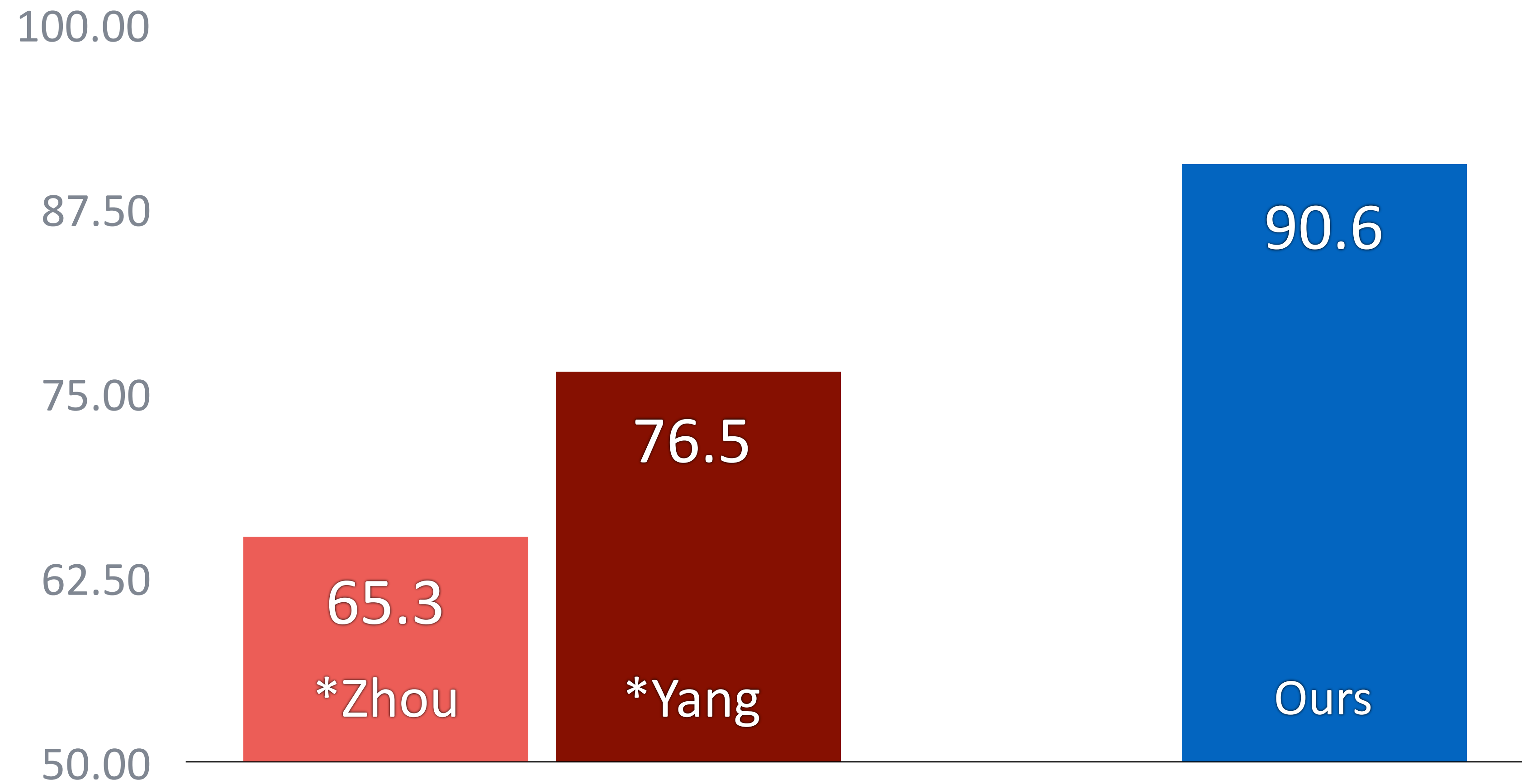
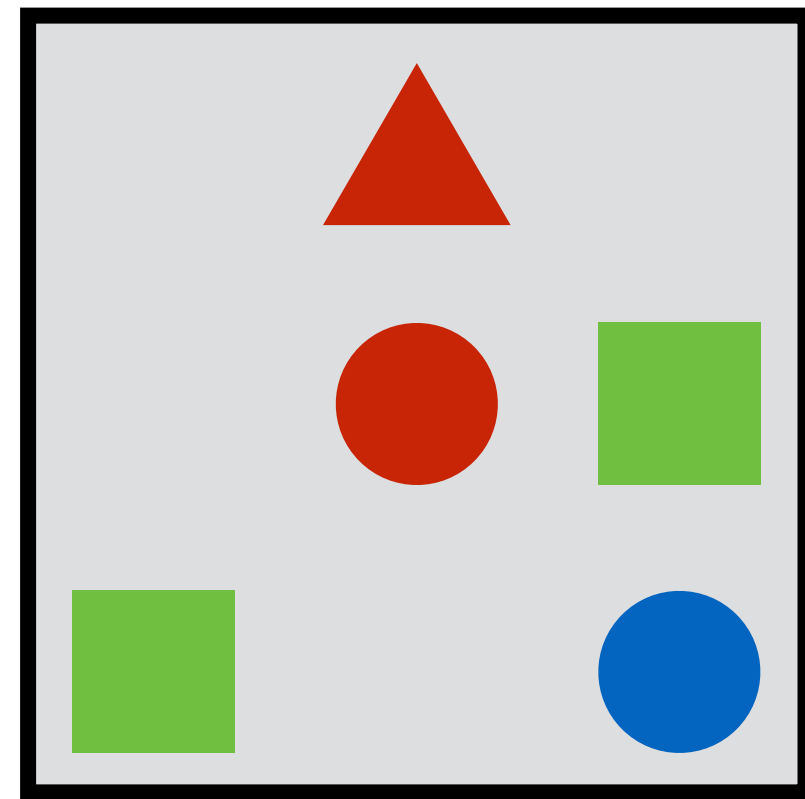
# Experiments: VQA dataset

