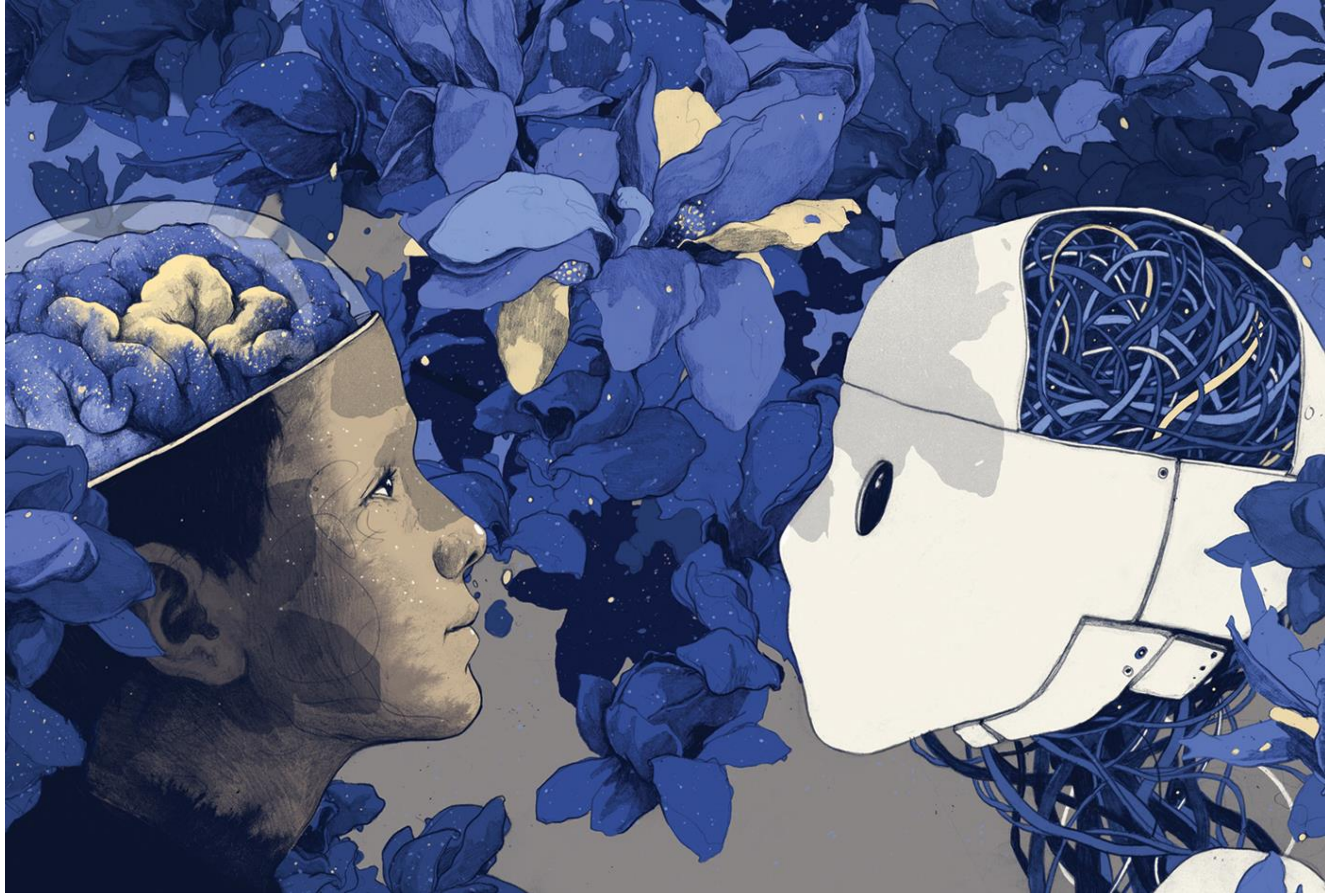
Analogical Reasoning in Children and AI

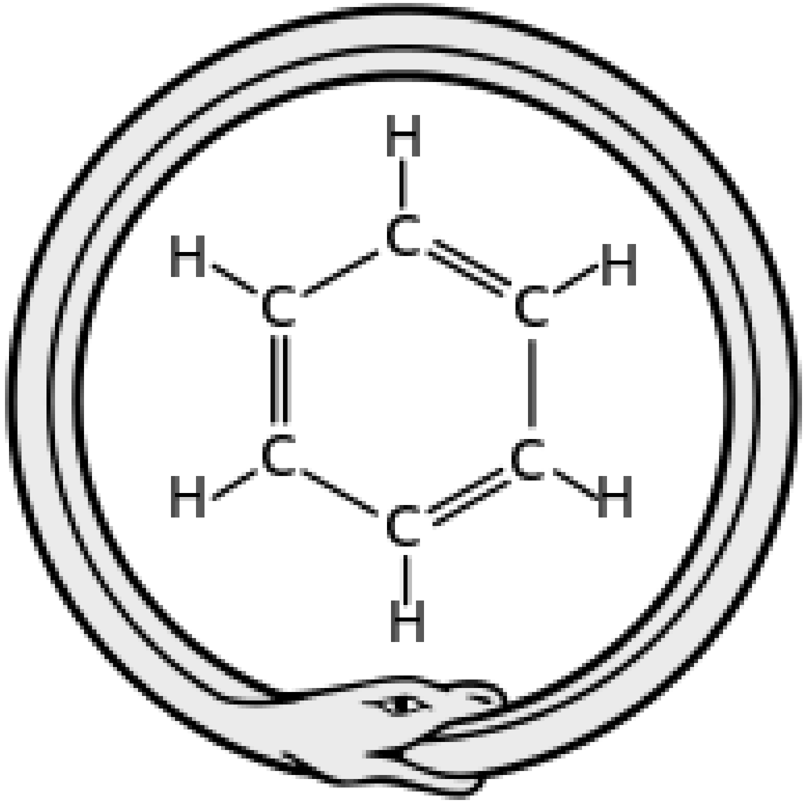
Alison Gopnik, Eunice Yiu, Mariel Goddu

“Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's?” Alan Turing, 1950.

Analogical Reasoning and Generalization

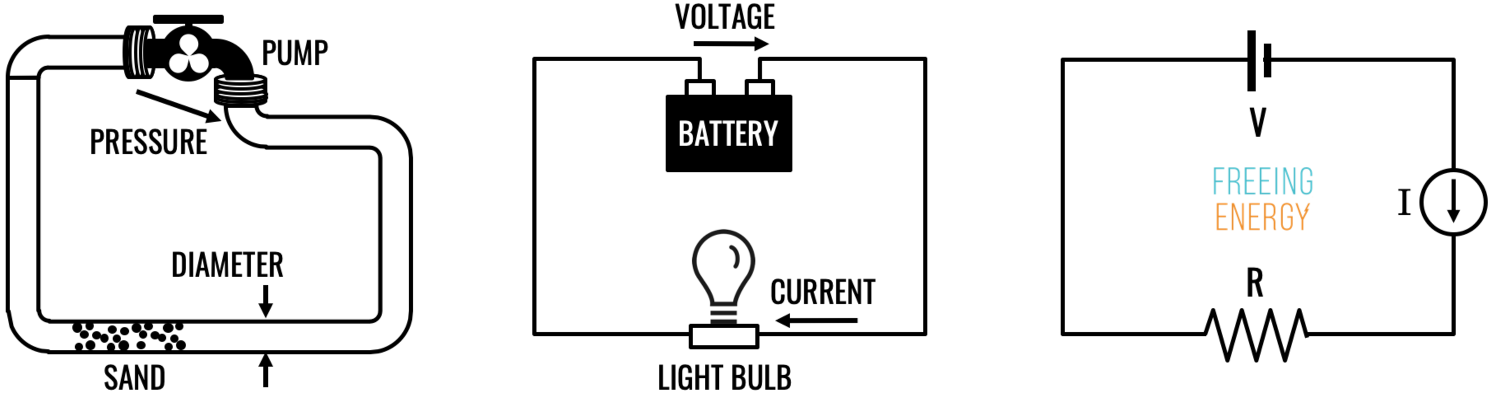
Lombrozo (2019);
Prystawski et al. (2024)



Ouroboros:
The cyclic structure of benzene is analogous to a snake eating its own tail.
image source: Wikipedia

Voltage = Current x Resistance
($V = I \times R$)

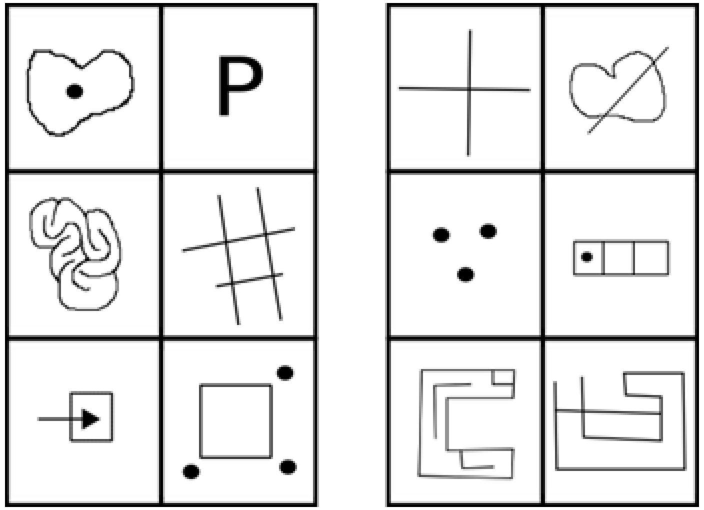
Water | Electricity | Circuit Diagram



Electricity flows like water.
image source: FreeingEnergy

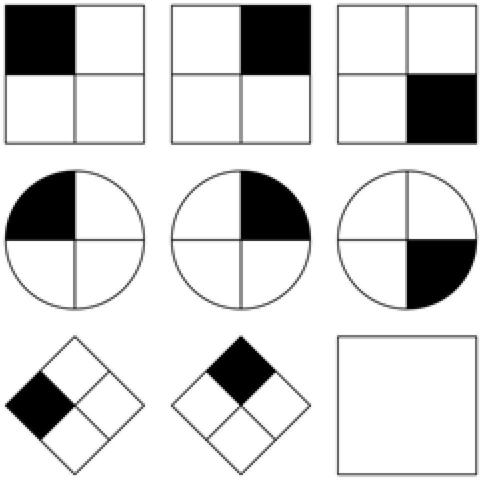
Analogical Reasoning as a Challenge for AI

