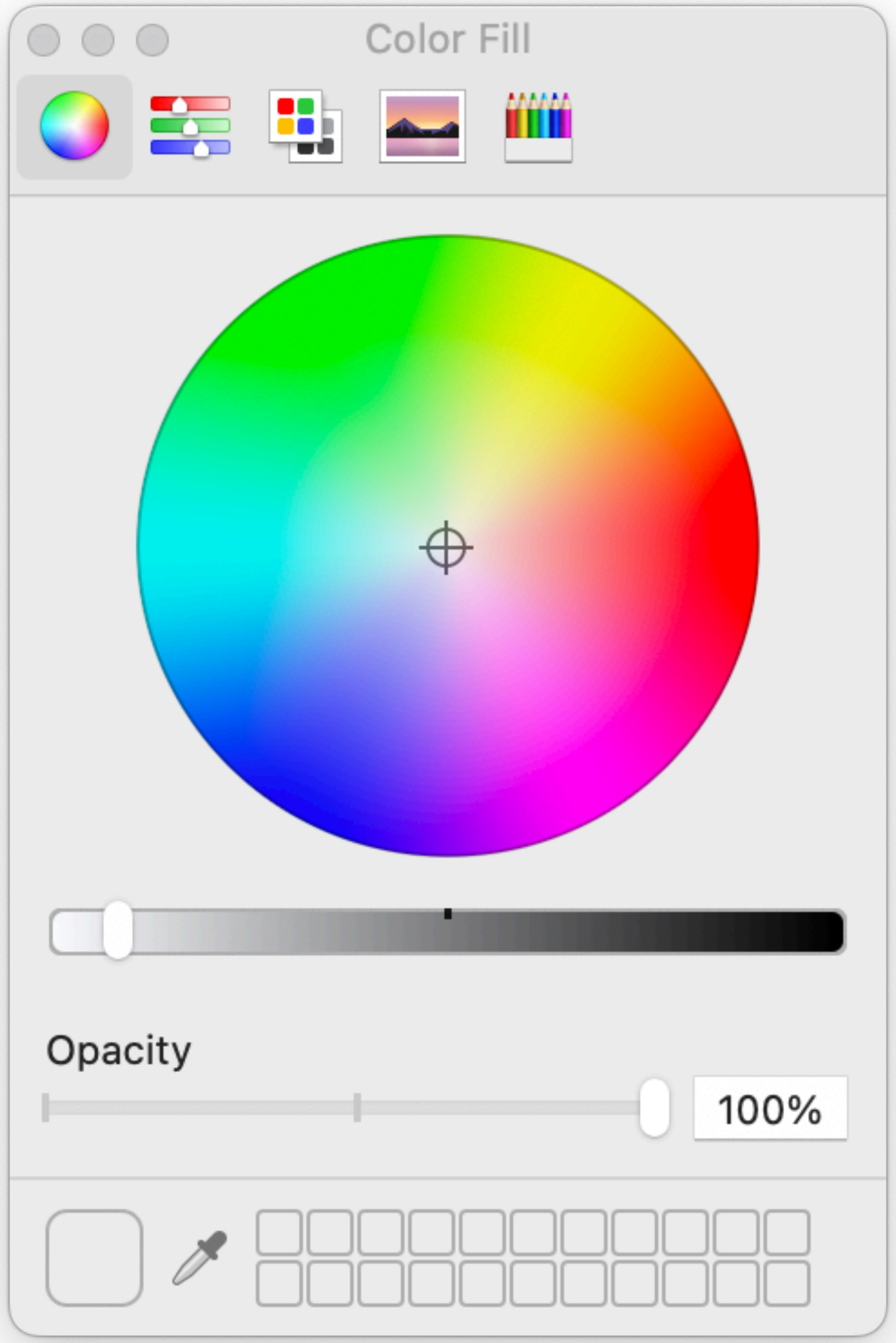
Better Computer Graphics Through Understanding of Human Visual System

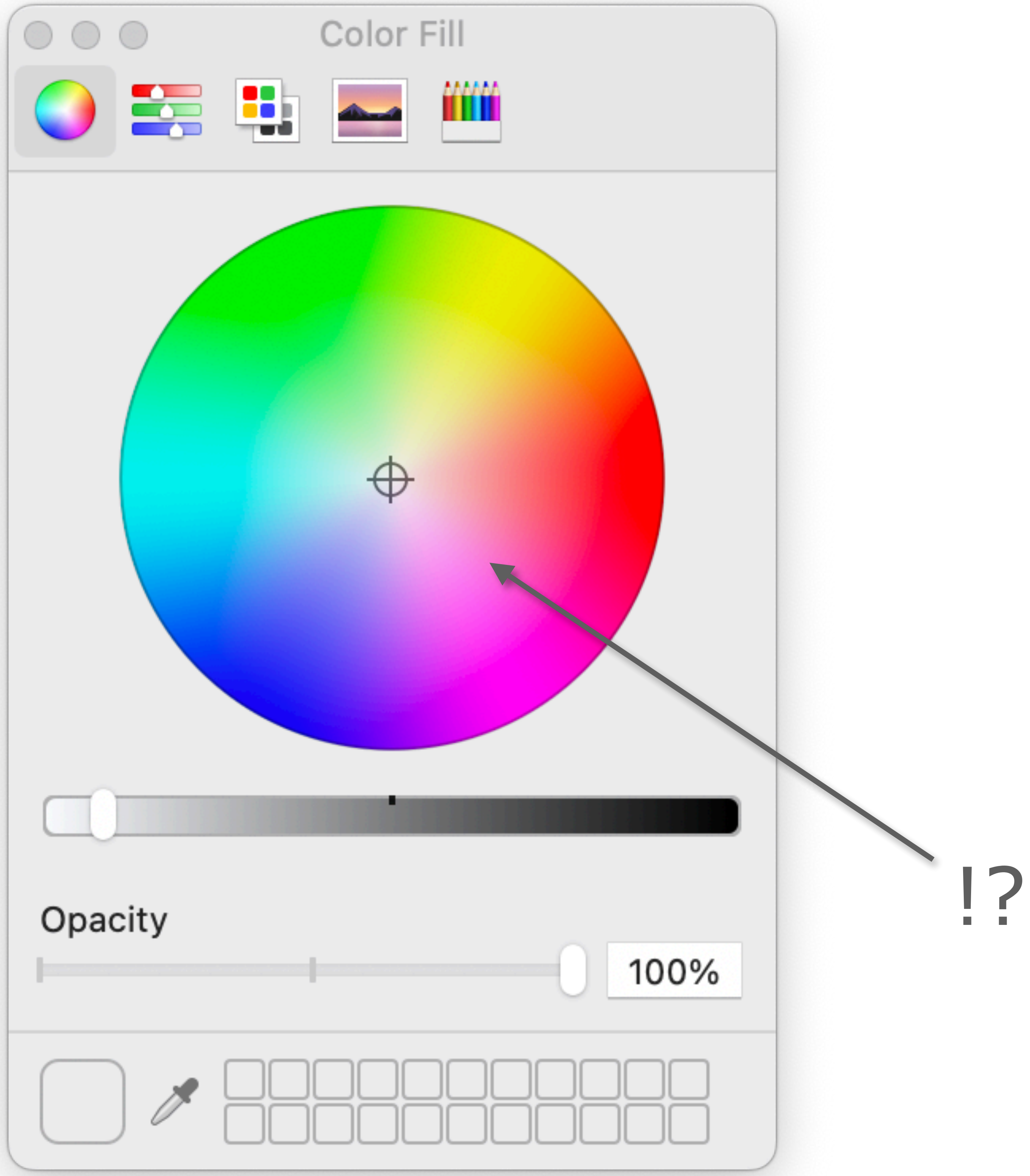
Sergei Savchenko

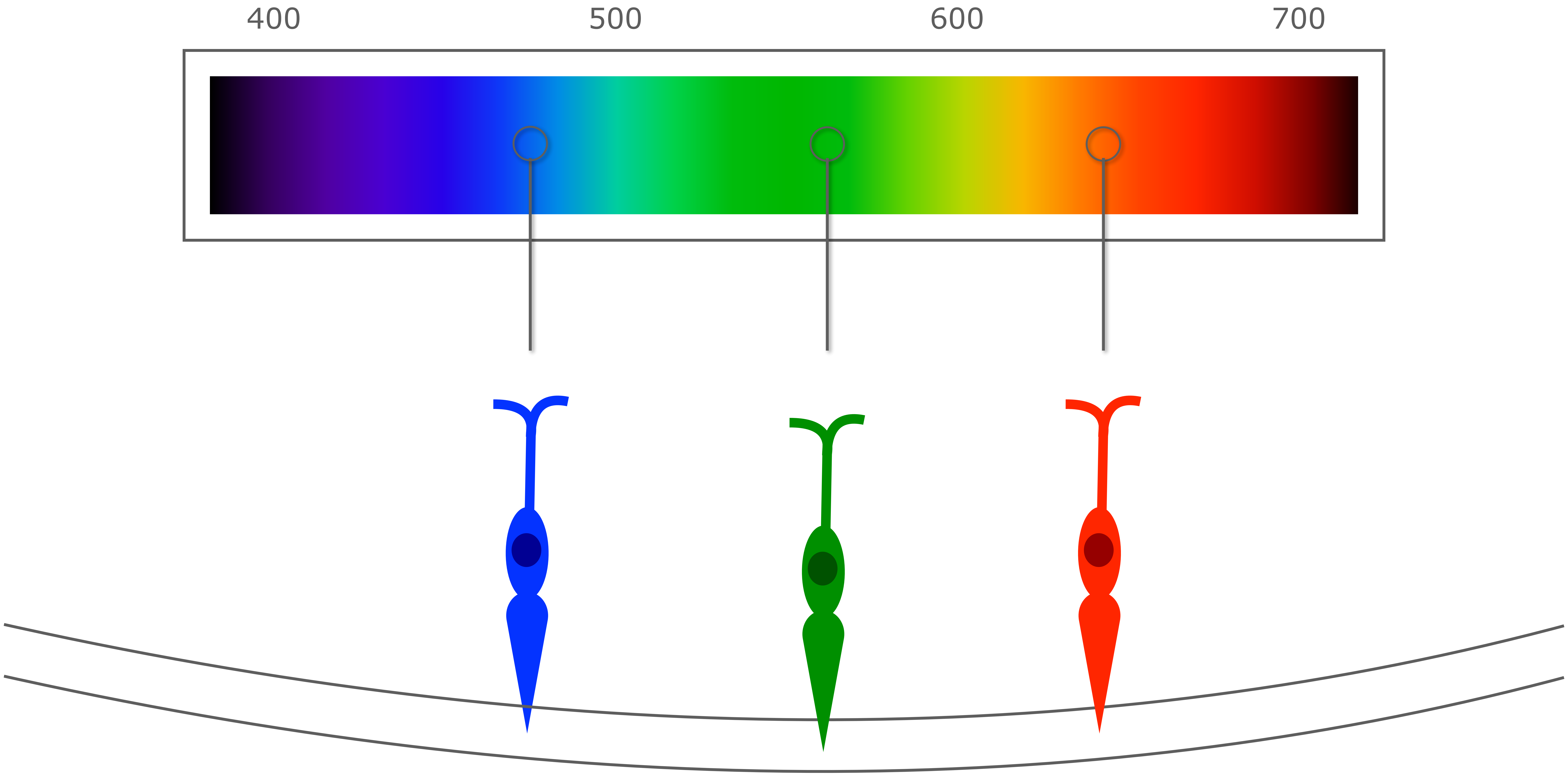
Technical Lead
Bethesda Game Studios

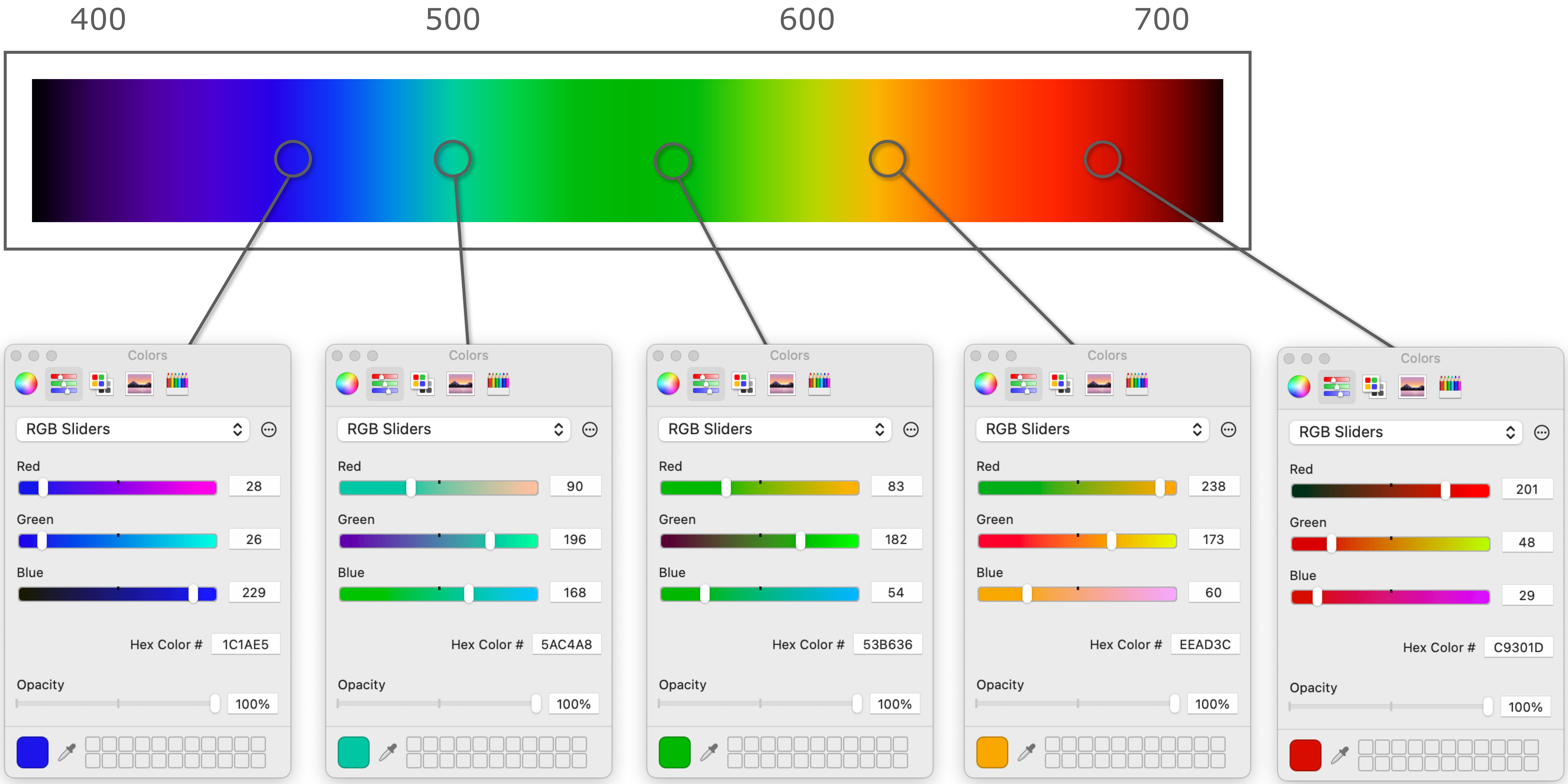
Sergei Savchenko

- Technical Lead at Bethesda Game Studios
- Also worked at 3DO, EA and Warner Bros.
- Consoles, Handhelds and Mobiles
- Wrote a book on graphics years back


Color Fill
Opacity
100%


Color Fill
Opacity
100%
!?


400
500
600
700


400
500
600
700
Colors
RGB Sliders
Red 28
Green 26
Blue 229
Hex Color # 1C1AE5
Opacity 100%
Colors
RGB Sliders
Red 90
Green 196
Blue 168
Hex Color # 5AC4A8
Opacity 100%
Colors
RGB Sliders
Red 83
Green 182
Blue 54
Hex Color # 53B636
Opacity 100%
Colors
RGB Sliders
Red 238
Green 173
Blue 60
Hex Color # EEAD3C
Opacity 100%
Colors
RGB Sliders
Red 201
Green 48
Blue 29
Hex Color # C9301D
Opacity 100%


400
500
600
700
!?
Colors
RGB Sliders
Red 72
Green 9
Blue 159
Hex Color # 48099F
Opacity 100%
Colors
RGB Sliders
Red 28
Green 26
Blue 229
Hex Color # 1C1AE5
Opacity 100%
Colors
RGB Sliders
Red 90
Green 196
Blue 168
Hex Color # 5AC4A8
Opacity 100%
Colors
RGB Sliders
Red 83
Green 182
Blue 54
Hex Color # 53B636
Opacity 100%
Colors
RGB Sliders
Red 238
Green 173
Blue 60
Hex Color # EEAD3C
Opacity 100%
Colors
RGB Sliders
Red 201
Green 48
Blue 29
Hex Color # C9301D
Opacity 100%

Not the only weird thing about human visual perception…

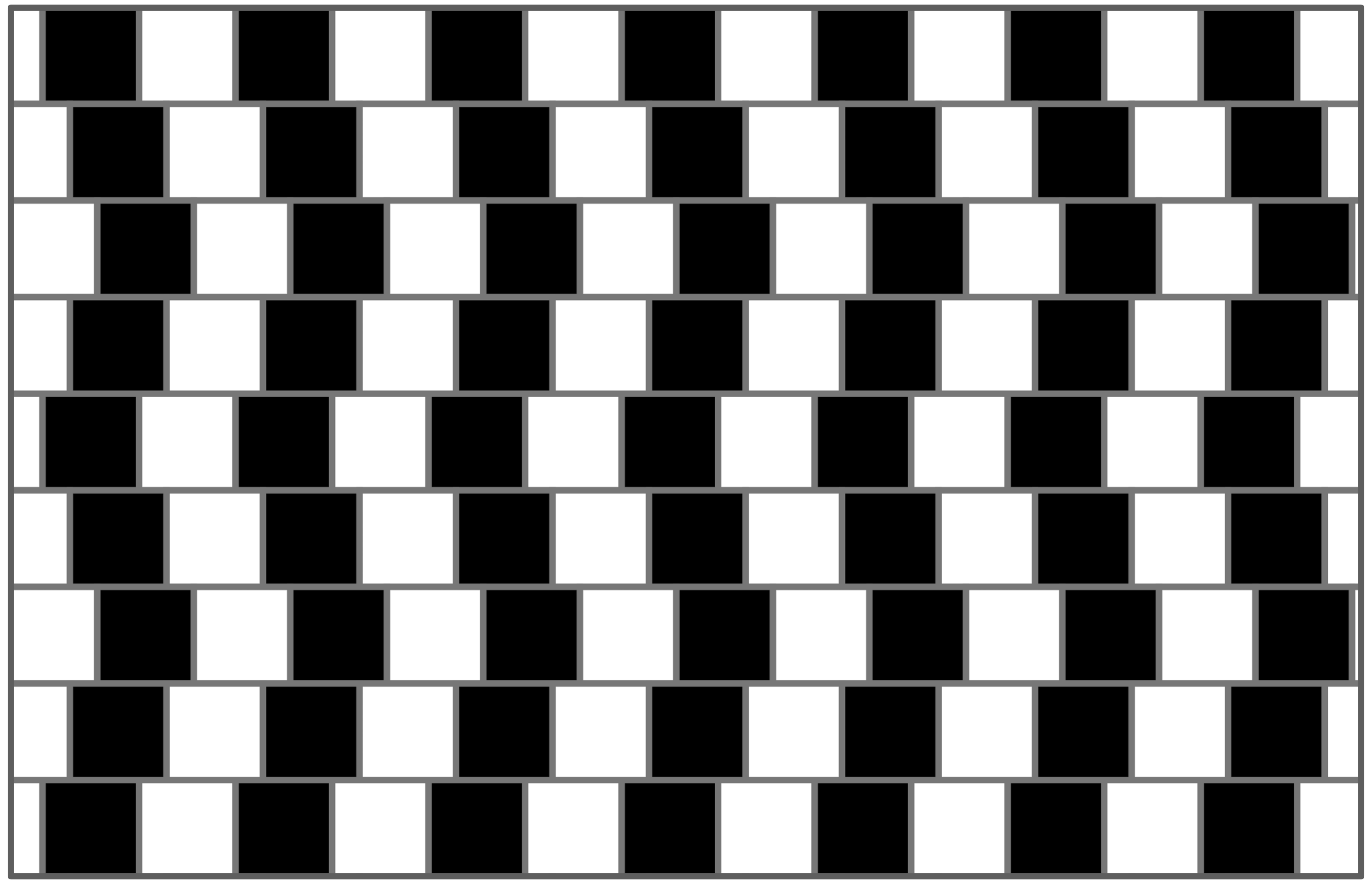
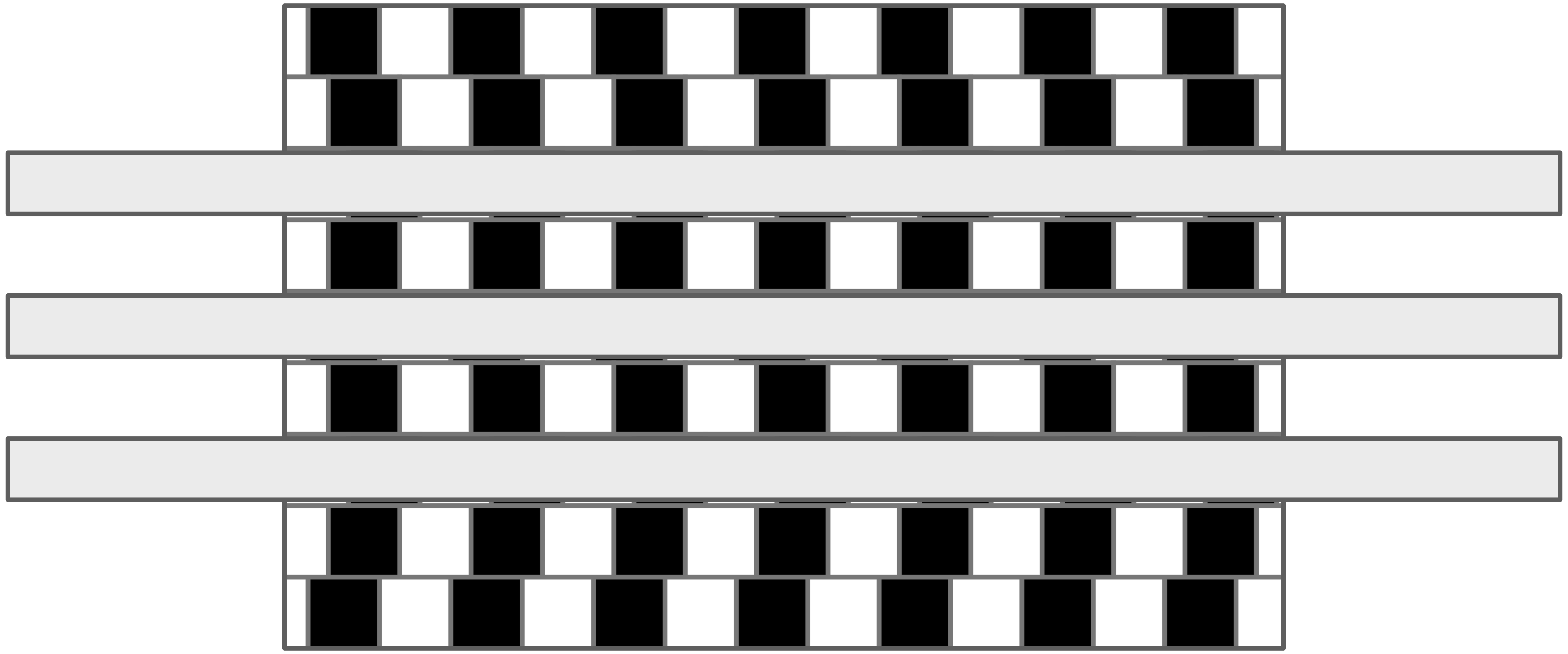
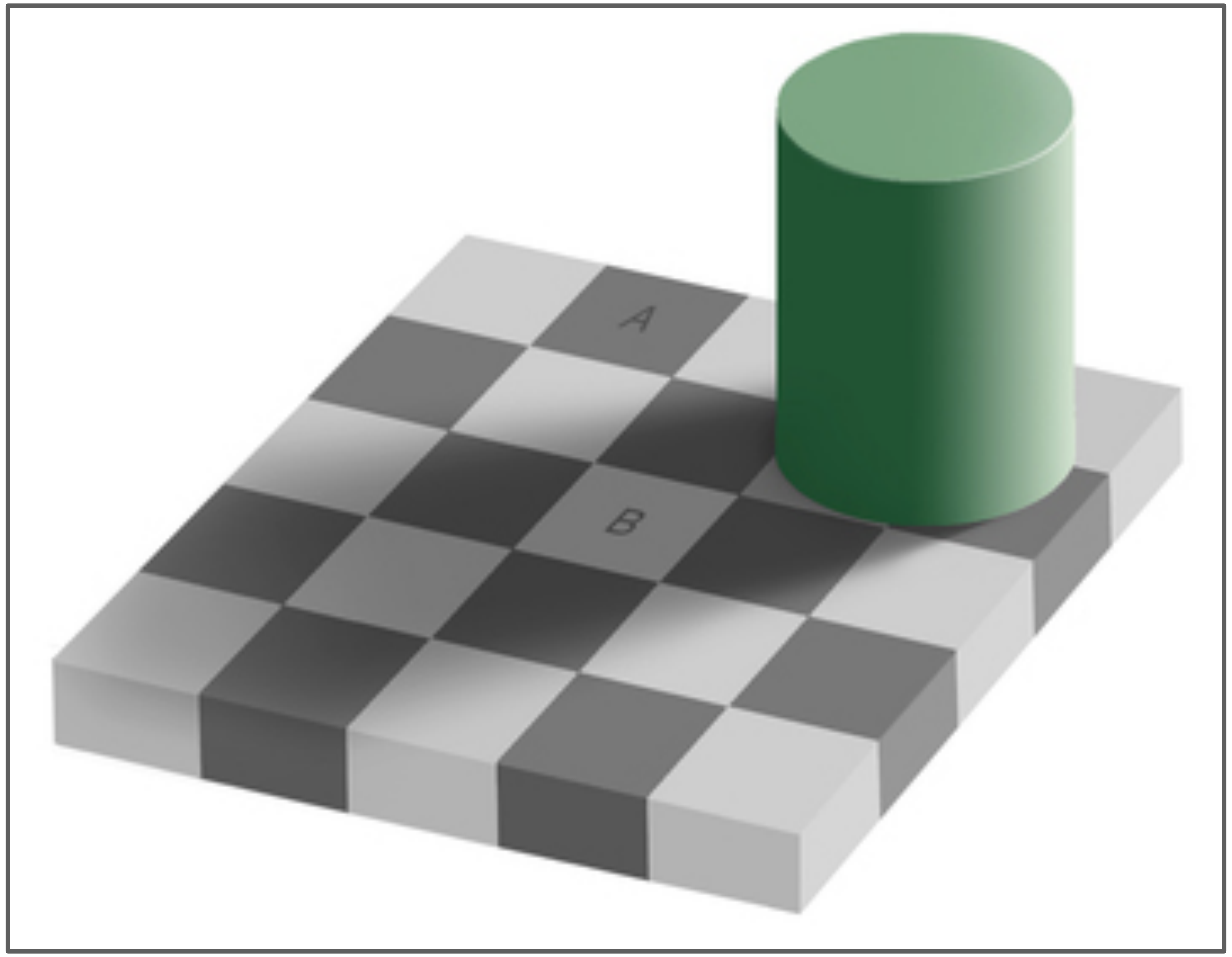
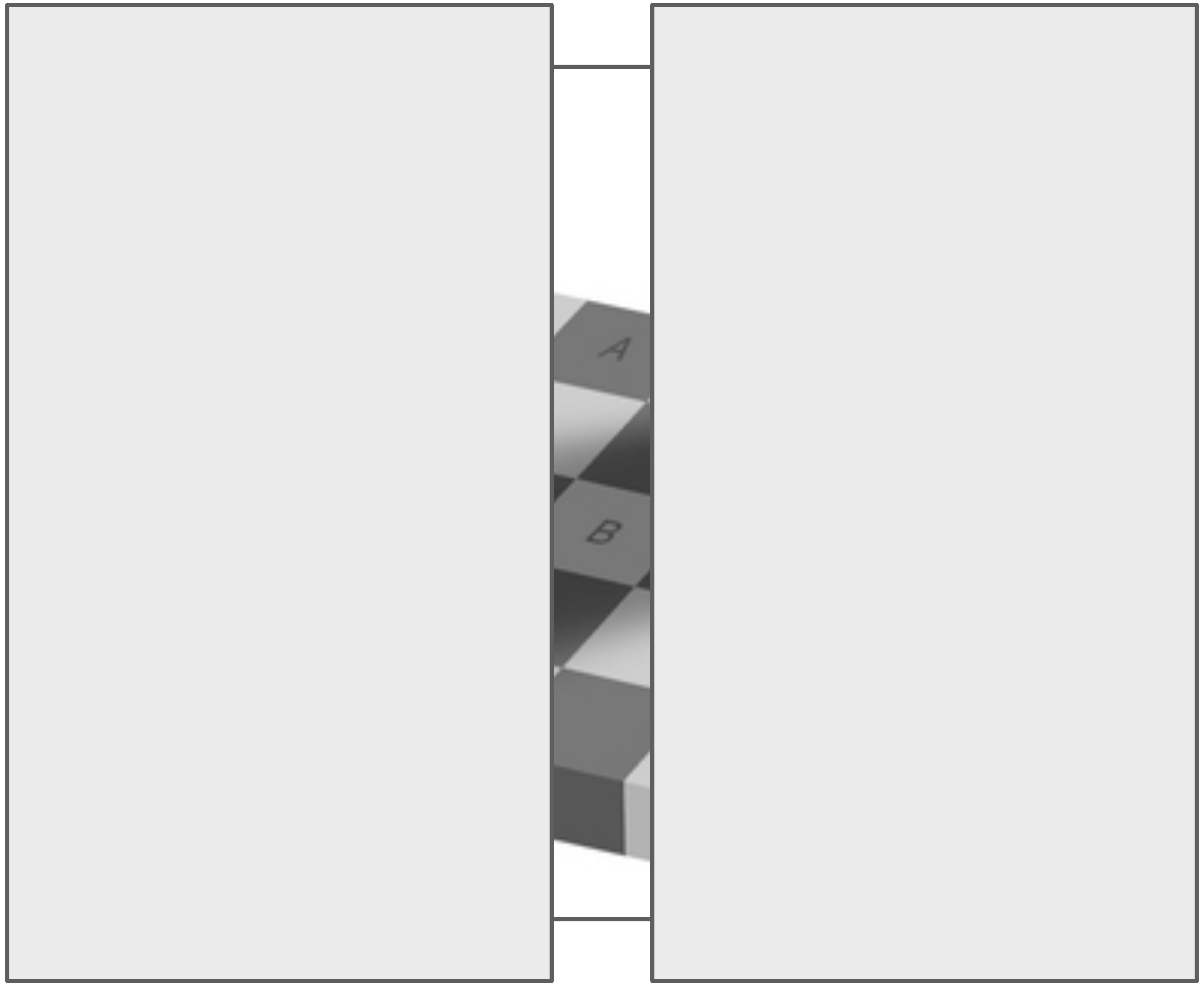


