Neural Module Networks

Jacob Andreas, Marcus Rohrbach,
Trevor Darrell, Dan Klein

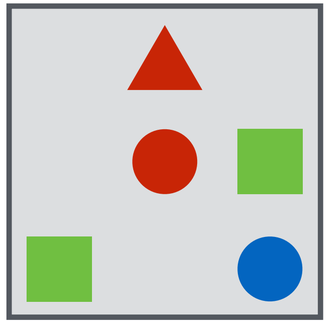
Visual question answering

What color is the necktie?

yellow

Visual question answering

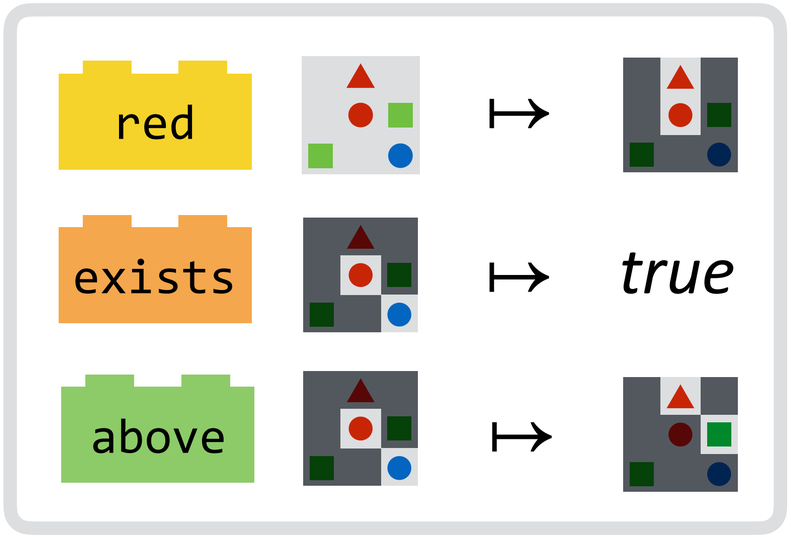
Is there a red shape above a circle?



yes

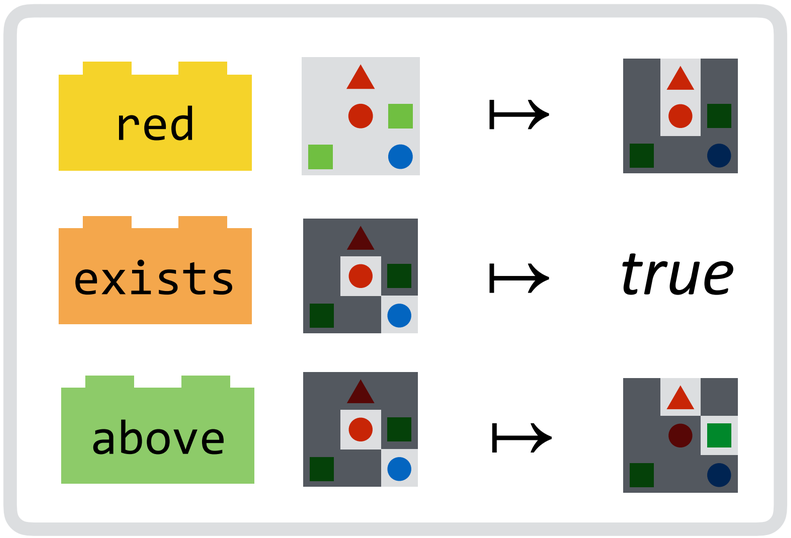
Neural module networks

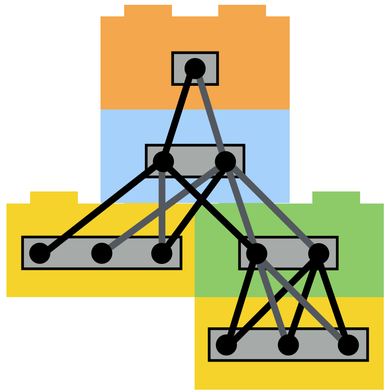
Is there a red shape above a circle?



Neural module networks

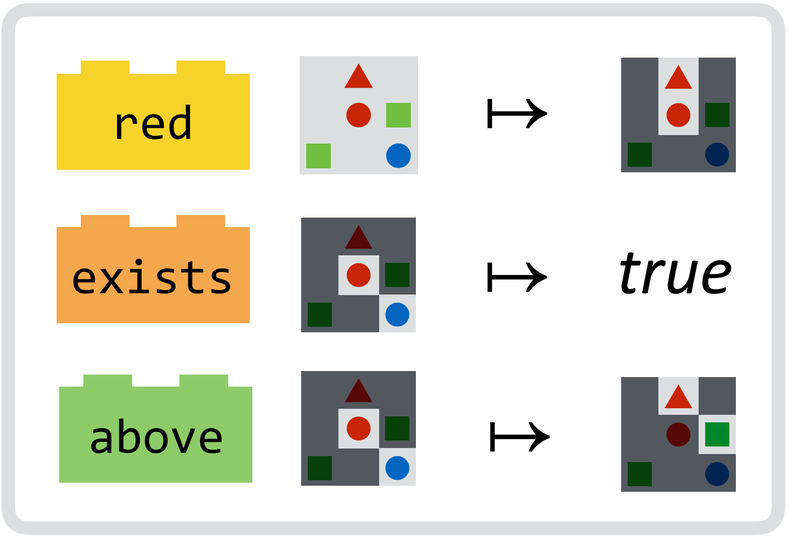
Is there a red shape above a circle?

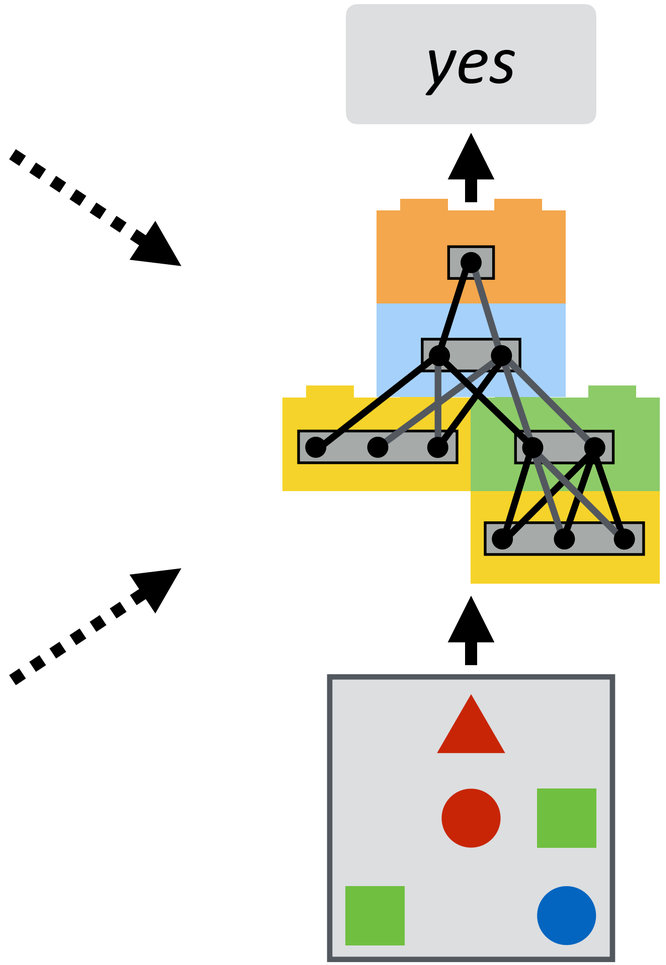




Neural module networks

Is there a red shape above a circle?





Nearest neighbors

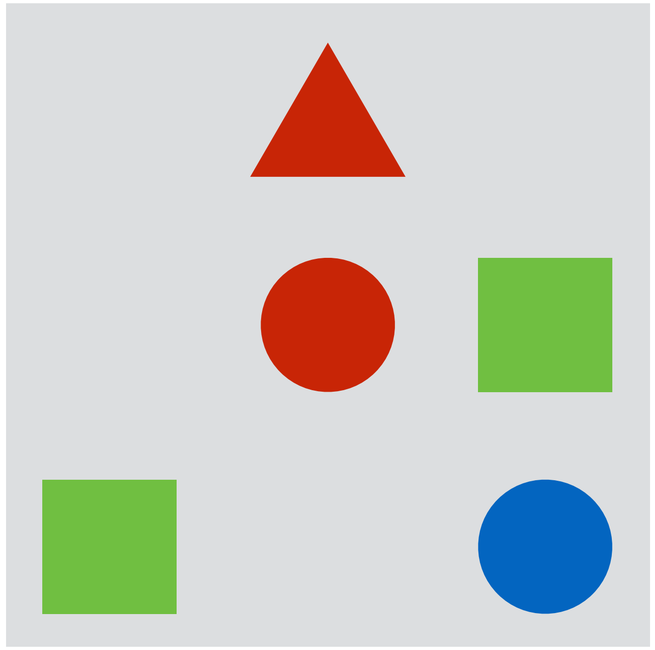
Structured neural models

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

Probabilistic formal semantics / predicate learning

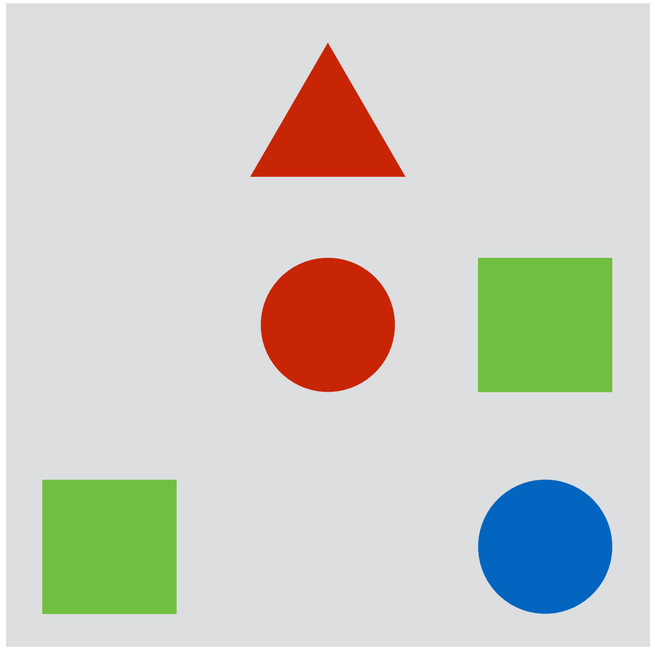
- [Beltagy et al. 2013, Lewis & Steedman 2013, Malinowski & Fritz 2014]

Representing meaning



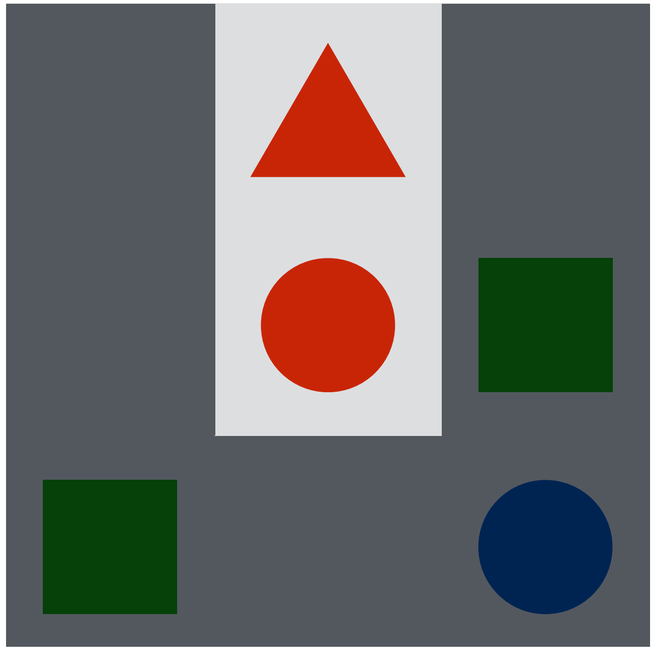
Is there a red shape above a circle?

Representing meaning



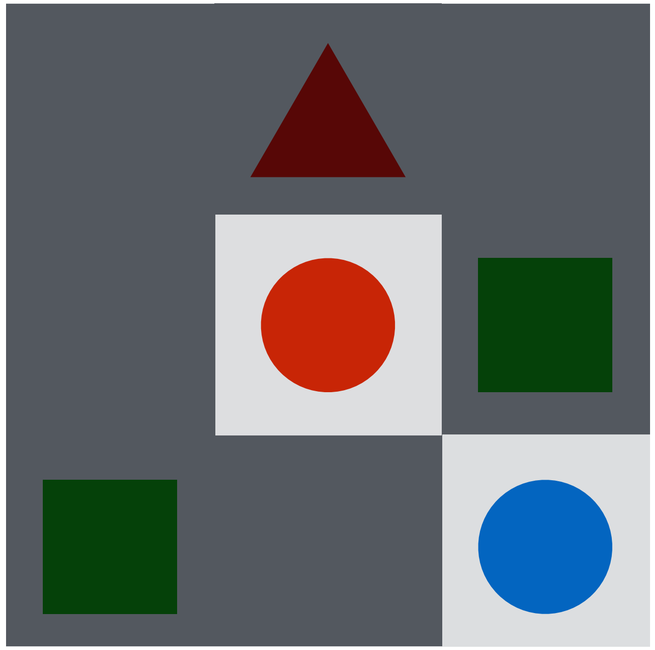
Is there a red shape above a circle?