Bongard Problems



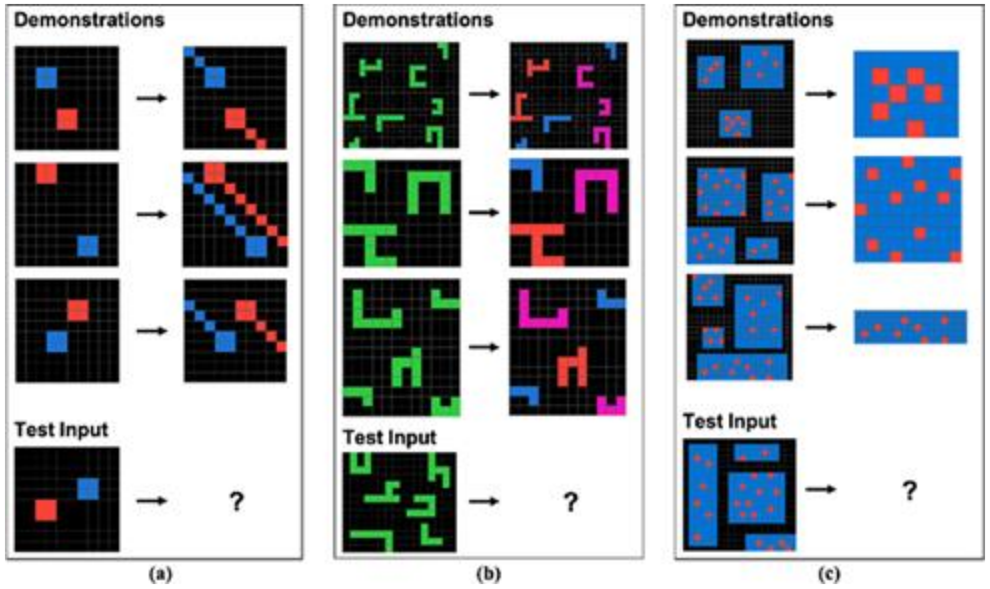
Sonwane et al., 2021; Nie et al., 2020

Raven's Progressive Matrices



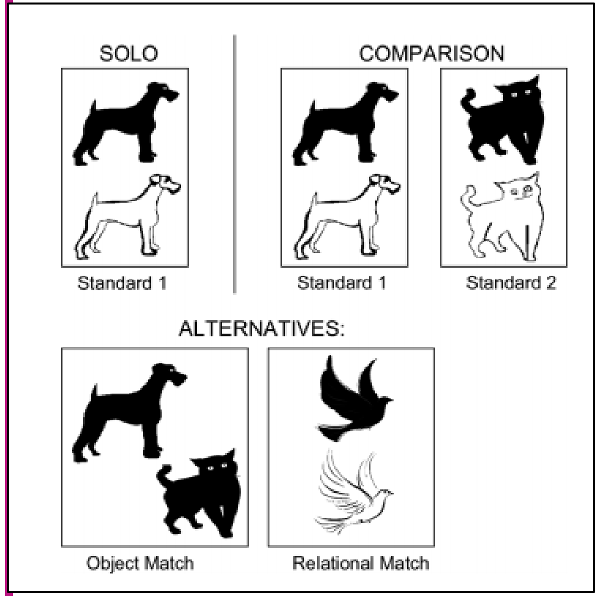
Zhang et al., 2019; Barrett et al., 2018

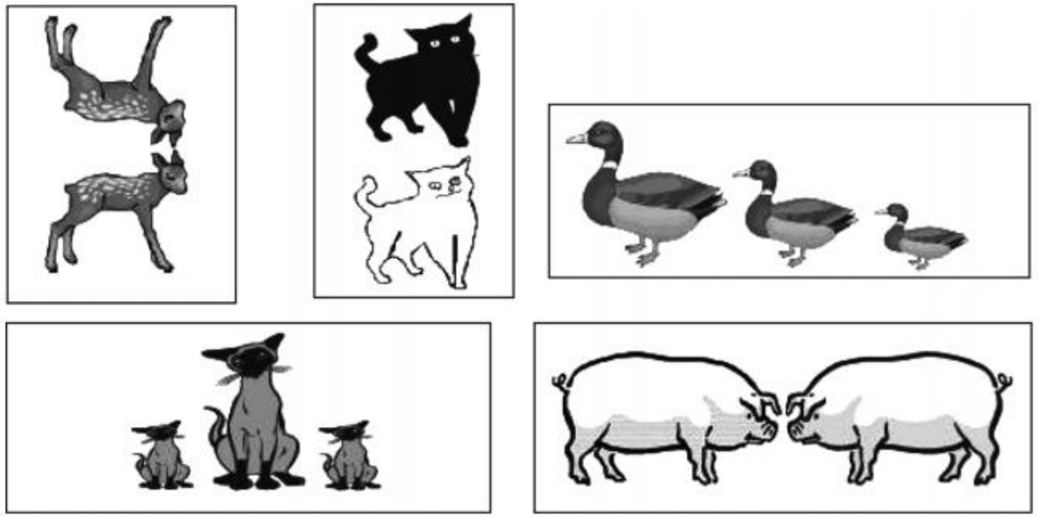
Abstract & Reasoning Corpus; ConceptARC



Mitchell et al., 2023; Moskichev et al., 2023 ; Chollet, 2019

Can Young Children Reason Analogically?





Christie & Gentner, 2010

Transformations and transfer: Preschool children infer abstract relations and reason analogically in a causal task.

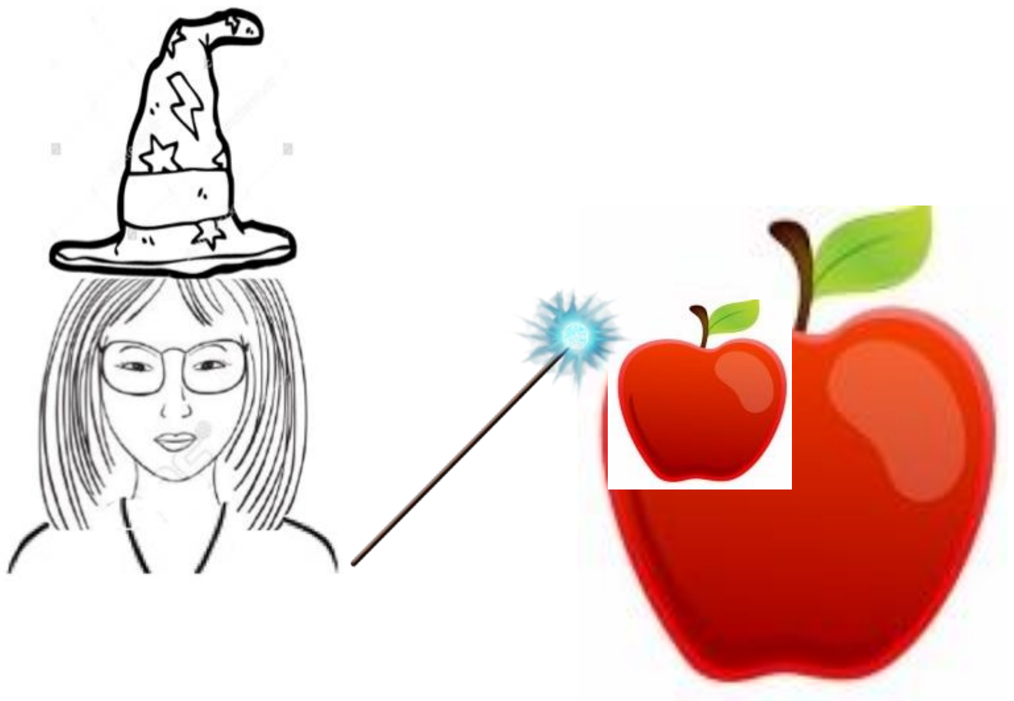
Goddu, Lombrozo, & Gopnik

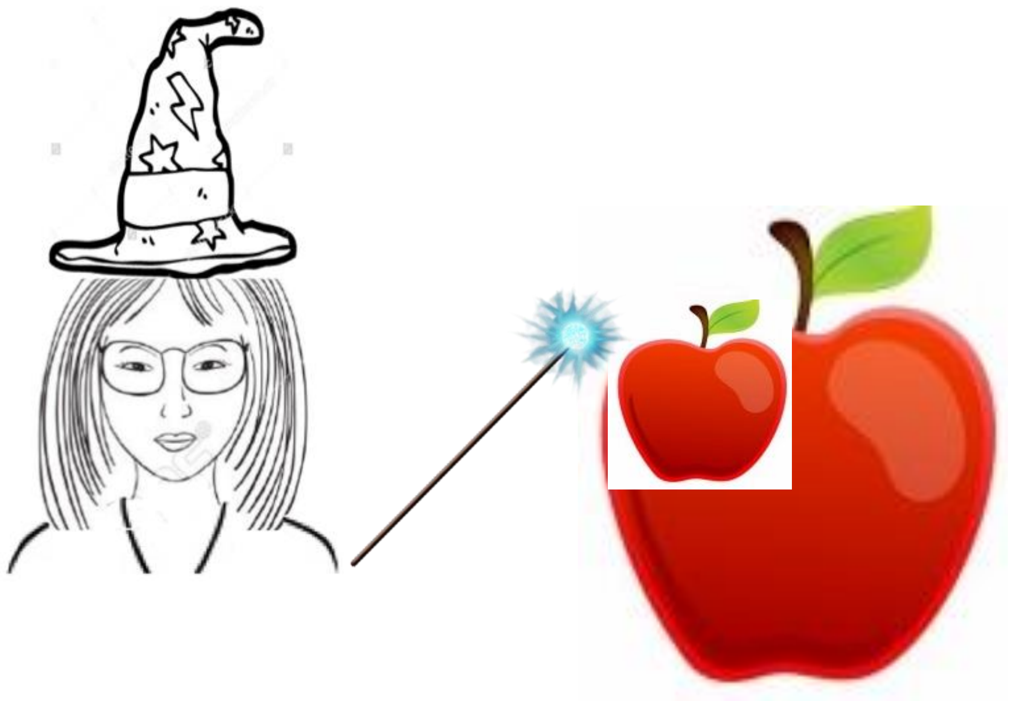
(Child Development 2020)

