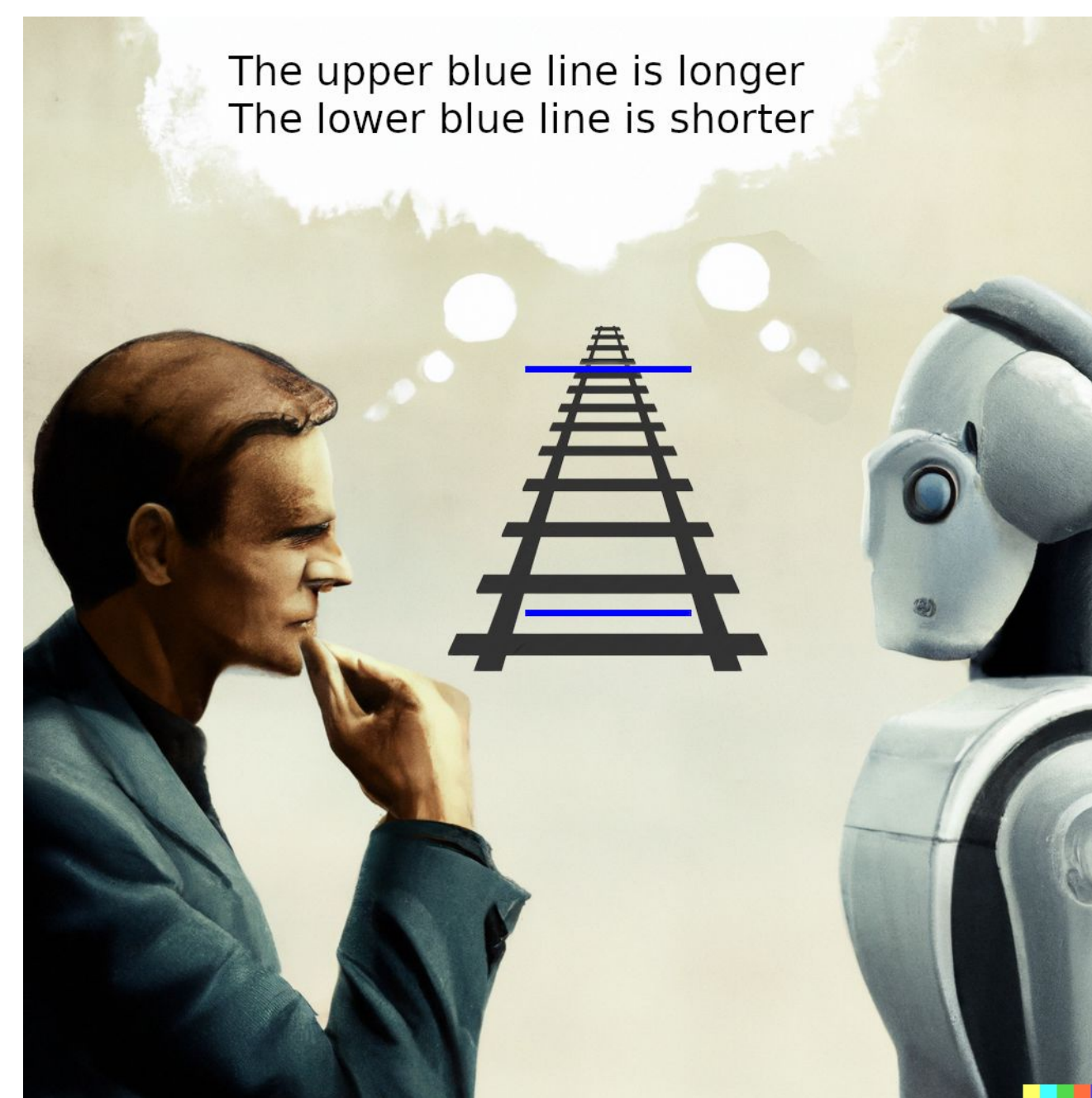


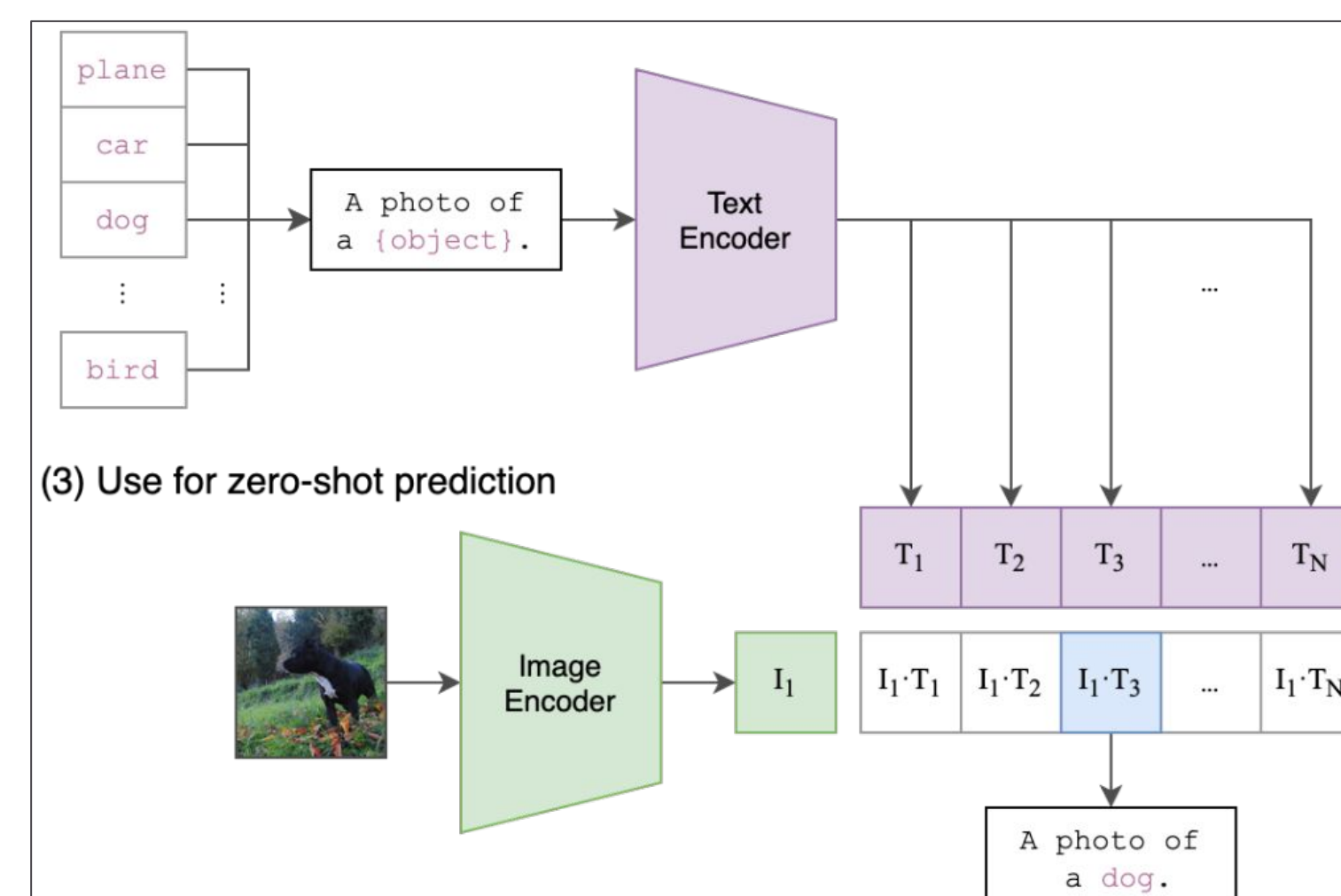
Introduction

- Recent big machine learning shows shared similarities with human perception
- Optical illusions are the product of human biology, learning, and perception
- Do machine learning models get fooled by optical illusion?**