Attentions encode meaning



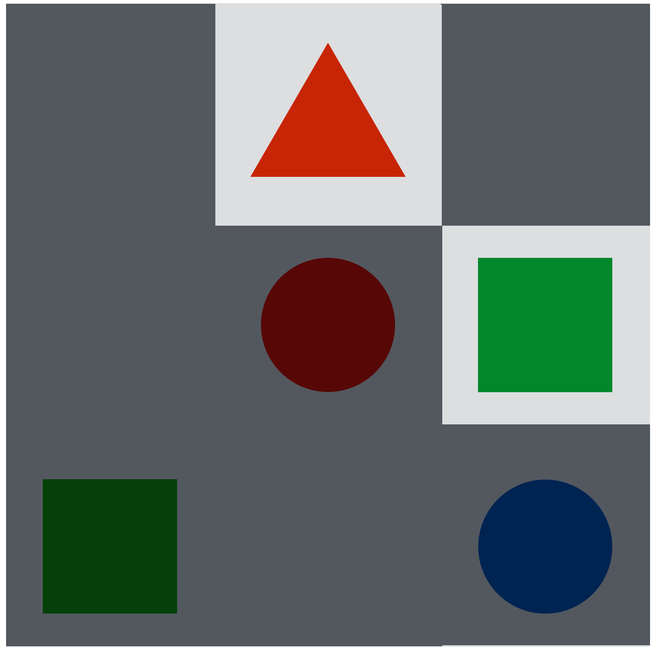
Is there a red shape above a circle?

Attention transformations encode meaning



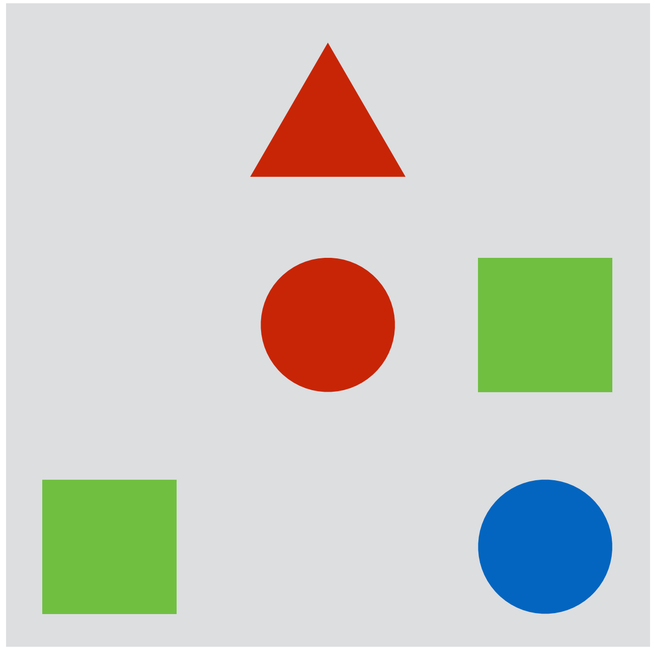
Is there a red shape above a circle?

Set transformations encode meaning

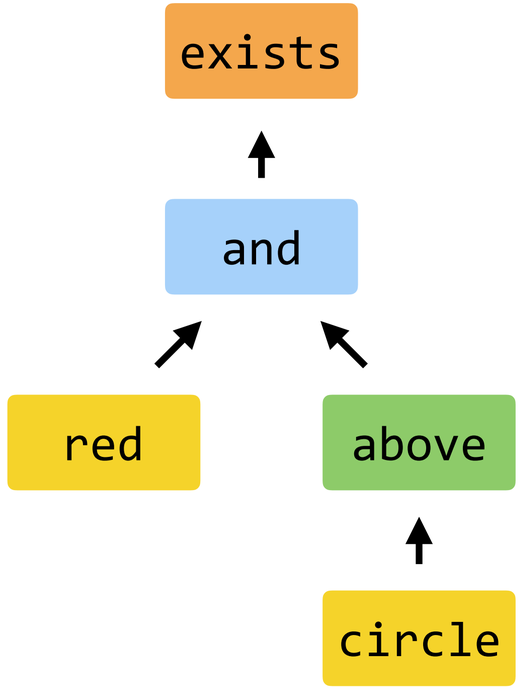


Is there a red shape above a circle?

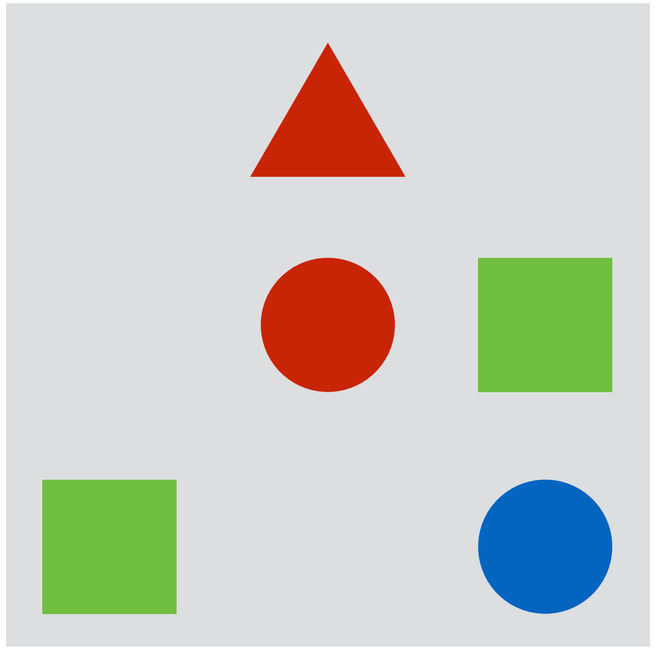
Sentence meanings are computations



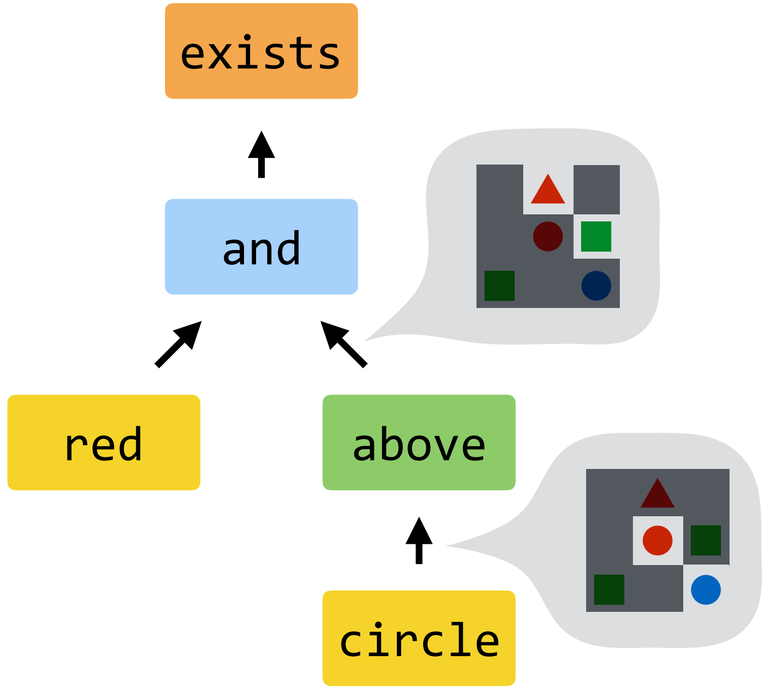
Is there a red shape above a circle?



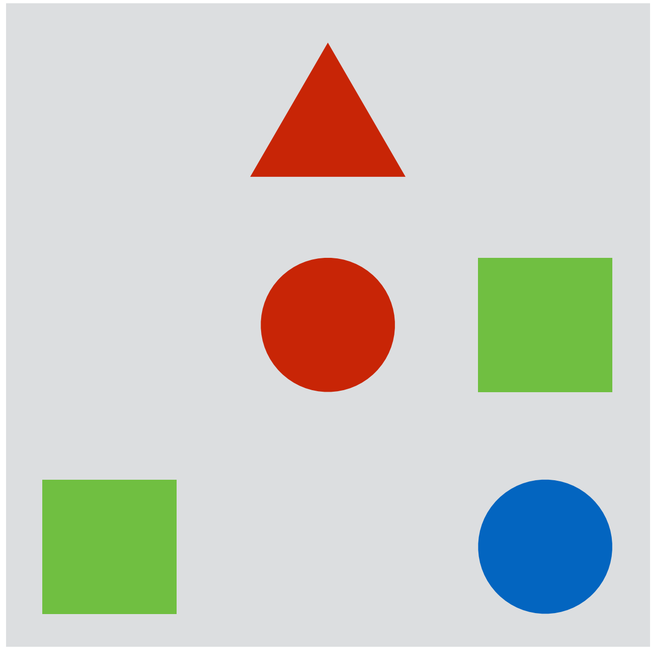
Sentence meanings are computations

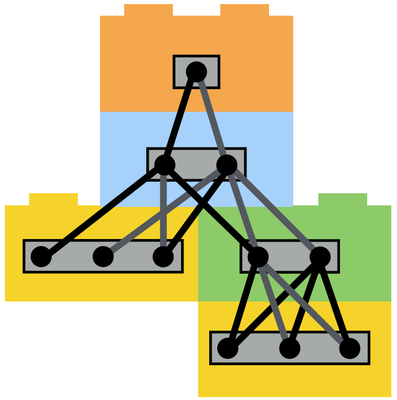


Is there a red shape above a circle?



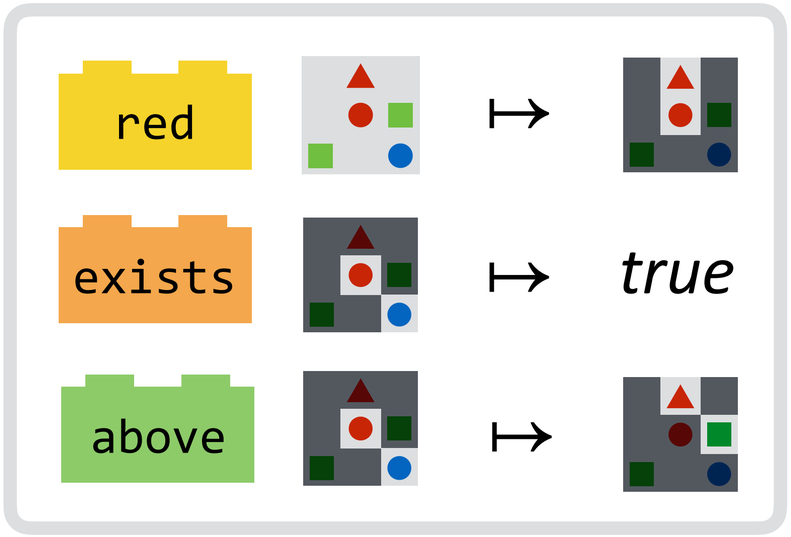
Compositions of vector functions are neural nets

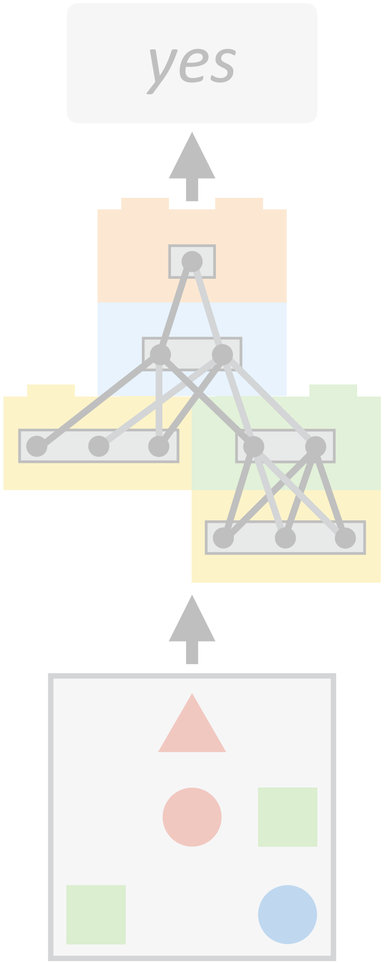




Outline

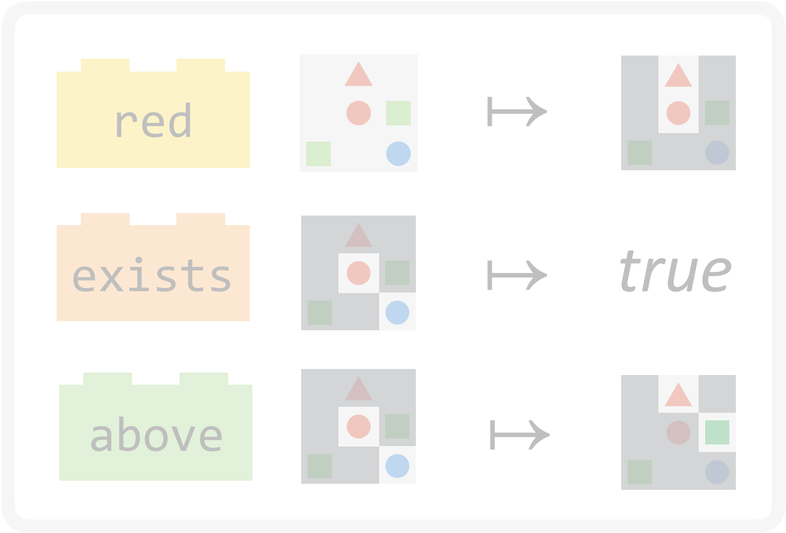
Is there a red shape above a circle?