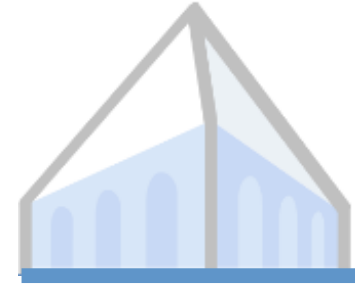






# Experiments: SHAPES dataset





# Experiments: VQA dataset

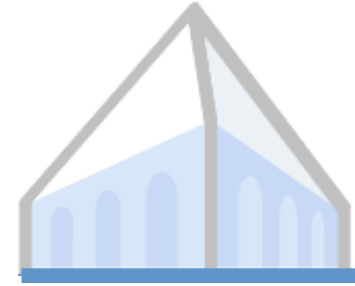
*What color is she wearing?*



color

wear

*white*



# Experiments: VQA Dataset

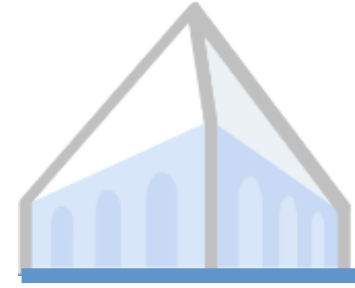
*What color is she wearing?*



color

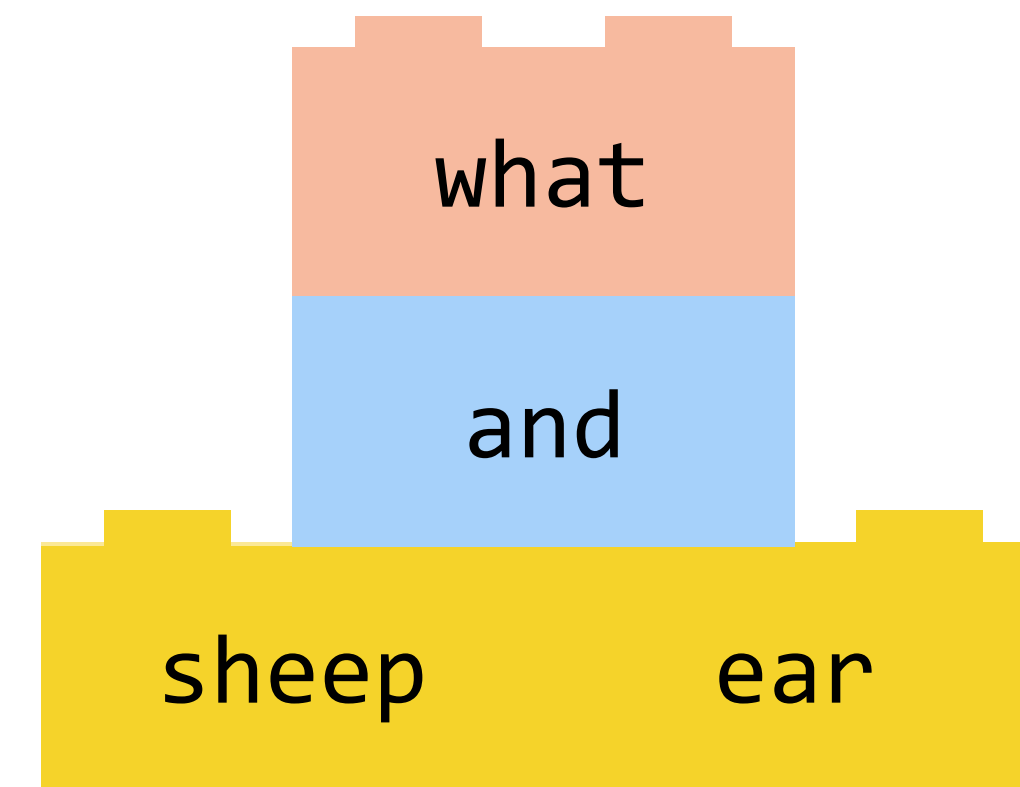
wear

*white*

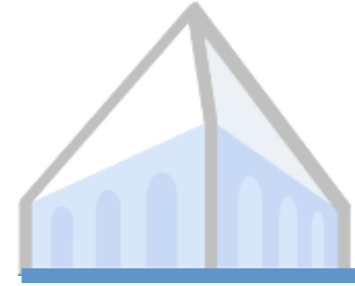


# Experiments: VQA Dataset

*What is in the sheep's ear?*



*tag*



# Experiments: VQA Dataset

What is the color of the sheep's tag?