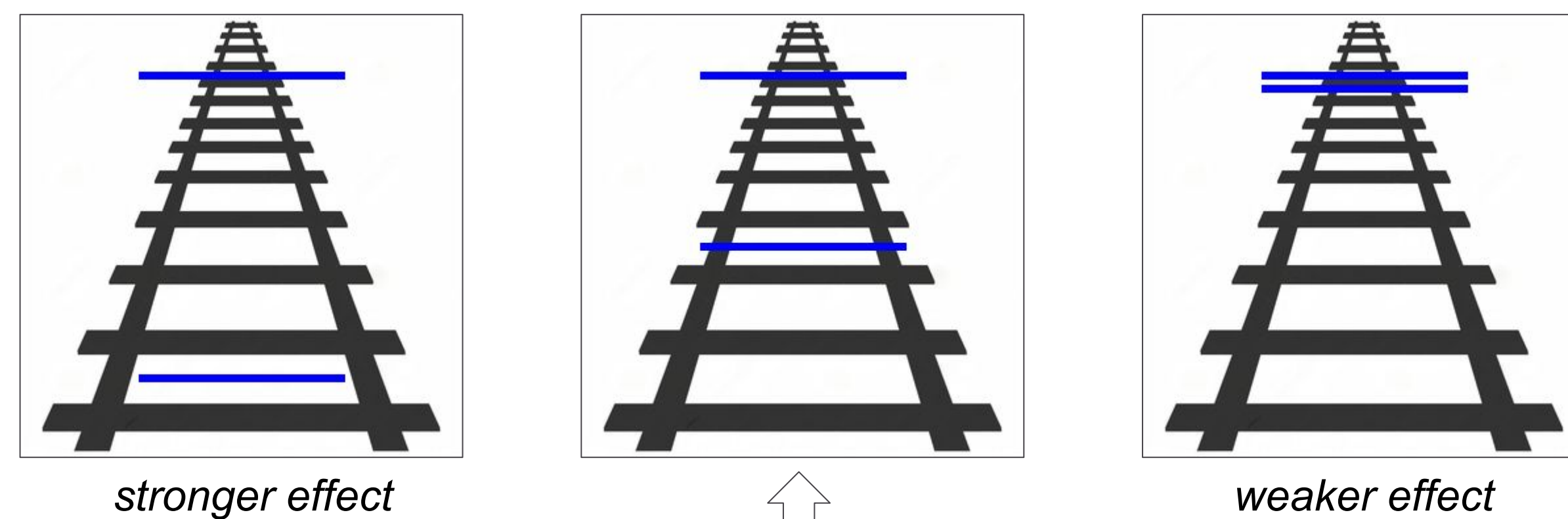


Methods

- We measure the effect on the CLIP models



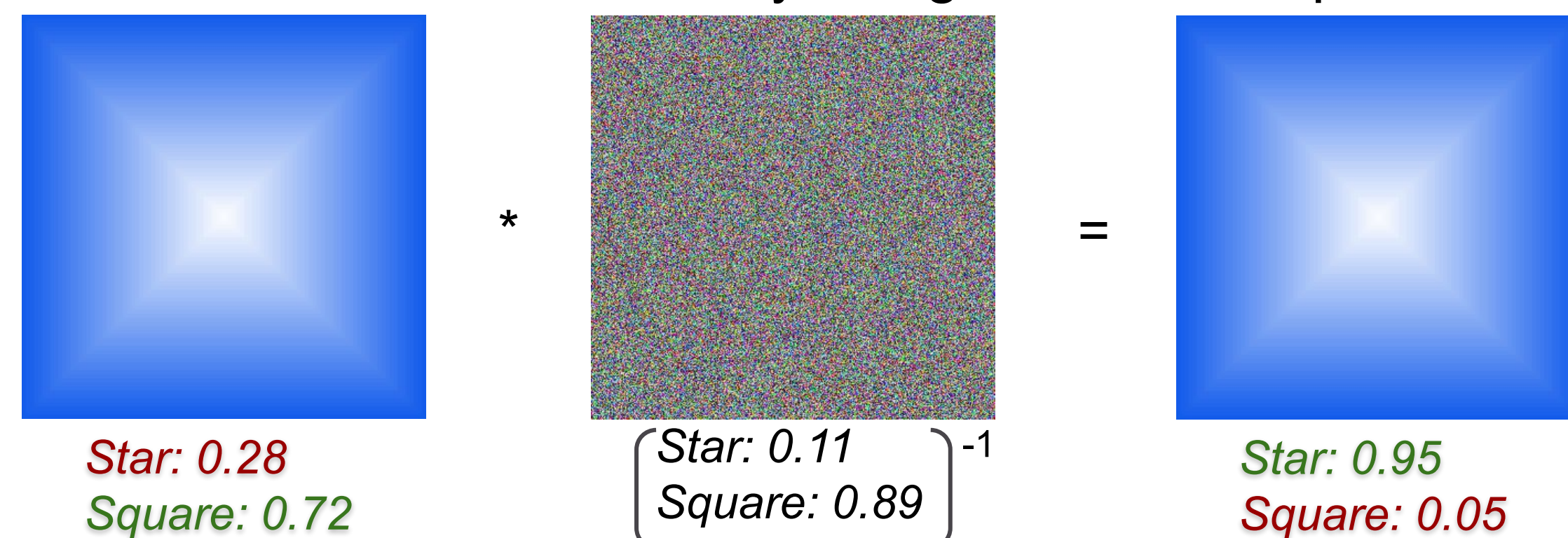
by using images of illusions with different illusion strengths



and classifying them into illusion or non-illusion prompts.