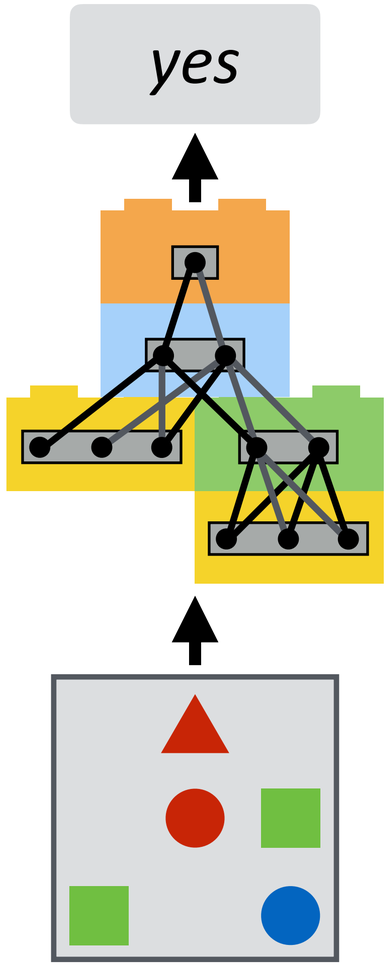




Outline

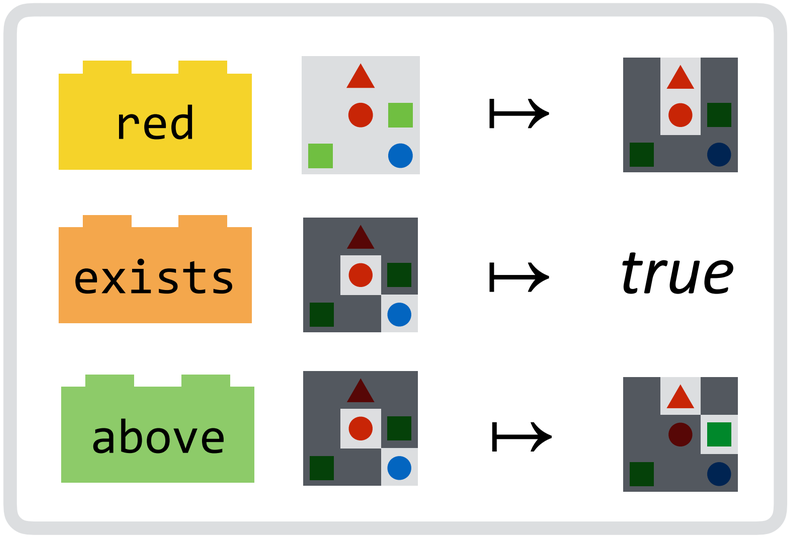
Is there a red shape above a circle?

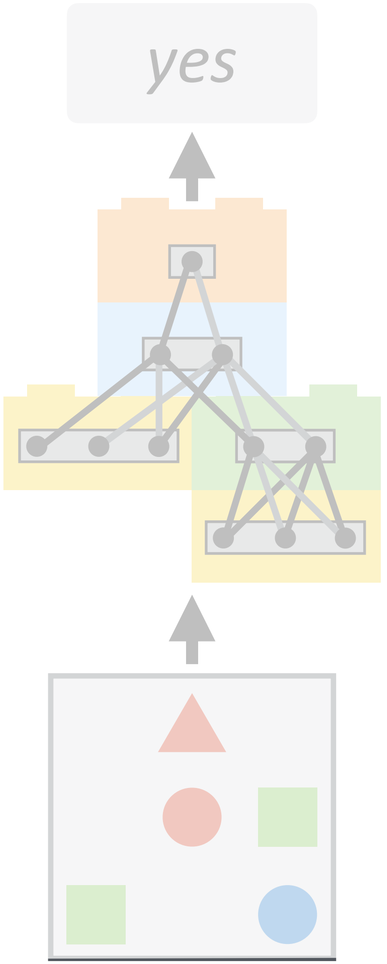




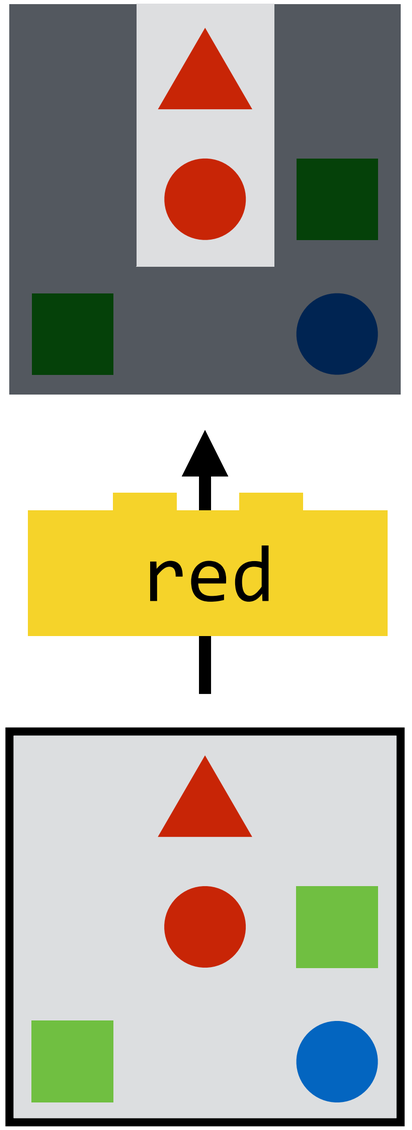
Outline

Is there a red shape above a circle?





