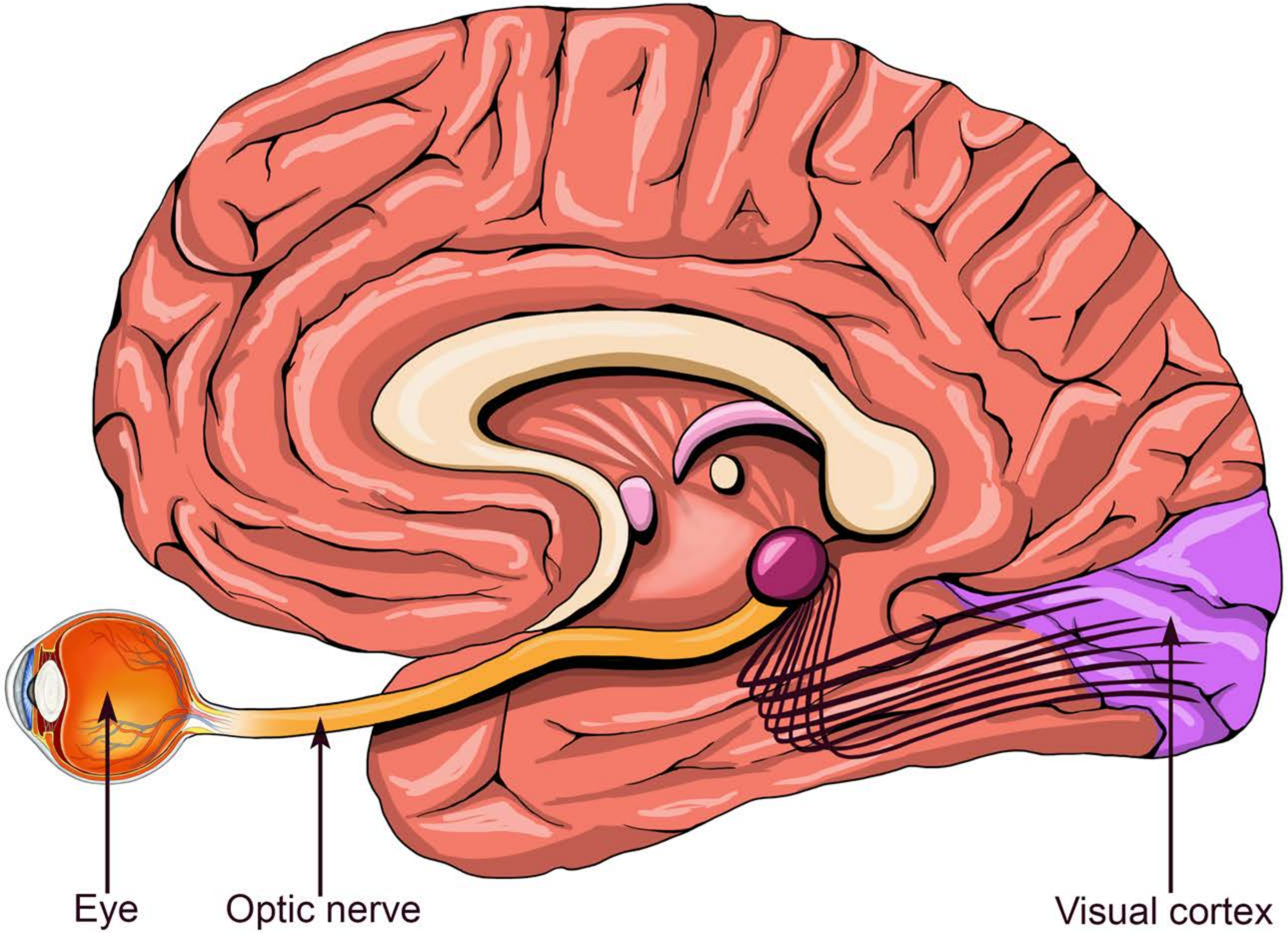
Human visual system and perception

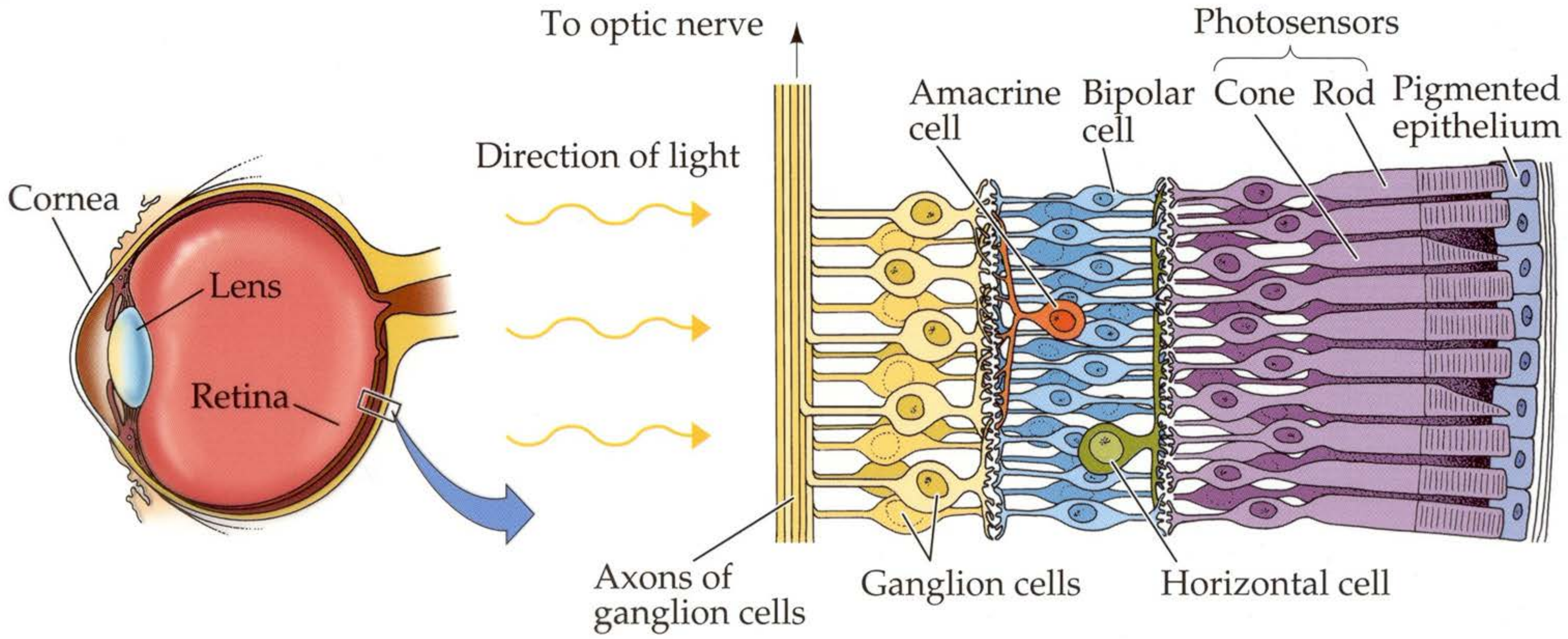
Javier Vazquez-Corral
Universitat Pompeu Fabra, Spain

javier.vazquez@upf.edu

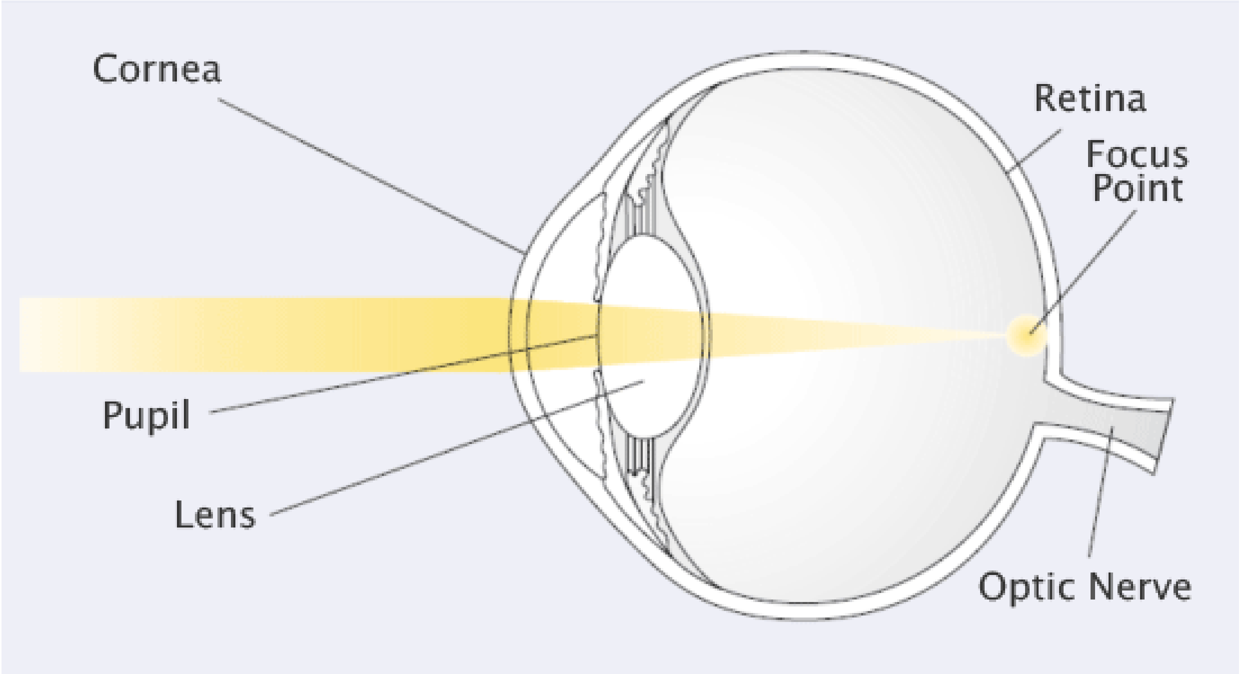
Slide credits: Marcelo Bertalmío and David Kane

Biological basis

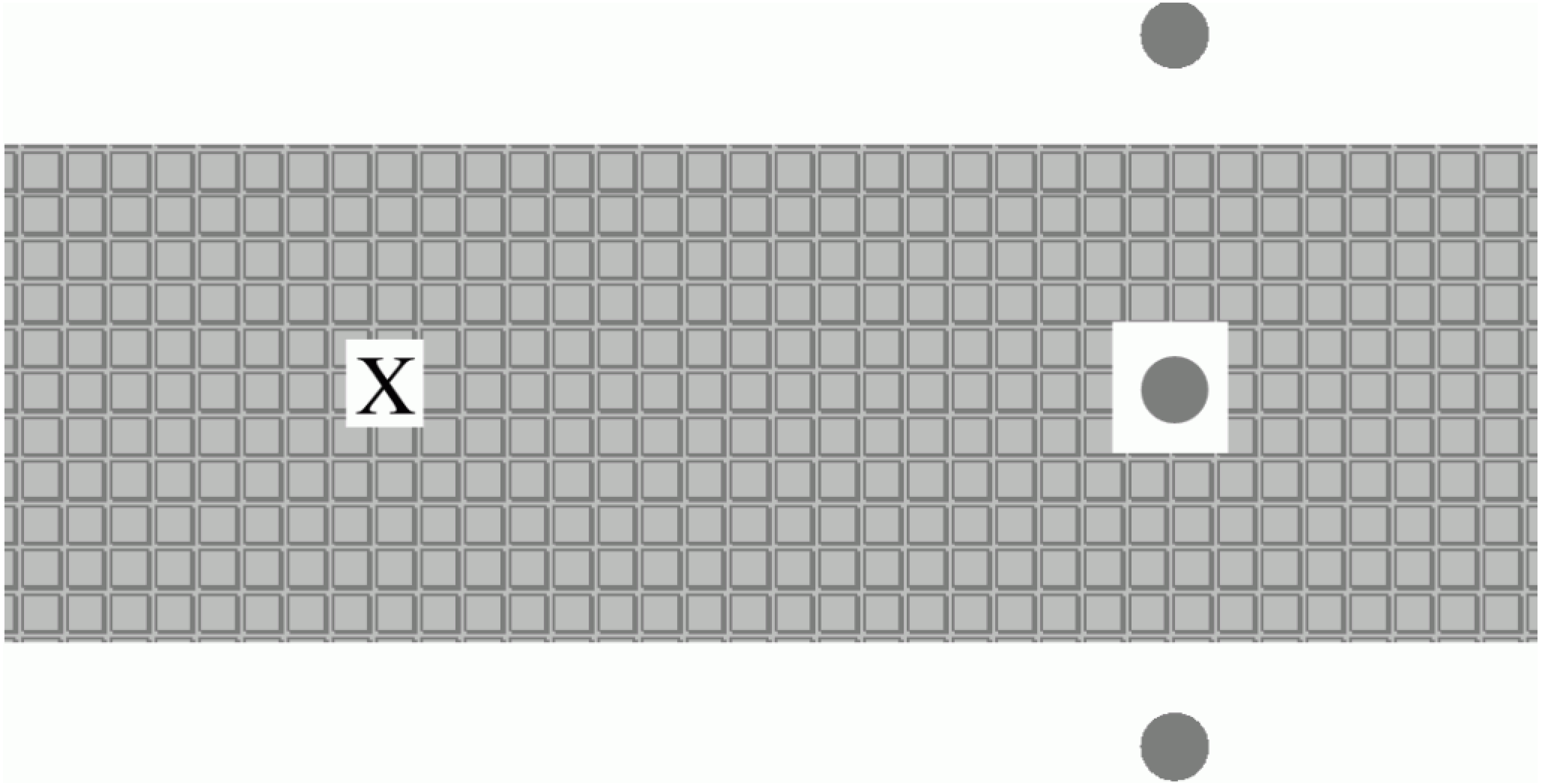

Eye
Optic nerve
Visual cortex


To optic nerve
Photosensors
Amacrine cell
Bipolar cell
Cone
Rod
Pigmented epithelium
Direction of light
Cornea
Lens
Retina
Axons of ganglion cells
Ganglion cells
Horizontal cell

The eye

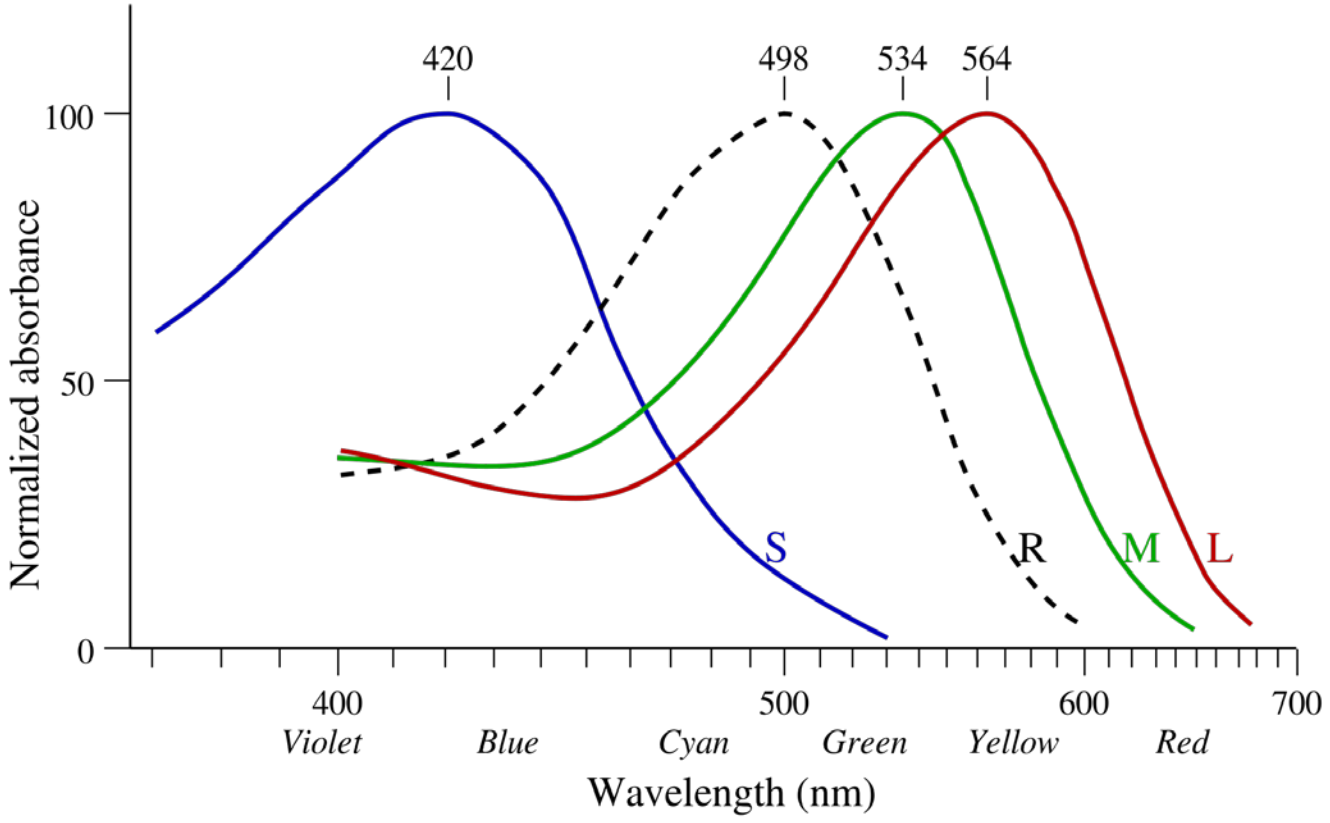


The blindspot



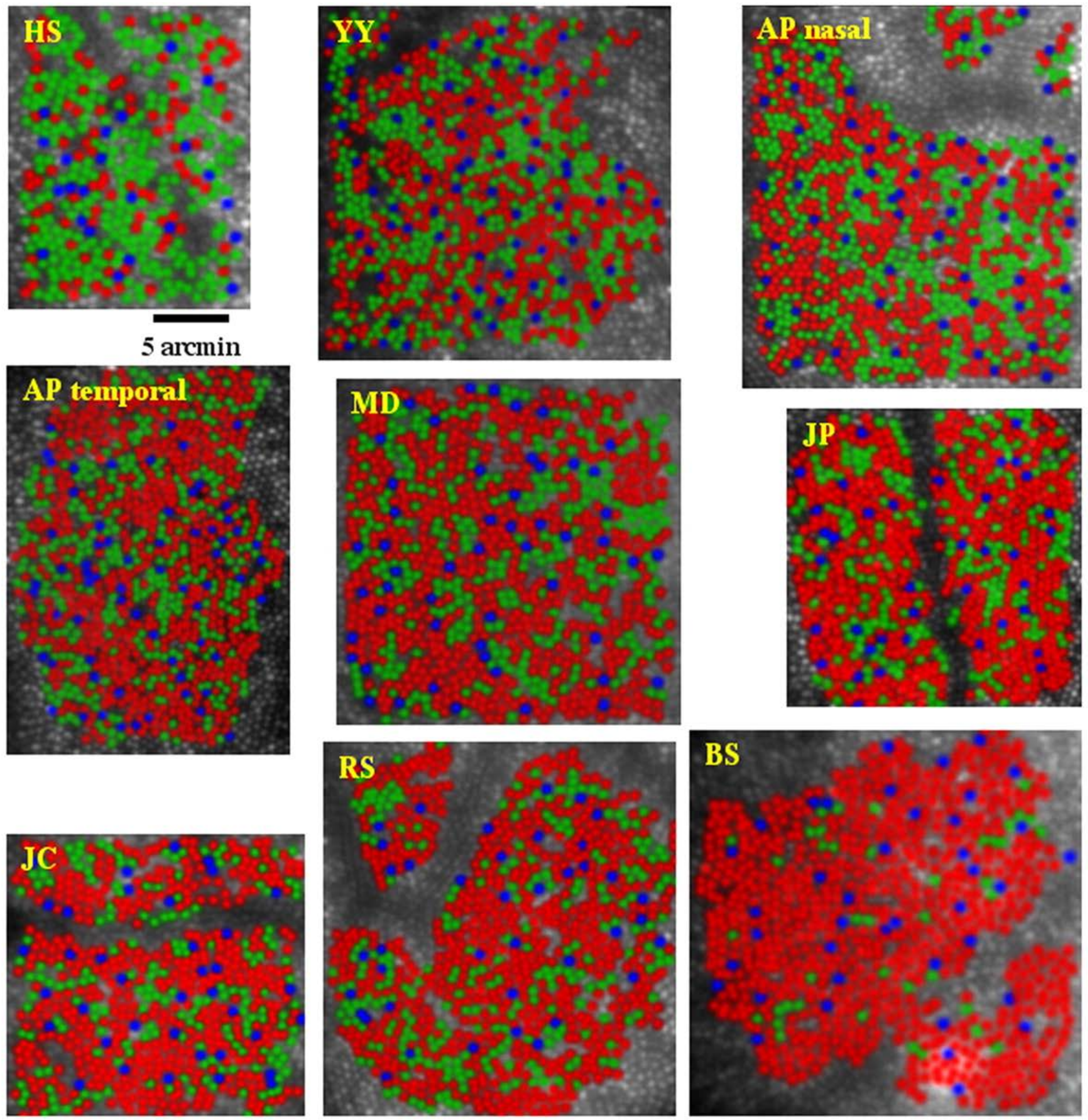
Filling in

Rod and Cone vision



- Rods are more sensitive to low luminance levels but to only one wavelength
- Cones are sensitive to different wavelength
  - Short, medium and long, roughly corresponding to blue green and red