Image by Fibonacci, CC BY-CA 3.0




A
B


A
B

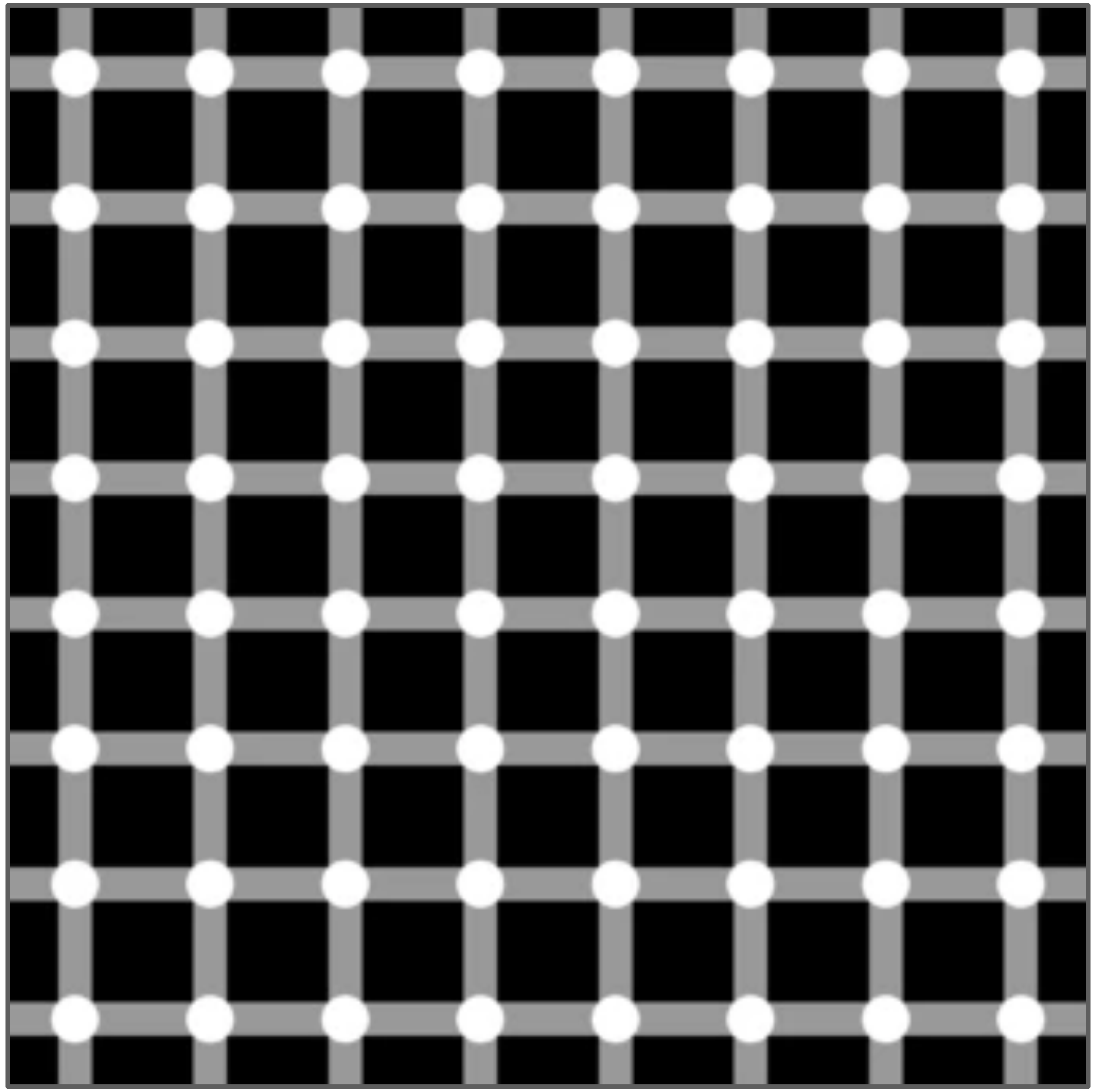


Image by António Miguel de Campos, Public Domain

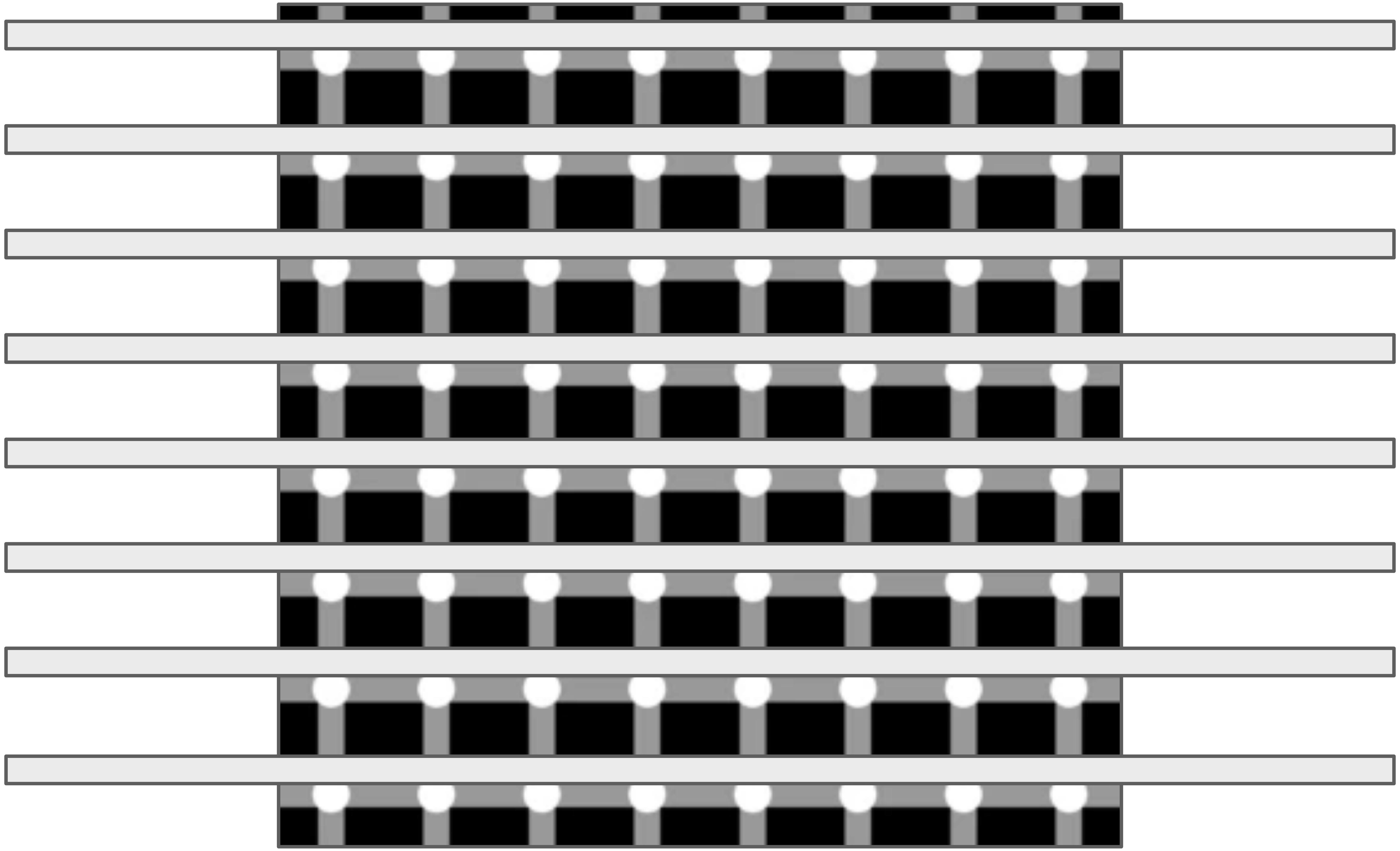
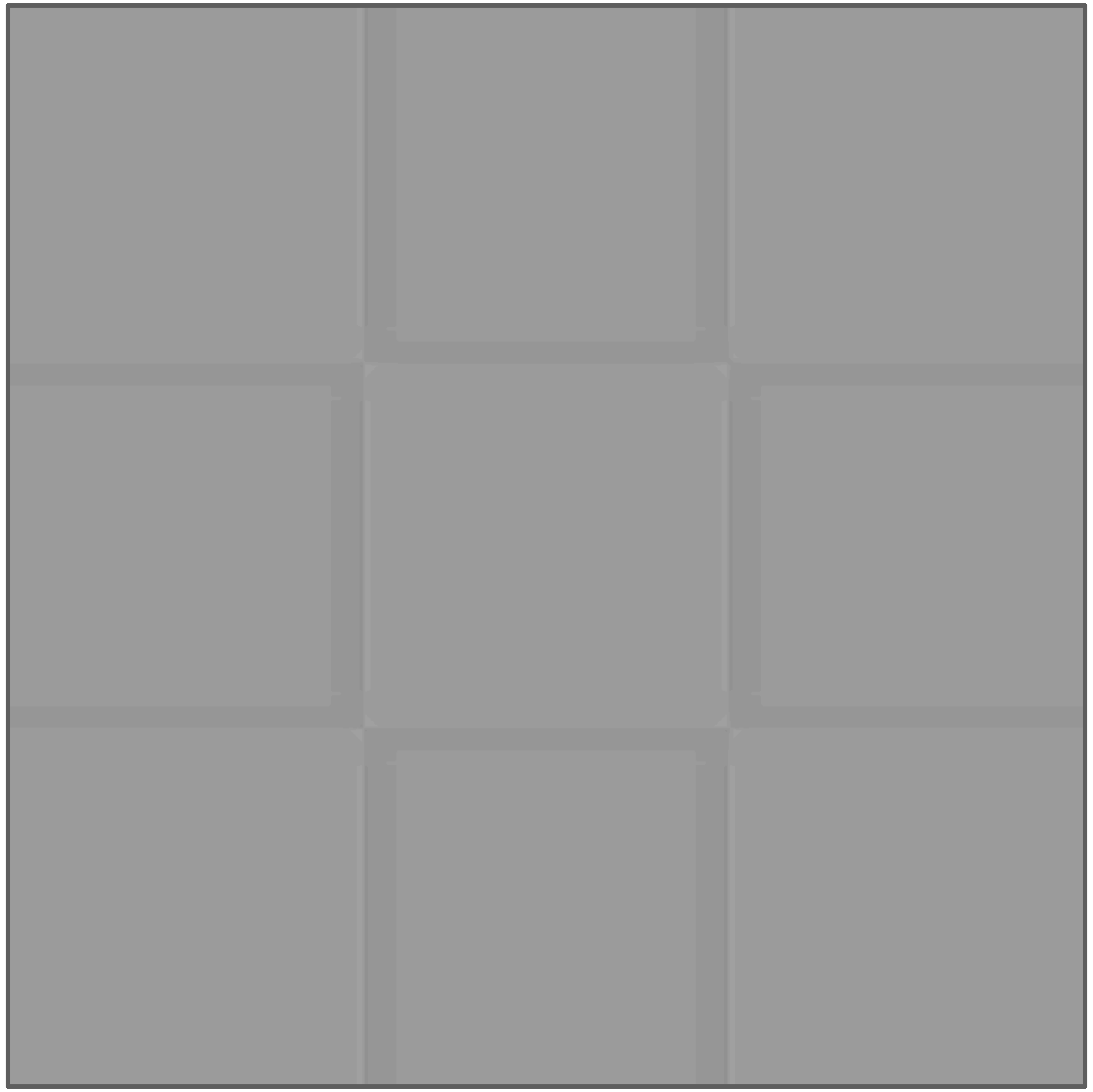
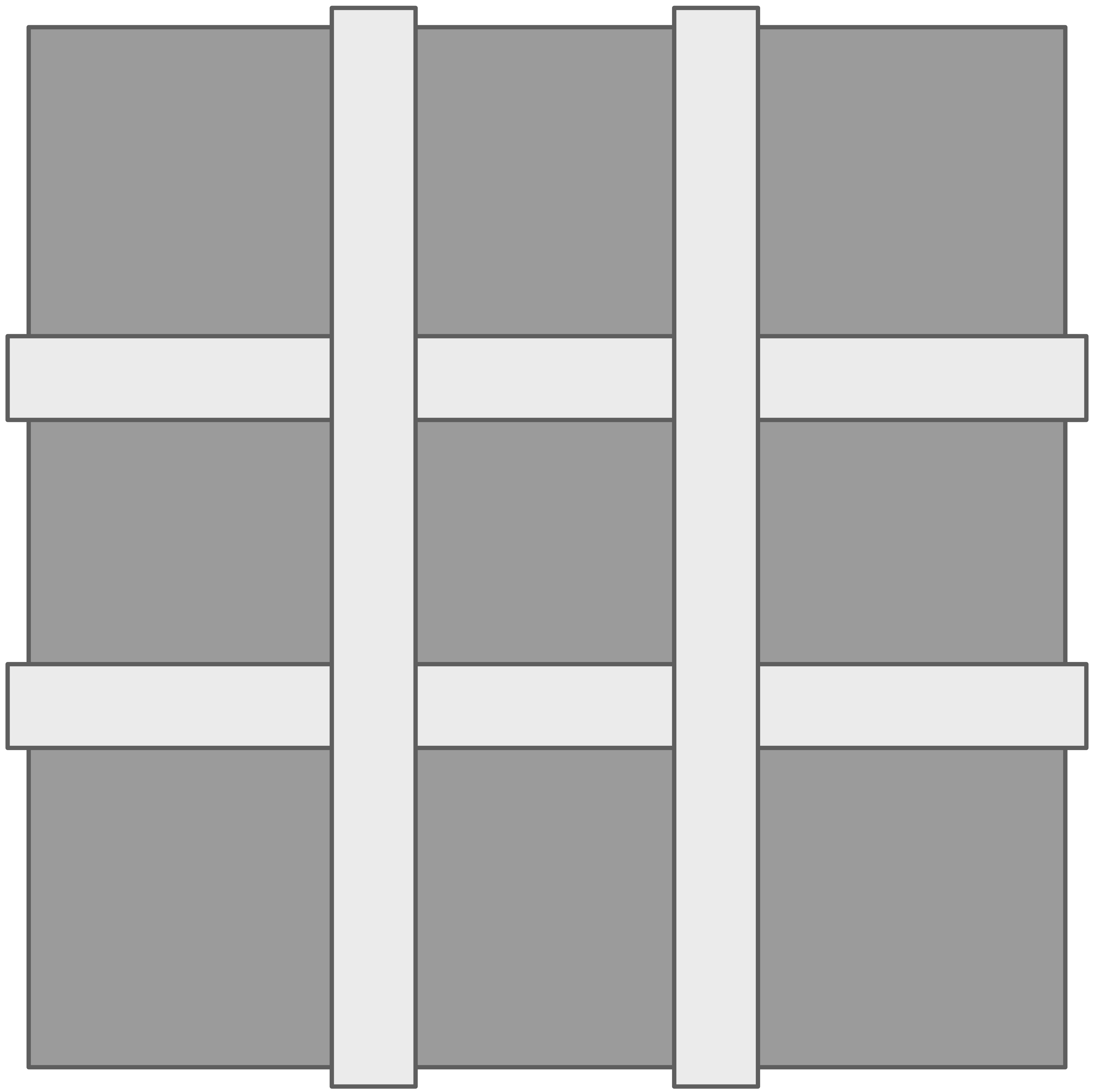
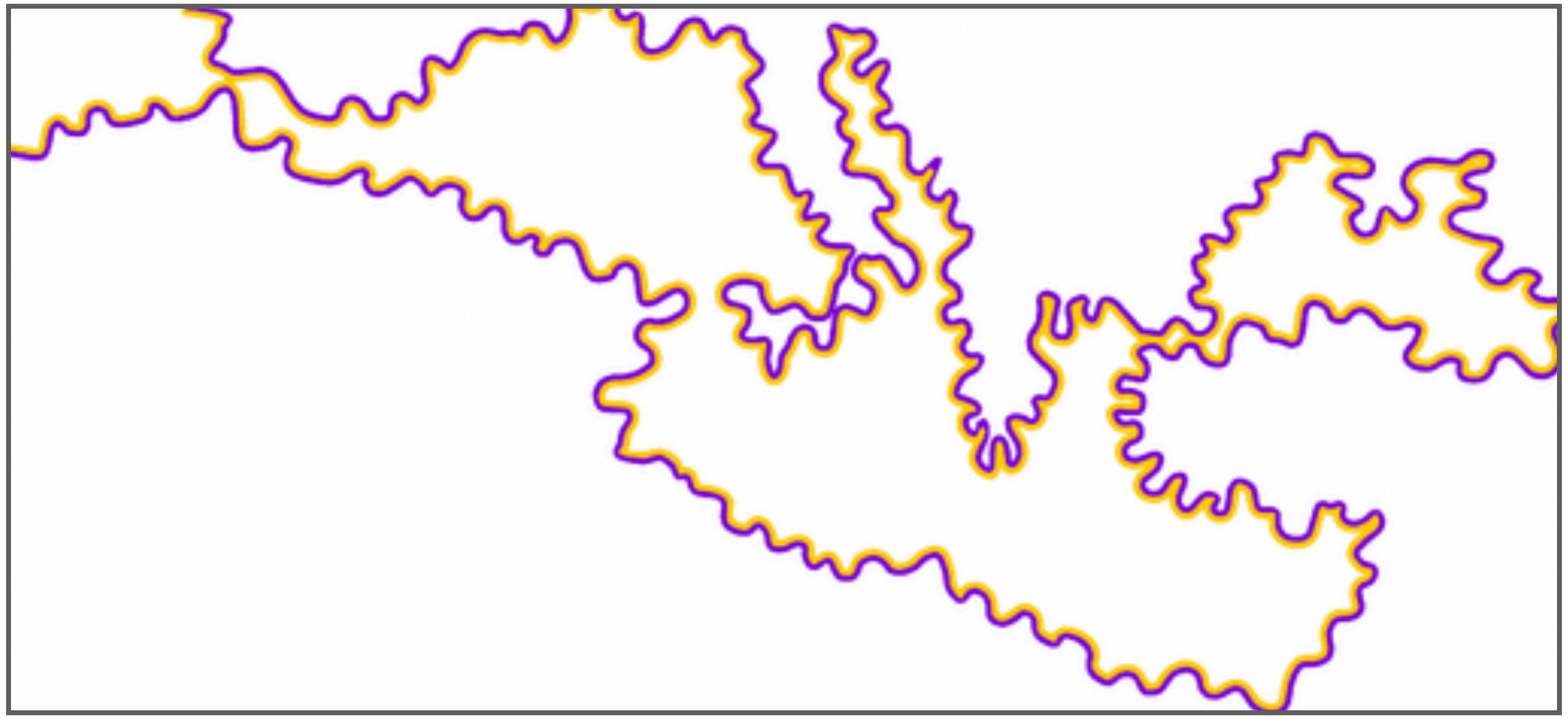


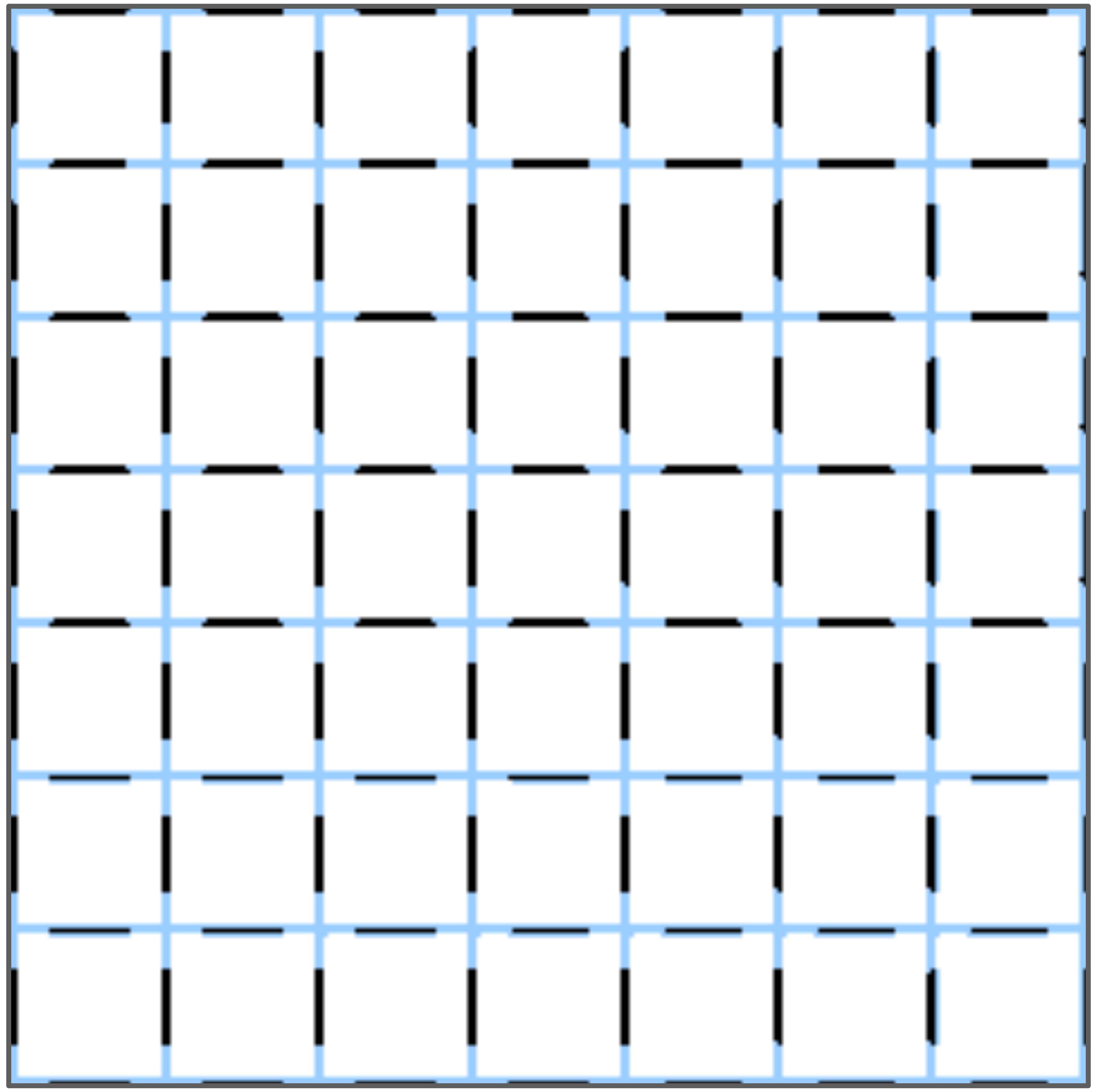
Image by António Miguel de Campos, Public Domain

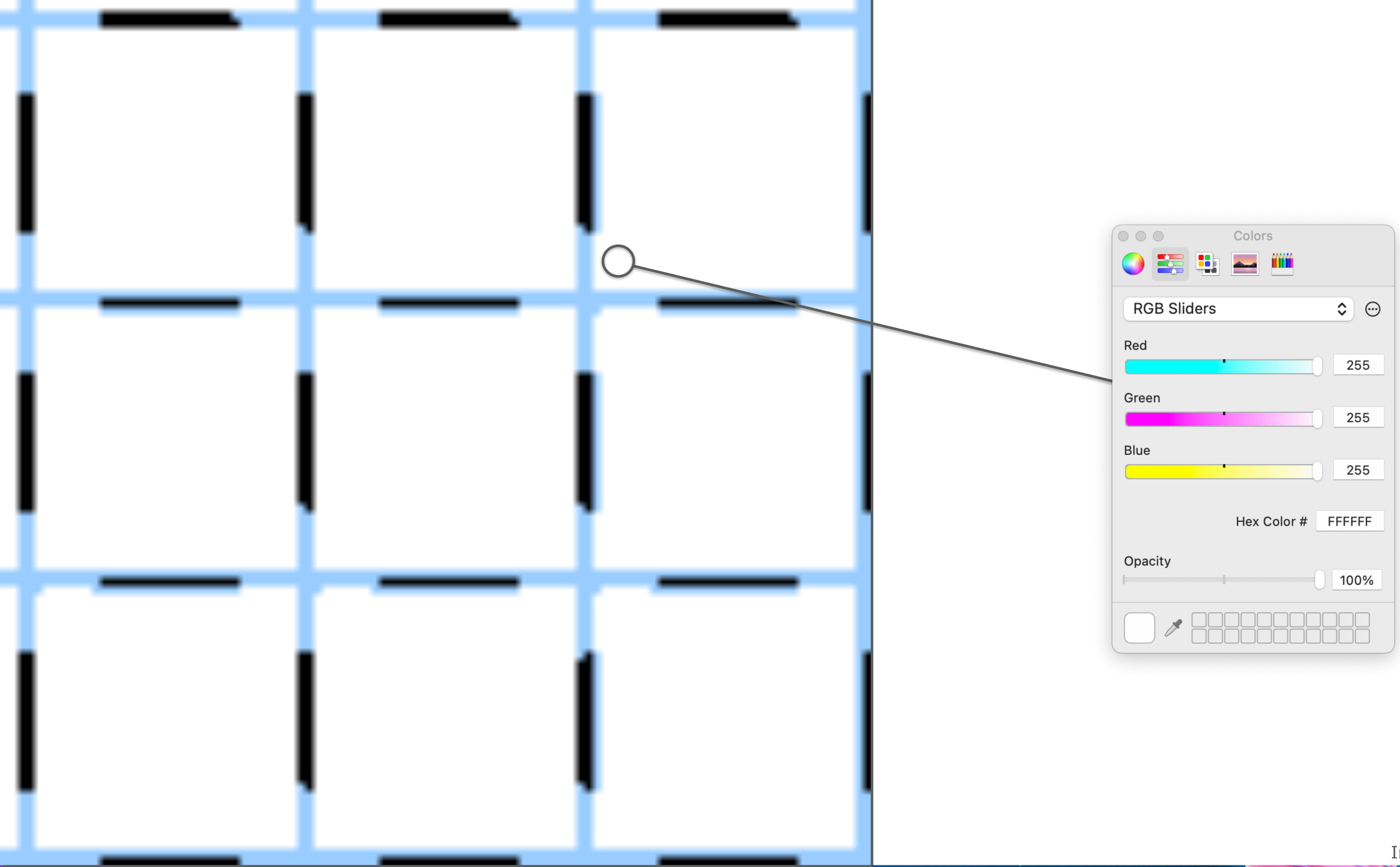







Colors
RGB Sliders
Red
255
Green
255
Blue
255
Hex Color #
FFFFFF
Opacity
100%



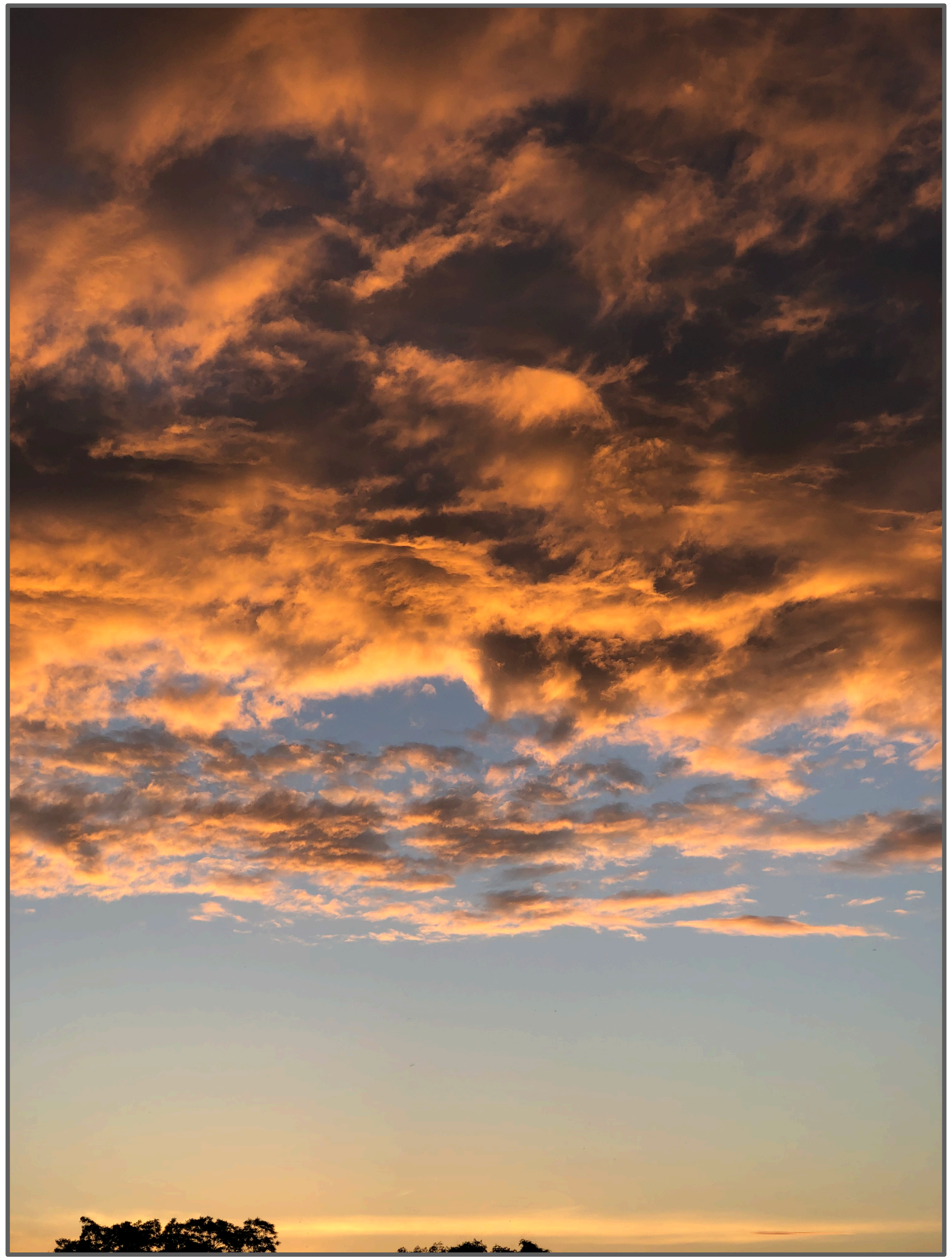

Colors
RGB Sliders
Red
255
Green
255
Blue
255
Hex Color #
FFFFFF
Opacity
100%

Why does it matter?

- Human vision, while effective, is not always accurate…
- A finer physics based simulation of reality has perceptual limits…

Image by Vincent Gerbouin from Pexels

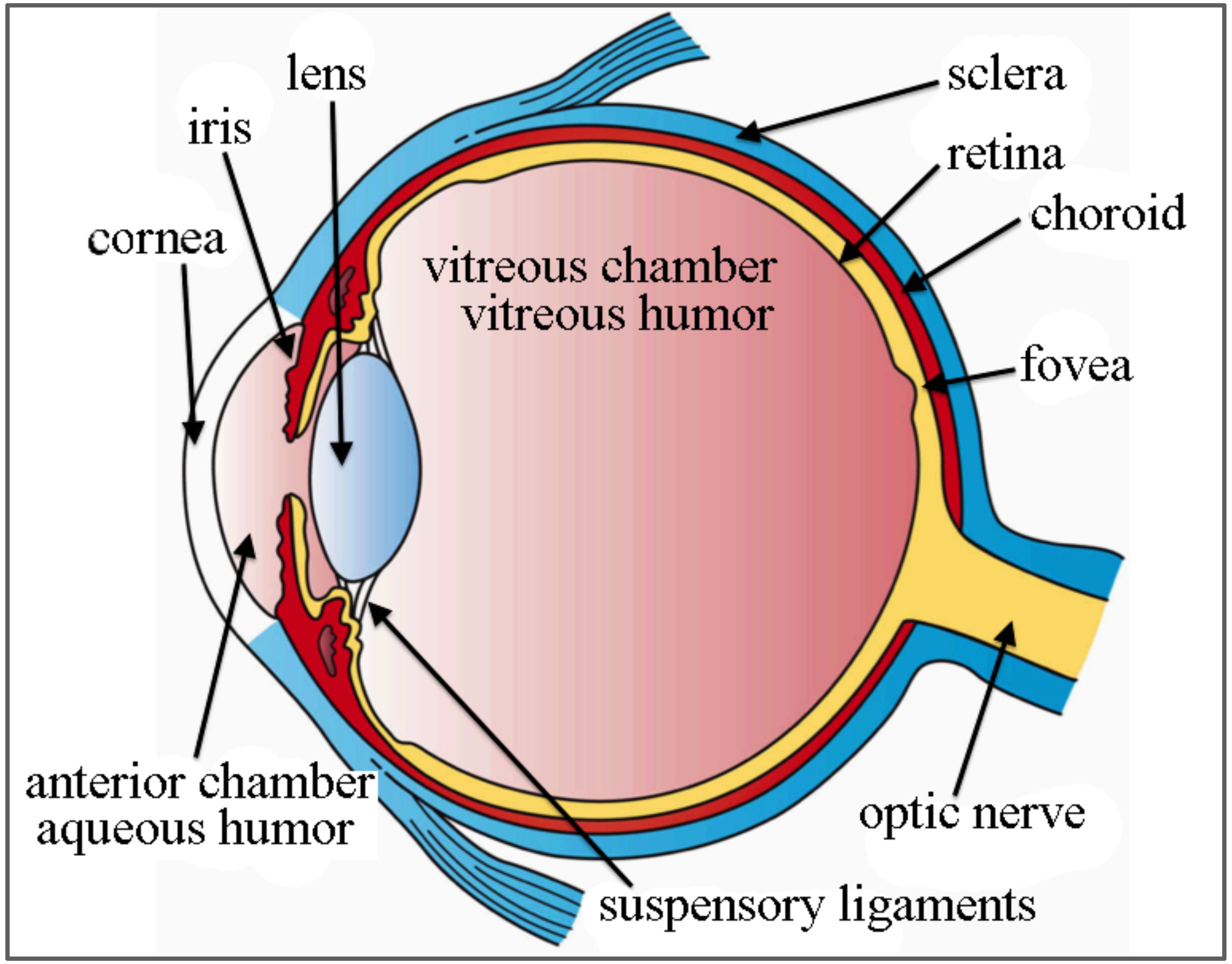
Image by Kanenori from Pixabay



Human Vision Hardware

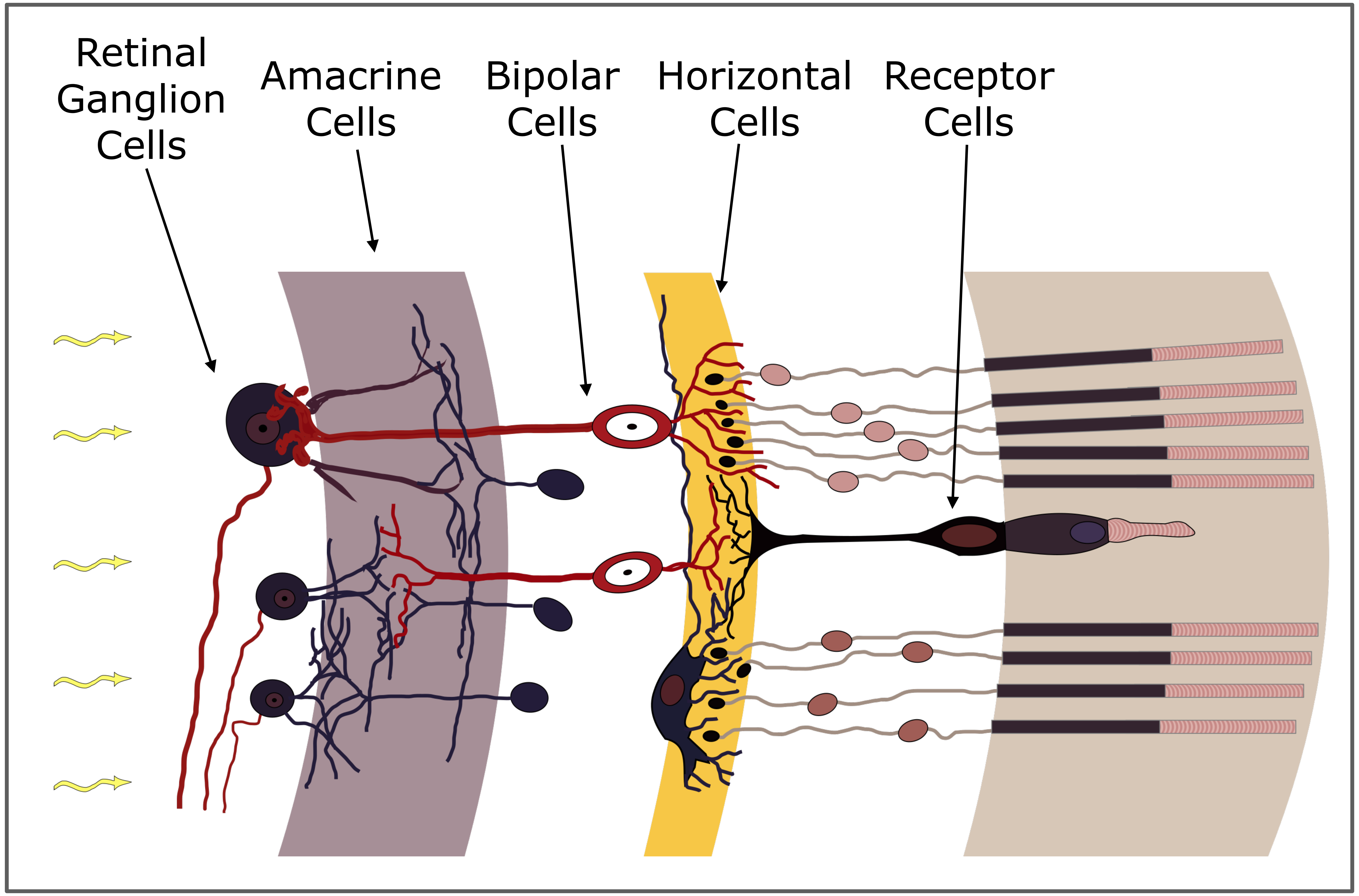
The Eye

- Incoming light is focused by a two lens system of cornea and the lens of variable curvature
- Light is projected on to the retina, primarily focused in the foveal pit
- Fovea is depressed to minimize light pollution and reduce light scattering
- There is a blind spot that we are psychologically unaware of



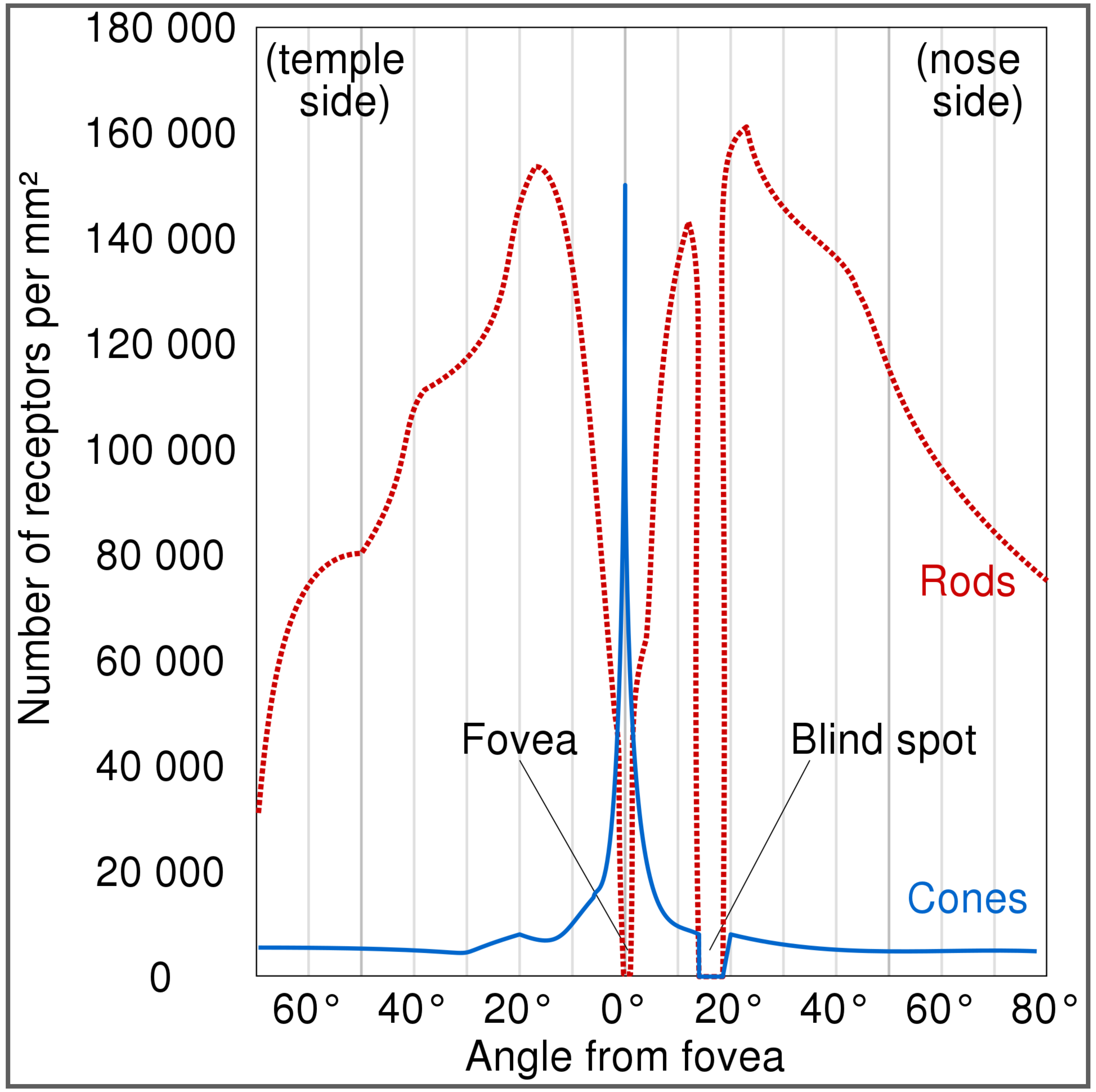
Retina

- Retina is multilayer
- The receptor layer is at the back (light passes through other layers first)
- Rod receptors (night or scotopic vision)
- Cone receptor (day or photopic vision)
- Horizontal and Amacrine cells provide lateral connections



Receptors

- Receptor density is the highest in the fovea and subsides in the periphery
- ~130M Receptors
  - ~120M Rods
  - ~6M Cones
- ~1M Ganglion cells (and hence optic channels to the brain)
  - In fovea about 1:1 Retinal Ganglion Cells (RGC) to receptors
  - Many Cones to a Ganglion cell in the periphery



Central and Peripheral Vision

- Foveal vision <5°
- At 2°, acuity is down to ~50%
- At 4°, it is down to ~30%
- Color perception strong at 20° but declining at 40° (not equally for all hues though)