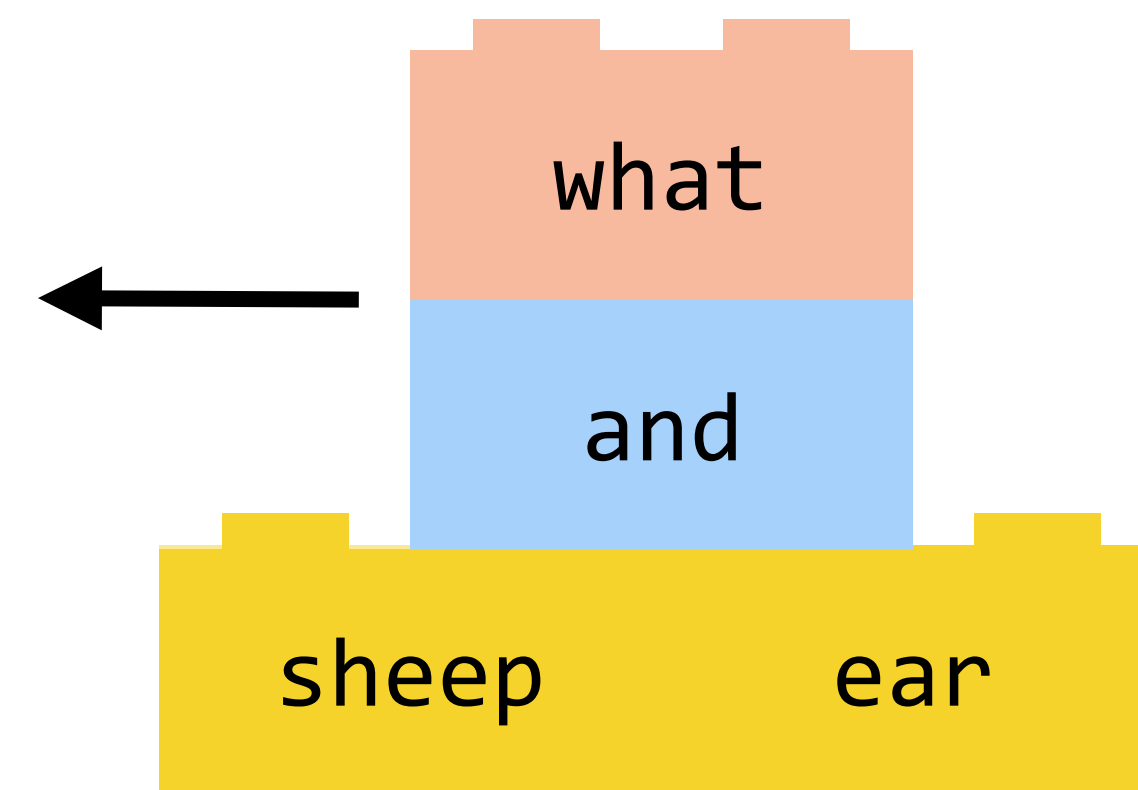


tag

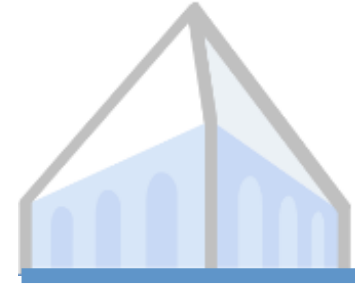


# Experiments: VQA Dataset

*What is in the sheep's ear?*



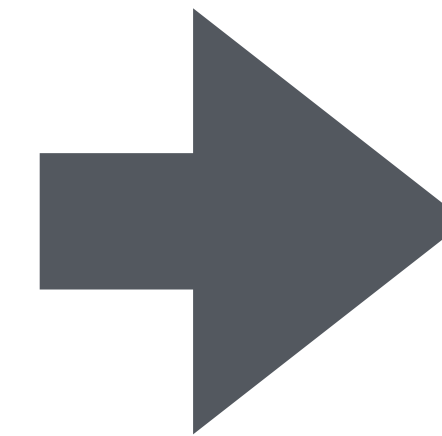
*tag*



# Experiments: GeoQA dataset

*What are some beaches in Florida?*

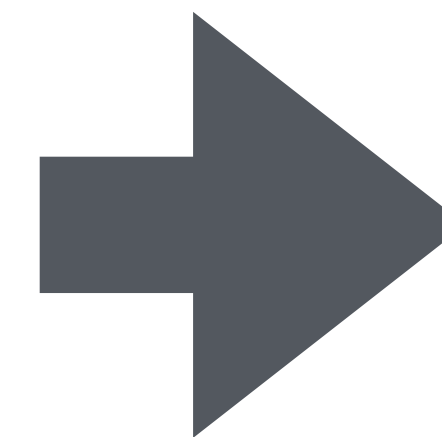
name	type	coastal
<i>Columbia</i>	city	no
<i>Cooper</i>	river	yes
<i>Charleston</i>	city	yes



*Daytona Beach*

*Is Key Largo an island?*

name	type	coastal
<i>Miami</i>	city	no
<i>Daytona Beach</i>	city	yes
<i>Everglades</i>	park	no