"A photo of a long blue line above a short blue line" - *Illusion prompt*
 "A photo of two equal length blue lines" - *Non-illusion prompt*

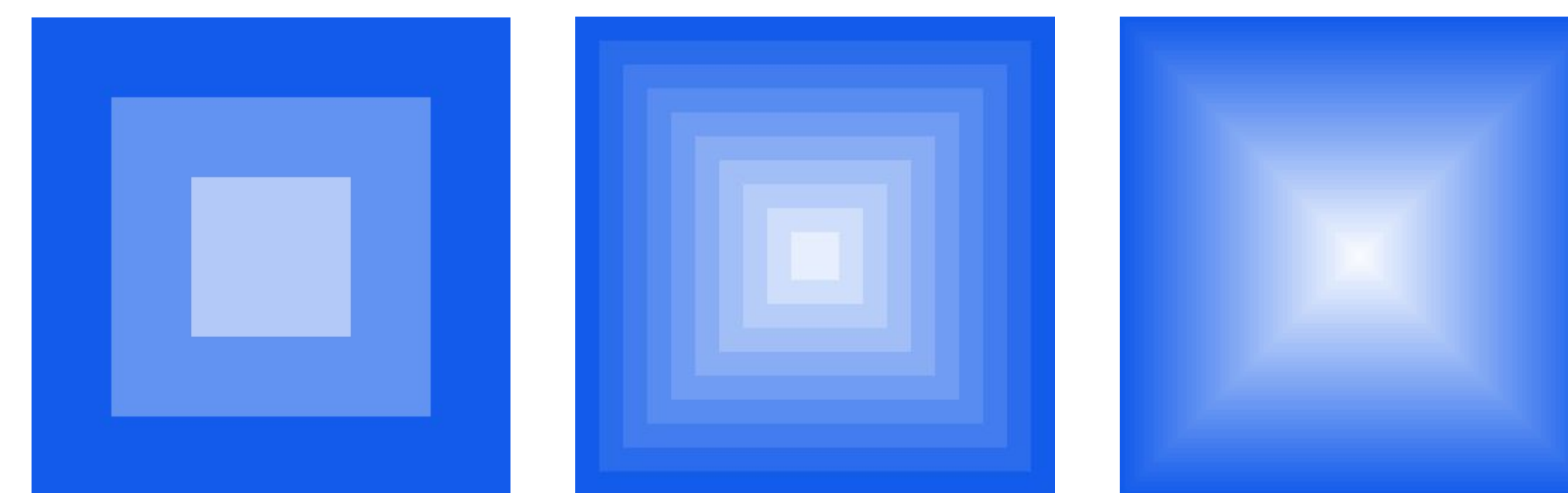
- We calibrate the association by using content-free probabilities