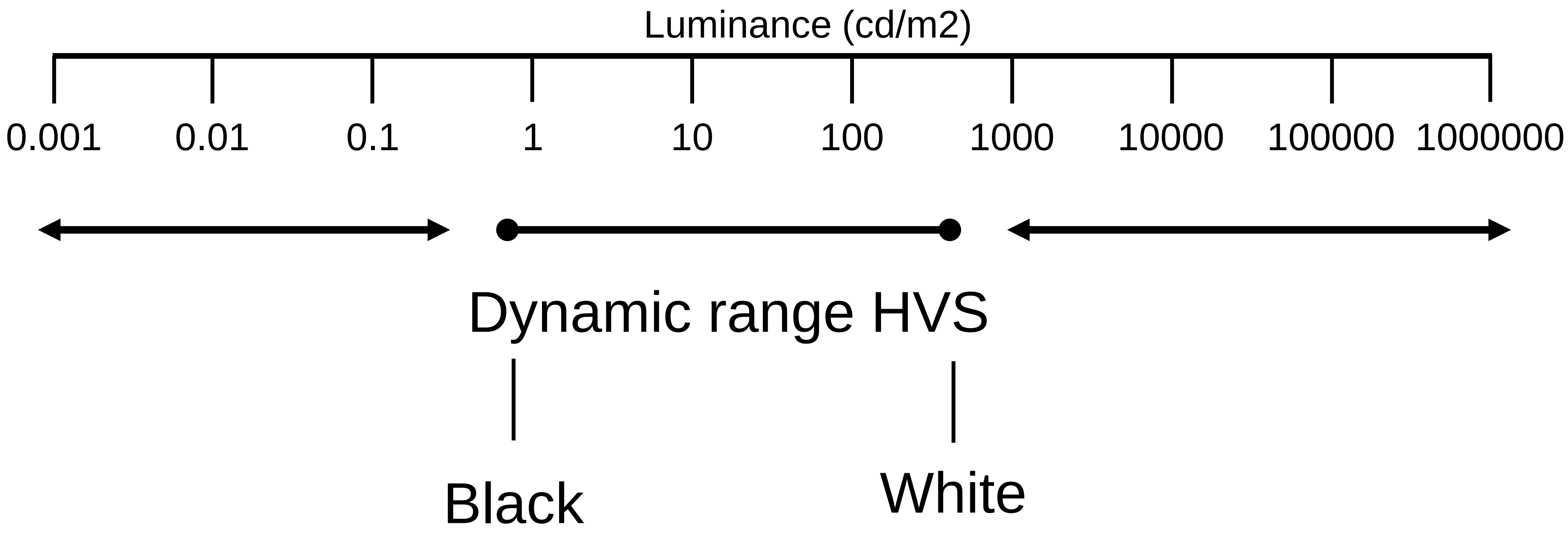
Rod and Cone vision



Scotopic, mesozoic & photopic vision

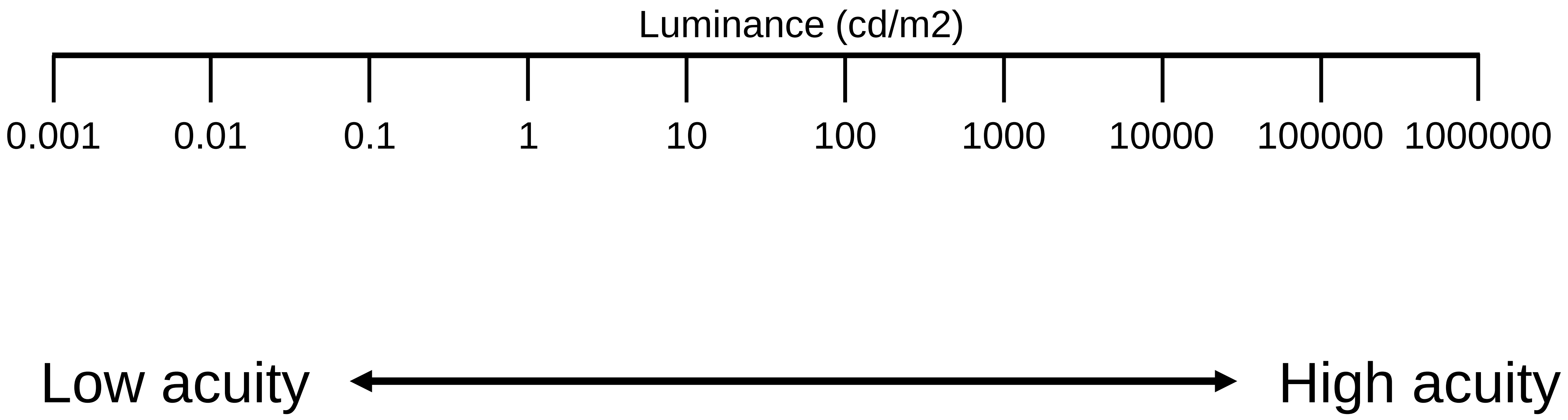
Condition	cd/m2		Cells
Clear night sky	0.001	Scotopic	Rods
Quarter moon	0.01	Scotopic	Rods
Full moon	0.1	Mesozoic	Rods & Cones
Late twilight	1	Mesozoic	Rods & Cones
Twilight	10	Photopic	Cones
Heavy overcast	100	Photopic	Cones
Overcast sky	1000	Photopic	Cones
Full daylight	10000	Photopic	Cones
Direct sunlight	100000	Photopic	Cones

Dynamic range/luminance adaption

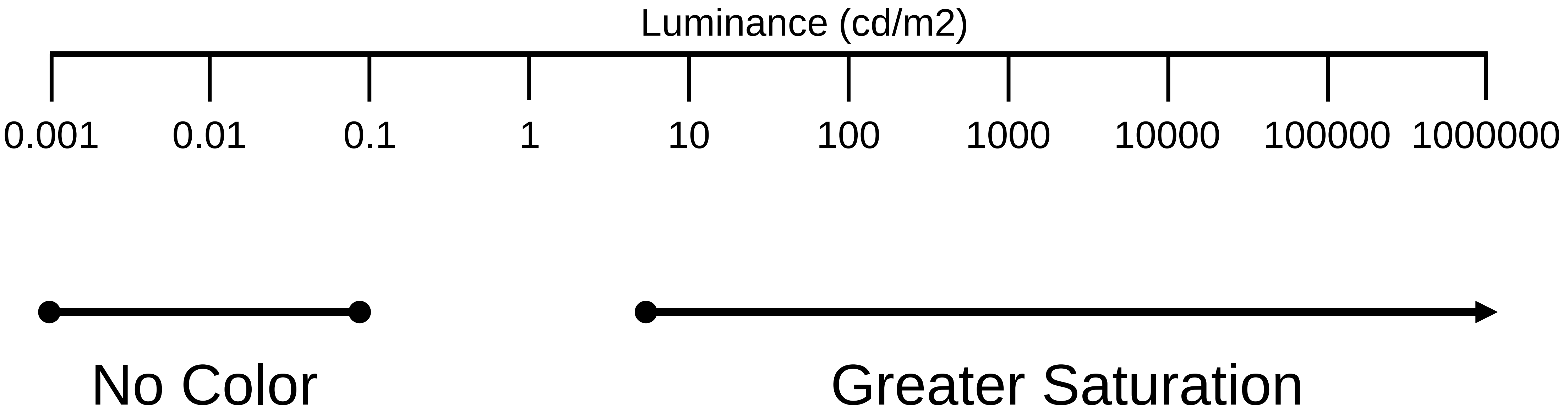


- Insensitive to absolute luminance

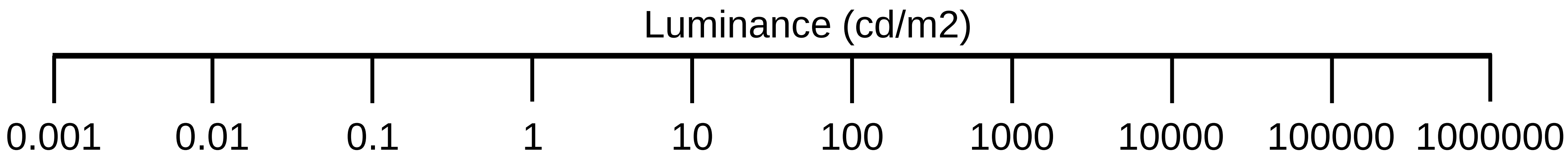
Dynamic range/luminance adaption



Dynamic range/luminance adaption



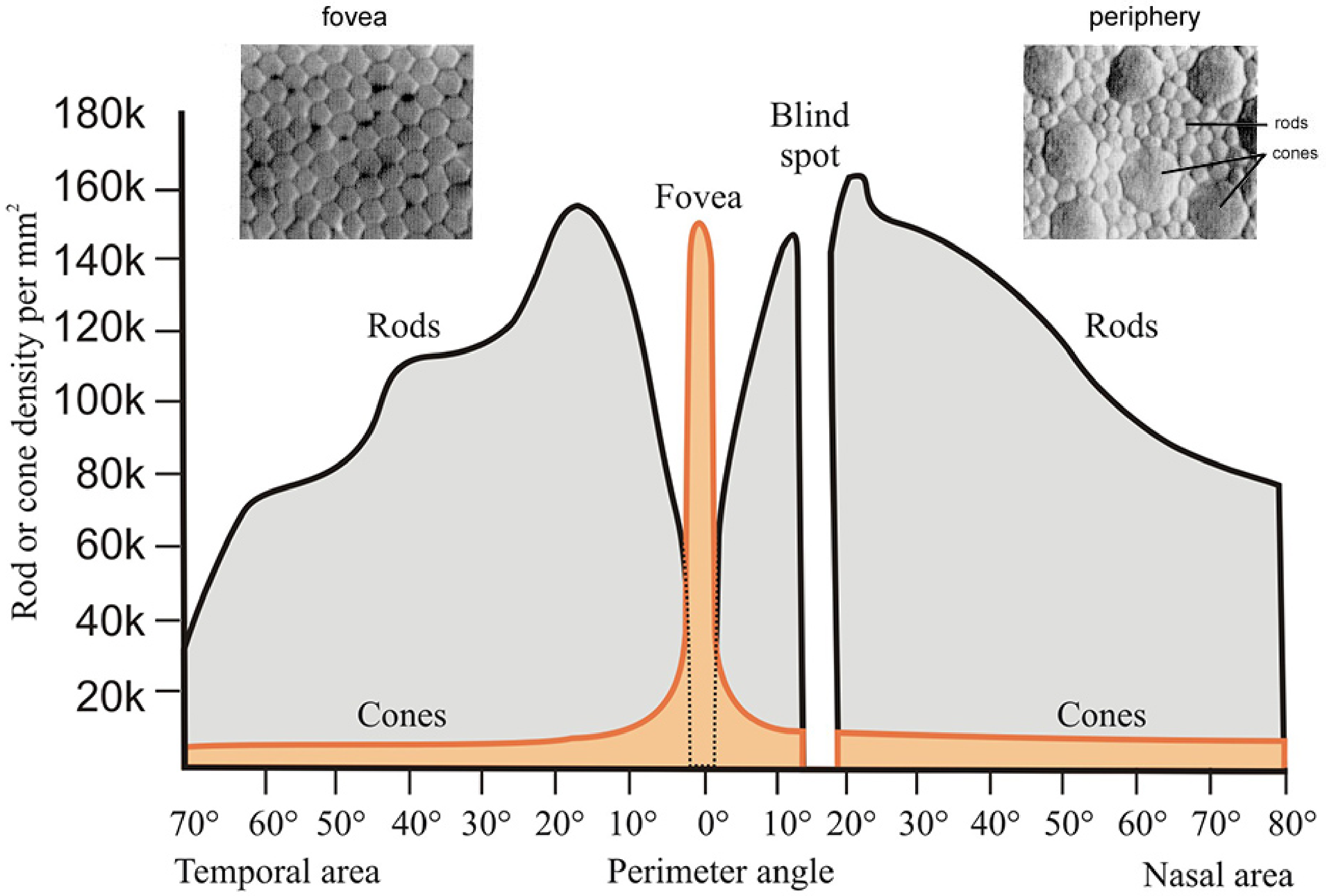
Dynamic range/luminance adaption



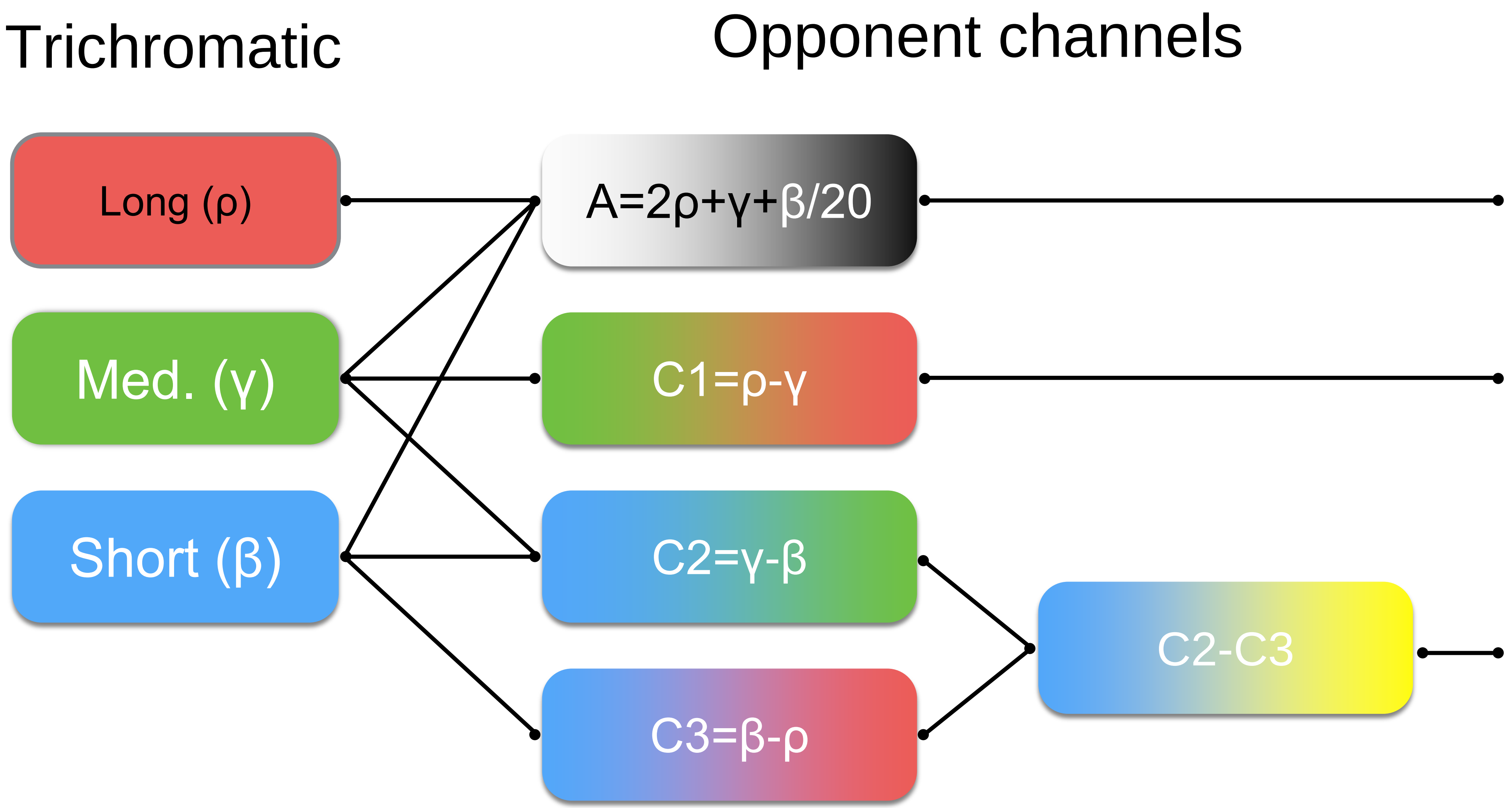
Purkinje effect

As light intensity decreases, red objects are perceived to fade faster than blue objects of the same brightness.

