

# PTR: A Benchmark for ParT-based Conceptual, Relational, and Physical Reasoning

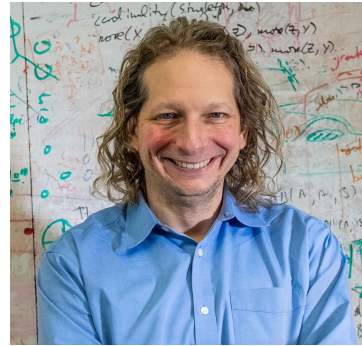
<http://ptr.csail.mit.edu>



Yining Hong



Li Yi



Josh Tenenbaum



Antonio Torralba



Chuang Gan



MIT-IBM  
Watson  
AI Lab

# Visual Question Answering Datasets



**Q:** *What is the mustache made of?*

**A:** *Banana*

VQA [Antol et al. ICCV 2015]

# Visual Question Answering Datasets



**Q:** *What is the mustache made of?*

**A:** *Banana*

VQA [Antol et al. ICCV 2015]

- Natural image datasets
  - Heavy human annotation

# Visual Question Answering Datasets



**Q:** *What is the mustache made of?*

**A:** *Banana*

VQA [Antol et al. ICCV 2015]

- Natural image datasets
  - Heavy human annotation
  - Noisy

# Visual Question Answering Datasets



**Q:** *What is the mustache made of?*

**A:** *Banana*

VQA [Antol et al. ICCV 2015]

- Natural image datasets
  - Heavy human annotation
  - Noisy
  - Biased

# Visual Question Answering Datasets

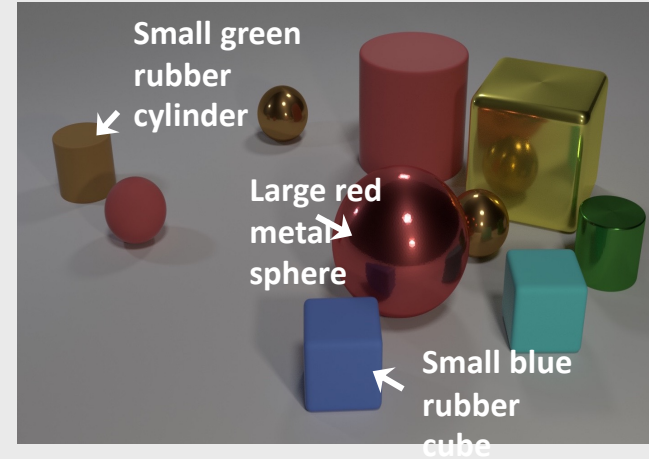


**Q:** *What is the mustache made of?*

**A:** *Banana*

VQA [Antol et al. ICCV 2015]

- Natural image datasets
  - Heavy human annotation
  - Noisy
  - Biased



**Q:** *Are there an equal number of large things and metal spheres?*

**A:** *Yes*

CLEVR [Johnson et al. CVPR 2017]

- Synthetic datasets

# Visual Question Answering Datasets

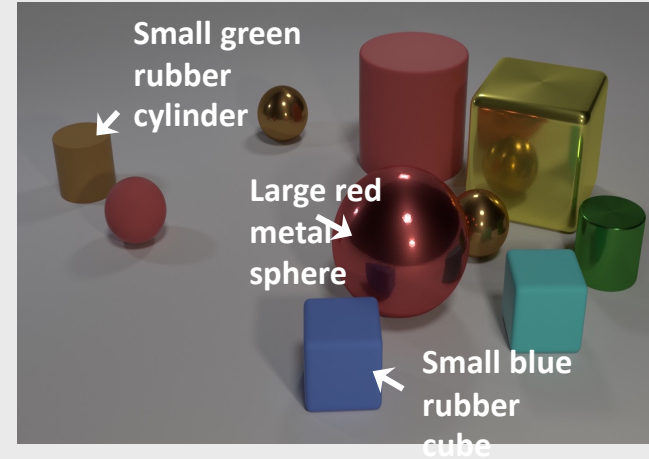


**Q:** *What is the mustache made of?*

**A:** *Banana*

VQA [Antol et al. ICCV 2015]

- Natural image datasets
  - Heavy human annotation
  - Noisy
  - Biased



**Q:** *Are there an equal number of large things and metal spheres?*

**A:** *Yes*

CLEVR [Johnson et al. CVPR 2017]

- Synthetic datasets
  - Fully-controlled

# Visual Question Answering Datasets

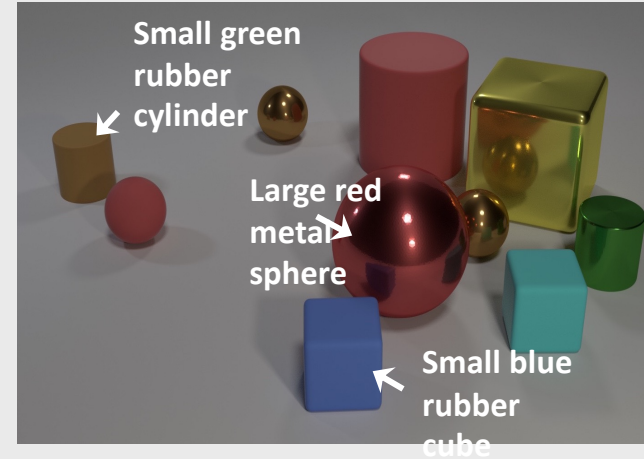


**Q:** *What is the mustache made of?*

**A:** *Banana*

VQA [Antol et al. ICCV 2015]

- Natural image datasets
  - Heavy human annotation
  - Noisy
  - Biased



**Q:** *Are there an equal number of large things and metal spheres?*

**A:** *Yes*

CLEVR [Johnson et al. CVPR 2017]

- Synthetic datasets
  - Fully-controlled
  - Less bias



# Visual Question Answering Datasets

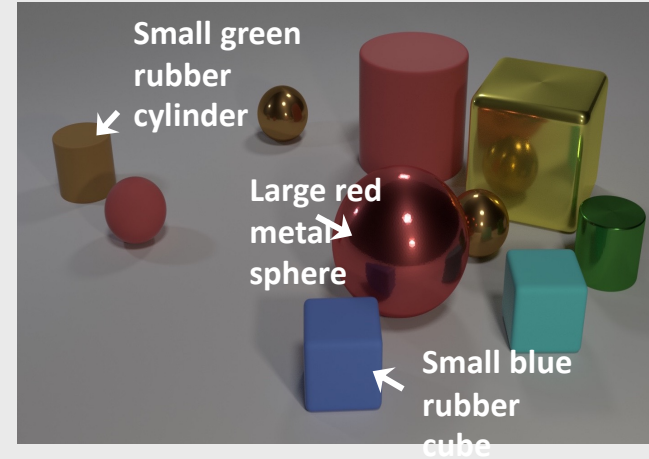


**Q:** *What is the mustache made of?*

**A:** *Banana*

VQA [Antol et al. ICCV 2015]

- Natural image datasets
  - Heavy human annotation
  - Noisy
  - Biased



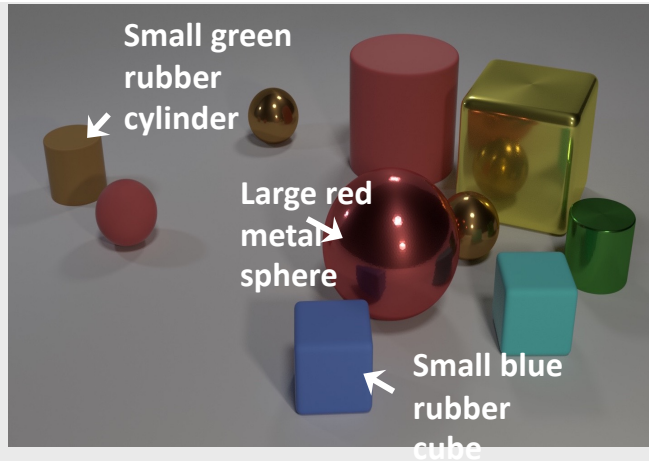
**Q:** *Are there an equal number of large things and metal spheres?*

**A:** *Yes*

CLEVR [Johnson et al. CVPR 2017]

- Synthetic datasets
  - Fully-controlled
  - Less bias
  - Easy to diagnose

# CLEVR - Diagnostic Visual Reasoning Dataset



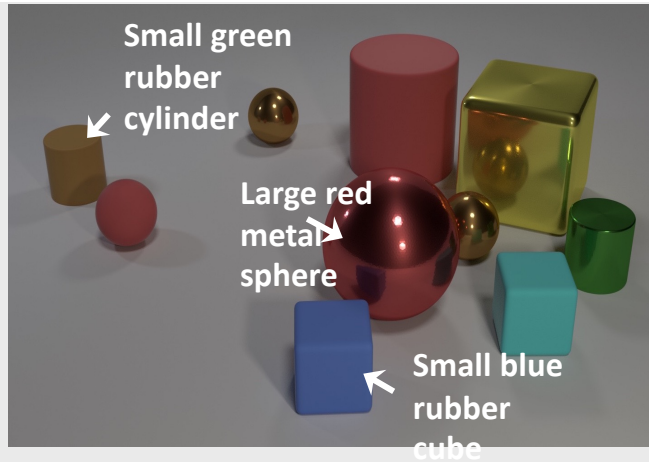
**Q:** *Are there an equal number of large things and metal spheres?*

**A:** *Yes*

CLEVR [Johnson et al. CVPR 2017]

- Fully-controlled
- Less bias
- Disentanglement of the reasoning process, easy to diagnose

# CLEVR - Diagnostic Visual Reasoning Dataset

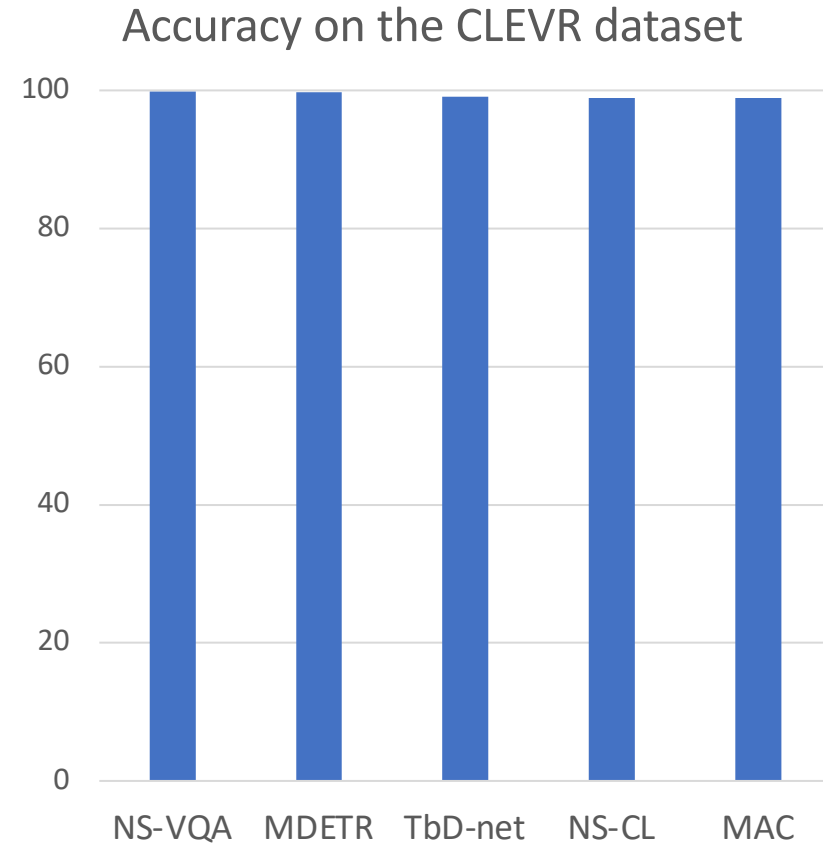


**Q:** *Are there an equal number of large things and metal spheres?*

**A:** *Yes*

CLEVR [Johnson et al. CVPR 2017]

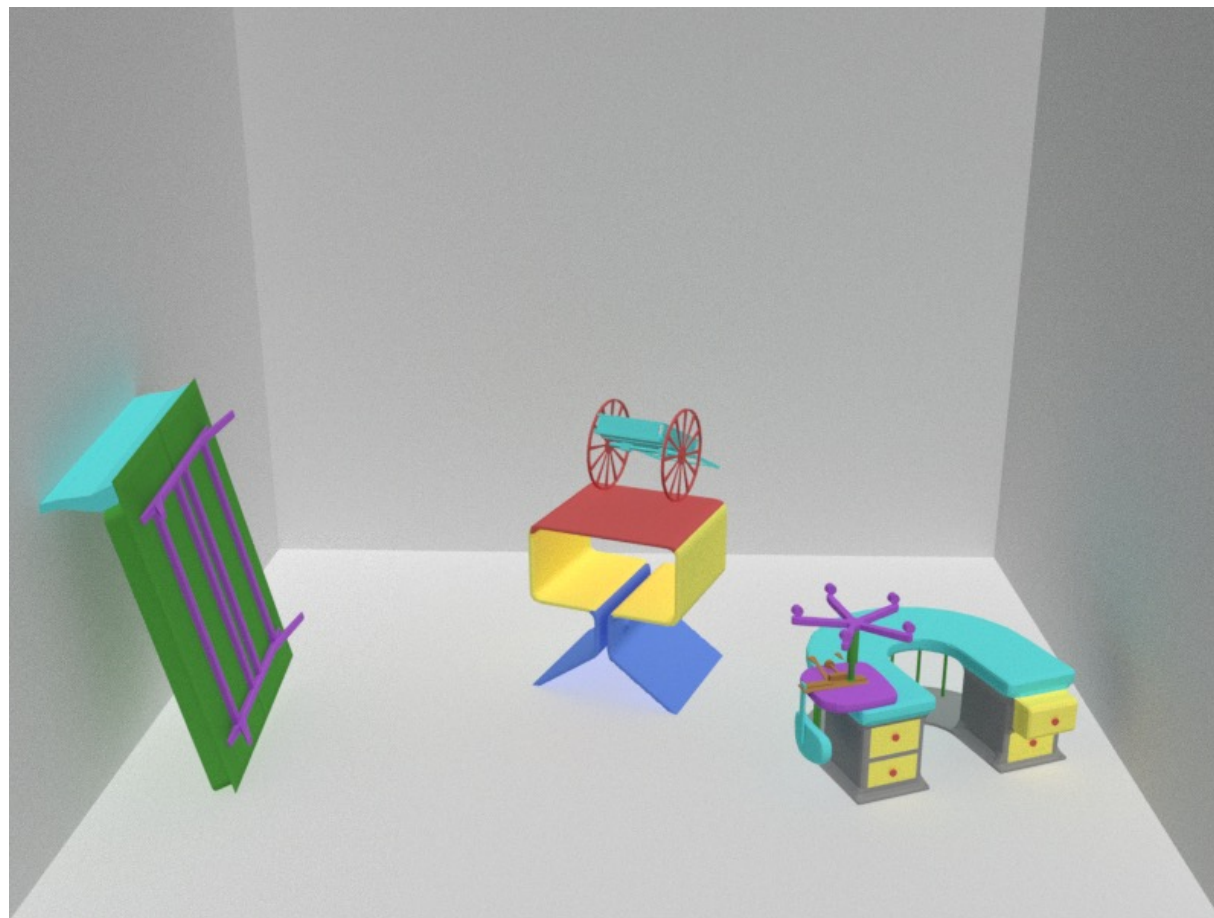
- Fully-controlled
- Less bias
- Disentanglement of the reasoning process, easy to diagnose



What's next for diagnostic visual reasoning?

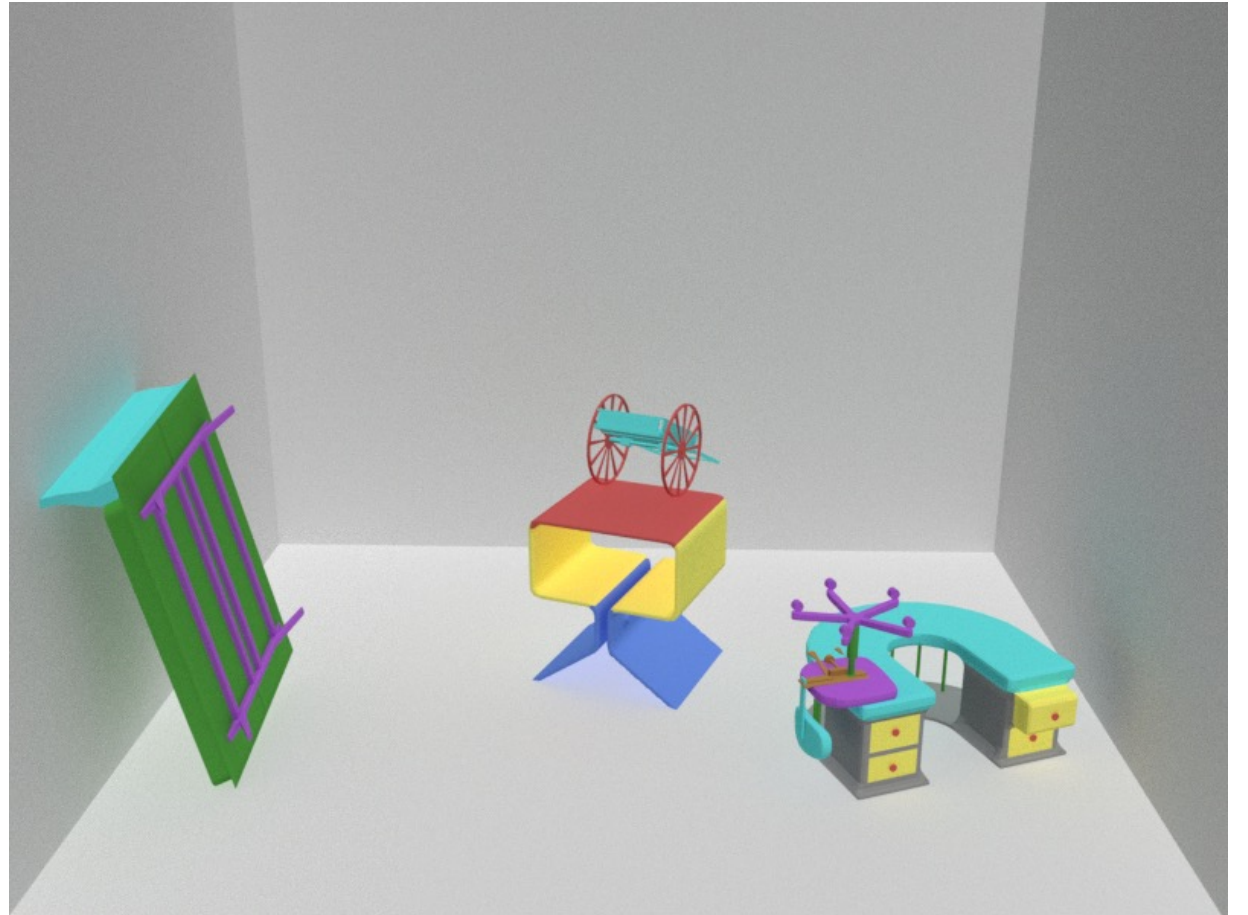
# The PTR dataset

- ~70k images & ~700k questions



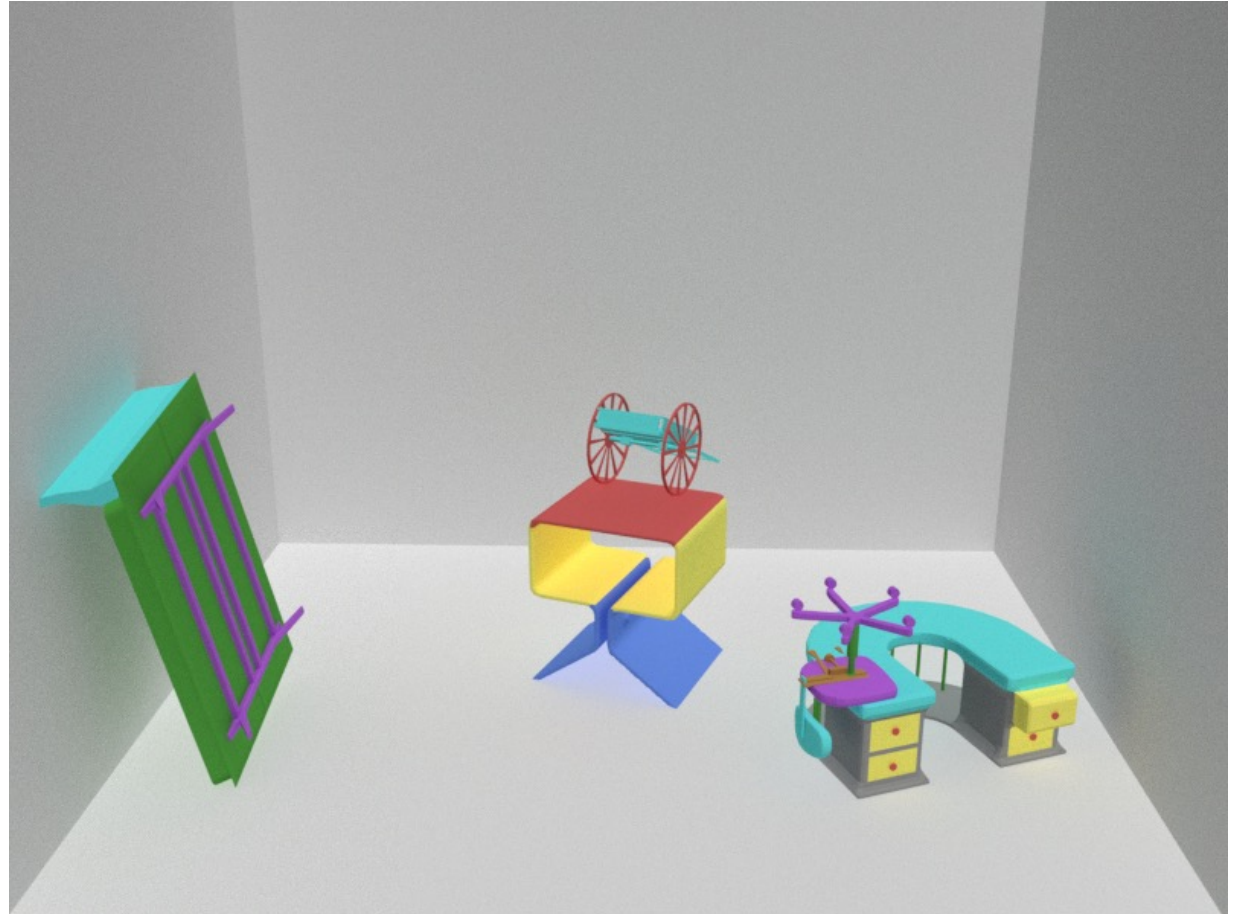
# The PTR dataset

- ~70k images & ~700k questions
- Five Question types:



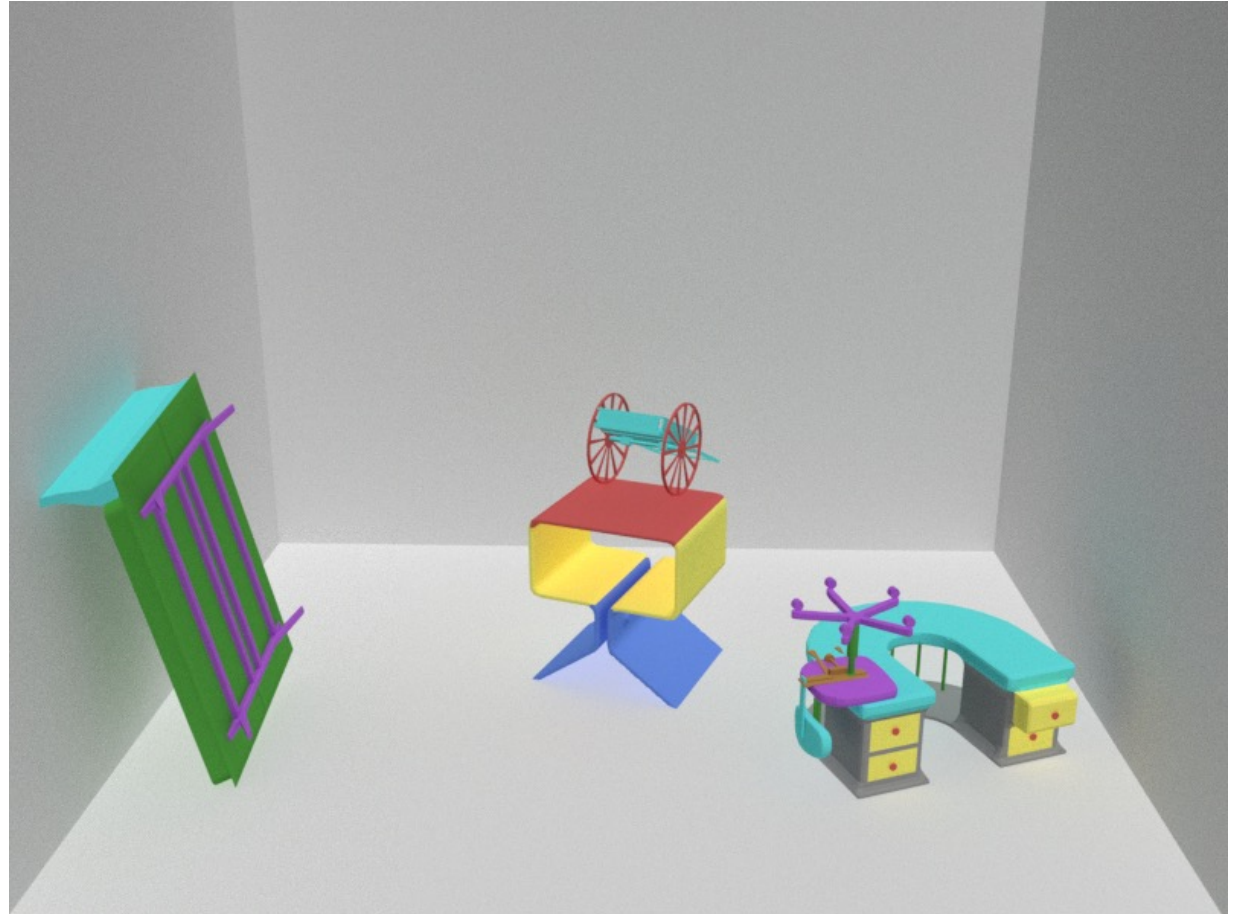
# The PTR dataset

- ~70k images & ~700k questions
- Five Question types:
  - Concept



# The PTR dataset

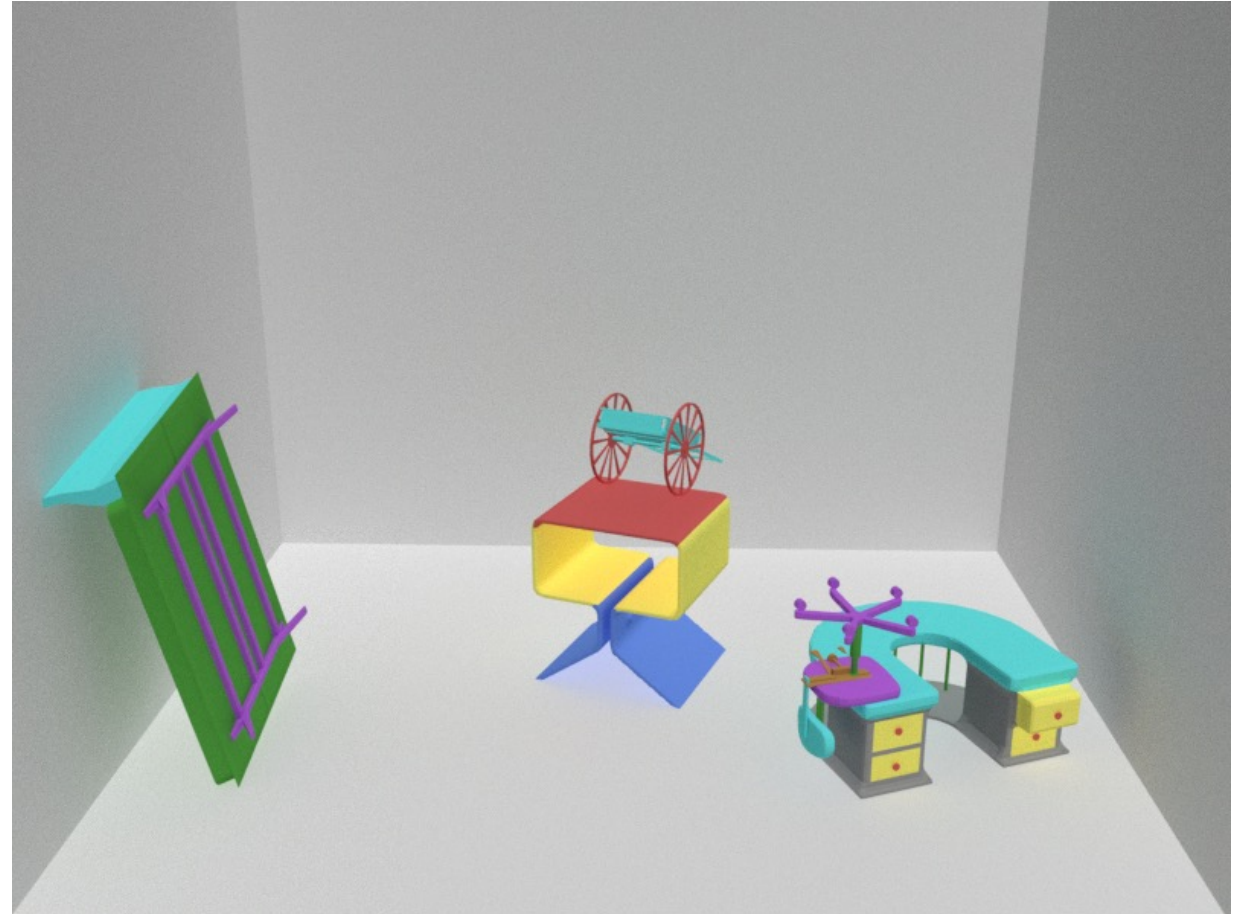
- ~70k images & ~700k questions
- Five Question types:
  - Concept
  - Relation





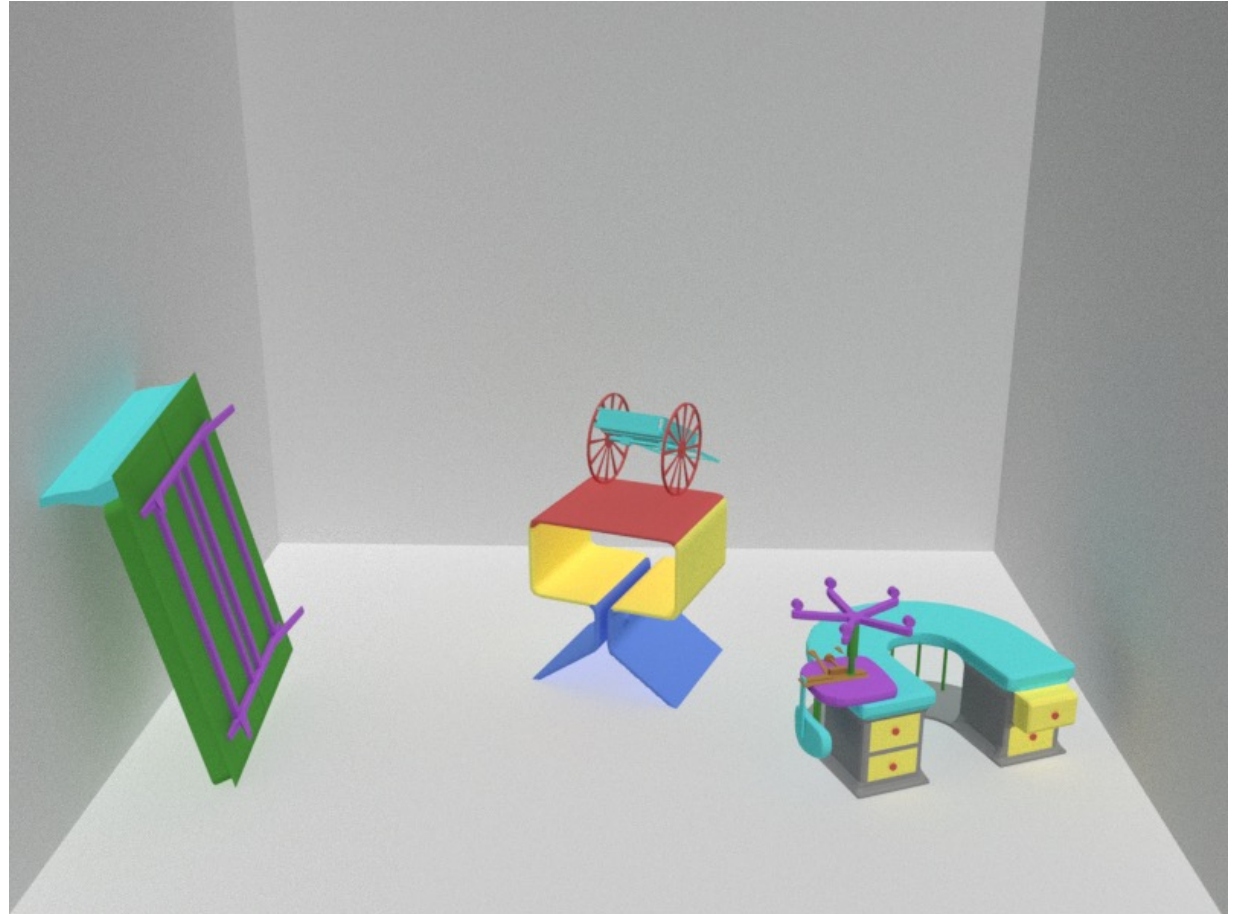
# The PTR dataset

- ~70k images & ~700k questions
- Five Question types:
  - Concept
  - Relation
  - Analogy



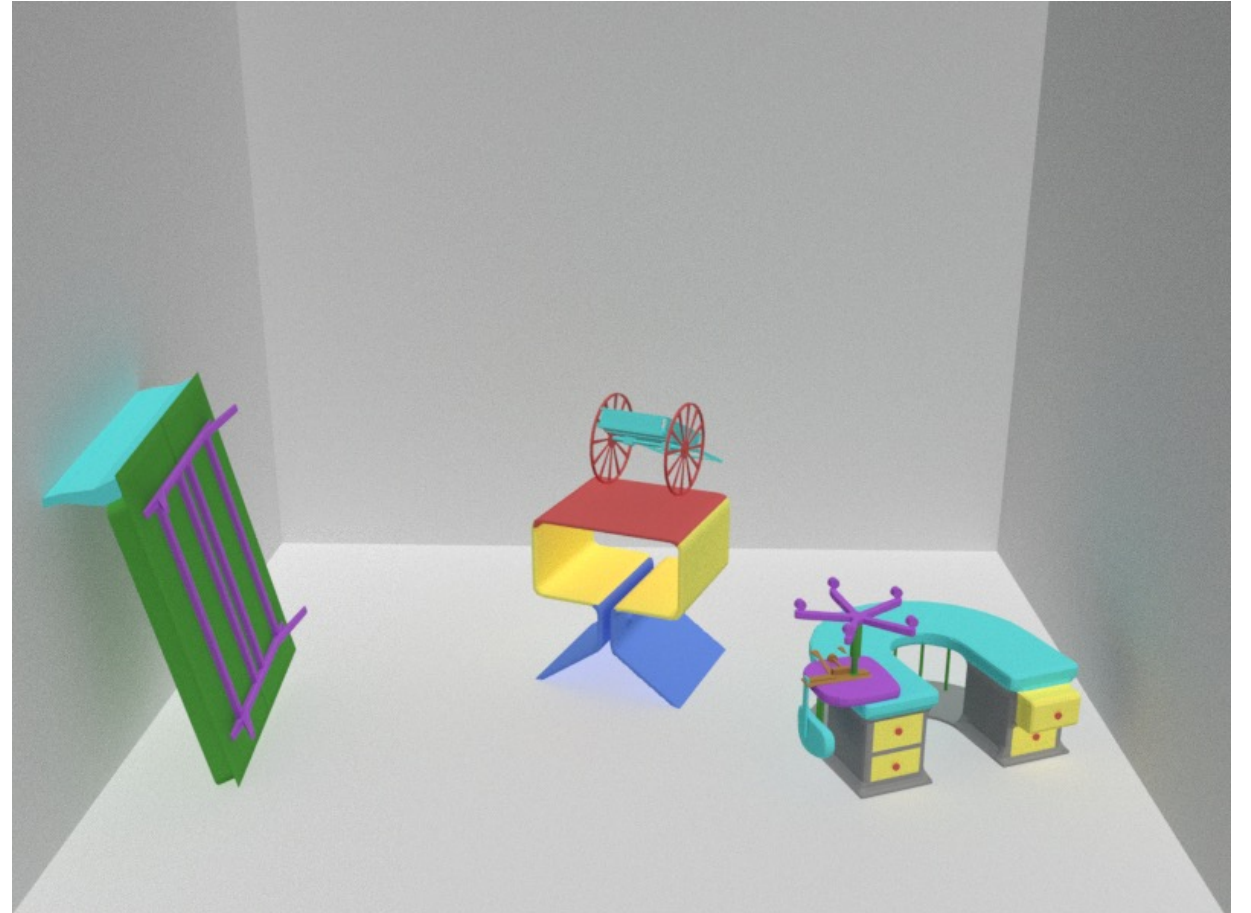
# The PTR dataset

- ~70k images & ~700k questions
- Five Question types:
  - Concept
  - Relation
  - Analogy
  - Arithmetic



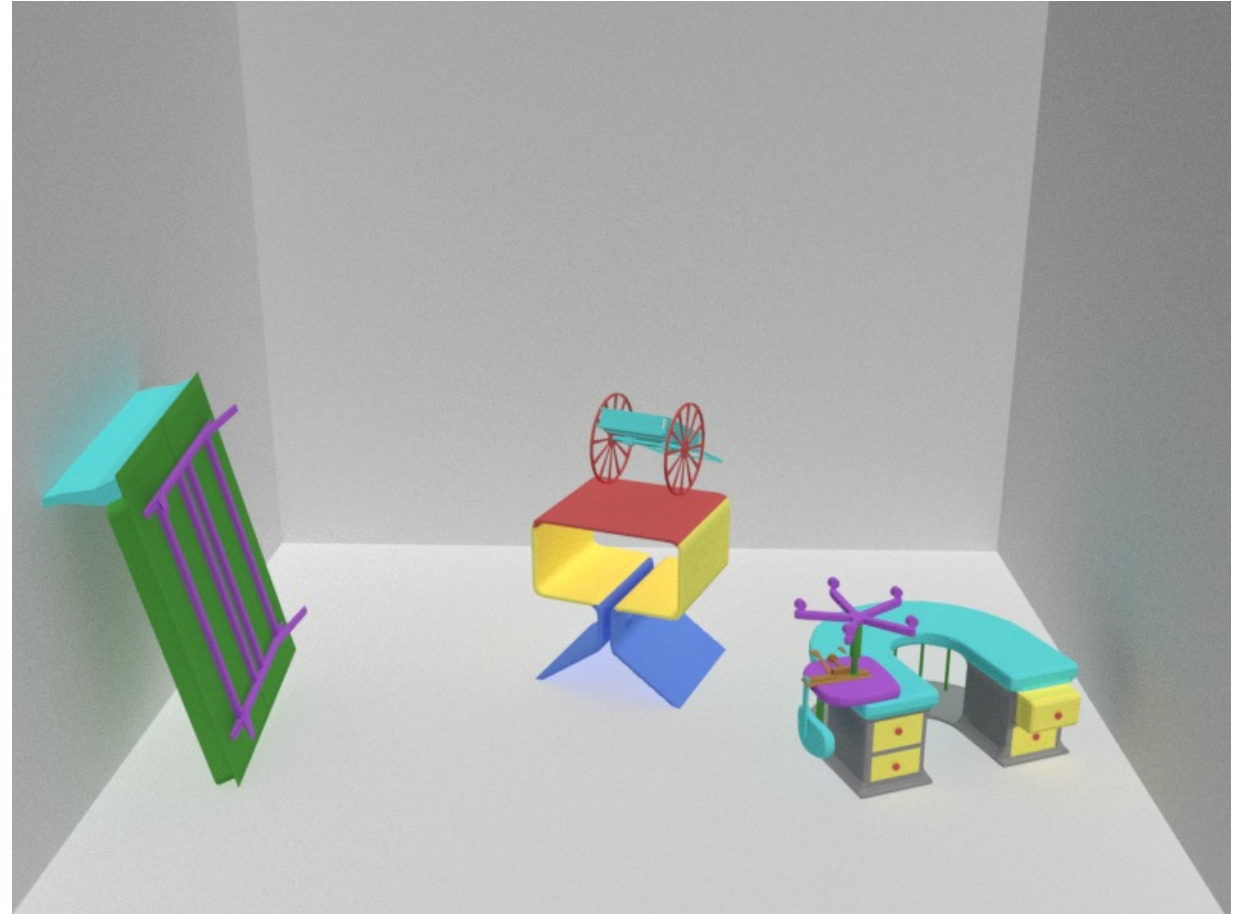
# The PTR dataset

- ~70k images & ~700k questions
- Five Question types:
  - Concept
  - Relation
  - Analogy
  - Arithmetic
  - Physics