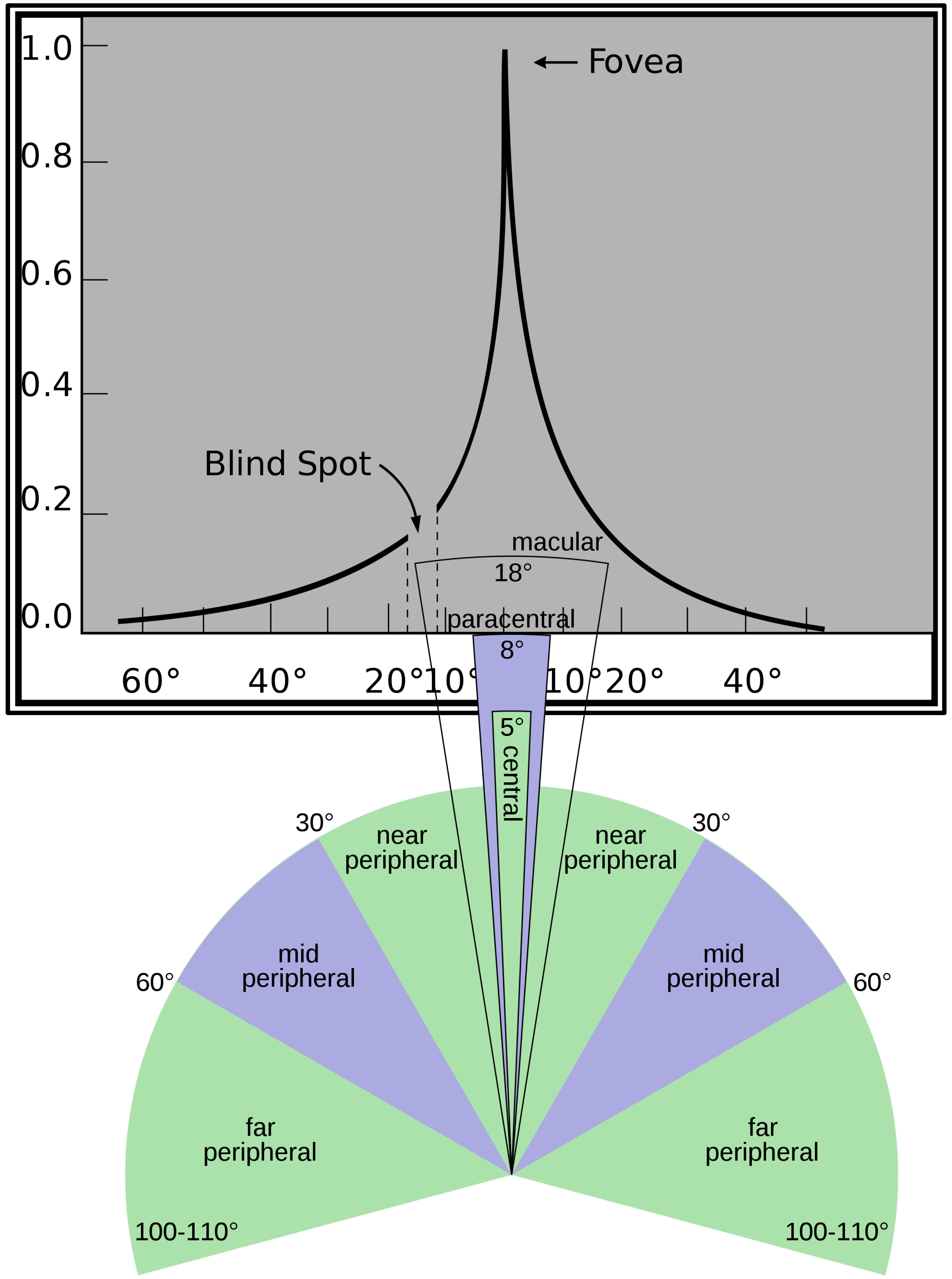
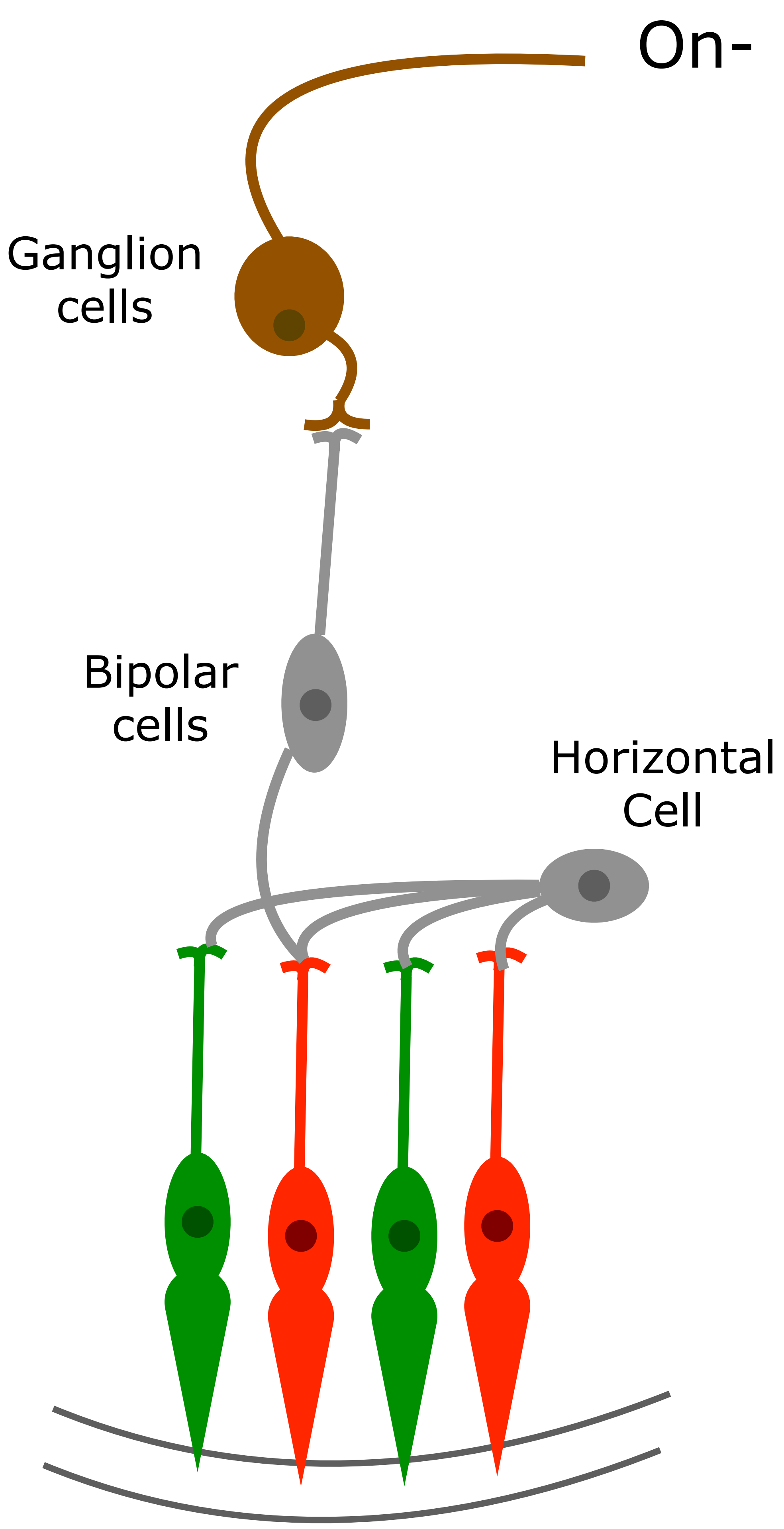
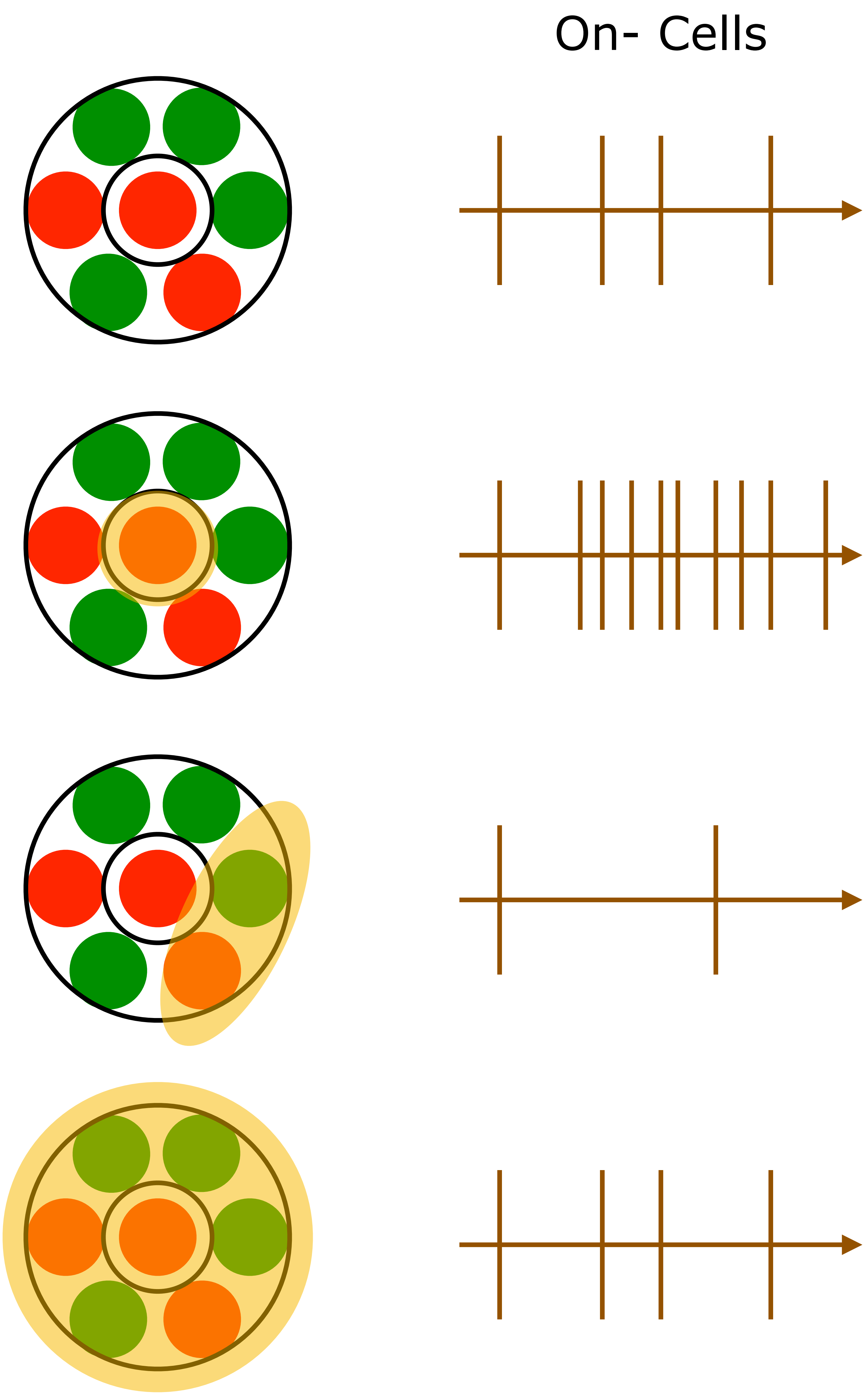


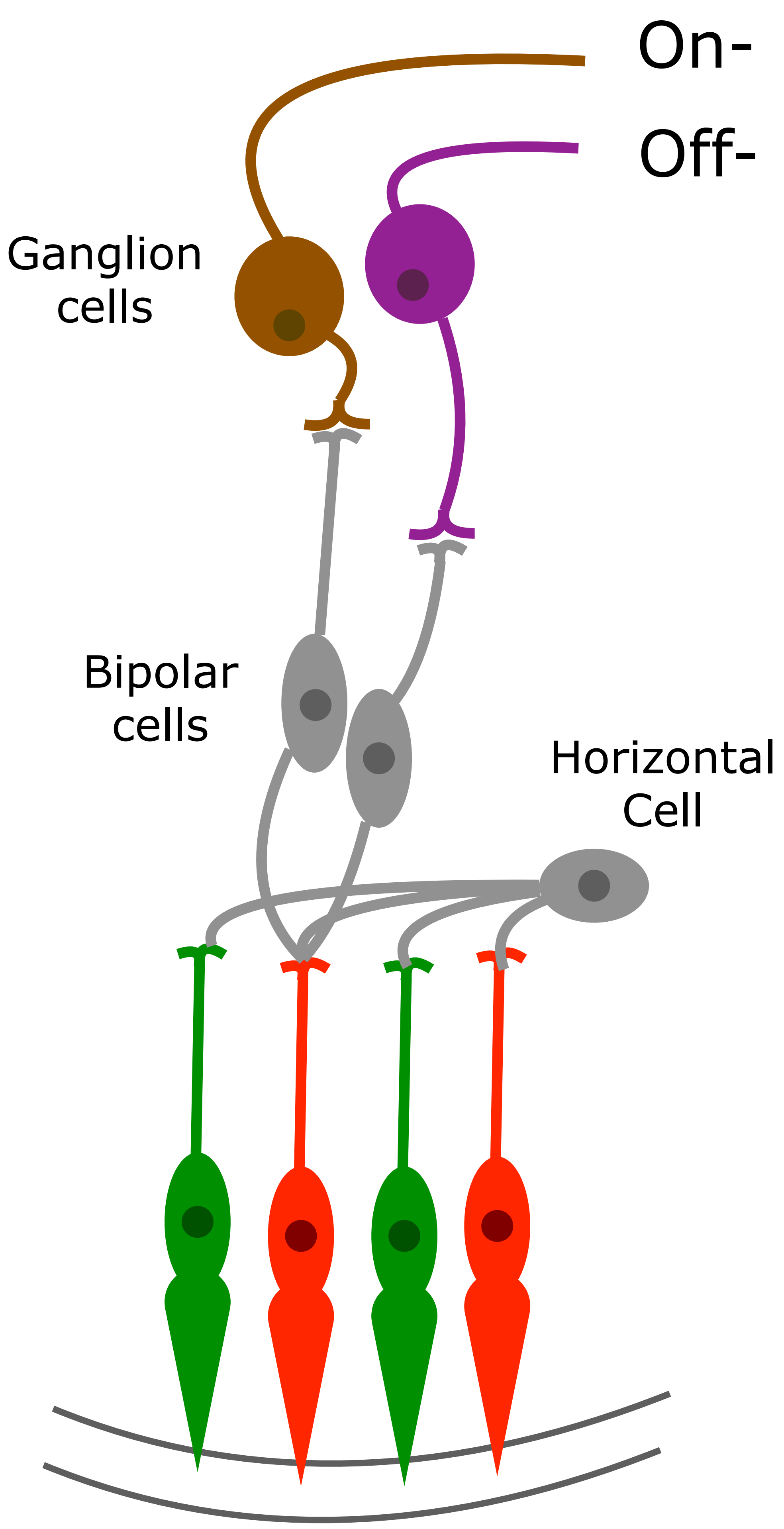
Image by Vanessa Ezekowitz BY-SA 3.0. Image by Zyxwv99 CC BY-SA 4.0

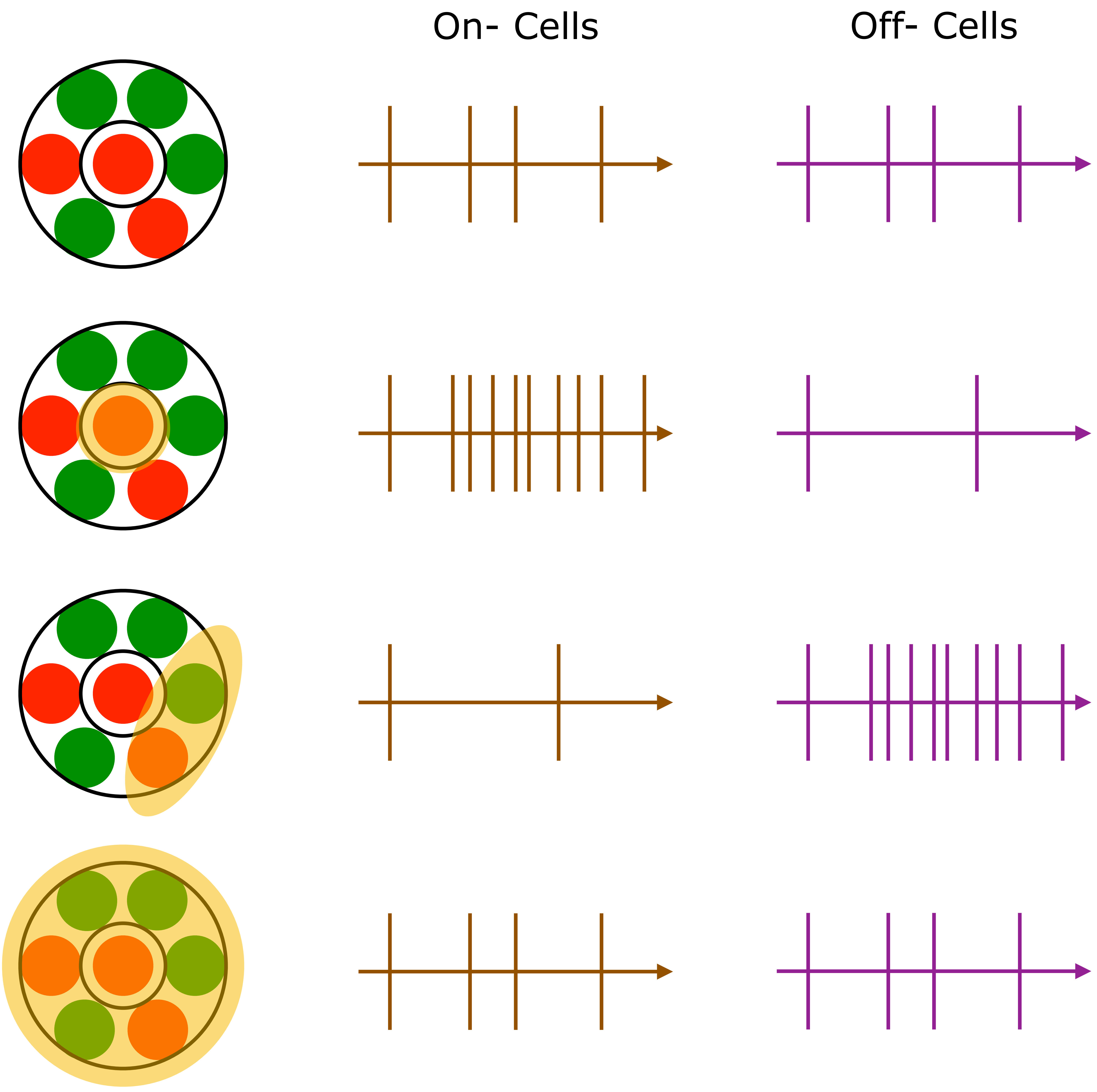
Cell Signals





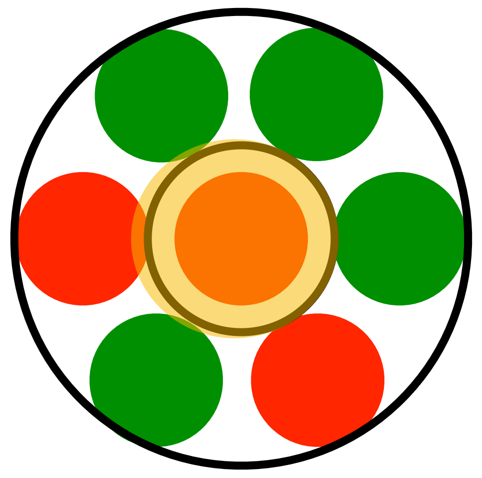
Cell Signals

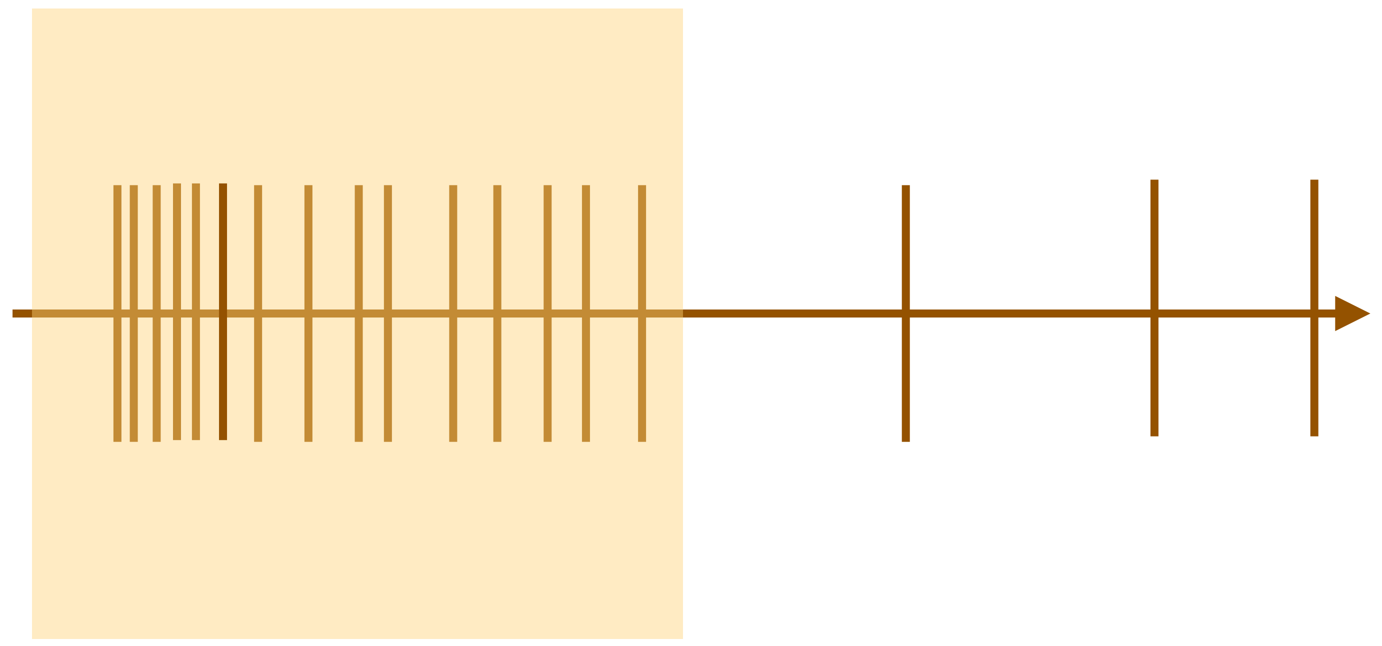




Pathways and Receptive Fields

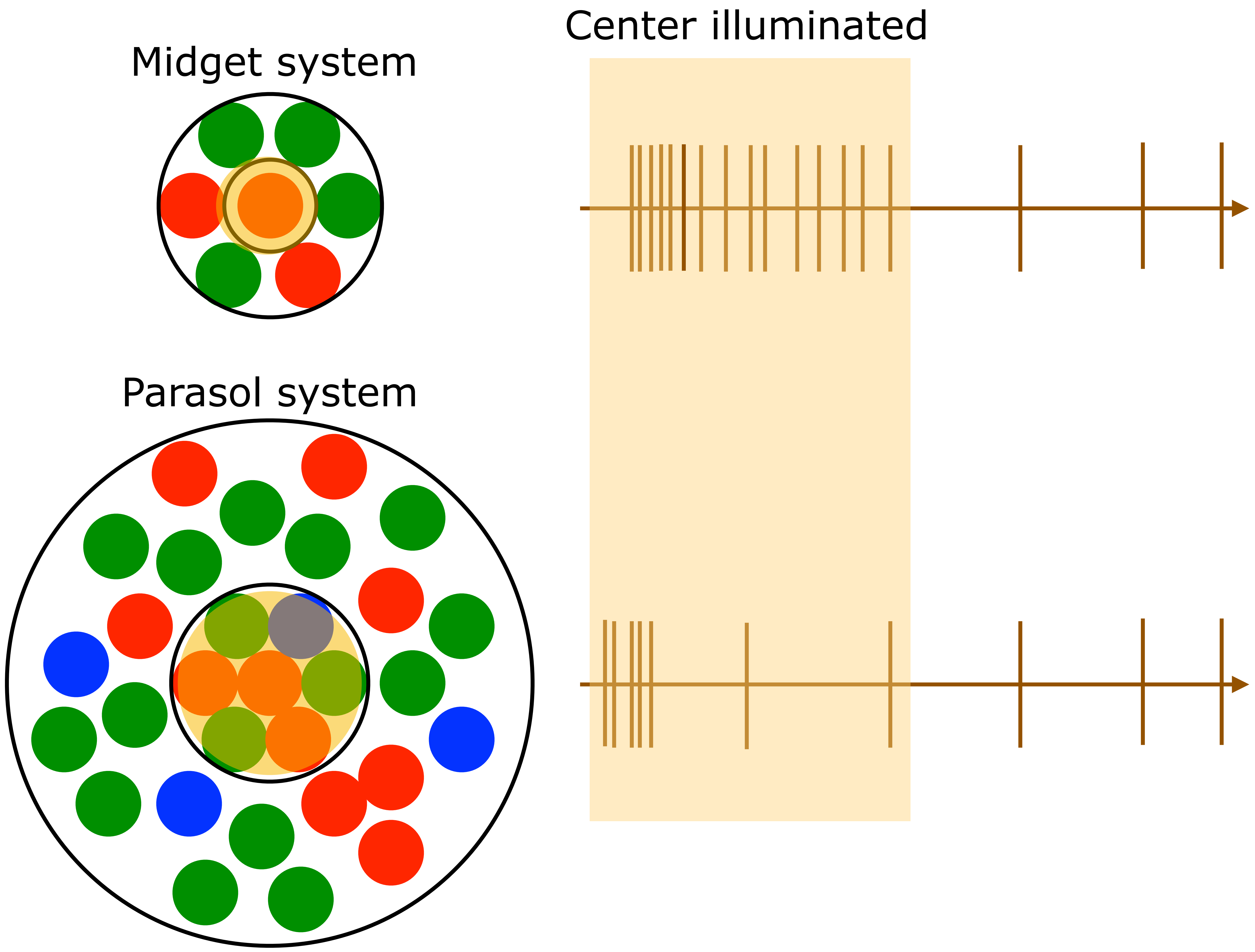
- There about 20 known types of Retinal Ganglion Cells (RGCs) with different receptive fields
- Midget/parvocellulare pathways (~90% of all RGCs)
  - Achromatic and chromatic vision
  - Slow temporal response
- Parasol/magnocellulare pathways (~5% of all RGCs)
  - Motion, change
  - Fast temporal response





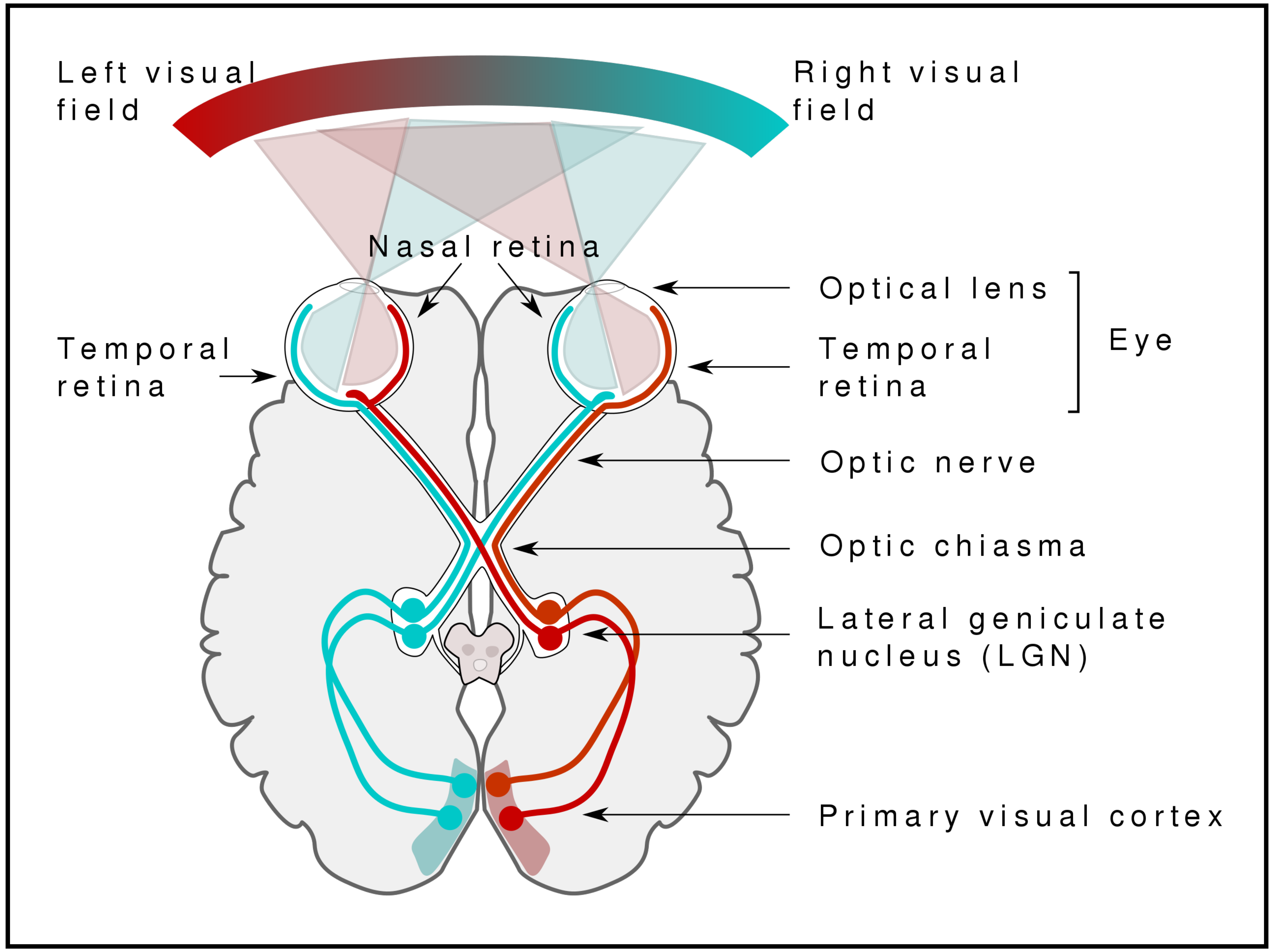
Pathways and Receptive Fields

- There about 20 known types of Retinal Ganglion Cells (RGCs) with different receptive fields
- Midget/parvocellulare pathways (~90% of all RGCs)
  - Achromatic and chromatic vision
  - Slow temporal response
- Parasol/magnocellulare pathways (~5% of all RGCs)
  - Motion, change
  - Fast temporal response



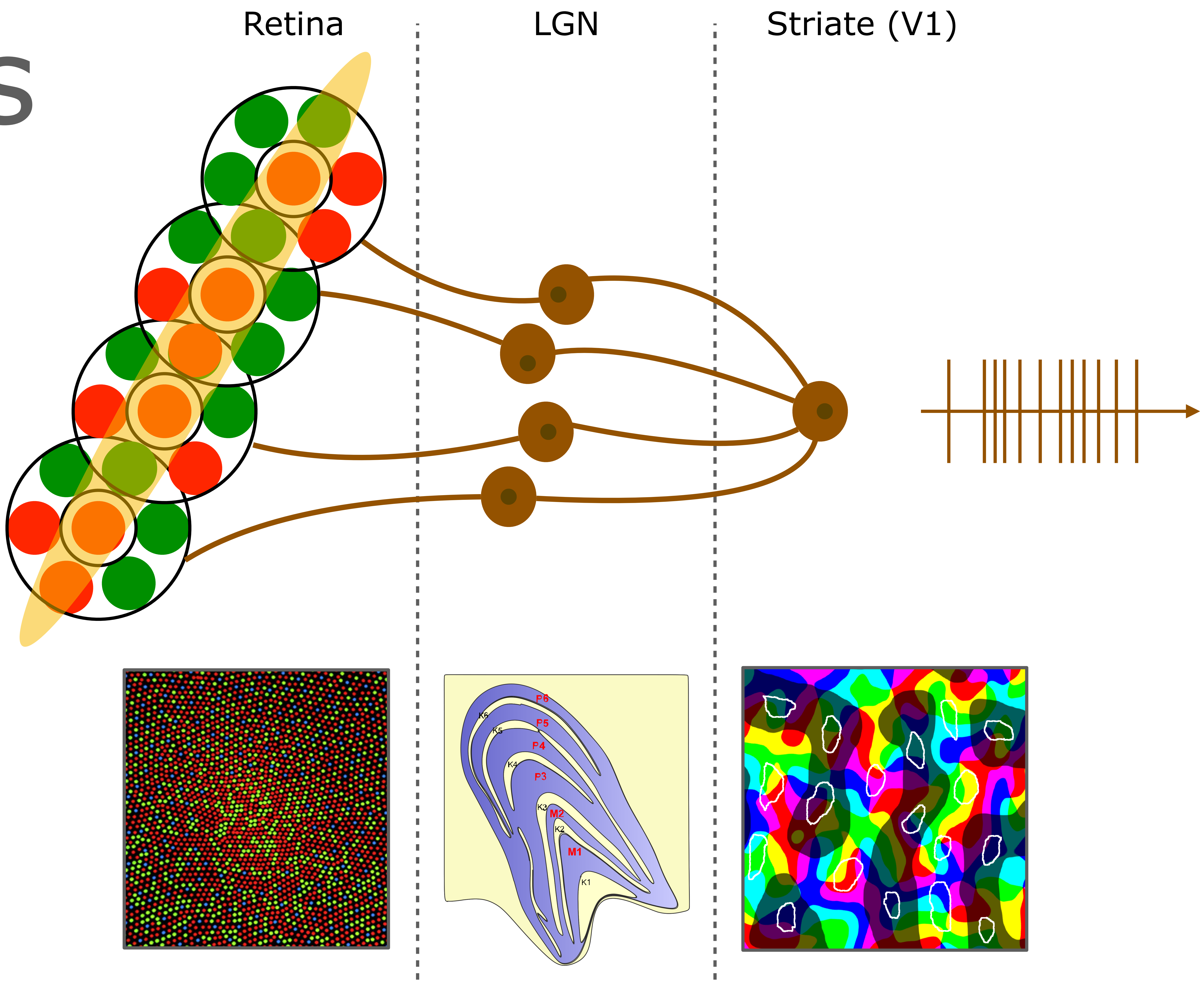
LGN and Striate Cortex

- Midget and Parasol cells project to different layers of Lateral Geniculate Nucleus (LGN)
- LGN projects into Striate cortex (V1 area)
- Left and Right visual fields are localized in opposite hemispheres
- Beyond Striate there is evidence for several visual pathways: ventral or “what” and dorsal or “where”



Simple Cells

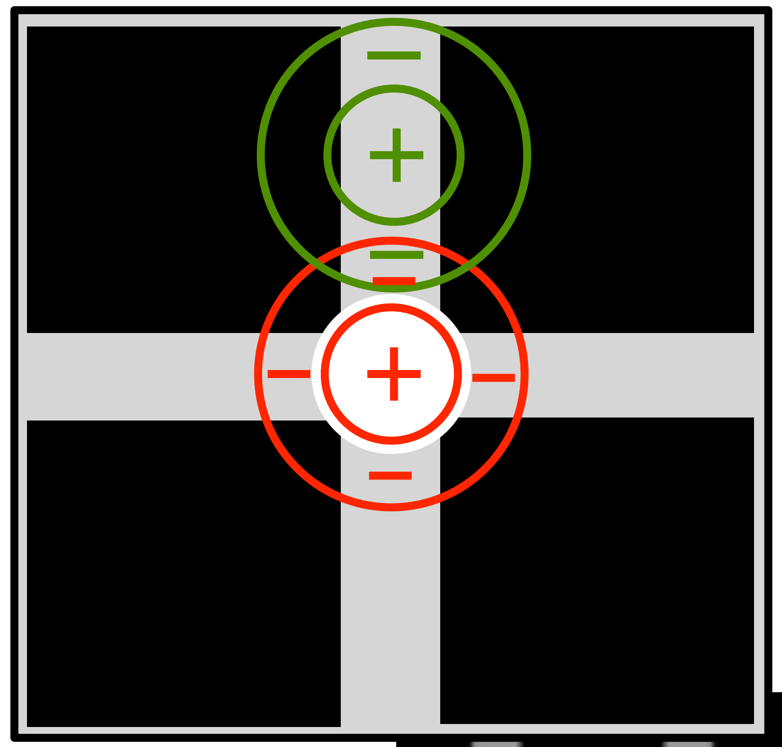
- Beyond LGN neural pathways feed into simple and complex cells in Striate cortex
- These cells respond to oriented stimuli and often to moving oriented stimuli
- There may be more cells responding to vertical and horizontal stimuli



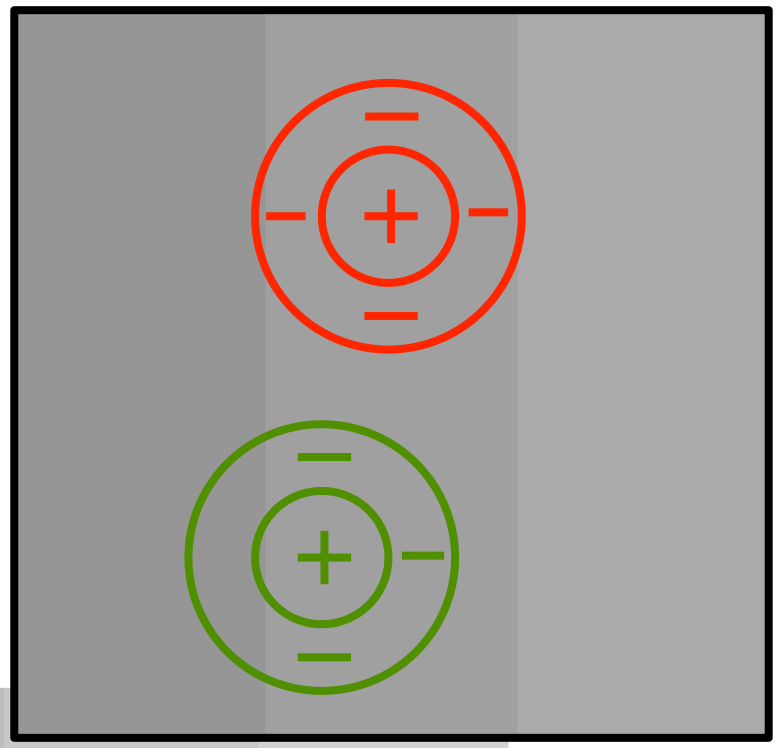
Why does it matter?