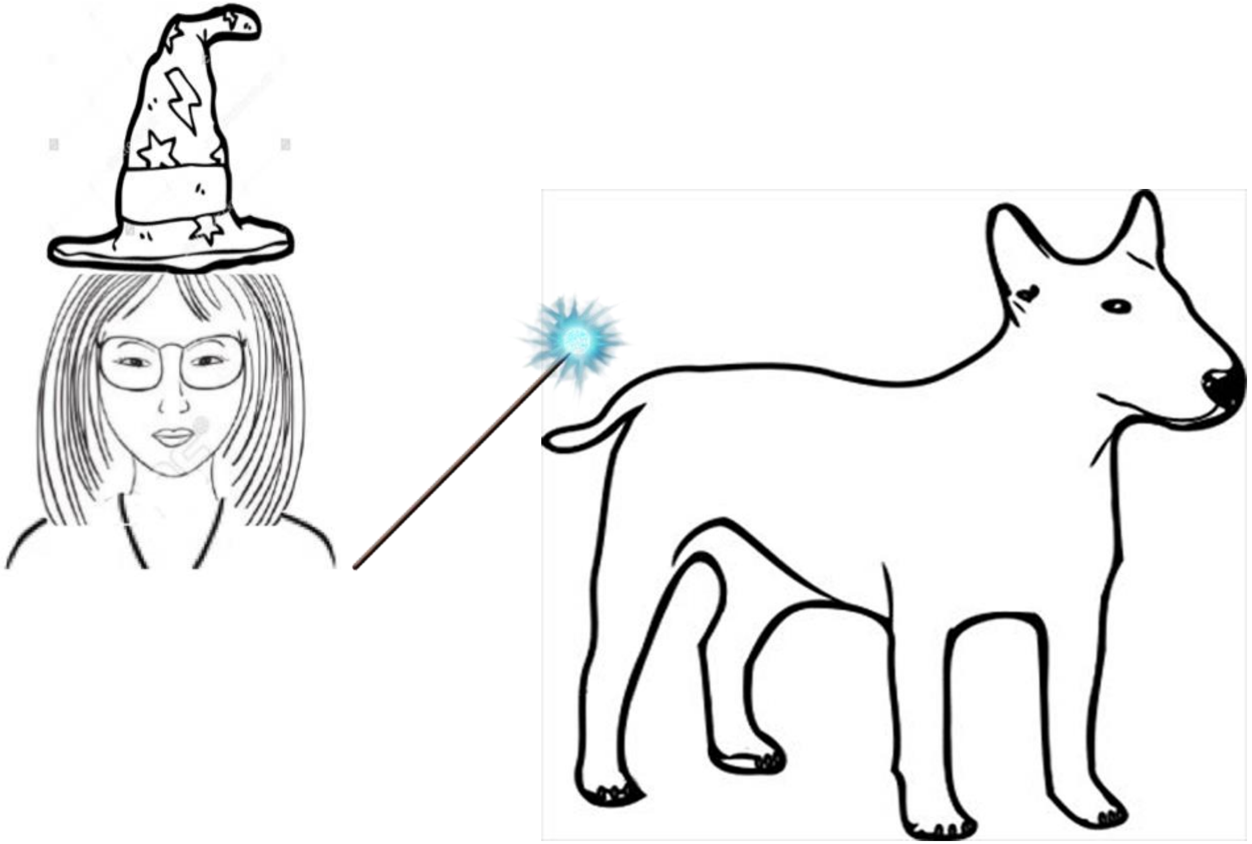
Mariel Goddu

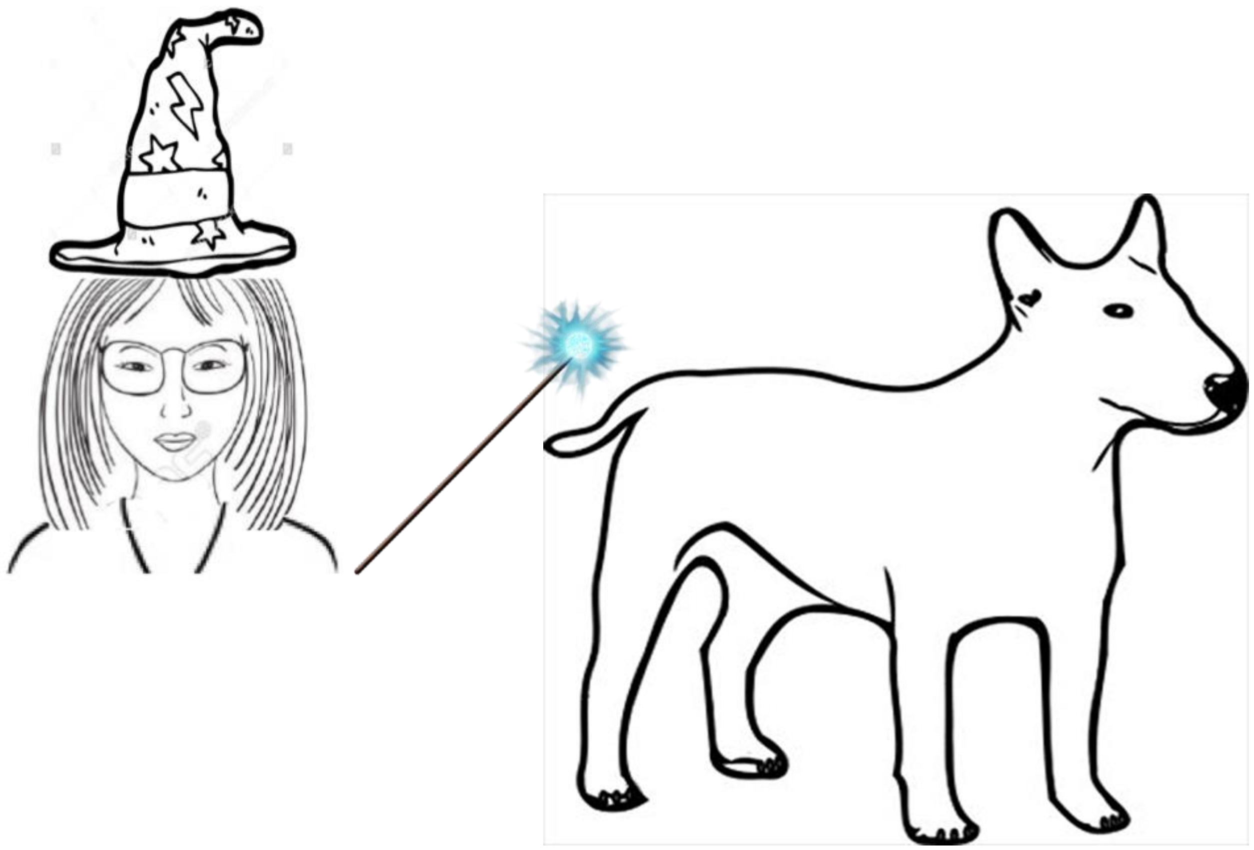
The *Wizards* Game!

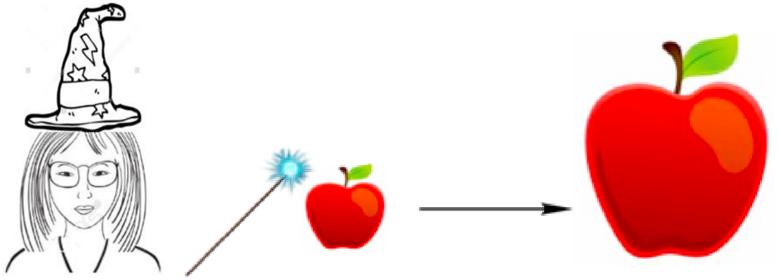
Gwen

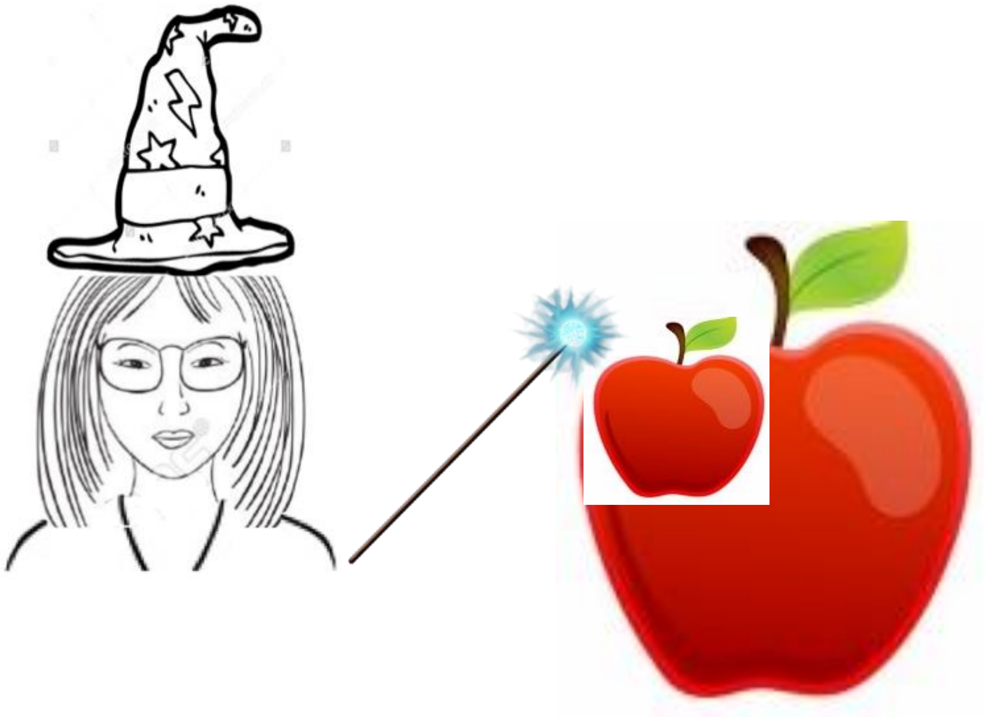


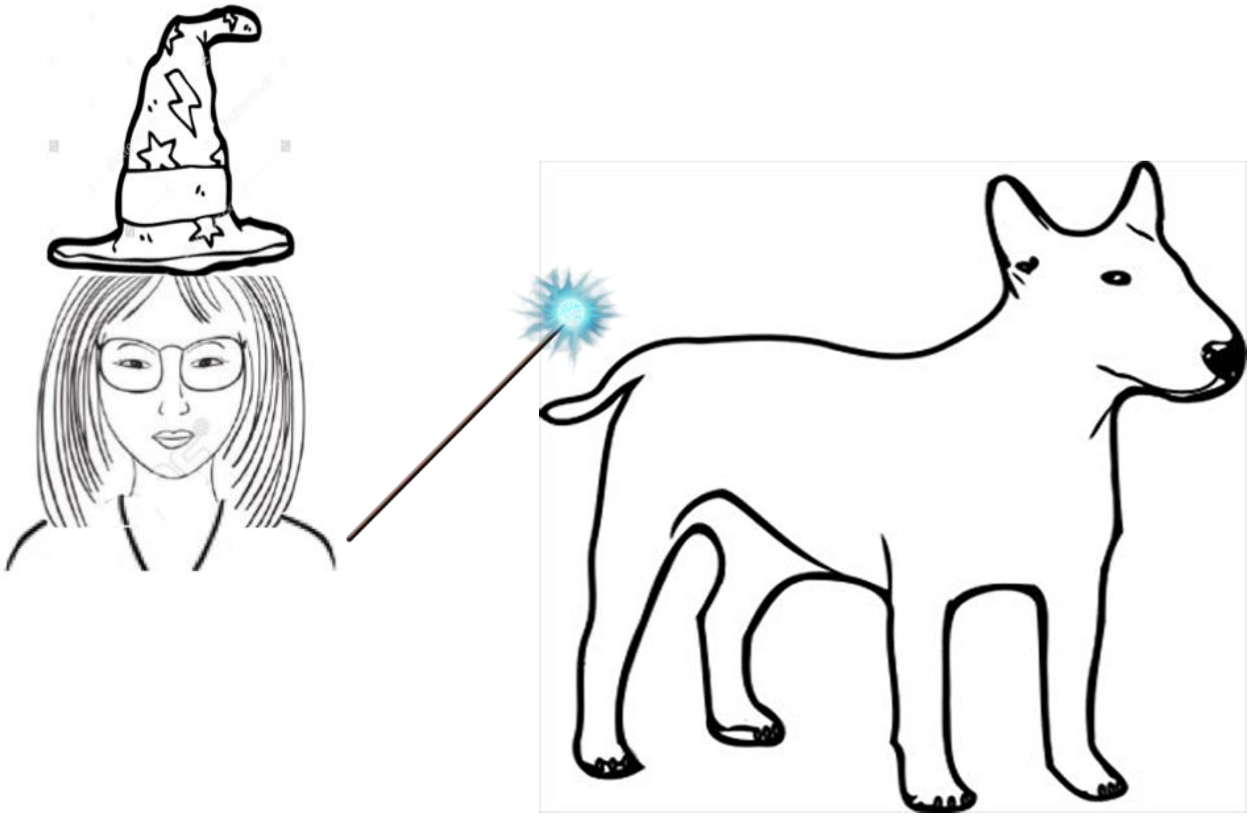


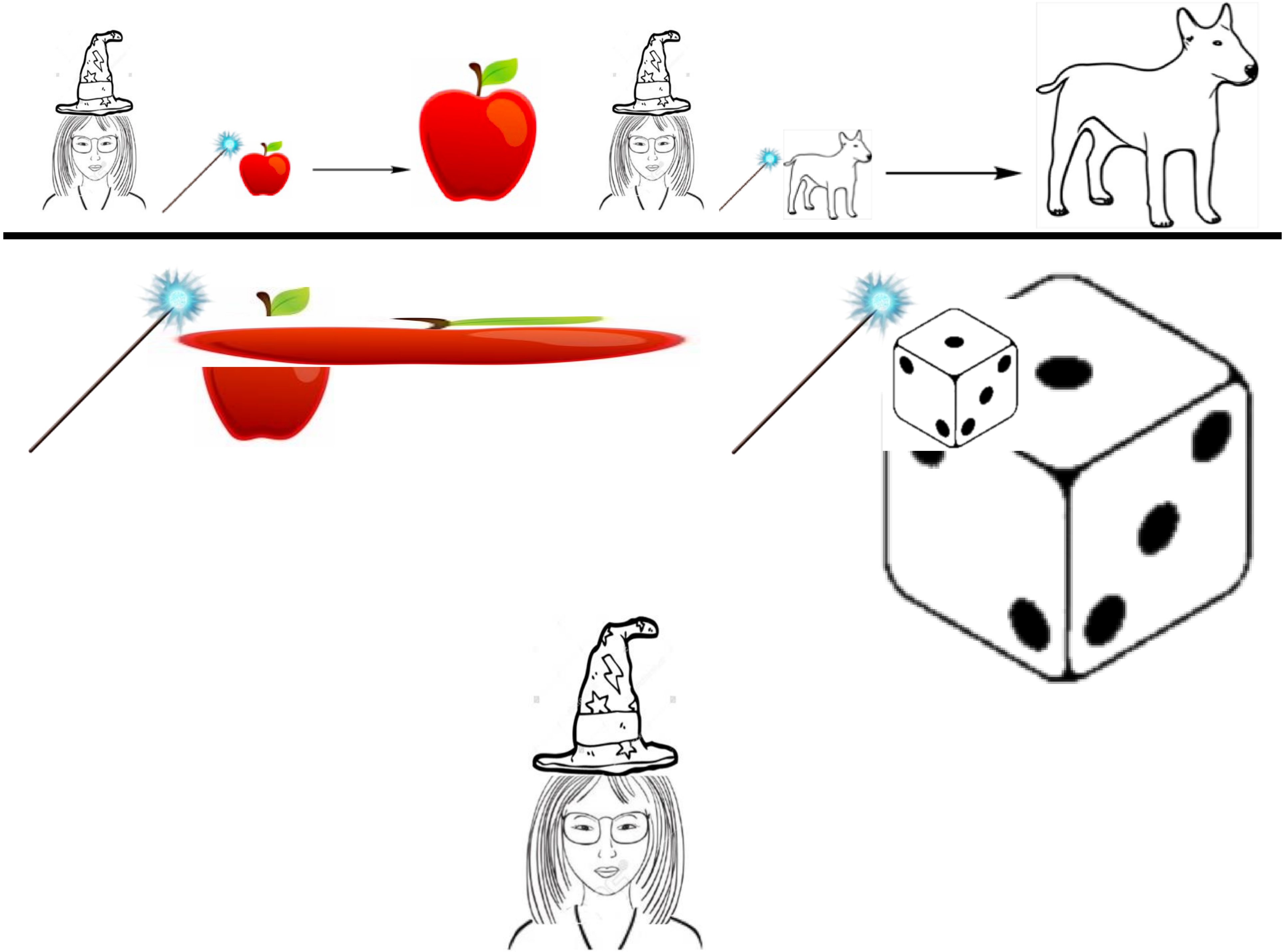












Causal relational problem solving in toddlers
Goddu, Yiu and Gopnik, Cognition, 2025