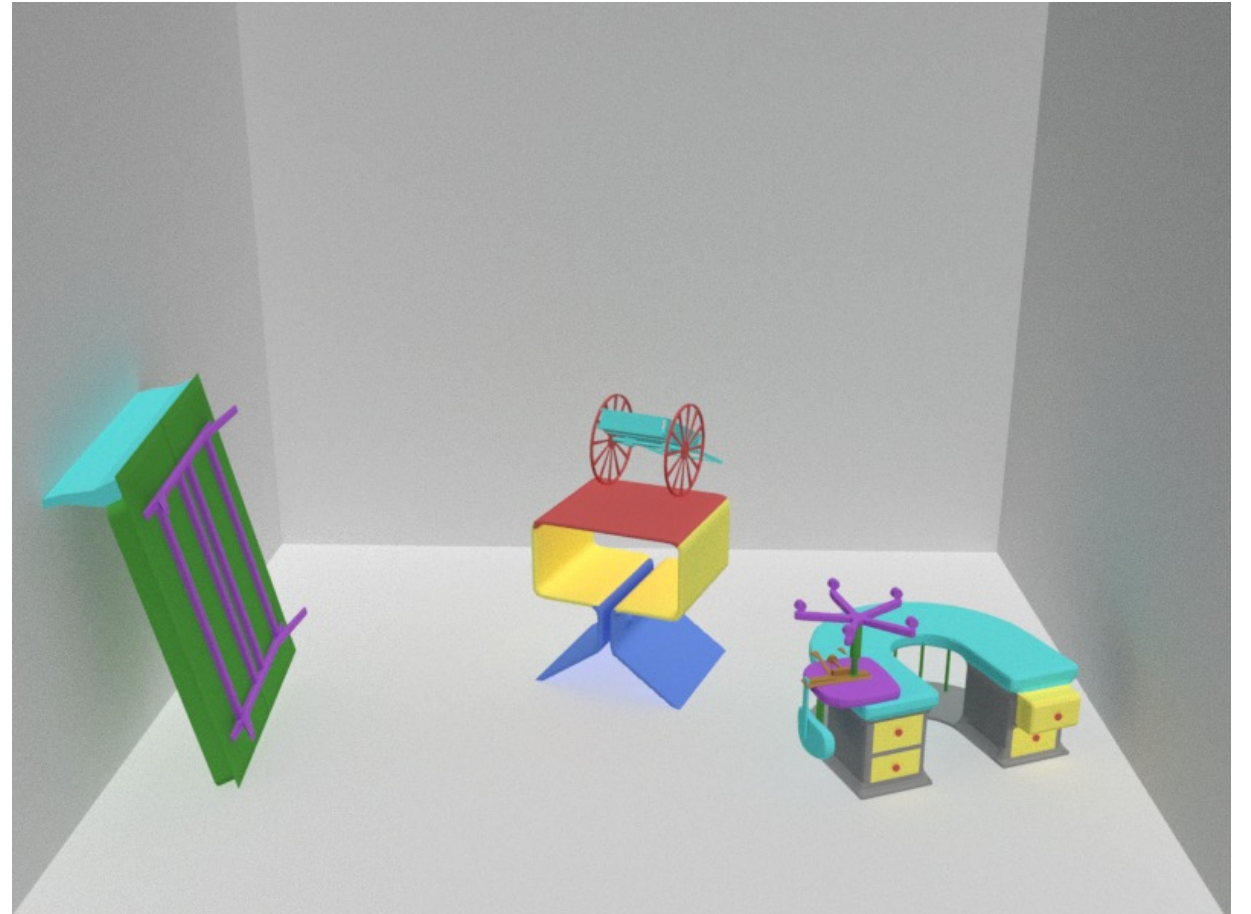
# The PTR dataset

- ~70k images & ~700k questions
- Five Question types:
  - Concept
  - Relation
  - Analogy
  - Arithmetic
  - Physics
- Bias Control

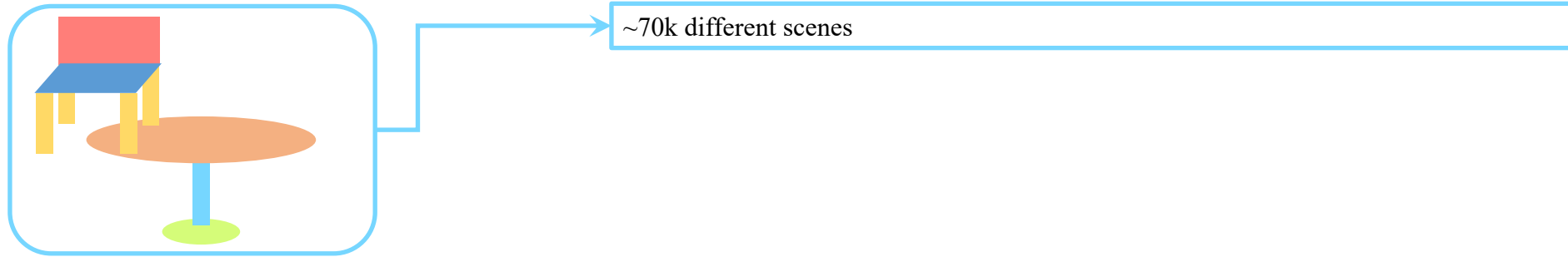


# The PTR dataset

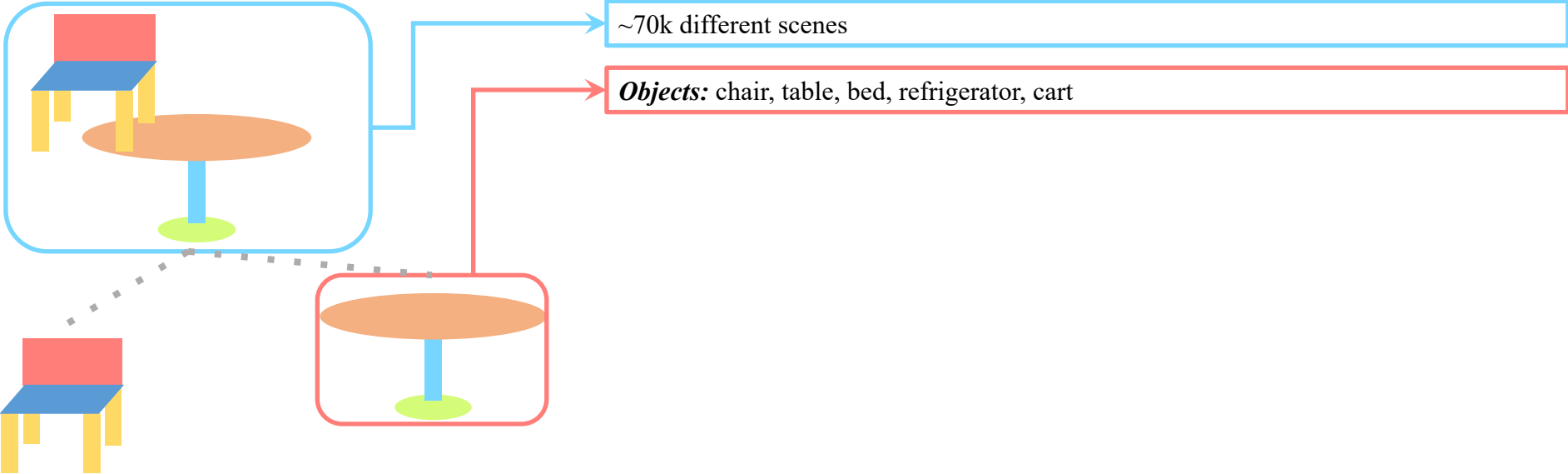
- ~70k images & ~700k questions
- Five Question types:
  - Concept
  - Relation
  - Analogy
  - Arithmetic
  - Physics
- Bias Control
- Diagnostic Annotations