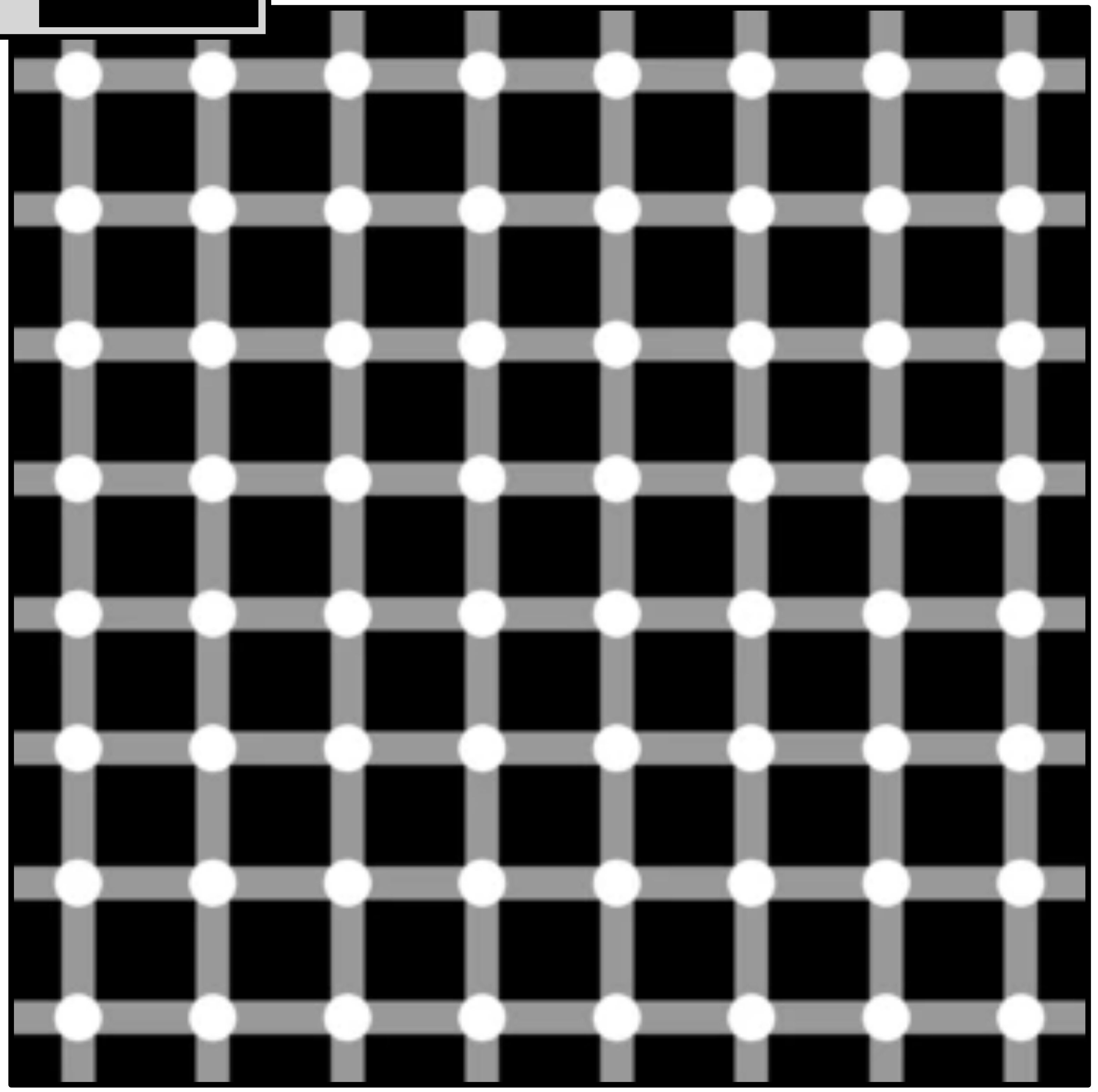
- Detail vision is very narrow, central and slow
- Vision is differential and over emphasizes edges
- Peripheral vision is not sensitive to detail but quite sensitive to motion and change

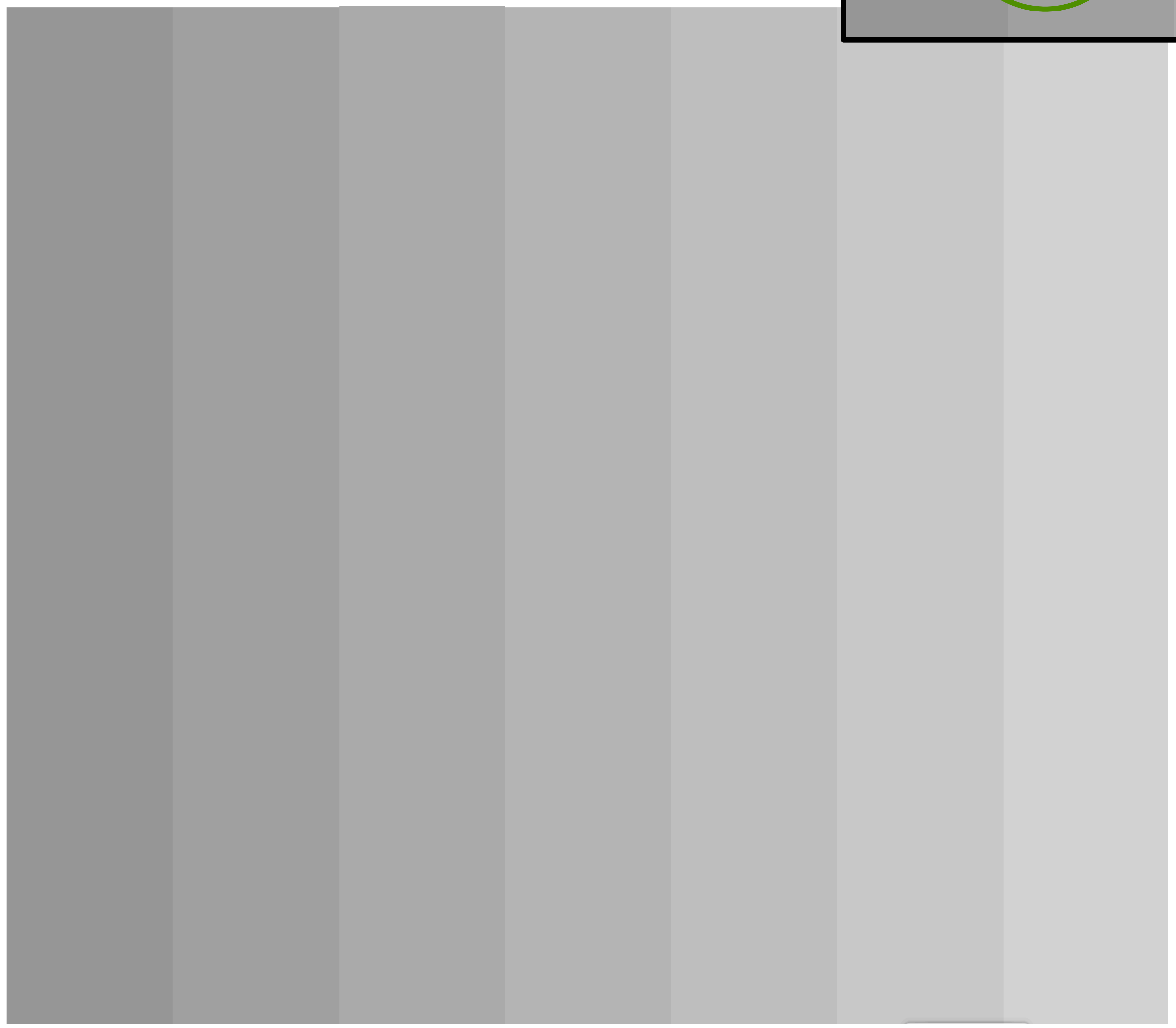
Lateral Inhibition

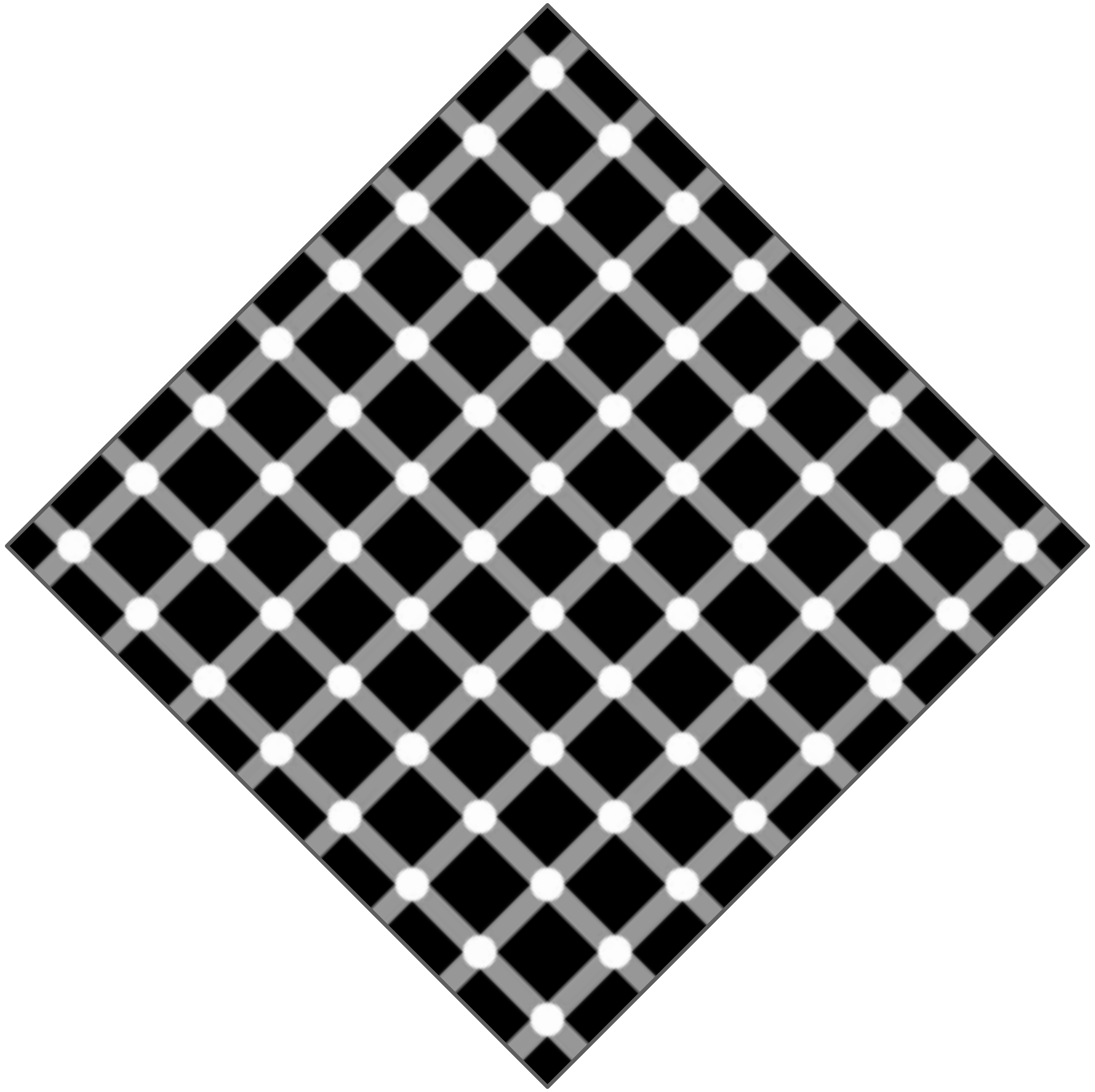


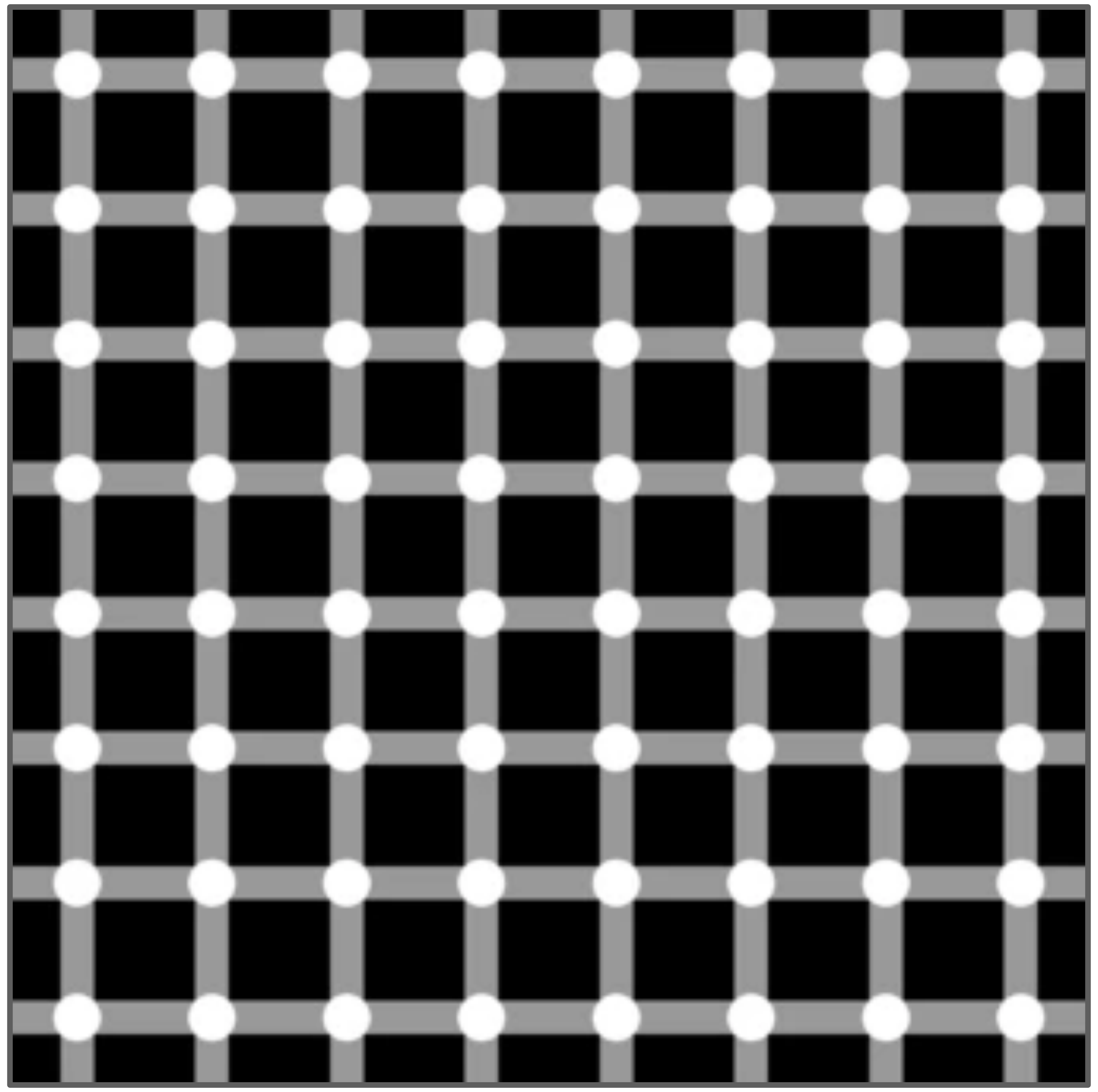
Lateral Inhibition?

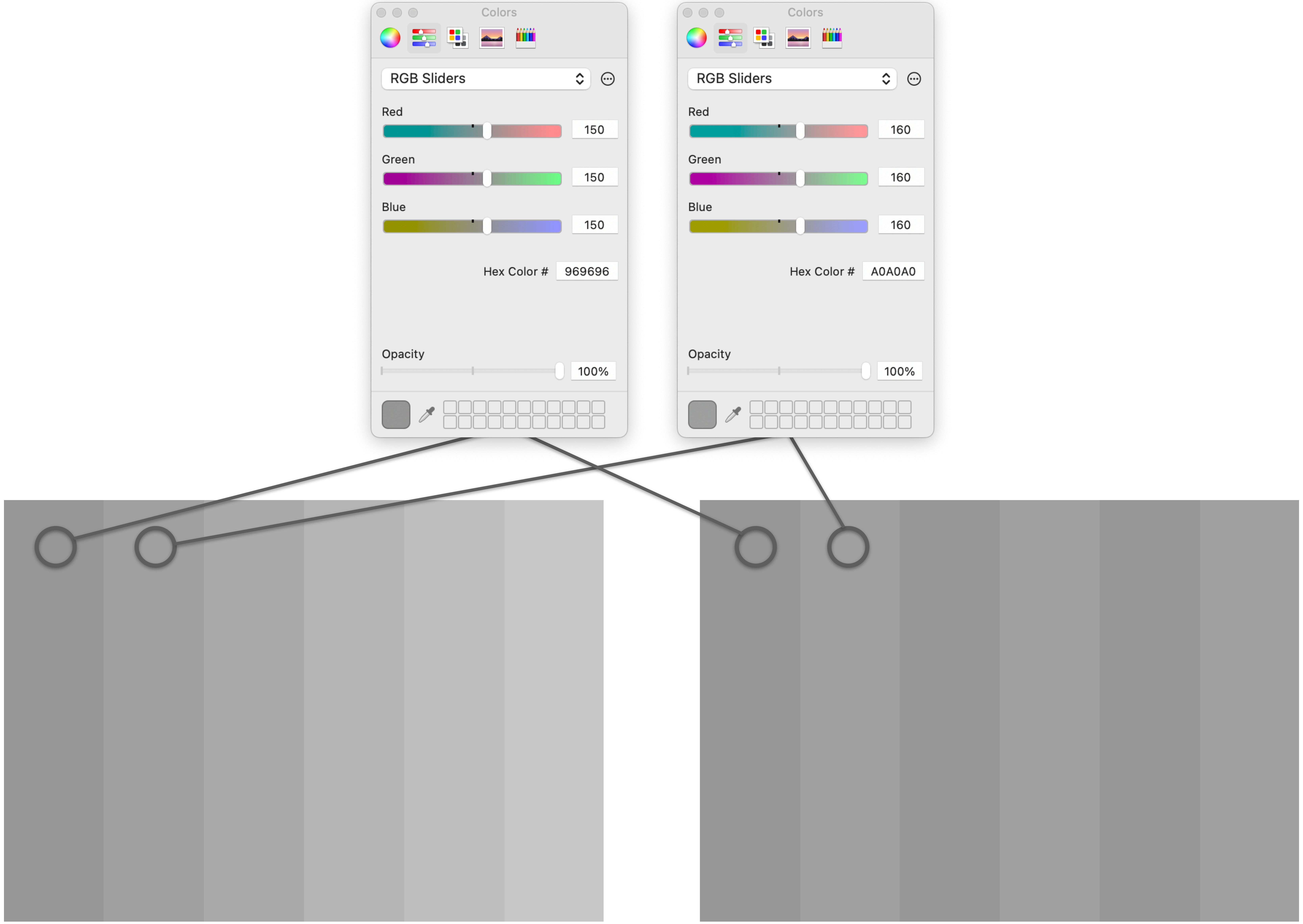


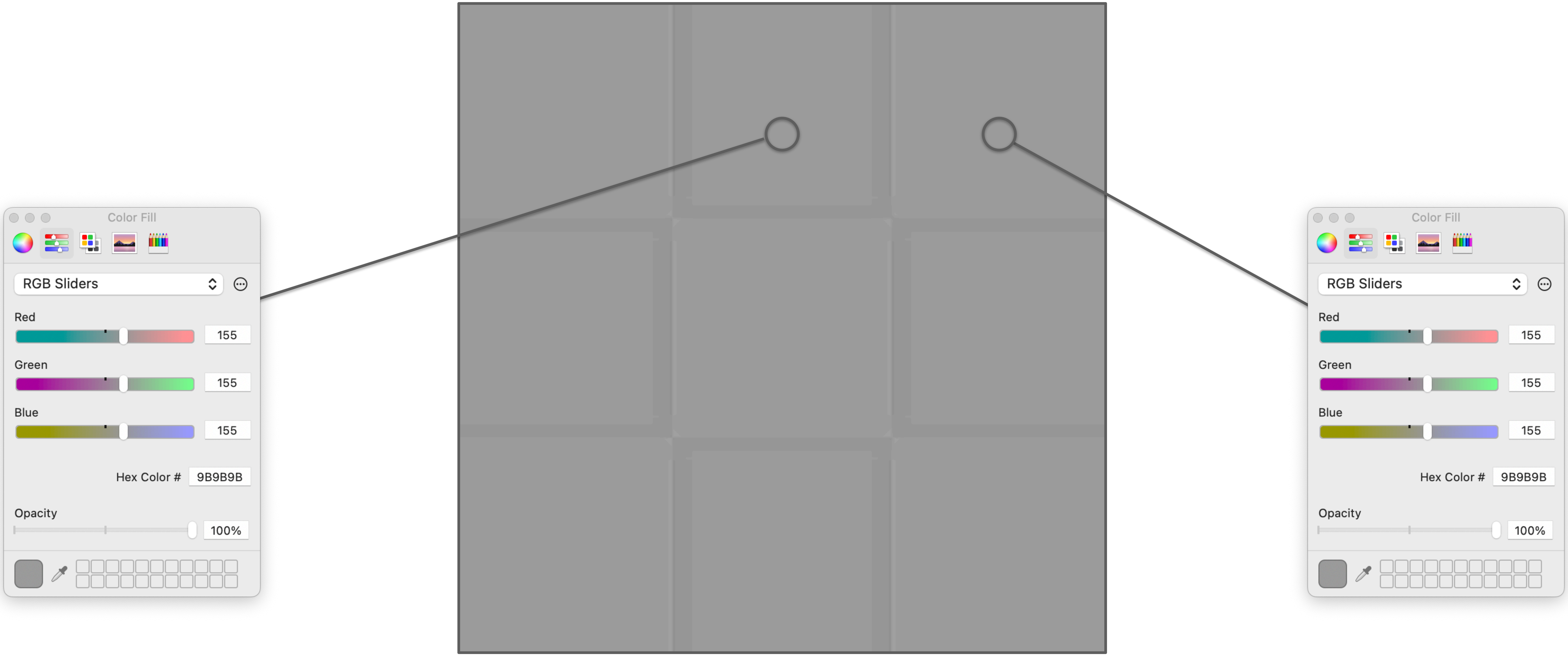


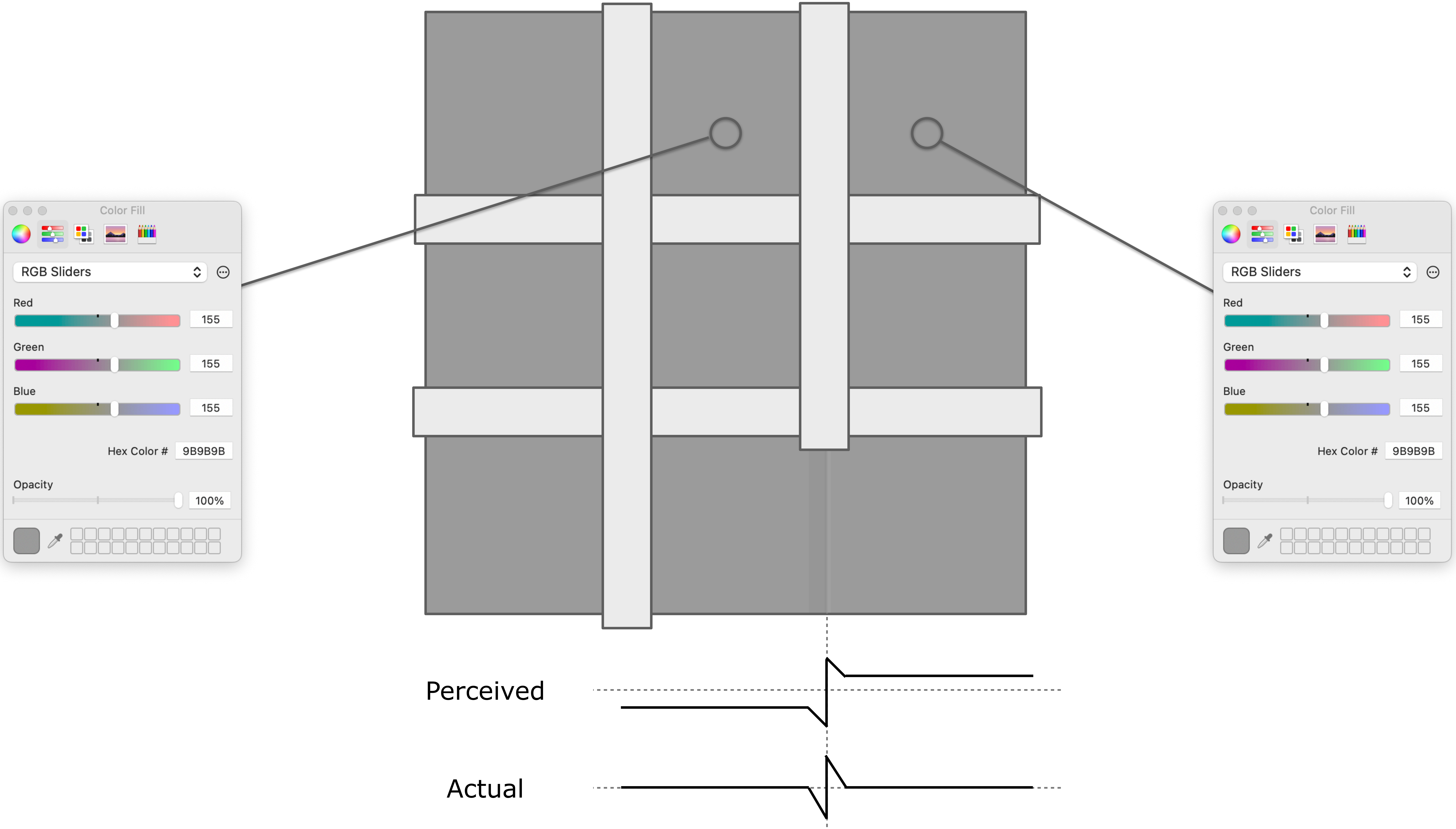







Colors
RGB Sliders
Red
150
Green
150
Blue
150
Hex Color # 969696
Opacity
100%
Colors
RGB Sliders
Red
160
Green
160
Blue
160
Hex Color # A0A0A0
Opacity
100%


Color Fill
RGB Sliders
Red
155
Green
155
Blue
155
Hex Color # 9B9B9B
Opacity
100%
Color Fill
RGB Sliders
Red
155
Green
155
Blue
155
Hex Color # 9B9B9B
Opacity
100%


Color Fill
RGB Sliders
Red
155
Green
155
Blue
155
Hex Color # 9B9B9B
Opacity
100%
Color Fill
RGB Sliders
Red
155
Green
155
Blue
155
Hex Color # 9B9B9B
Opacity
100%
Perceived
Actual

Reference

The Herman Grid Illusion Revisited

Peter H Schiller, Christina E Carvey

Perception, 2005

Perception, 2005, volume 34, pages 1375 – 1397

DOI:10.1068/p5447

The Hermann grid illusion revisited

Peter H Schiller, Christina E Carvey
Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; e-mail: phschill@mit.edu
Received 12 October 2004, in revised form 12 January 2005; published online 23 September 2005

Abstract. The Hermann grid illusion consists of smudges perceived at the intersections of a white grid presented on a black background. In 1960 the effect was first explained by a theory advanced by Baumgartner suggesting the illusory effect is due to differences in the discharge characteristics of retinal ganglion cells when their receptive fields fall along the intersections versus when they fall along non-intersecting regions of the grid. Since then, others have claimed that this theory might not be adequate, suggesting that a model based on cortical mechanisms is necessary [Lingelbach et al, 1985 *Perception* **14**(1) A7; Spillmann, 1994 *Perception* **23** 691 – 708; Geier et al, 2004 *Perception* **33** Supplement, 53; Westheimer, 2004 *Vision Research* **44** 2457 – 2465]. We present in this paper the following evidence to show that the retinal ganglion cell theory is untenable: (i) varying the makeup of the grid in a manner that does not materially affect the putative differential responses of the ganglion cells can reduce or eliminate the illusory effect; (ii) varying the grid such as to affect the putative differential responses of the ganglion cells does not eliminate the illusory effect; and (iii) the actual spatial layout of the retinal ganglion cell receptive fields is other than that assumed by the theory. To account for the Hermann grid illusion we propose an alternative theory according to which the illusory effect is brought about by the manner in which S1 type simple cells (as defined by Schiller et al, 1976 *Journal of Neurophysiology* **39** 1320 – 1333) in primary visual cortex respond to the grid. This theory adequately handles many of the facts delineated in this paper.

1 Introduction

The Hermann grid illusion (1870) in its best-known form consists of intersecting vertical and horizontal white bars superimposed on a black background, thereby forming an array of evenly spaced black squares. At the intersection of the bars, ghostly gray smudges are perceived comprising the illusion. The grid in this form is displayed in figure 1a. The smudges are seen everywhere except at the center of gaze. In figure 1b, the grid is displayed in reverse contrast; in this case white smudges are perceived at the intersections.

Over the years, the Hermann grid illusion has received considerable attention (for examples see Hering 1920; Baumgartner 1960; Spillmann and Levine 1971; Spillmann 1994; Ninio and Stevens 2000; De Lafuente and Ruiz 2004). Interest in the illusion was heightened when a clever hypothesis was advanced to explain the perception of the phantasmal smudges (Baumgartner 1960). We shall refer to this hypothesis as the retinal ganglion cell theory. As cited in several publications, the theory is laid out in figures 1c and 1d (Wolfe 1984; Sekuler and Blake 1994; Spillmann 1994). The explanation suggested is based on the findings made by neurophysiologists demonstrating that retinal ganglion cells have antagonistic center/surround organization (Kuffler 1953; Werblin and Dowling 1969; Schiller 1996). Consequently, when the grid consists of black squares and white bars, an ON-center retinal ganglion cell responds much more vigorously to a small bright spot placed into its receptive field center than to a large bright spot that activates both the center and the surround of the receptive field. The argument advanced was that smaller responses are elicited in the ON-center retinal ganglion cells whose receptive field centers fall into the intersections of the white bars than in cells whose receptive fields fall along non-intersecting regions of the bars.

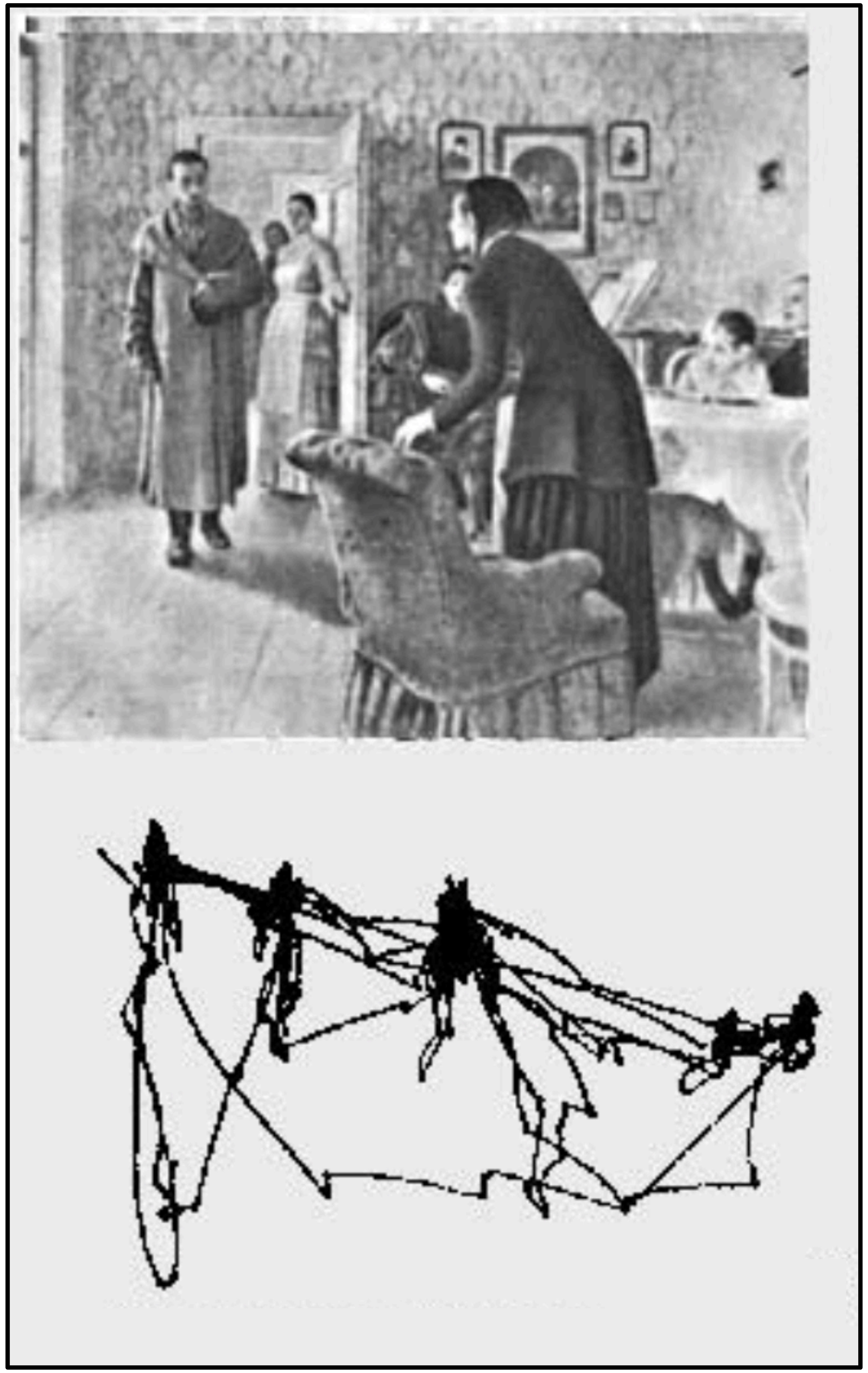
Why does it matter?

- Perception of a shape filler can be affected by the edges while perception of edges can be affected by overall scene
- Image artifacts are, sometimes, orientation dependent

Eye Movement

Saccadic Movement

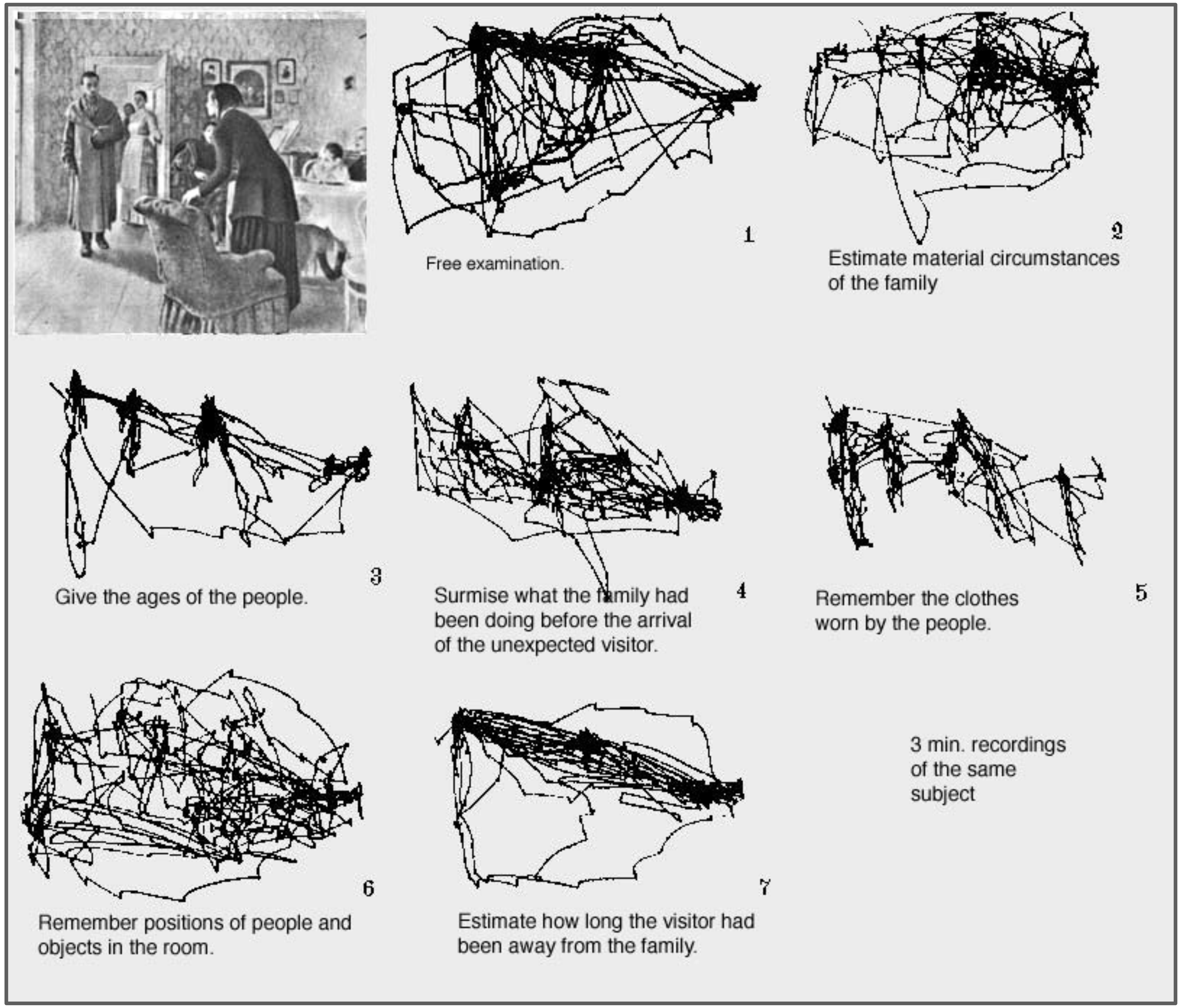
- Average saccade ~30ms
- functional blindness during the saccade
- Average fixation (with micro-saccade drift) ~200-300ms
- Smooth pursuit movement to chase moving targets



From Alfred Yarbus Eye Movement and Vision, 1967. Image by Lucs-Kho, Public Domain

Fixations

- Eye movement patterns are task specific
- Faces are almost always salient (even in peripheral vision!)
- Feet position may also be salient
- Shape interiors are rarely fixated on…



From Alfred Yarbus Eye Movement and Vision, 1967. Image by Lucs-Kho, Public Domain

Why does it matter?