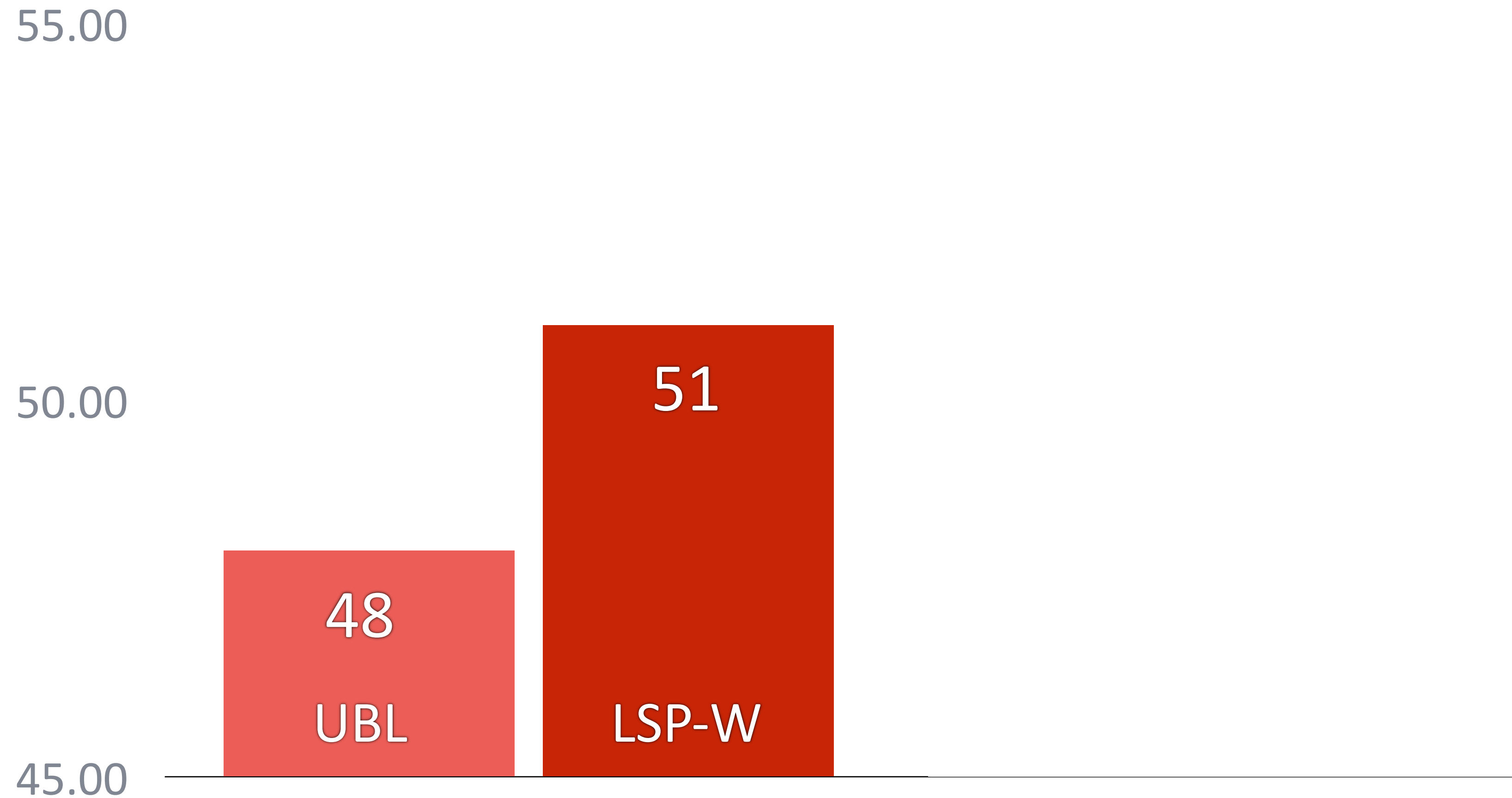


*Yes*



# Experiments: GeoQA dataset

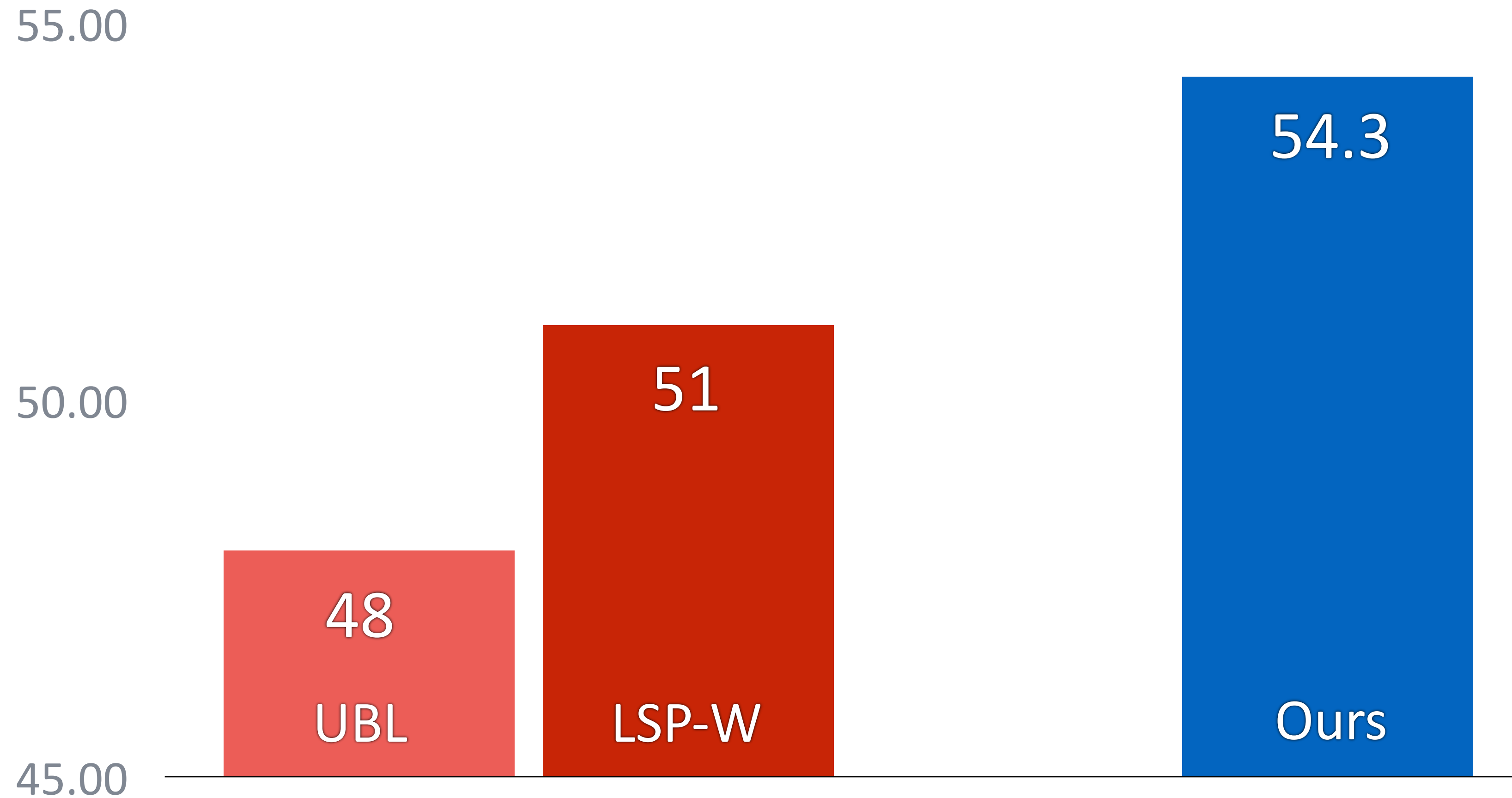
---

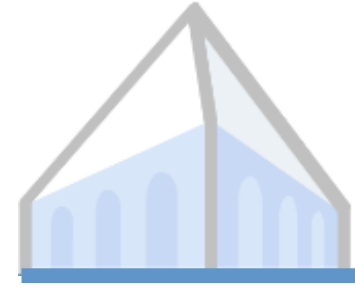