- Design:
  - **Exp. 1:** 3-year-olds learn about “change machines”; see demos showing that one machine makes objects bigger, and the other makes things smaller
  - At test, must use the machines to solve a novel problem (i.e., a monkey’s hat is ill-fitting)

Goddu & Gopnik, *CogSci 2020*

Goddu & Gopnik, *CogSci 2020*

Goddu & Gopnik, *CogSci 2020*

Goddu & Gopnik, *CogSci 2020*

Goddu & Gopnik, *CogSci 2020*

Goddu & Gopnik, *CogSci 2020*

Goddu & Gopnik, *CogSci 2020*

Goddu & Gopnik, *CogSci 2020*

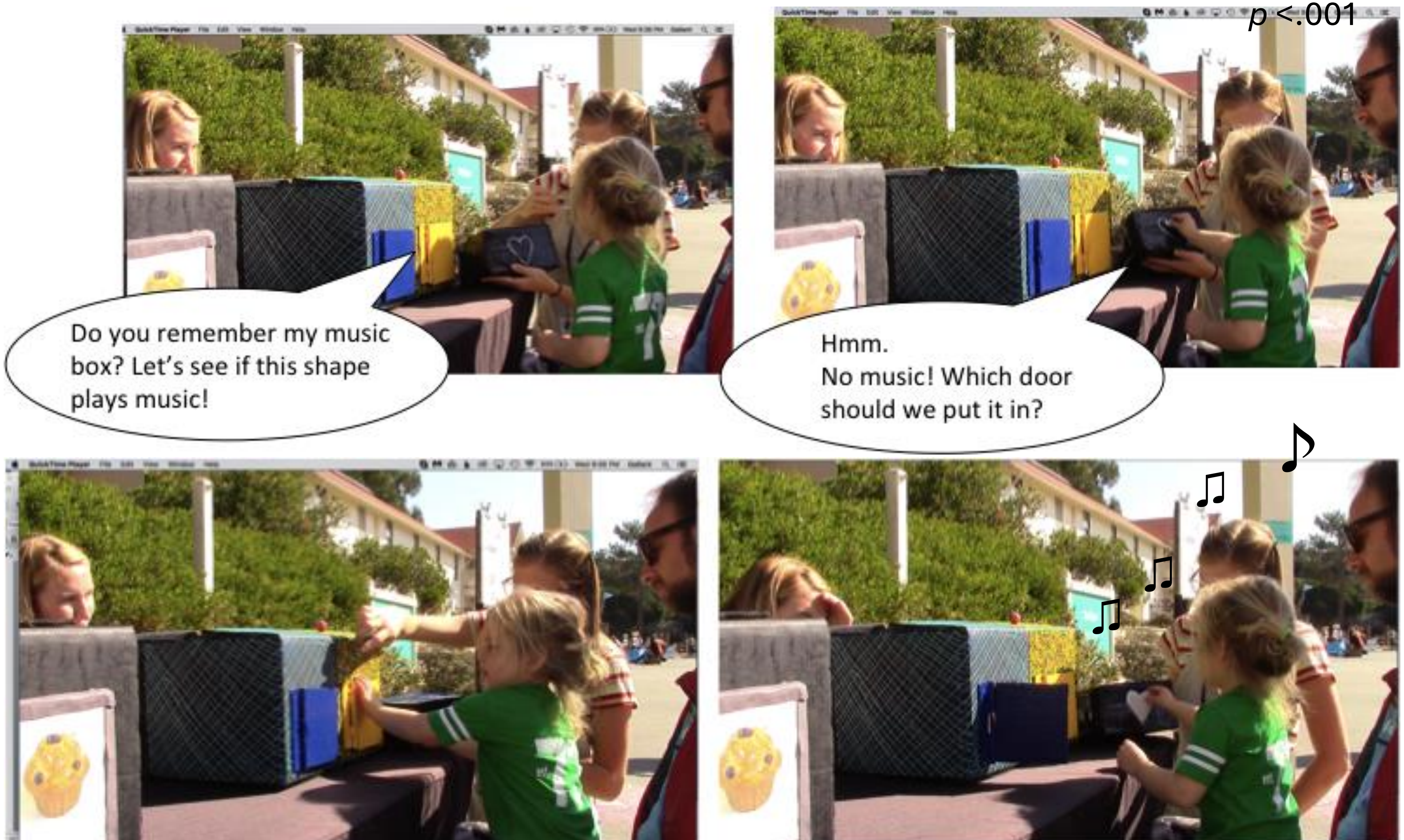
Change machines, Exp. 1

- A significant majority (72.2 %) of participants chose the appropriate machine to solve the problem (SD = 0.45, 95 % CI [56.9 %, 87.6 %]), significantly above chance (50 %), $t(35) = 2.94$, $p = 0.006$, $d = 0.49$, with no difference between performance in the “too big” and “too small”
  conditions, $t(34) = 0.73$, $p = 0.47$, $d = 0.24$. N=31
  (three-year-olds only)

ASPREDICTED

4. Test With 24-Month-Olds

N = 100, p <.001



KiVA: Kid-inspired Visual Analogies for Large Multimodal Models

Eunice Yiu, Maan Qraitem, Charlie Wong, Anisa Noor Majhi, Yutong Bai, Shiry Ginosar, Alison Gopnik, Kate Saenko

Kate Saenko

Shiry Ginosar

Alison Gopnik

Maan Qraitem

Anisa Noor Majhi

Charlie Wong

Yutong Bai

BOSTON UNIVERSITY

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

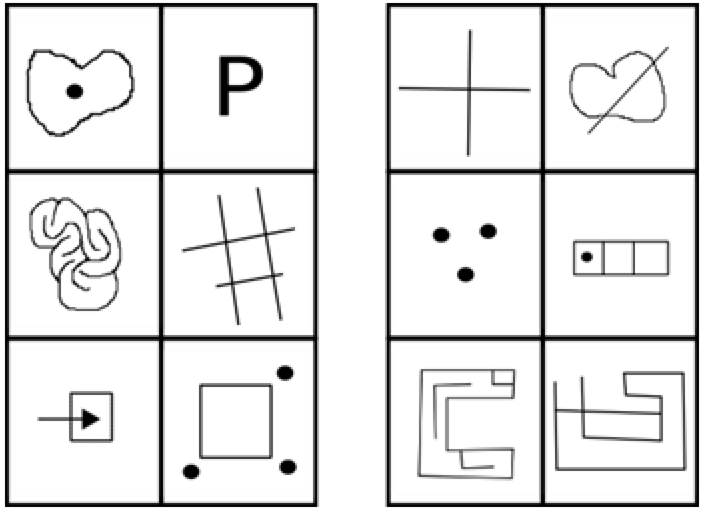
Berkeley UNIVERSITY OF CALIFORNIA Psychology

Meta AI

Funding sources: 5th Meta AI-BAIR Commons, DARPA 047498-002 Machine Common Sense, John Templeton Foundation 61475 The Development of Curiosity, Templeton World Charity Foundation TWCF0434 Play: A Computational Account, DOD ONR MURI Self-Learning Perception Through Real World Interaction