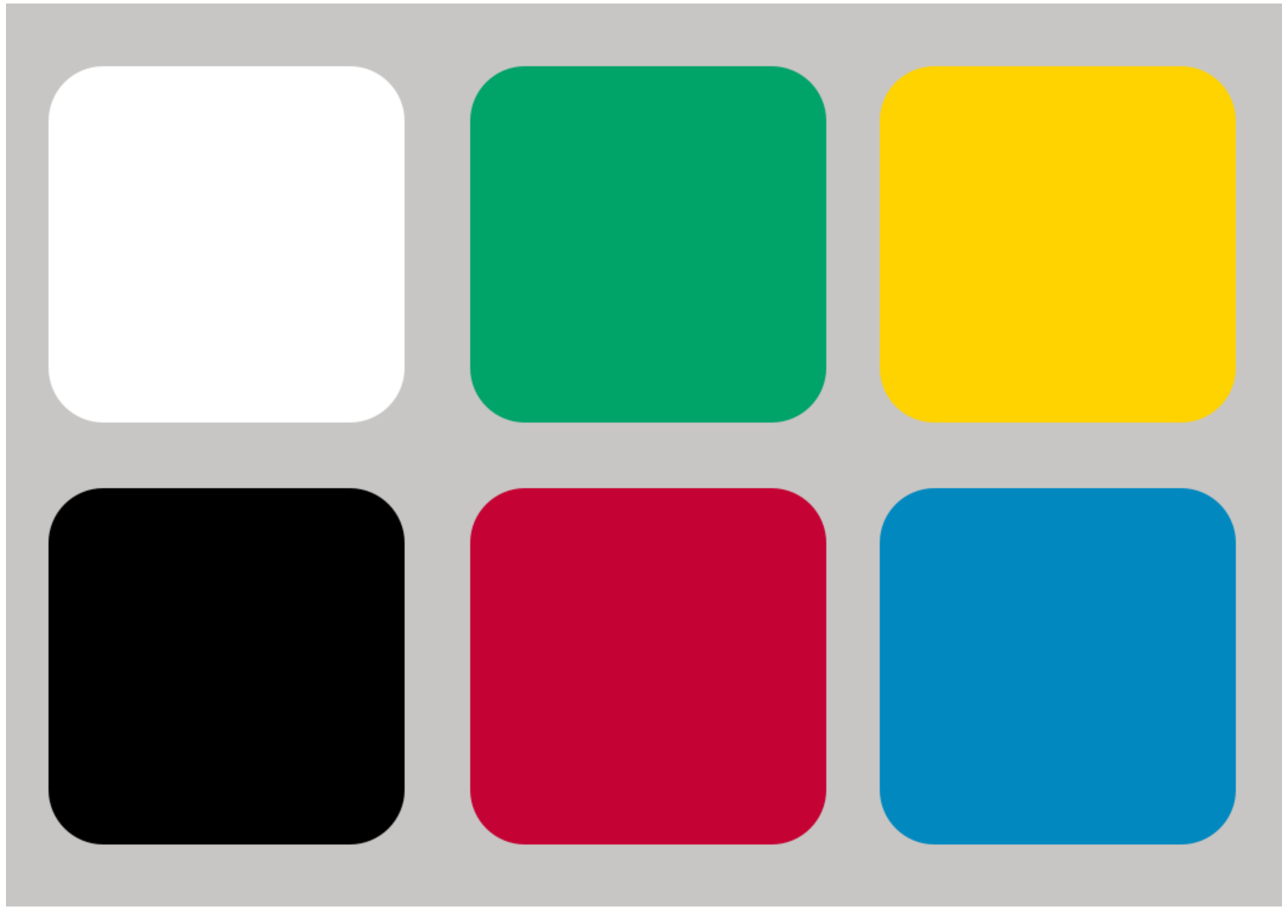
Rod and Cone Vision



Opponent Colors

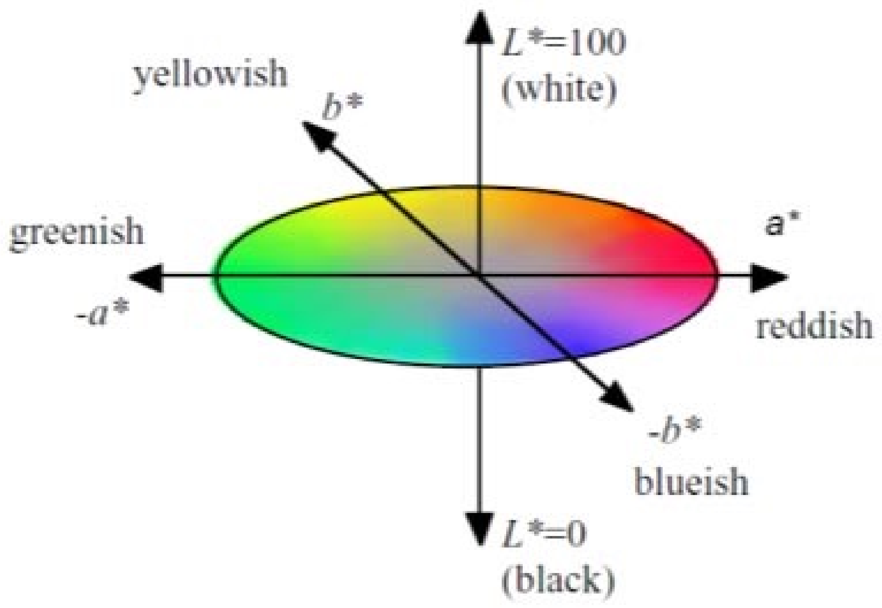


Opponent Colors



Opponent Colors

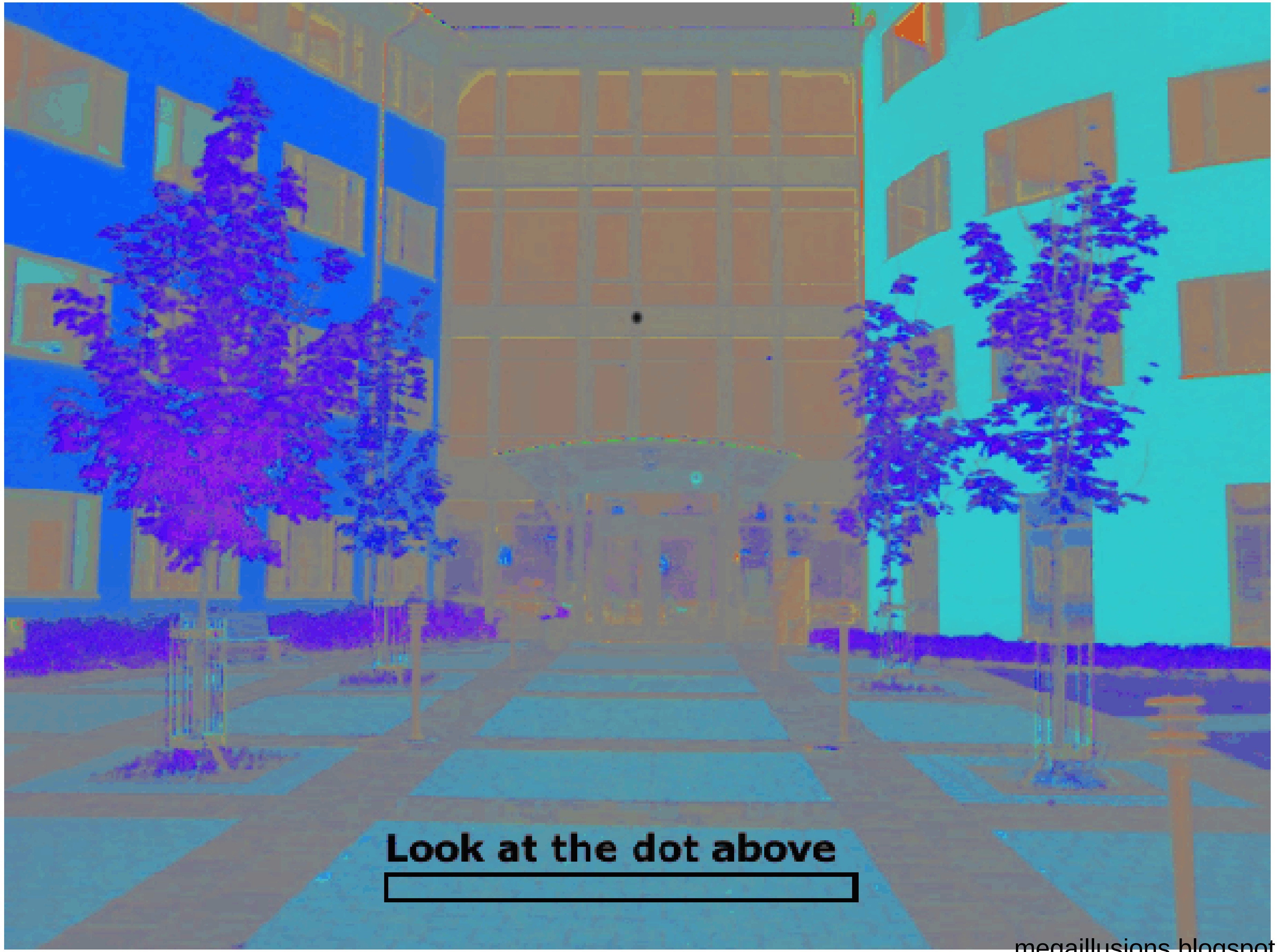
- People don’t perceive
  - Reddish-greens
  - Blueish-yellows
- CIE-LAB color space



Rod and Cone Vision

Rod and Cone Vision

Rod and Cone Vision



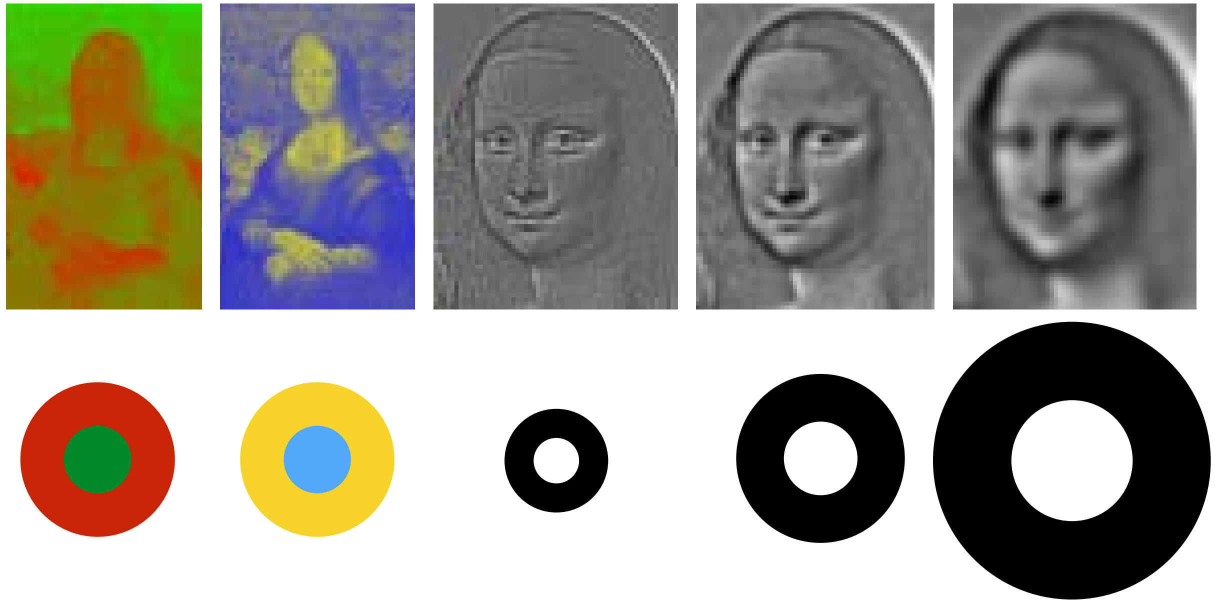
megaillusions.blogspot.com

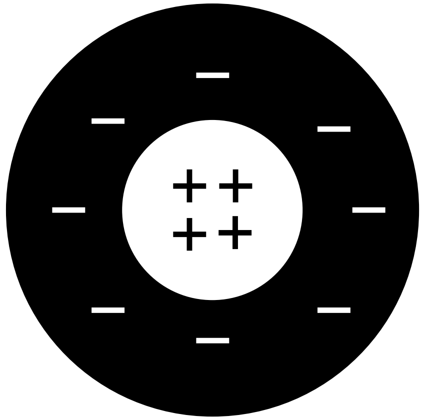
Basic Transformations

<u>The retina</u>

- Cones: Red/Green/Blue
- Rods: Grayscale

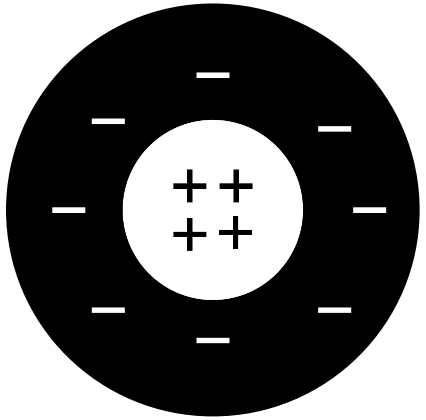
- Ganglion Cells:
  Centre-Surround

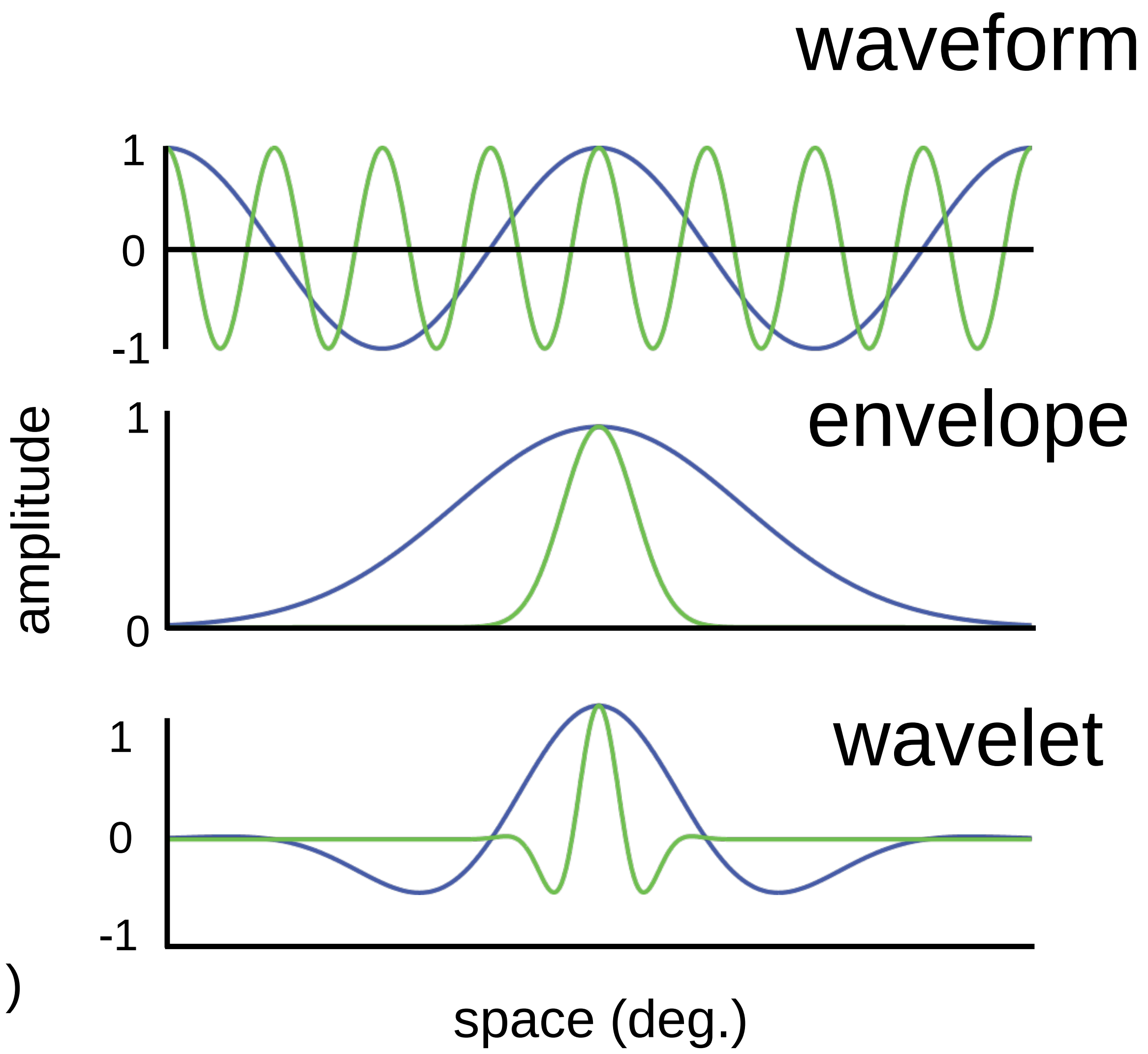




Basic Transformations

The centre surround

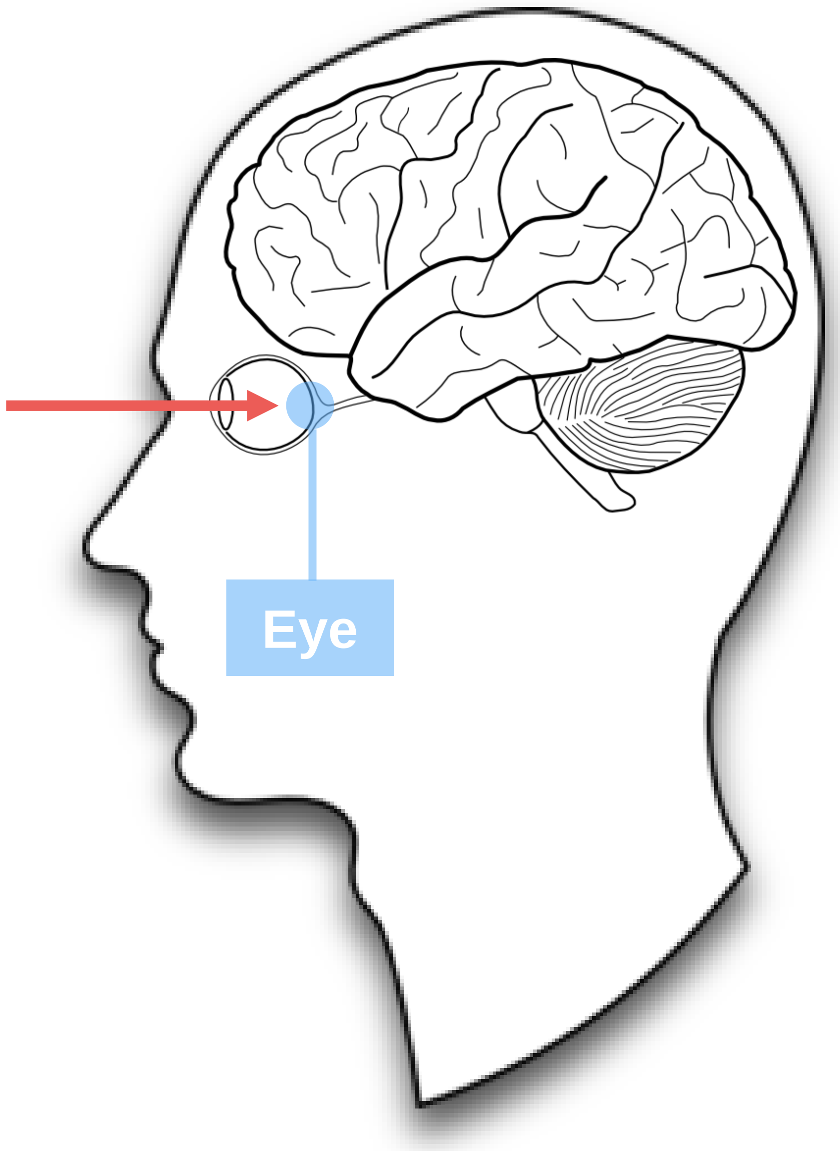




Low spatial frequency (coarse detail)

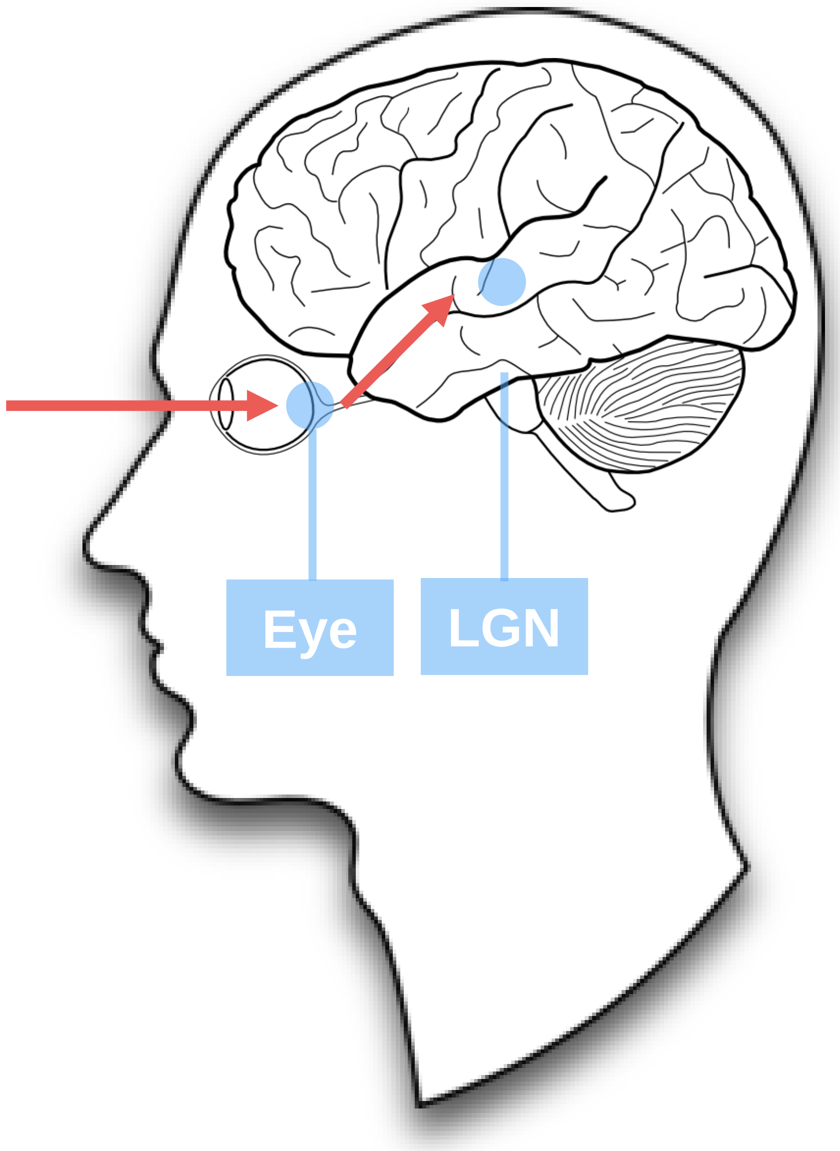
High spatial frequency (fine detail)