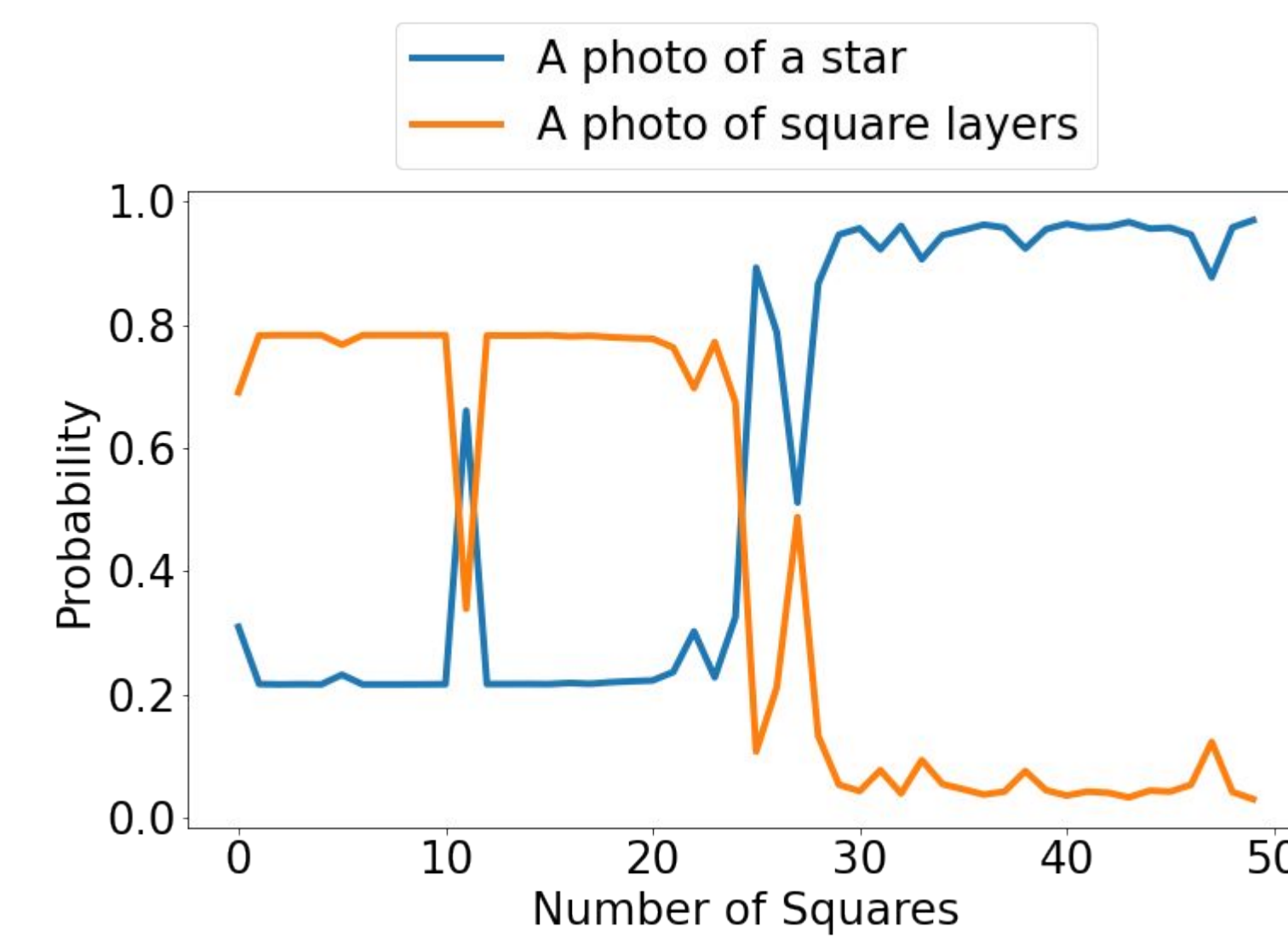
Results

Edge Detection Illusion

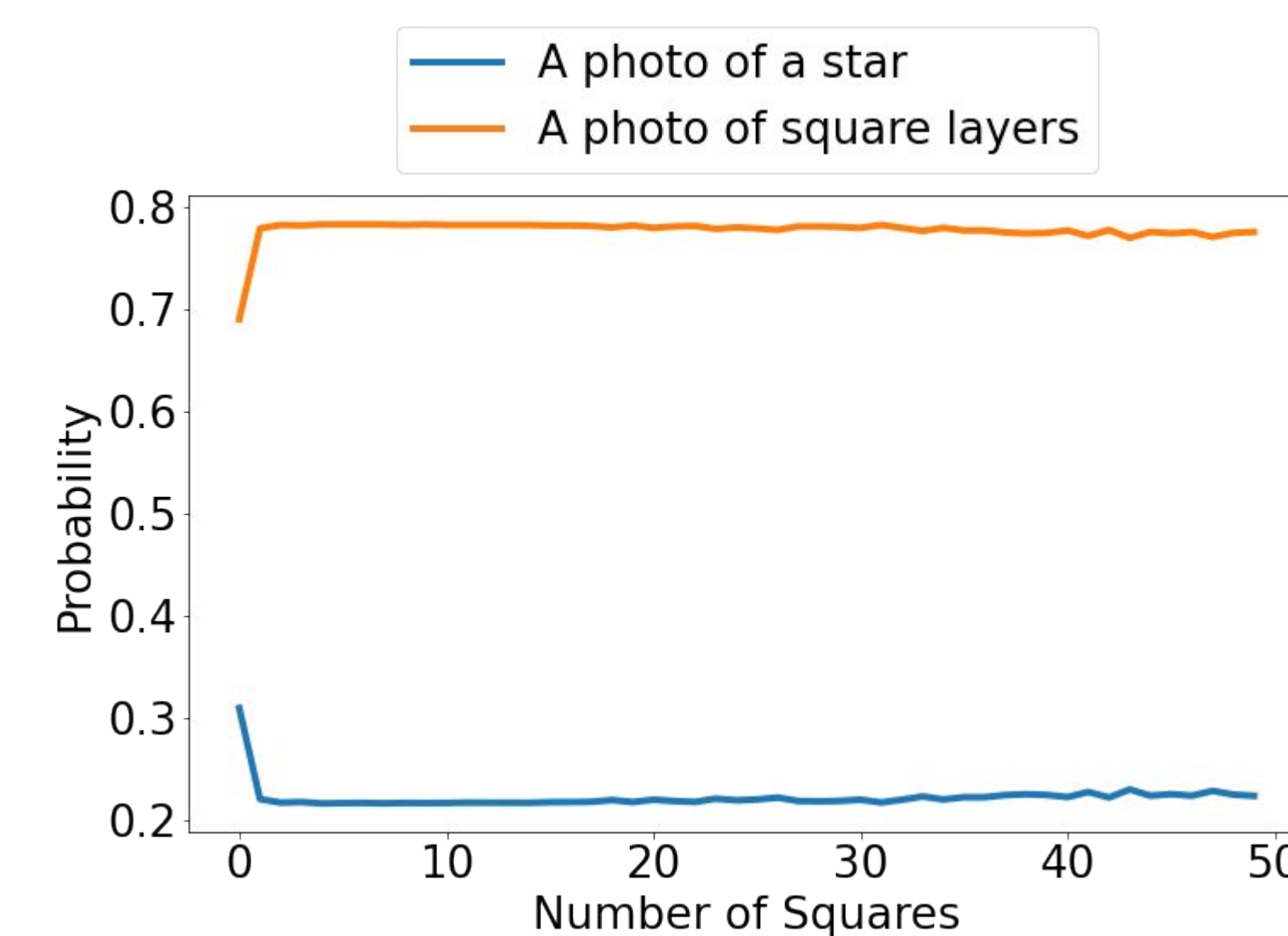
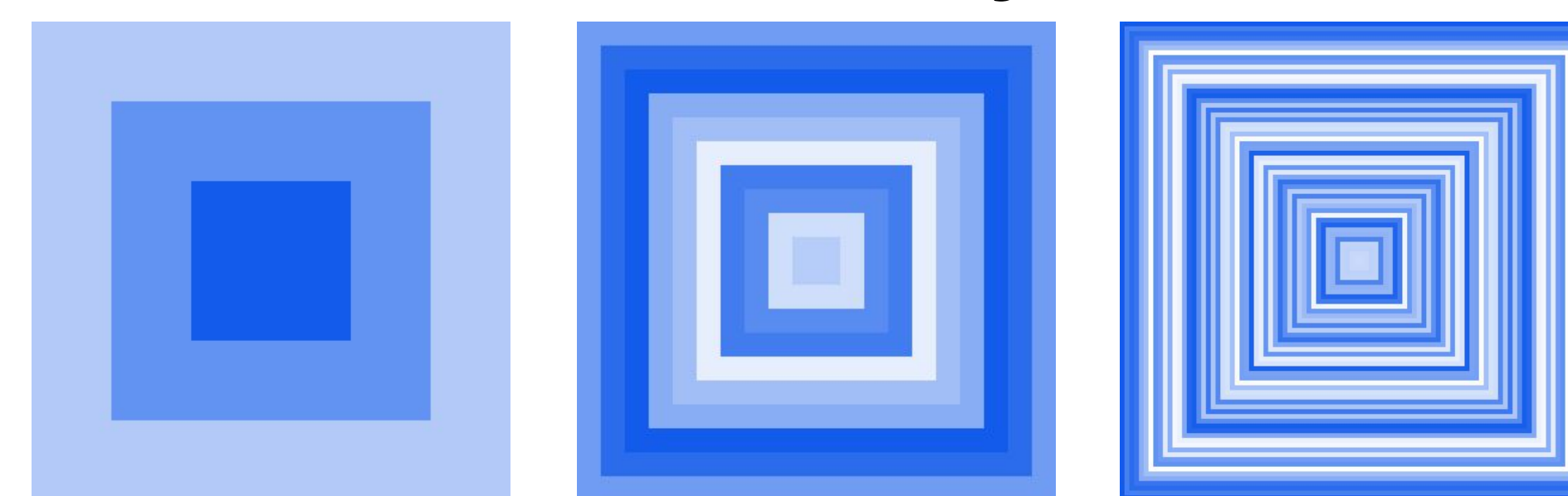
Vasarely Illusion