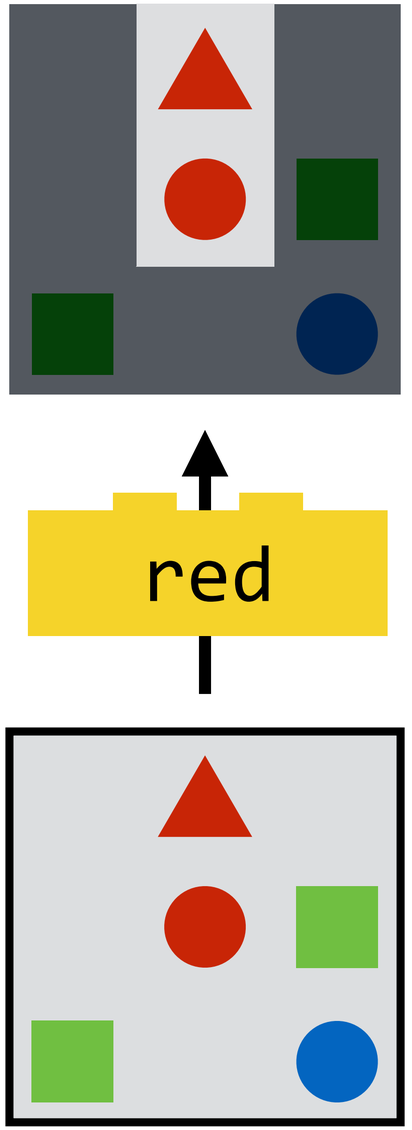
The [find] module



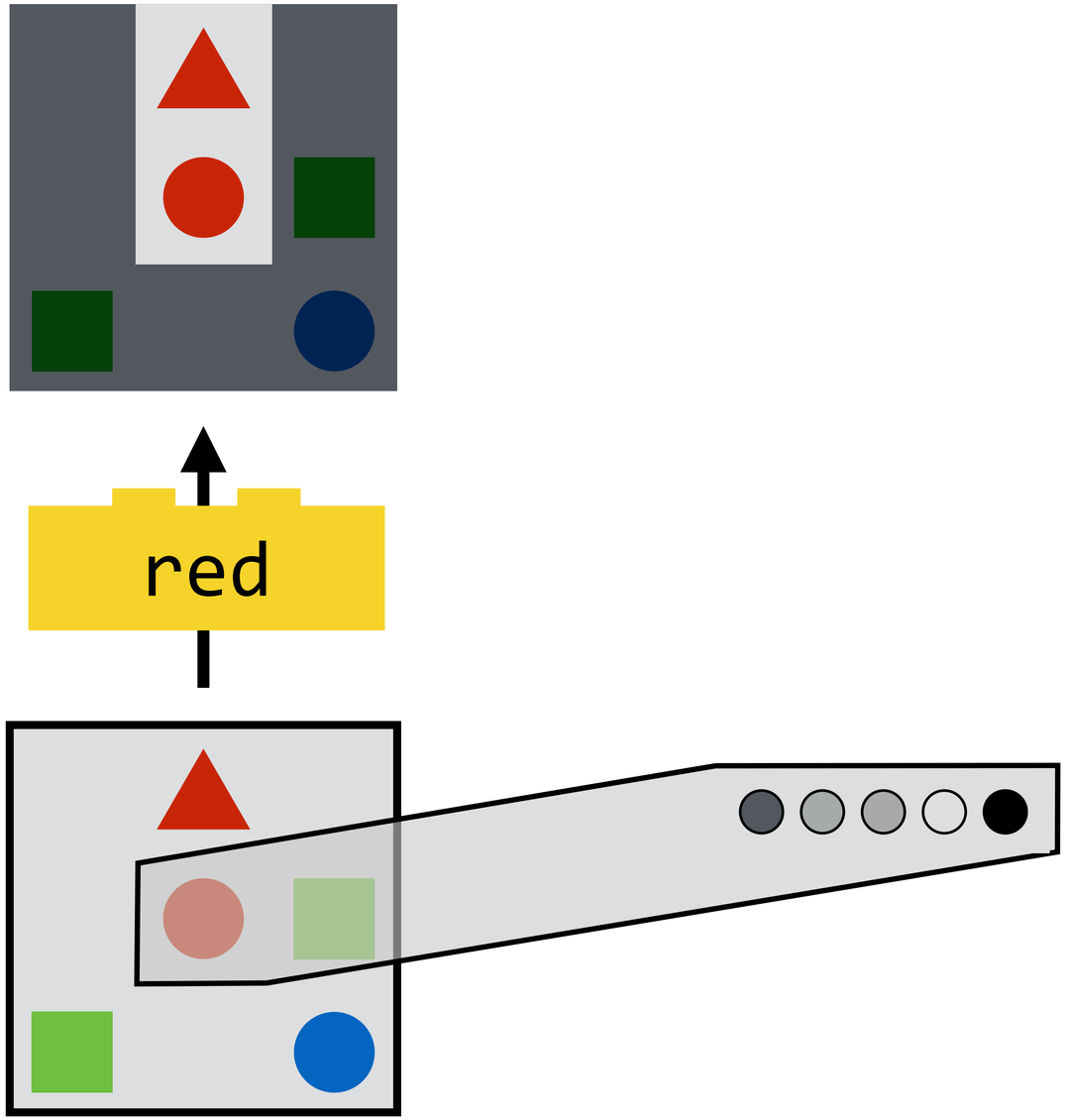
The [find] module

necktie

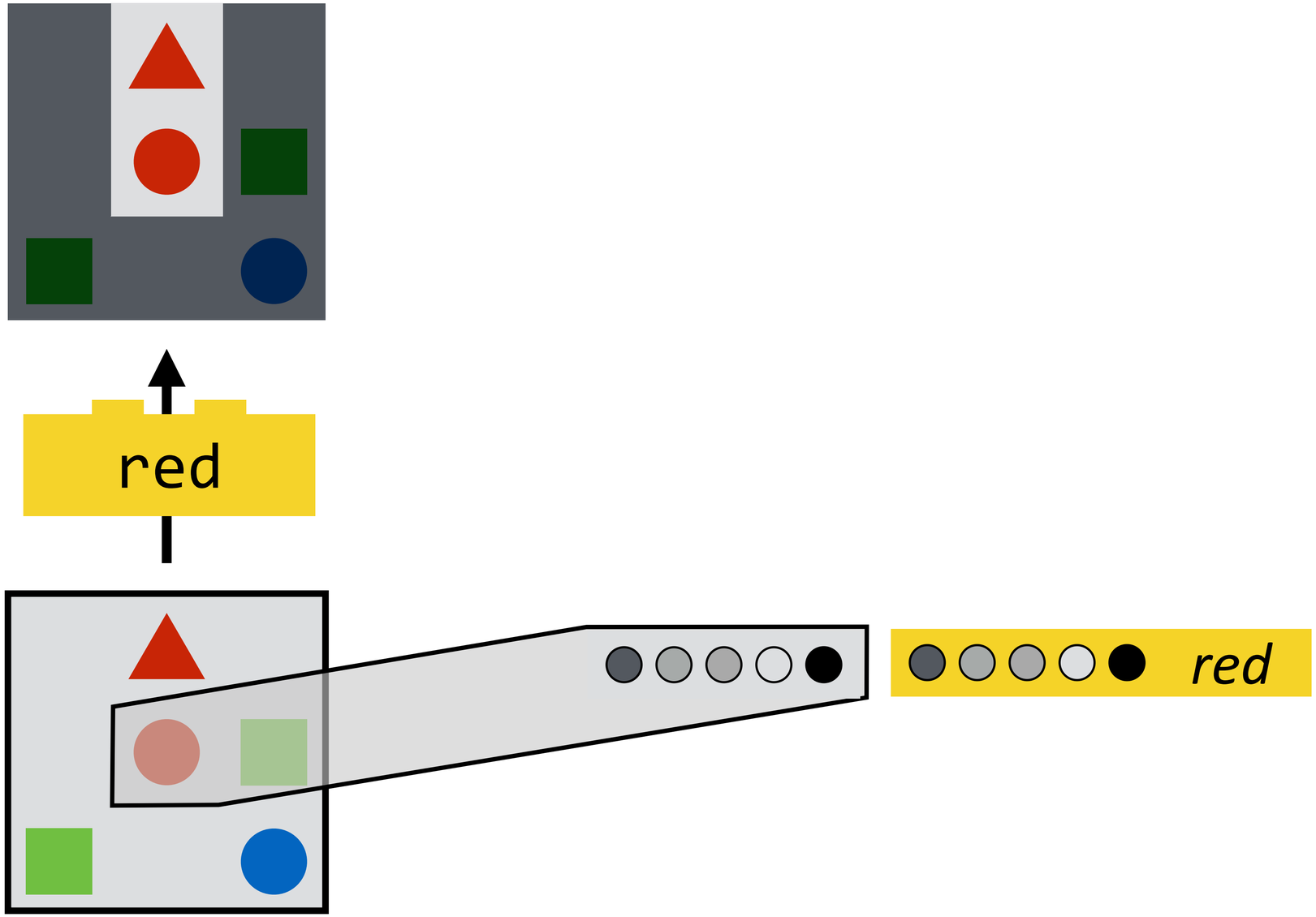
The [find] module



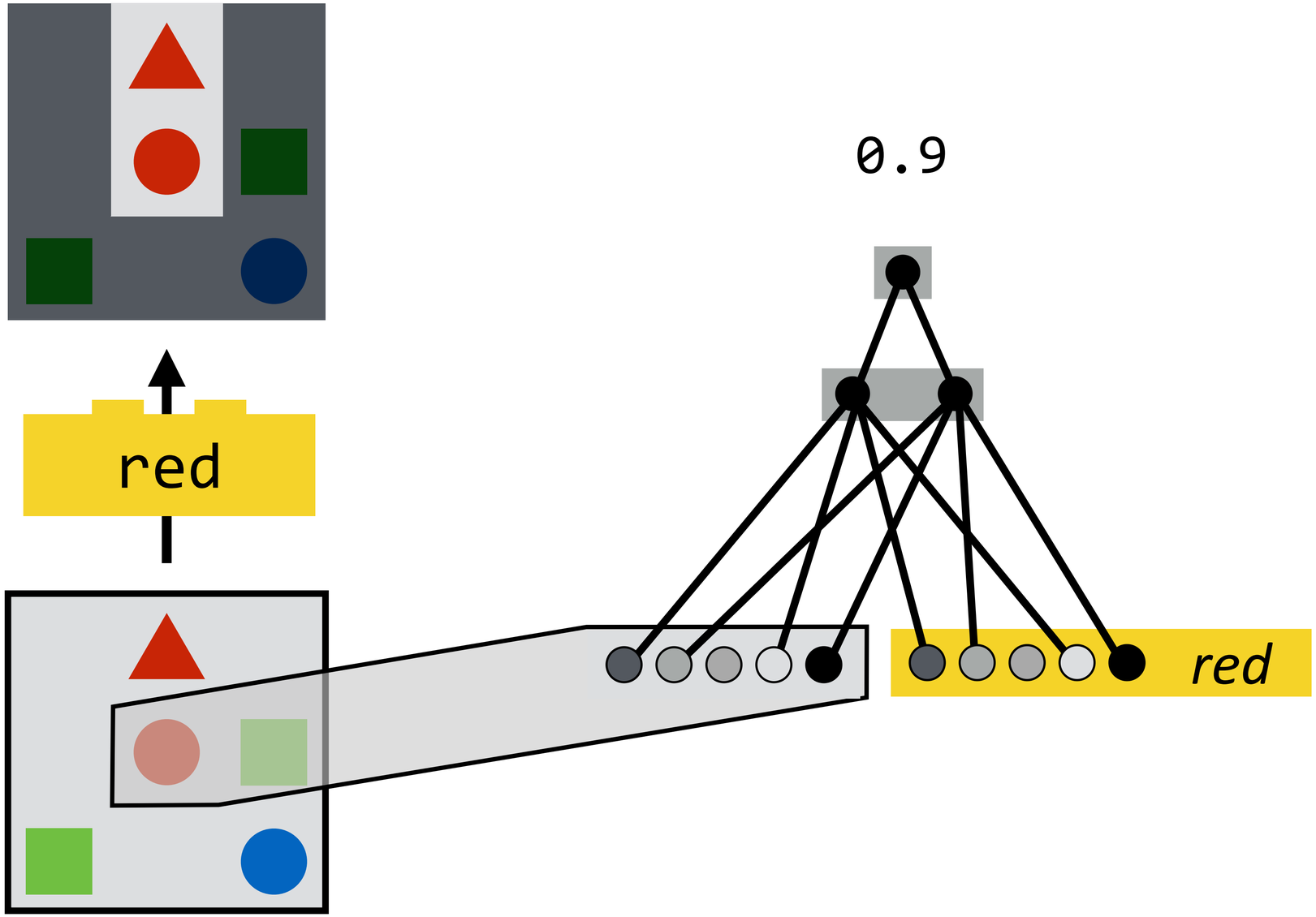
The [find] module



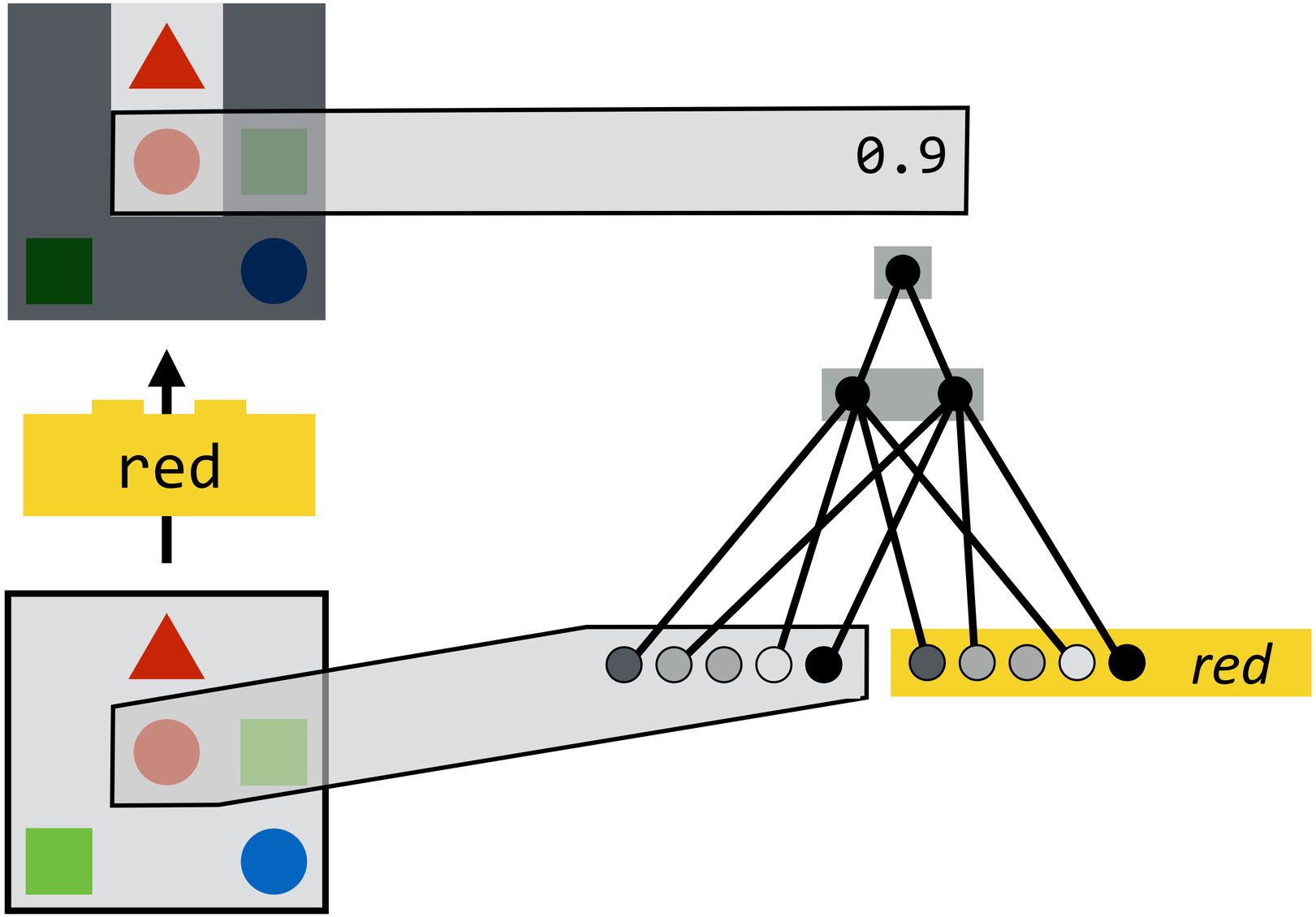
The [find] module



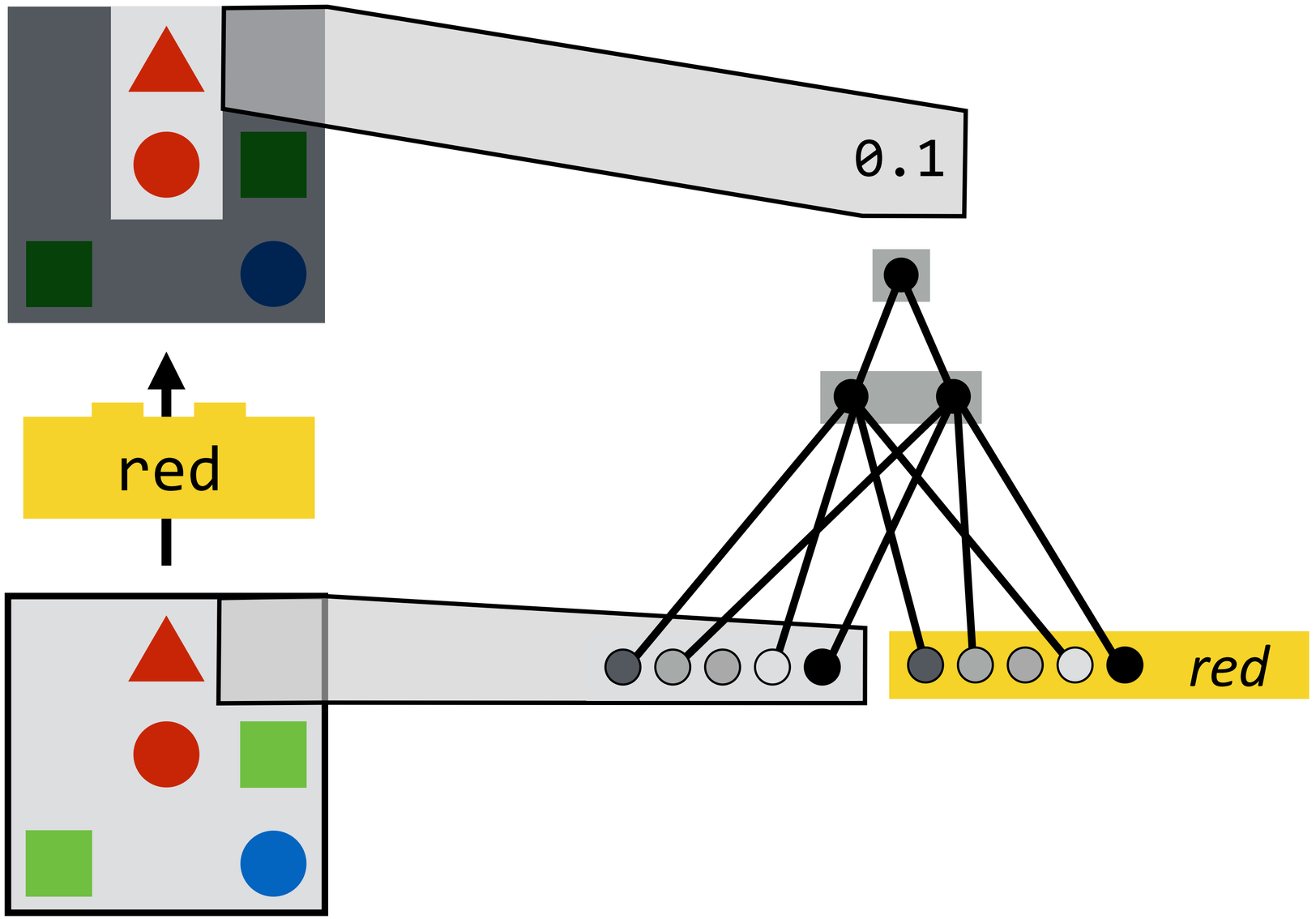
The [find] module



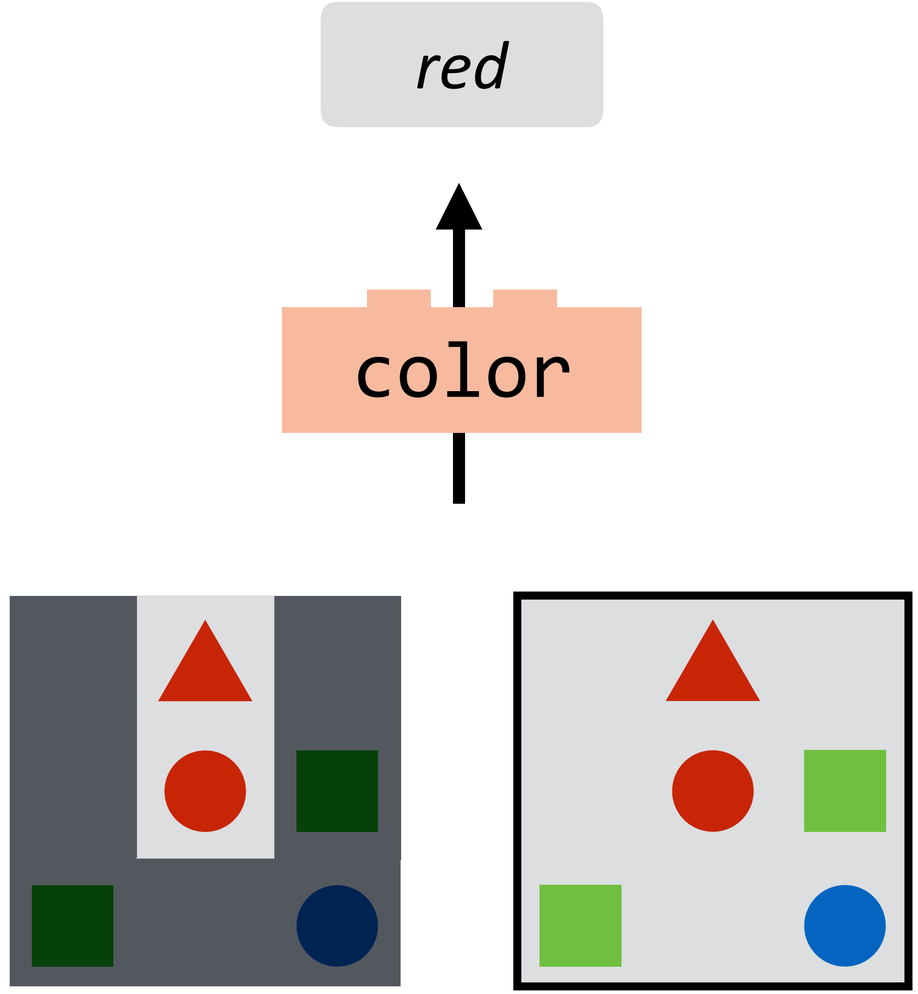
The [find] module



The [find] module

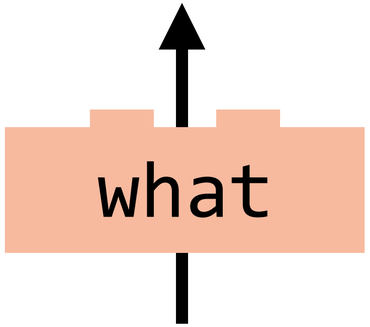


The [describe] module

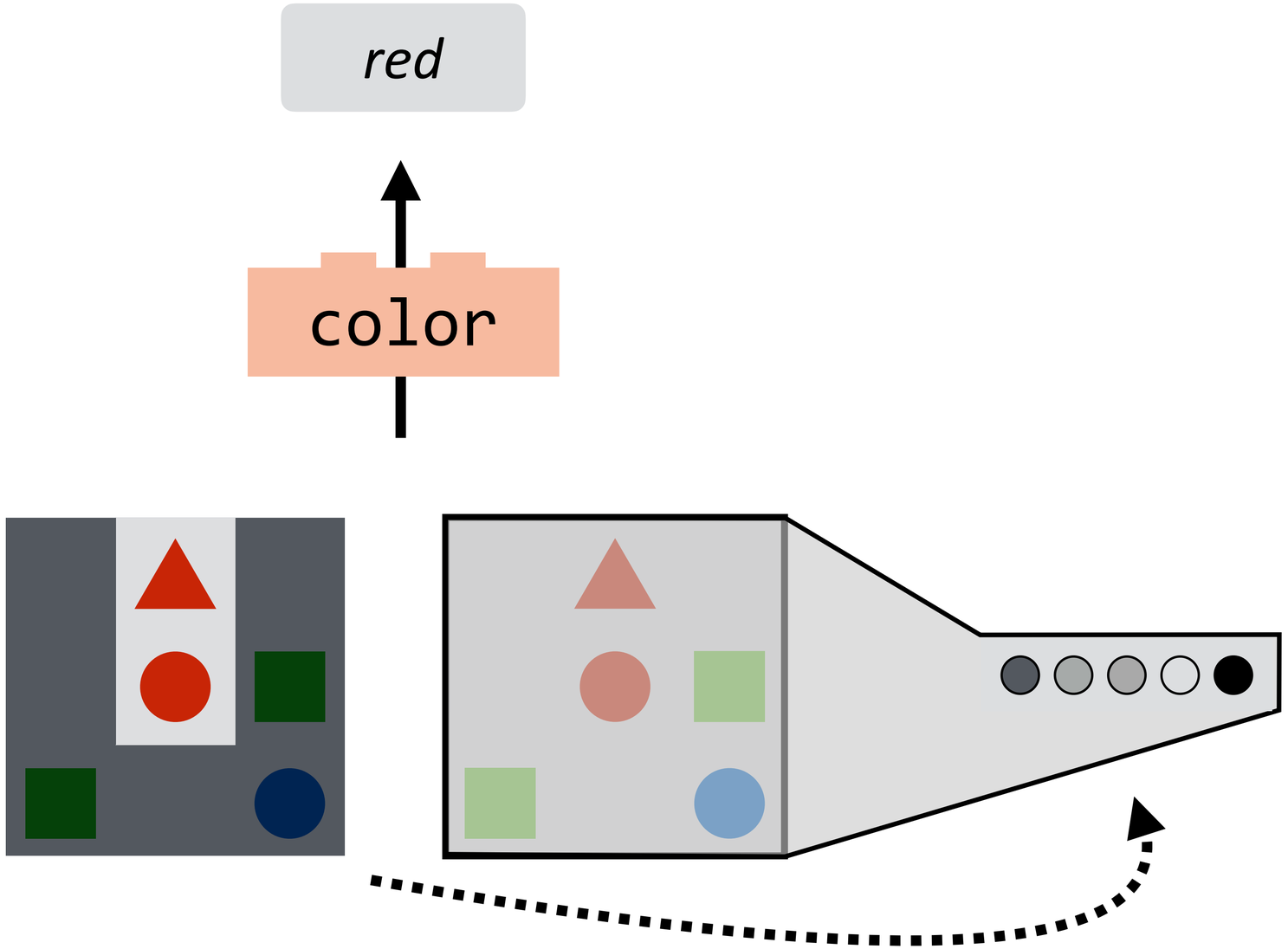


The [describe] module