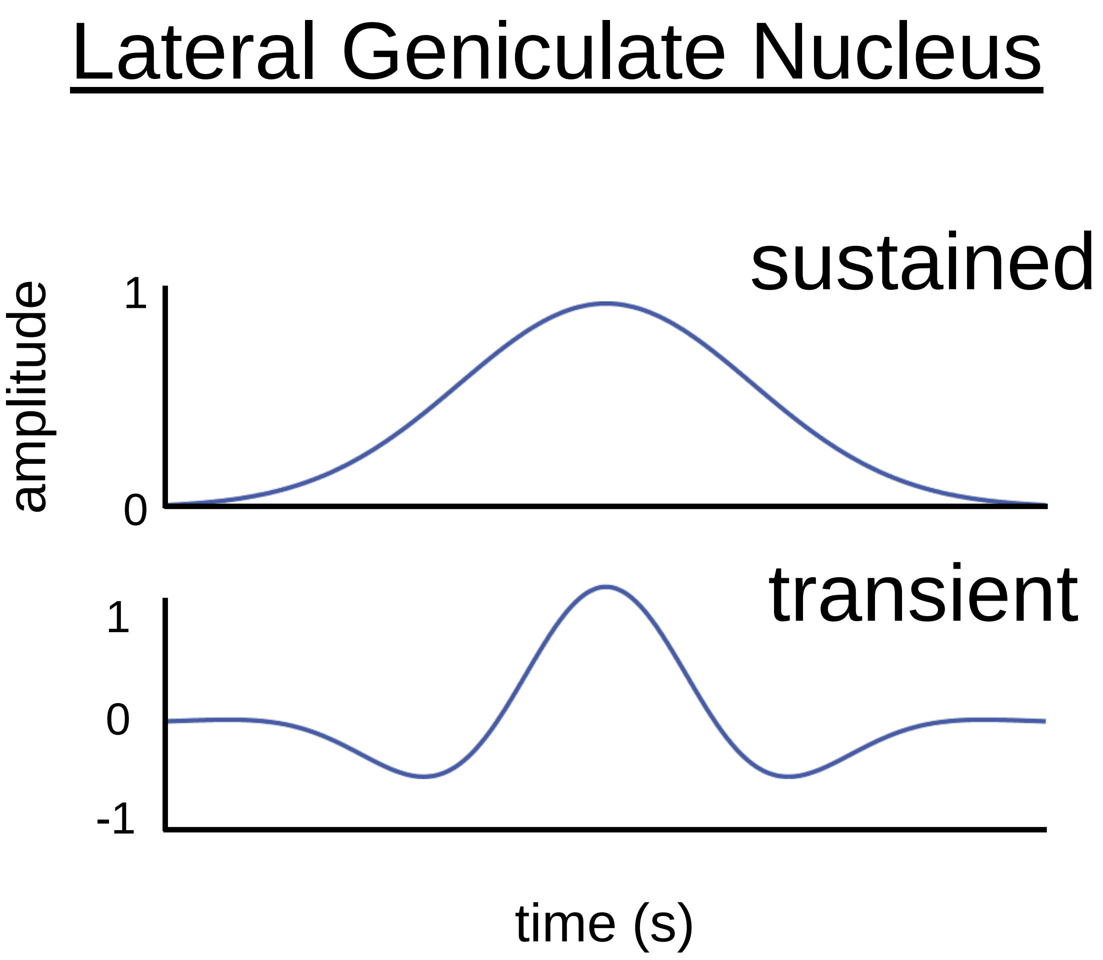
(very) Basic Overview



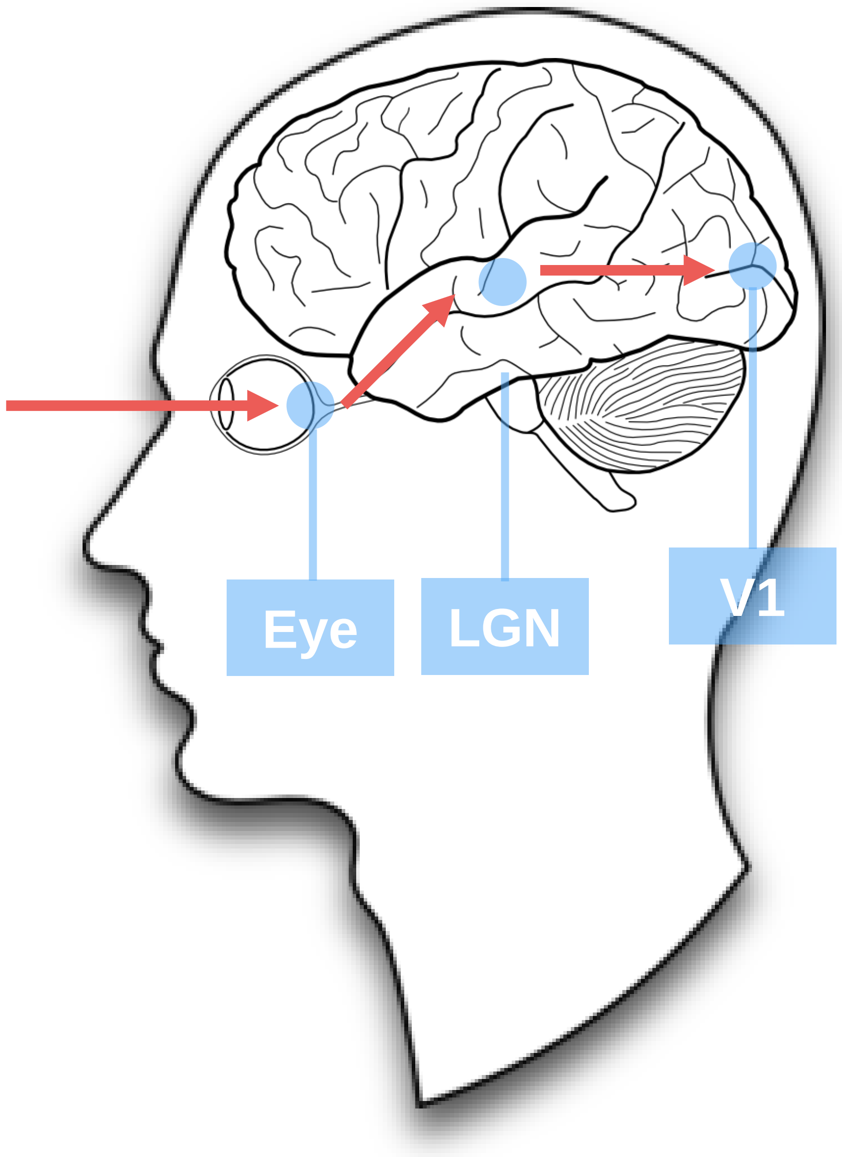
Retina

(very) Basic Overview

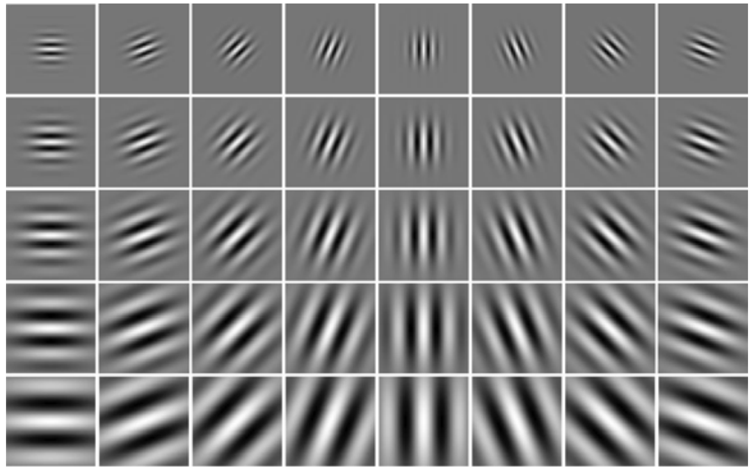




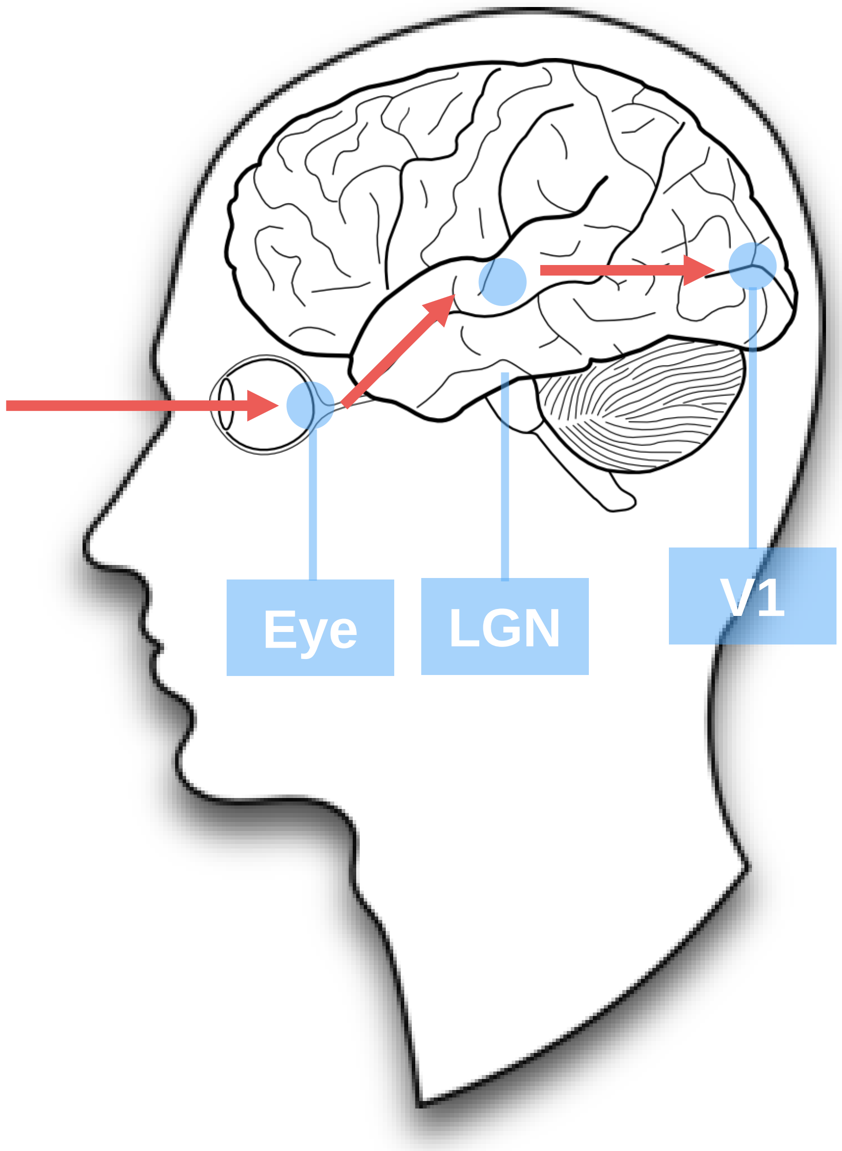
(very) Basic Overview



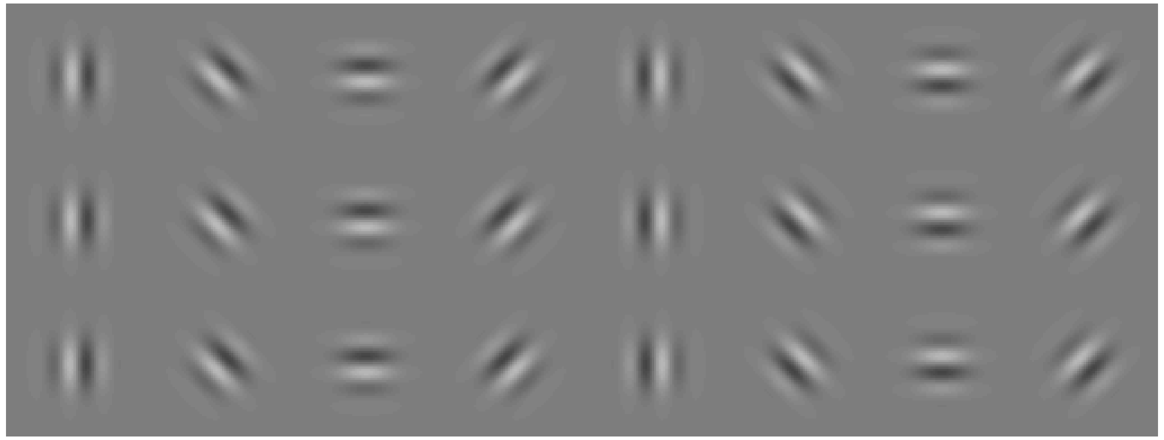
Visual Cortex: Area V1



(very) Basic Overview



Visual Cortex: Area V1

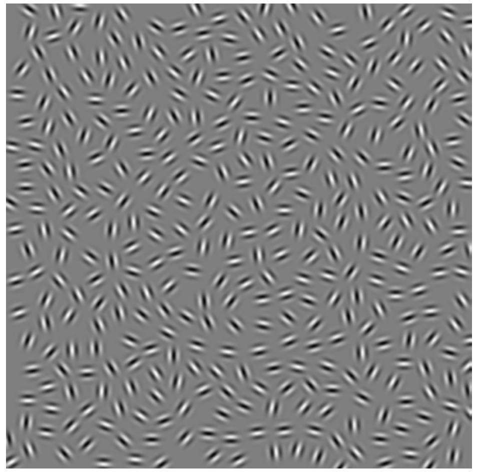


Motion is detected not by tracking objects, but through spatio-temporal energy.

(very) Basic Overview

<u>Examples</u>

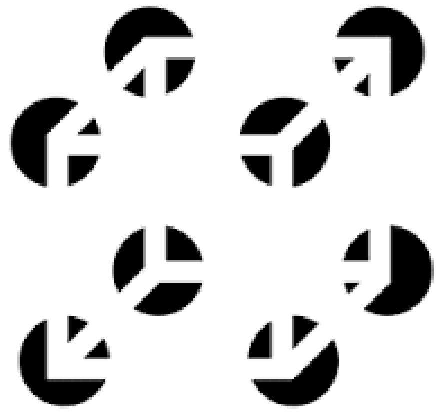
- Contours



(very) Basic Overview

Examples

- Illusory contours



(very) Basic Overview

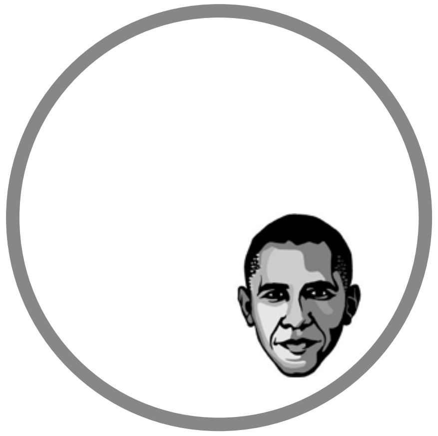
Examples

- Spatial invariance

(very) Basic Overview

Examples

- Spatial invariance

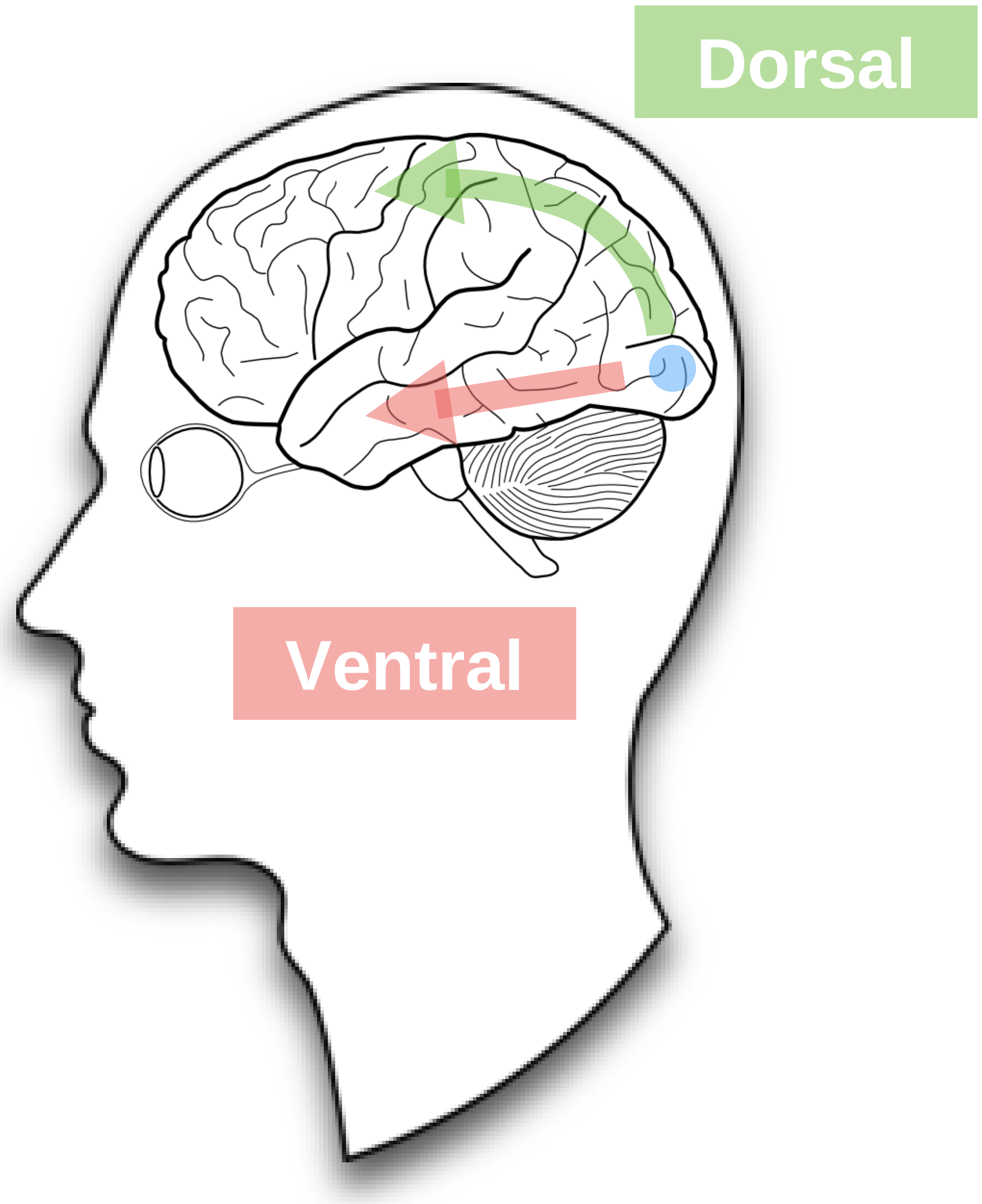


(very) Basic Overview

Examples

- Viewpoint invariance

(very) Basic Overview



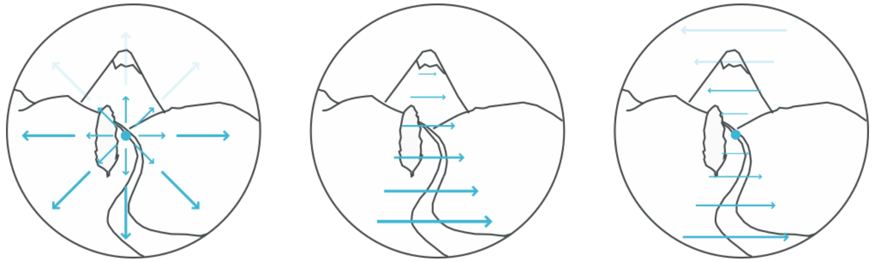
Feed Forward

- Dorsal stream
  - Identification
  - What?
- Ventral stream
  - Localisation
  - Where?

