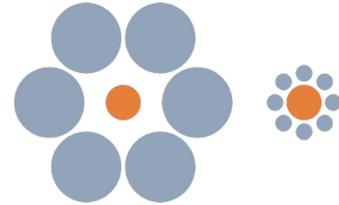# Question:
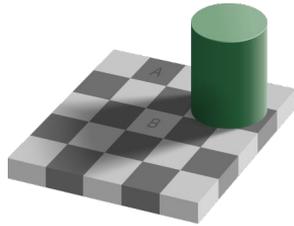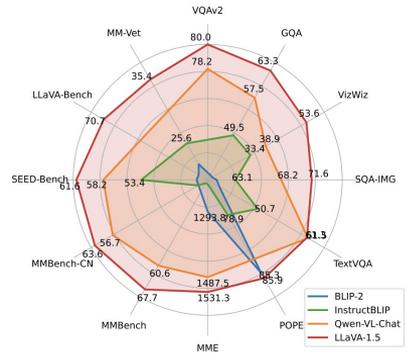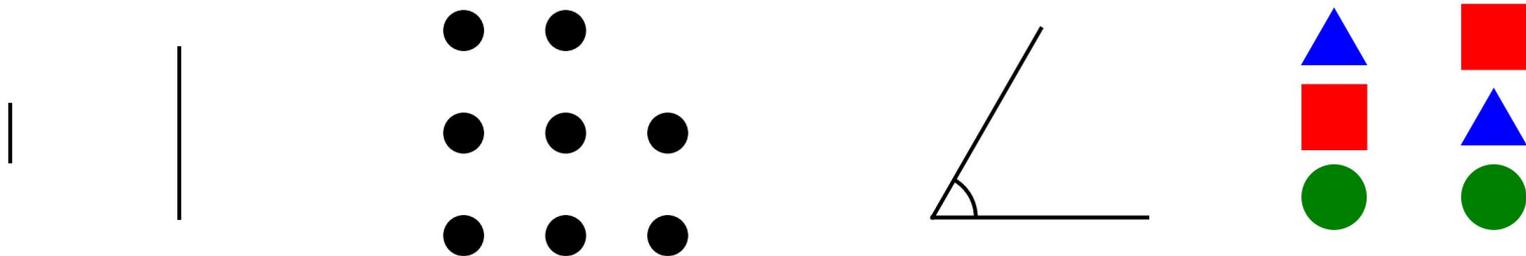
~~Do Large Multimodal Models fall for Optical Illusions?~~
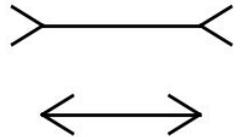
What does this tell us about how their visual perception compares to humans'?

Question:
Do Large Multimodal Models Possess Human-Like Geometrical Intuition?

Humans and LMMs both see similar natural images during cognitive development/pretraining. Are the underlying abstractions they form for solving VQA benchmarks the same as humans'?

Method: Prompt LMMs to answer questions about geometrical properties of simple synthetic images.

- 2 models: LLaVA-1.5 7B and InstructBLIP-2 7B
- 13 datasets testing size-comparison, counting, angle-identification and symmetry-identification.
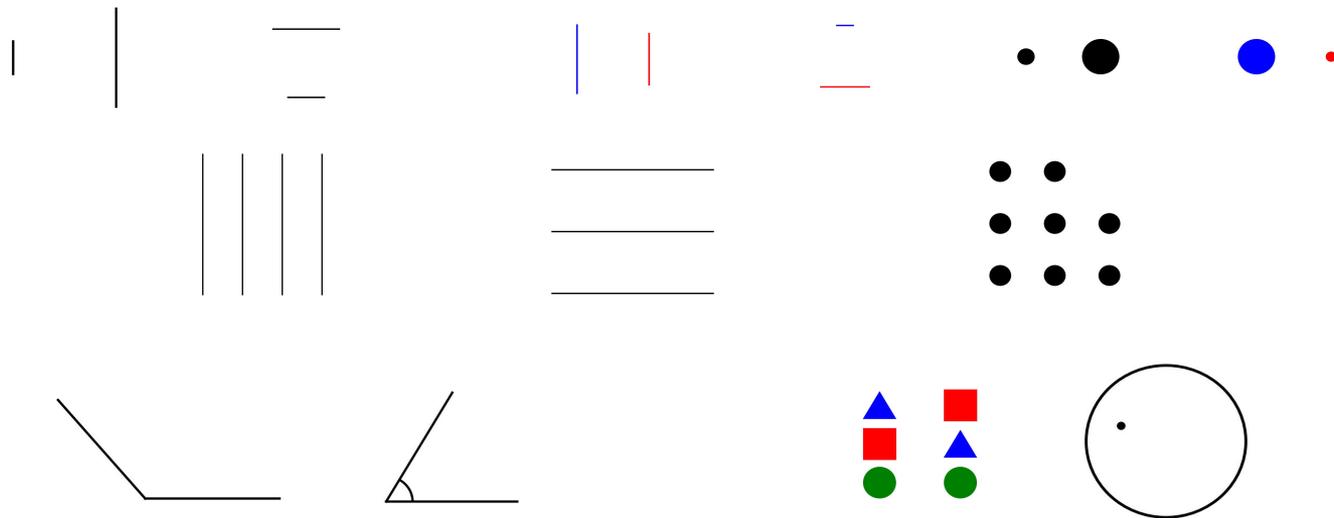
Results: No. But there some interesting emergent abilities.

1. Usually fail at size-comparison.
   a. But show ability to tell if one size is much bigger than another (without retaining identity of the larger entity).

2. Can perform approximate counting.

3. Cannot detect symmetry
   a. But show an ability to distinguish the interior of a shape from its exterior.