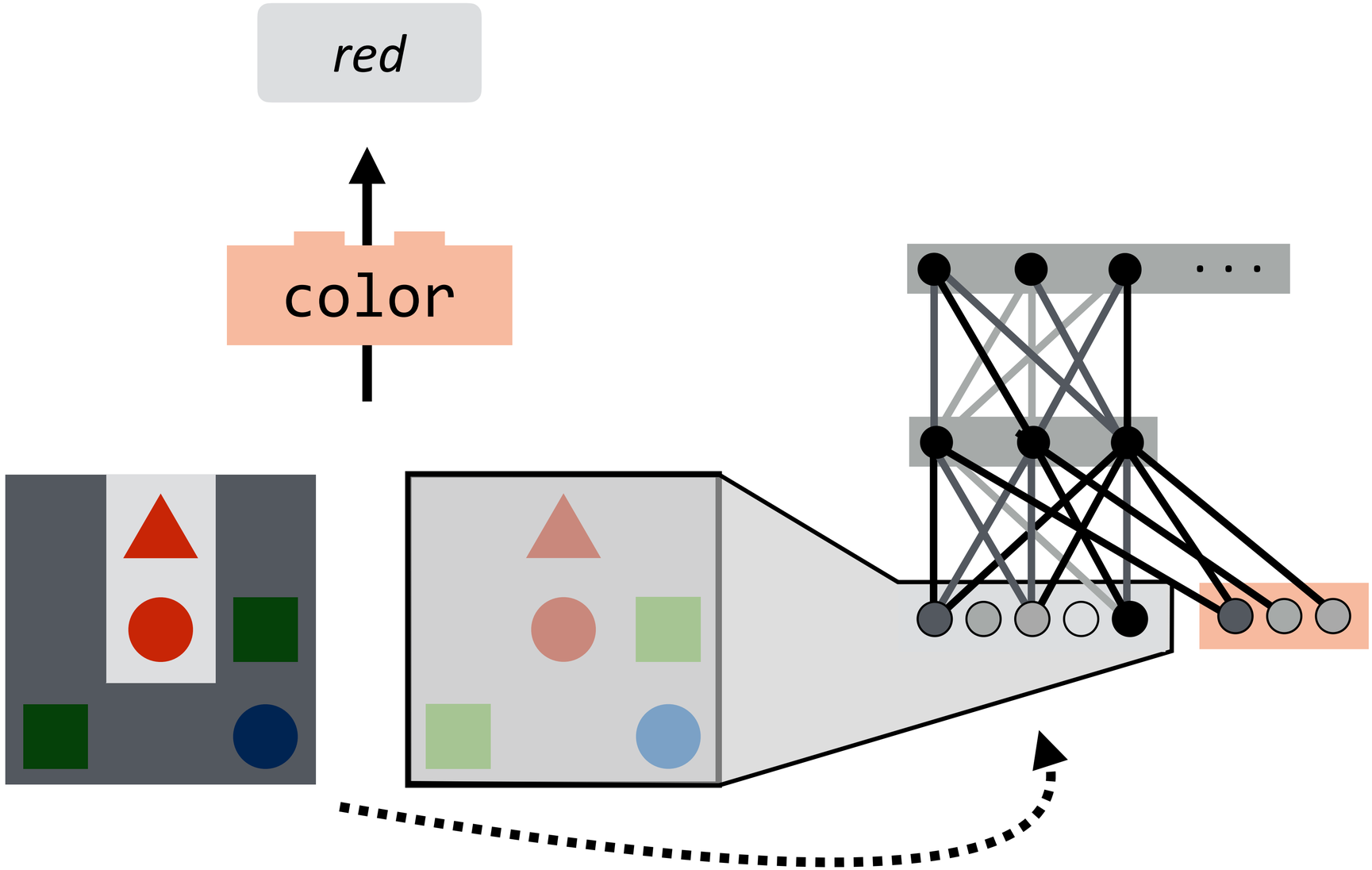
necktie



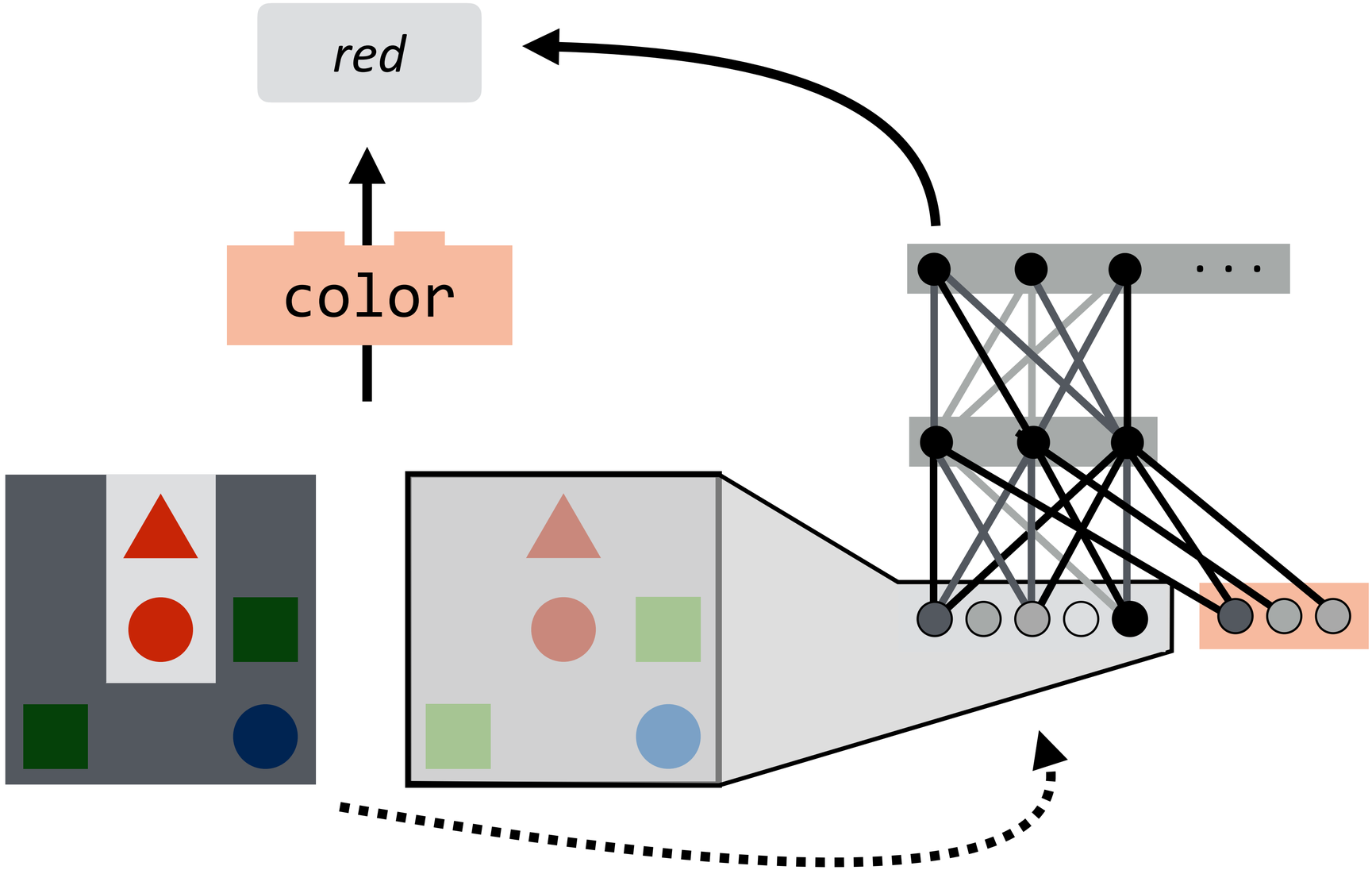
The [describe] module



The [describe] module

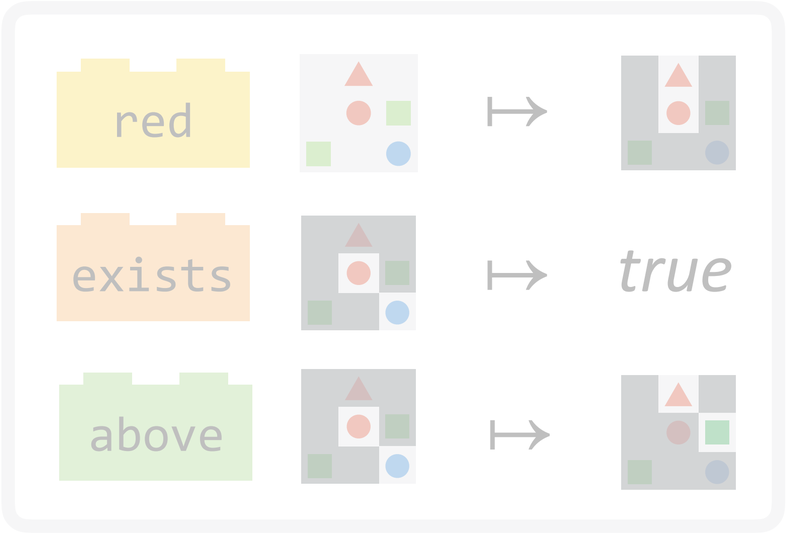


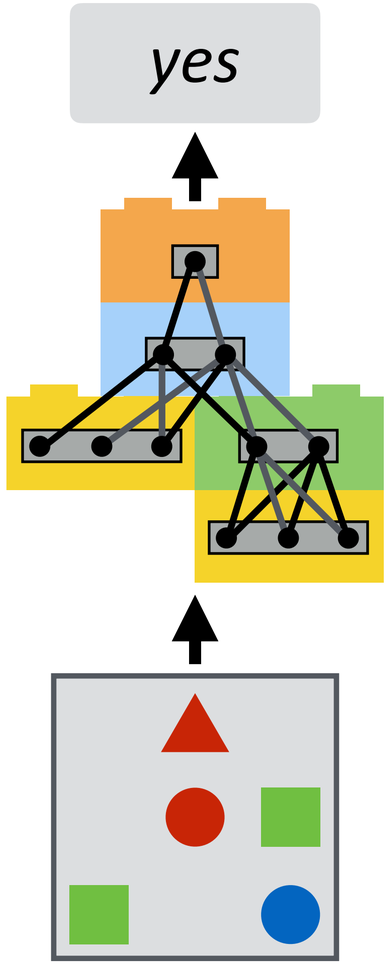
The [describe] module



Outline

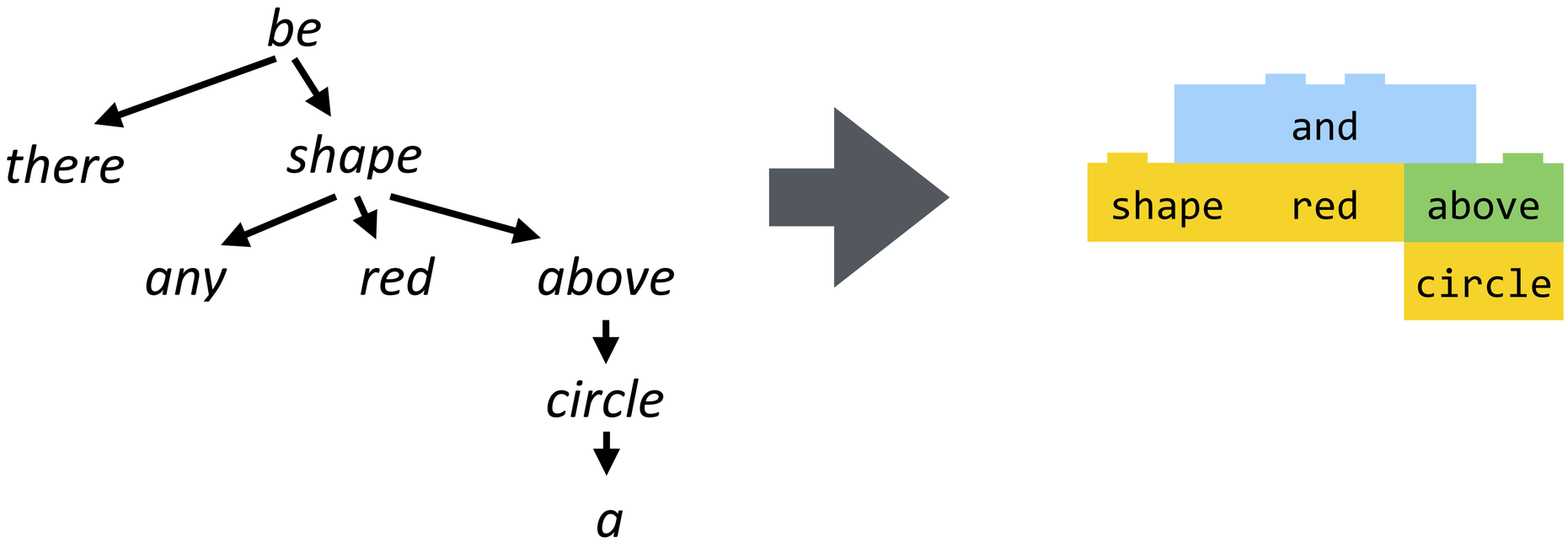
Is there a red shape above a circle?





Where do layouts come from?

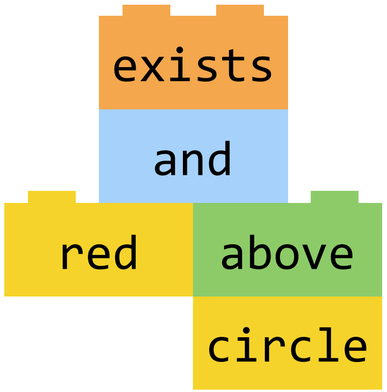
Is there a red shape above a circle?

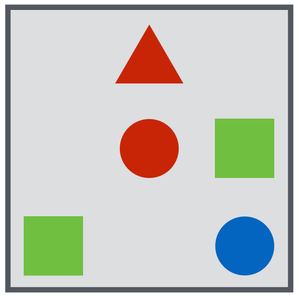


[Reddy et al. 2016]

Learning

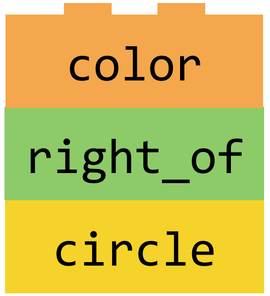
yes

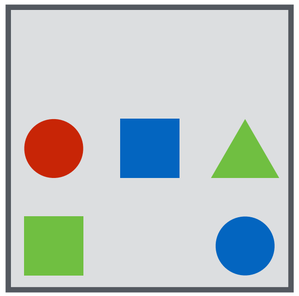