# Experiments: GeoQA dataset

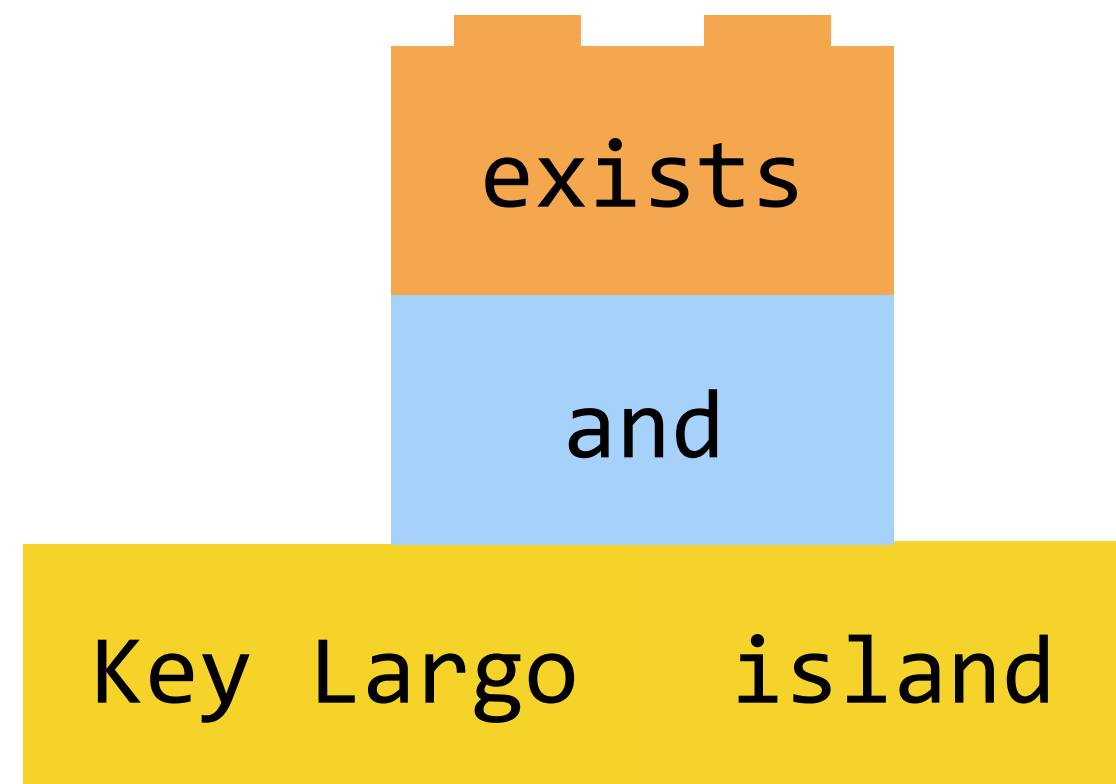




# Experiments: GeoQA dataset

---

*Is Key Largo an island?*

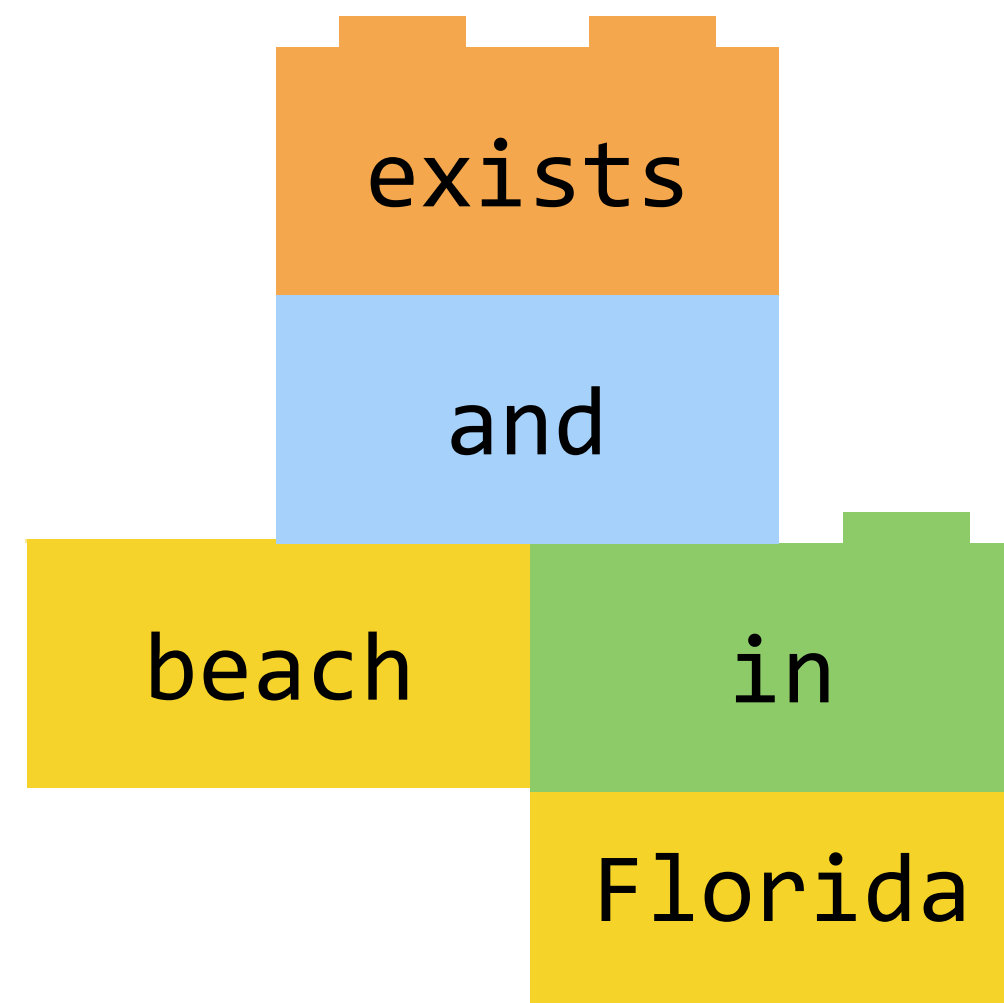


*yes*