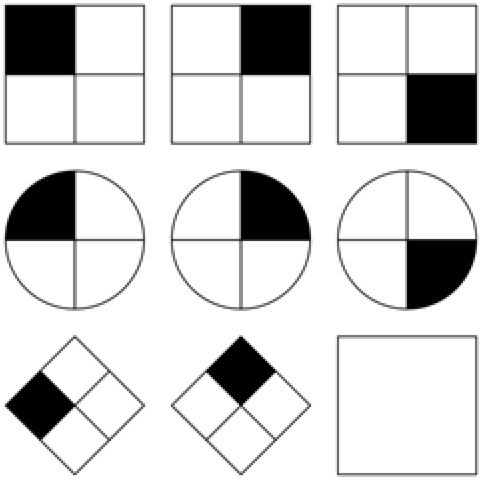
Visual Analogies out there

Bongard Problems



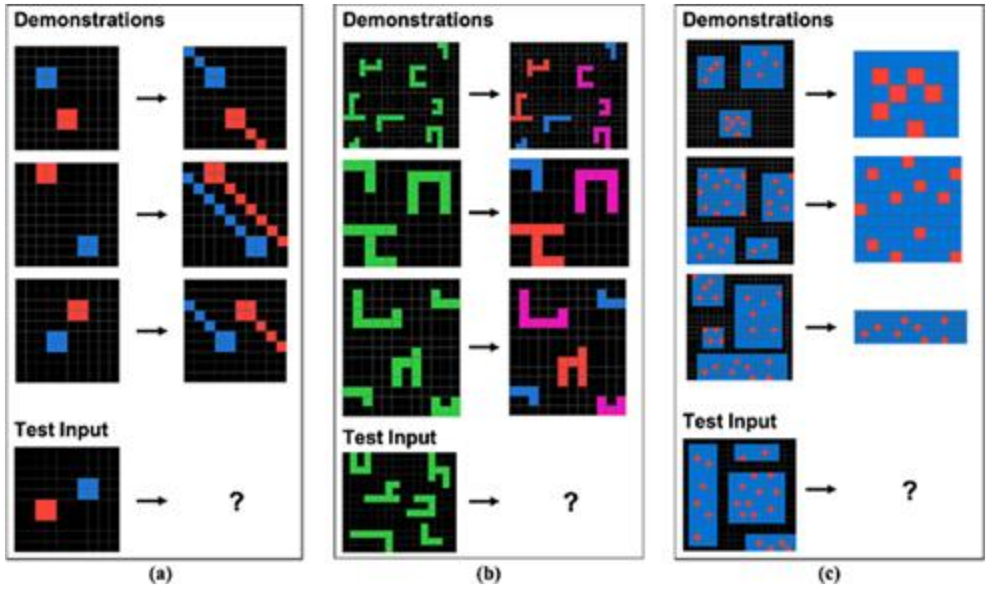
Sonwane et al., 2021; Nie et al., 2020

Raven's Progressive Matrices



Zhang et al., 2019; Barrett et al., 2018

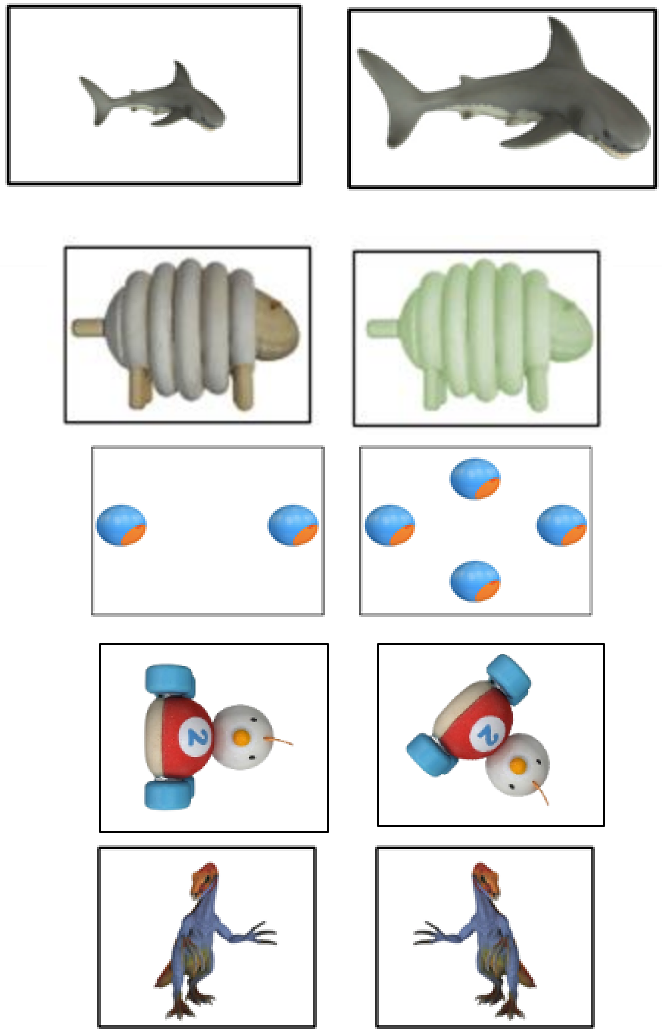
Abstract & Reasoning Corpus; ConceptARC



Mitchell et al., 2023; Moskichev et al., 2023 ; Chollet, 2019

Something even a 3-year-old can understand

- **Size** Change
- **Color** Change
- **Number** Change
- **Rotation**
- **Reflection**



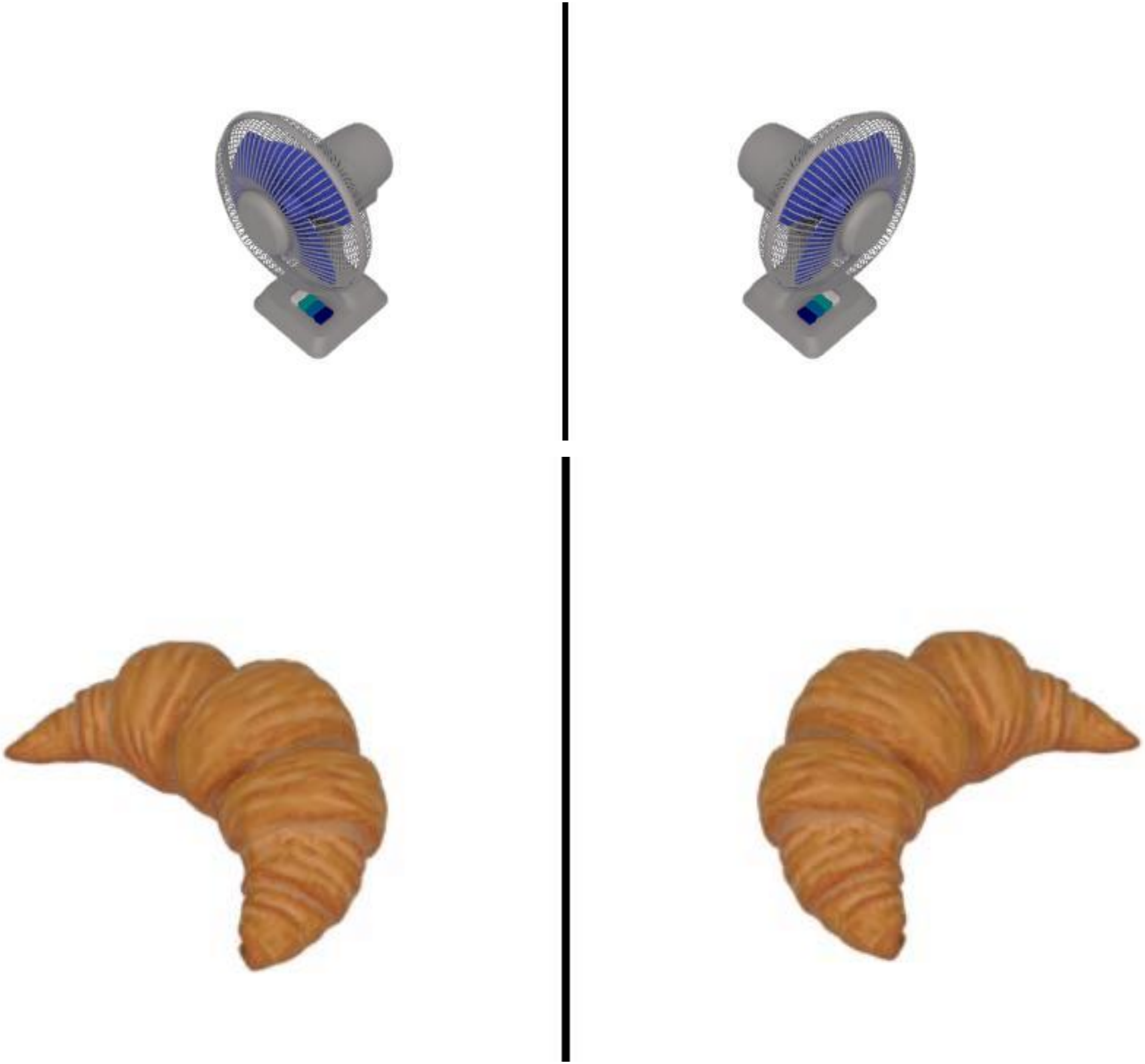
Goddu et al. (2020)

Benchmark curated from transforming common, child-friendly 3D objects

Google Scanned Objects (Downs et al., 2022)

Toys4K (Stojanov et al., 2021)

Our benchmark consists of 3000+ transformation stimuli

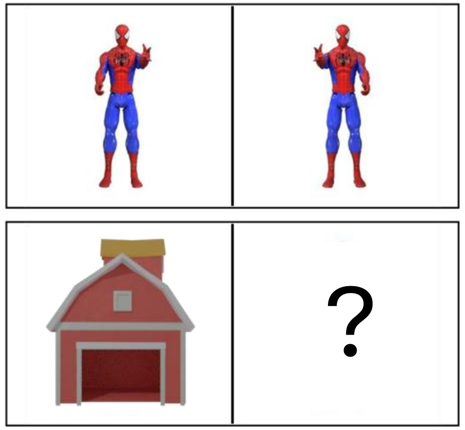


Participants

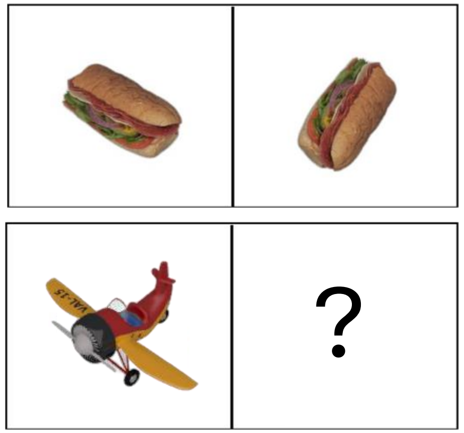
- 250 adults (21 to 35 years old)
  - Each trial in the benchmark is annotated by >= 3 adults
- 50 children (3-7 years old, mean: 4.43, SD = 1.12)
  - Healthy vision, no colorblindness
  - Completed a random subset of trials sampled across the 5 domains
- GPT-4V, LLaVA-1.5 & MANTIS (Interleaved Multi-Image Instruction Tuning)
  - Each trial is repeated queried 3 times and the scores are averaged
  - Completed all trials across the 5 transformation domains

Transformations examined

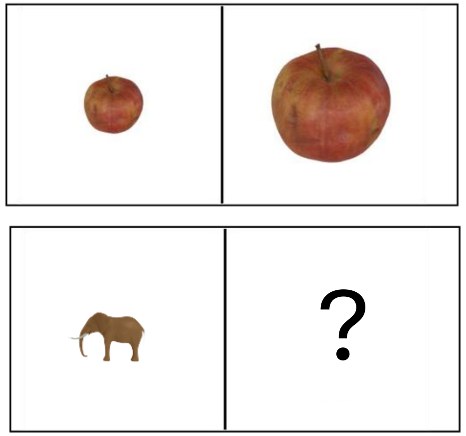
Reflection



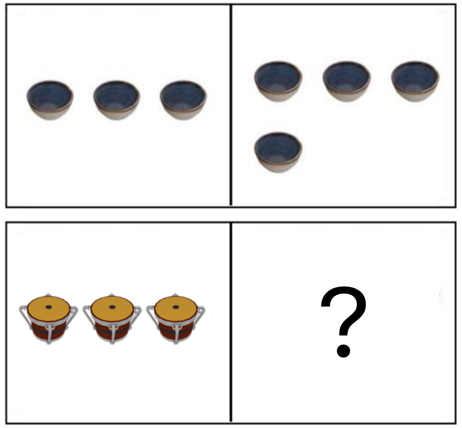
Rotation



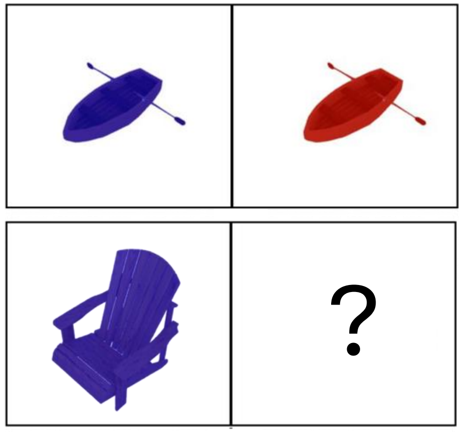
Size



Number



Color

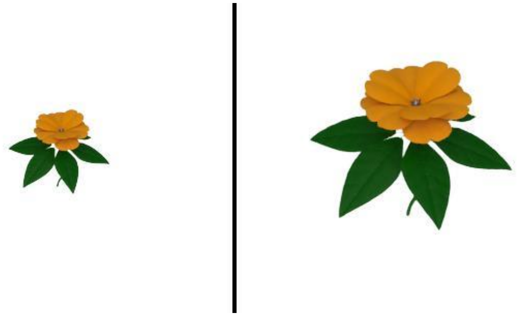


automated, step-by-step evaluation process

System message:

You are an **excellent visual puzzle solver**! You will be given a visual puzzle that requires using visual analogical reasoning. You will think *step-by-step* and carefully examine the visual evidence before providing an answer. The puzzle features a left-to-right transformation of an object. The transformation involves a change in *the size, orientation, number, or color of an object*. Note that each transformation consists of a left image (the starting position) and a right image (the ending position).