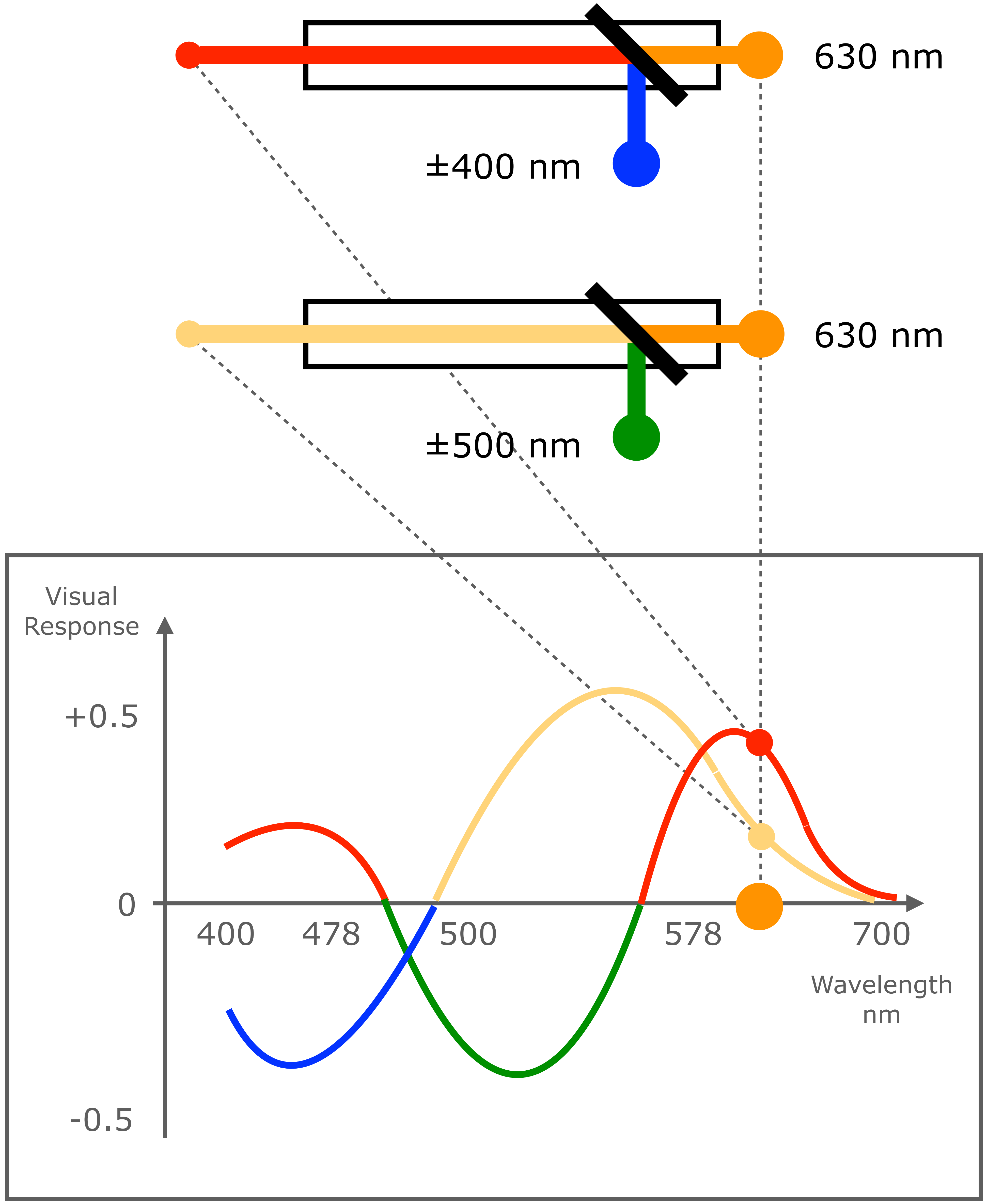
- We fixate on very few elements of the visual scene
- We are blind during saccadic eye movements and have limited memory of the visual scene before and after saccades and eye blinks

Color Vision

Psychophysics of hue perception

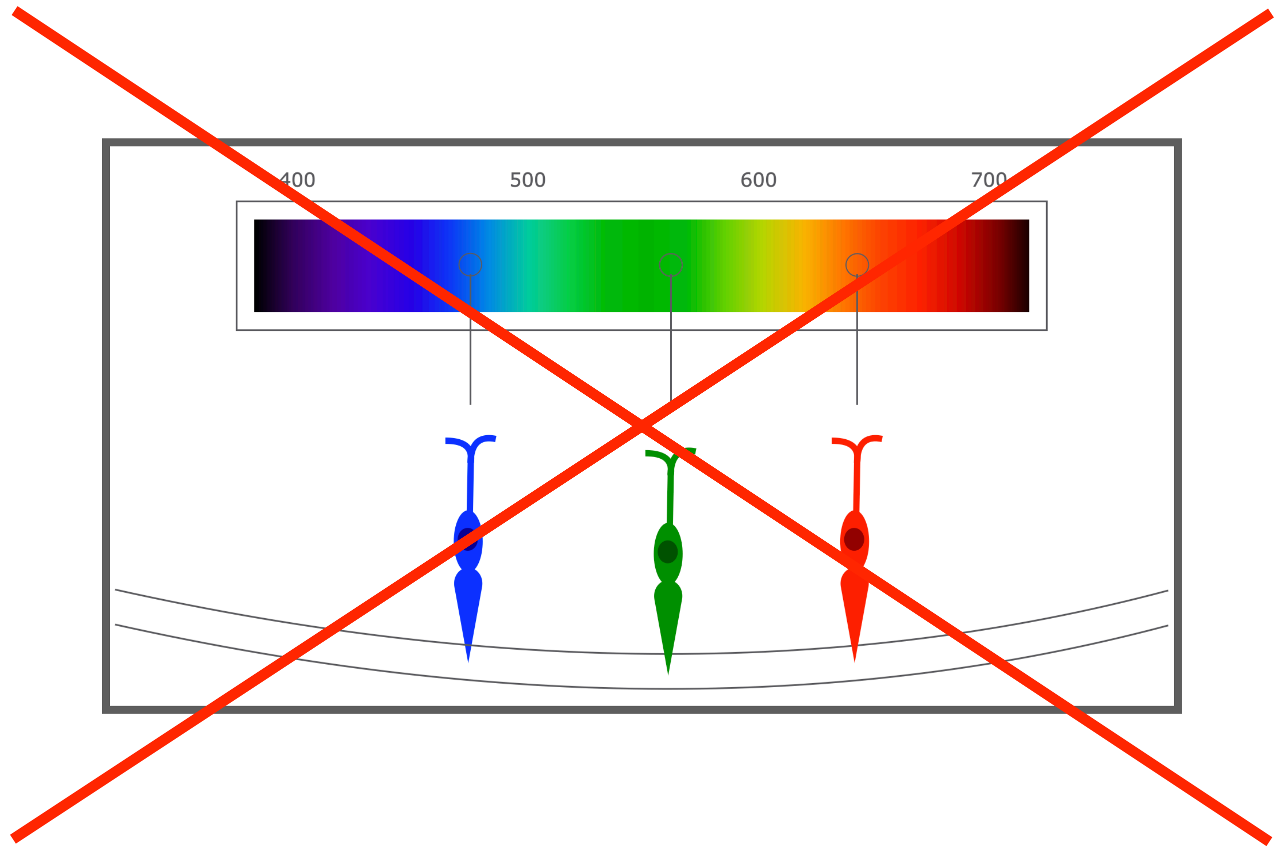
- Trichromatic color theory Young-Helmholtz 1802-1850
- Opponent color theory Herring 1892
- Color cancellation experiments Hurvich-Jameson 1957

- There are three main color receptor types and six main percepts (red, green, blue, yellow, white and black)
- There is perception opponency for red-green, blue-yellow

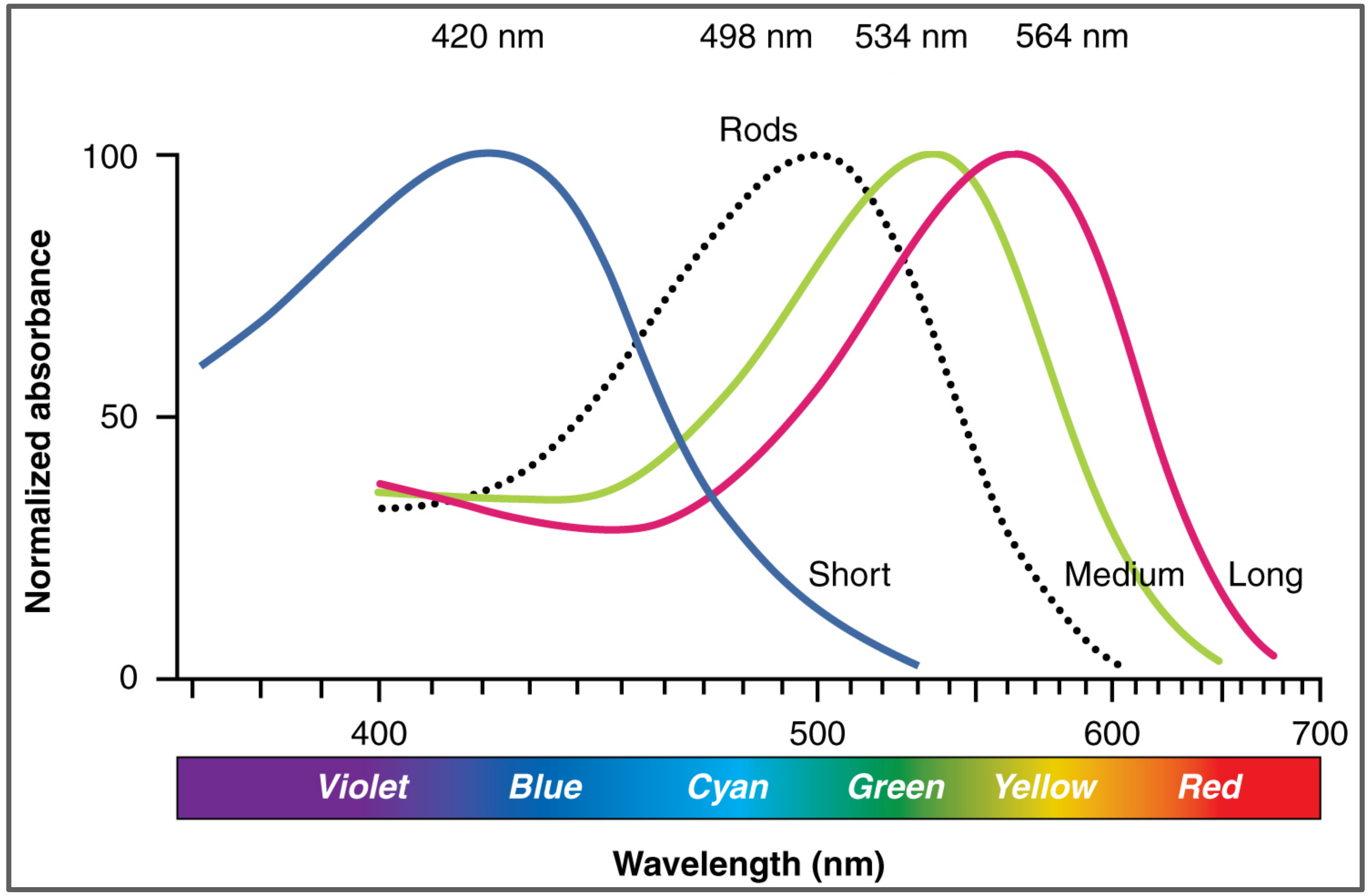


Adapted from Hurvich-Jameson 1957

Cone Receptors

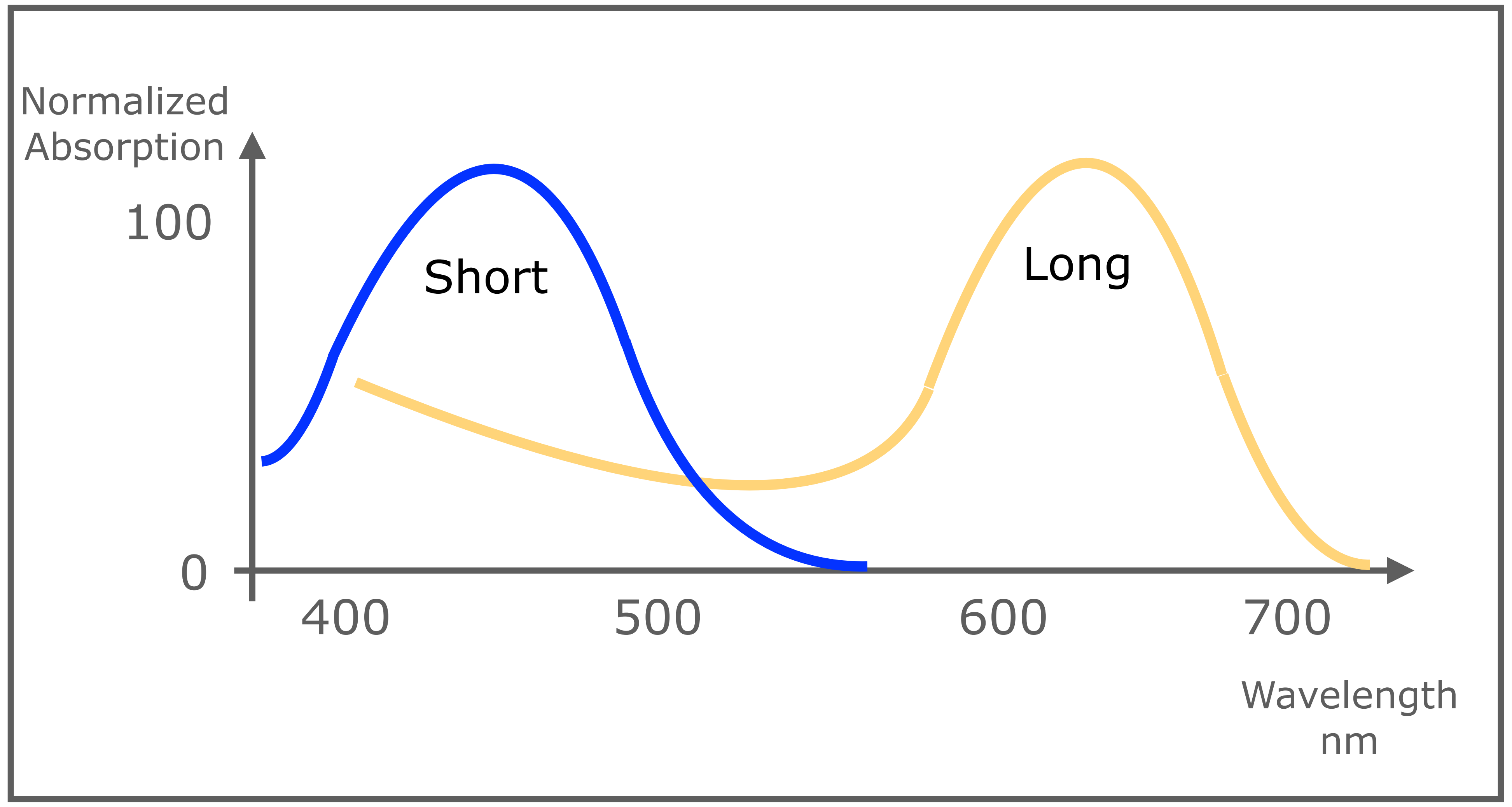


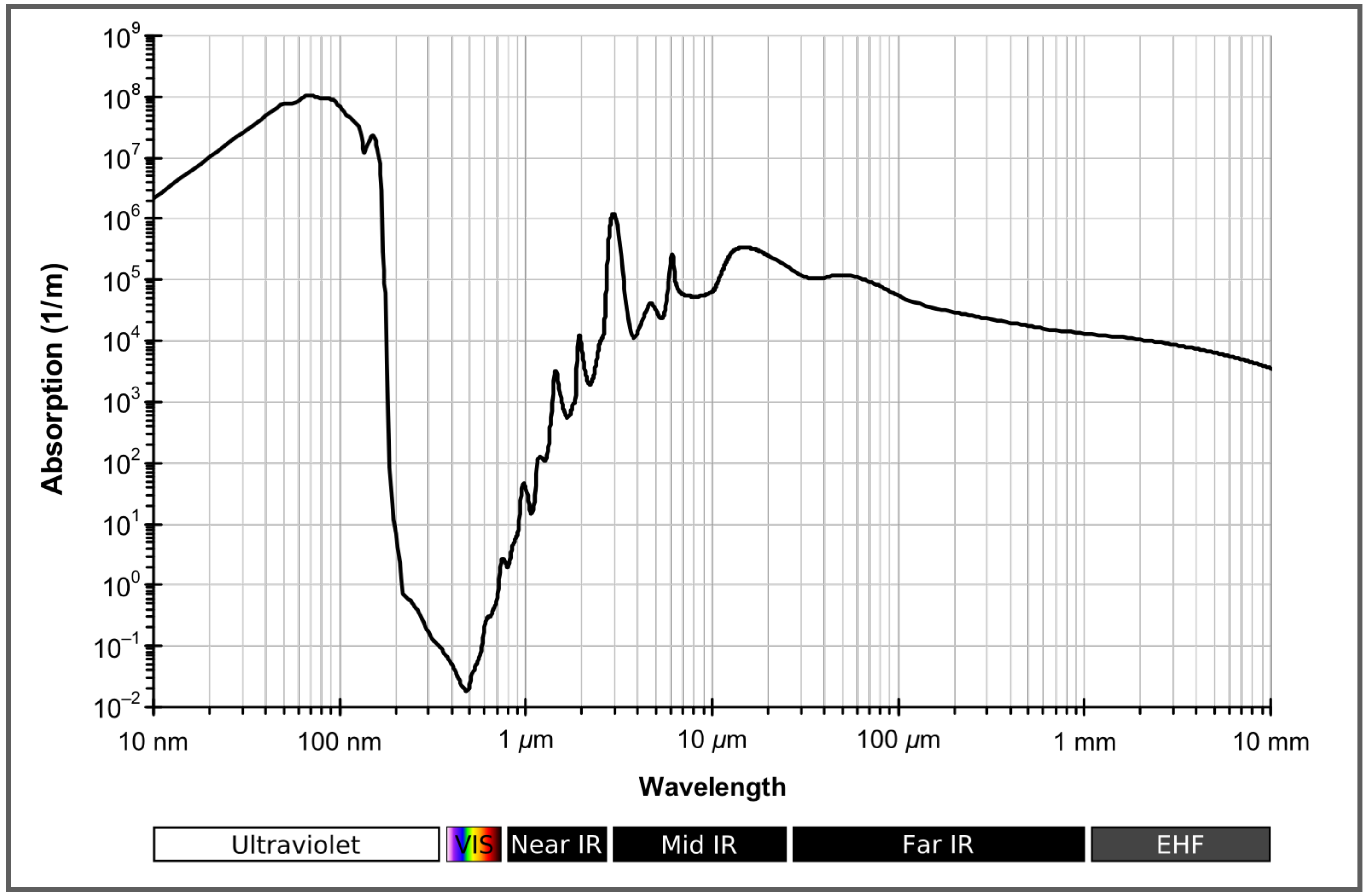
- Rod receptors (night or scotopic vision)
- Cone receptors (day or photopic vision)
- Long, Medium, Short wavelength cone receptors rather than Red, Green and Blue
- Chromatic aberration issues: S vs L and M



Visible Spectrum

- Most mammals are dichromates with Blue/Yellow vision
- This probably first appeared in ocean environment
- Visible spectrum doesn't get absorbed by water as much (particularly shorter wavelengths)





Reference

Evolution of the circuitry for conscious color vision in primates

J Neitz and M Neitz

Eye, 2017

Eye (2017) 31, 286–300

www.nature.com/eye

RCOPHTH EPONYMOUS LECTURE

Evolution of the circuitry for conscious color vision in primates

J Neitz and M Neitz

Department of Ophthalmology, School of Medicine, Vision Sciences at South Lake Union/ Ophthalmology, University of Washington, Seattle, WA, USA

Correspondence: J Neitz, Department of Ophthalmology, School of Medicine, Vision Sciences at South Lake Union/ Ophthalmology, University of Washington, Box 358058, 750 Republican Street, Building E Room 184, Seattle, WA 98109, USA
Tel: +1 206 5438065;
Fax: +1 206 6859315.
E-mail: jneitz@uw.edu

Received: 16 September 2016
Accepted: 6 October 2016
Published online: 9 December 2016

Abstract

There are many ganglion cell types and subtypes in our retina that carry color information. These have appeared at different times over the history of the evolution of the vertebrate visual system. They project to several different places in the brain and serve a variety of purposes allowing wavelength information to contribute to diverse visual functions. These include circadian photoentrainment, regulation of sleep and mood, guidance of orienting movements, detection and segmentation of objects. Predecessors to some of the circuits serving these purposes presumably arose before mammals evolved and different functions are represented by distinct ganglion cell types. However, while other animals use color information to elicit motor movements and regulate activity rhythms, as do humans, using phylogenetically ancient circuitry, the ability to appreciate color appearance may have been refined in ancestors to primates, mediated by a special set of ganglion cells that serve only that purpose. Understanding the circuitry for color vision has implications for the possibility of treating color blindness using gene therapy by recapitulating evolution. In addition, understanding how color is encoded, including how chromatic and achromatic percepts are separated is a step toward developing a complete picture of the diversity of ganglion cell types and their functions. Such knowledge could be useful in developing therapeutic strategies for blinding eye disorders that rely on stimulating elements in the retina, where more than 50 different neuron types are organized into circuits that transform signals from photoreceptors into specialized detectors many of which are not directly involved in conscious vision.
Eye (2017) **31,** 286–300; doi:10.1038/eye.2016.257; published online 9 December 2016

Introduction

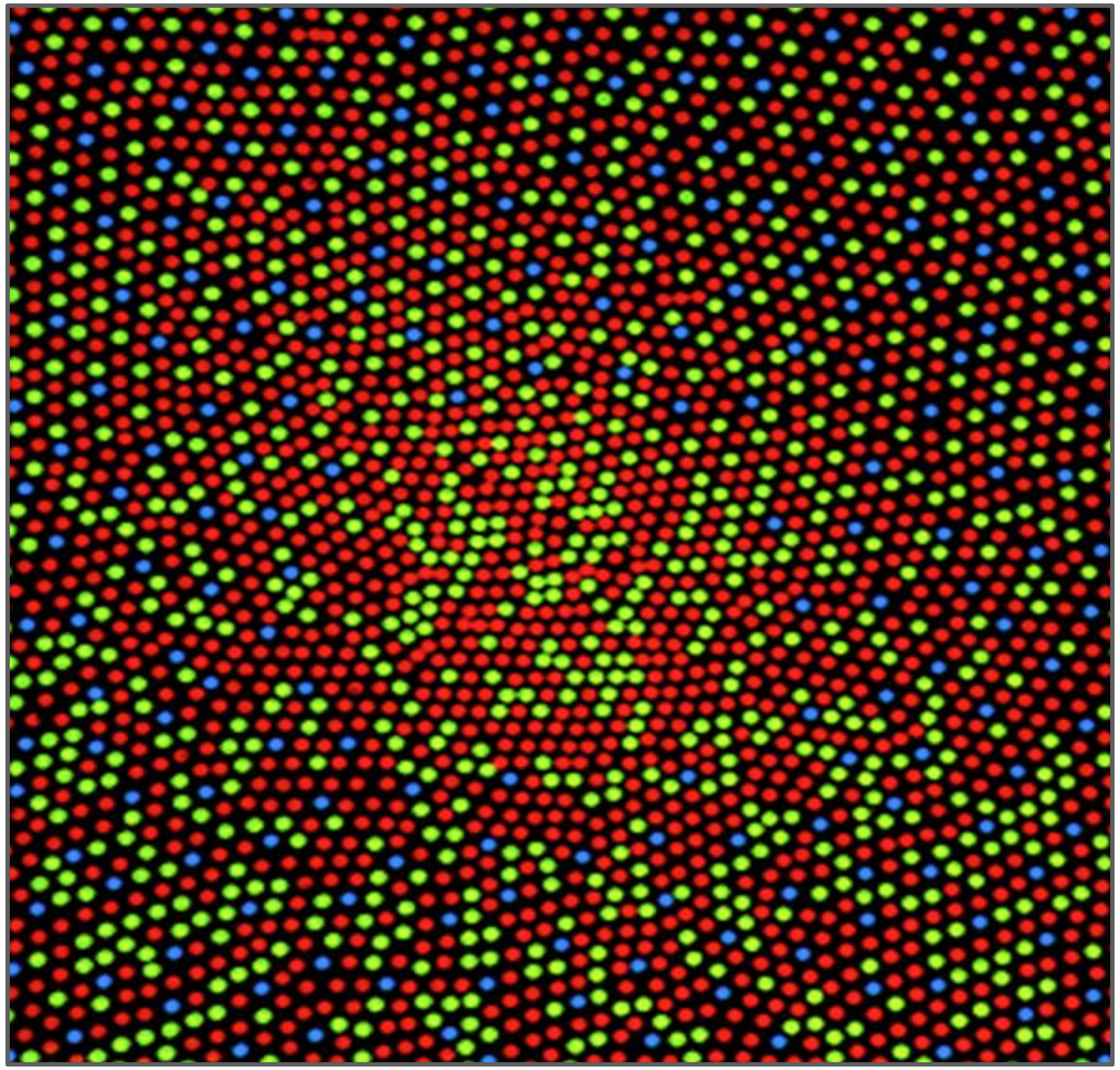
Color is the perception associated with the spectral composition of light. The question addressed here is what is the circuitry responsible for conscious color 'perception'. We are concerned with the neural machinery responsible for the hues, red, green blue and yellow, and how they are separated from black and white. Countless ideas have been proposed relating to the neural underpinnings for human color perception. The goal here is not to add new ideas, but rather to examine evidence from experiments in combination with consideration of constraints from evolution to determine which ideas are most likely to be true. From those we attempt to synthesize the best possible current explanation of the physiological mechanisms underlying human color perception.

Understanding the circuitry for color vision is important because it helps explain our conscious and unconscious reactions to colored stimuli. It can also explain the remarkable agreement across people about the appearance of some colors in the face of differences in our physiology and the exceptional disagreement we have about other colors. In addition, we can better understand how color vision deficiencies differ from normal vision and consider the prospects for curing them with gene therapy.

Taking an empirical approach to differentiating theories of color vision was a theme of the work of Frederick William Edridge-Green (1863–1953) as he performed experiments illuminating what it means to be color blind and tried to discover the best ways color vision deficiencies should be tested. We were very privileged to give the Edridge-Green lecture at The Royal College of Ophthalmologists Annual Congress at Birmingham in May 2016. This article is based on the material presented at that occasion.

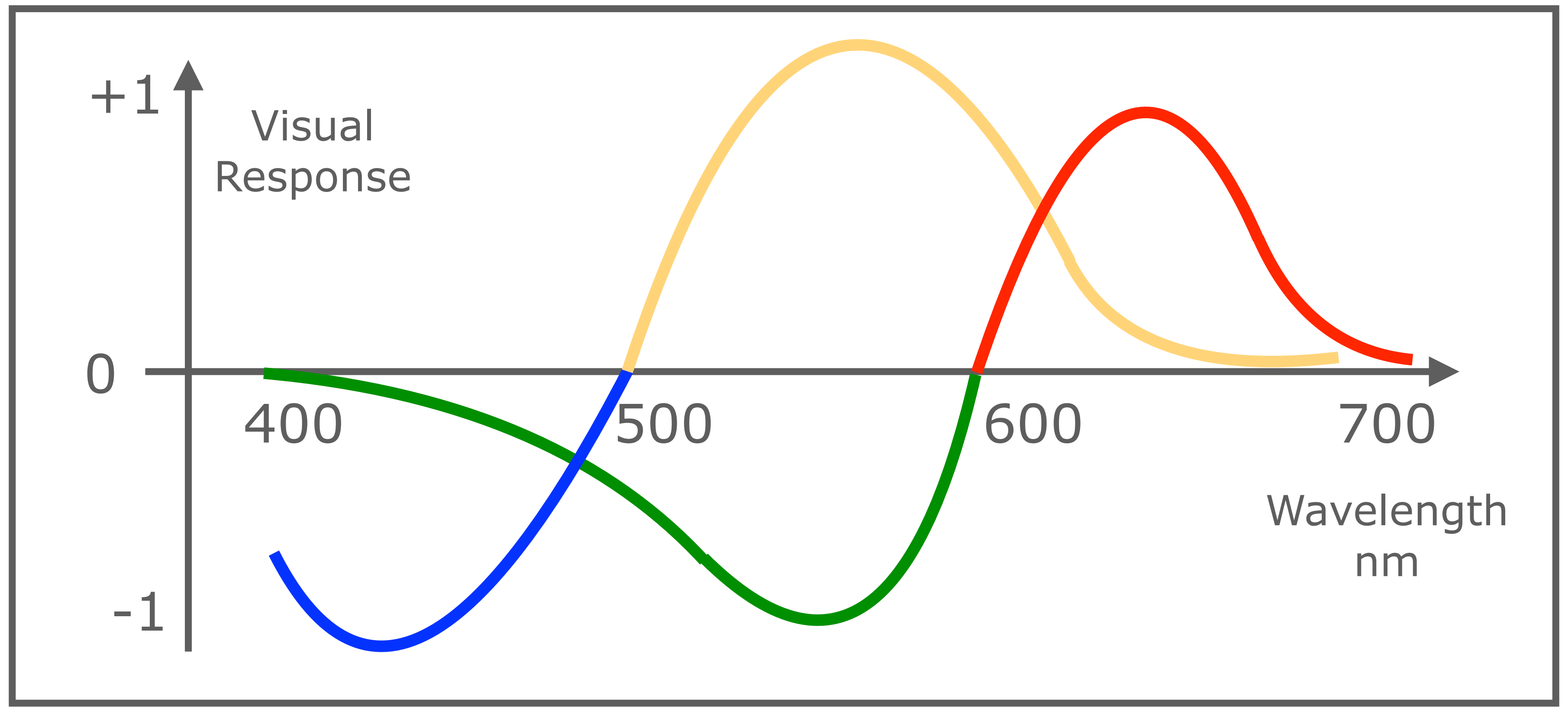
Cone Receptors

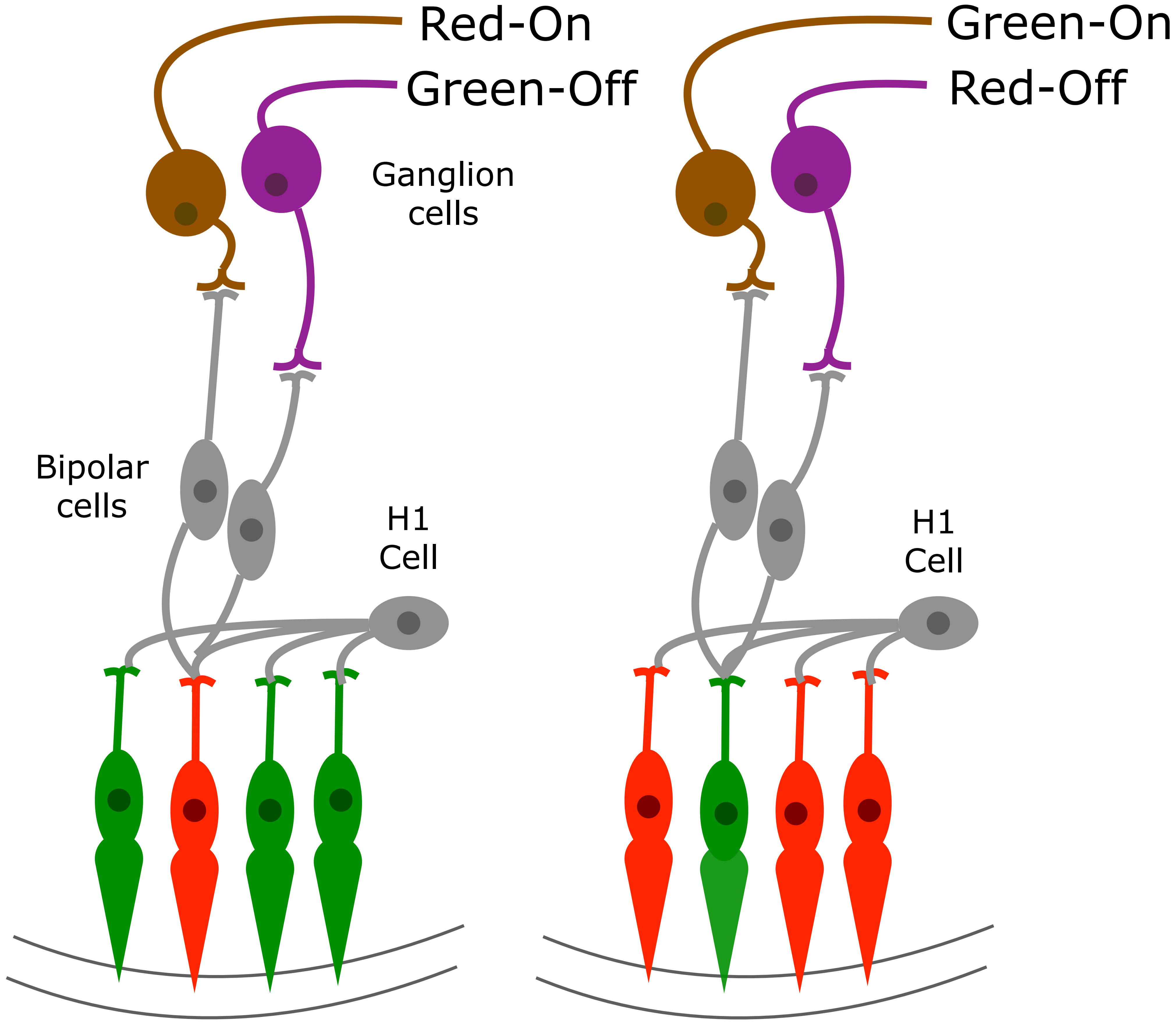
- Fovea has the highest density of L and M cones but no S cones (other than foveal walls) and no Rods
- Overall, there are very few S cones and usually the highest number of L cones
- ~6M Cones
  - S cones, only about 100-120K
  - M cones, ~2M
  - L cones, ~3-4M (ratio of L:M is quite variable)



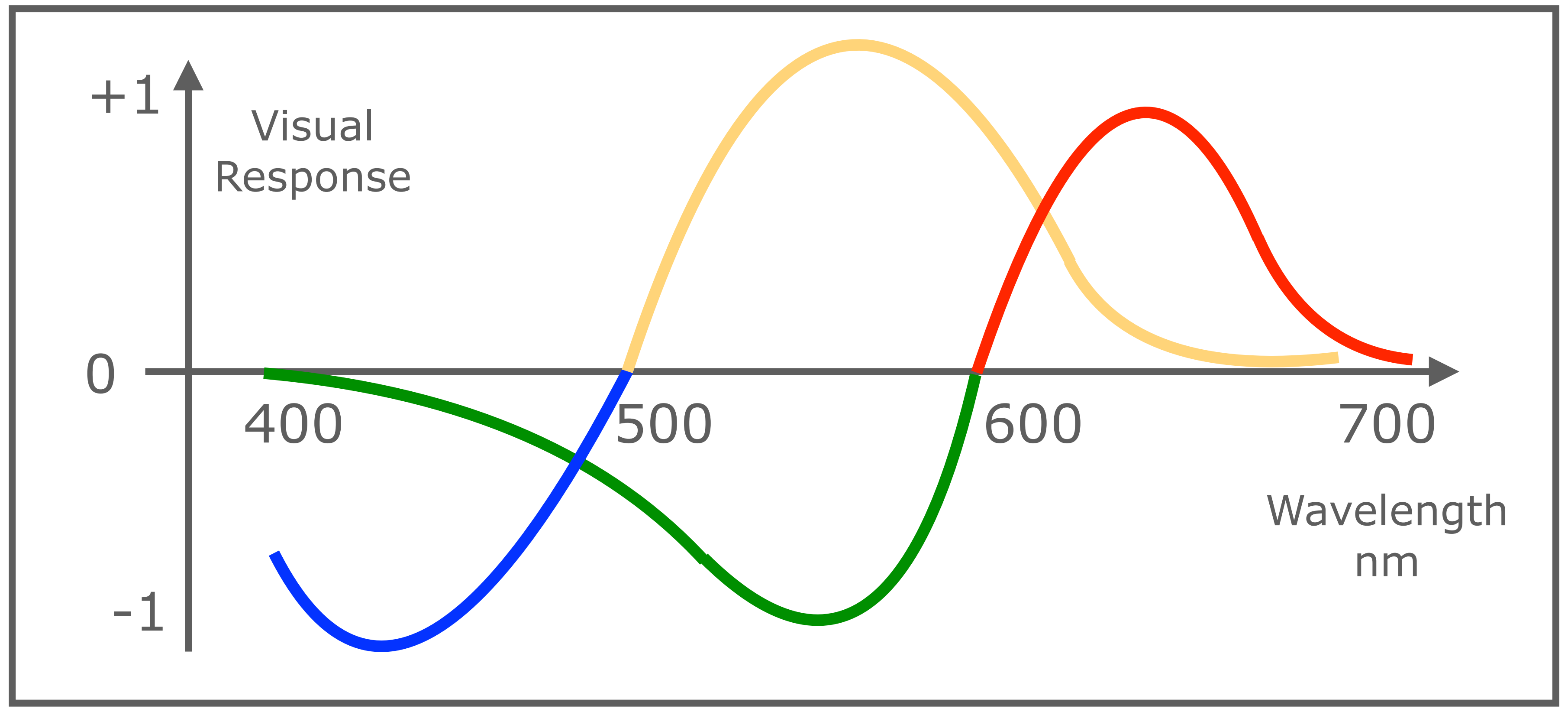
Hue Perception

- There is an ongoing debate about specifics of the mechanism…
- Many textbooks will outline parvocellulare mechanism:
  - R-G = L-M
  - B-Y = S-(M+L)
- though this is most probably wrong…

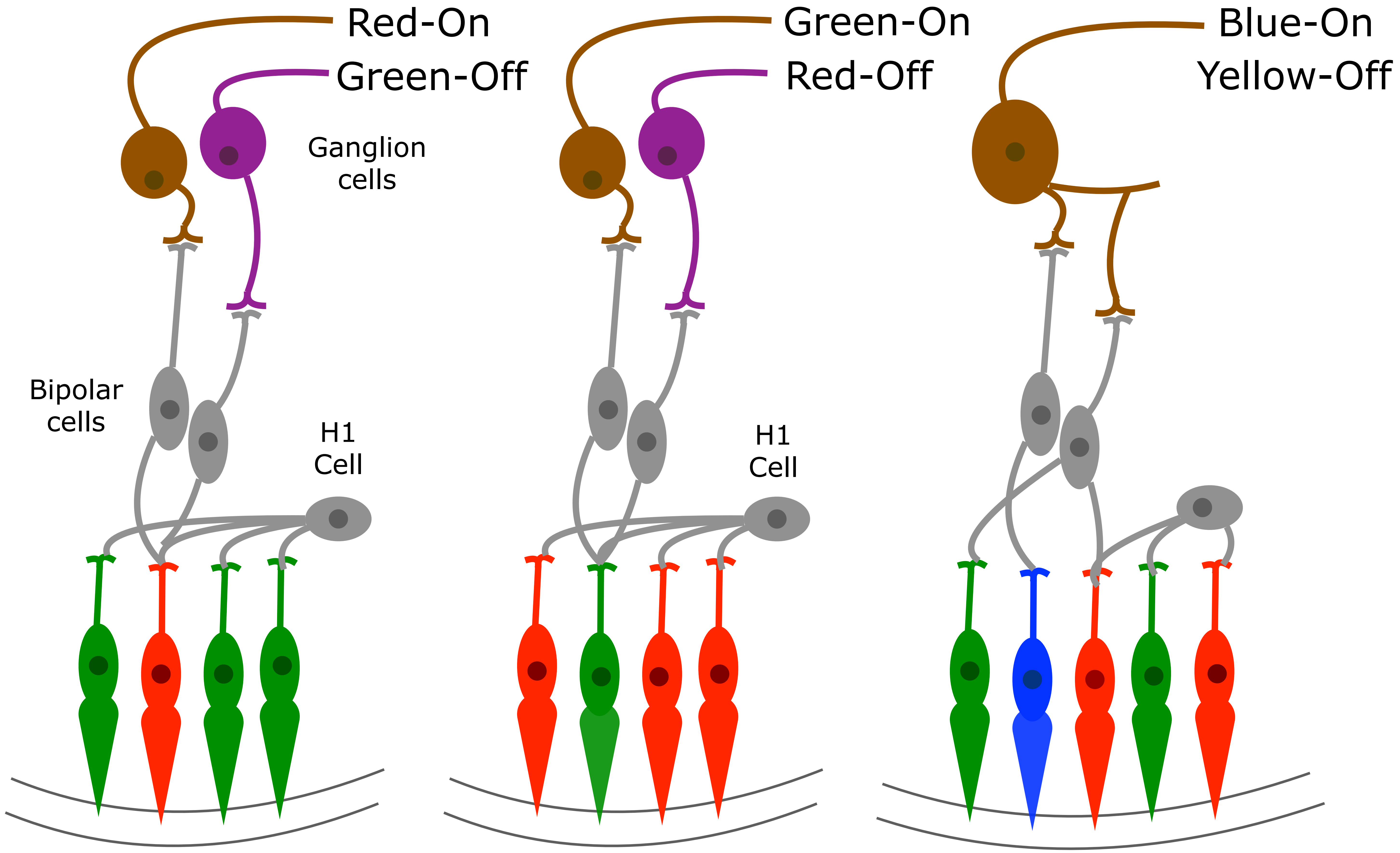




Hue Perception

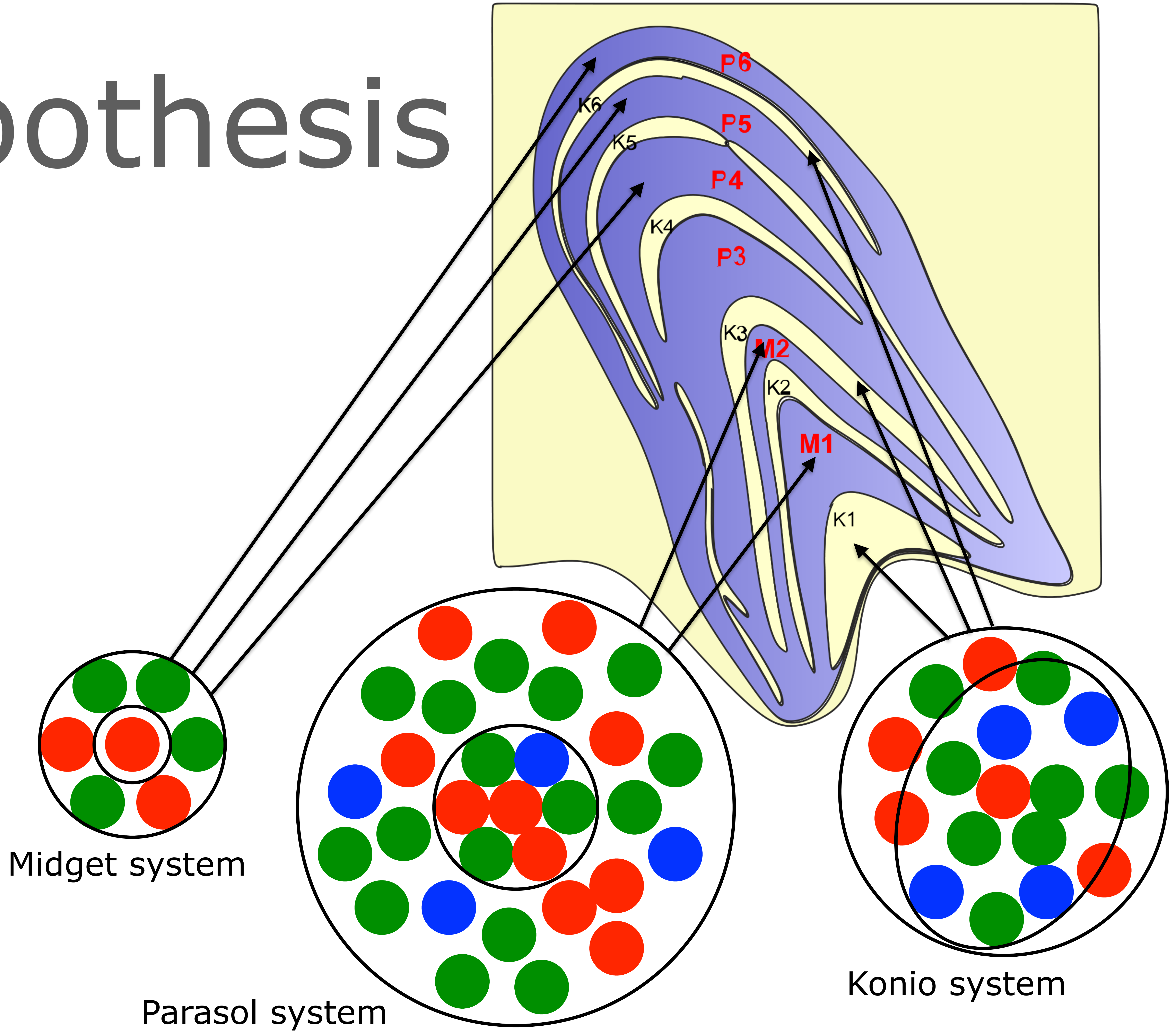


- There is an ongoing debate about specifics of the mechanism…
- Many textbooks will outline parvocellulare mechanism:
  - R-G = L-M
  - B-Y = S-(M+L)
- though this is most probably wrong…