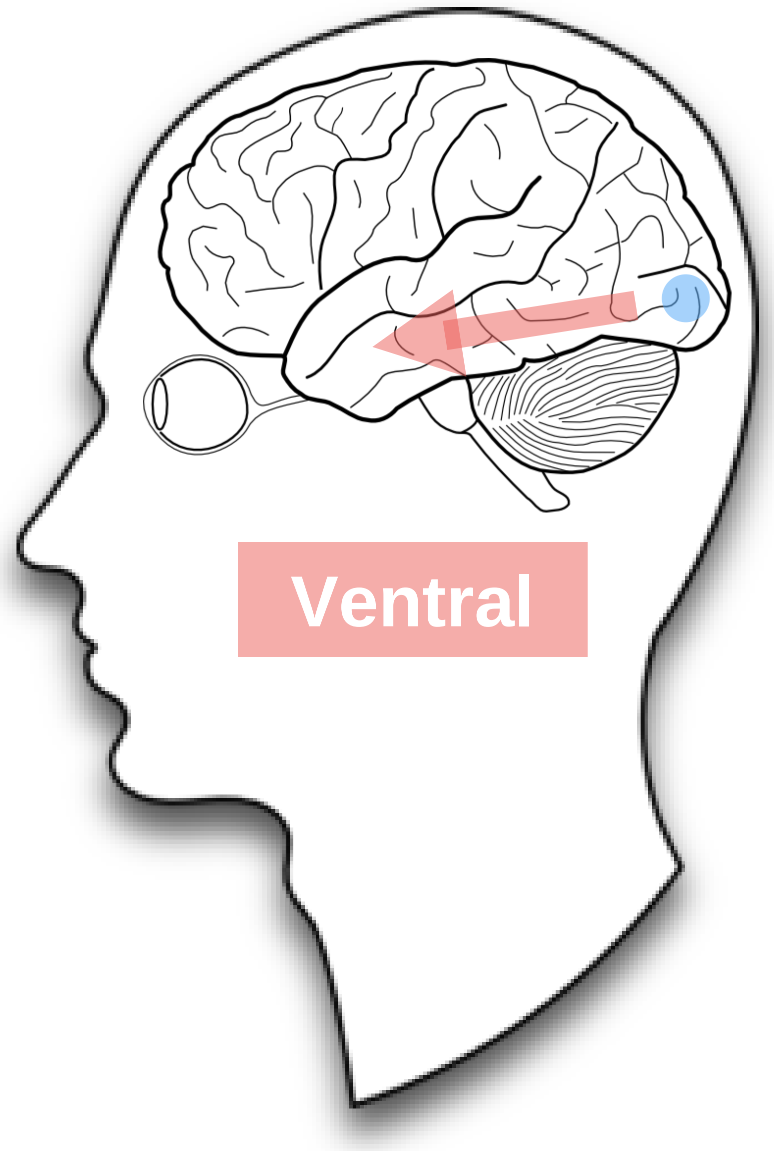
(very) Basic Overview

<u>Dorsal steam</u>

- V1-V2-MT-MST
- Parietal cortex
- Visually guided behaviour
- Eye movements
- Arm movements

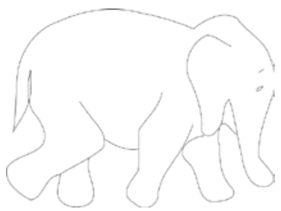


(very) Basic Overview



<u>Ventral Stream</u>

- V1-V2-V4-
- Inferior temporal cortex (IP)
- Form processing
- Object recognition
- Long term memory



(very) Basic Overview

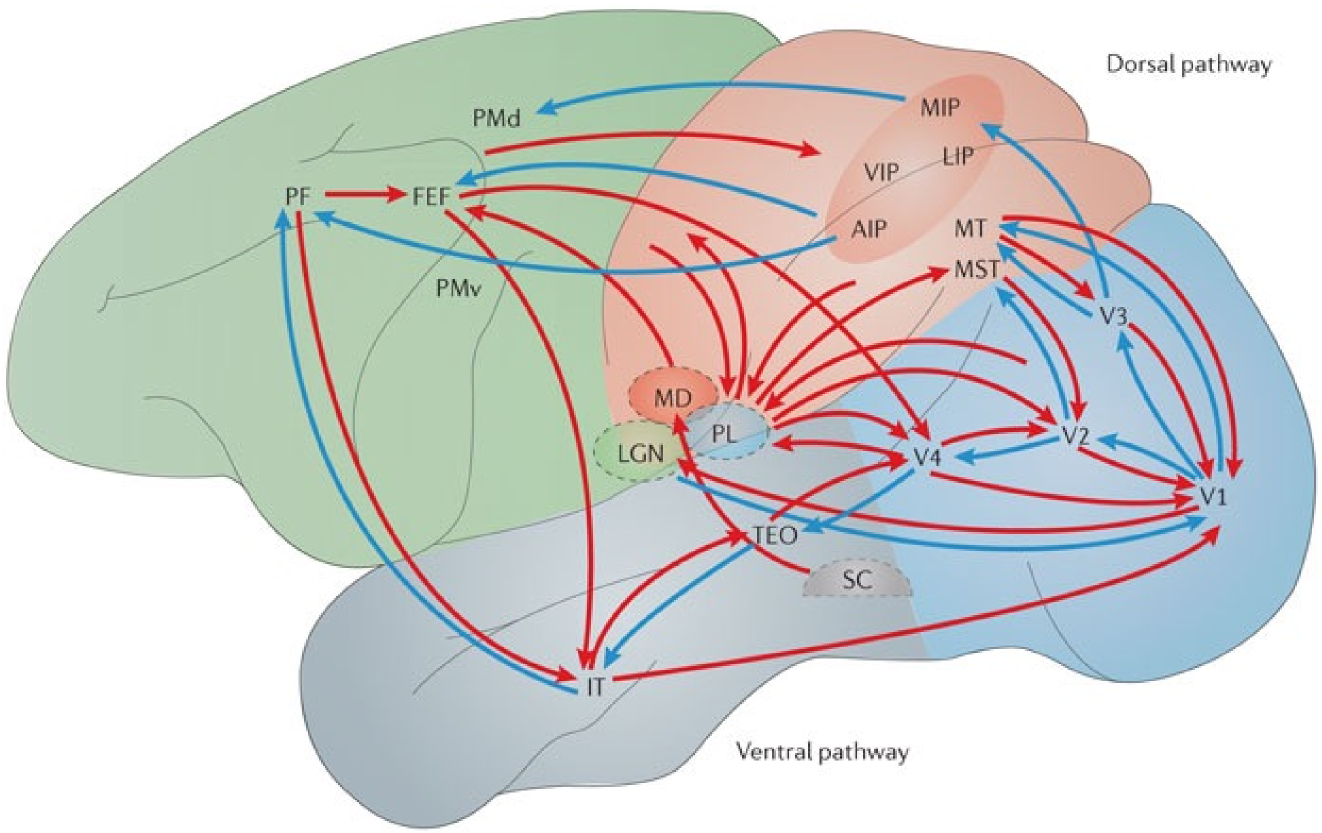
Downstream

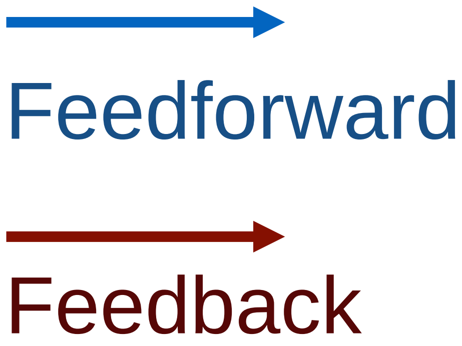
- Retinatopic
- Simple features

- Increasing complexity
- Spatial invariance
- More abstract qualities

Upstream

(very) Basic Overview

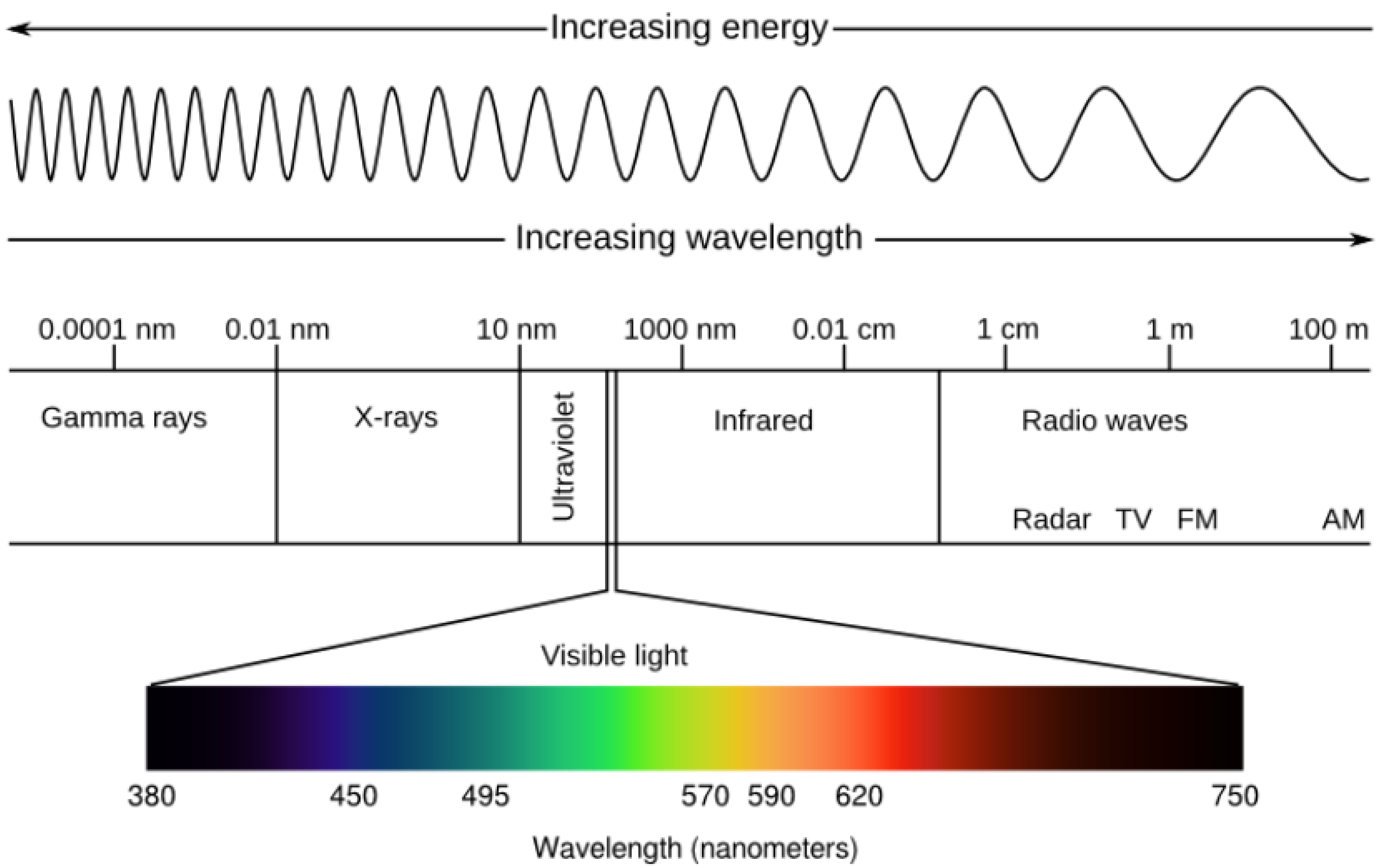




Gilbert & Wu Li (2013). Top-down influences on visual processing. Nature Reviews Neuroscien

Light and color

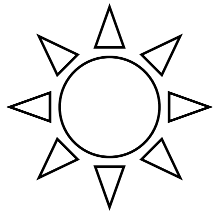
Electromagnetic spectrum



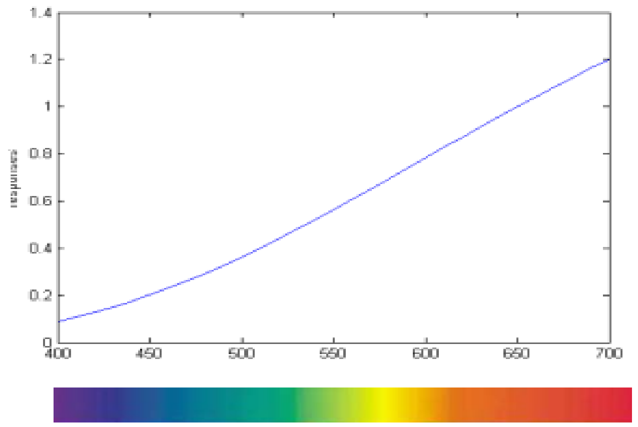
How is colour formed?