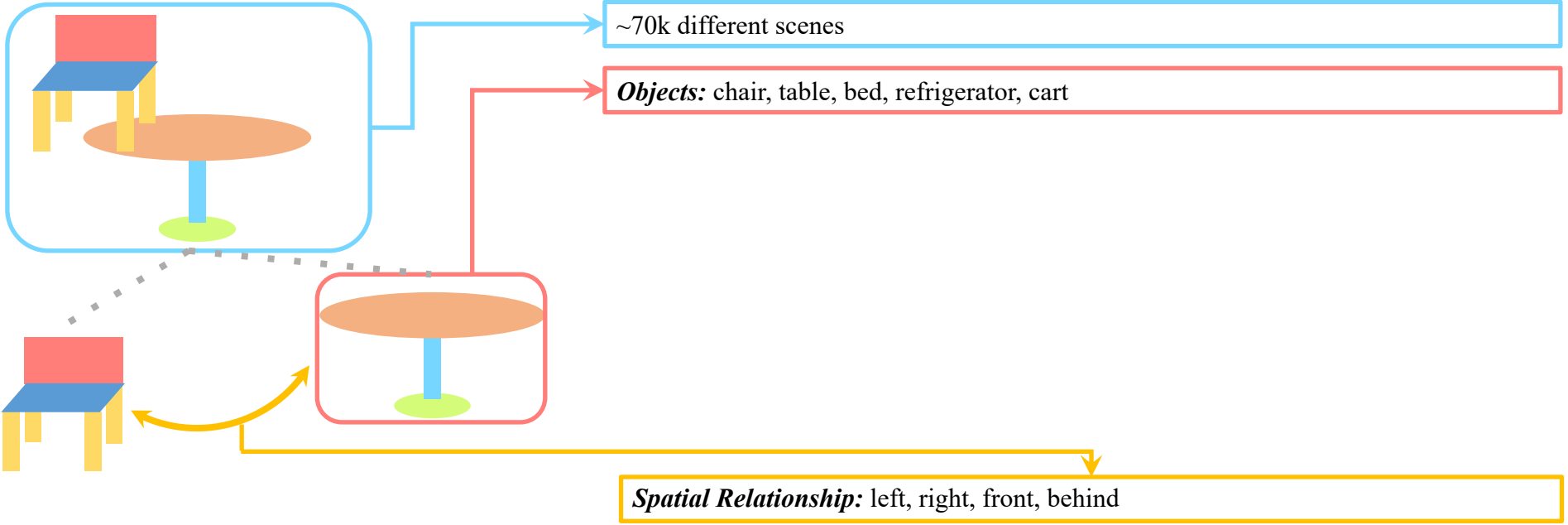
# The PTR Universe



# The PTR Universe

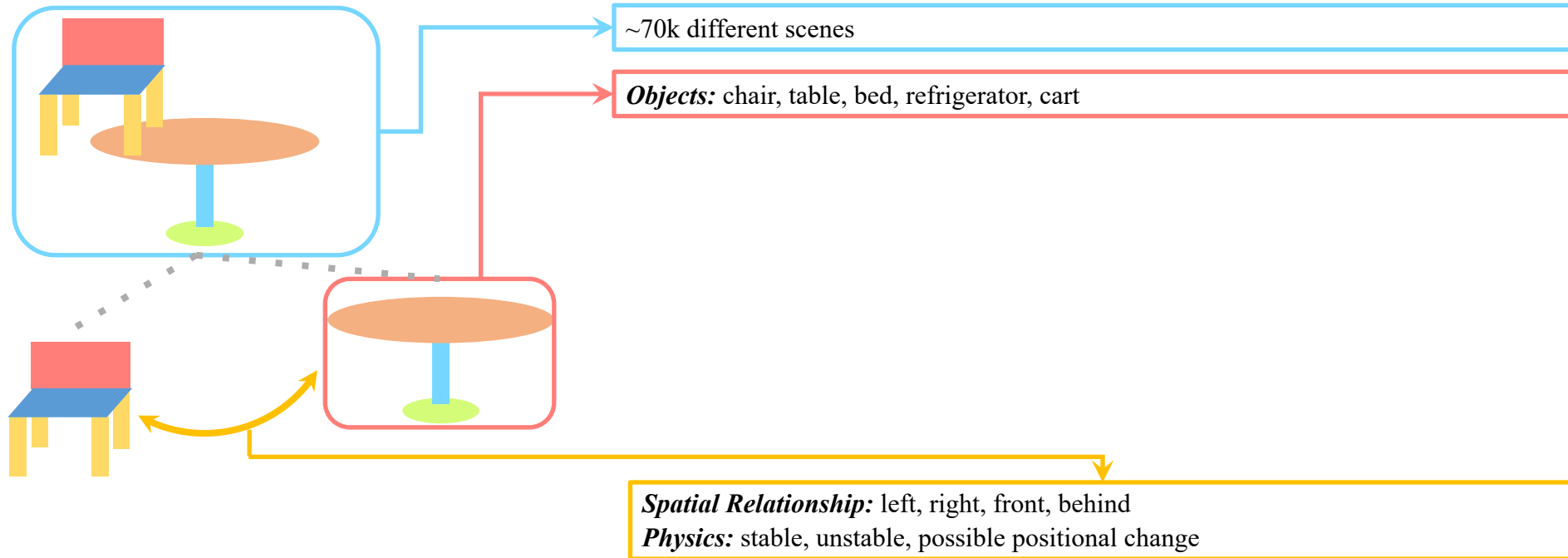


# The PTR Universe

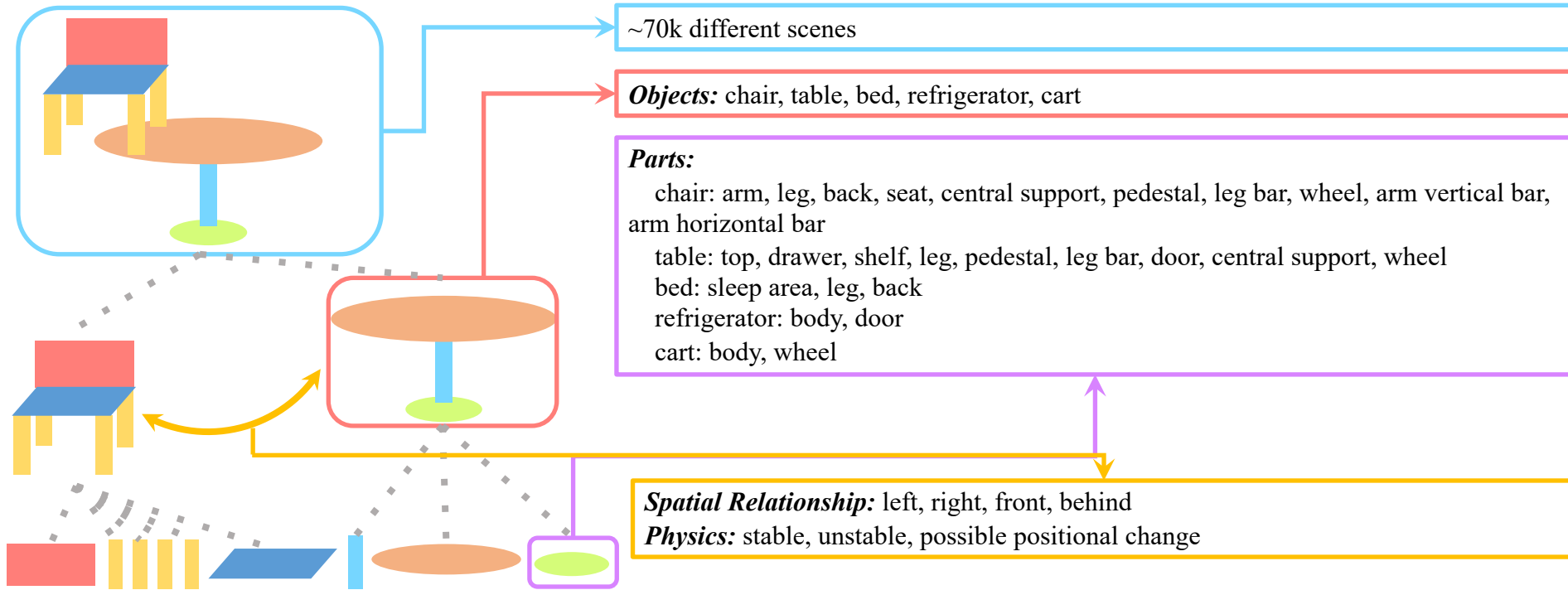




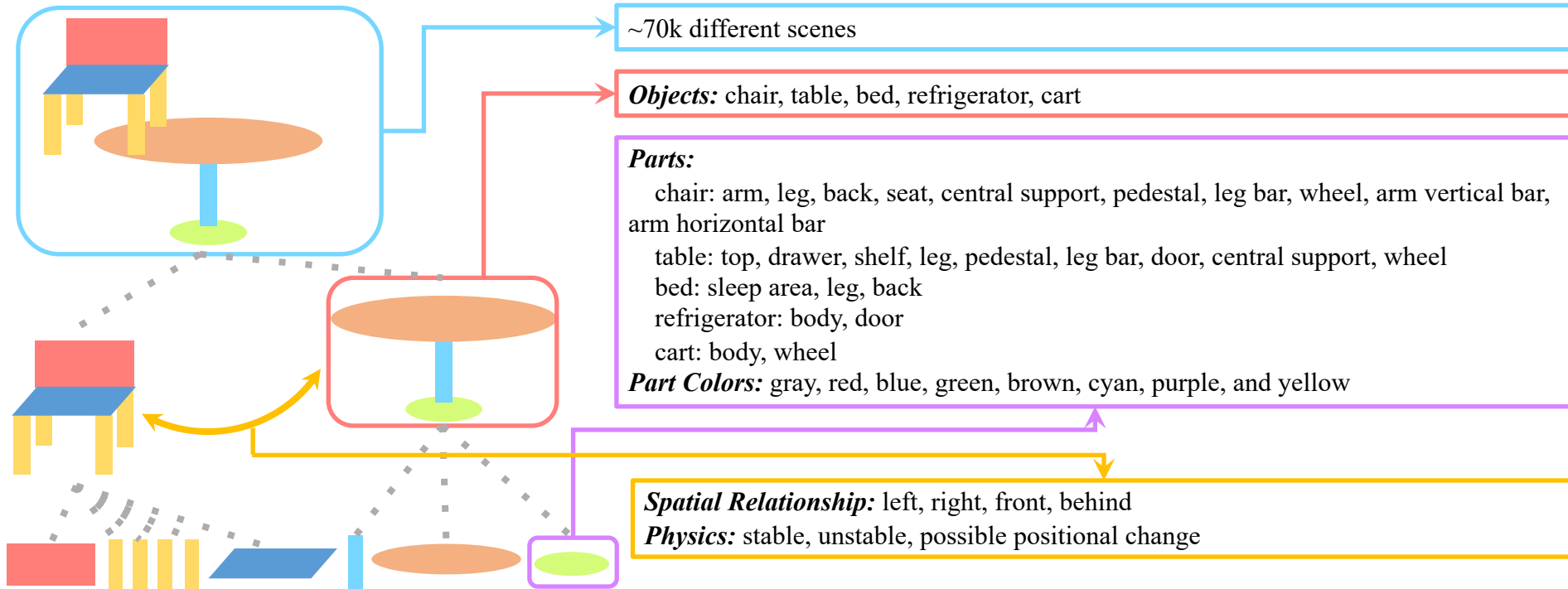
# The PTR Universe



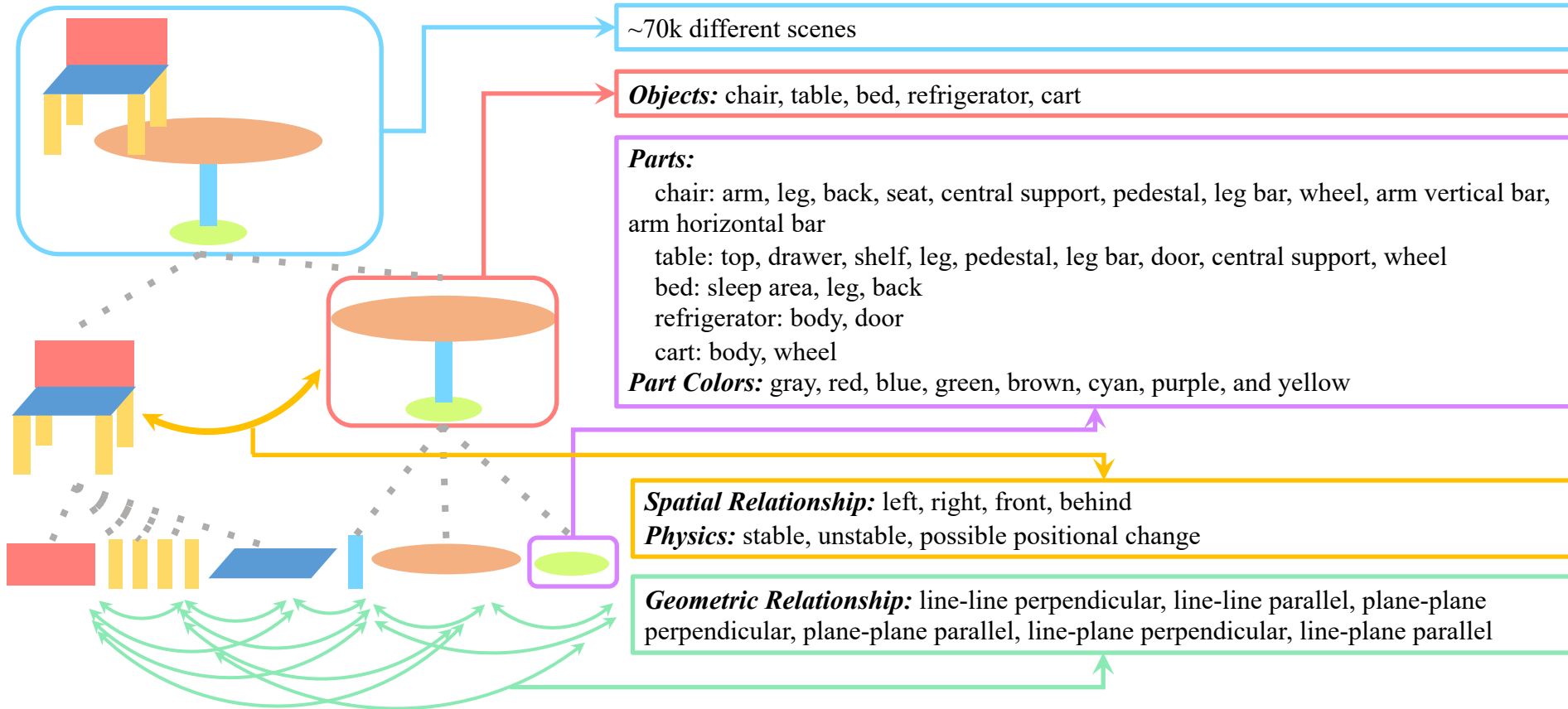
# The PTR Universe



# The PTR Universe

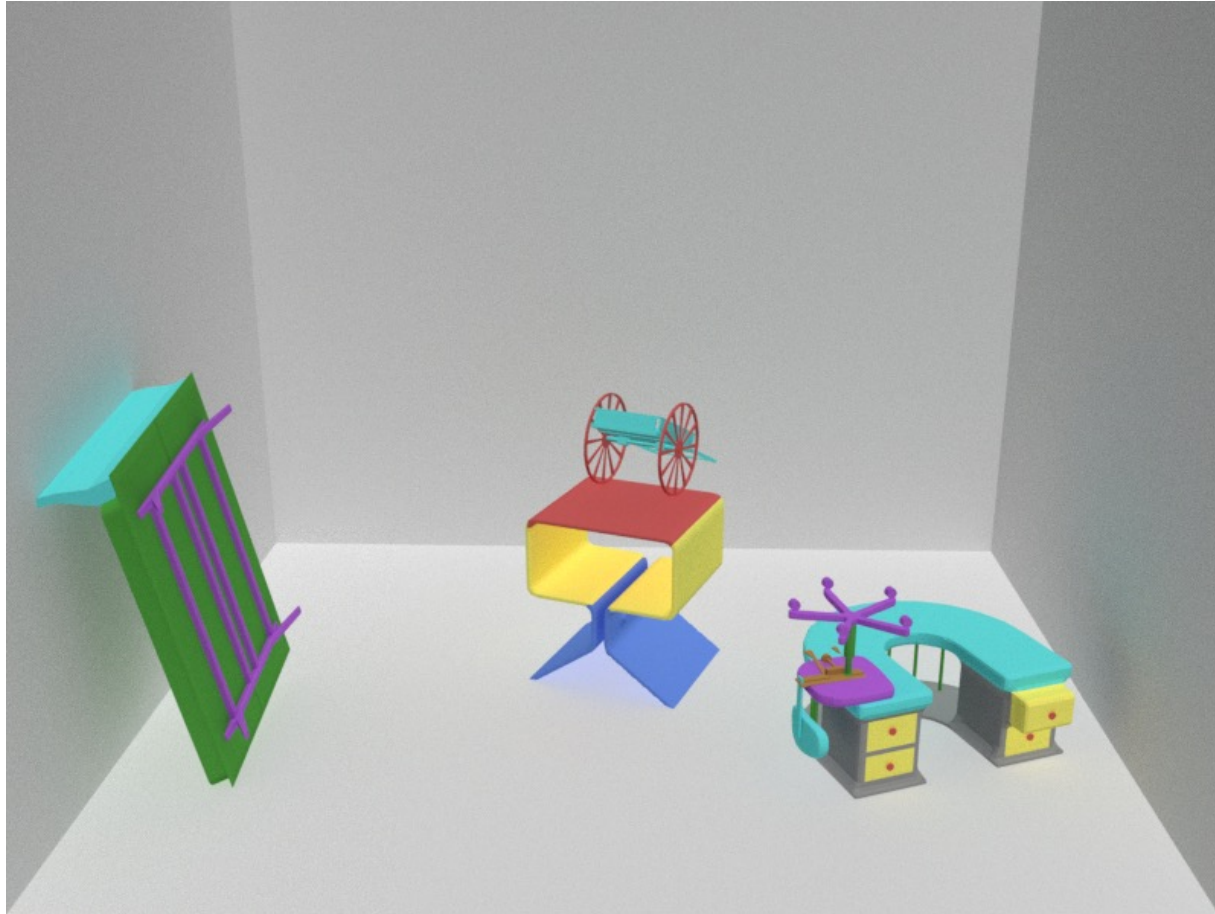


# The PTR Universe



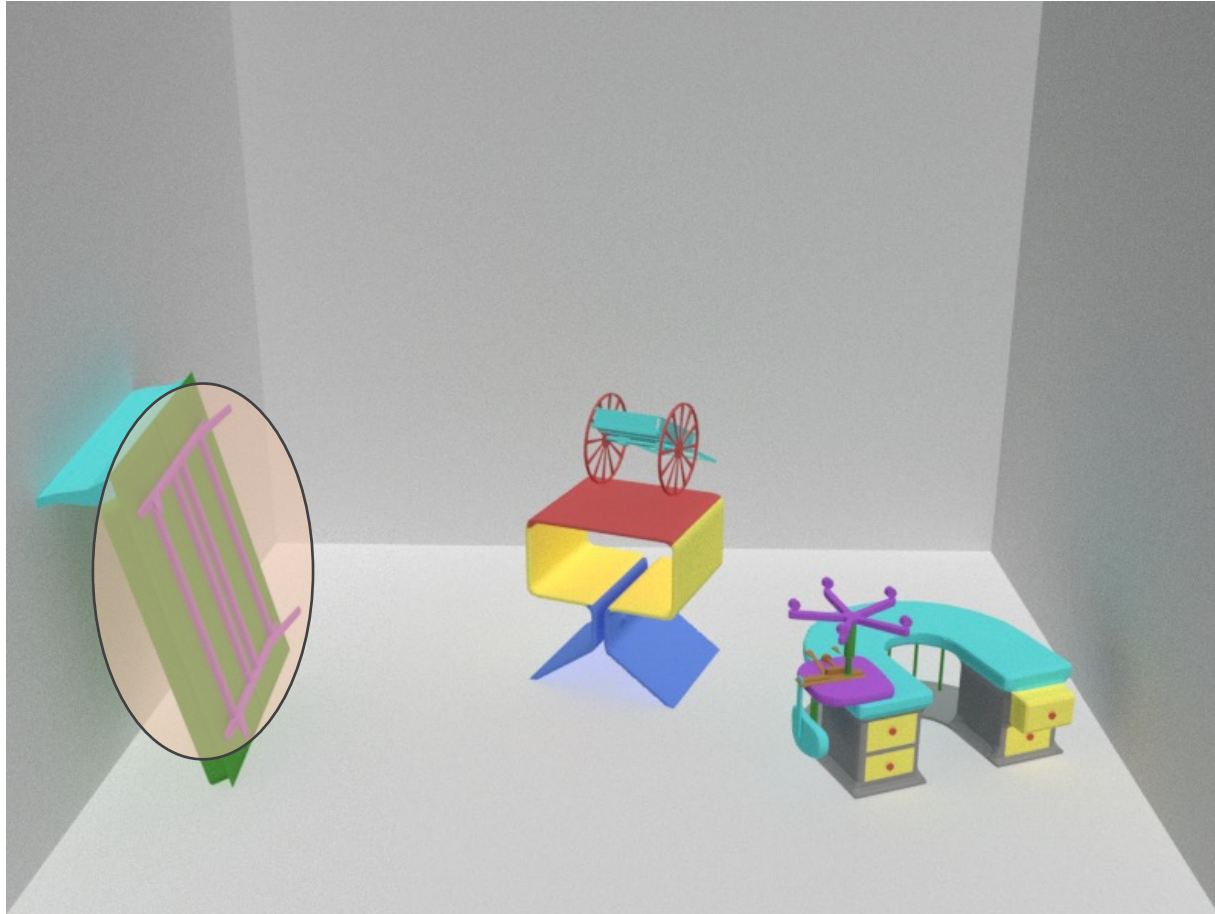
# The PTR Dataset - Examples

# Conceptual Reasoning



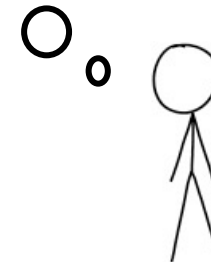
*Question: How many objects with purple legs are there?*