Is there a red shape above a circle?

blue

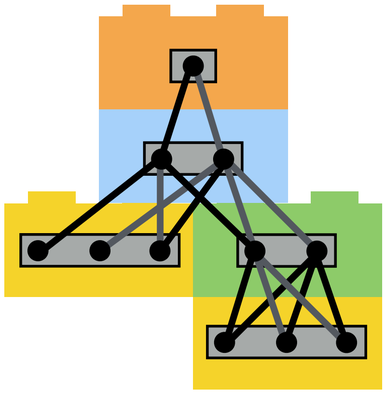


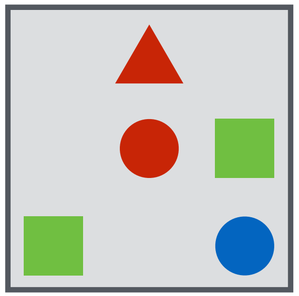


What color is the shape right of a circle?

Learning

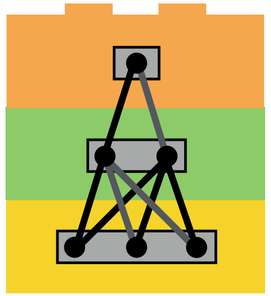
yes

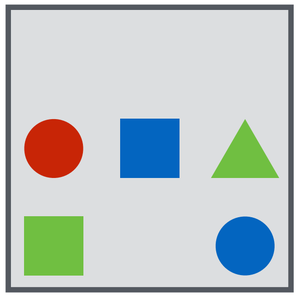




Is there a red shape above a circle?

blue

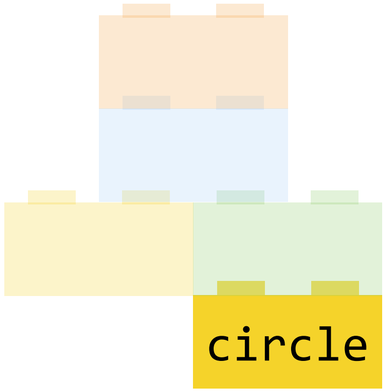


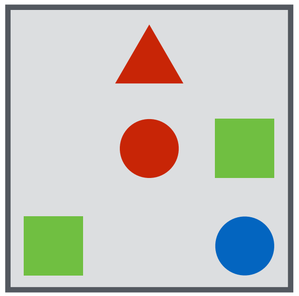


What color is the shape right of a circle?

Parameter tying

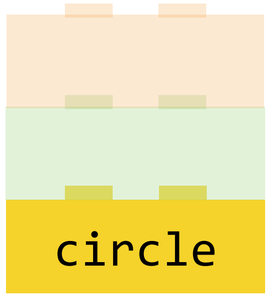
yes

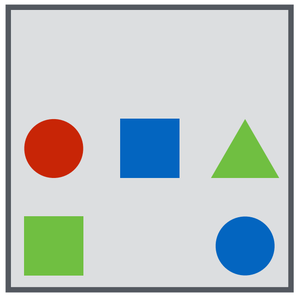




Is there a red shape above a circle?