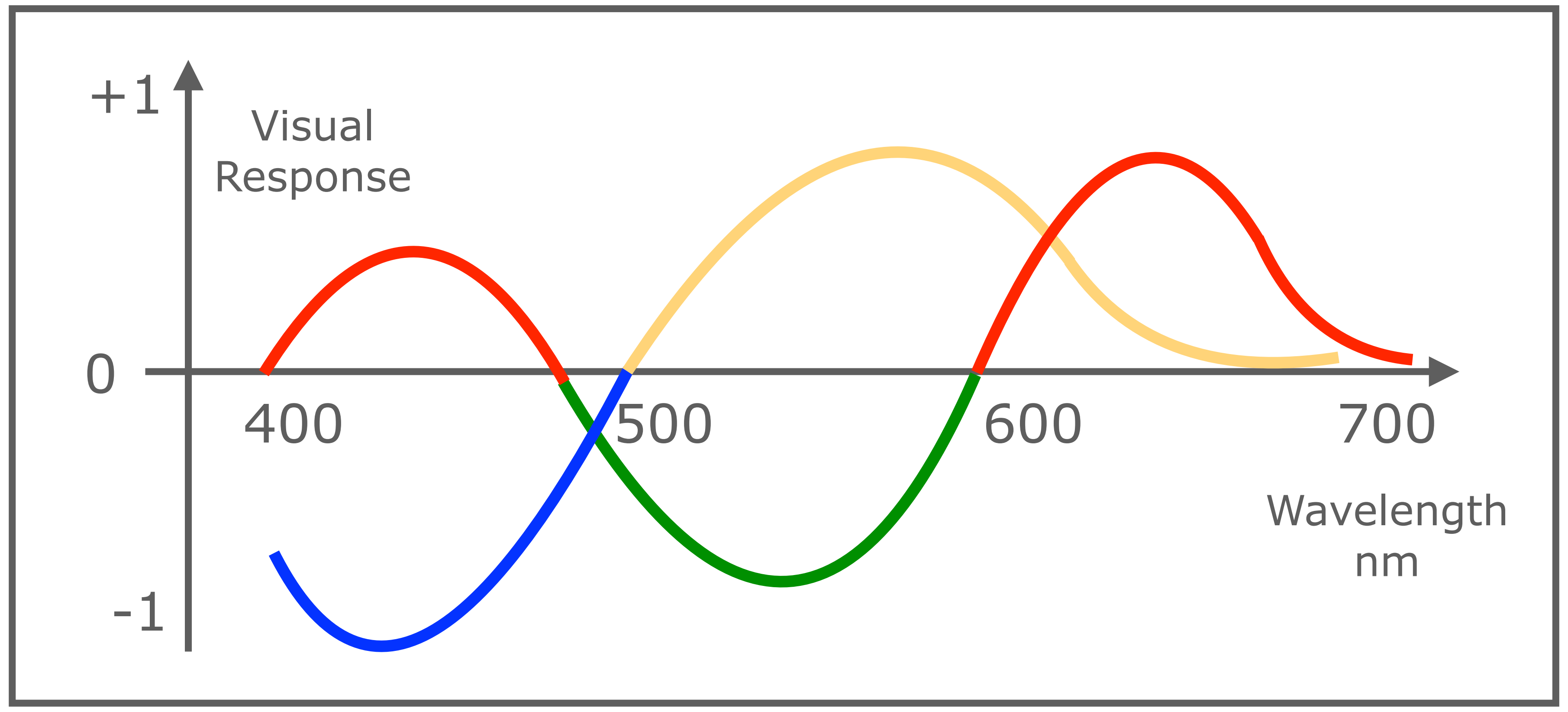


Another hypothesis

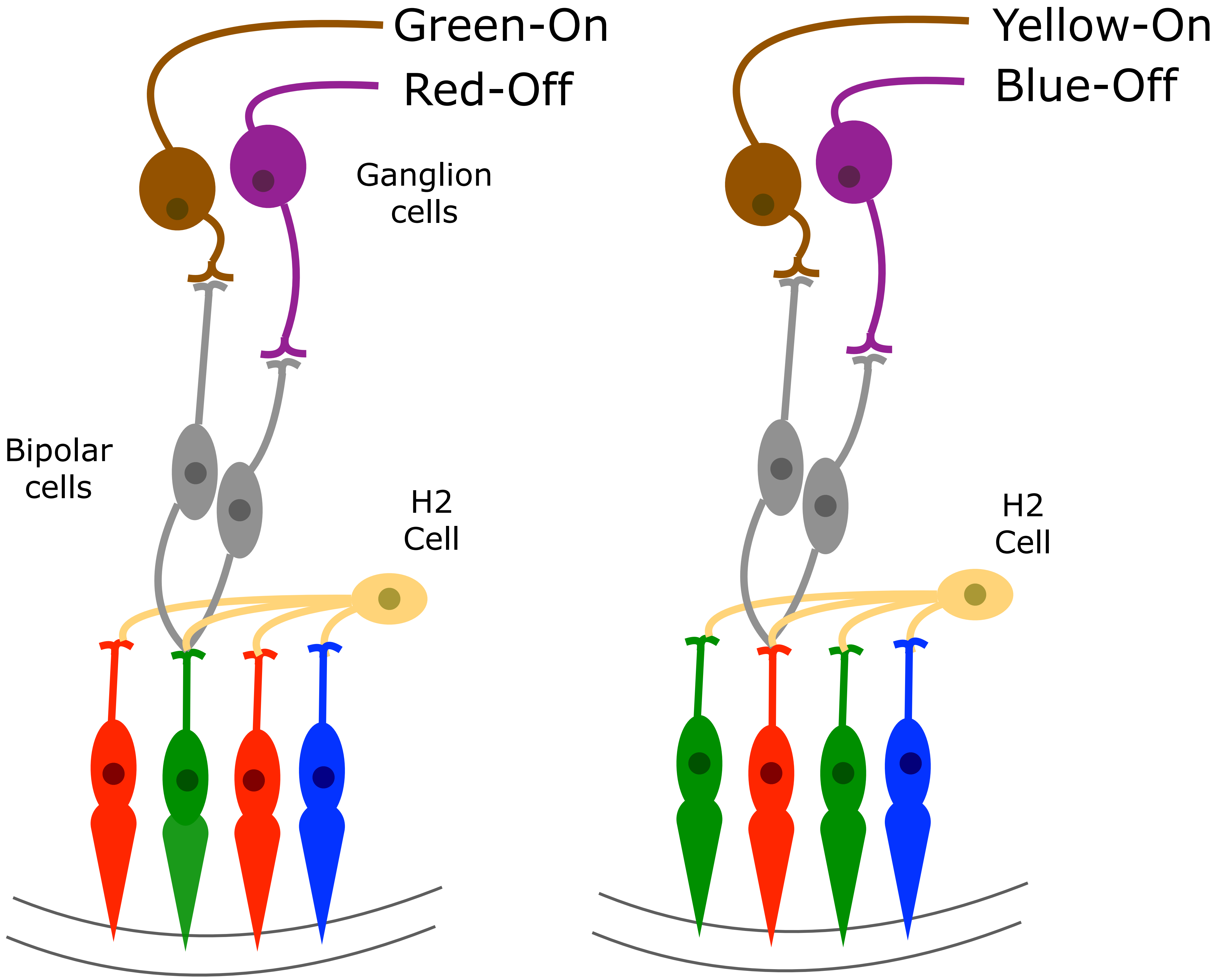
- Perhaps midget RGCs send both achromatic and chromatic data?
- Midget RGCs send L-M, M-L, L+M and -L-M to LGN P layers
- Bistratified RGCs provide additional +S signal through Koniocellular system to K layers



Retinal Hypothesis



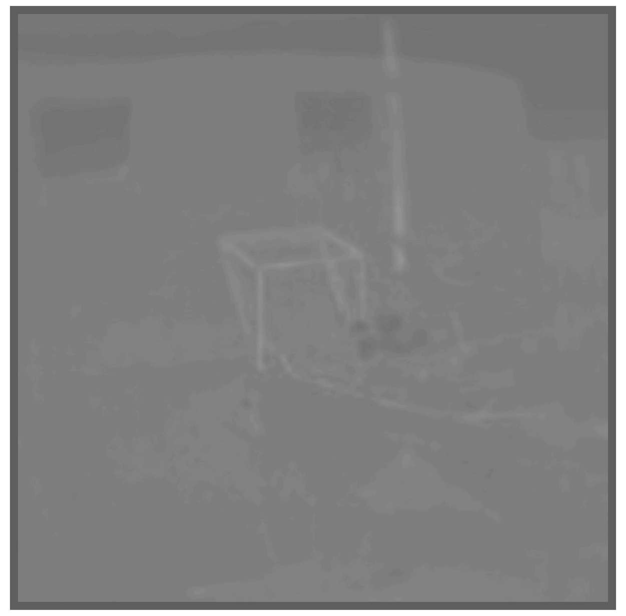
- Adaptive optics experiments: Most midgets RGCs produce achromatic perception…
- Some are tuned for chromatic edges but produce achromatic perception
- Relatively few midget RGCs carry opponent chromatic signal:
  - R = (L+S)-M
  - G = M-(L+S)
  - B = (M+S)-L
  - Y = L-(M+S)



Midget/Parvocellulare pathways retinal data

Edges

Red/Green



Blue/Yellow

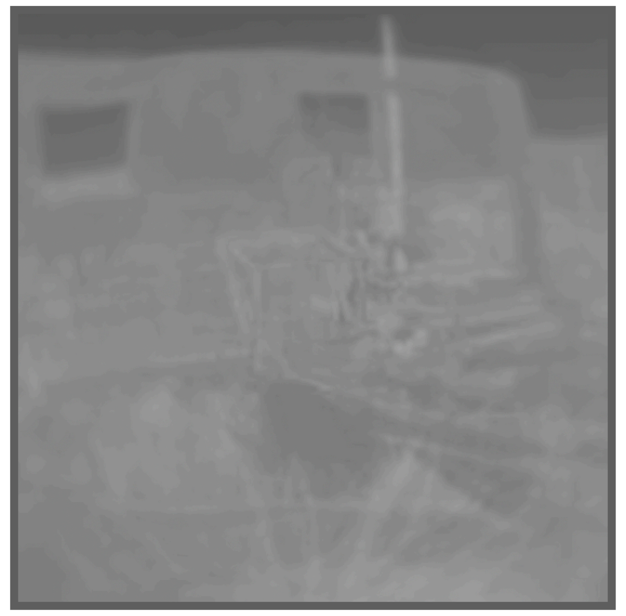
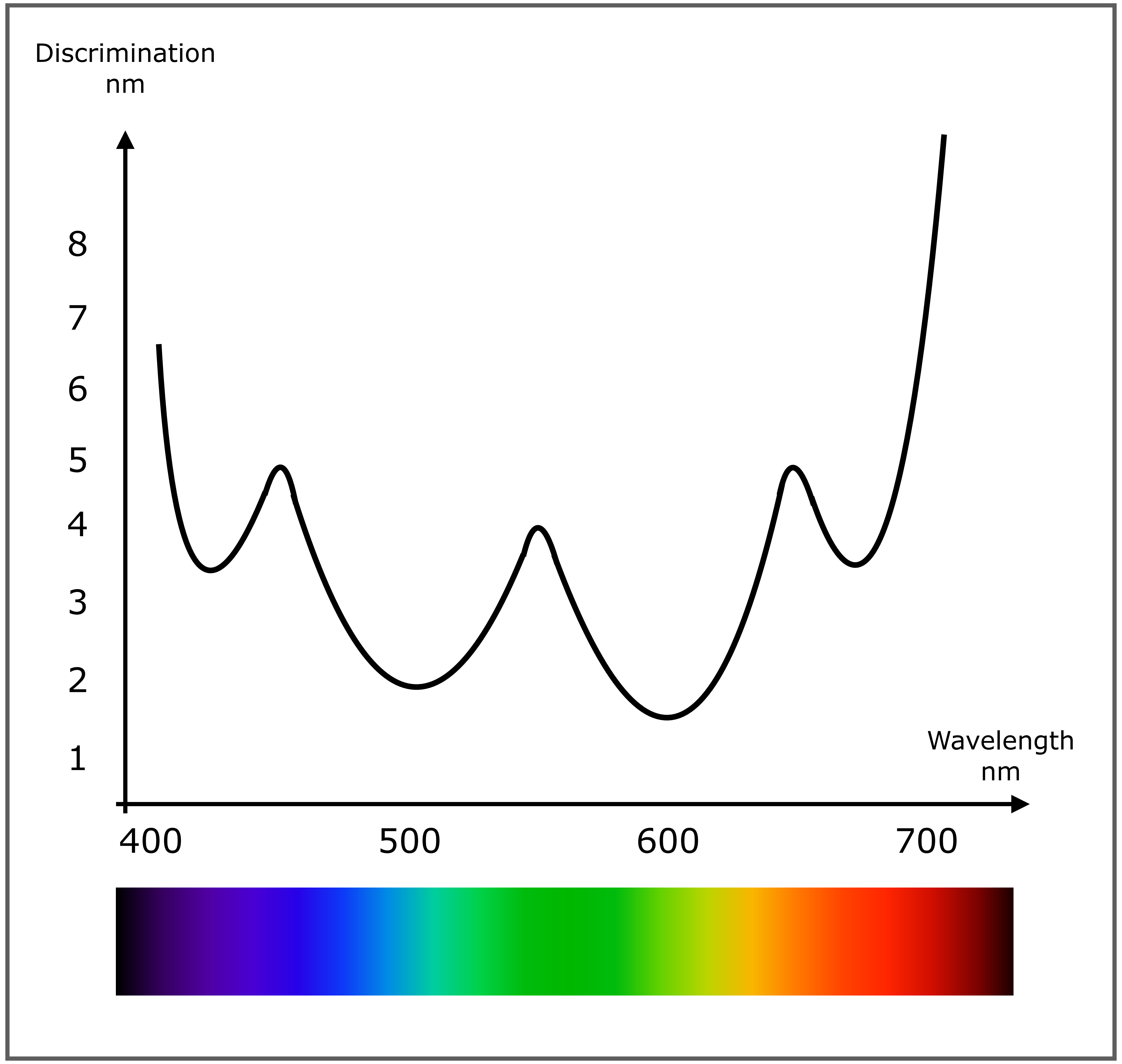


Image by Jeff Perry, Public Domain

Wavelength Color Discrimination

- Humans can differentiate a 2-4 nm step in monochromatic stimuli
- Humans can see anywhere from 1M-10M distinct colors
- A 24 bit image contains 16.7M colors



Adopted from Pokorny and Smith, JOSA 1970

Divergences from “Normal” Color Vision

- About 8% of population have color vision abnormalities
- L or M receptors (rarely S) can be anomalous or missing
- Even with “normal” vision the ratio of L and M receptors vary significantly among the population leading to differences of perceiving pure green by as much as 50 nm

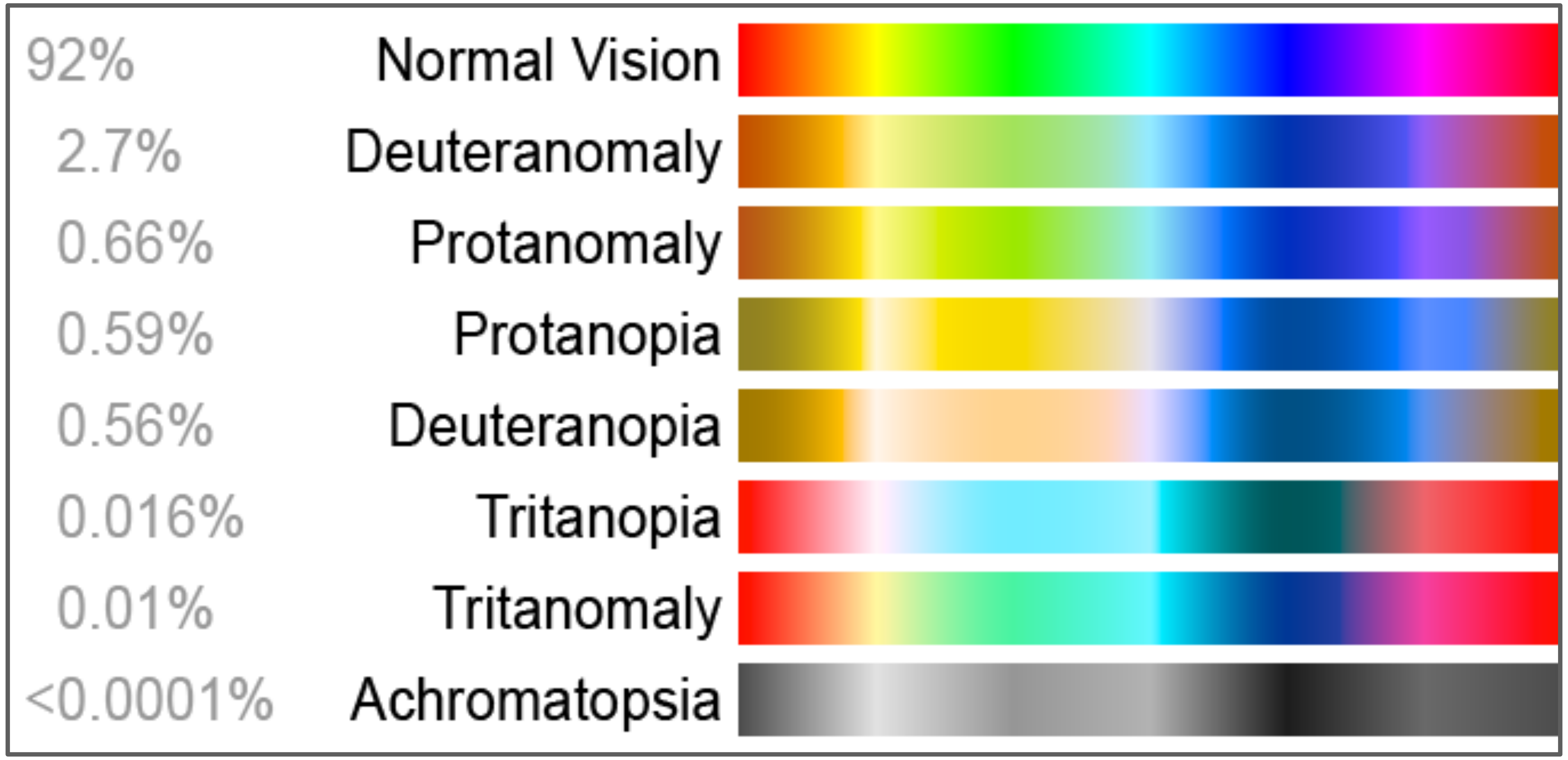
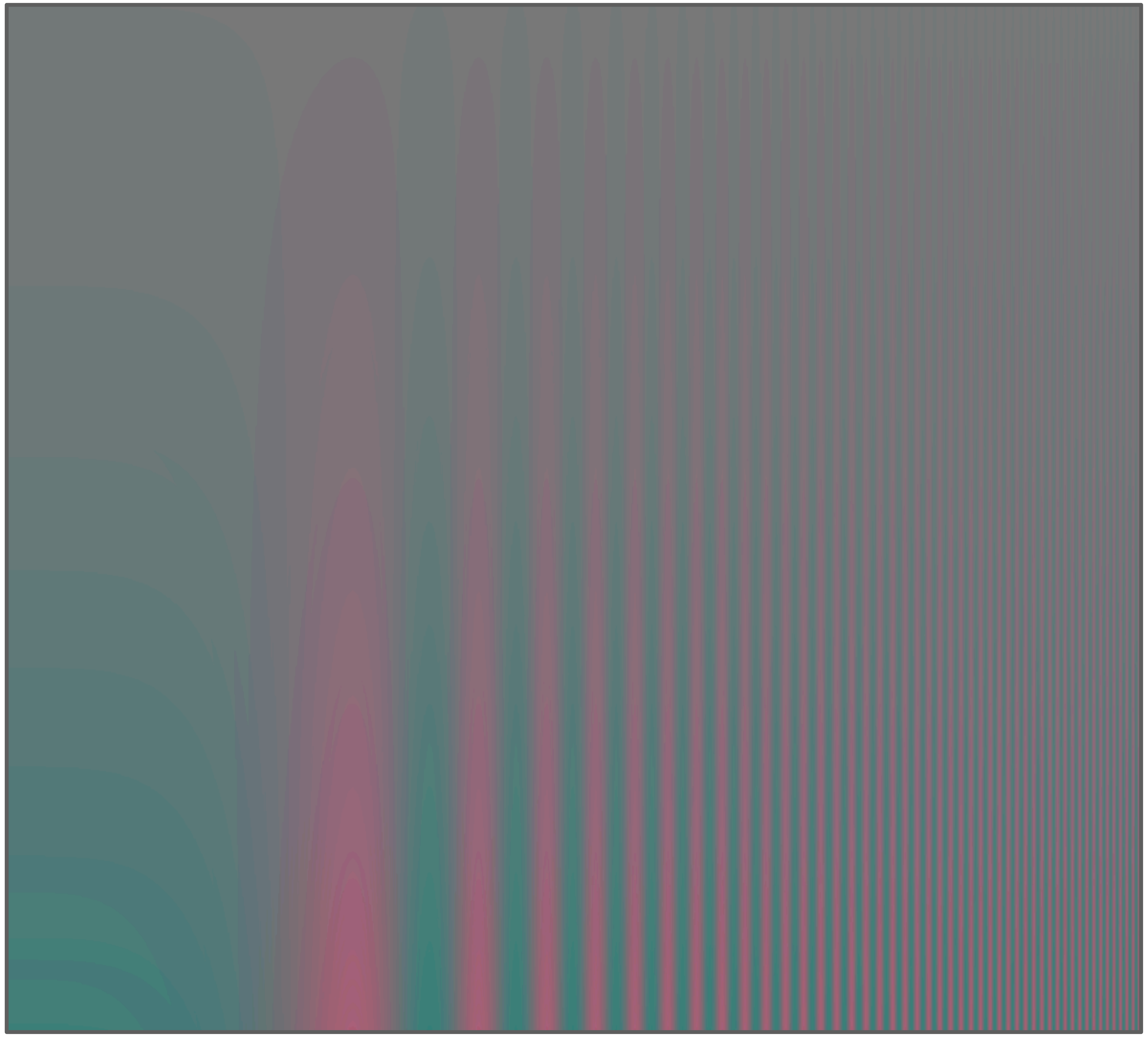


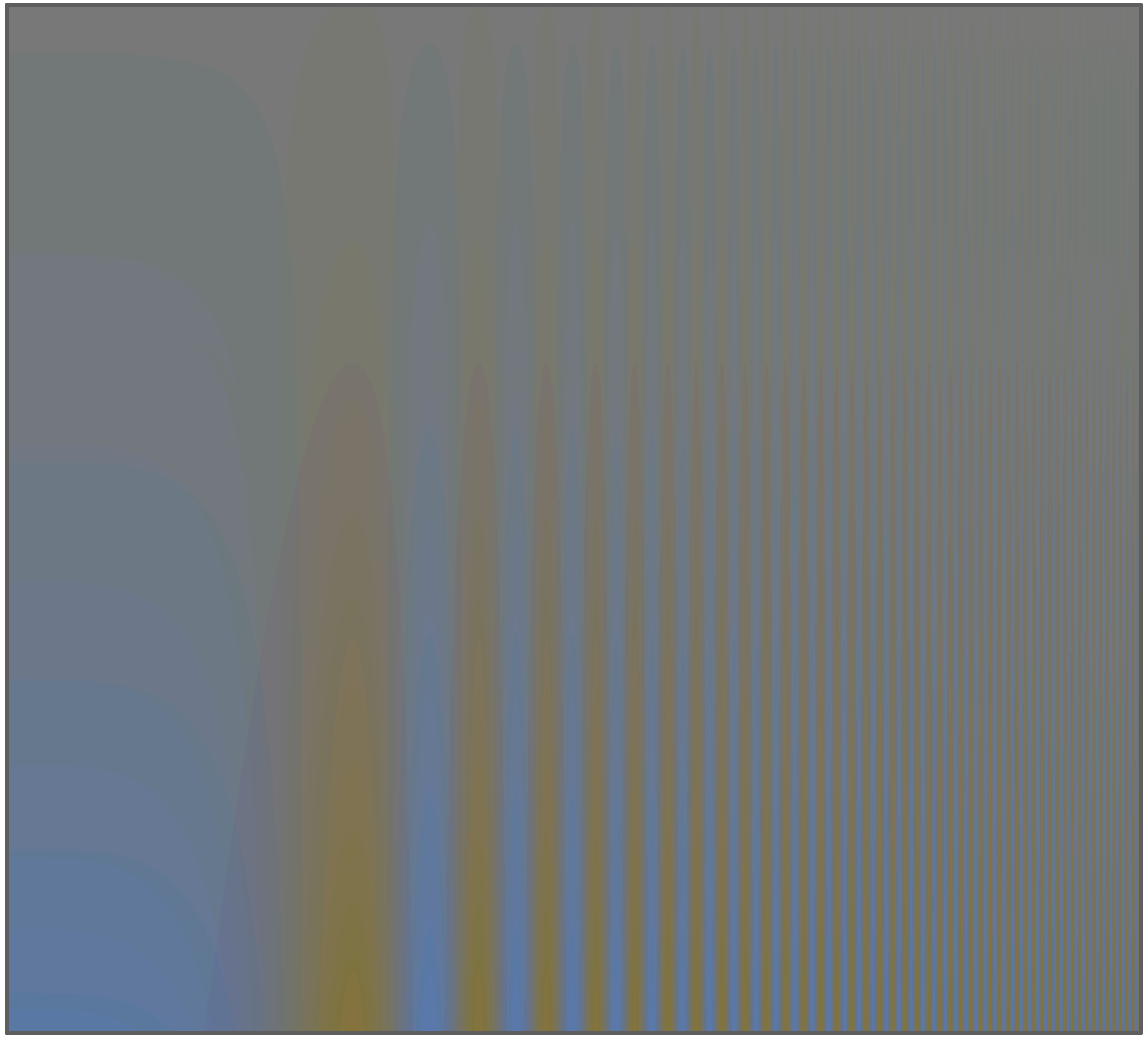
Image by Nanobot, Public Domain

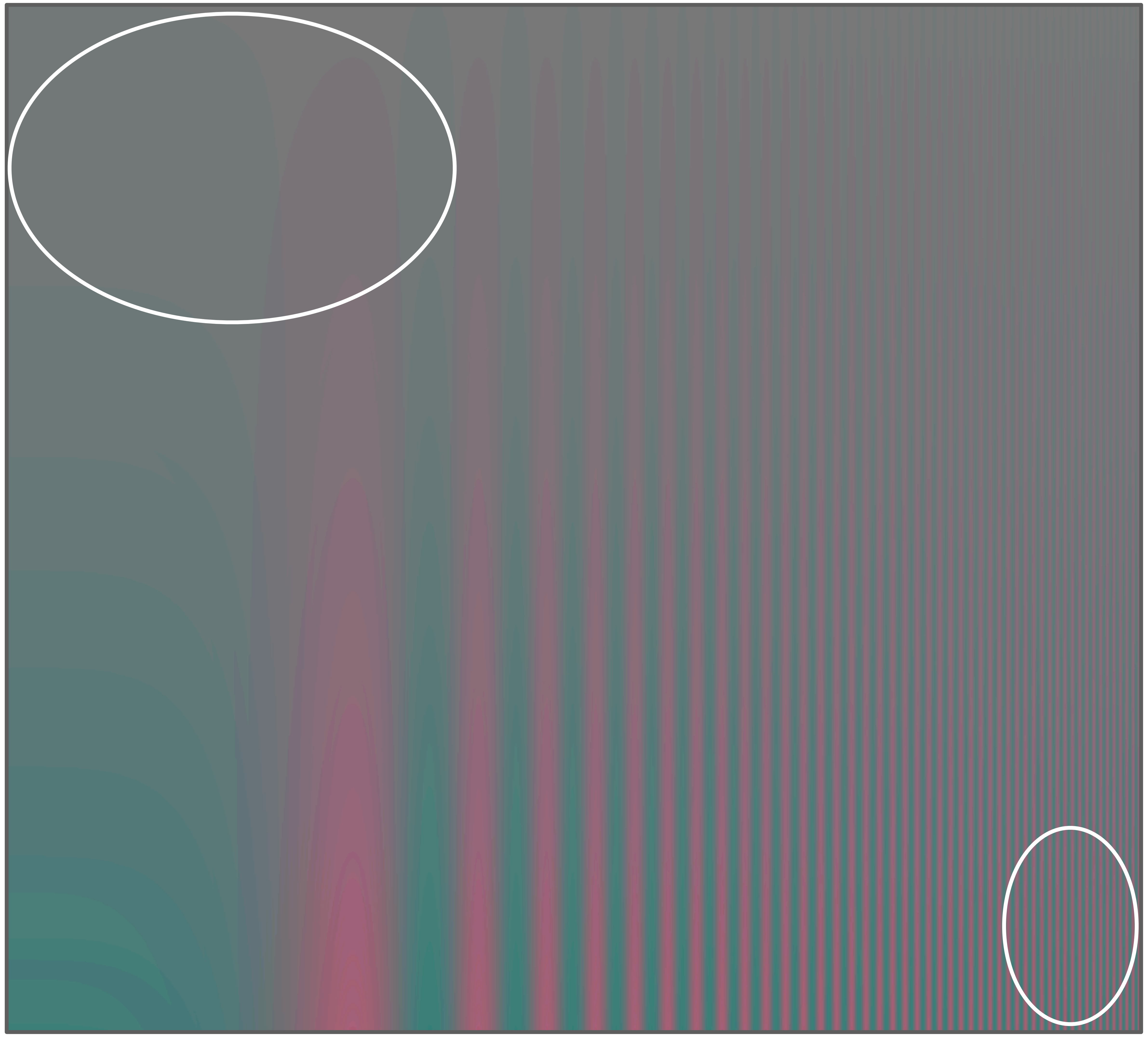
Why does it matter?

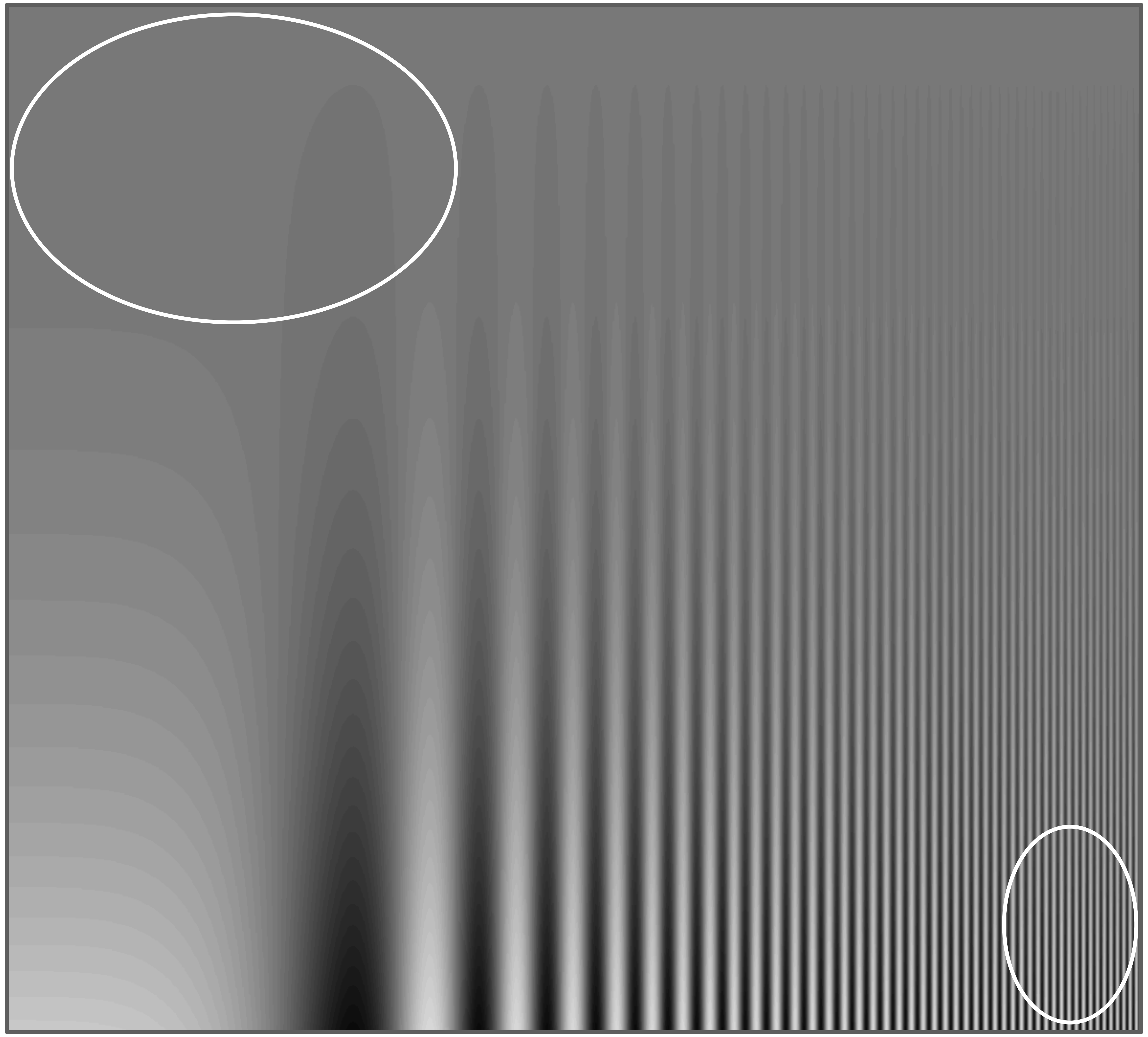
- We cannot see greenish reds or yellowish blues
- Changes in greenish intensities are easier to recognize
- Red and Blue may not be the best for small foreground elements
- Not everyone sees the same way!