# Conceptual Reasoning

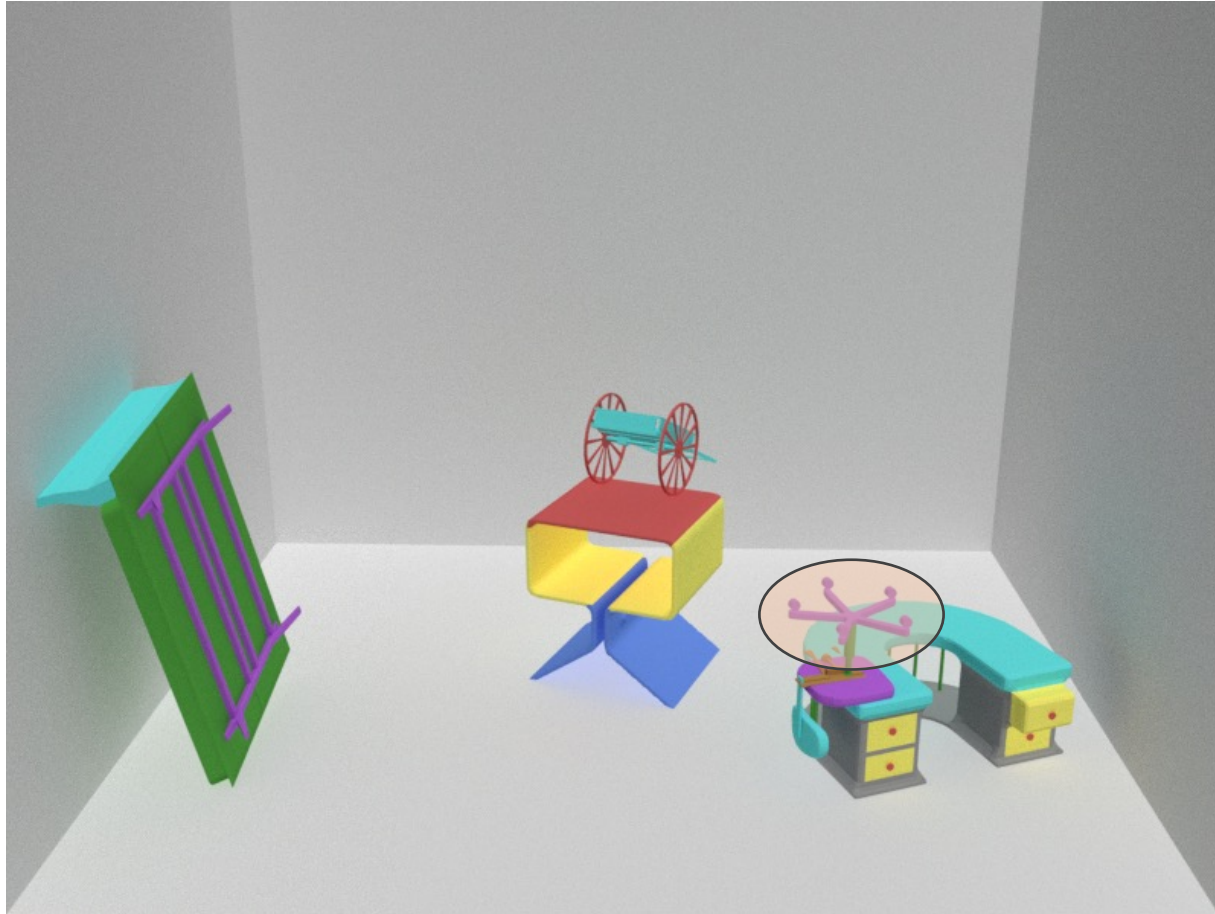


*Question: How many objects with purple legs are there?*

The bed has purple legs



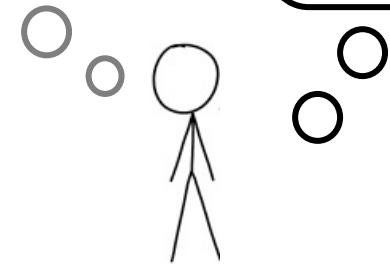
# Conceptual Reasoning



*Question: How many objects with purple legs are there?*

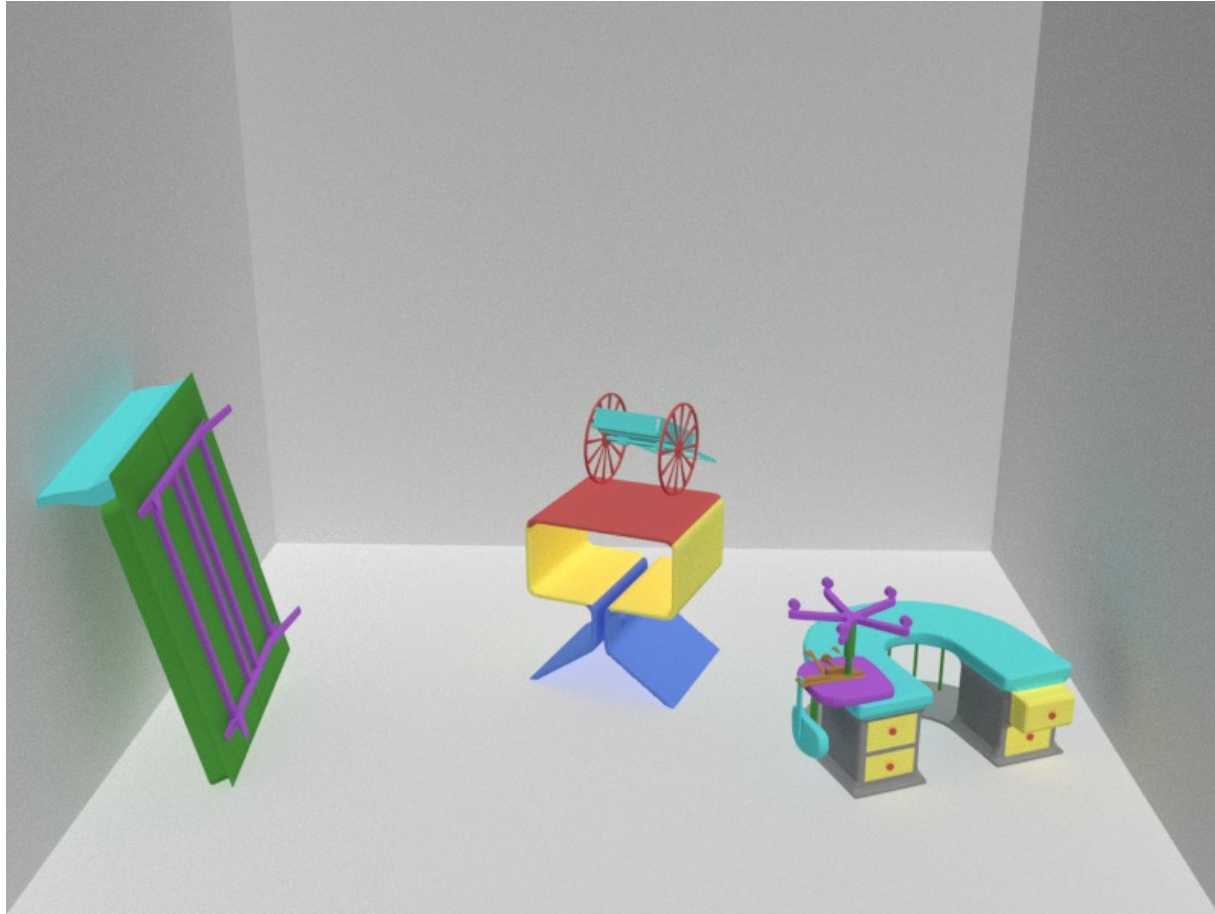
The bed has purple legs

The chair has purple legs

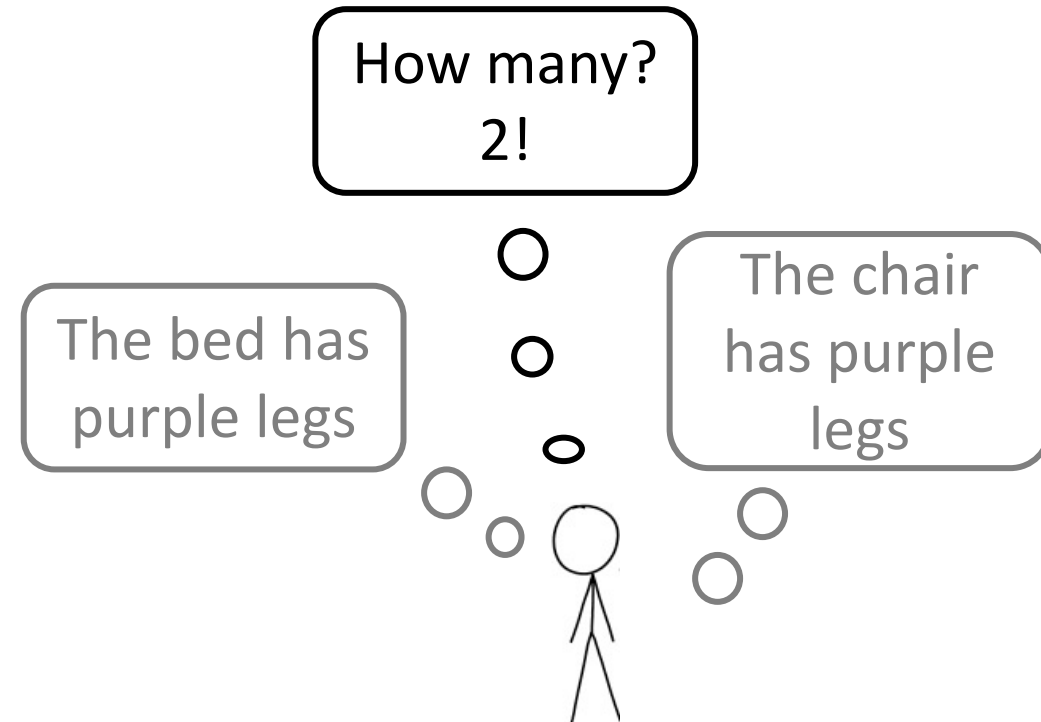




# Conceptual Reasoning

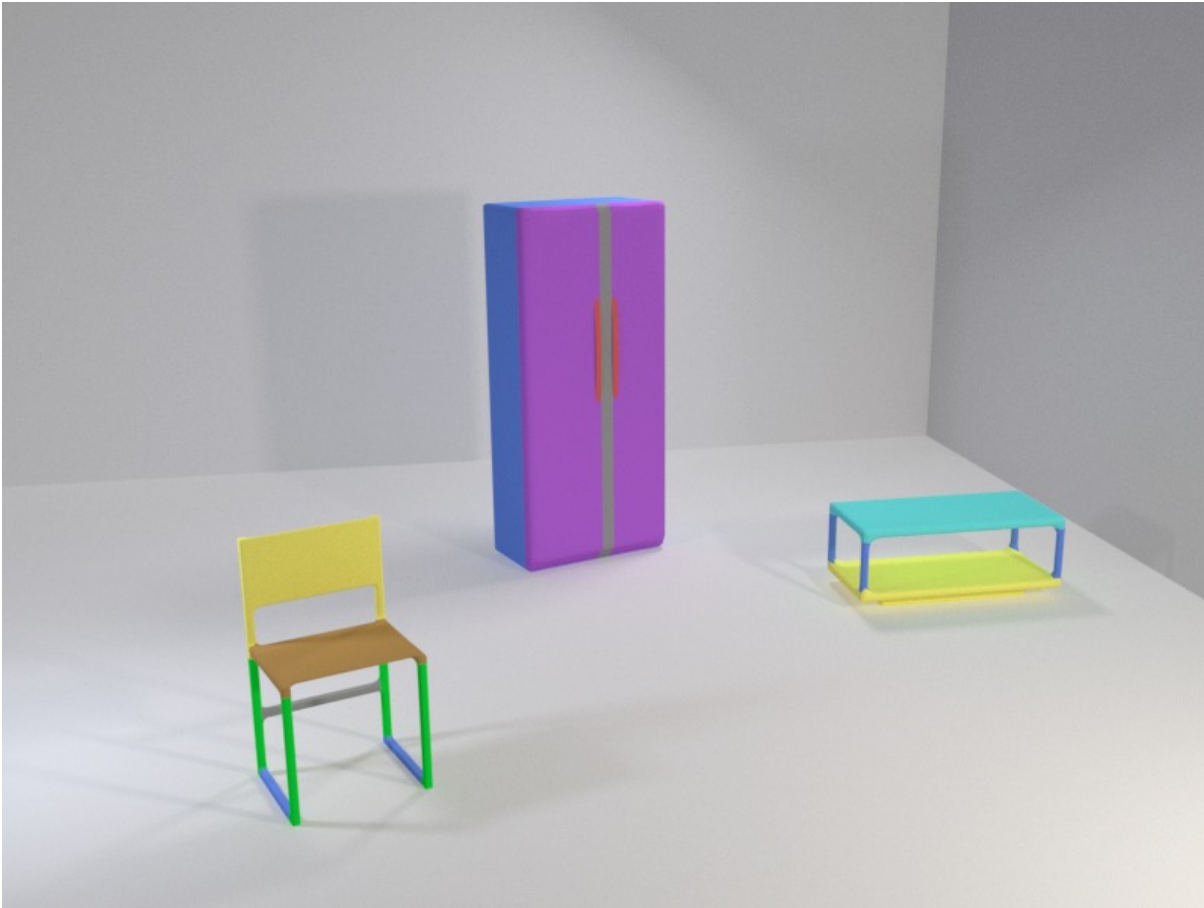


*Question: How many objects with purple legs are there?*



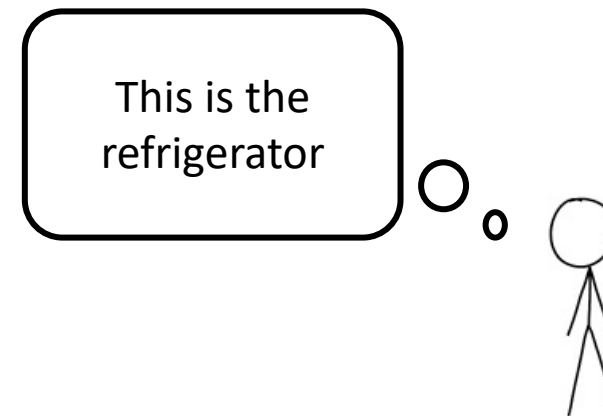
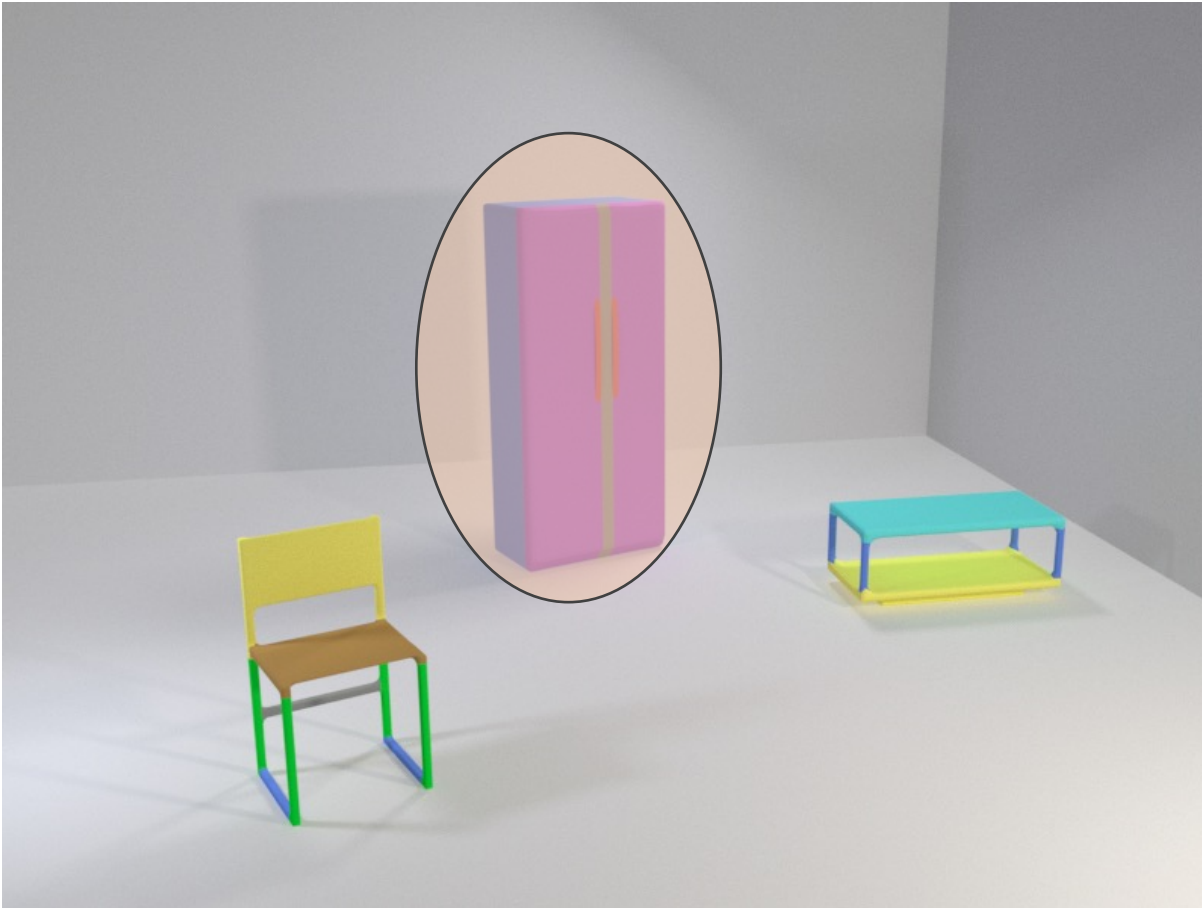
# Relational Reasoning

*Question: What is the color of the part in the chair that can be considered a line, and is perpendicular to the purple part of the refrigerator?*



# Relational Reasoning

*Question: What is the color of the part in the chair that can be considered a line, and is perpendicular to the purple part of the refrigerator?*



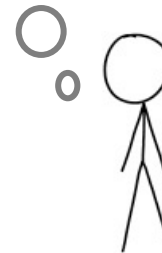
# Relational Reasoning

*Question: What is the color of the part in the chair that can be considered a line, and is perpendicular to the purple part of the refrigerator?*



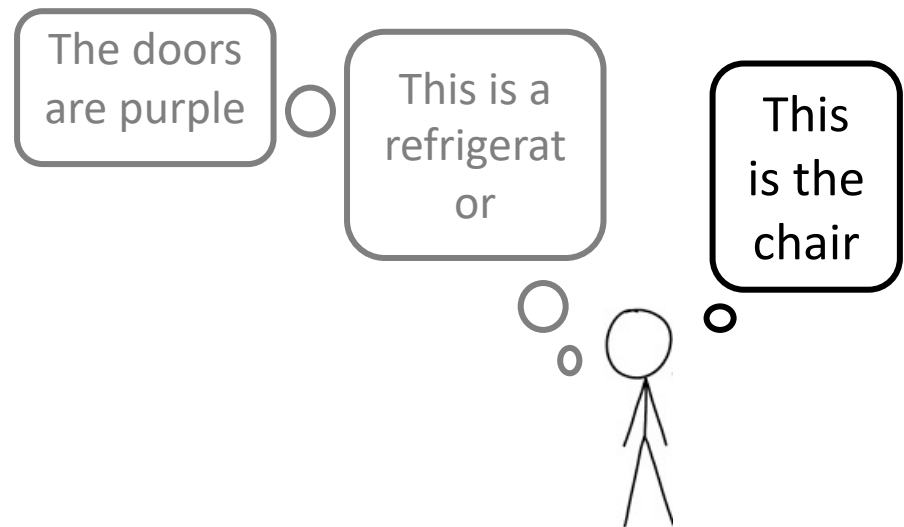
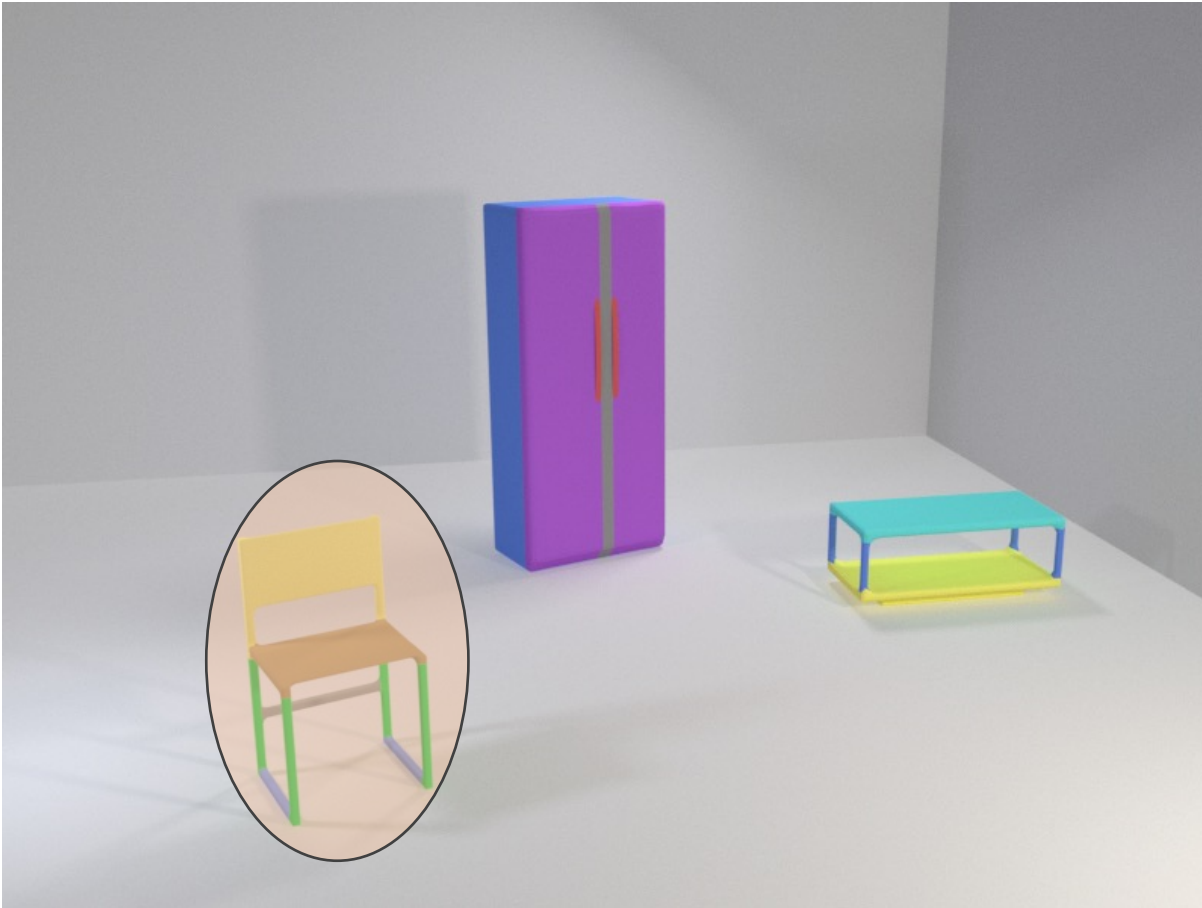
The doors  
are purple

This is a  
refrigerat  
or



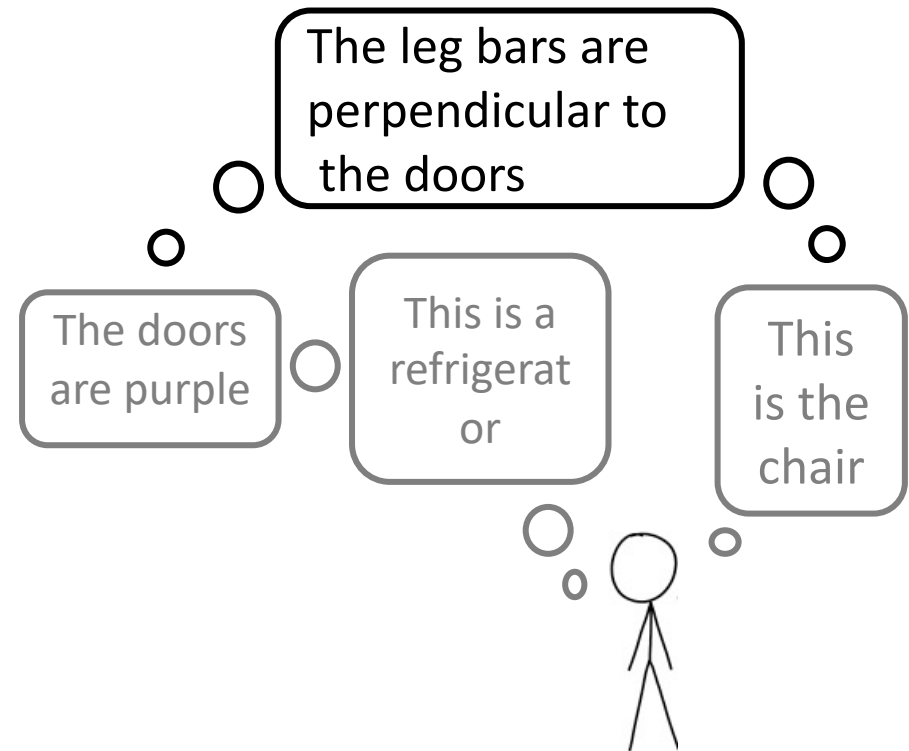
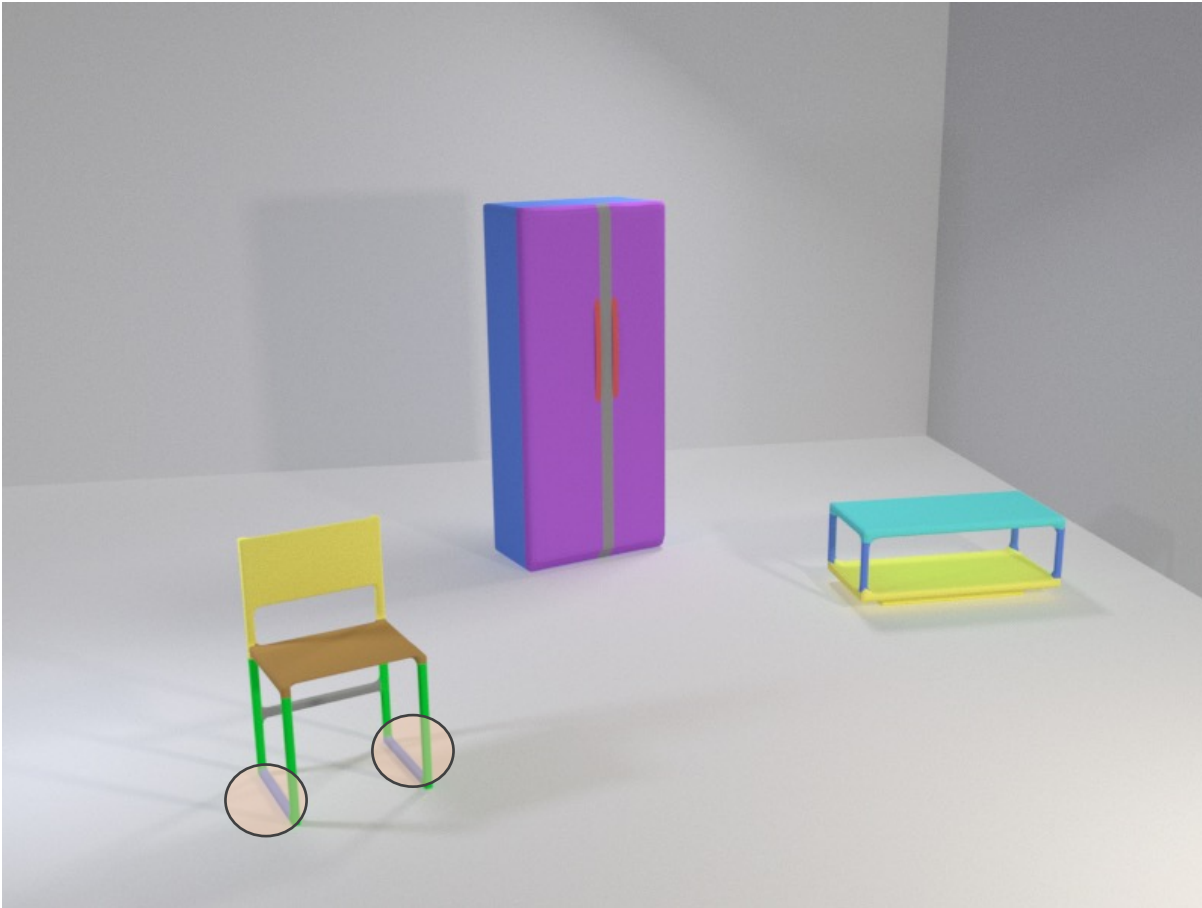
# Relational Reasoning

*Question: What is the color of the part in the chair that can be considered a line, and is perpendicular to the purple part of the refrigerator?*



# Relational Reasoning

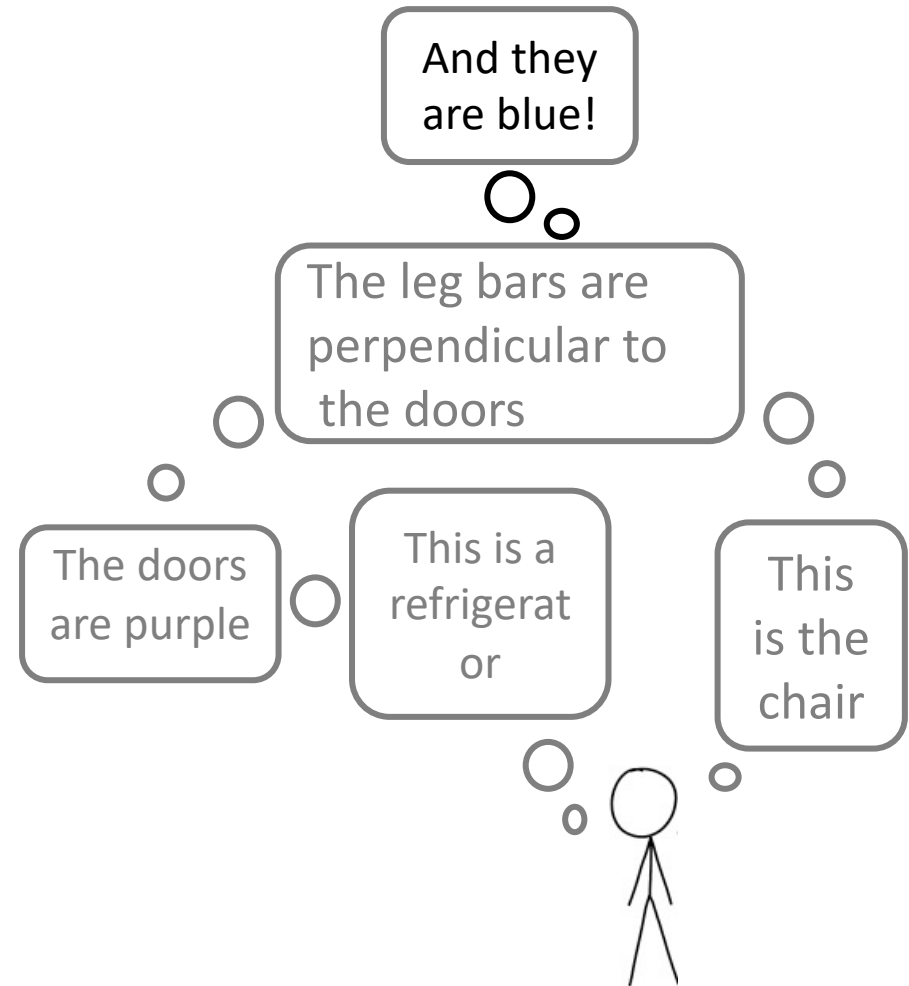
*Question: What is the color of the part in the chair that can be considered a line, and is perpendicular to the purple part of the refrigerator?*



# Relational Reasoning

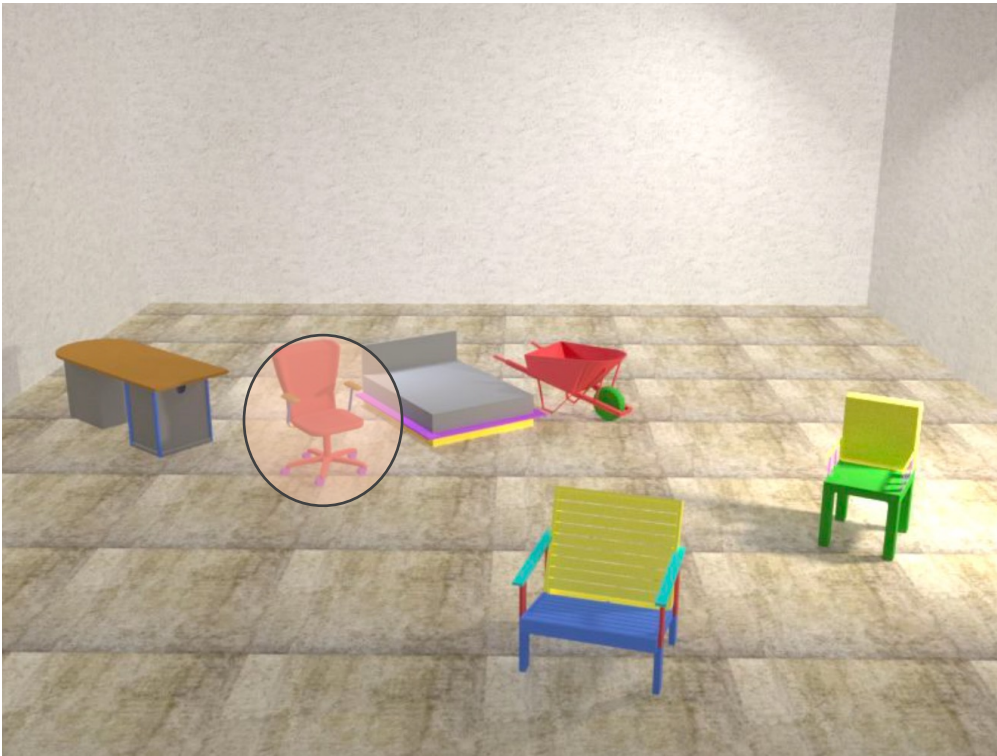


*Question: What is the color of the part in the chair that can be considered a line, and is perpendicular to the purple part of the refrigerator?*