Colour is based on 3 properties

The incoming light

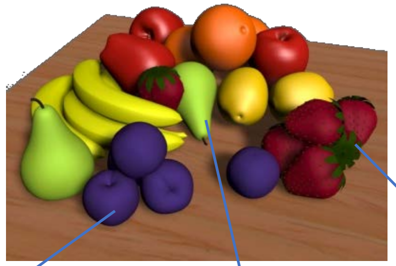


$I(\lambda)$

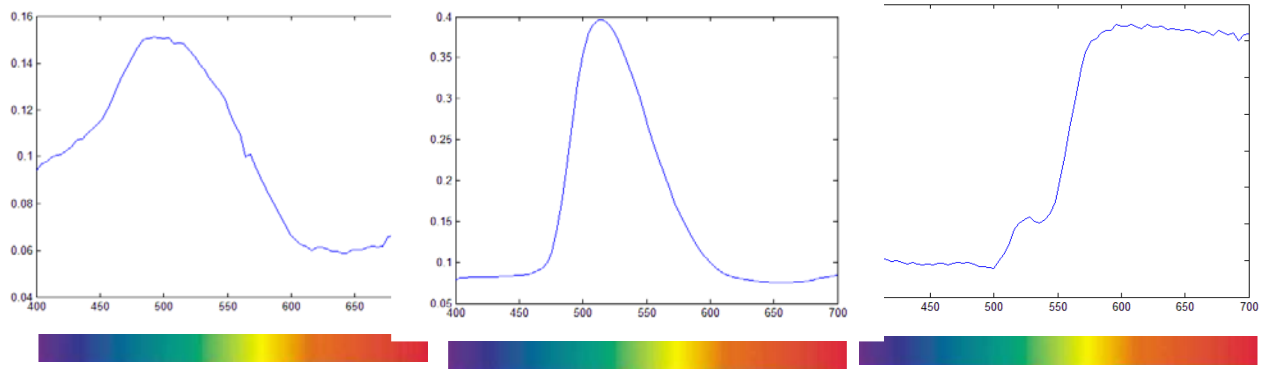


400 700

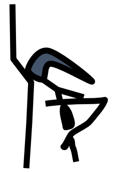
The reflectance of the object

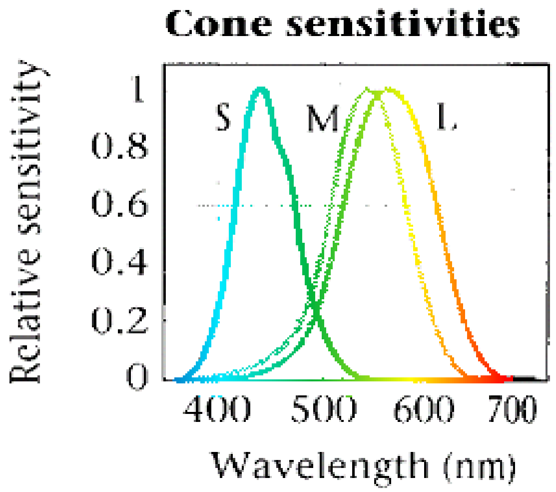


$R(\lambda)$



The cons in the retina





$L(\lambda)$, $M(\lambda)$, $S(\lambda)$

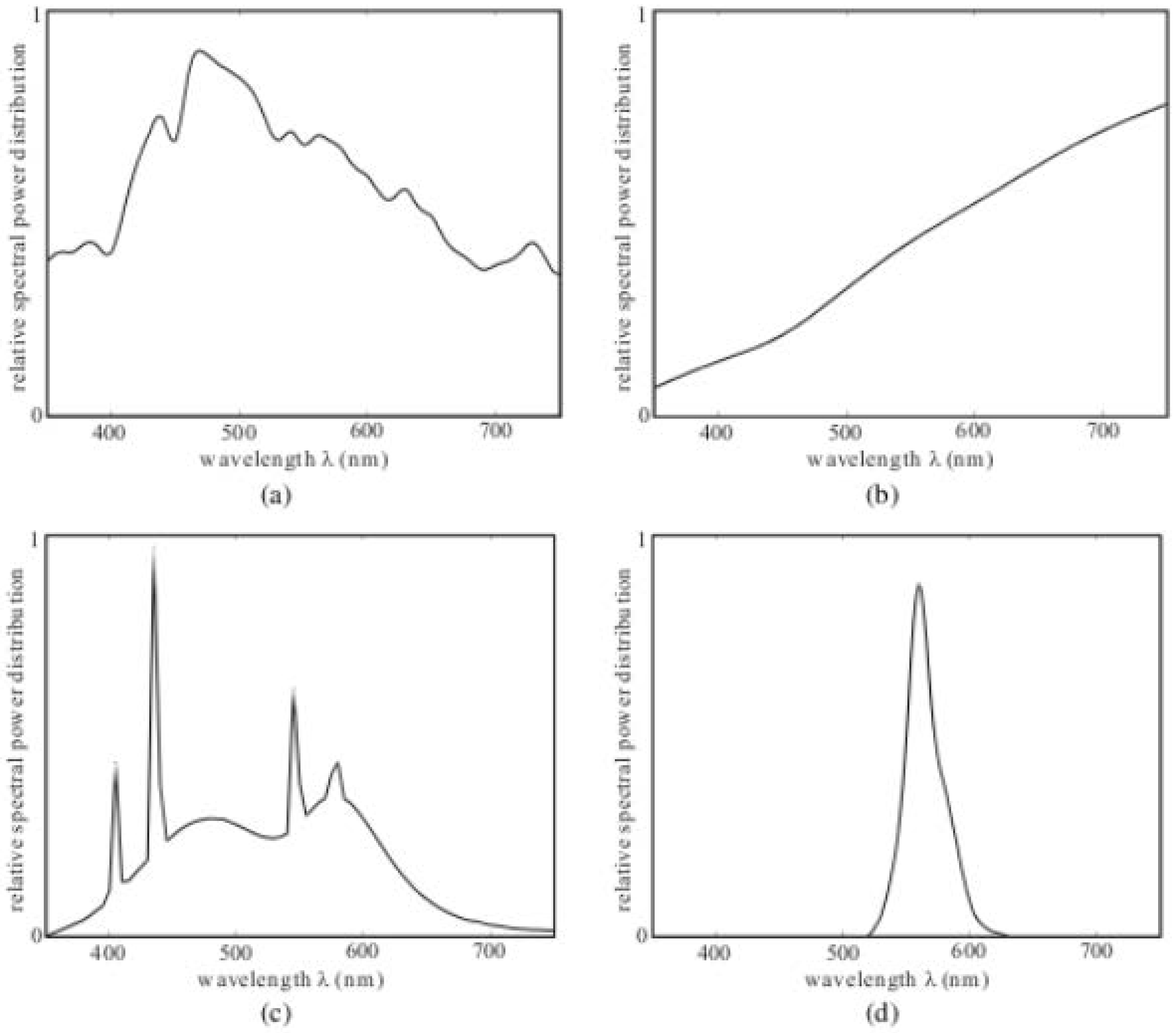
Some illuminants

Daylight

Tungsten

Fluorescent

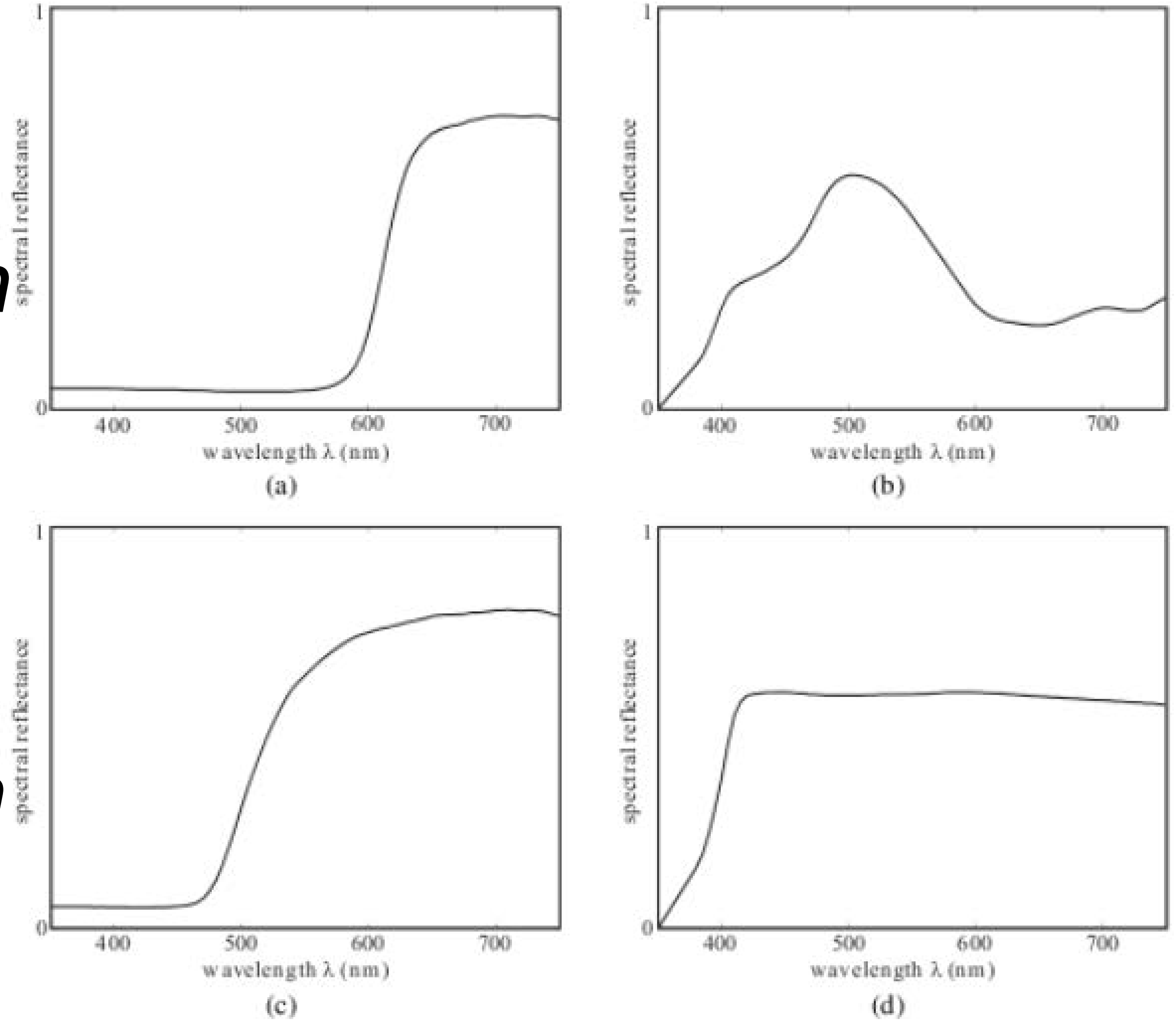
LED



Some reflectances

Red patch

Blue patch



Yellow patch

Gray patch

How the color is formed?

The colour signal (i.e. the light reaching at our eye) is the point-wise multiplication of illuminant and reflectance

$$E(\lambda) = I(\lambda) \times R(\lambda)$$

How the color is formed?

The colour signal (i.e. the light reaching at our eye) is the point-wise multiplication of illuminant and reflectance

$$E(\lambda) = I(\lambda) \times R(\lambda)$$

This value is point-wise multiplied by the cones, and integrated over the visual spectra

$$L = \int_{380}^{740} l(\lambda)E(\lambda)d\lambda$$

$$M = \int_{380}^{740} m(\lambda)E(\lambda)d\lambda$$

$$S = \int_{380}^{740} s(\lambda)E(\lambda)d\lambda.$$

Tristimulus values

How the color is formed?

The colour signal (i.e. the light reaching at our eye) is the point-wise multiplication of illuminant and reflectance

$$E(\lambda) = I(\lambda) \times R(\lambda)$$

This value is point-wise multiplied by the cones, and integrated over the visual spectra

$$L = \int_{380}^{740} l(\lambda)E(\lambda)d\lambda$$

$$M = \int_{380}^{740} m(\lambda)E(\lambda)d\lambda$$

$$S = \int_{380}^{740} s(\lambda)E(\lambda)d\lambda.$$

Practically, this integration is done by sampling over the different wavelengths

$$L = \sum_{i=380}^{740} l(\lambda_i)E(\lambda_i)$$

$$M = \sum_{i=380}^{740} m(\lambda_i)E(\lambda_i)$$

$$S = \sum_{i=380}^{740} s(\lambda_i)E(\lambda_i),$$

Corollarys from the colour formation equation

1) Two different reflectances under two different illuminants can have exactly the same tristimulus values:

This is called the metamerism problem. It is supposed to occur around a 3% in natural images.

Corollarys from the colour formation equation

1) Two different reflectances under a particular illuminant can have exactly the same tristimulus values:

This is called the metamerism problem. It is supposed to occur around a 3% in natural images.

2) From the colour formation equation it can be proved (see notes) that we can generate any color by mixing three given colors –called primaries-, just by adjusting the amount of each one:

This is called the trichromacy property. It is a fundamental property of human color vision.

The first color spaces

Deriving the CIE R,G,B colour space

Guild and Wright in two independent studies.

Very few observers (10 and 7). All male. All Brits. Is it really then a good standard?

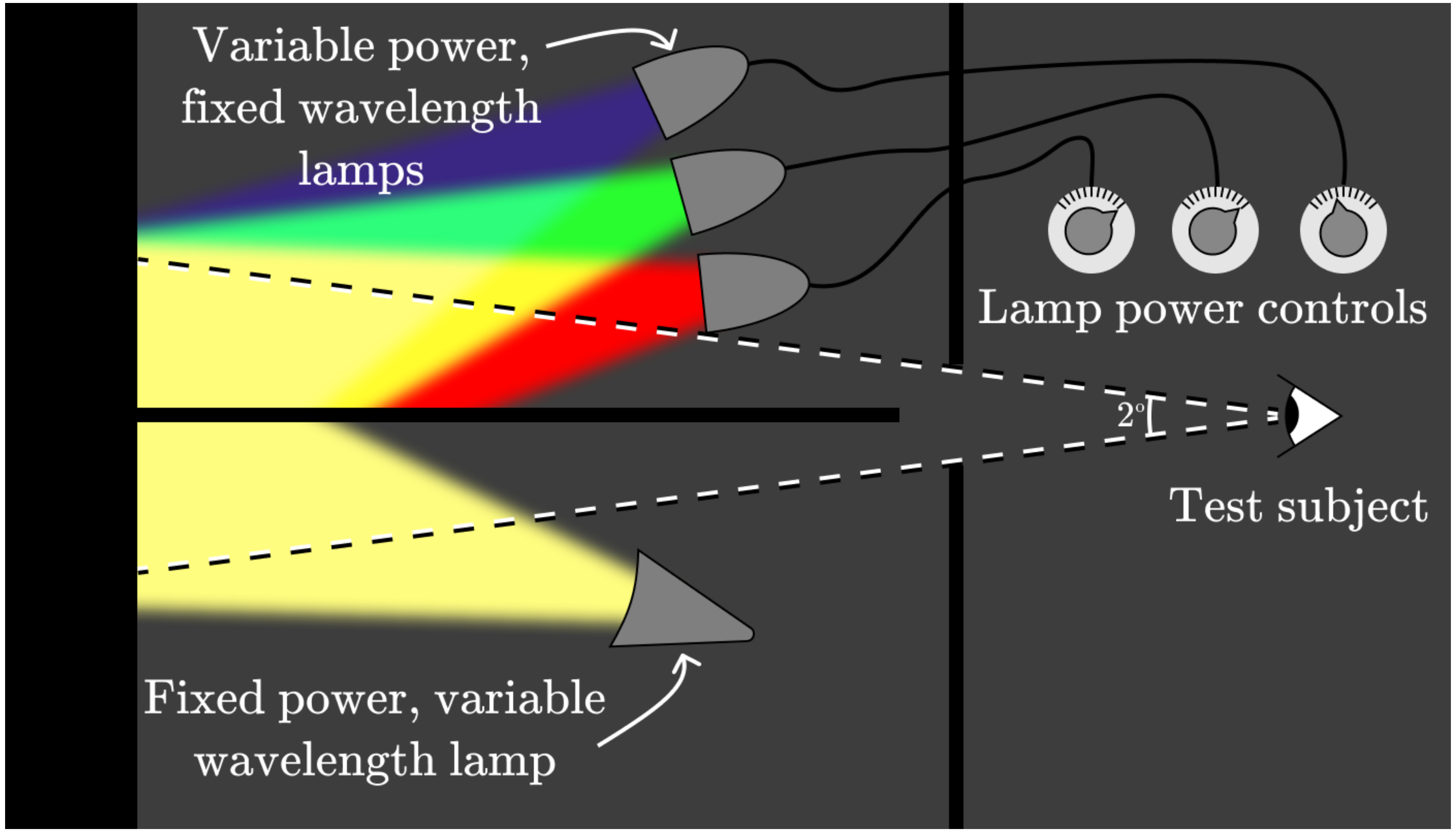


Image from: http://jamie-wong.com/post/color/

Deriving the CIE R,G,B colour space

Guild and Wright in two independent studies.

Very few observers (10 and 7). All male. All Brits. Is it really then a good standard?

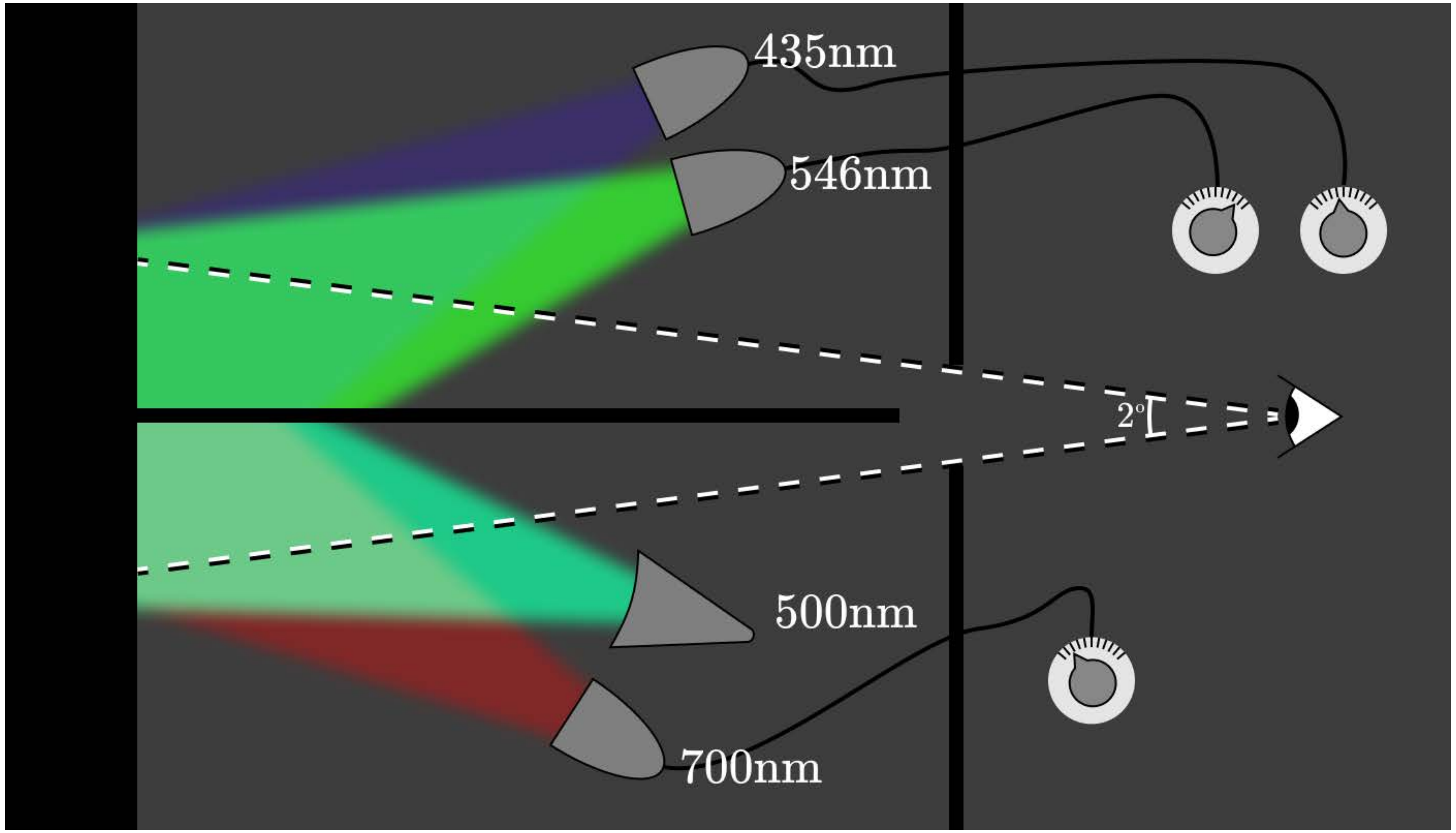


Image from: http://jamie-wong.com/post/color/

Deriving the CIE R,G,B colour space

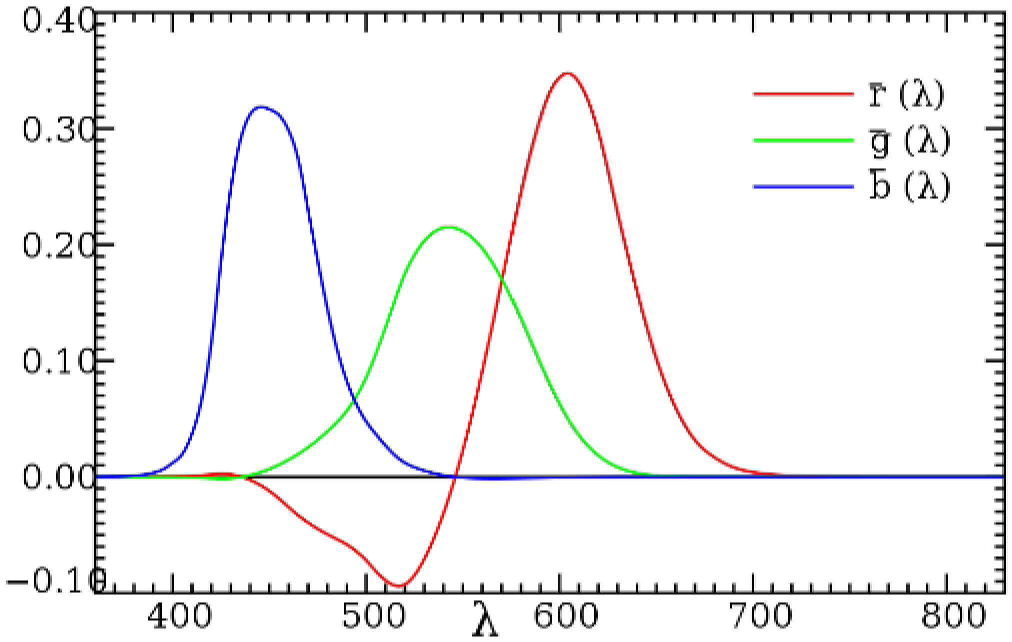


FIGURE 1.5: Color matching functions. Figure from [13].

It can be seen that the CIE R,G,B color marching functions and the L,M,S cone functions are just a linear correction apart (see notes).

Deriving the CIE R,G,B colour space

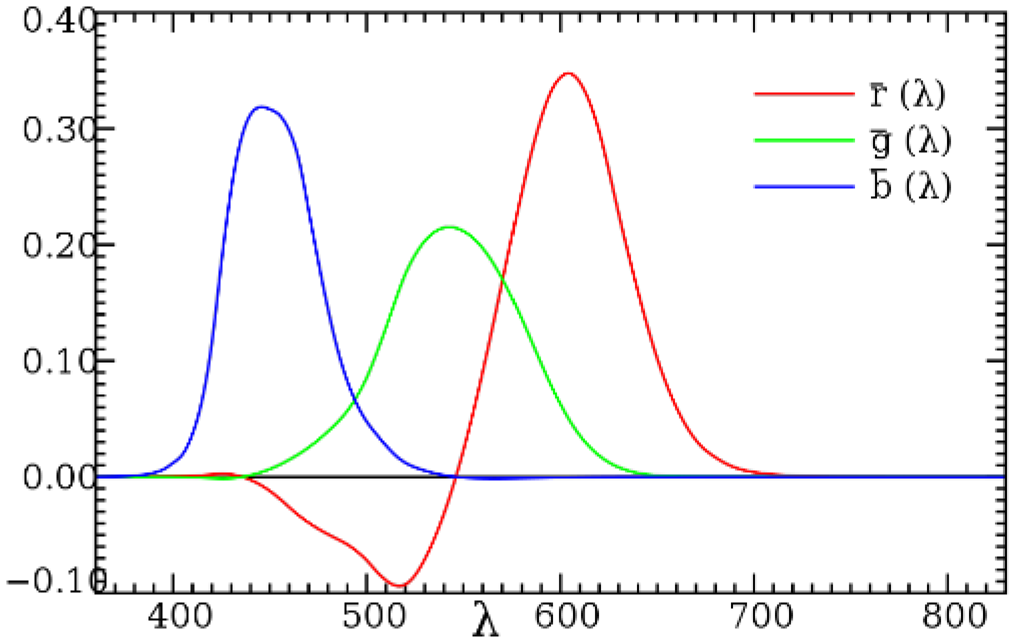


FIGURE 1.5: Color matching functions. Figure from [13].