Shape Perception



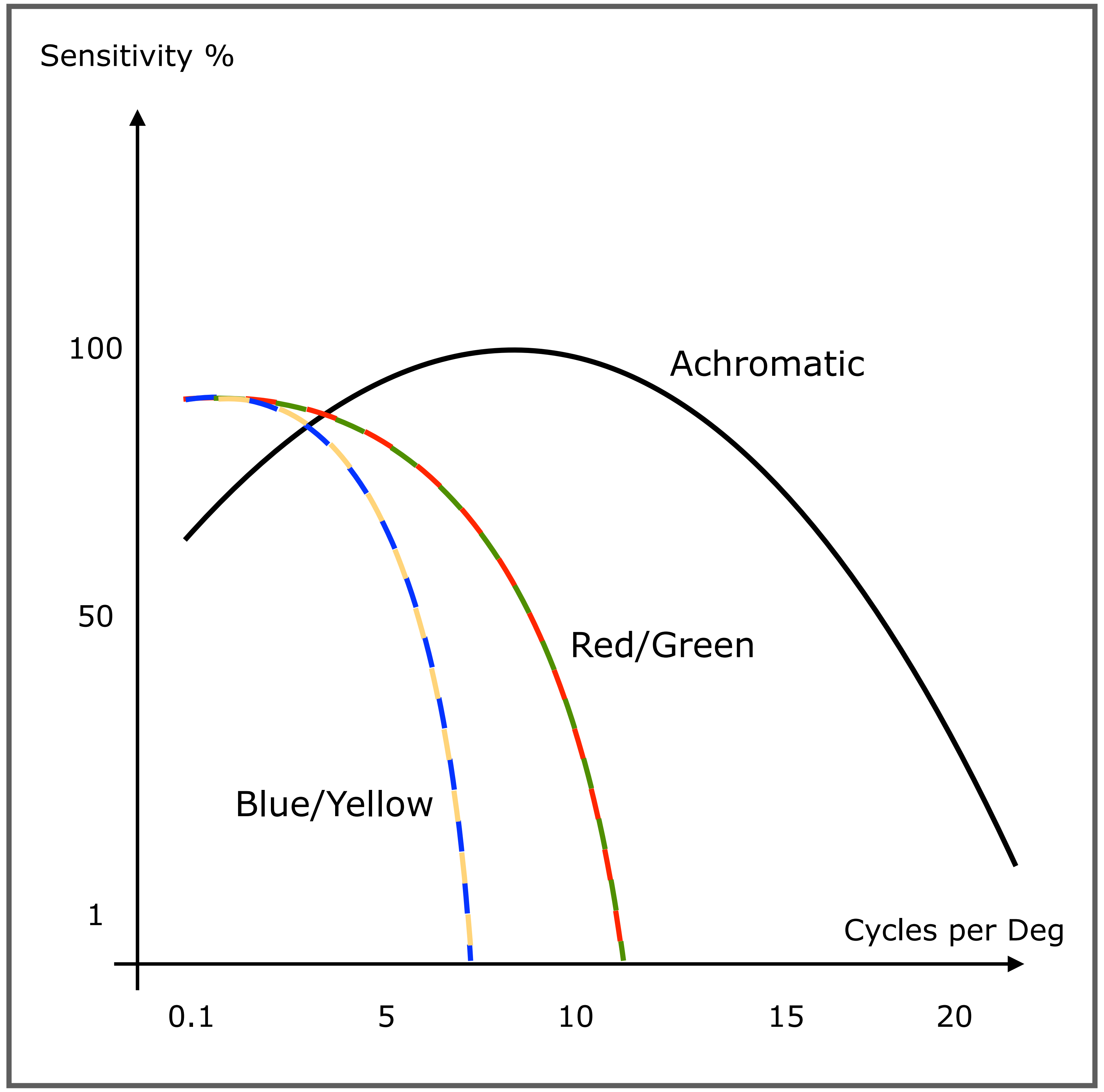






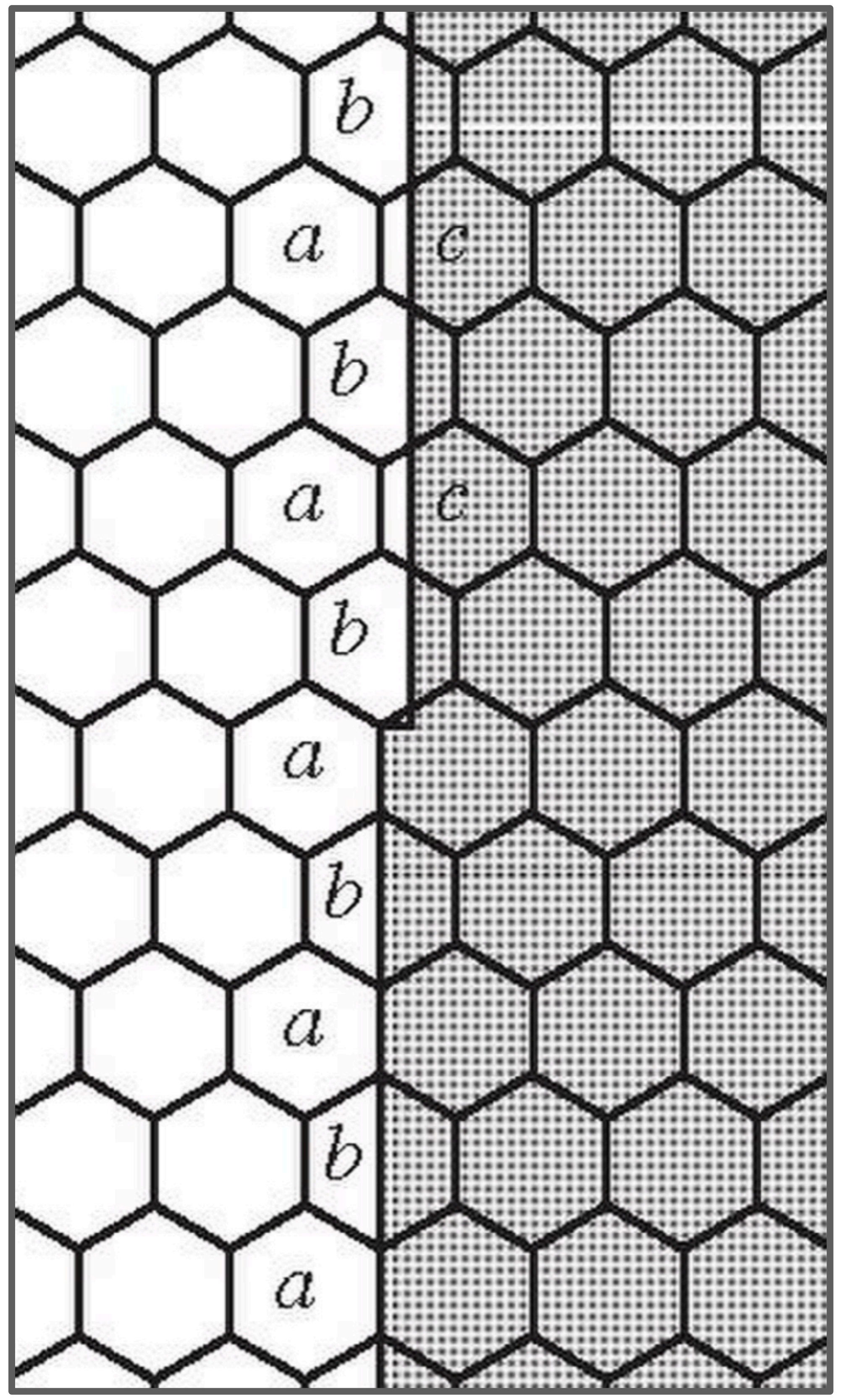
Contrast Perception

- Achromatic contrast sensitivity peaks at 8-12 cycles per degree
- Chromatic contrast sensitivity declines much faster compared to achromatic but it is stronger at low cycles per degree
- May not be the same for all orientations…



Foveal Hyperacuity

- Humans can perceive line discontinuities at a rate exceeding foveal acuity - essentially sub cone size…
- It is impressive but exaggerates aliasing artifacts



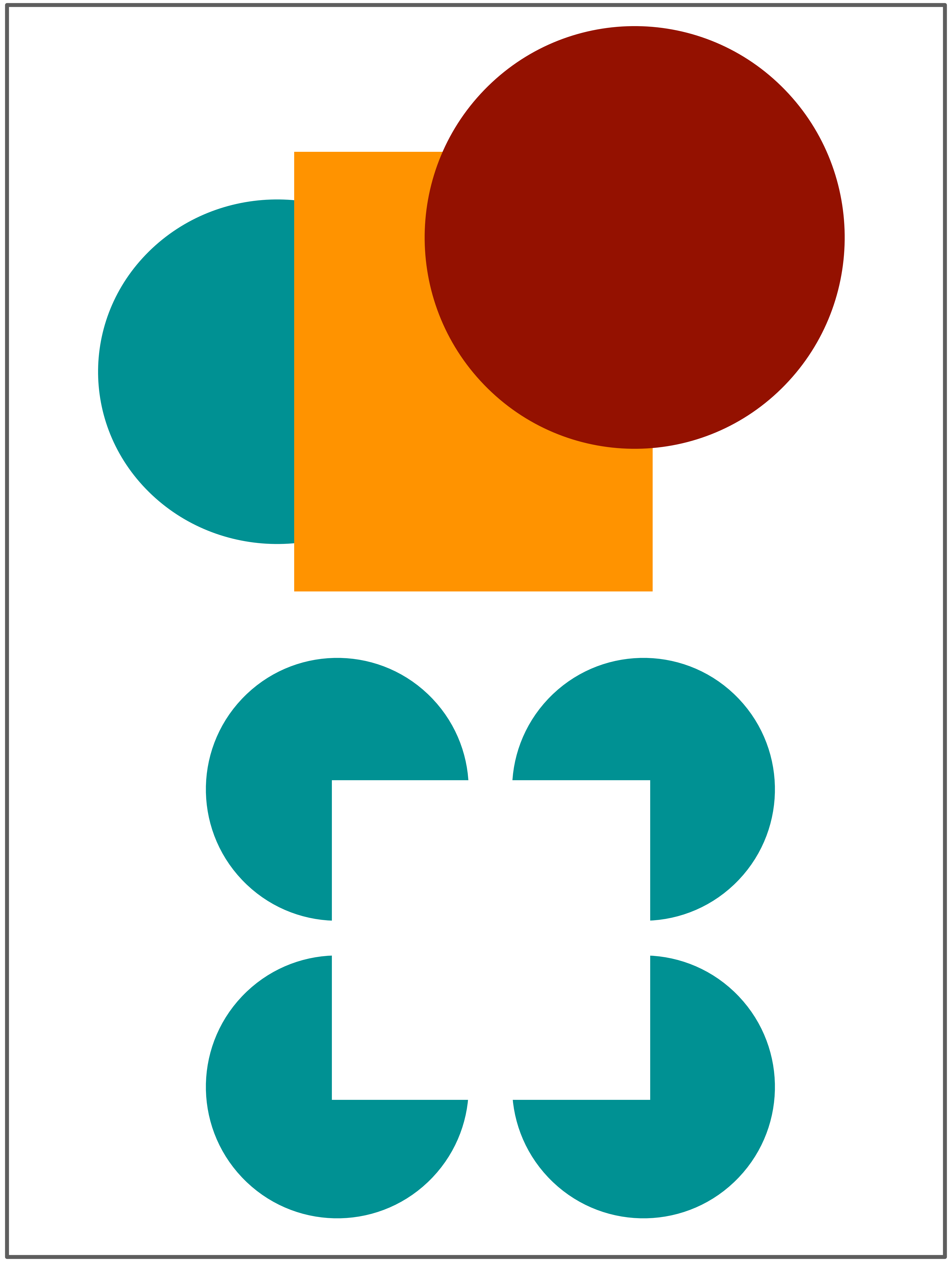
Figure/Ground

- Important automatic step to identify what are not objects
- Probabilistic, using multiple clues:
  - Edge density, convexity, color saturation, movement
- Can be ambiguous
- Importantly, less attention will be paid to the ground and less details remembered

Image by Klaus-Dieter Keller, Public Domain

Contours

- Contours are autocompleted
- There is a bias towards concave shape interpretation
- We can perceive imaginary occluders to a point of perceiving fine completion edges where there are none



Grouping

- We perceive objects hierarchically and in groups
- Gestalt psychology explored this extensively (many UI/UX implications)
- Attention spreads for the entire object or group of objects when we fixate on its part

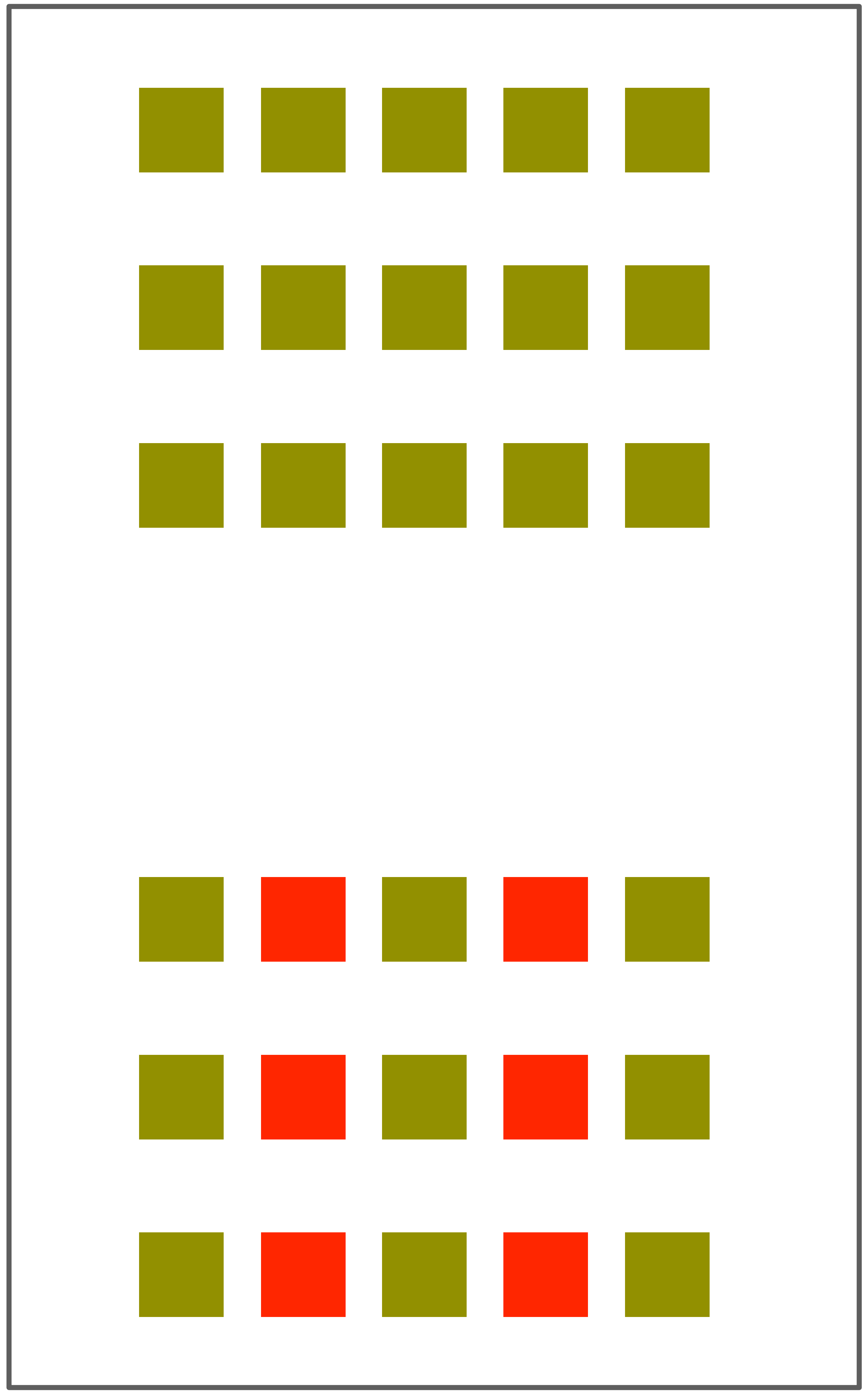


Image by Rob Bogaerts Image manipulation: Phonebox, Public Domain

Image by Rob Bogaerts Image manipulation: Phonebox, Public Domain

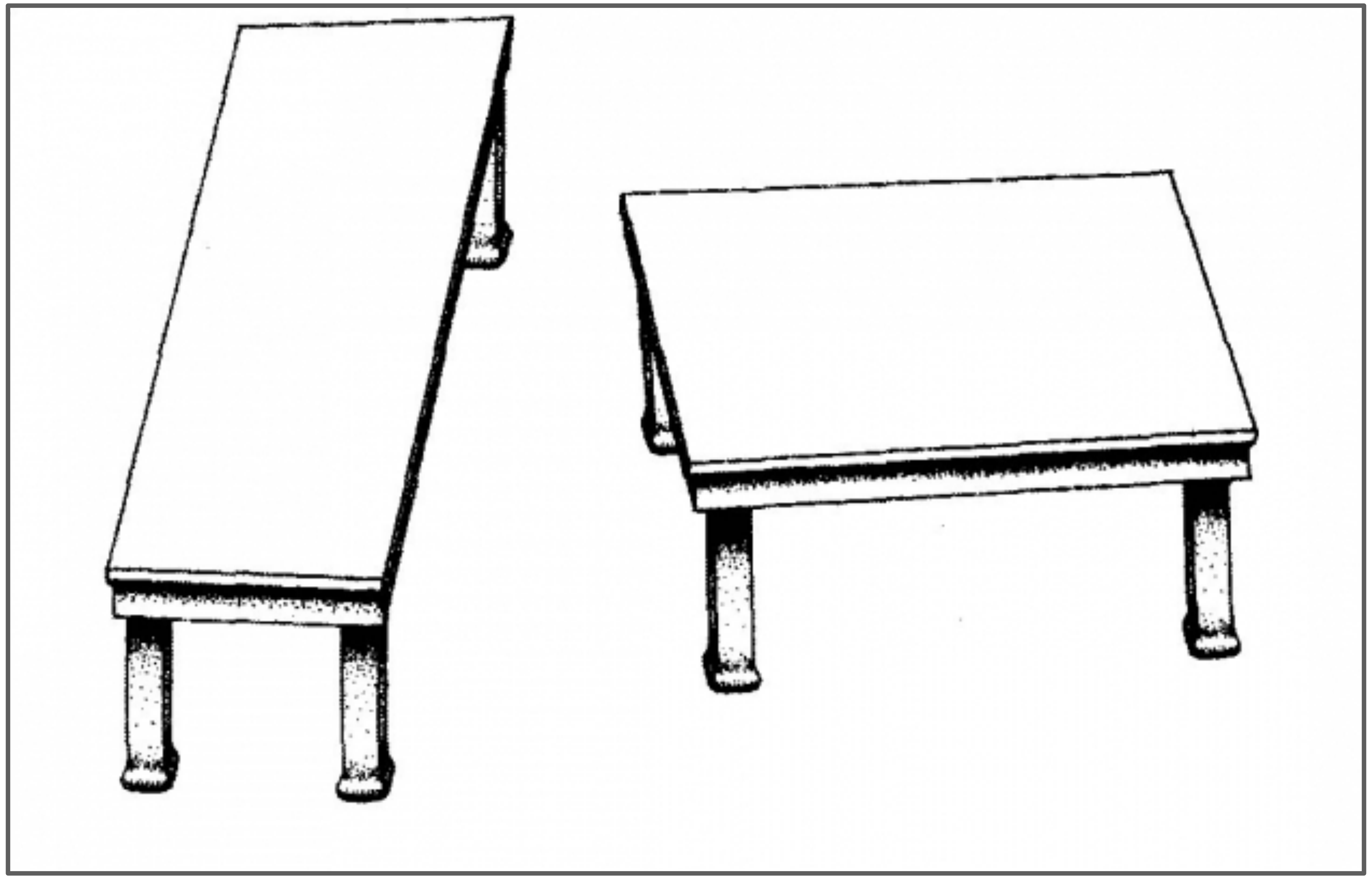


Image by Xenia Nedelina based on Roger N Shepardwork,SS BY-SA 4.0

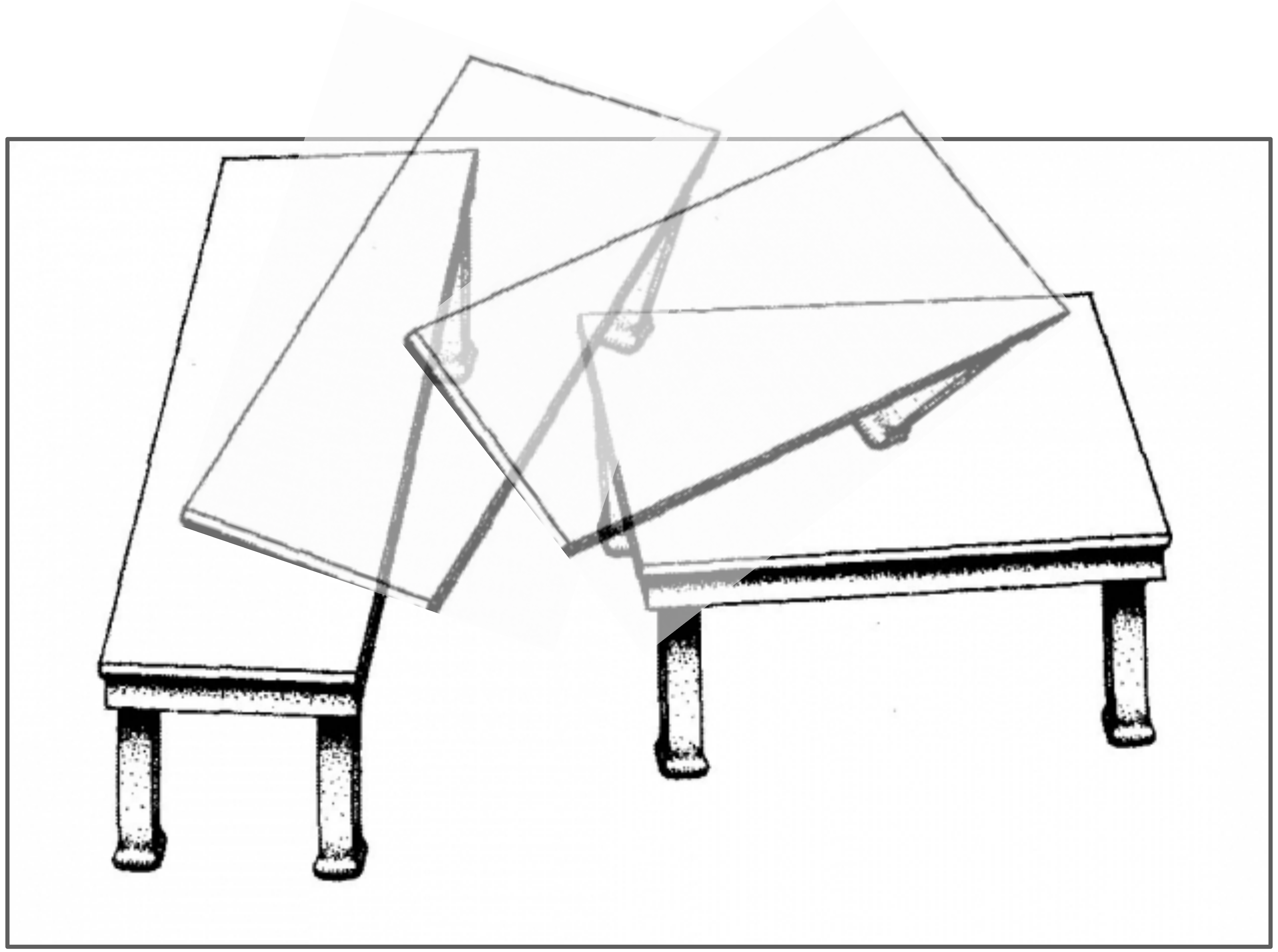


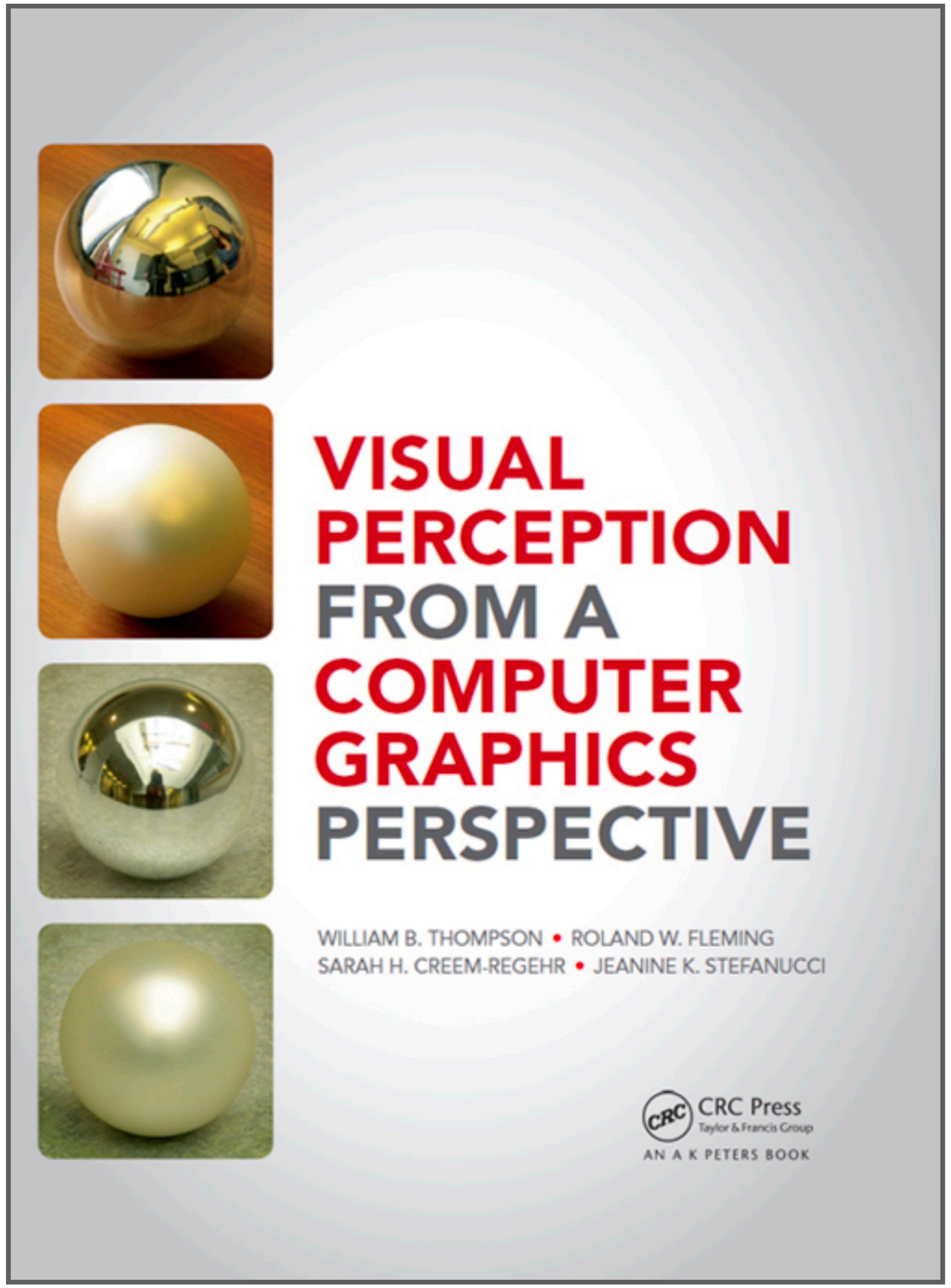
Image by Xenia Nedelina based on Roger N Shepardwork,SS BY-SA 4.0

Reference

Visual Perception From a Computer Graphics Perspective

William Thompson, Jeanine Kelly Stefanucci, Sarah Creem-Regehr, Roland Fleming

CRC Press, 2011



Why does it matter?

- Humans have different limits to chromatic and achromatic high frequency details
- We are hyper acute to visual discontinuities
- Many continuity and perspective clues are used for shape recognition and classification
- There is a specialized brain area responsible for face recognition

Lighting

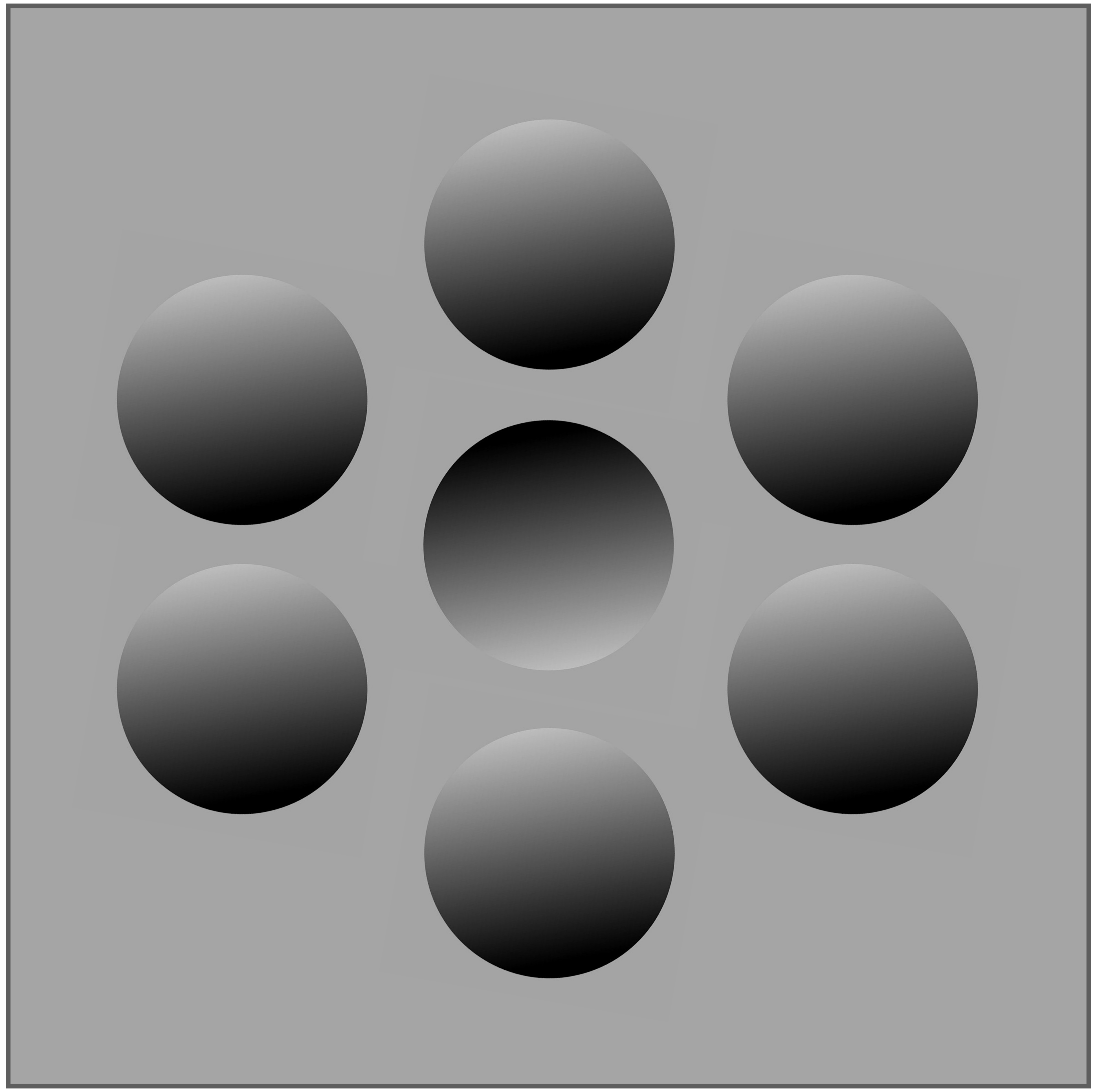
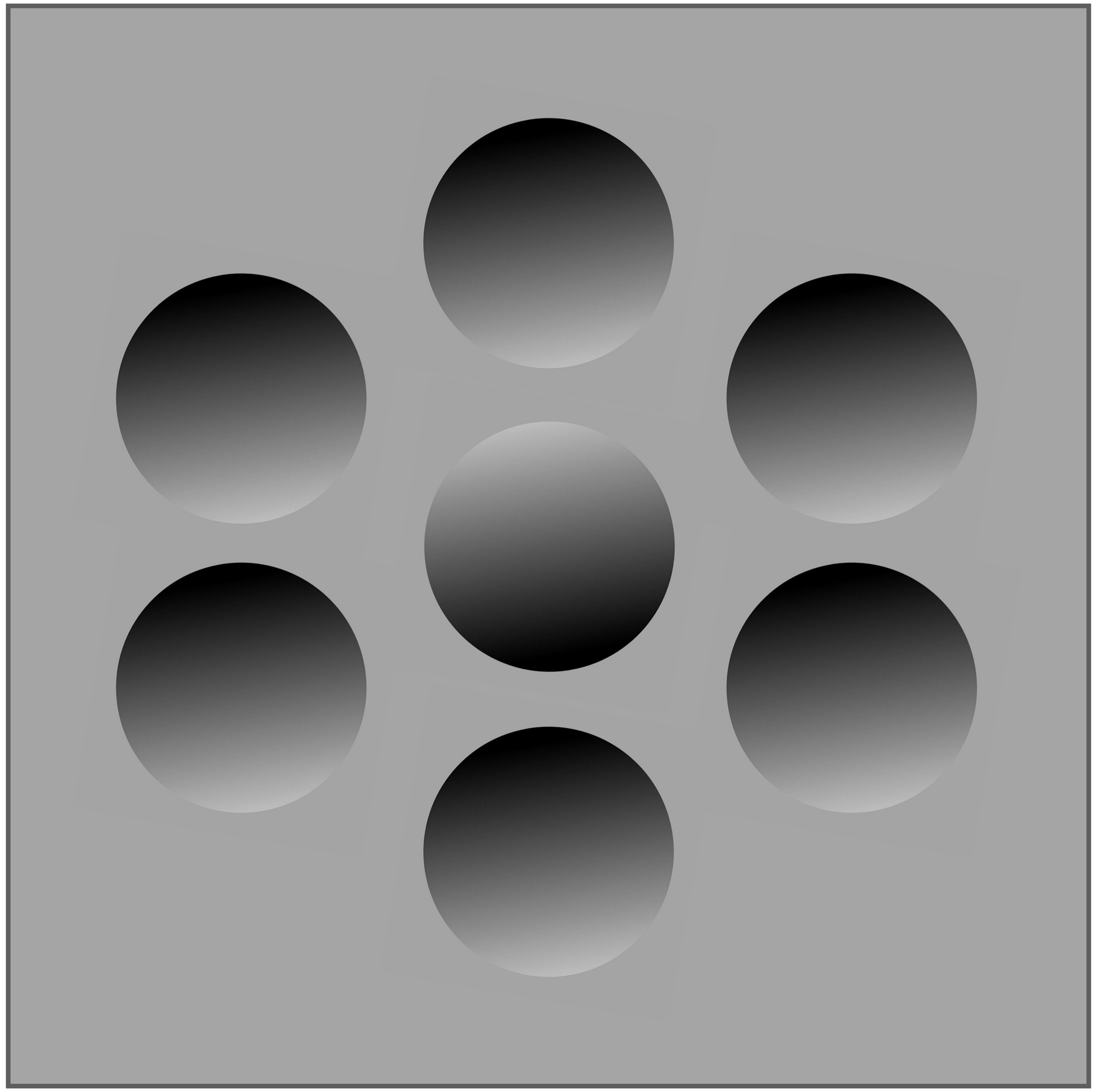


Image by Friedrich Frisch, CC BY-SA 4.0



Lighting from Above

- Shape perception is trained on natural lighting stimuli (hence lighting from above perceptual bias)
- Lighting from above bias is not absolute and easily compensated by perspective and shadows clues

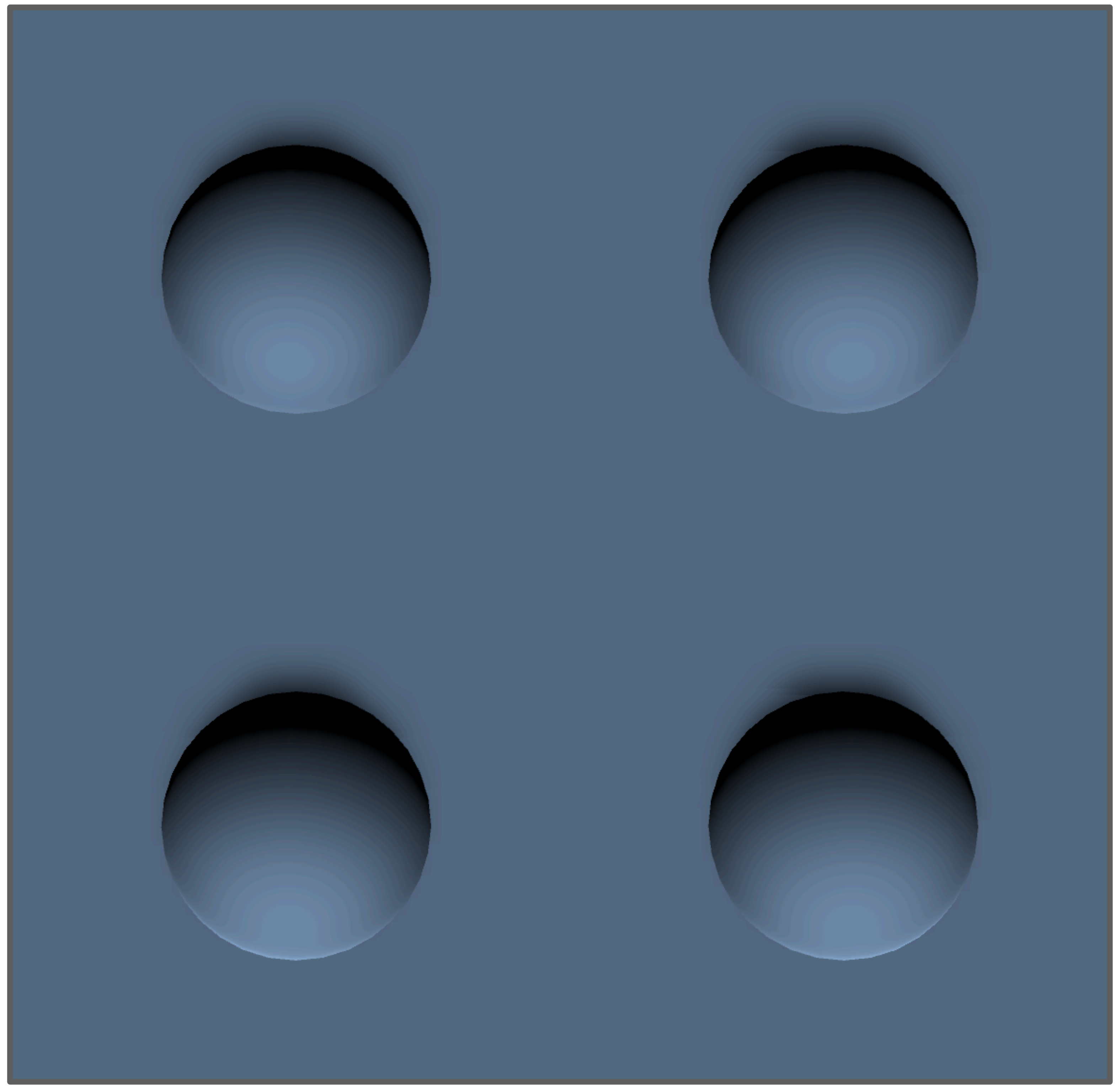


Image courtesy of Akiyoshi Kitaoka, http://www.ritsumei.ac.jp/~akitaoka/index-e.html