- Imposing squares with different luminance level
- An invisible X shape gradually appears

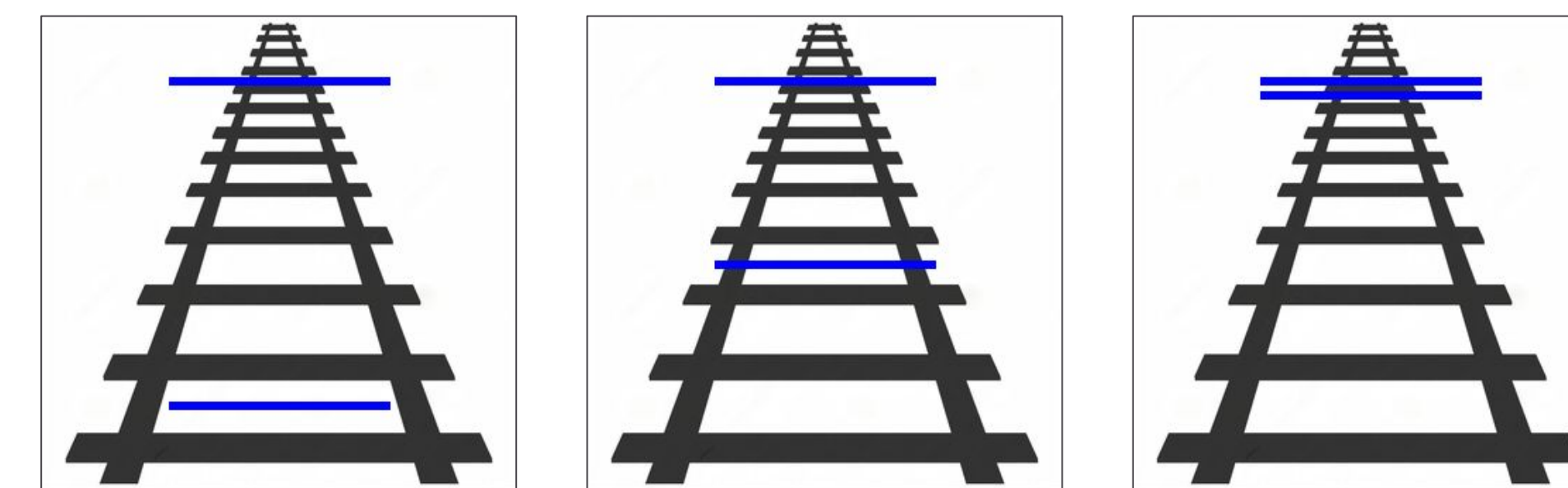


Distorted Vasarely Illusion

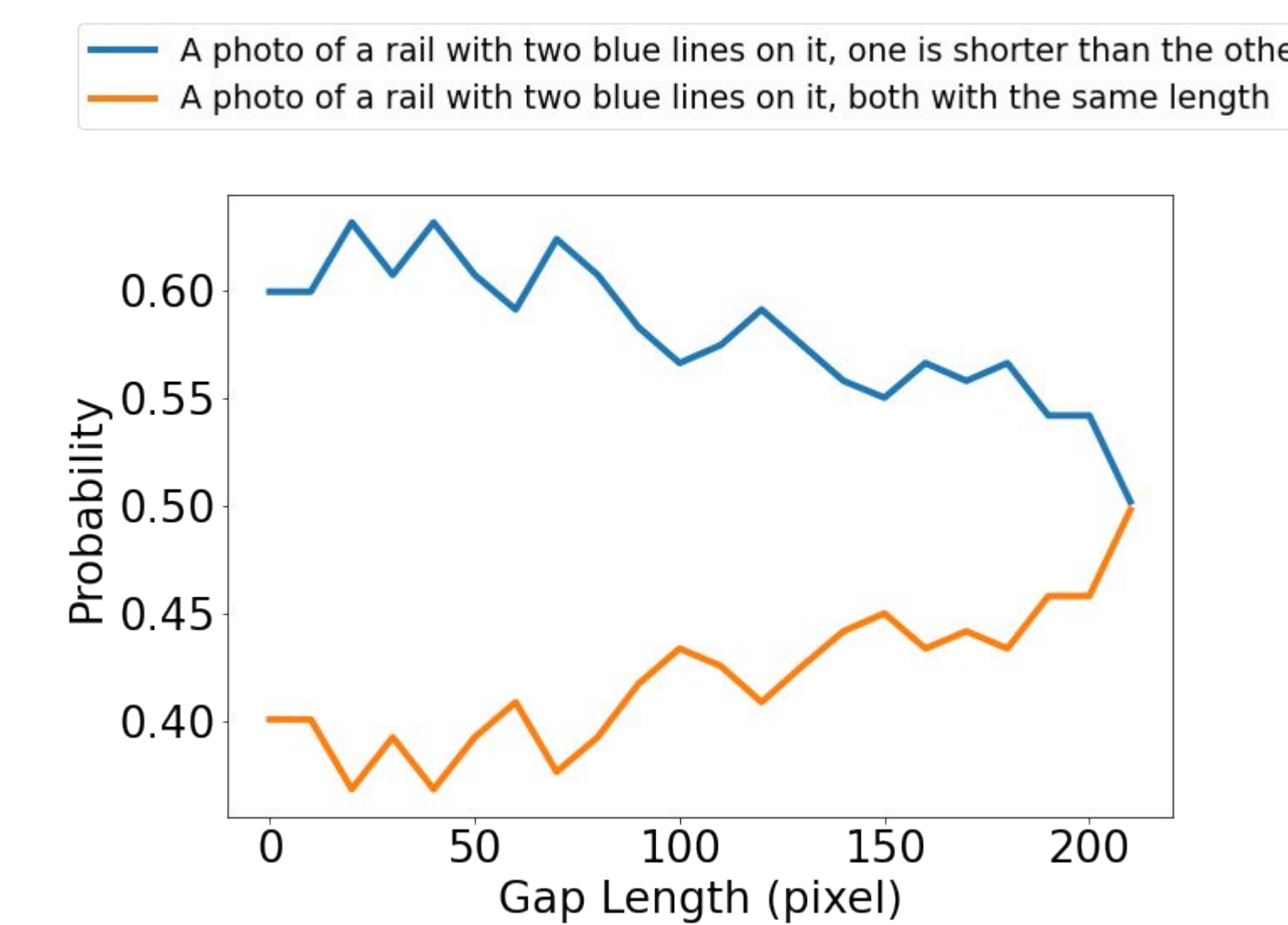