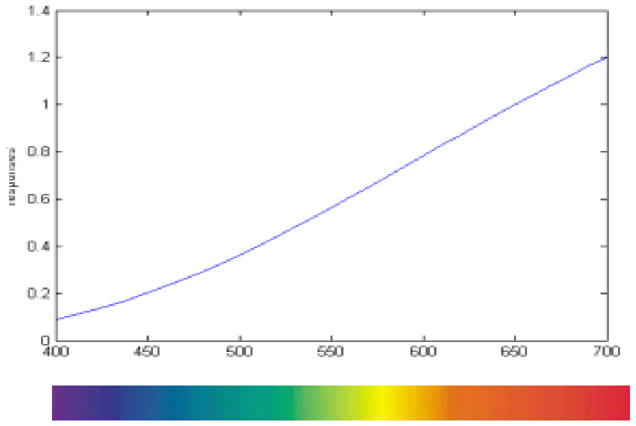
It can be seen that the CIE R,G,B color matching functions and the L,M,S cone functions are just a linear correction apart (see notes).

In other words, they generate the same 3-dimensional space, the space of the colours we are able to see.

Deriving the CIE R,G,B colour space

The incoming light

$I(\lambda)$

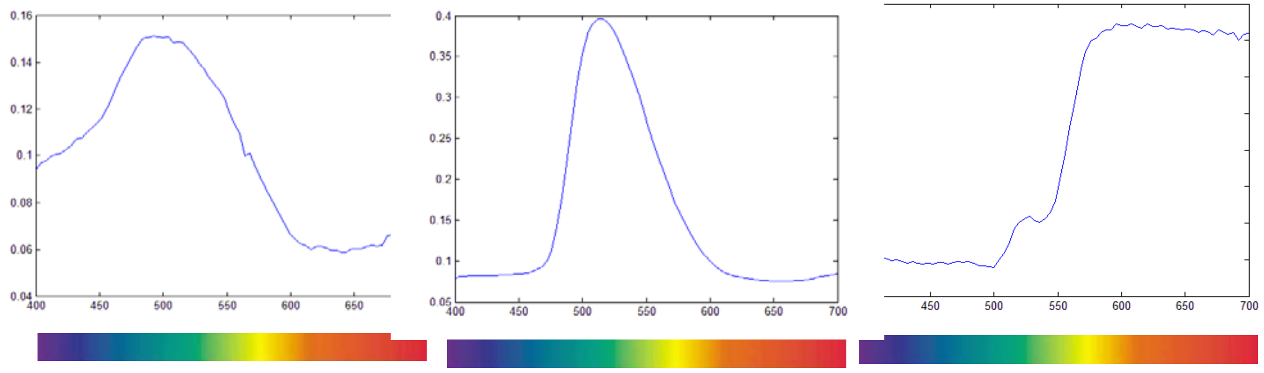


400

700

The reflectance of the object

$R(\lambda)$



The CIE RGB

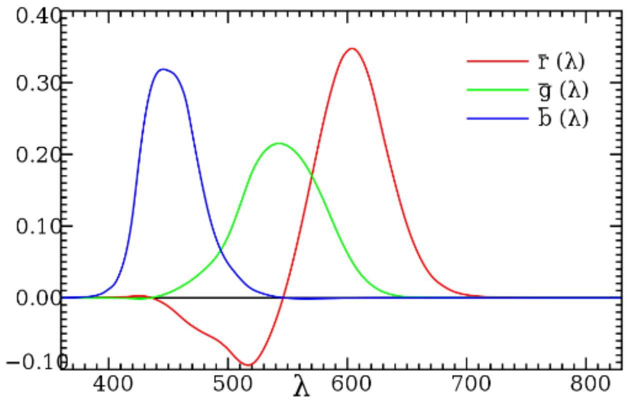
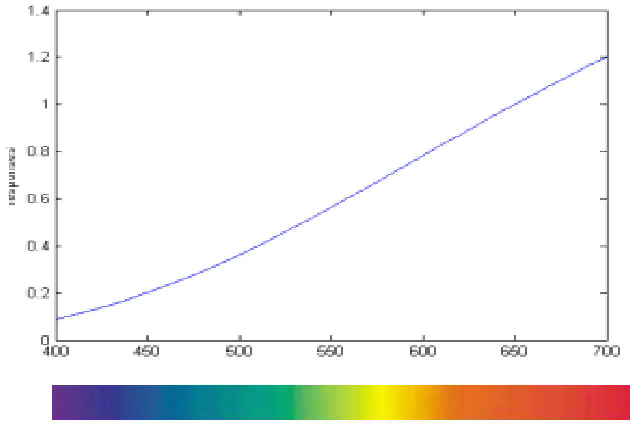


FIGURE 1.5: Color matching functions. Figure from [13].

Deriving the CIE R,G,B colour space

The incoming light

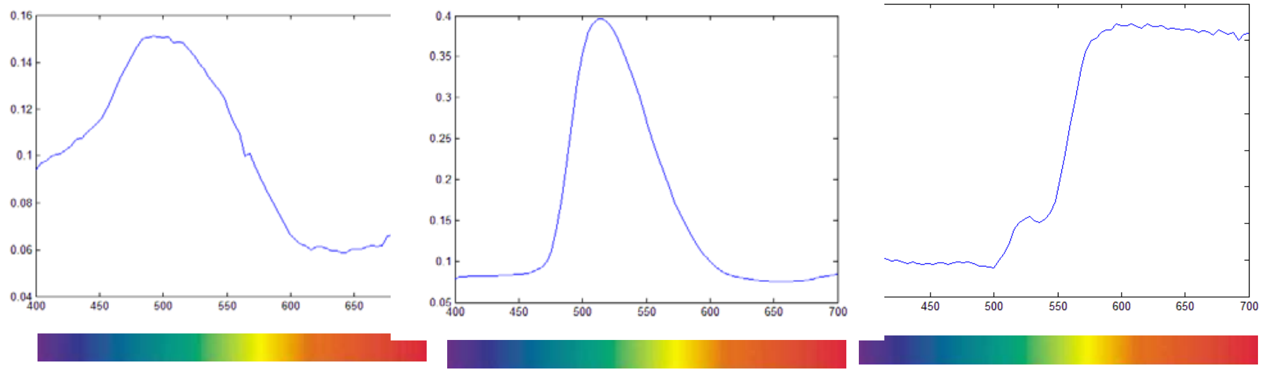
$I(\lambda)$



400 700

The reflectance of the object

$R(\lambda)$



The CIE RGB

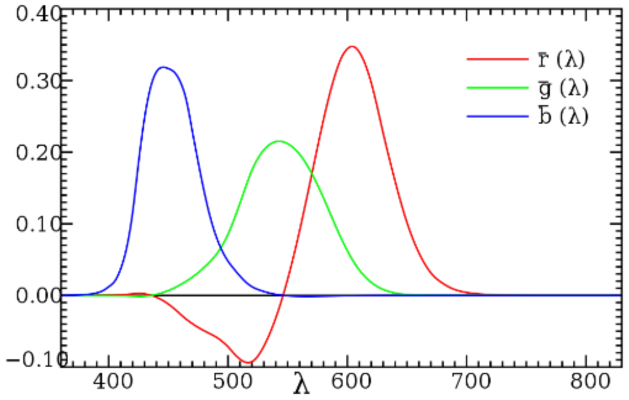


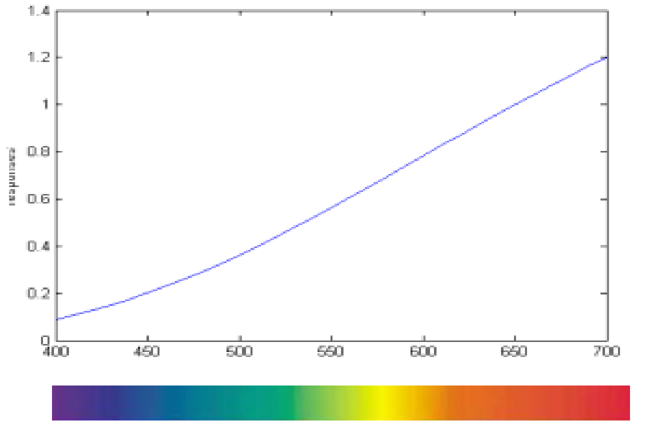
FIGURE 1.5: Color matching functions. Figure from [13].

$$E(\lambda) = I(\lambda) \times R(\lambda)$$

Deriving the CIE R,G,B colour space

The incoming light

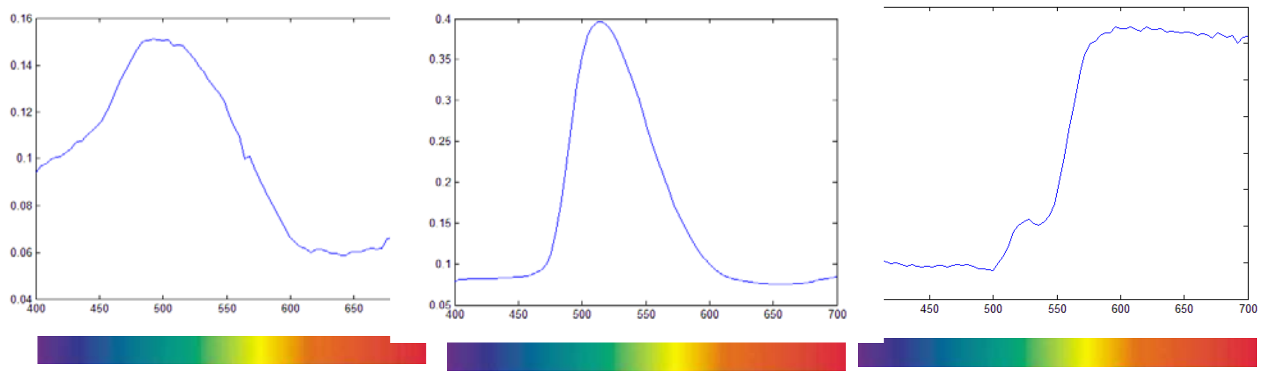
I(λ)



400 700

The reflectance of the object

R(λ)



The CIE RGB

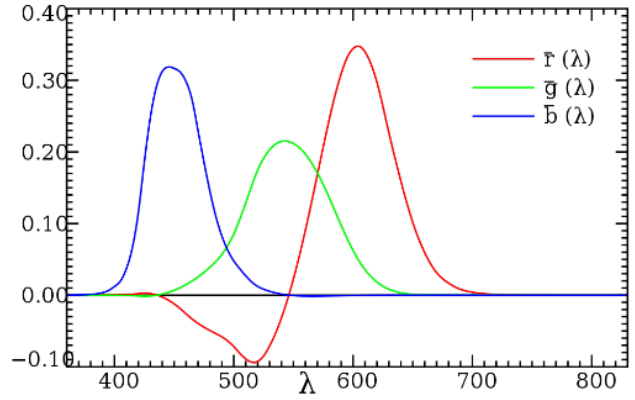


FIGURE 1.5: Color matching functions. Figure from [13].

$$E(\lambda) = I(\lambda) \times R(\lambda)$$

$$R = \int_{380}^{740} \bar{r}(\lambda)E(\lambda)d\lambda \qquad R = \sum_{i=380}^{740} \bar{r}(\lambda_i)E(\lambda_i)$$

$$G = \int_{380}^{740} \bar{g}(\lambda)E(\lambda)d\lambda \qquad G = \sum_{i=380}^{740} \bar{g}(\lambda_i)E(\lambda_i)$$

$$B = \int_{380}^{740} \bar{b}(\lambda)E(\lambda)d\lambda, \qquad B = \sum_{i=380}^{740} \bar{b}(\lambda_i)E(\lambda_i).$$

From CIE R,G,B colour space to CIE XYX colour space

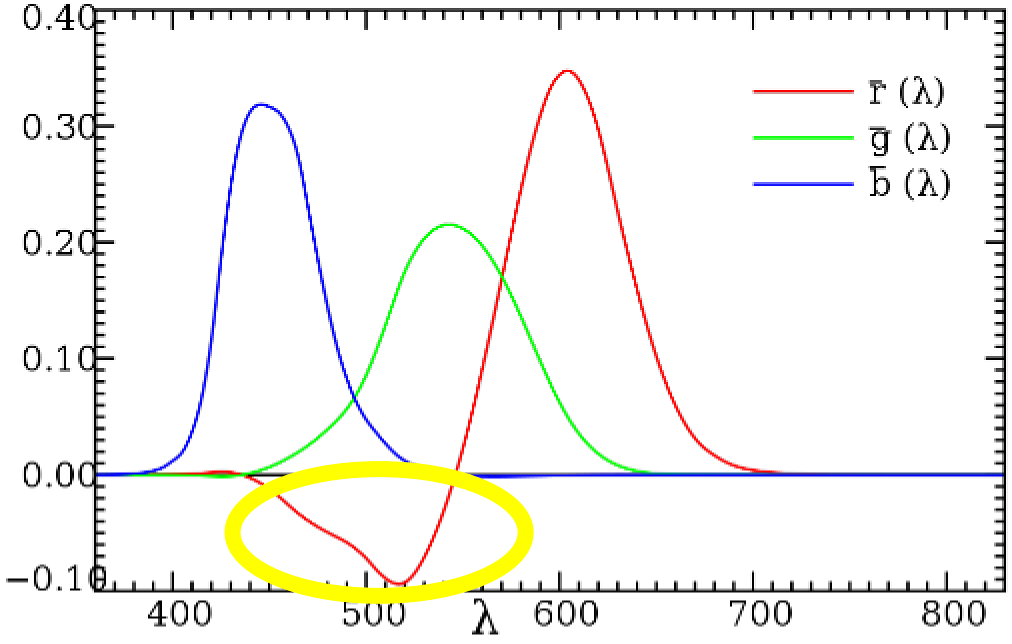
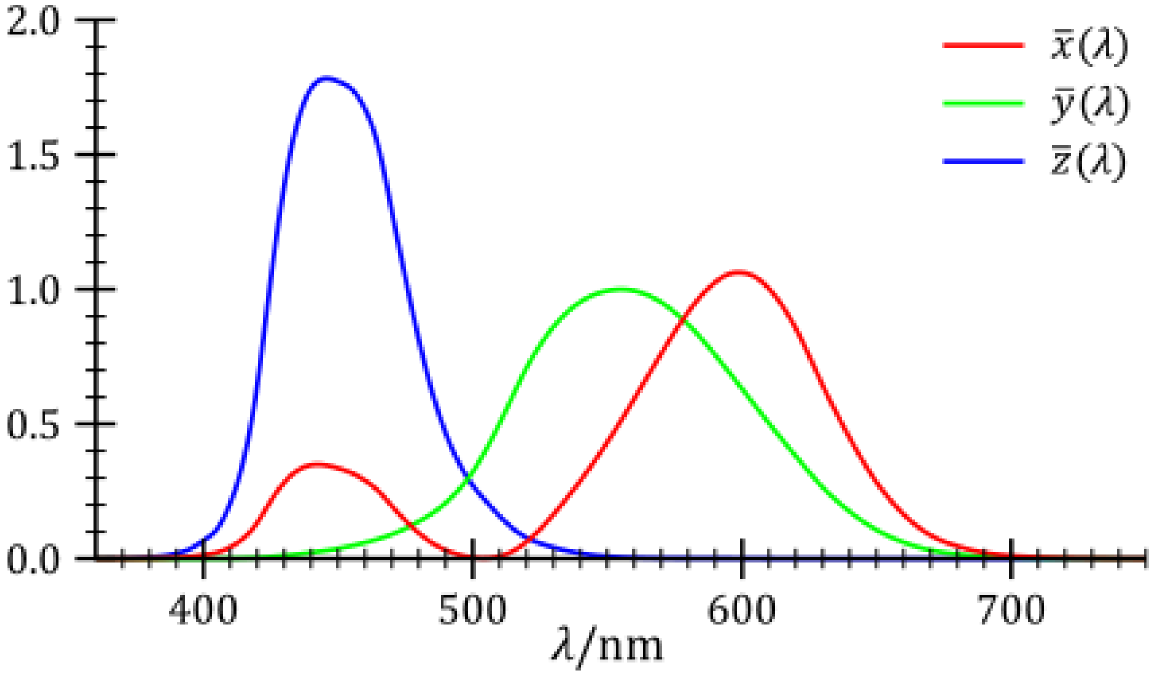


FIGURE 1.5: Color matching functions. Figure from [13].



Also just a linear correction apart (and forcing Y to be the luminosity function).

From CIE X,Y,Z to xyY

Color has two main parts: chromaticity (2-dimensional), and intensity (Black-White axis).

How can we create a color space that represent this?

xyY color space

$\left\{ \begin{array}{l} \text{x, y chromatic coordinates} \\ \\ \text{Y intensity} \end{array} \right.$

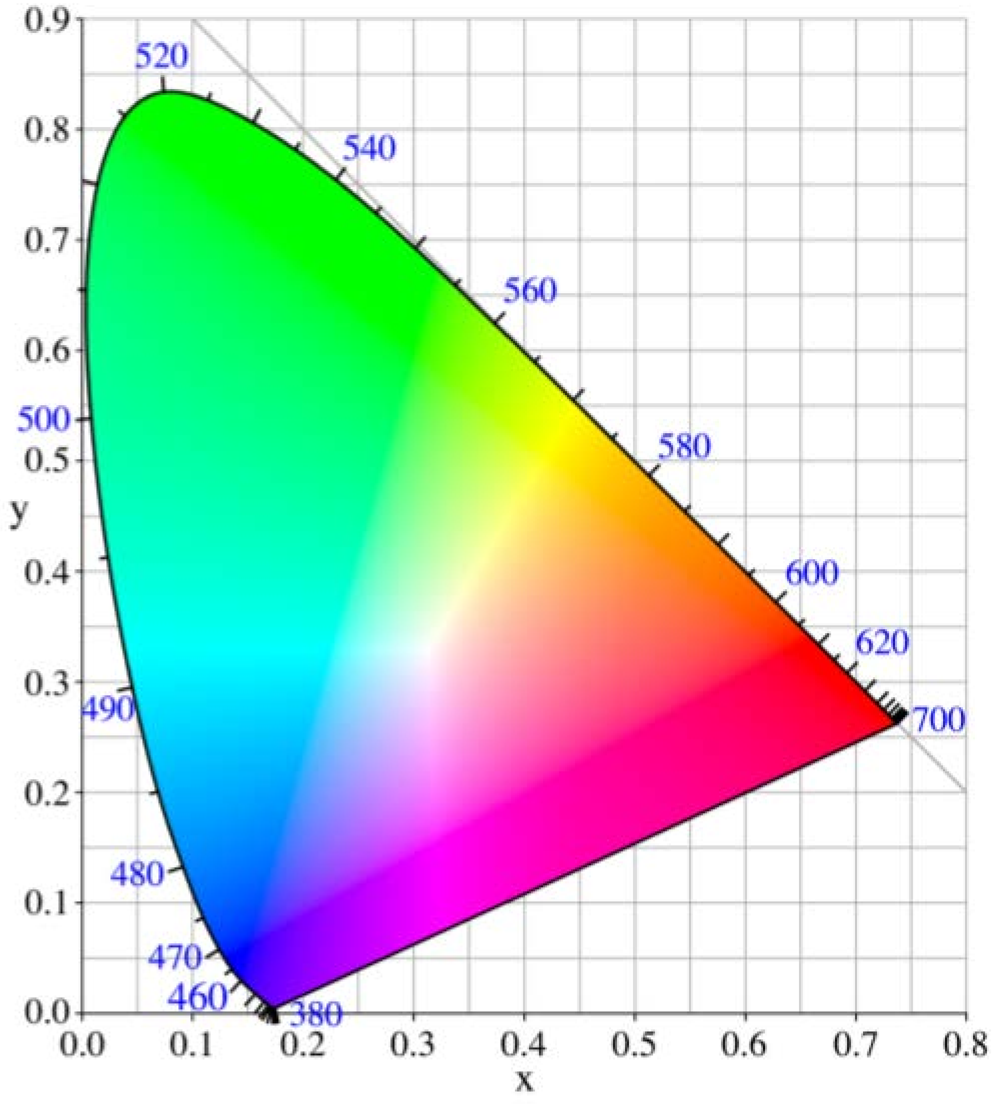
$$x = \frac{X}{X+Y+Z}$$

$$y = \frac{Y}{X+Y+Z}$$

The x,y chromaticity diagram

Represents all the colors humans are able to see.