Verbal Classification:

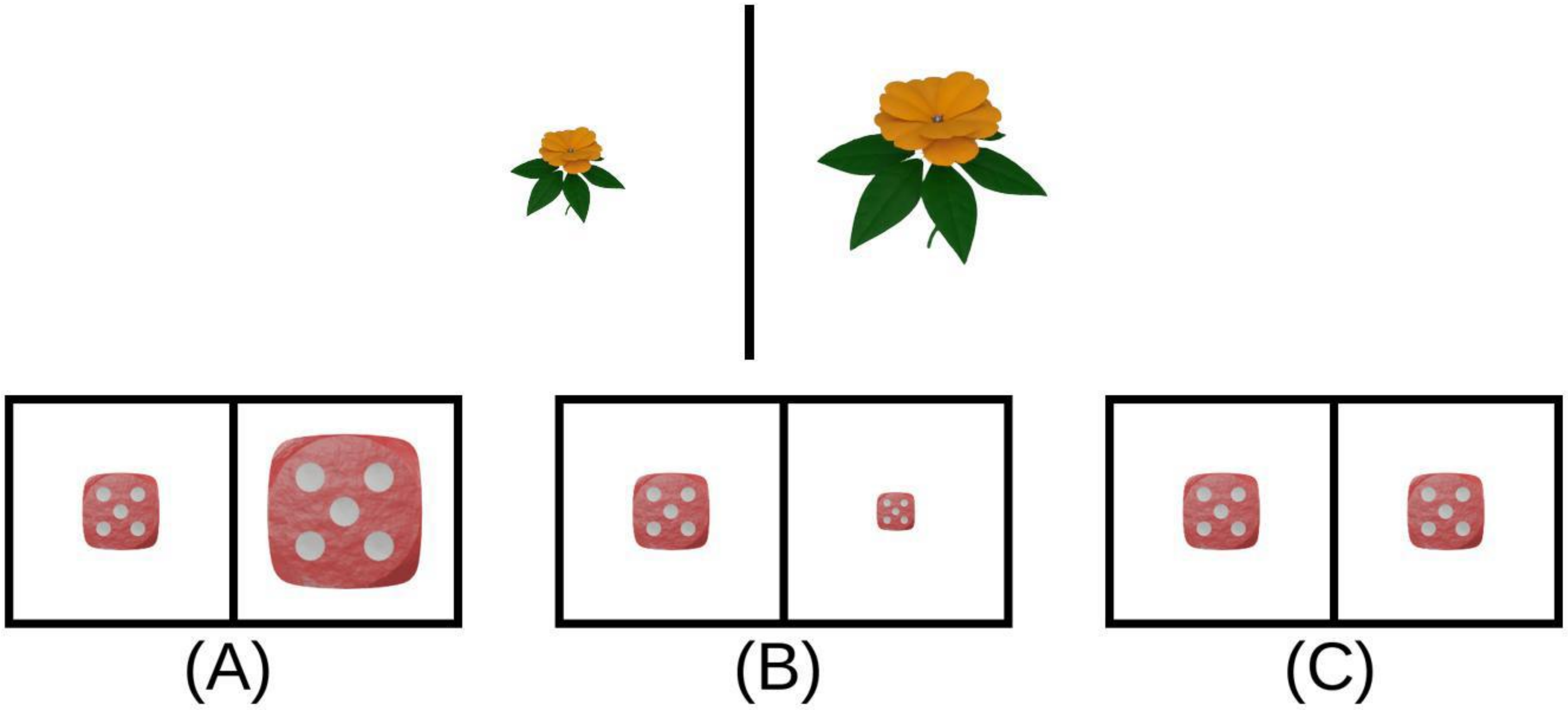


(1) Change in **color** of objects
(2) Change in **size** of objects
(3) Change in **number** of objects
(4) No change
(5) Doesn't apply

(if Verbal Classification is correct) **Verbal Specification**:

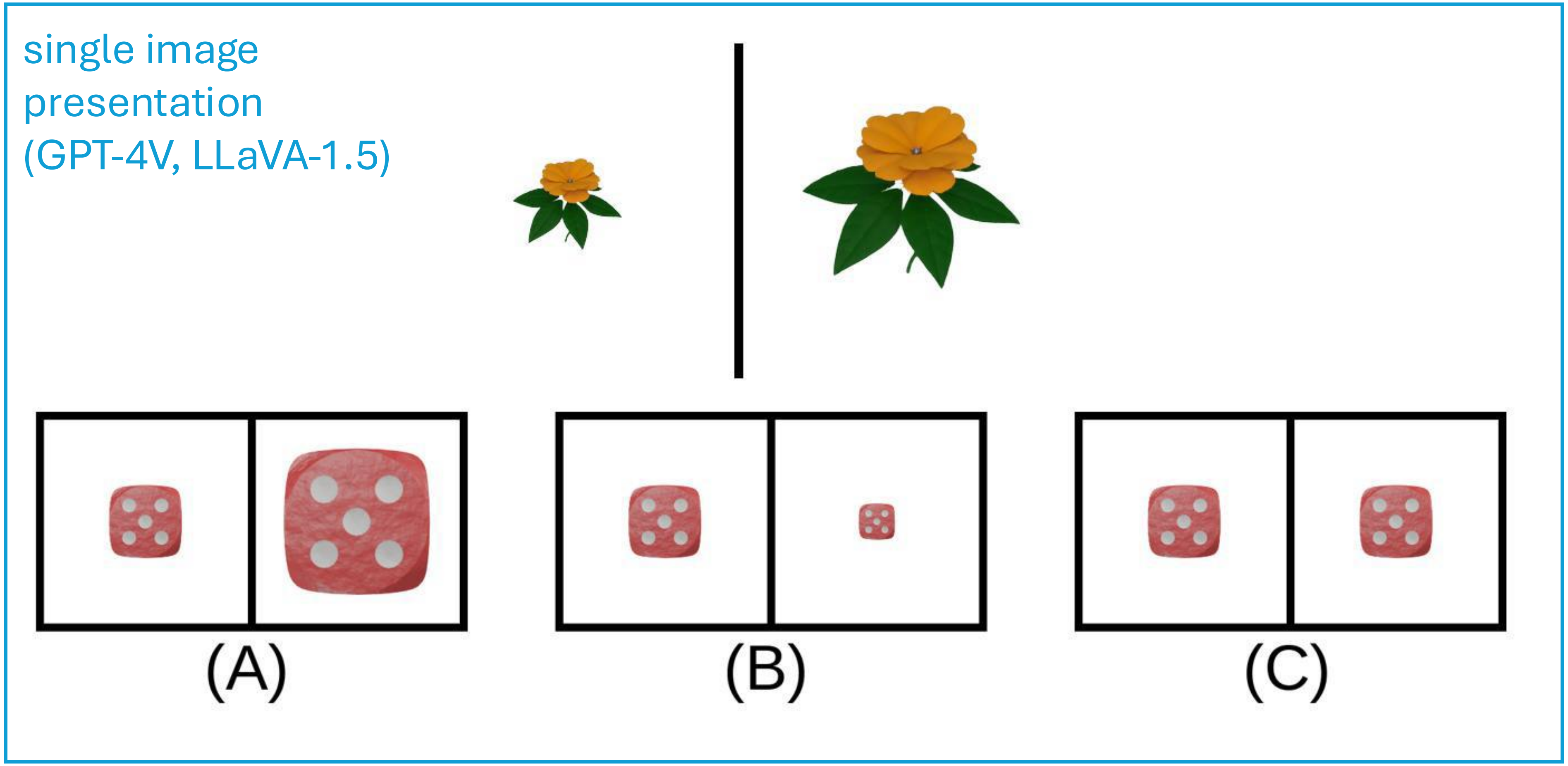
(1) Objects become **bigger**
(2) Objects become **smaller**
(3) No change
(4) Doesn't apply

Visual Extrapolation:



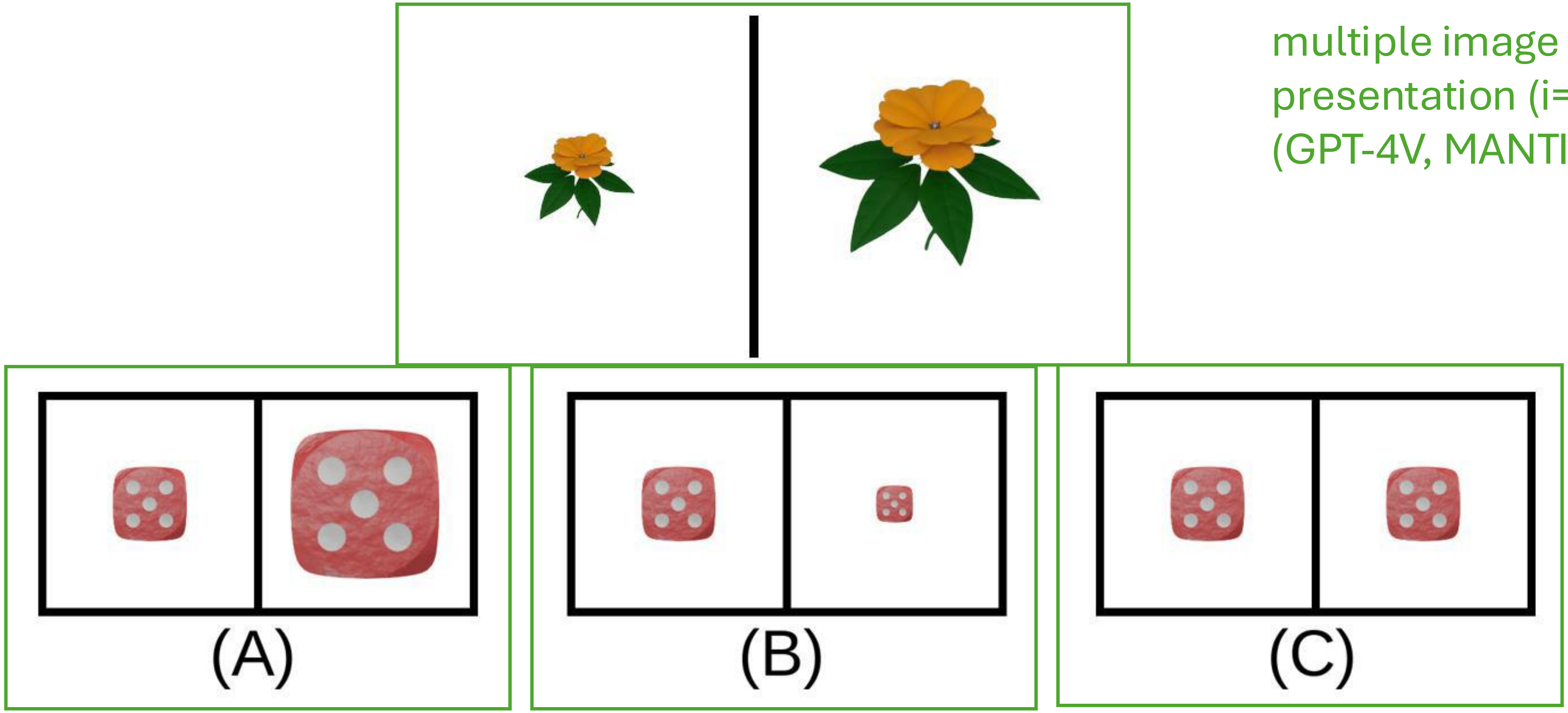
(D) Doesn’t apply

Visual Extrapolation:



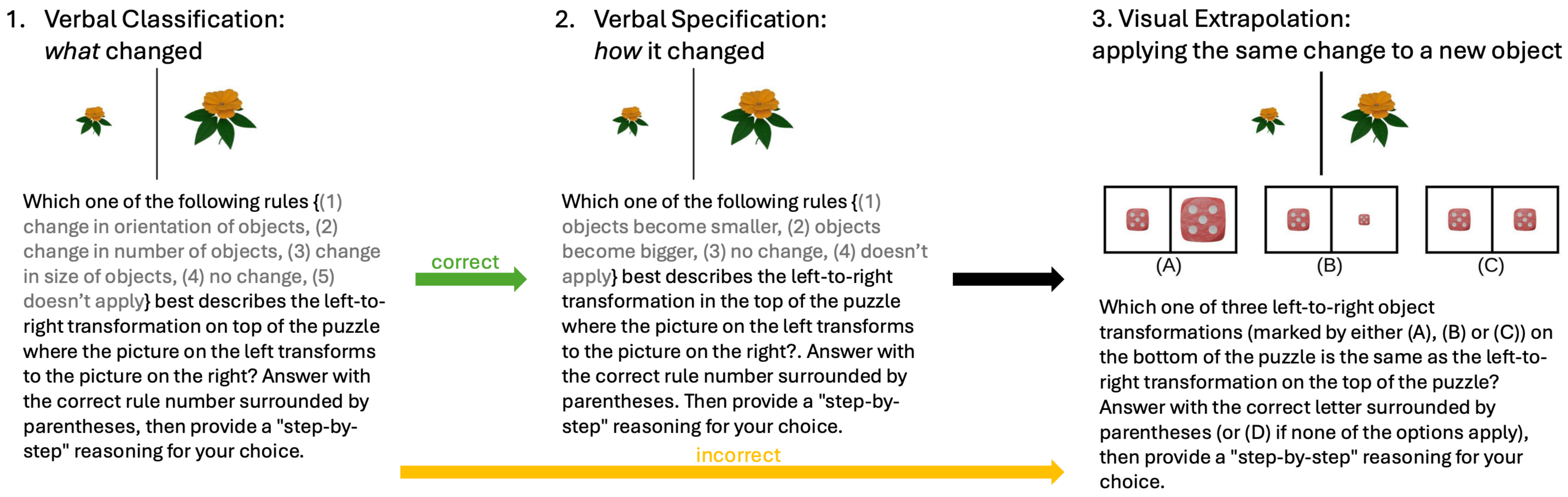
(D) Doesn’t apply

Visual Extrapolation:



multiple image
presentation (i=4)
(GPT-4V, MANTIS)