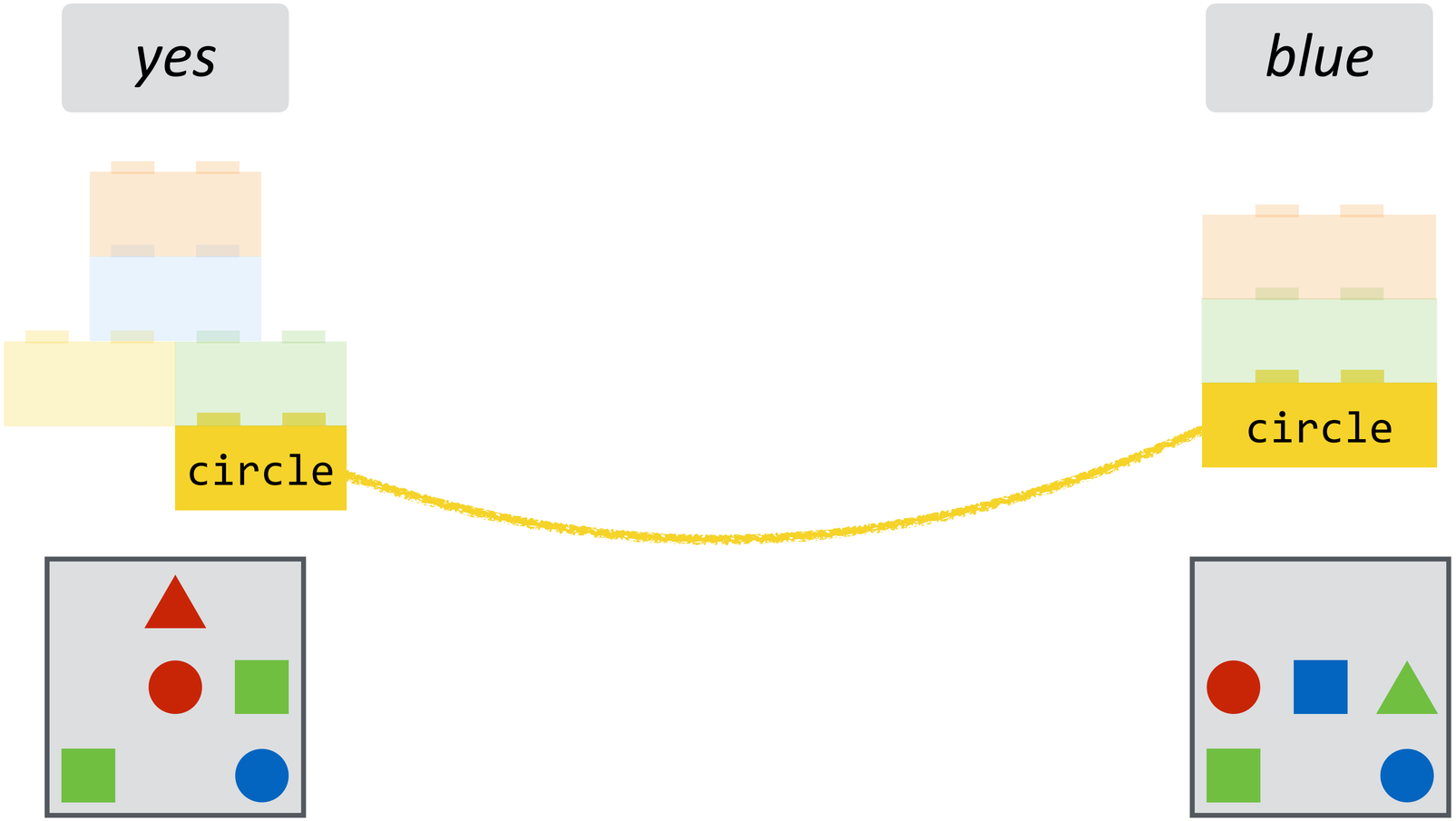
blue





What color is the shape right of a circle?

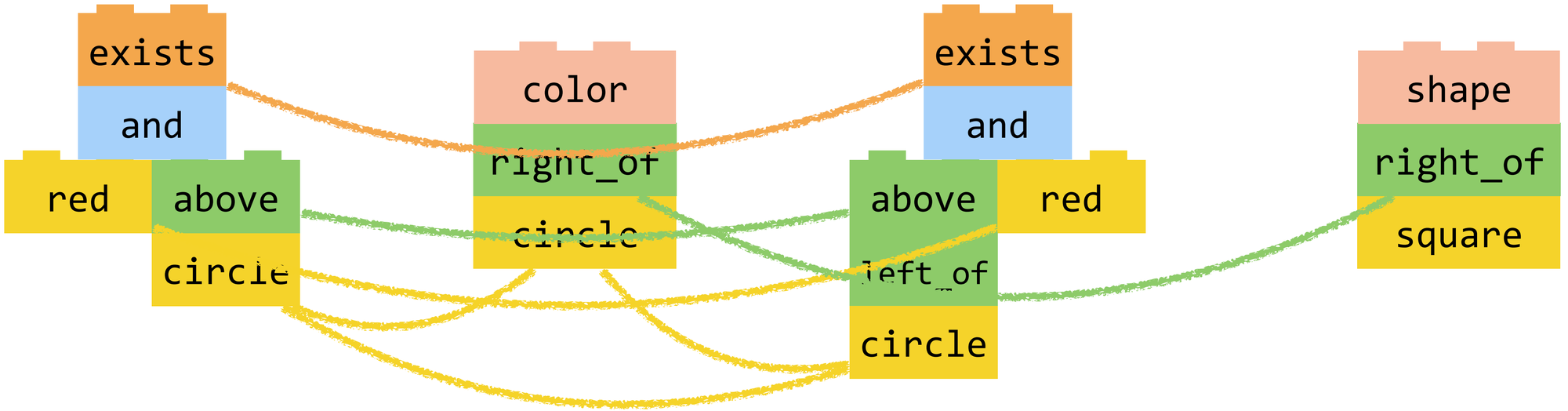
Parameter tying



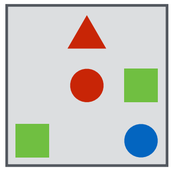
Is there a red shape above a circle?

What color is the shape right of a circle?

Extreme parameter tying

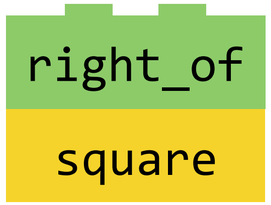


Learning with fixed layouts is easy!

$$\arg\max_{W} \sum \mathrm{p}(\ \textit{yes}\ \mid \ \text{  }\ ,\ \text{  }\ ; W)$$

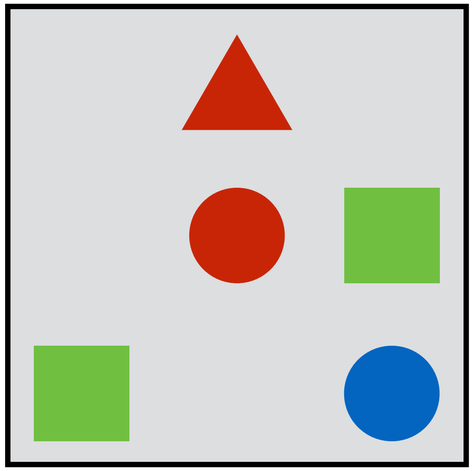
(where every root module outputs a distribution over answers and W is the set of all module parameters)

Learning module behaviors



Module specialization is driven entirely by context!

Experiments



Experiments: VQA dataset

What color is the necktie?

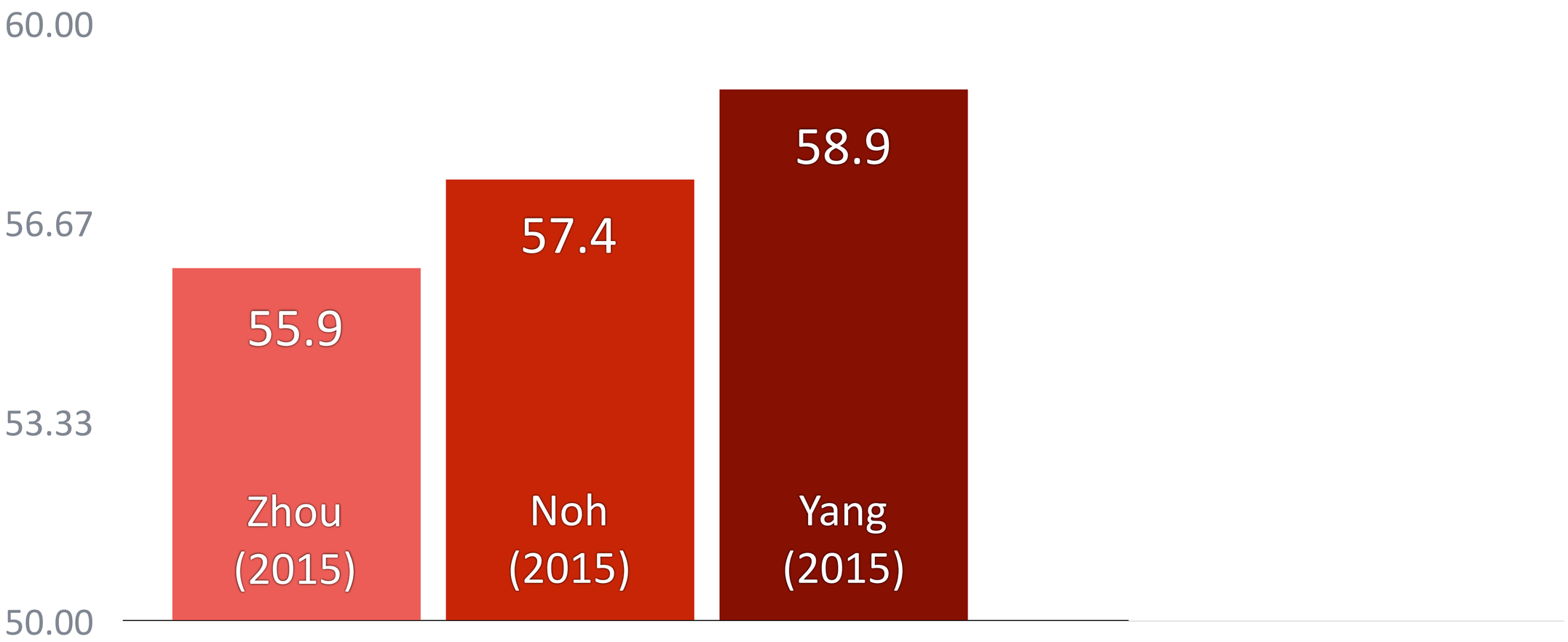
yellow

What is in the sheep's ear?

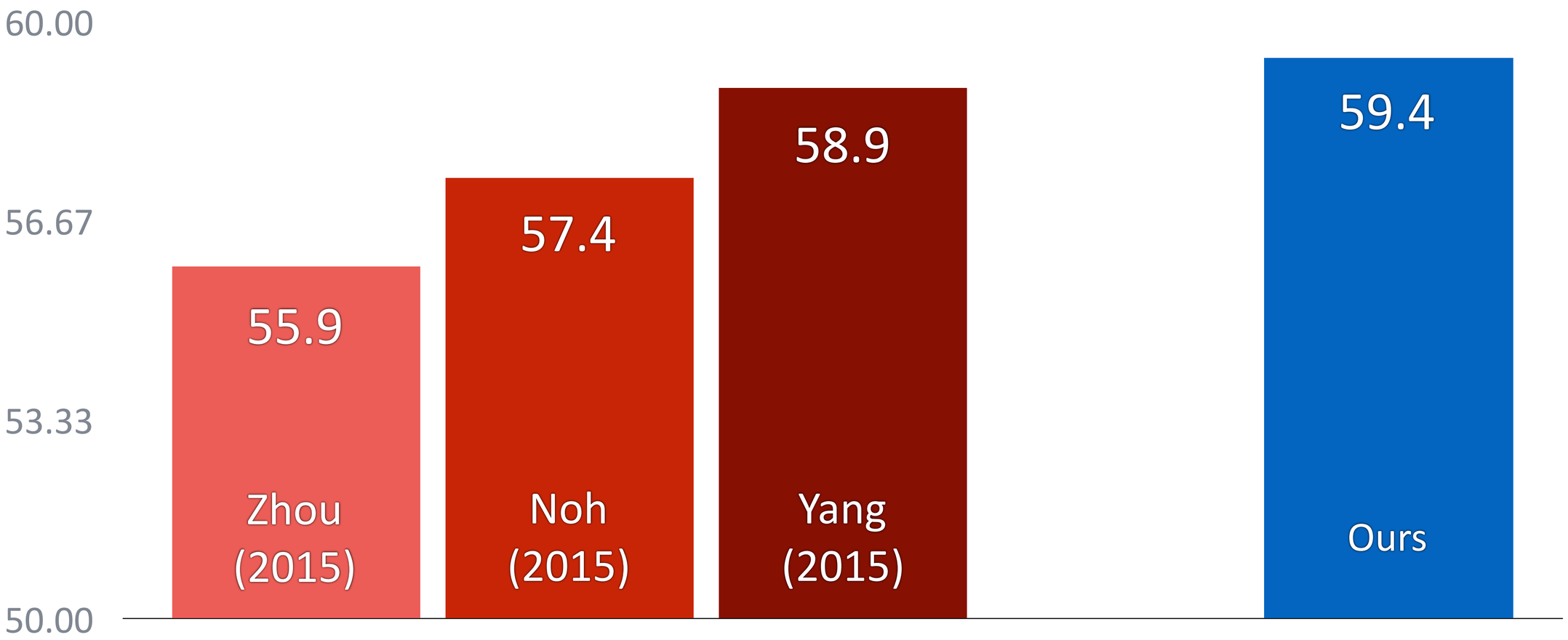
tag

[Antol et al. 2015]

Experiments: VQA dataset



Experiments: VQA dataset



Experiments: VQA dataset

What color is she wearing?