# Analogical Reasoning

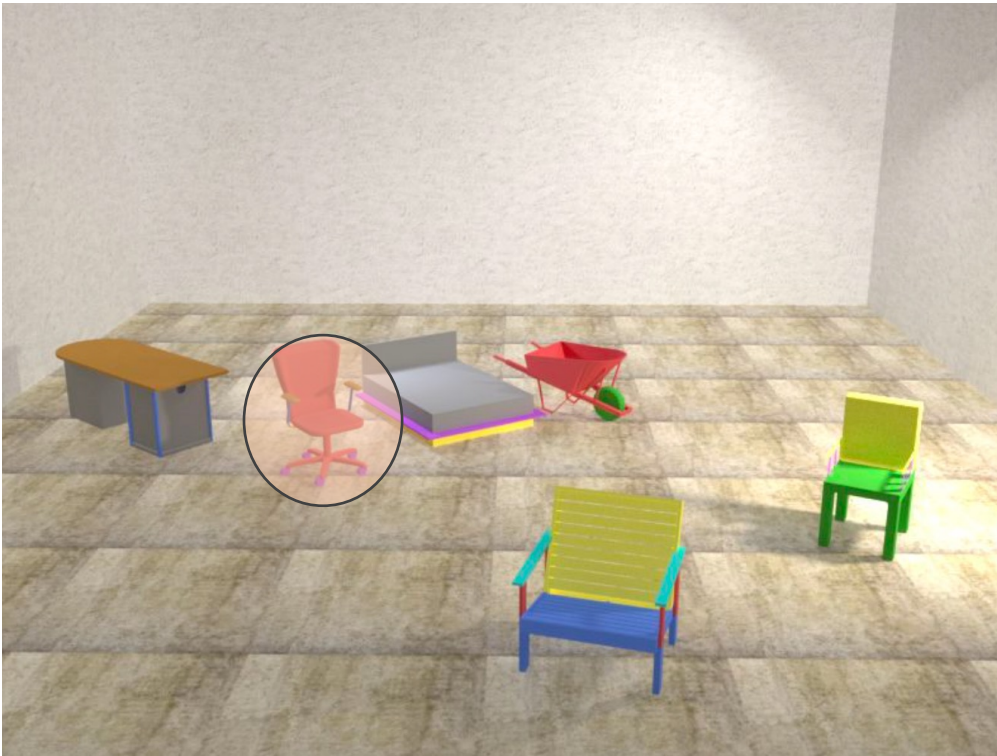
*Question: the thing with five legs has certain positional relation to the object with blue seat. By analogy, how many objects does the bed have the same positional relation to ?*





# Analogy Reasoning

*Question: the thing with five legs has certain positional relation to the object with blue seat. By analogy, how many objects does the bed have the same positional relation to ?*

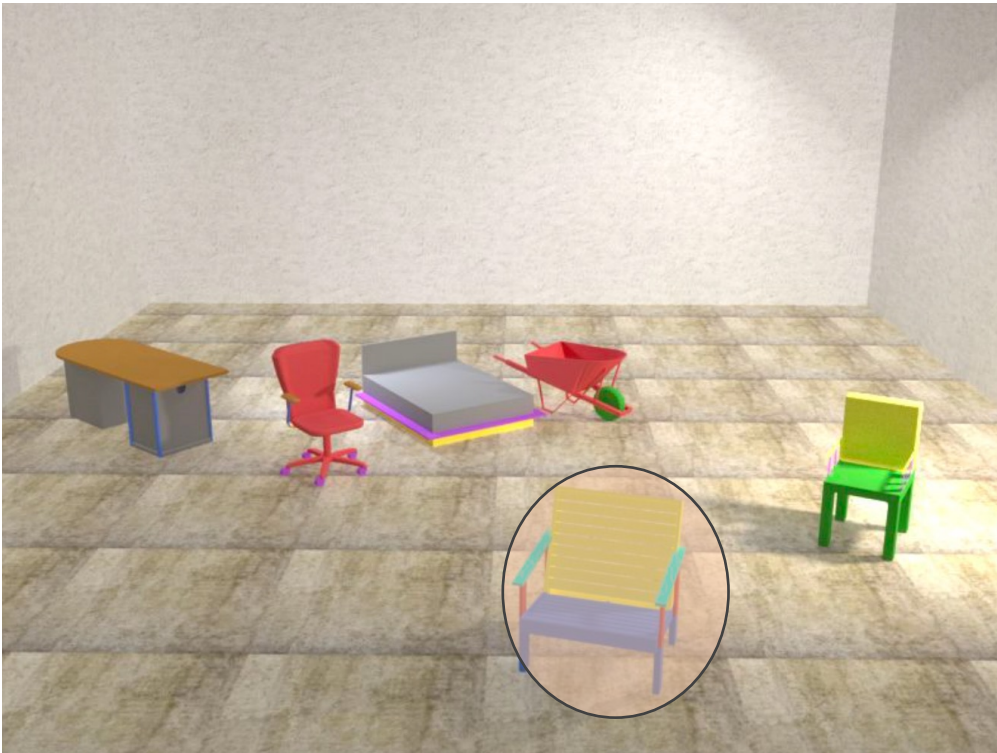


This chair has  
five legs



# Analogical Reasoning

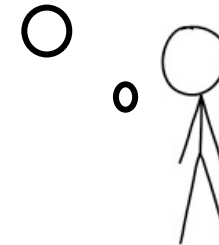
*Question: the thing with five legs has certain positional relation to the object with blue seat. By analogy, how many objects does the bed have the same positional relation to ?*



This chair has  
five legs

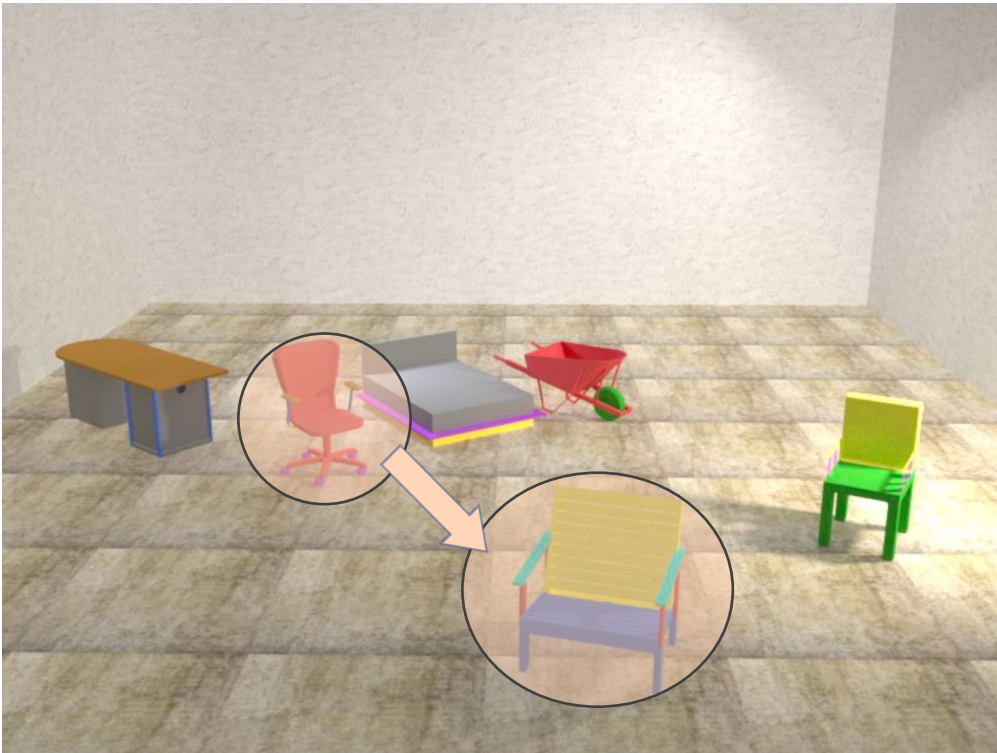


This chair has a  
blue seat



# Analogical Reasoning

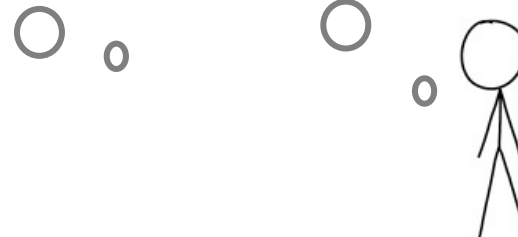
*Question: the thing with five legs has certain positional relation to the object with blue seat. By analogy, how many objects does the bed have the same positional relation to ?*



The first object is behind and to the left of the second

This chair has five legs

This chair has a blue seat



# Analogical Reasoning

*Question: the thing with five legs has certain positional relation to the object with blue seat. By analogy, how many objects does the bed have the same positional relation to ?*

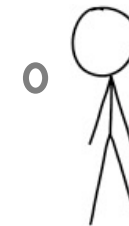


The first object is behind and to the left of the second

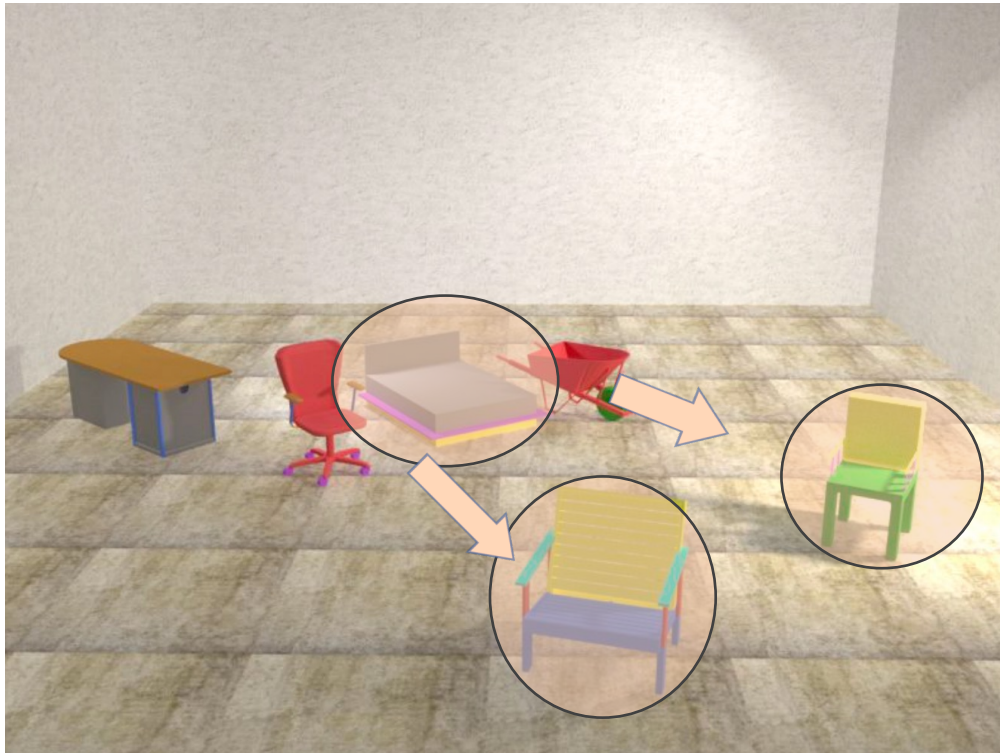
This chair has five legs

This chair has a blue seat

This is the bed



# Analogical Reasoning



*Question: the thing with five legs has certain positional relation to the object with blue seat. By analogy, how many objects does the bed have the same positional relation to ?*

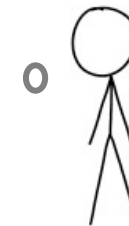
The first object is behind and to the left of the second

It's behind and to the left of TWO objects!

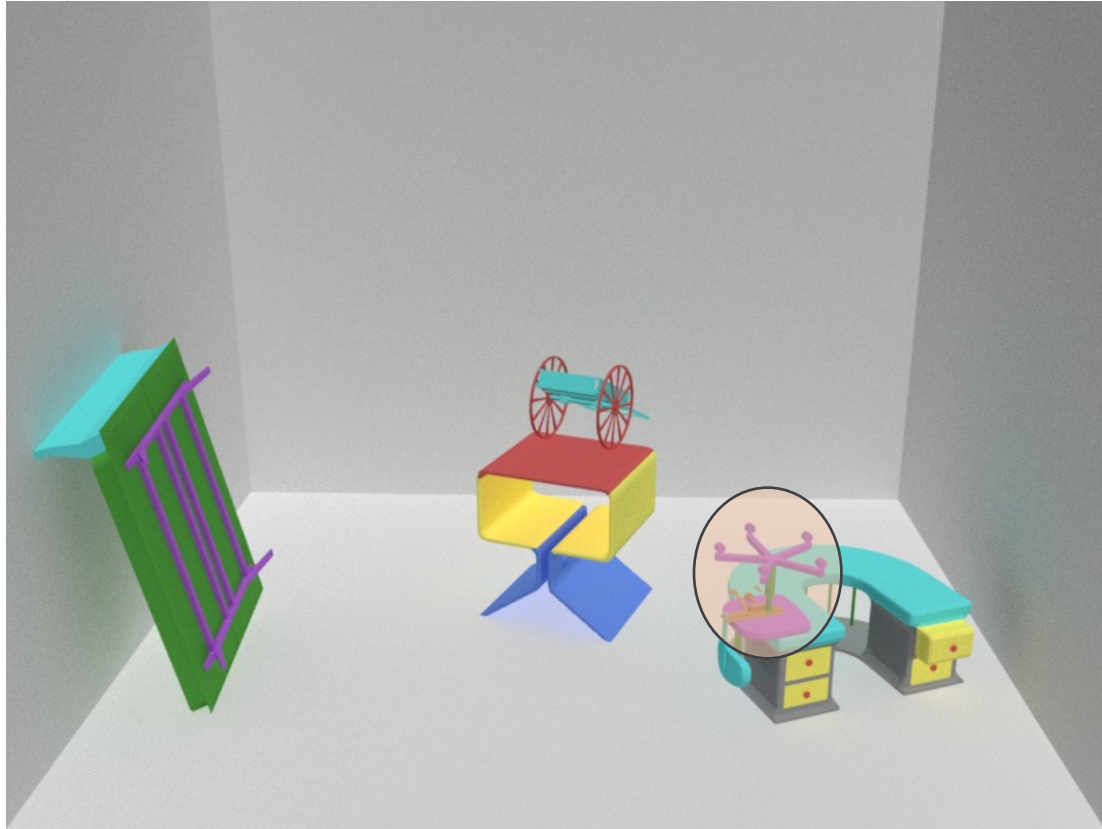
This chair has five legs

This chair has a blue seat

This is the bed

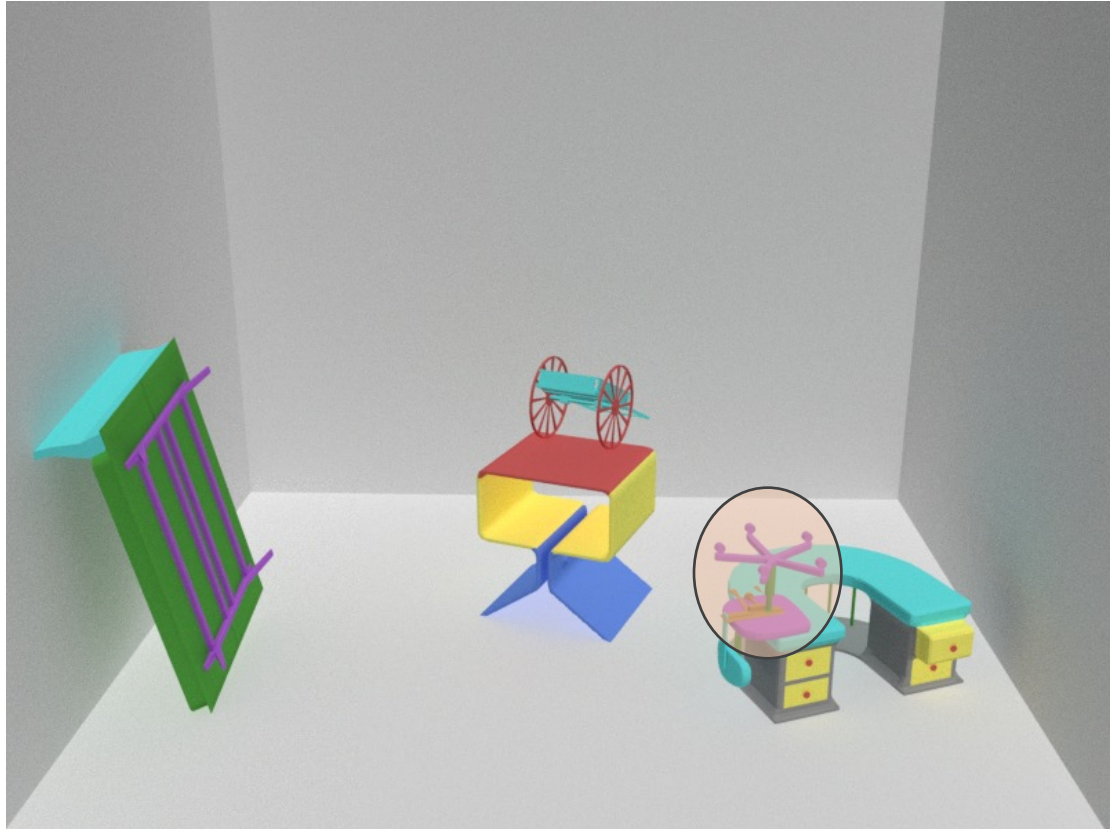


# Arithmetic Reasoning



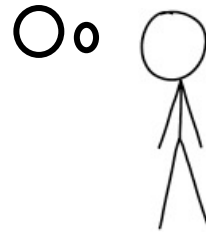
- *What is the sum of the number of legs in the chair, and number of wheels in the cart?*

# Arithmetic Reasoning

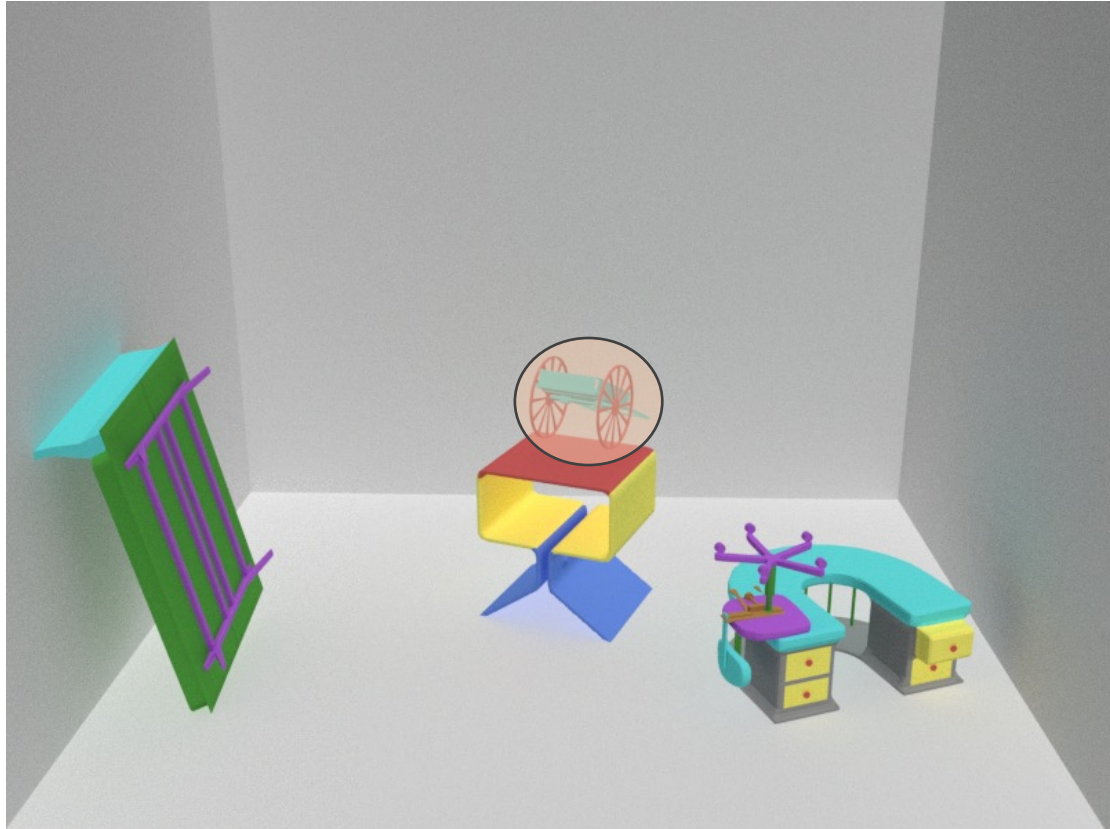


- *What is the sum of the number of legs in the chair, and number of wheels in the cart?*