(D) Doesn’t apply


1. Verbal Classification:
what changed
Which one of the following rules {(1) change in orientation of objects, (2) change in number of objects, (3) change in size of objects, (4) no change, (5) doesn't apply} best describes the left-to-right transformation on top of the puzzle where the picture on the left transforms to the picture on the right? Answer with the correct rule number surrounded by parentheses, then provide a "step-by-step" reasoning for your choice.
correct
2. Verbal Specification:
how it changed
Which one of the following rules {(1) objects become smaller, (2) objects become bigger, (3) no change, (4) doesn't apply} best describes the left-to-right transformation in the top of the puzzle where the picture on the left transforms to the picture on the right?. Answer with the correct rule number surrounded by parentheses. Then provide a "step-by-step" reasoning for your choice.
incorrect
3. Visual Extrapolation:
applying the same change to a new object
(A)
(B)
(C)
Which one of three left-to-right object transformations (marked by either (A), (B) or (C)) on the bottom of the puzzle is the same as the left-to-right transformation on the top of the puzzle? Answer with the correct letter surrounded by parentheses (or (D) if none of the options apply), then provide a "step-by-step" reasoning for your choice.

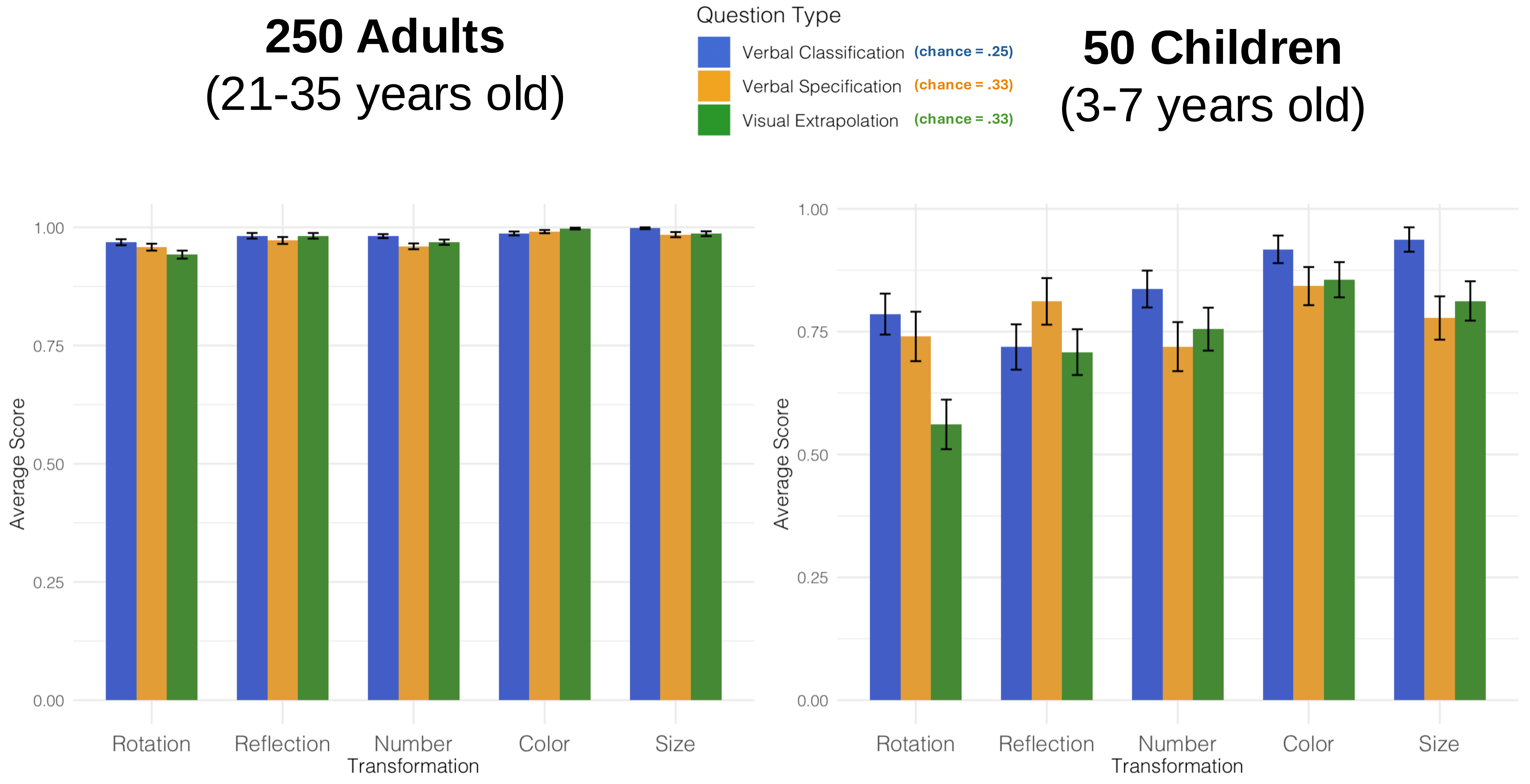
Prompt for children

You are on a mission as a picture detective. You will see how different pictures change.
Your job as a picture detective is to figure out how the pictures change, and to guess how a new picture would change based on that.

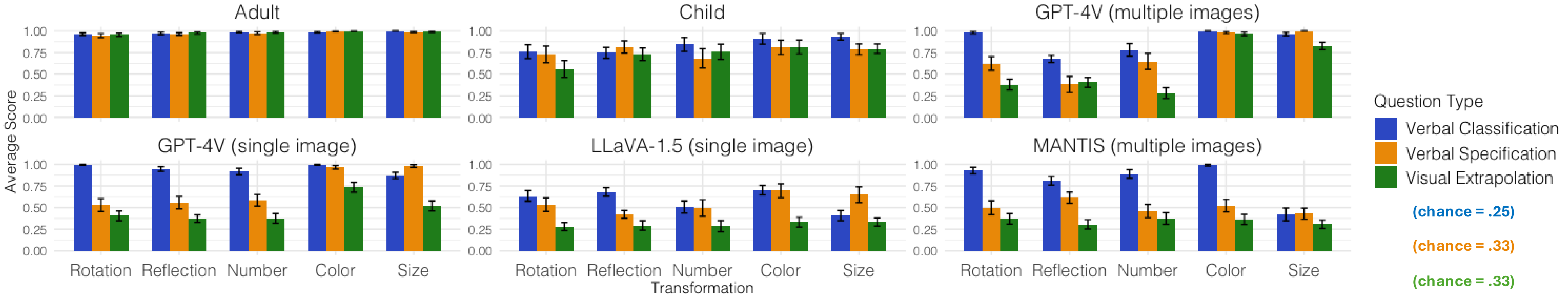
These pictures can change in size, where they face, number, or color. Every time you answer correctly, you will get a coin.

You won’t find out how many coins you get until the end of the game. At the end of the game, you will see the total number of coins you win.

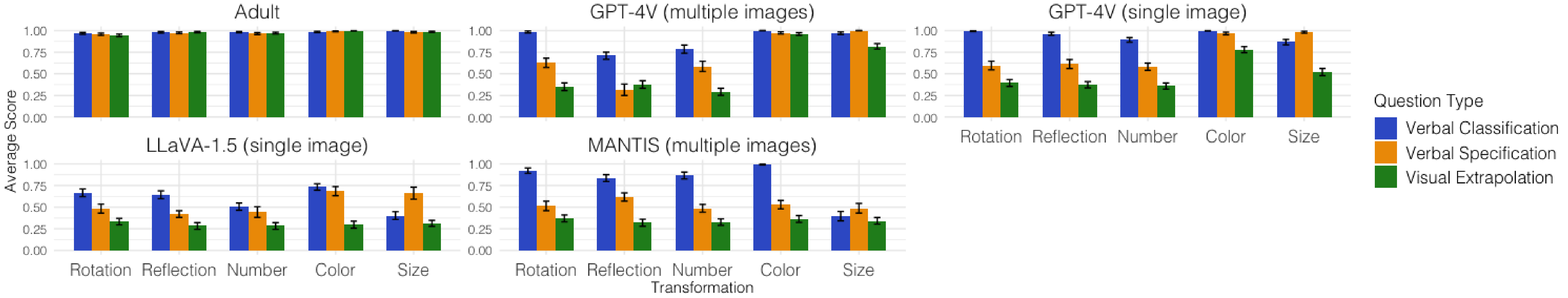
The more coins you get, the more stickers you win.


250 Adults
(21-35 years old)
Question Type
Verbal Classification (chance = .25)
Verbal Specification (chance = .33)
Visual Extrapolation (chance = .33)
50 Children
(3-7 years old)
Average Score
1.00
0.75
0.50
0.25
0.00
Rotation
Reflection
Number Transformation
Color
Size

subset of benchmark completed by 50 child participants

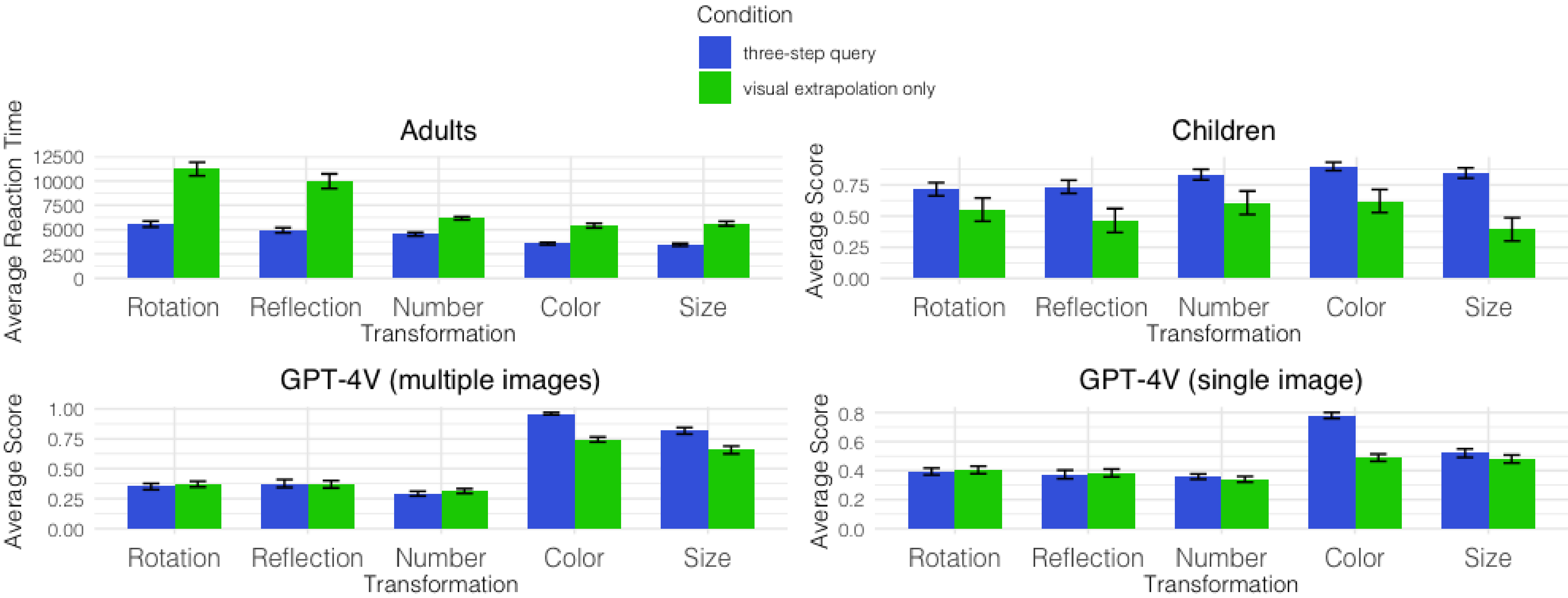


similar pattern on the full benchmark, excluding child participants



What if we just test directly on visual extrapolation?