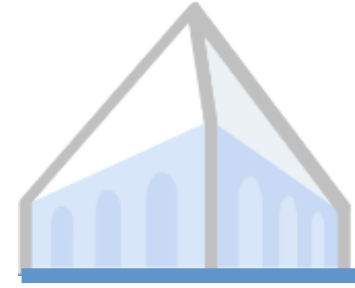
# Experiments: GeoQA dataset

*What are some  
beaches in Florida?*



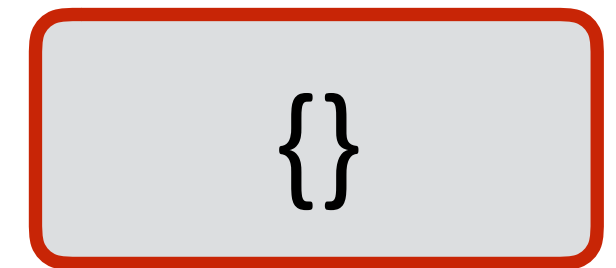
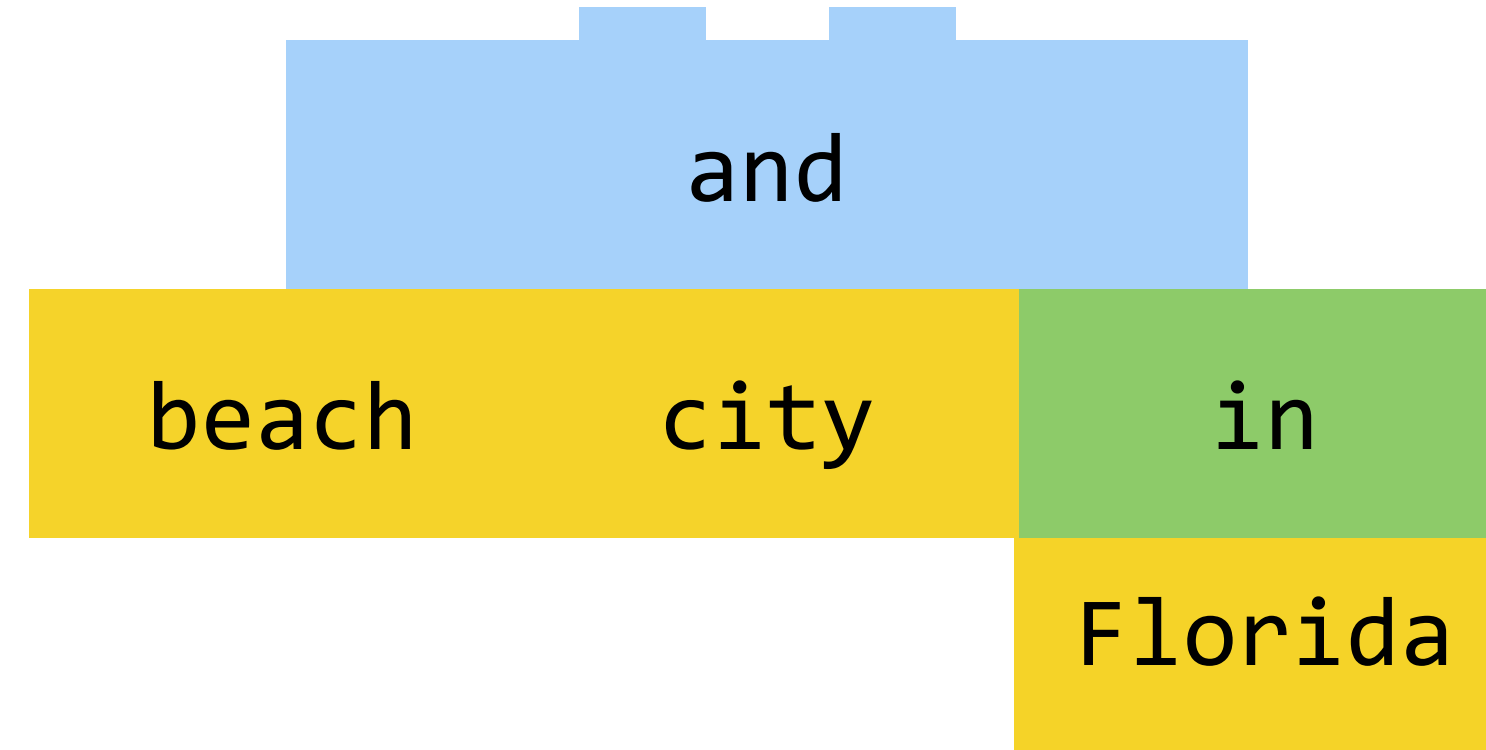
*yes*