Geometrical Illusion

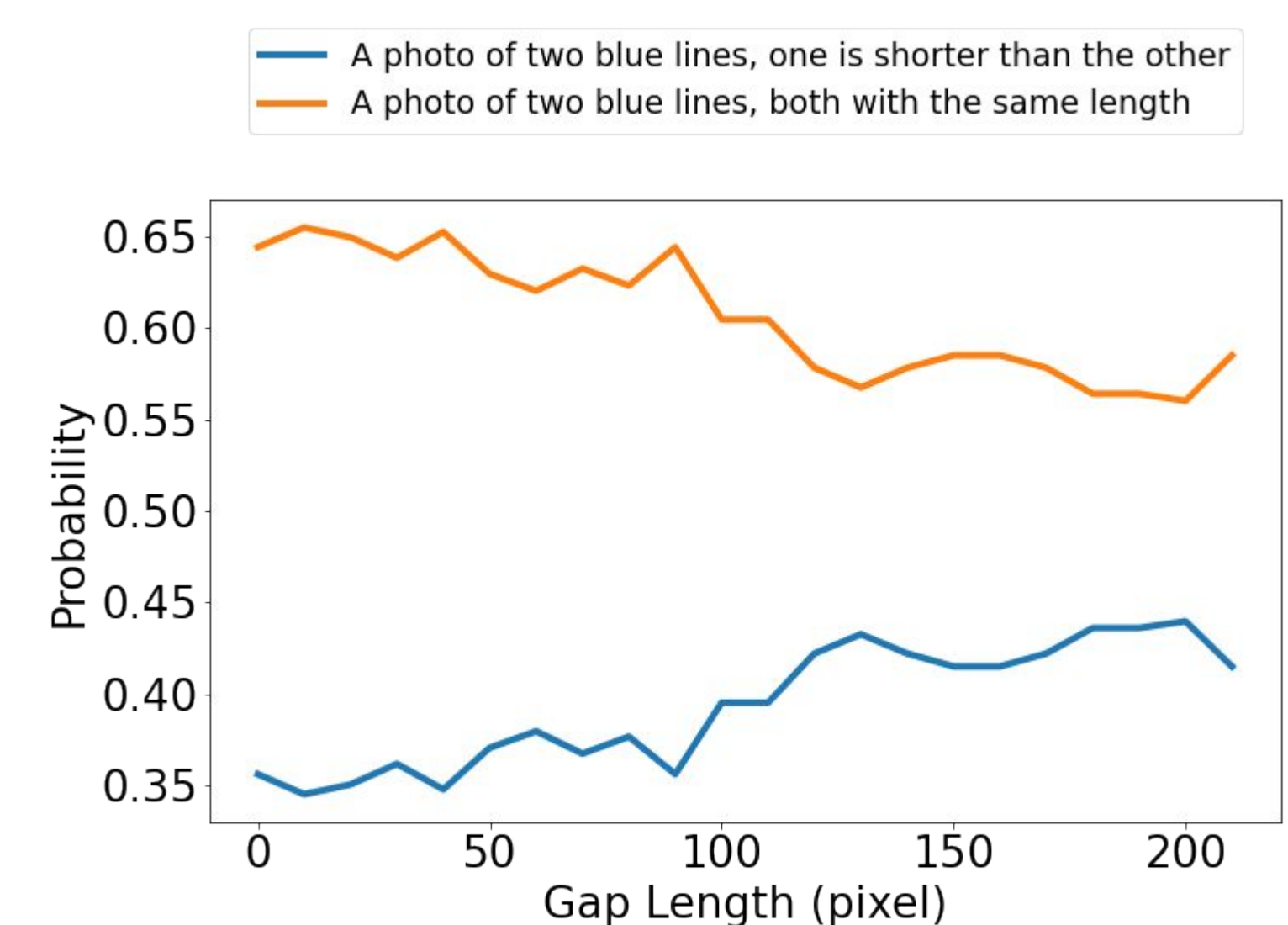
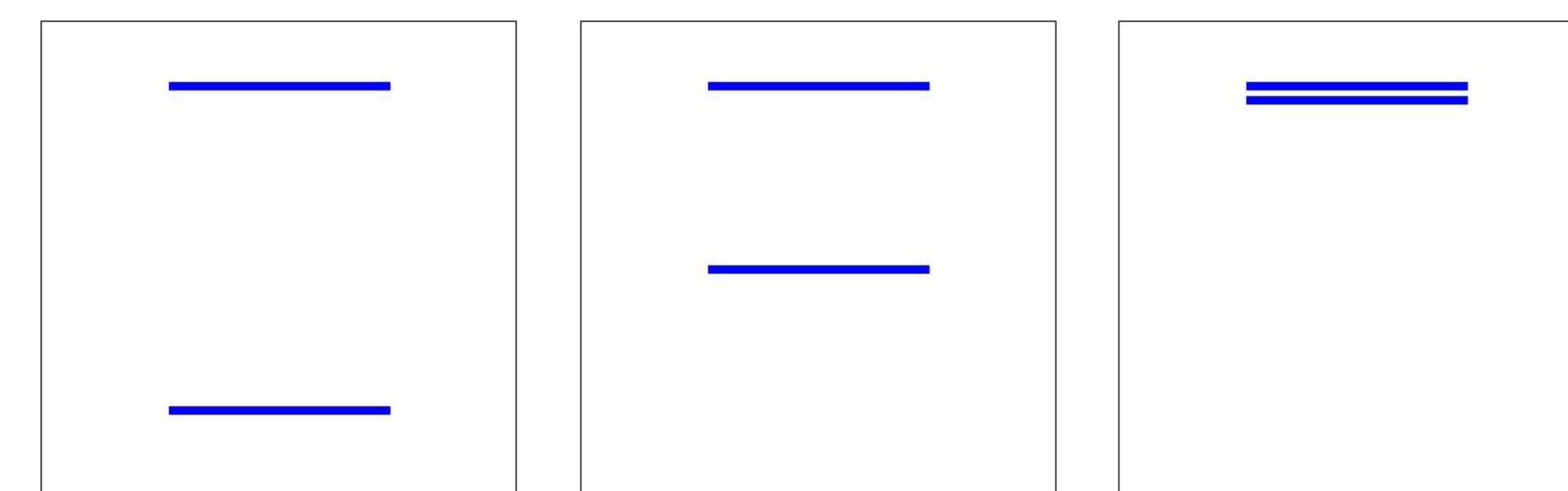
Ponzo Illusion



- Two same-length blue lines look unequal
- As two lines are closer, the illusion gets weaker