We recruited another 200 adults, 20 children and the best performing model, GPT-4V, on a visual-extrapolation-only task



Verbal questions help!

Can we make the models better at visual analogical reasoning?

What about presenting more relevant training data?

What about prompting in different ways?

#1 Alternative Text Prompting:

- Instruction Prompting (Yang et al., 2024; Zhang et al., 2023)

  Write instructions on how to establish if a transformation involves an object that rotates 90 degrees or 180 degrees. Now using the instructions from before...

- Inference through Code (Sharma et al., 2024)

  Generate python code using the package pillow that takes in the left image in the left-to-right transformation and outputs the right image. Denote this snippet as training snippet. Now using the insights from the training code snippet...

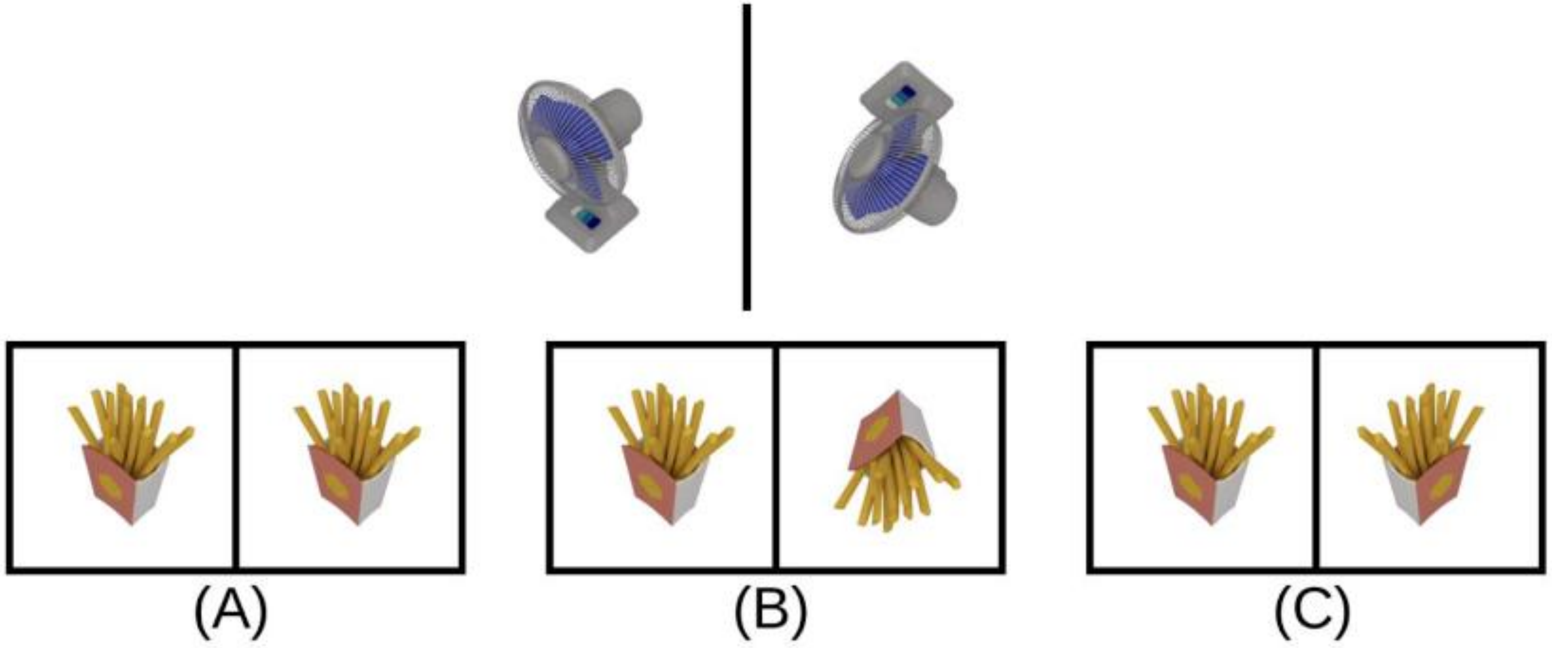
- Self Reflect (Li et al., 2024)

  Please reflect on your answer and provide a revised response if necessary. Start your response with your updated answer (xN)
  N = the number of times the model is instructed to self reflect)

No improvement!

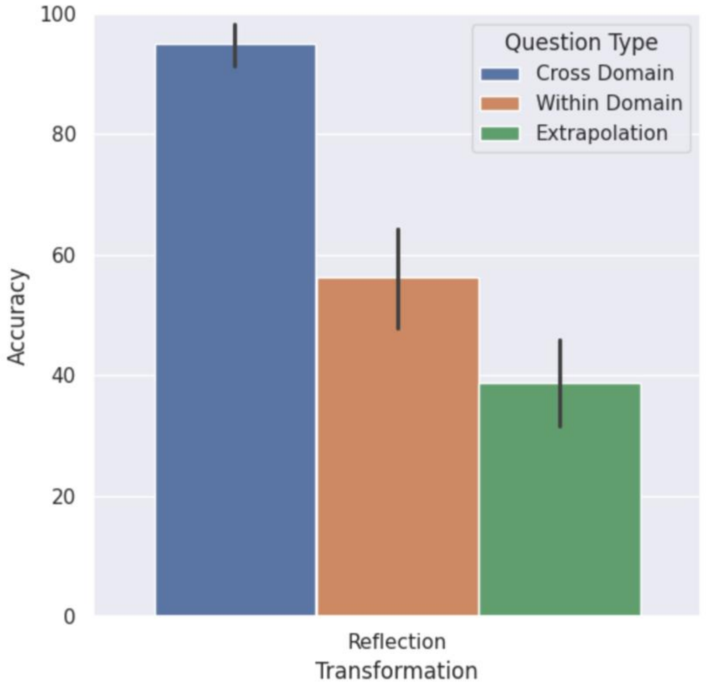
#2 In-context Learning:

- Provide 6-10 example puzzles from the training set along with an explanation

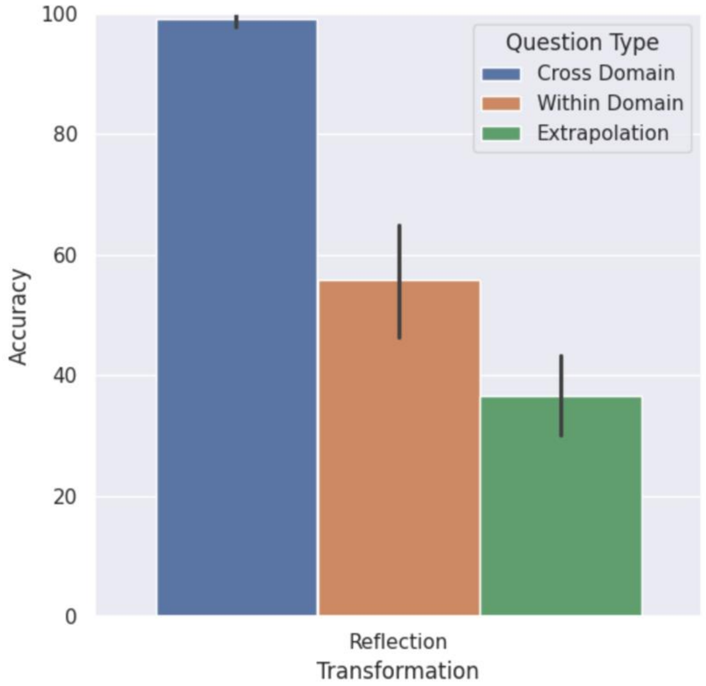


Solution is (B) because the object flips upside down.

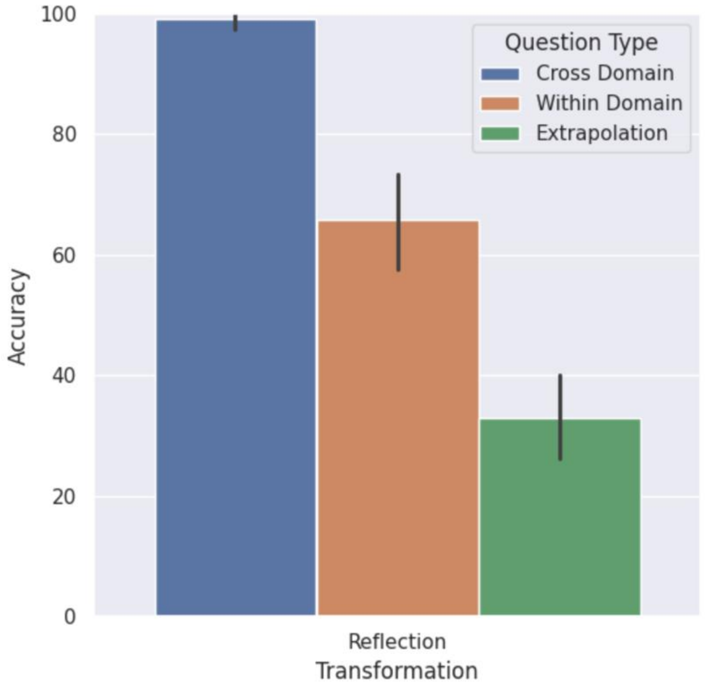
Did it improve results?



0 in-context examples



6 in-context examples



10 in-context examples

#3 Visual Prompting

- Changing background color
- Cropping, changing object scale

No improvement!

Summary

- Large multimodal models cannot perform visual analogical reasoning as well as a three-year-old can
  - sometimes it fails to recognize the specific visual details
  - other times it fails to extrapolate faithfully even after identifying correct visual details

- It may take more than simple in-context learning and “prompt engineering” to improve visual analogical reasoning

Ongoing Questions

- What does take for large multimodal models to do visual analogical reasoning?
  - Is it a limitation of the current visual encoder, or is there more to it?
  - Will generative vs. discriminative models make a difference?

- What does it take for a child to do visual analogical reasoning?
  - Does it come from encountering data for structural alignment (e.g., Gentner & Markman, 1997), or is it facilitated by their active predictions and interventions in the world (e.g., Goddu et al., 2020)?

- Will training multimodal/vision models on a "developmental curriculum" of KiVA enable them to acquire adult-like analogical reasoning capabilities?

- What about evaluating visual analogical reasoning in videos (e.g., Patraucean et al., 2024)?
  - Does temporality make it easier or harder?

Thank you for listening!

email: ey242@berkeley.edu
personal webpage: https://ey242.github.io


Alison Gopnik


Shiry Ginosar


Kate Saenko


Maan Qraitem


Yutong Bai


Anisa Noor Majhi


Charlie Wong

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Berkeley UNIVERSITY OF CALIFORNIA Psychology

BOSTON UNIVERSITY

Meta AI

Participants and families of participants, Lawrence Hall of Science, Children's Creativity Museum, Clark-Kerr Preschool, Berkeley Early Childhood Center, ChildrenHelpingScience
Funding sources: DOD ONR MURI Self-Learning Perception Through Real World Interaction, 5th Meta AI-BAIR Commons