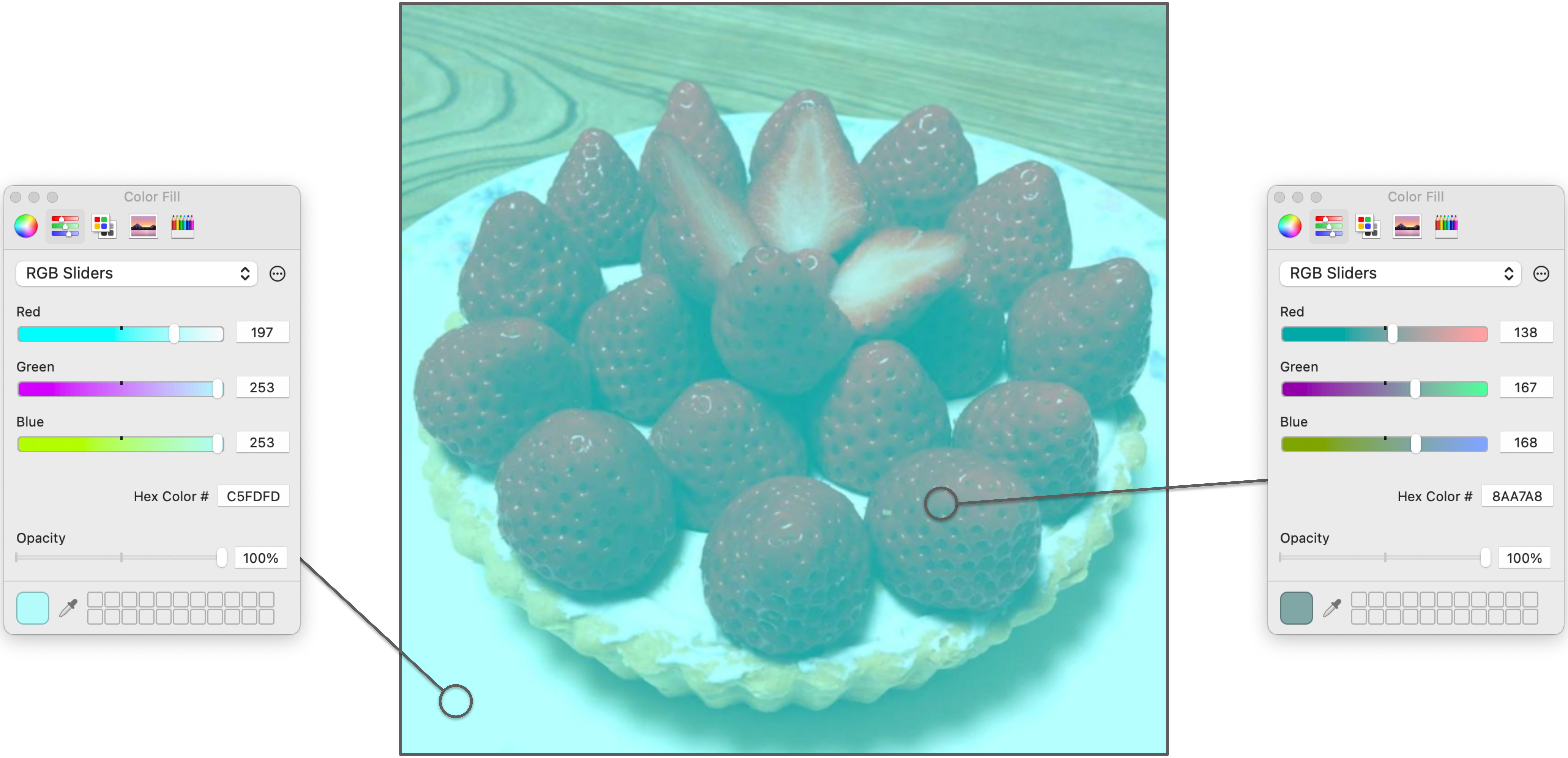
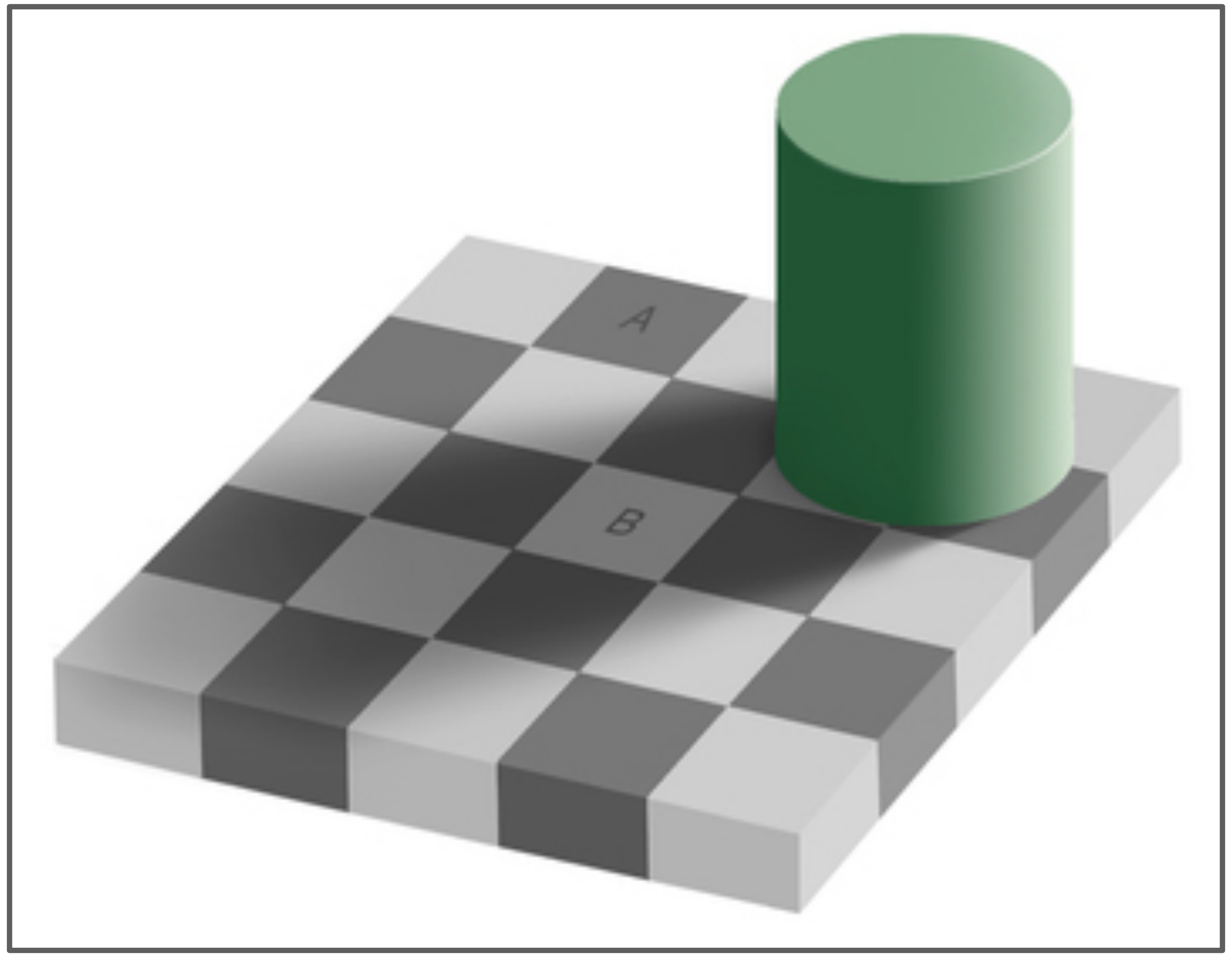
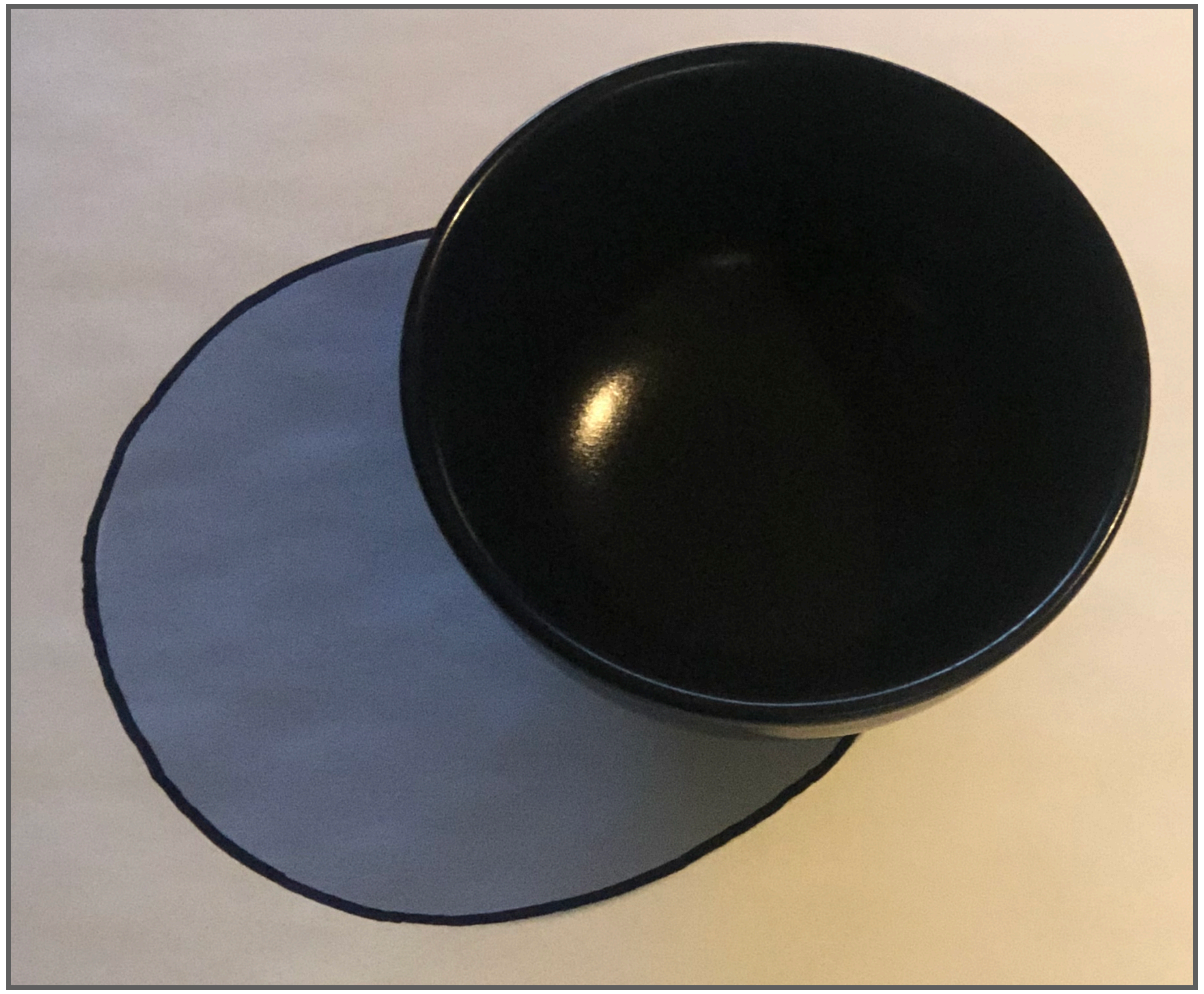


Image courtesy of Akiyoshi Kitaoka, http://www.ritsumei.ac.jp/~akitaoka/index-e.html

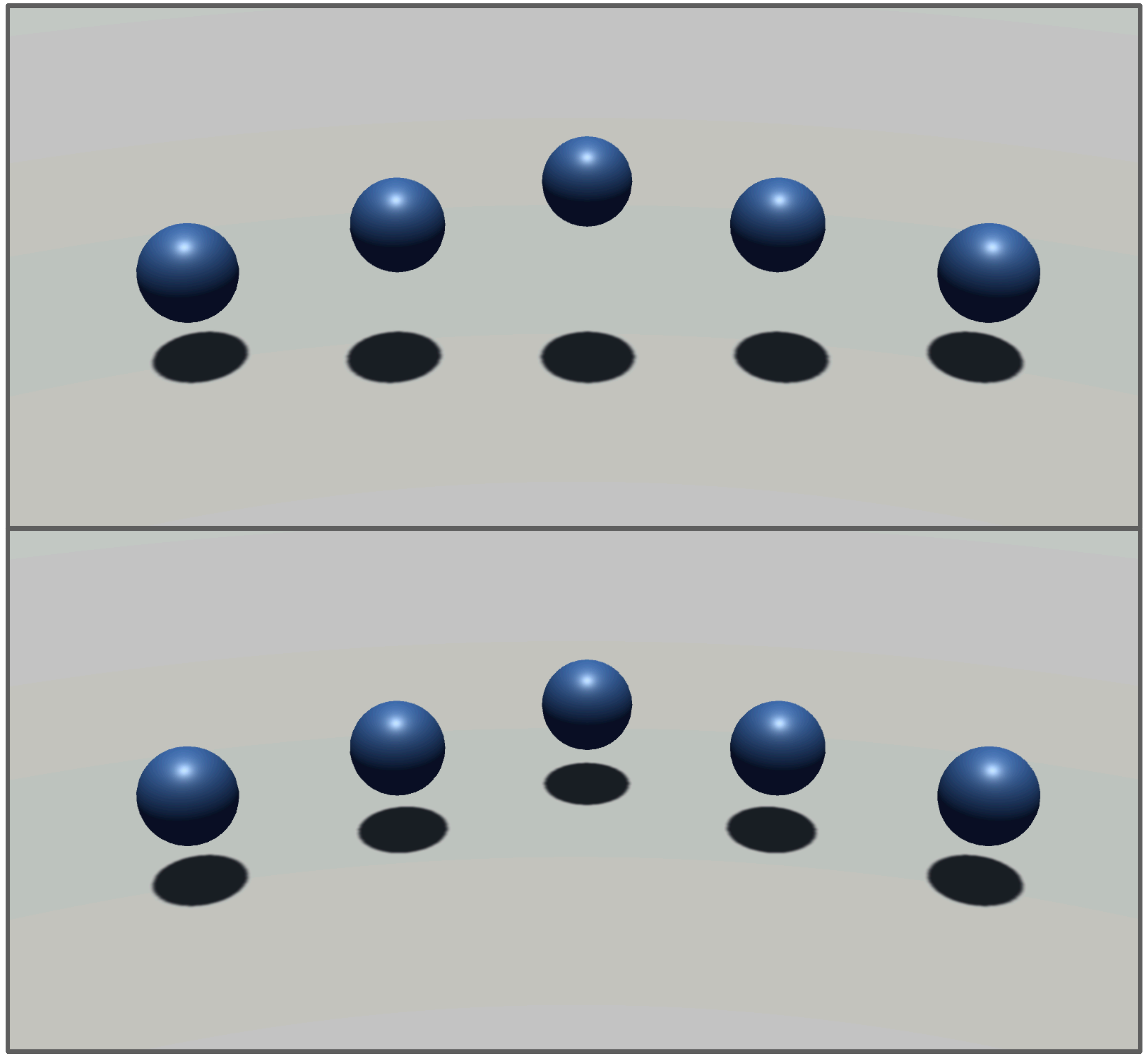
Color Constancy

- The goal of vision is recognizing objects' shapes and material properties under different lighting conditions
- Perception of "lightness or color constancy": e.g.: the checker board interpretation is strongly reinforced whether in shadows or not
- Color constancy is not universal and may fail with highly saturated lights or ambiguous context





Shadows



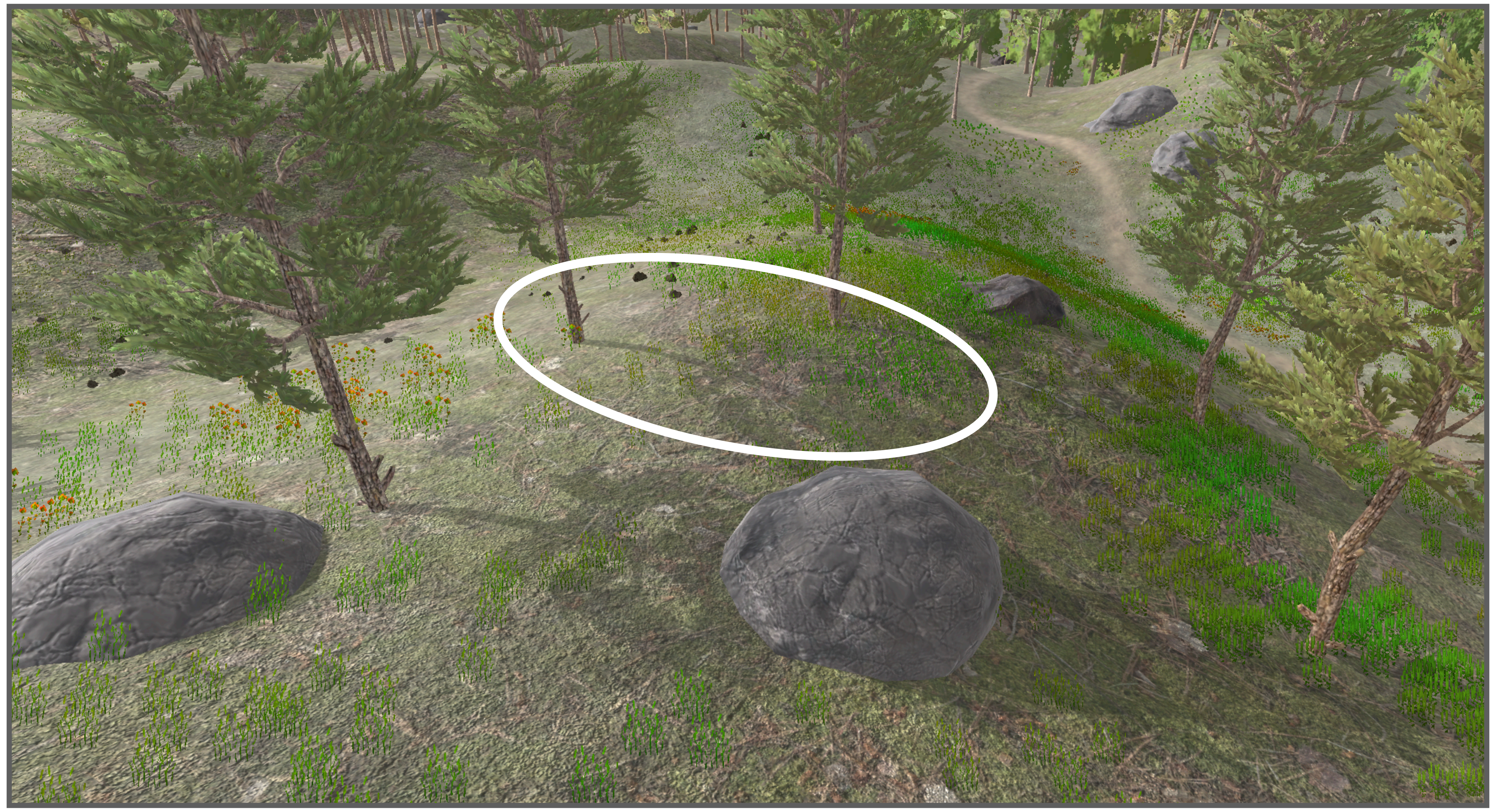
- Shadows are generally not interpreted as shapes
- Shadows provide strong positional clues (sometimes stronger than perspective clues)
- Shadows edges are nearly always softer compared to object edges
- Generally we have a perceptual bias of "darker means indented"

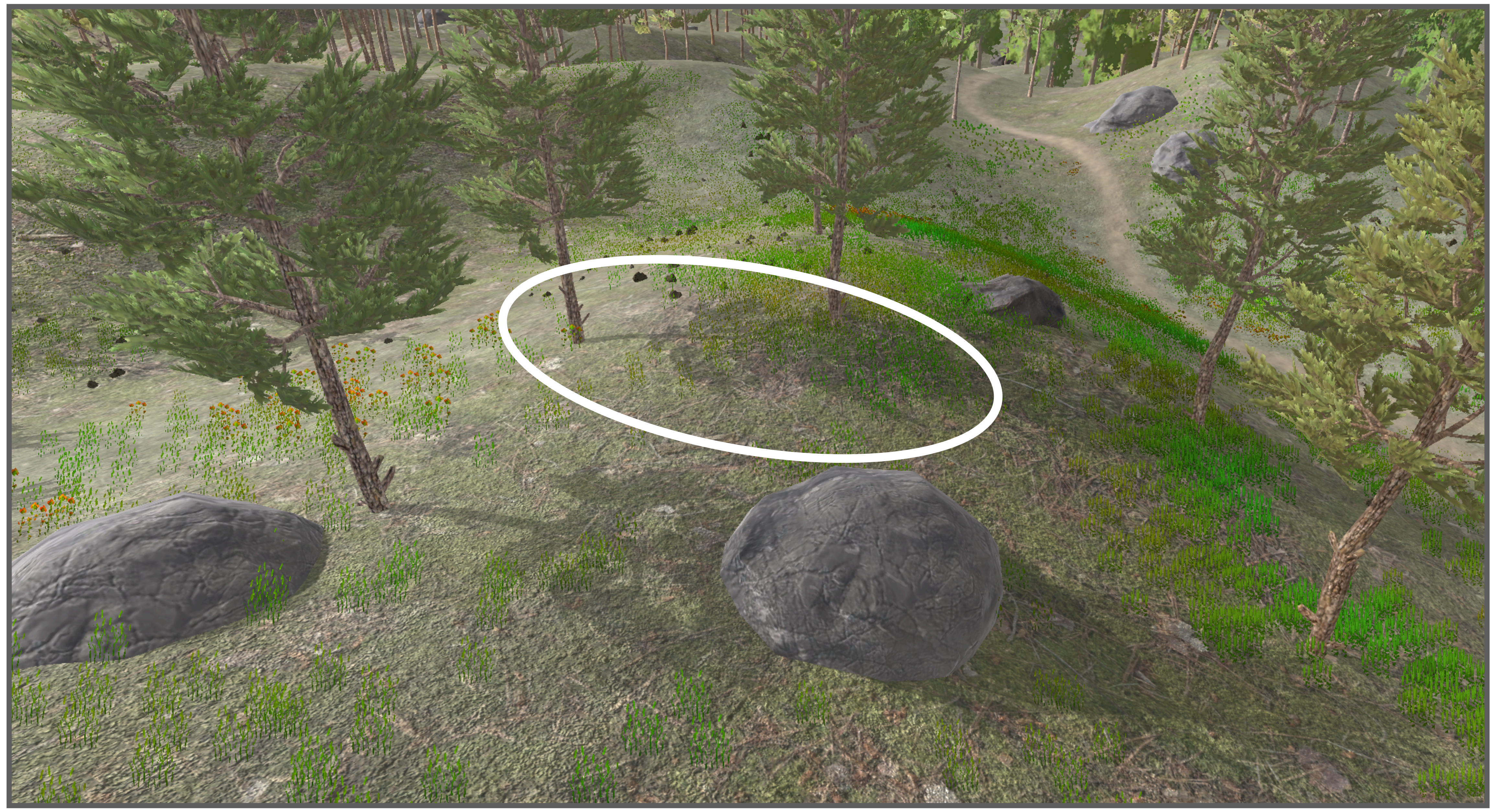
Shadows and Illumination Direction

- Directional inconsistencies in shadows and illumination direction are not always salient
- Humans in experiments may not pay attention to inconsistencies as large of 45°

Shadows and Illumination Direction

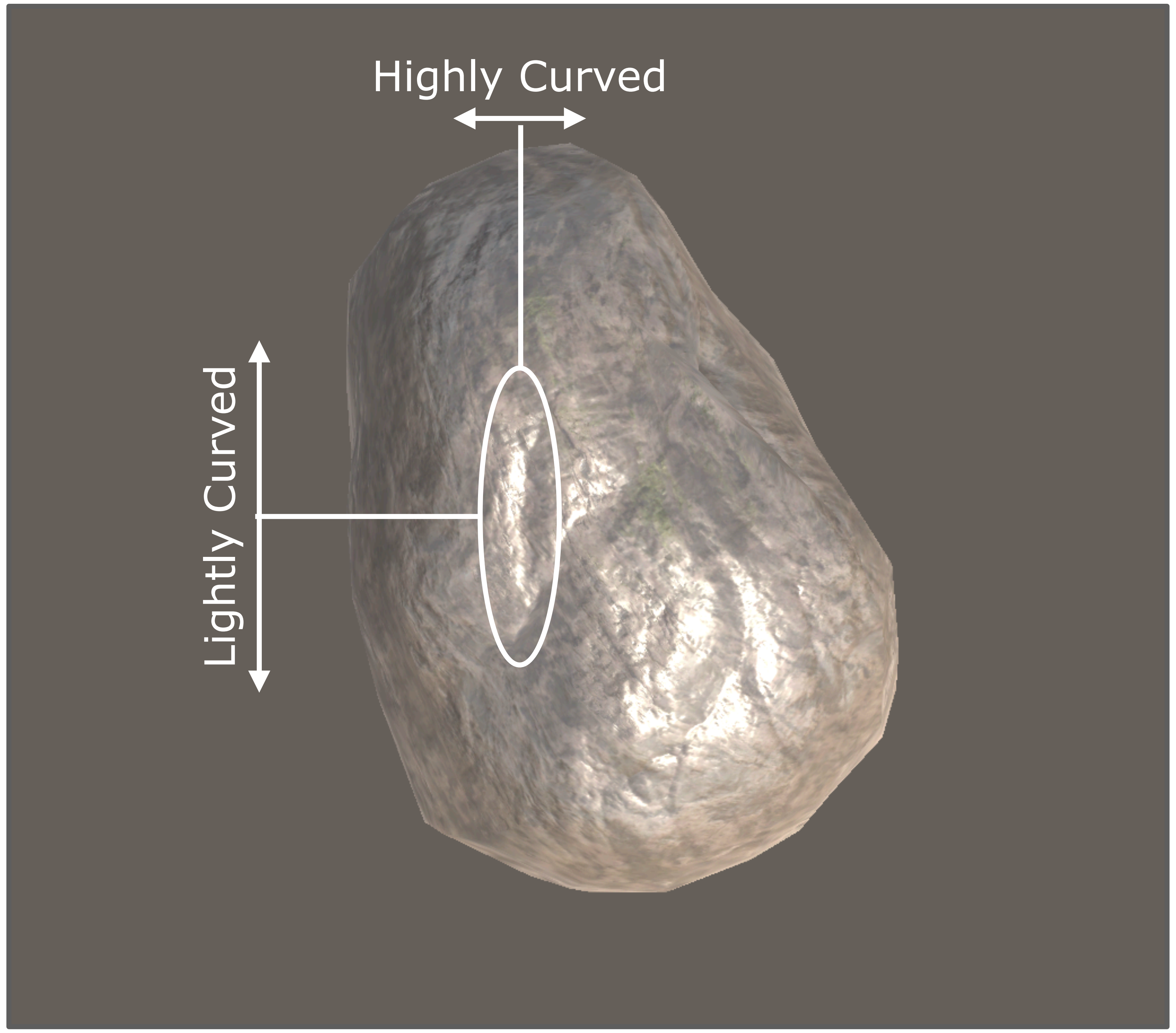
- Directional inconsistencies in shadows and illumination direction are not always salient
- Humans in experiments may not pay attention to inconsistencies as large of 45°





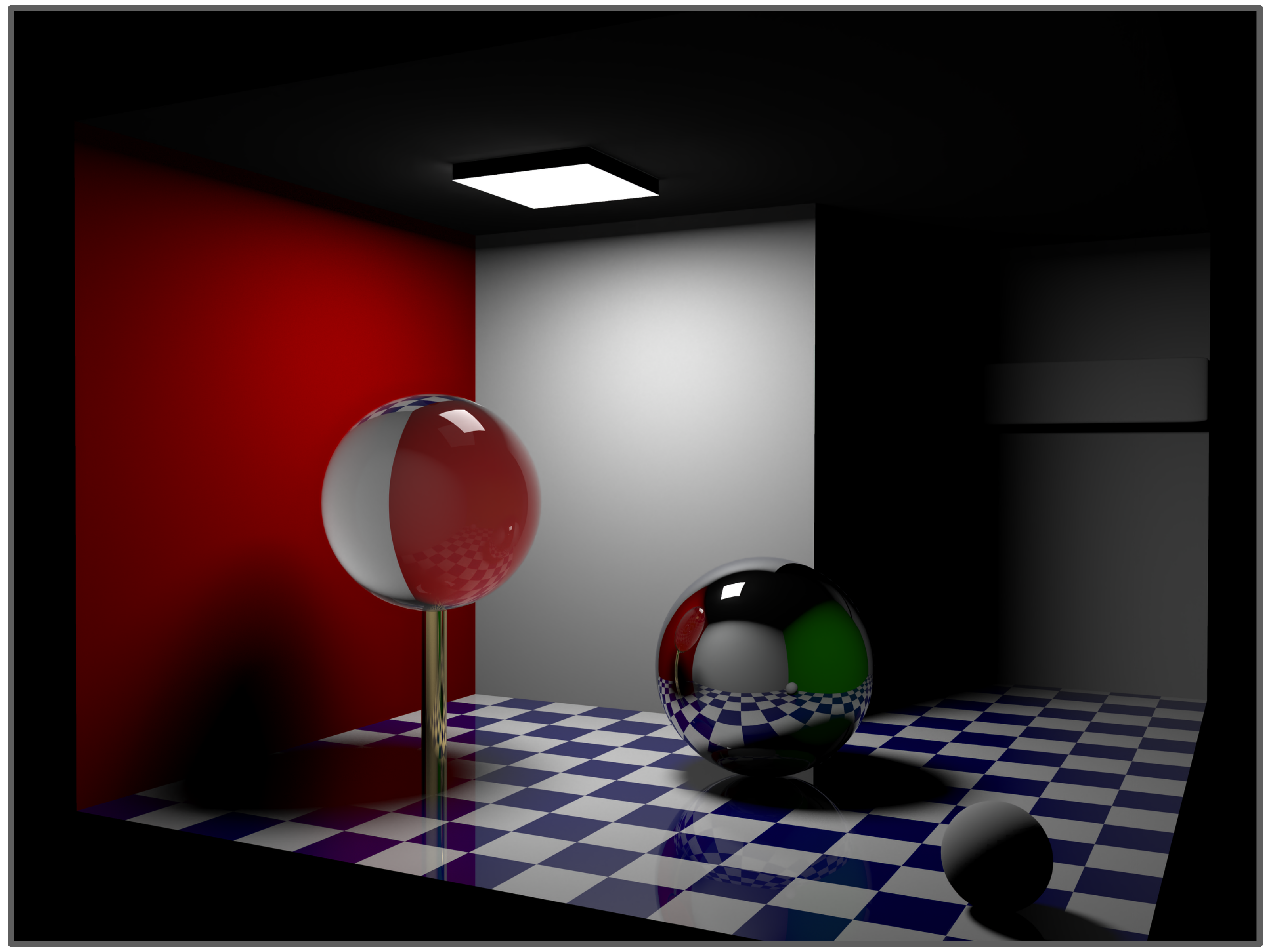
Specular Highlights

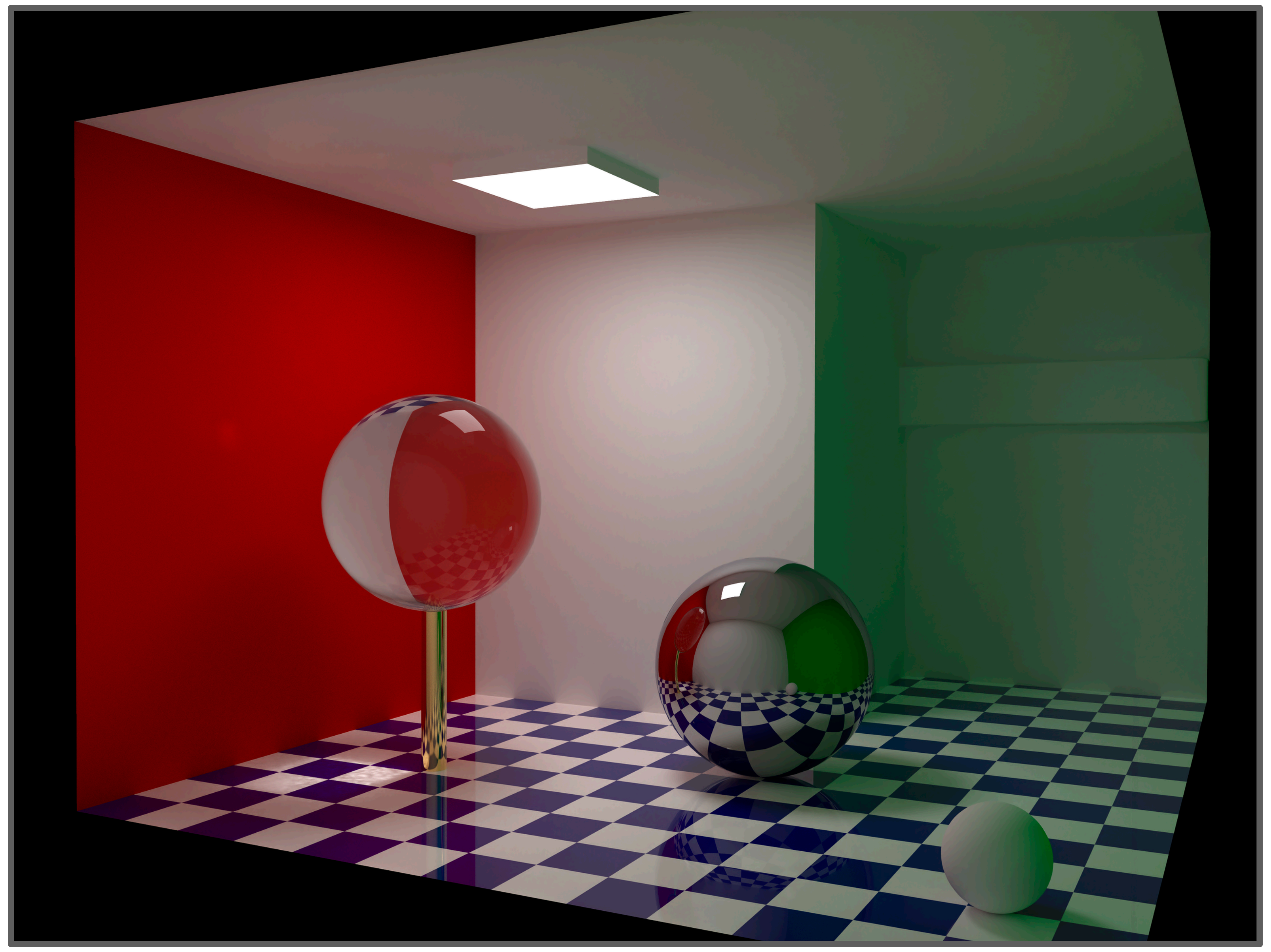
- Specular highlights are always much brighter compared to Lambertian type of reflection
- They help perceive object shape — being thinner in the direction of higher curvature
- Specular highlights spread perception of glossiness beyond their visual



Perception of Global Illumination

- Scenes are likely segmented in our perception into areas where illumination is perceived as mostly the same
- There is evidence that people mostly perceive influence from one or two light directions per object
- Diffuse to diffuse interactions (e.g. small color bleeding) may be less perceptually salient





Reference

Enhancing Photorealism Enhancement

Stephan R. Richter, Hassan Abu AlHaija, and Vladlen Koltun

2021

Enhancing photorealism enhancement

Stephan R. Richter, Hassan Abu AlHaija, and Vladlen Koltun

Abstract—We present an approach to enhancing the realism of synthetic images. The images are enhanced by a convolutional network that leverages intermediate representations produced by conventional rendering pipelines. The network is trained via a novel adversarial objective, which provides strong supervision at multiple perceptual levels. We analyze scene layout distributions in commonly used datasets and find that they differ in important ways. We hypothesize that this is one of the causes of strong artifacts that can be observed in the results of many prior methods. To address this we propose a new strategy for sampling image patches during training. We also introduce multiple architectural improvements in the deep network modules used for photorealism enhancement. We confirm the benefits of our contributions in controlled experiments and report substantial gains in stability and realism in comparison to recent image-to-image translation methods and a variety of other baselines.

arXiv:2105.04619v1 [cs.CV] 10 May 2021

1 Introduction

PHOTOREALISM has been the defining goal of computer graphics for half a century. In 1977, Newell and Blinn [1] surveyed a decade of work on this problem. In the ensuing four decades, substantial further progress has been made, due in part to physically based simulation of light transport [2], principled representation of material appearance [3], [4], and photogrammetric modeling [5]. These techniques and their approximations have been integrated into real-time rendering pipelines, substantially advancing the realism of computer games [6]. Nevertheless, a look at even the most sophisticated real-time games will quickly reveal that photorealism has not been achieved. An ineffable difference in the appearance of simulation and reality remains.

In recent years, a complementary set of techniques has been developed in computer vision and machine learning. These techniques, based on deep learning, convolutional networks, and adversarial training, bypass physical modeling of geometric layout, material appearance, and light transport. Instead, images are synthesized by convolutional networks trained on large datasets. These techniques have been used to synthesize representative images from a given domain [7], [8], [9], to convert semantic label maps to photographic images [10], [11], [12], [13], [14], [15], [16], and to attempt to bridge the appearance gap between synthetic and real images [17], [18], [19], [20], [21], [22], [23], [24], [25]. Images synthesized by these approaches capture aspects of photographic appearance that often elude even state-of-the-art computer games. On the flip side, these approaches are largely disconnected from the rendering pipelines that drive computer games, can be hard to control, and often produce jarring artifacts that would be unacceptable in production-quality media.

In this work, we take a step towards melding these two complementary routes to photorealism. We seek to build on the infrastructure developed in the production of modern games and enhance their photorealism via techniques developed in the deep learning community. Our starting point is a set of intermediate buffers (G-buffers) produced by game engines during the rendering process [6], [26]. These buffers provide detailed information on geometry, materials, and lighting in the scene. We train convolutional networks with these auxiliary inputs to enhance the realism of images produced by the rendering pipeline. To integrate these buffers into the photorealism enhancement flow, we design new network components that modulate features from a rendered image according to information extracted from the buffers.

We also seek to eliminate artifacts that can be seen in the results of prior deep-learning approaches, which often hallucinate objects. To this end, we analyze the datasets that are commonly used for photorealism enhancement. Our analysis reveals that their scene layouts differ in ways that can explain artifacts commonly seen in prior work. To better align the datasets and alleviate the artifacts, we propose a new strategy for sampling image patches during training. We further design a new adversarial training objective that facilitates enhancements that are geometrically and semantically consistent with the content of the input image.

Combining all of our contributions, our approach significantly enhances the photorealism of rendered images (Fig. 1). It can add gloss to cars (1^{st} row), green parched hills (2^{nd} row), and rebuild roads (3^{rd} row). Training it with different real-world image collections (e.g., Cityscapes [28], KITTI [29], or Mapillary Vistas [30]) expresses the corresponding visual styles in the output (Fig. 2).

Our analysis further suggests that standard metrics confound differences in style and content. Motivated by this observation, we develop a new family of metrics that mitigate the effect of mismatched scene layouts and provide a finer-grained assessment of realism at multiple levels.

We compare the presented approach against a broad array of strong baselines that represent diverse perspectives on photorealism enhancement. We also conduct a perceptual experiment to assess photorealism. The results indicate that our approach consistently produces the most realistic results, by a wide margin. In all experiments, our approach outperforms all baselines and sets a new state of the art in photorealism enhancement.

2 Related Work

Photorealistic images can be synthesized by simulating all the physical processes involved in image formation. However, this simulation is computationally expensive, may not be feasible at interactive rates, and requires physically accurate

Why does it matter?

- Human vision is somewhat biased to interpret scenes as lit from above and to interpret dark spots as indentations
- Specular highlights provide glossiness and curvature clues
- Shadows give strong positional clues and need to have soft edges
- Moderate inconsistencies in illumination direction and shadow direction are not immediately salient

Thanks!

sergei.savchenko@bethesdastudios.com