The chair has  
five legs



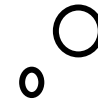
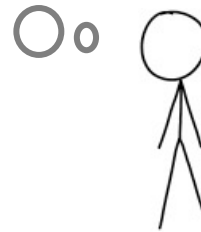
# Arithmetic Reasoning



- *What is the sum of the number of legs in the chair, and number of wheels in the cart?*

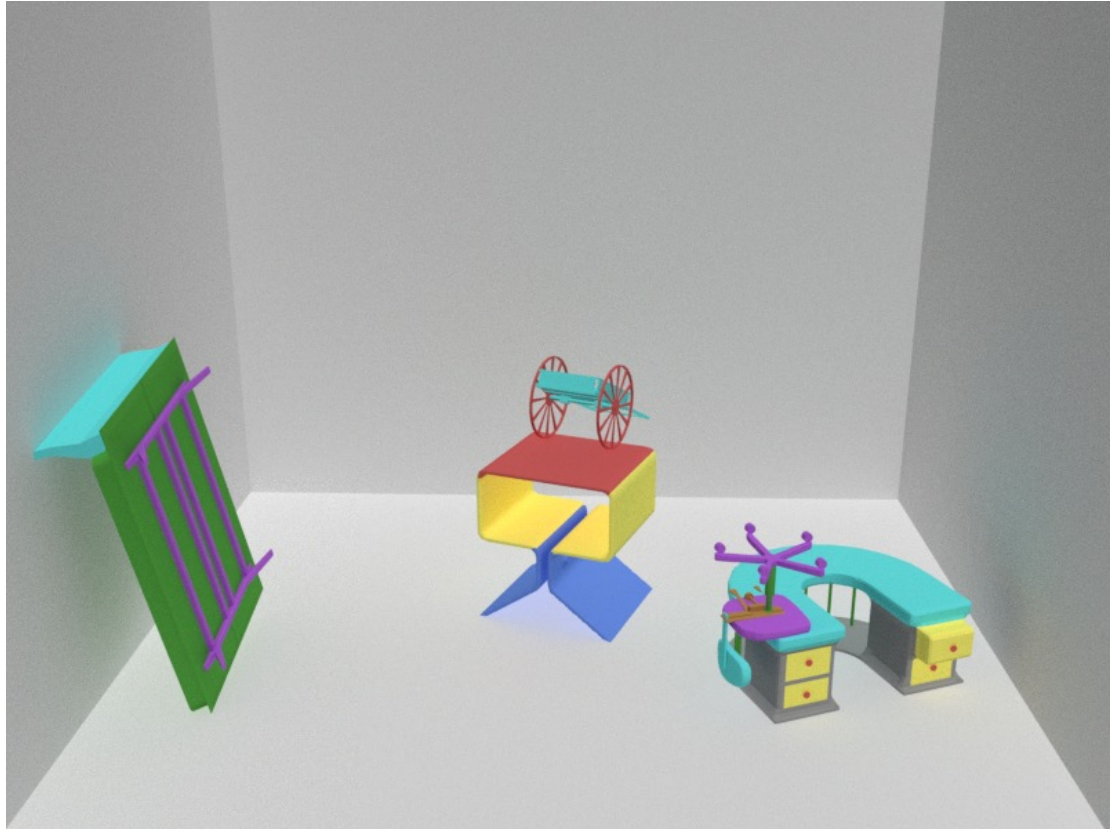
The chair has  
five legs

The cart has  
two wheels





# Arithmetic Reasoning

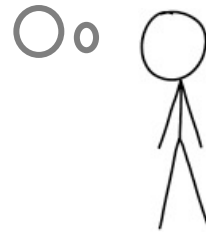


- *What is the sum of the number of legs in the chair, and number of wheels in the cart?*

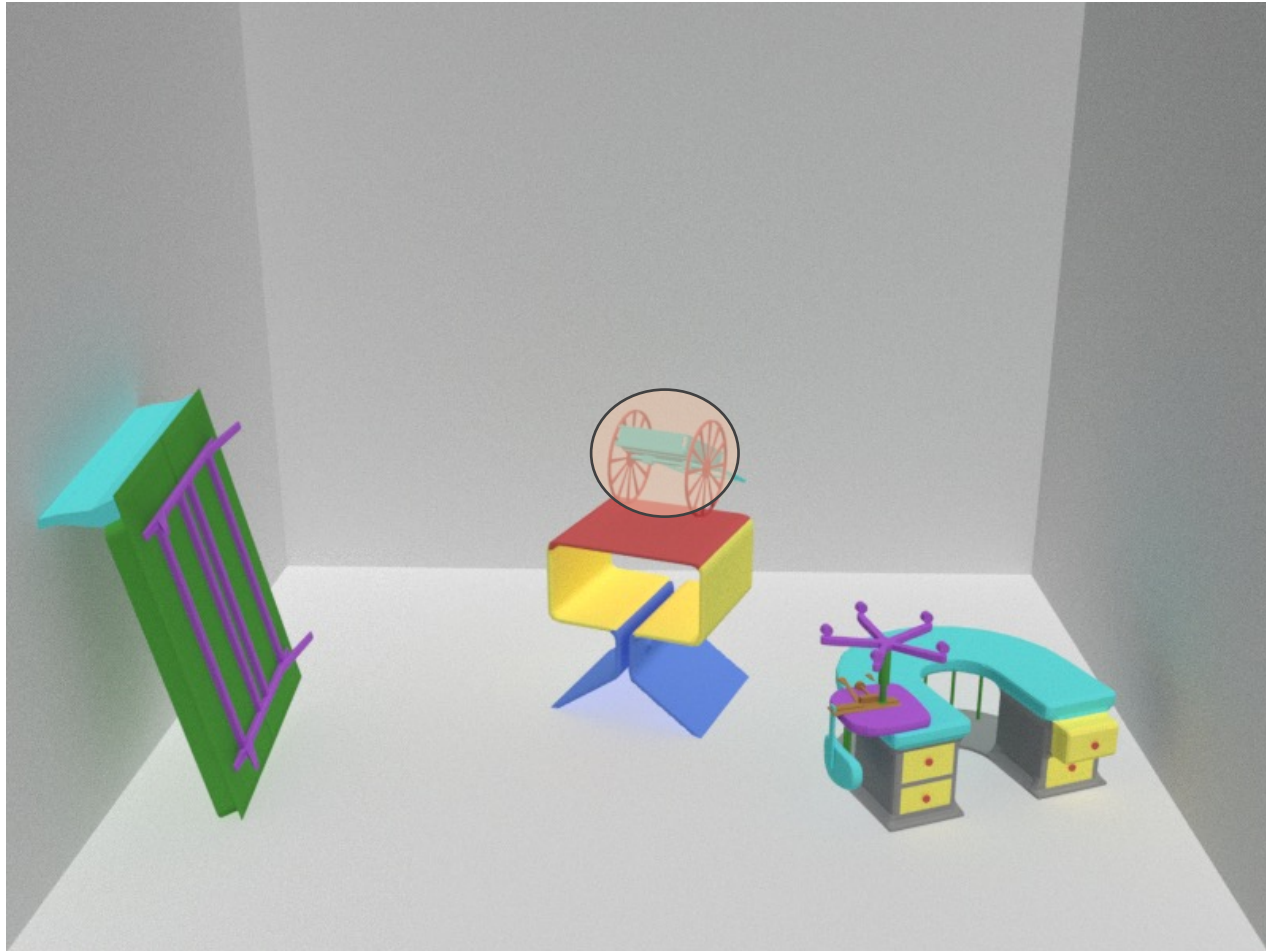
$$5 + 2 = 7 !$$

The chair has  
five legs

The cart has  
two wheels

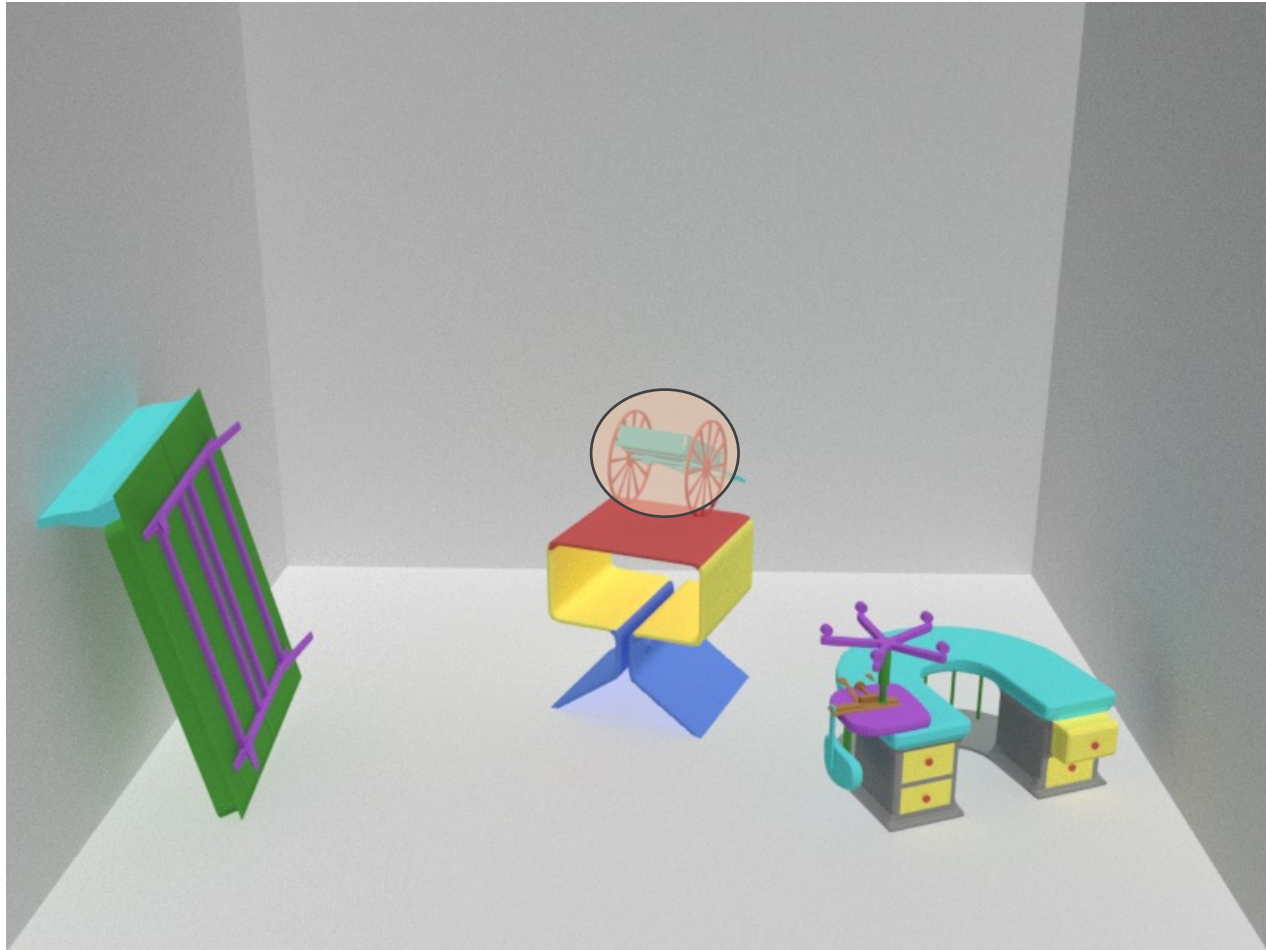


# Physical Reasoning



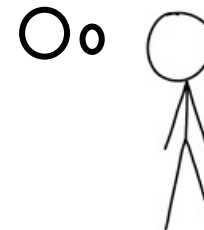
*Towards which direction shall the cart move to become stable?*

# Physical Reasoning

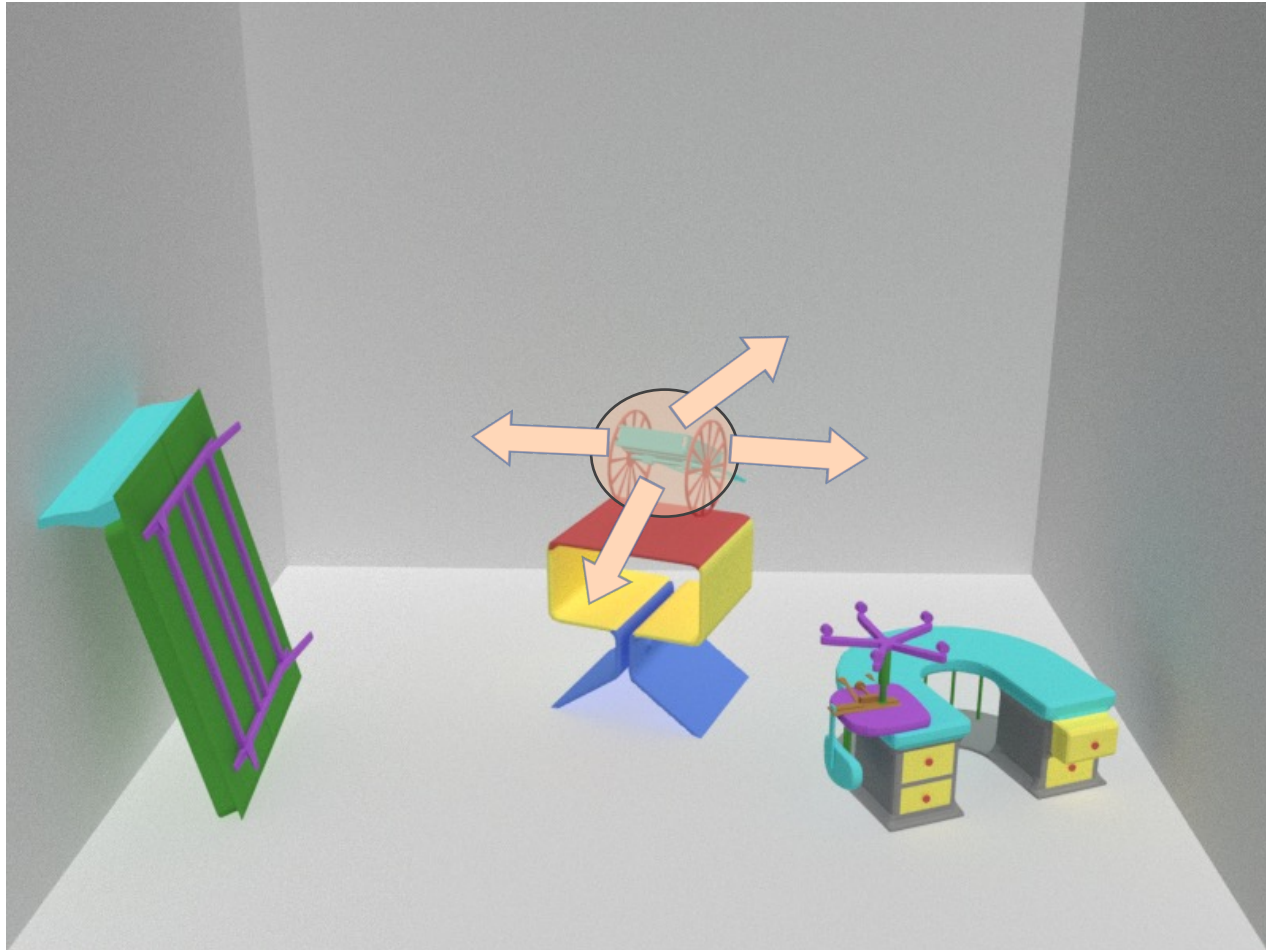


*Towards which direction shall the cart move to become stable?*

This is the cart  
and it will fall  
off the table



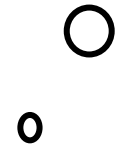
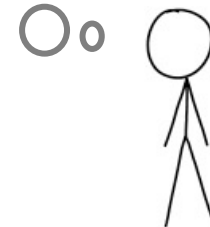
# Physical Reasoning



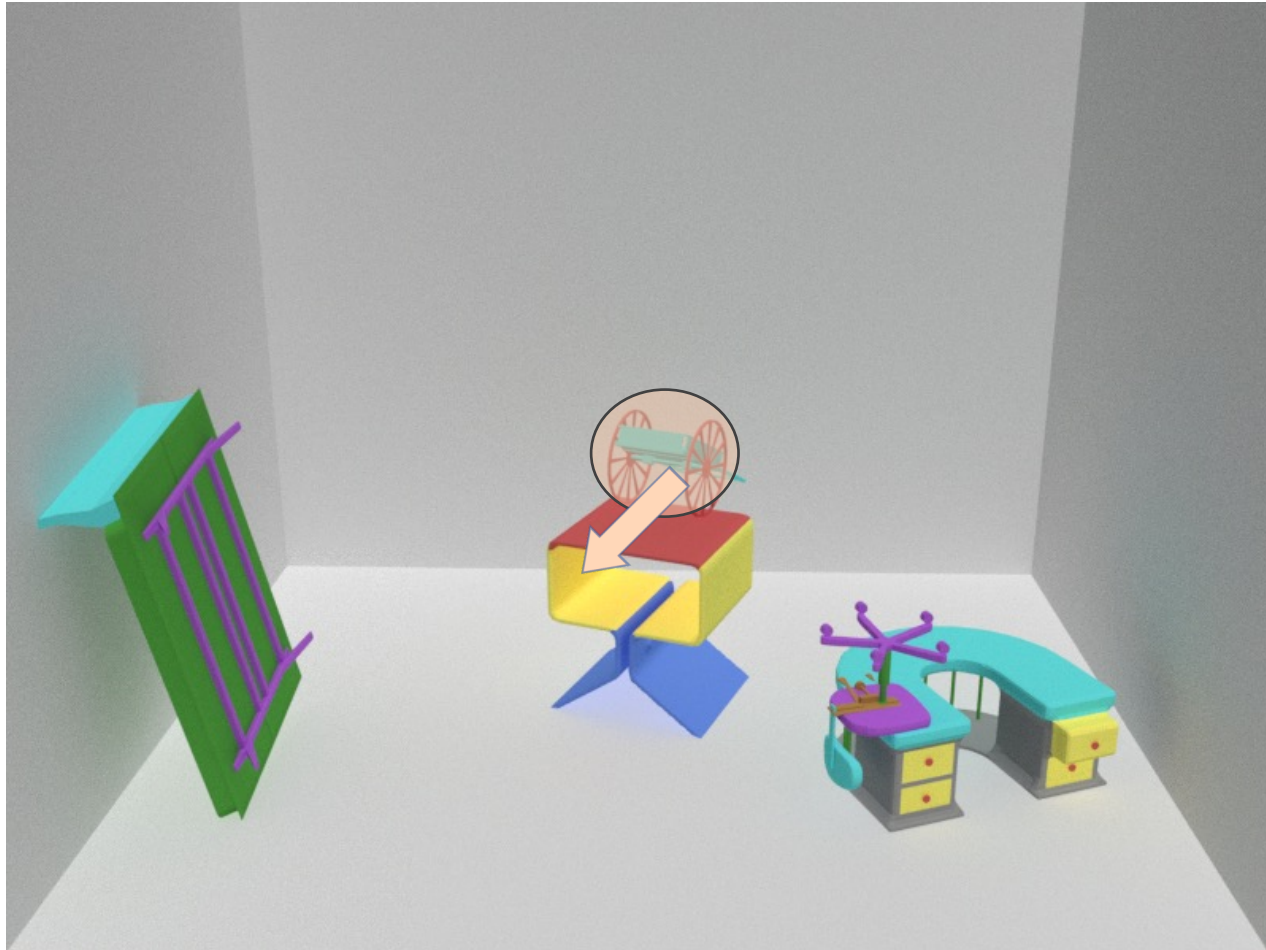
*Towards which direction shall the cart move to become stable?*

This is the cart and it will fall off the table

Mental Simulation of Possible Changes



# Physical Reasoning

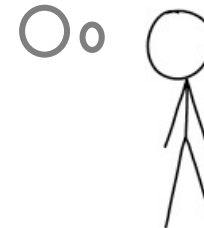


*Towards which direction shall the cart move to become stable?*

Moving front!