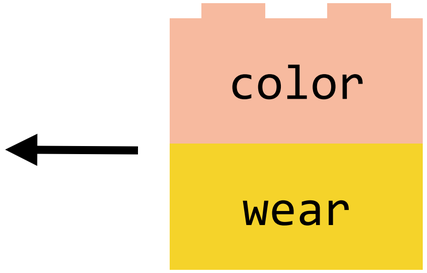
color

wear

white

Experiments: VQA Dataset

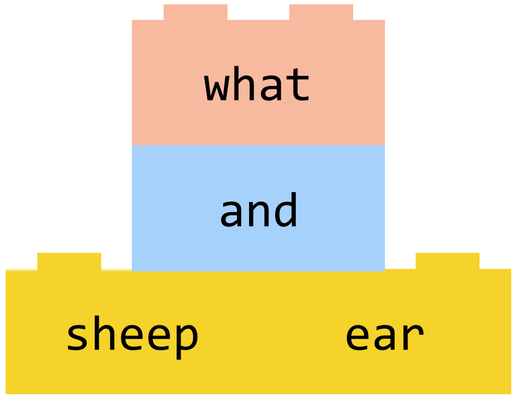
What color is she wearing?



white

Experiments: VQA Dataset

What is in the sheep's ear?



tag

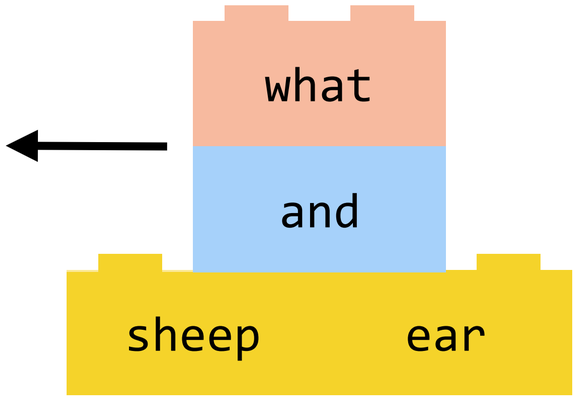
Experiments: VQA Dataset

What
sheep

tag

Experiments: VQA Dataset

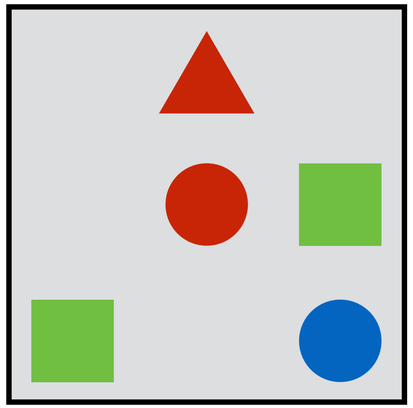
What is in the sheep's ear?



tag

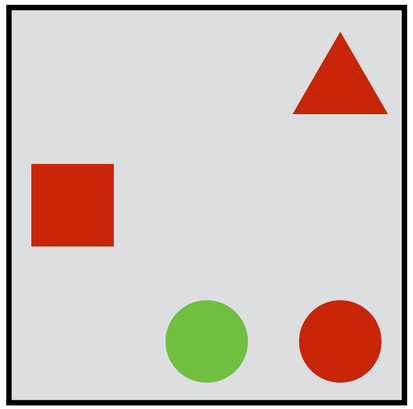
Experiments: SHAPES dataset

Is there a red shape above a circle?



yes

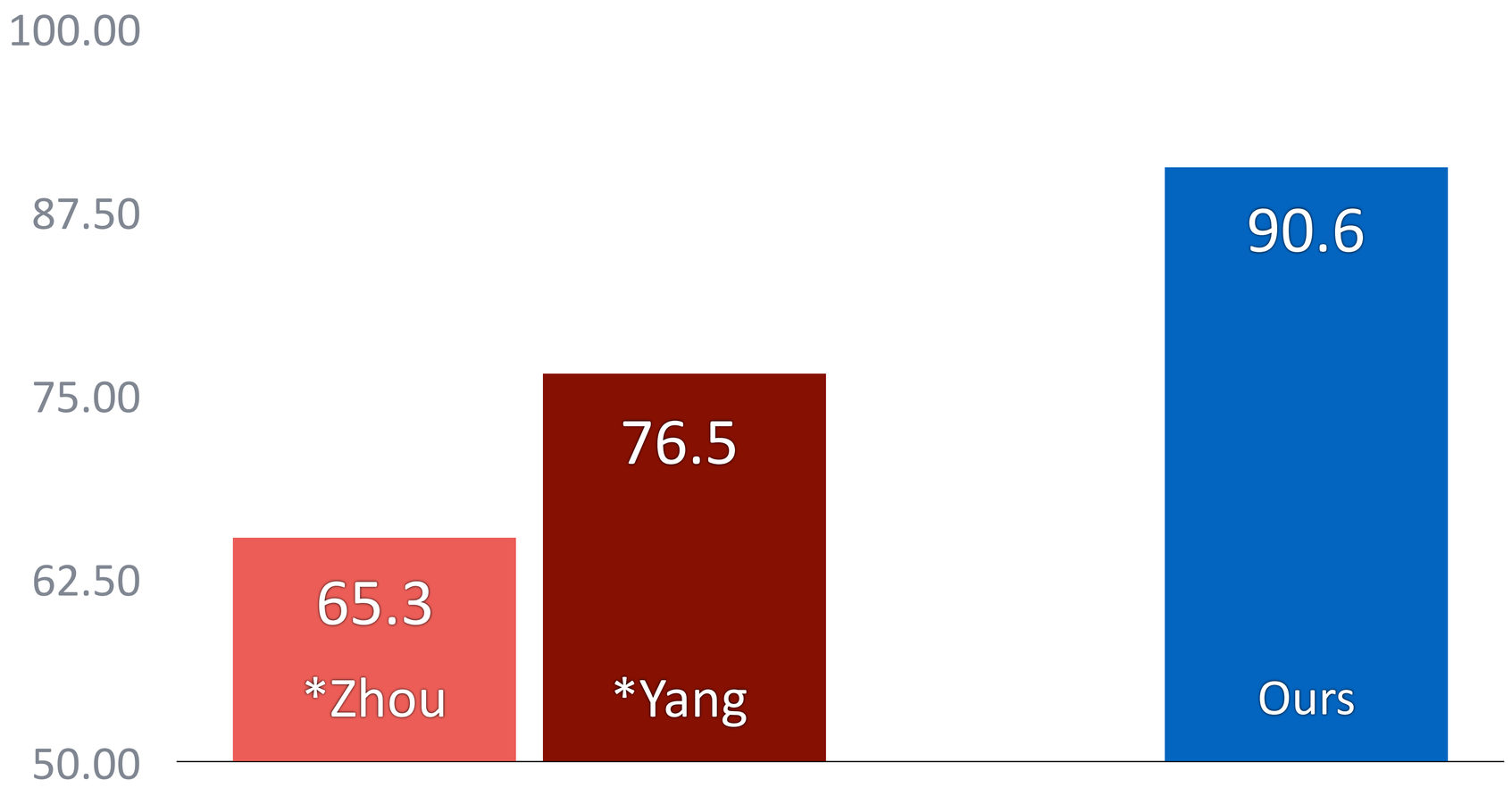
Is a green shape above left of a red shape?



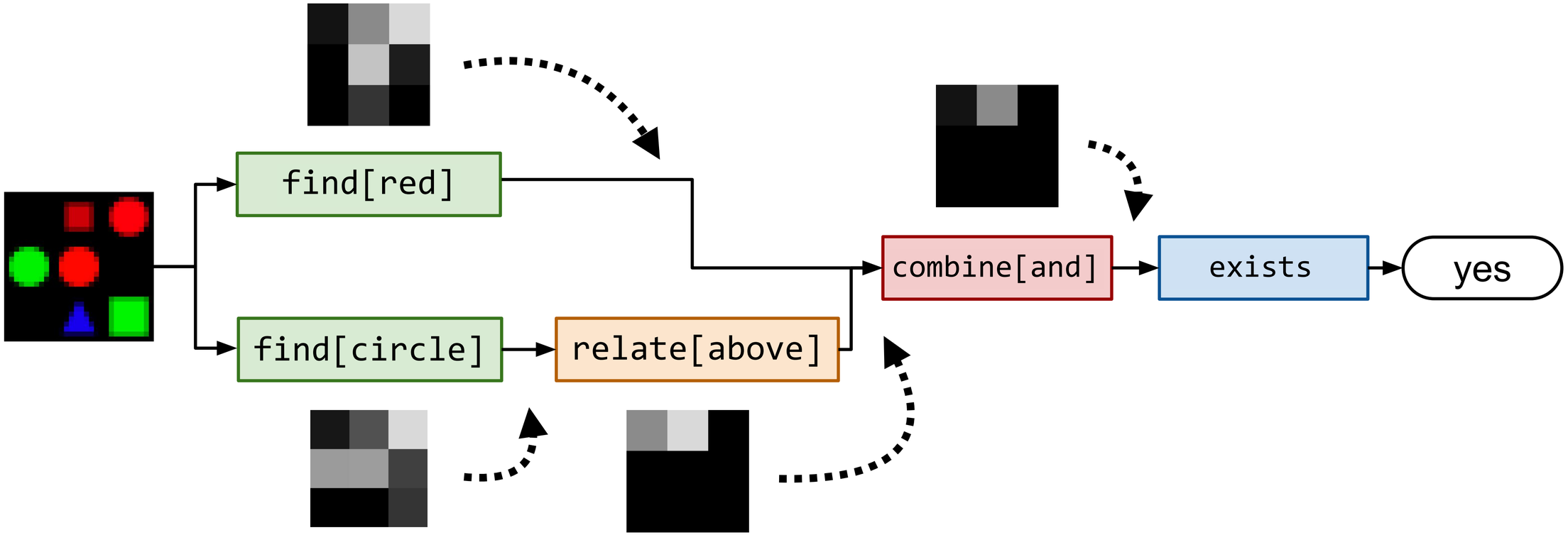
no

c.f. [Xu and Saenko, 2016]

Experiments: SHAPES dataset

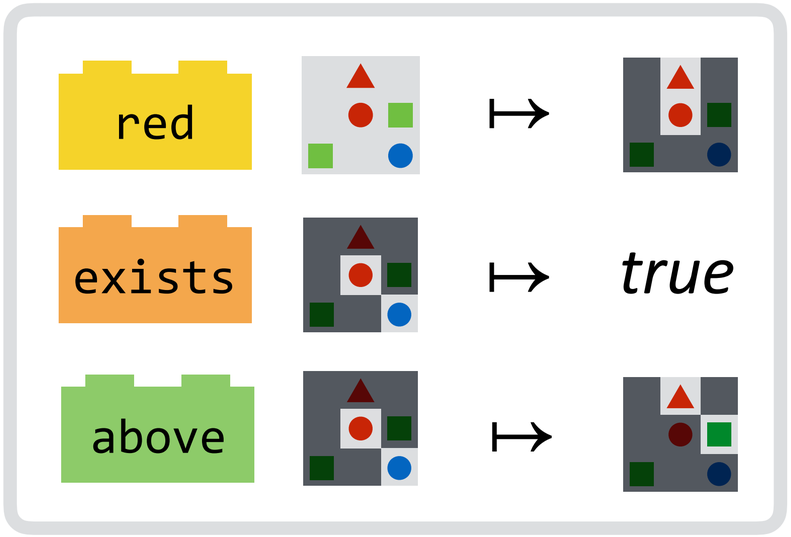


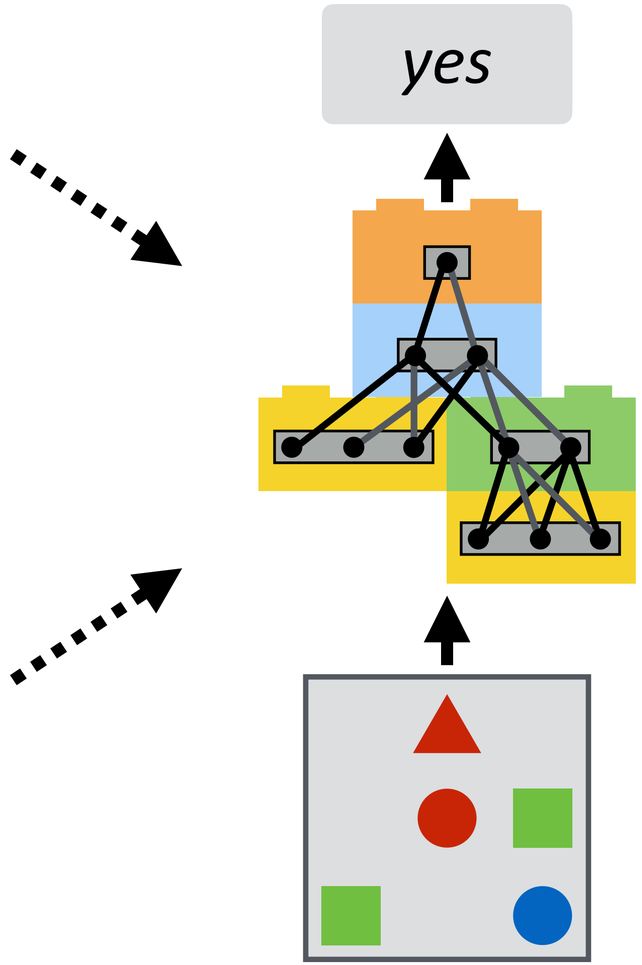
Experiments: SHAPES dataset



Neural module networks

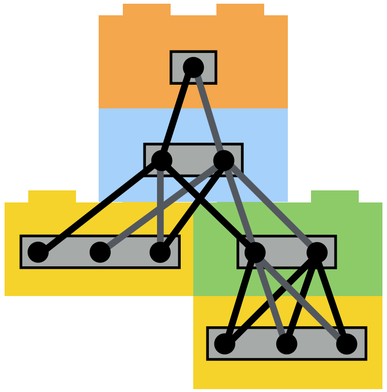
Is there a red shape above a circle?





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