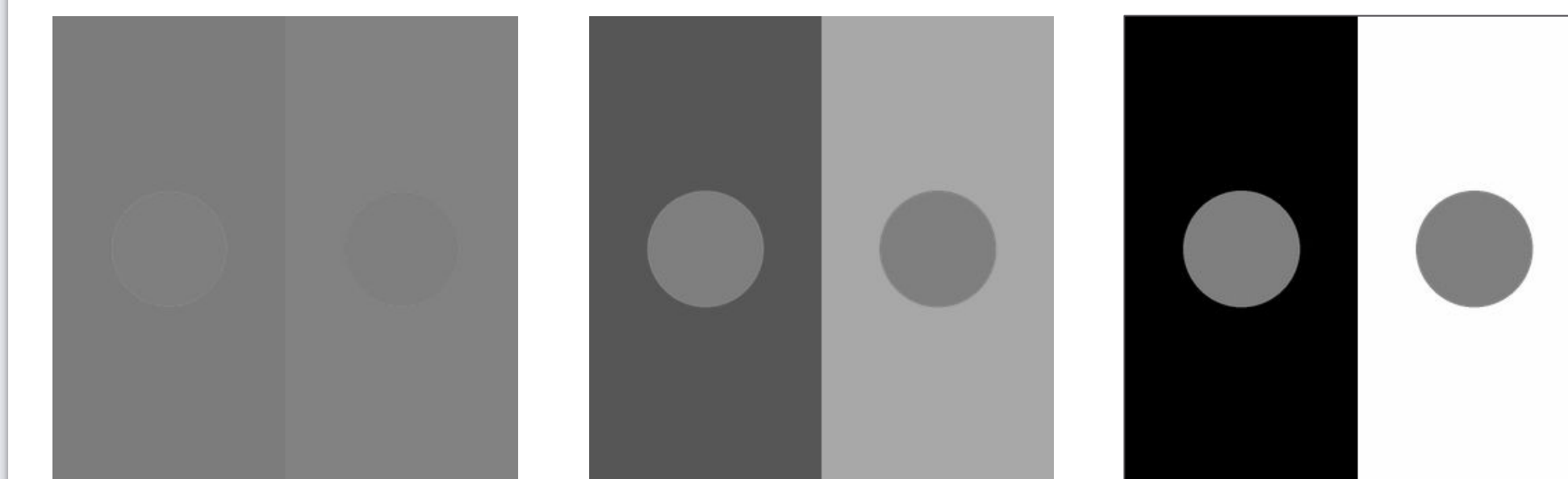


Distorted Ponzo Illusion

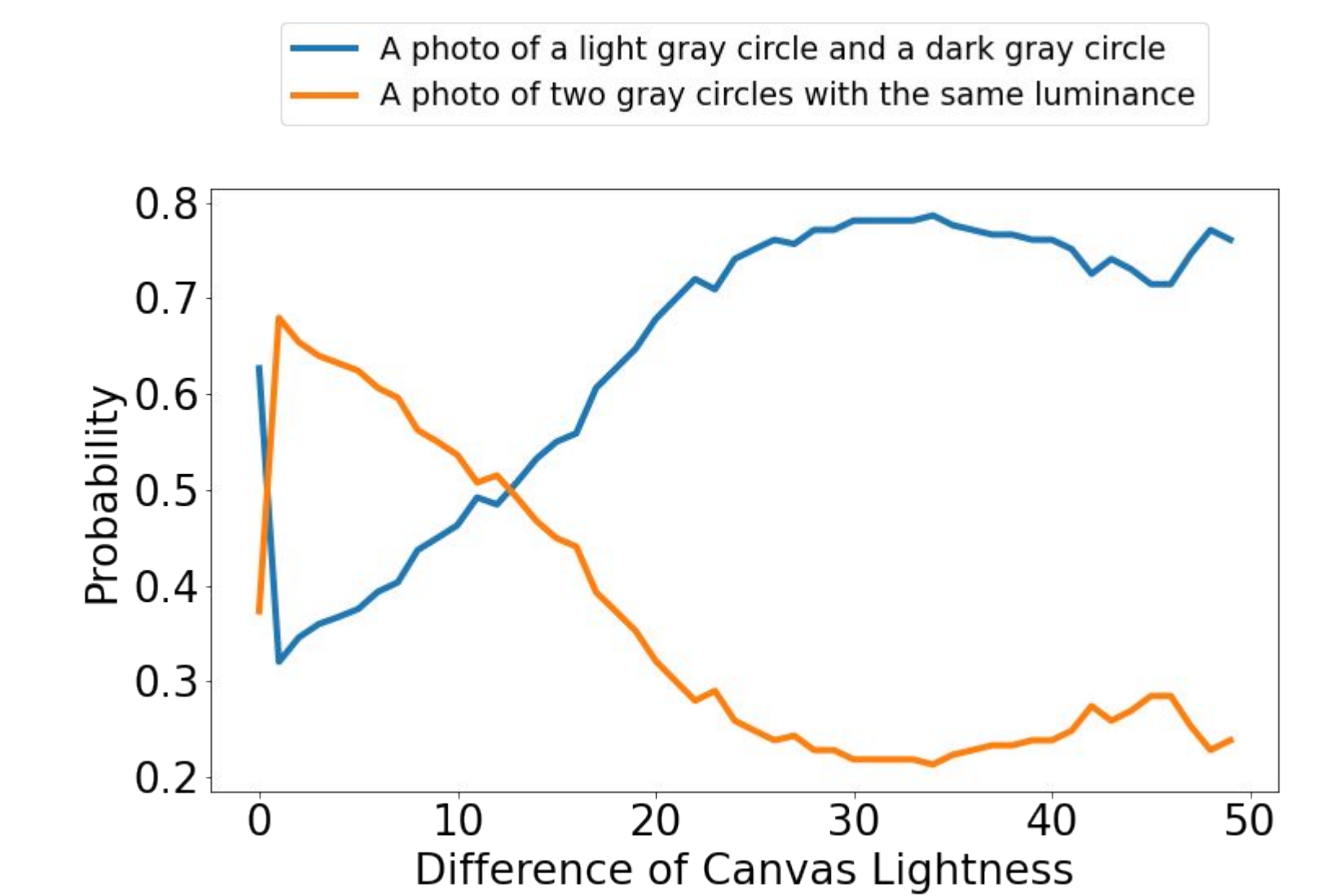


Lightness Illusion

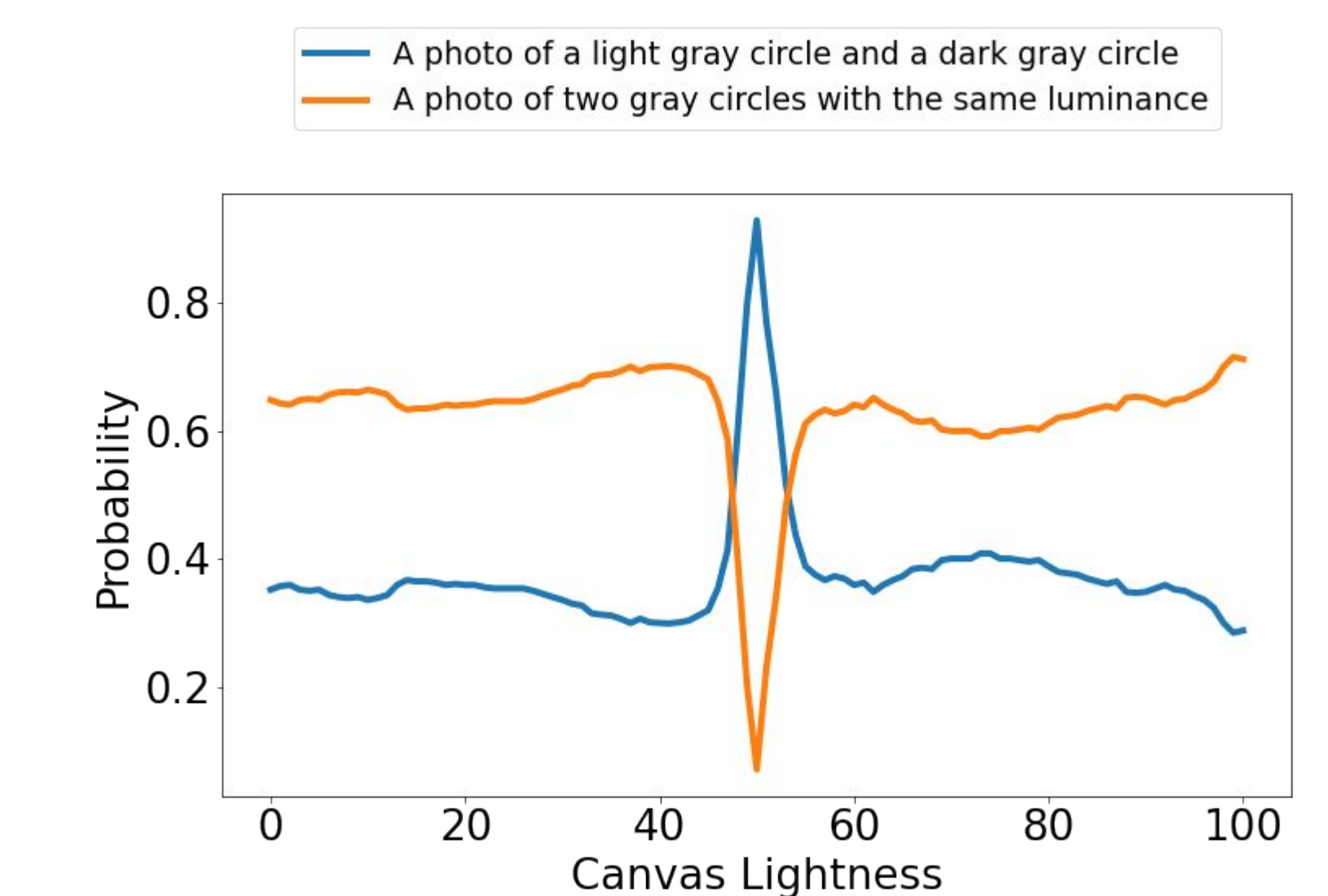
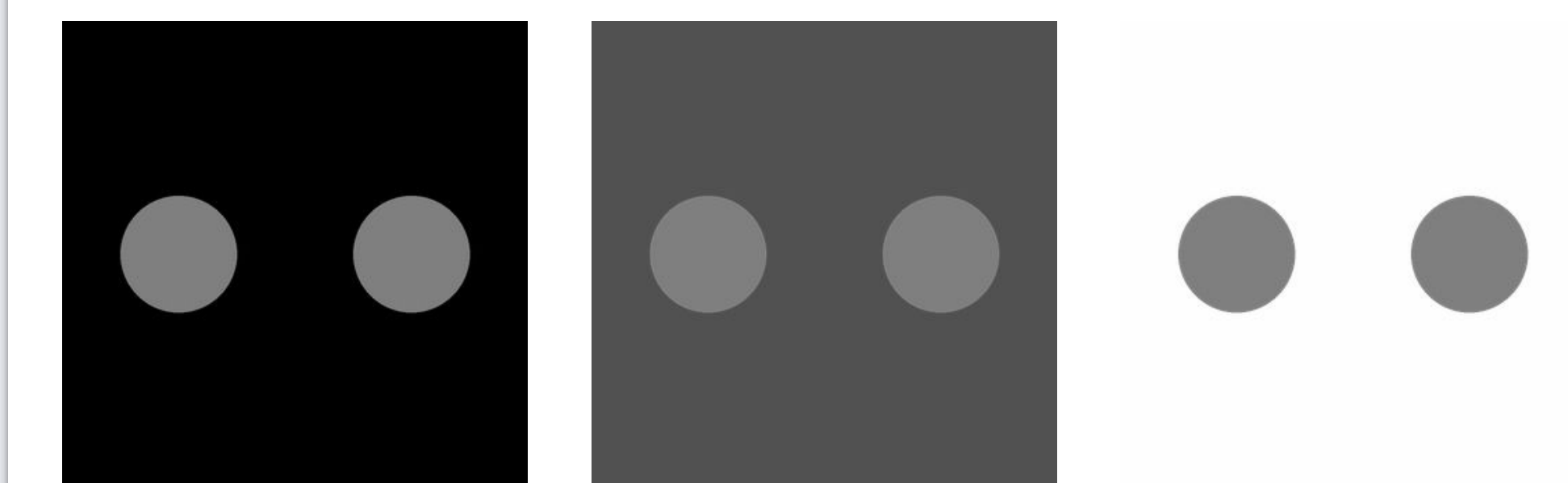
Simultaneous Contrast Illusion



- Two circles have the same level of luminance
- The left circle seems lighter than the right one



Distorted Simultaneous Contrast Ills



Conclusion

- A variety of human optical illusions fools CLIP
- This shows similarities between human biology/perception and machine learning
- This also poses potential vulnerabilities to these models
- There are multiple ways to describe an illusion, so a mass survey is needed

Acknowledgement

We appreciate Professor Yoon Kim's suggestion of using the calibration method to debias CLIP. He also helps adapt the original language model calibration to image-text calibration.

I am grateful to the Office of Graduate Education at MIT for organizing the MSRP program that enabled these research opportunities. I also thank all the interns in the cohort for their support and participation in the illusion-prompt survey.

Project Link

jerryngo.com/clip-illusion/