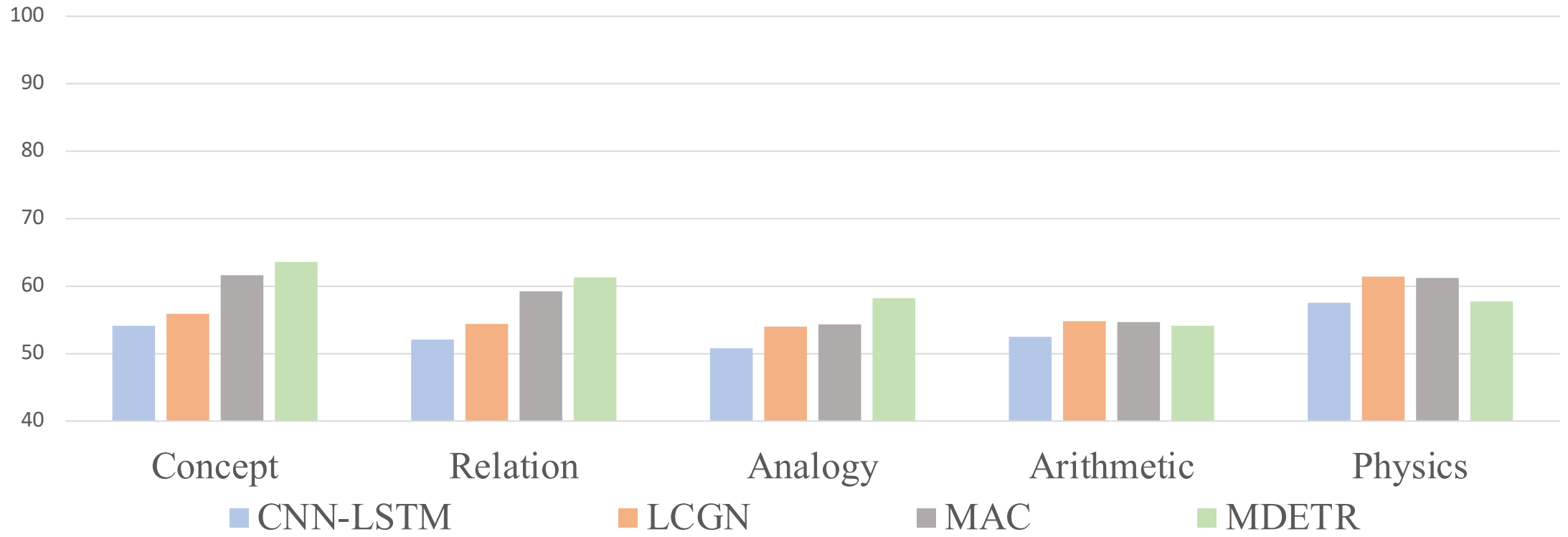
Mental  
Simulation of  
Possible  
Changes

This is the cart  
and it will fall of  
the table



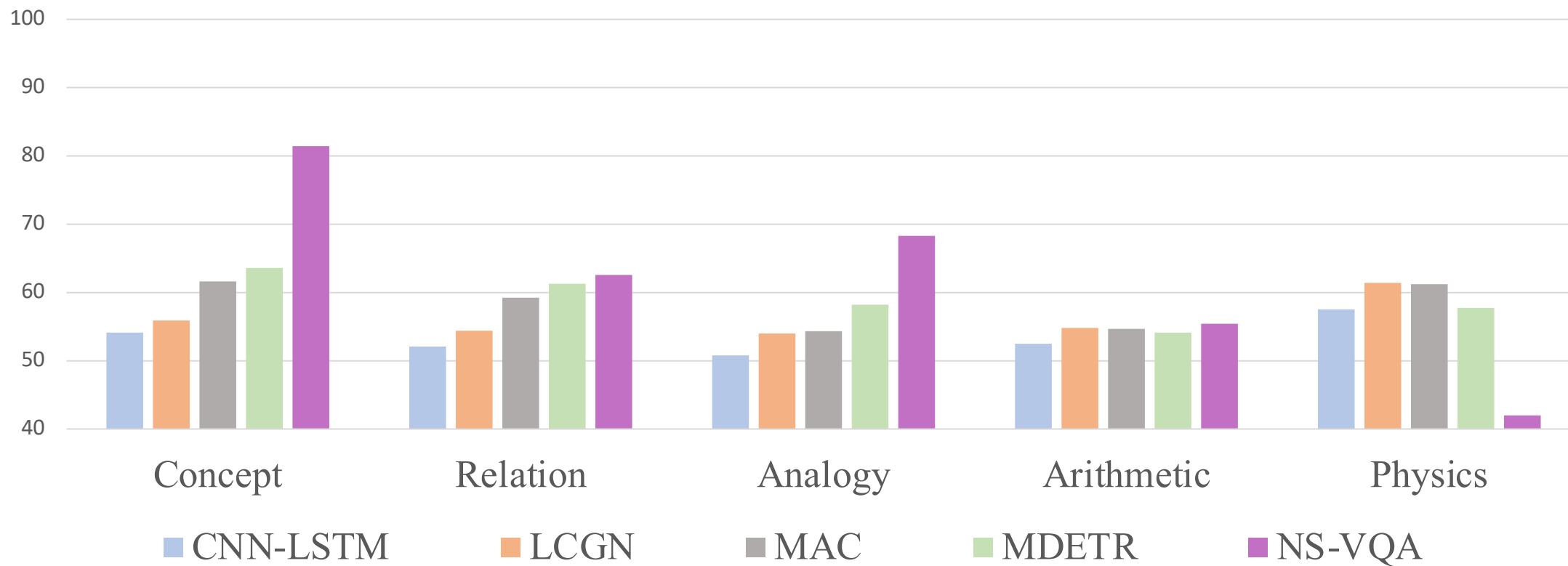
# Experimental Results

## PTR Test Accuracy



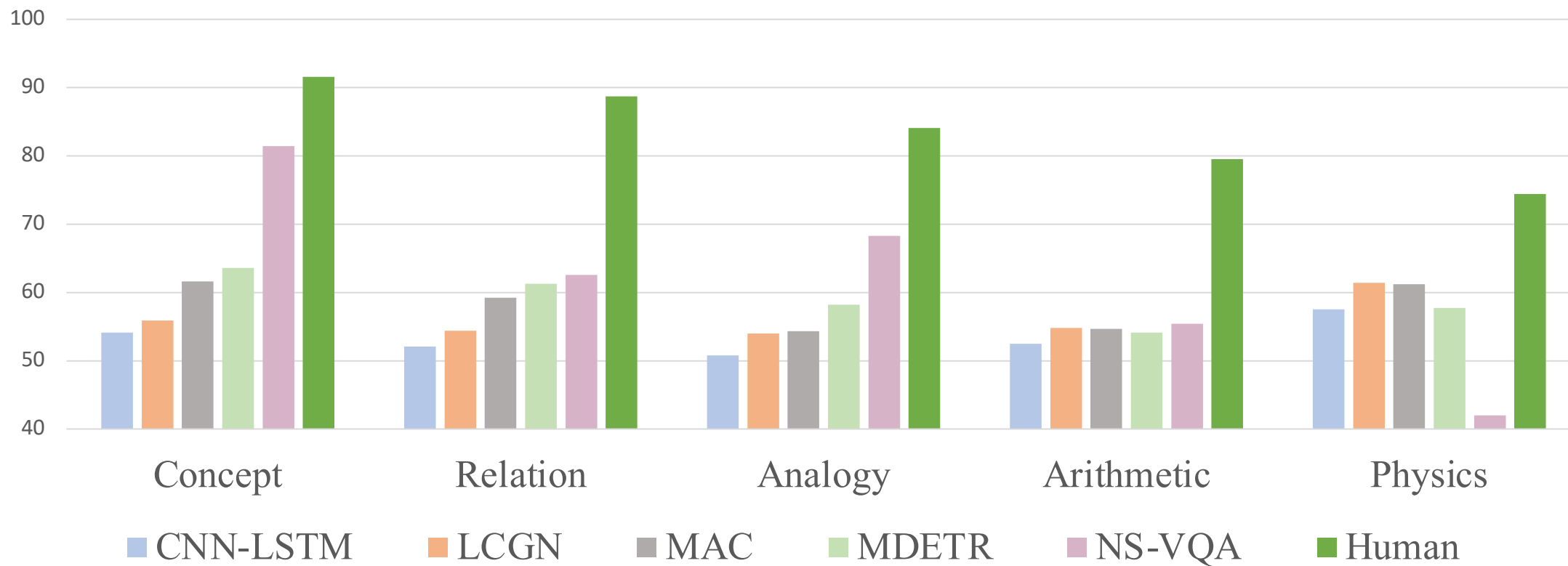
# Experimental Results

## PTR Test Accuracy



# Experimental Results

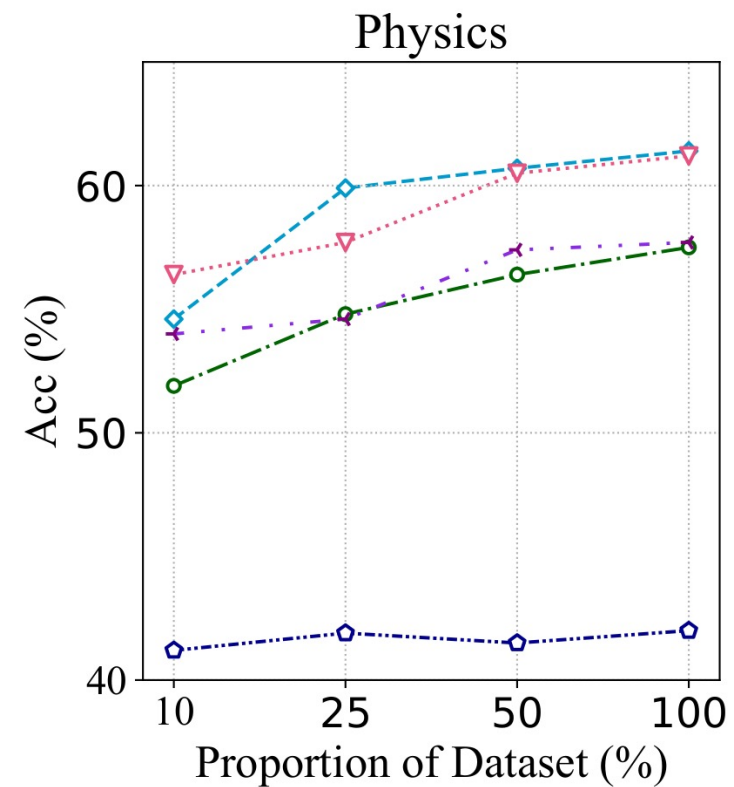
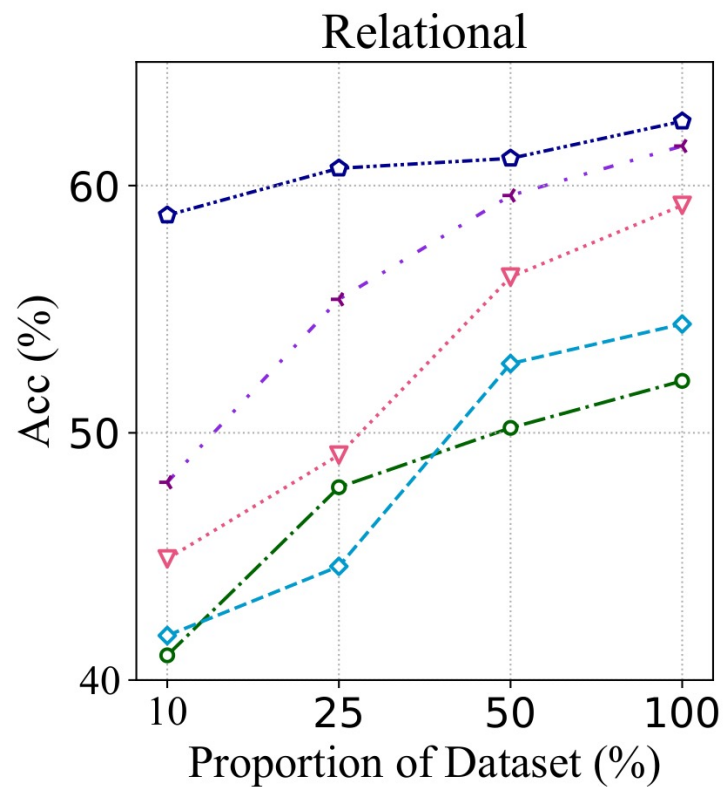
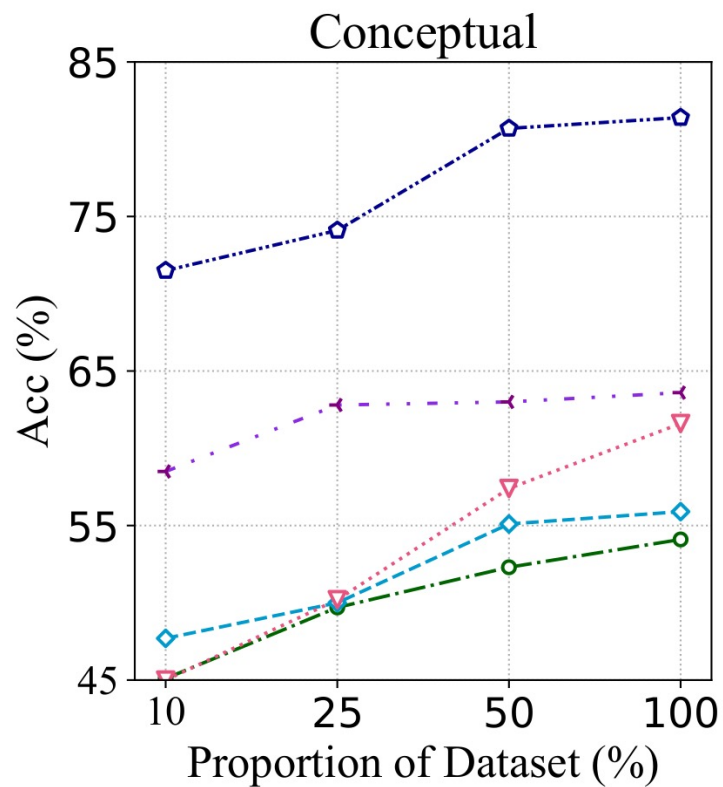
## PTR Test Accuracy



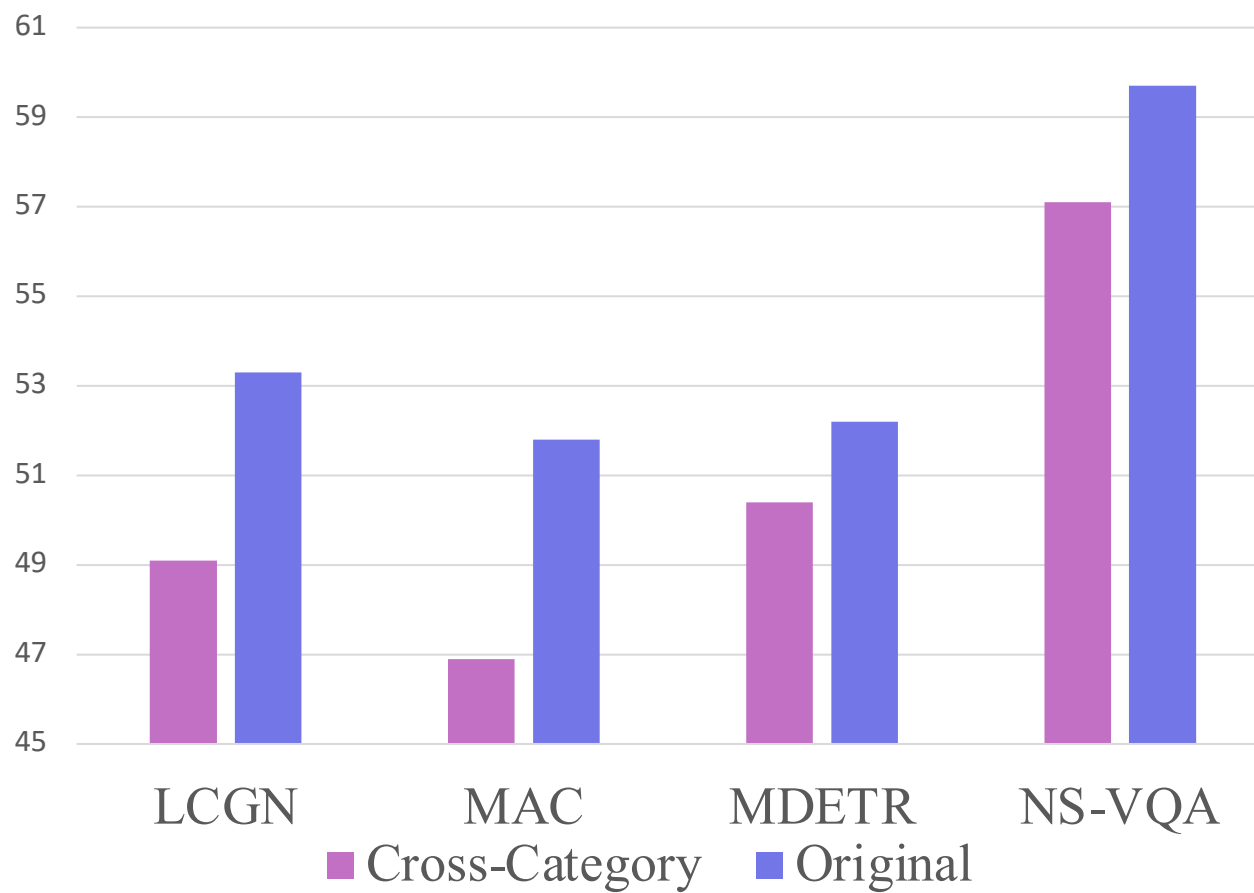


# Data Efficiency

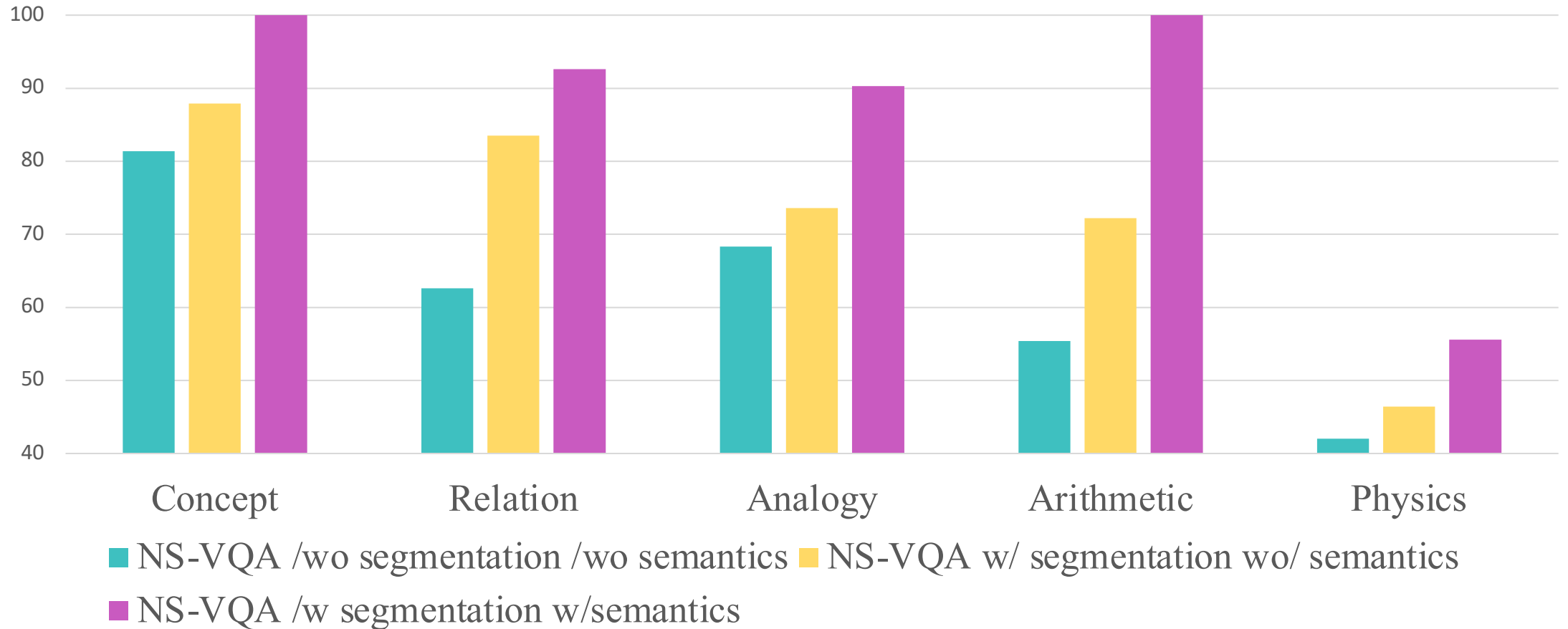
◇ CNN+LSTM    ○ LCGN    ◇ MAC    ▽ MDETR    ✕ NS-VQA



# Cross-Category Generalization



# Ablative Studies on NS-VQA



# Summary

- Next generation visual-reasoning benchmark on part-based reasoning
- Four question types:
  - Concept
  - Relation
  - Analogy
  - Arithmetic
  - Physics
- Baseline Analyses of neural / neural-symbolic models
- Long way to go compared with human performances

Thanks

---