(wrong parse)

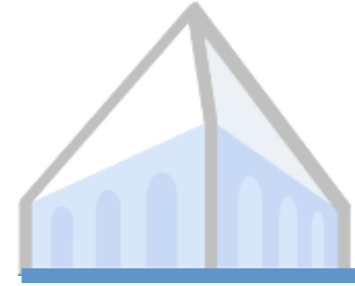


# Experiments: GeoQA dataset

*What beach city is there in Florida?*

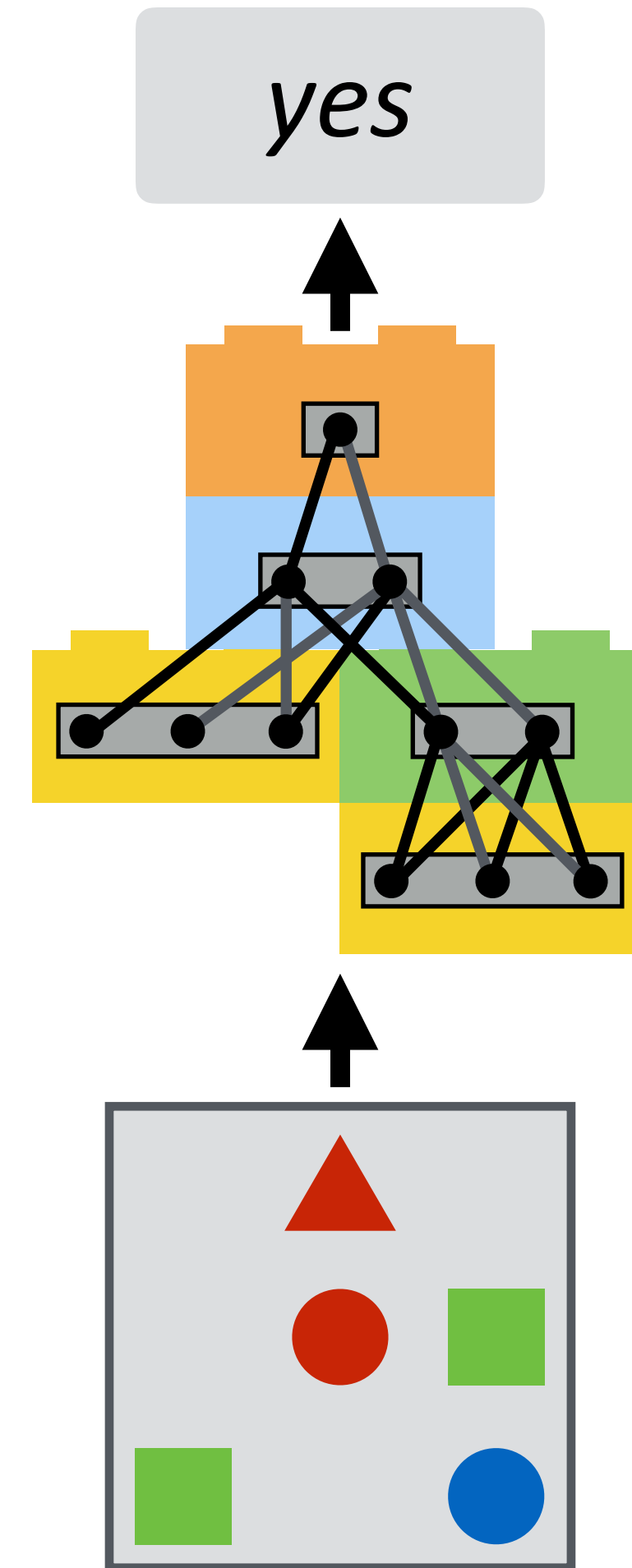
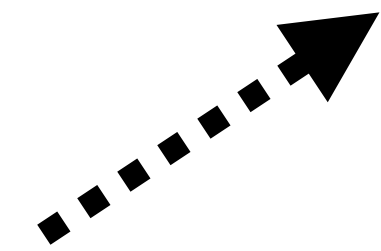
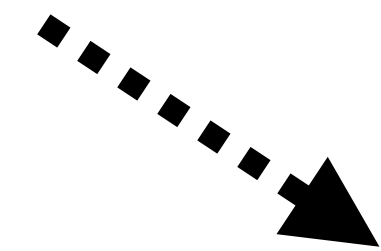
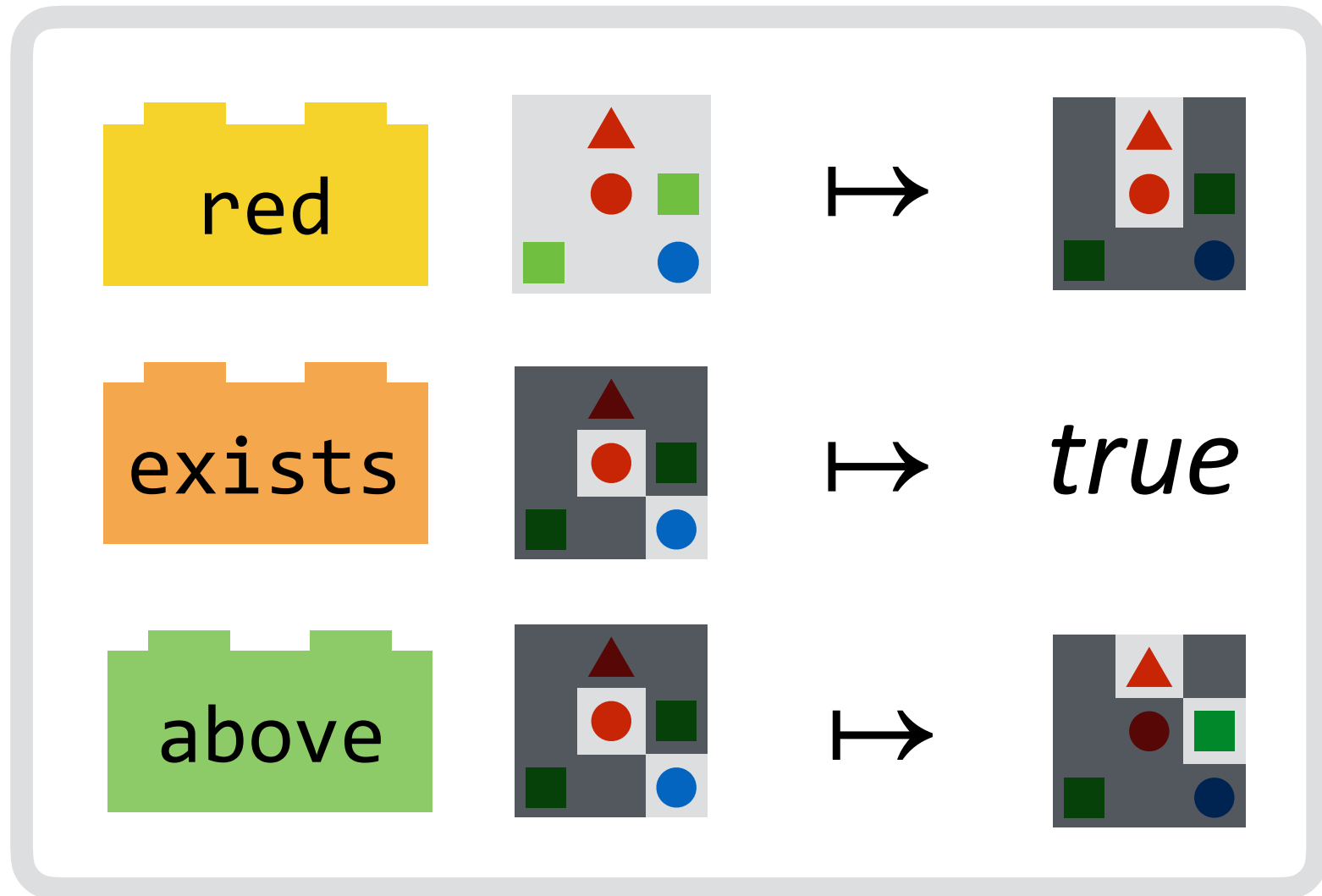


(wrong module behavior)



# Neural module networks

*Is there a red shape above a circle?*

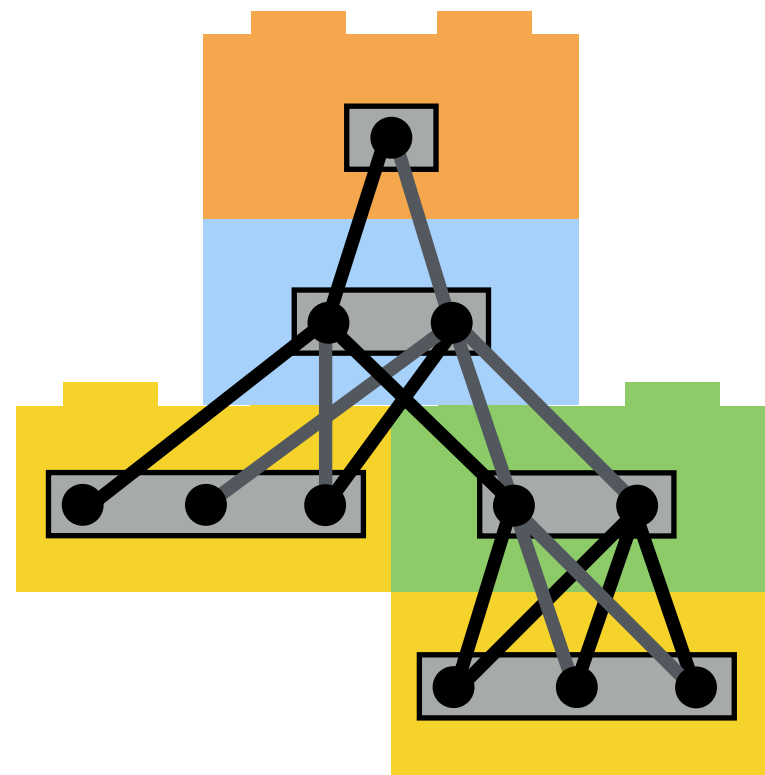




# Neural module networks

---

## Linguistic structure dynamically generates model structure



Combines advantages of:

- Representation learning (like a neural net)
- Compositionality (like a semantic parser)

thank you

Download our code at <http://github.com/jacobandreas/nmn2>