Pure colors are in the Edge of the horseshoe shape

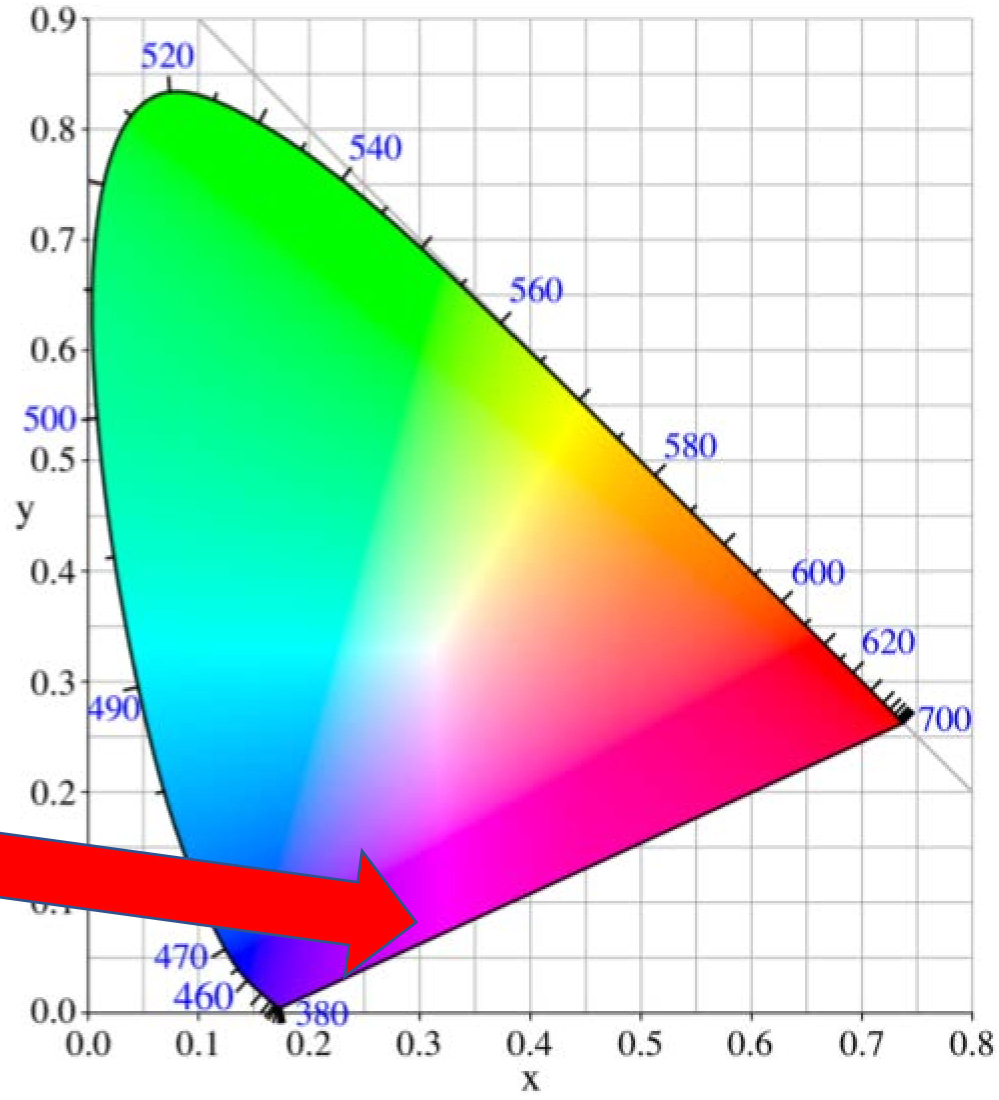


The x,y chromaticity diagram

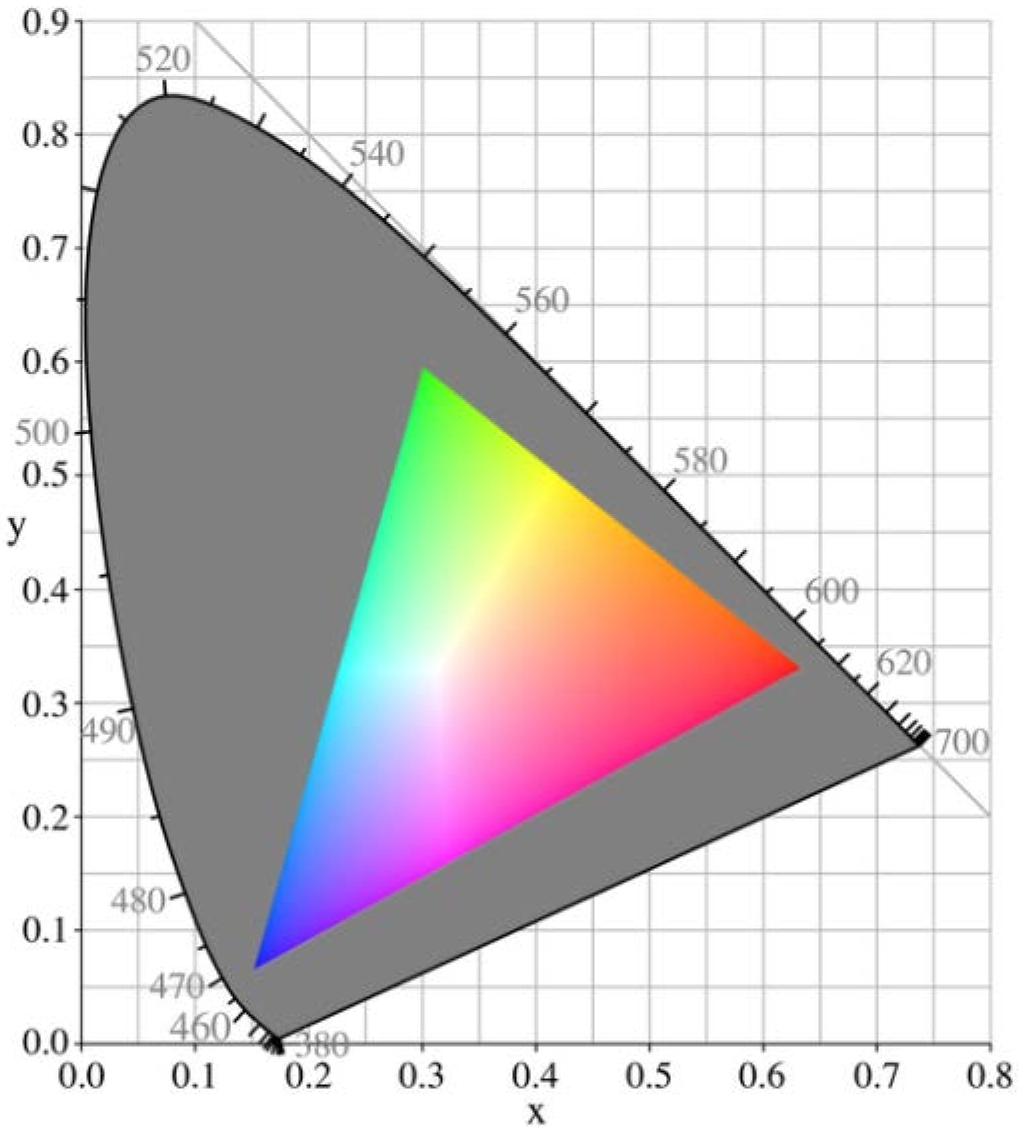
Represents all the colors humans are able to see.

Pure colors are in the Edge of the horseshoe shape

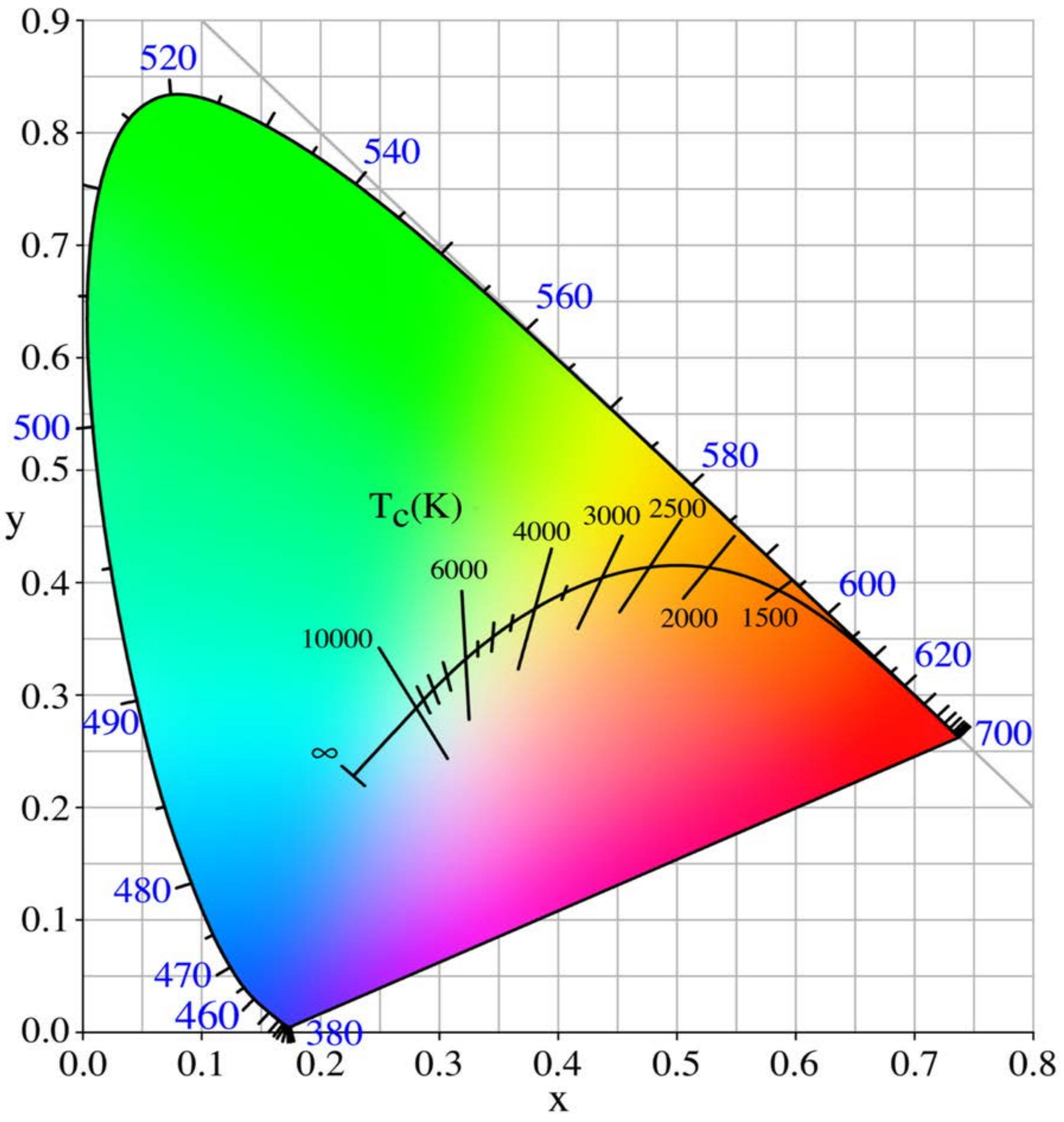
Line of purples



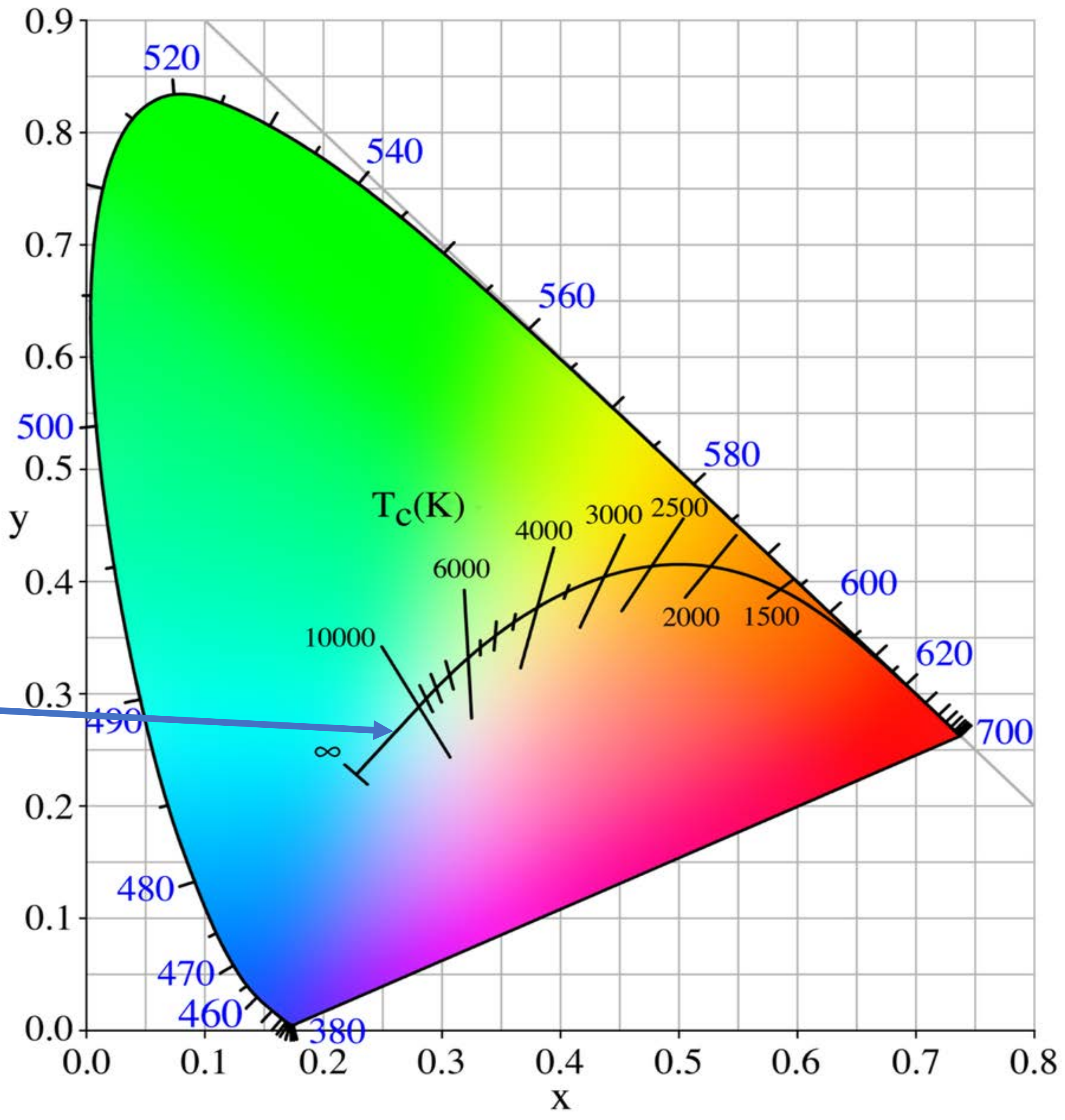
A standard gamut display in the x,y diagram



Daylight illuminants in the x,y diagram



Daylight illuminants in the x,y diagram



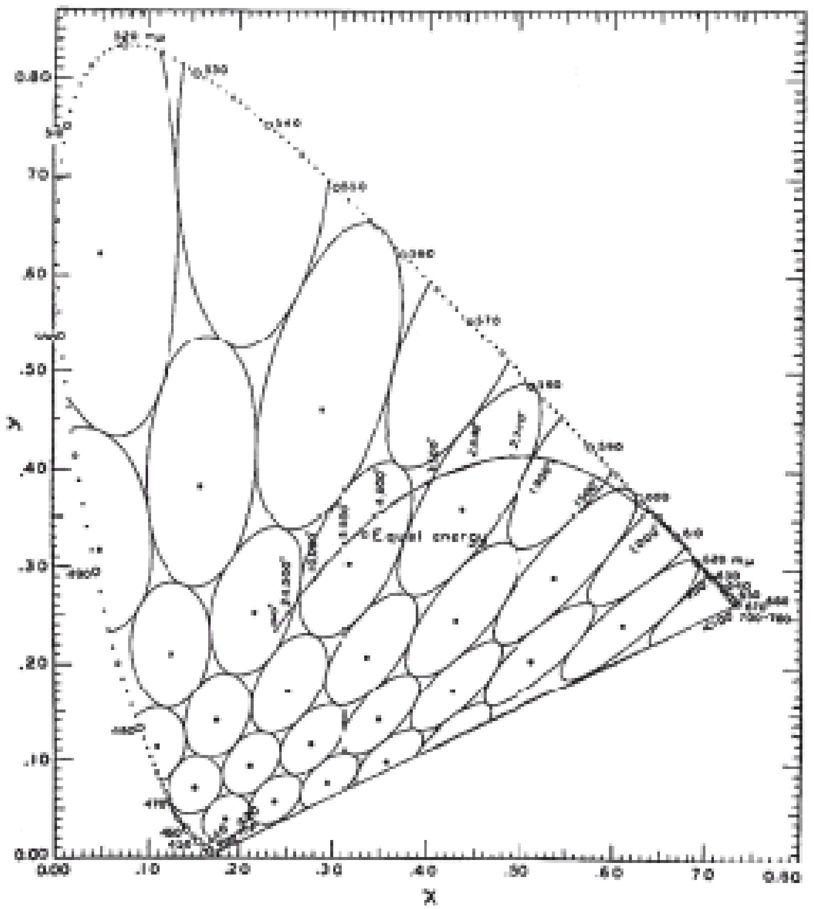
Planckian locus

A problem with CIE XYZ

CIE XYZ is not a perceptual color space, in the sense that distances in this space does not correspond to perceptual distances.

The same XYZ distance will be perceptually small in the green region, but perceptually large in the blue region

MacAdam ellipses

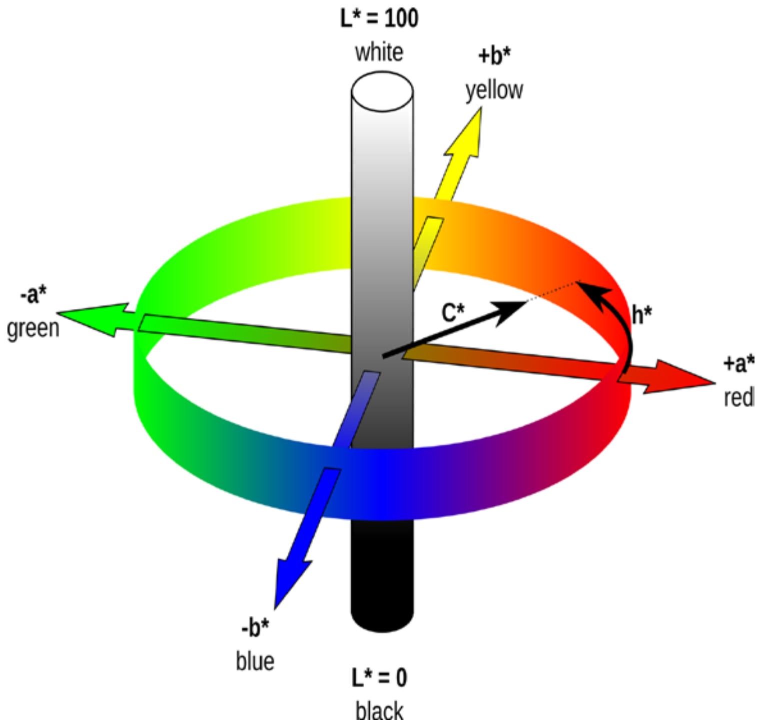


Solution: The CIE Lab colour space

CIE Lab was defined with the goal of being perceptually uniform.

It is based in the idea of opponency:

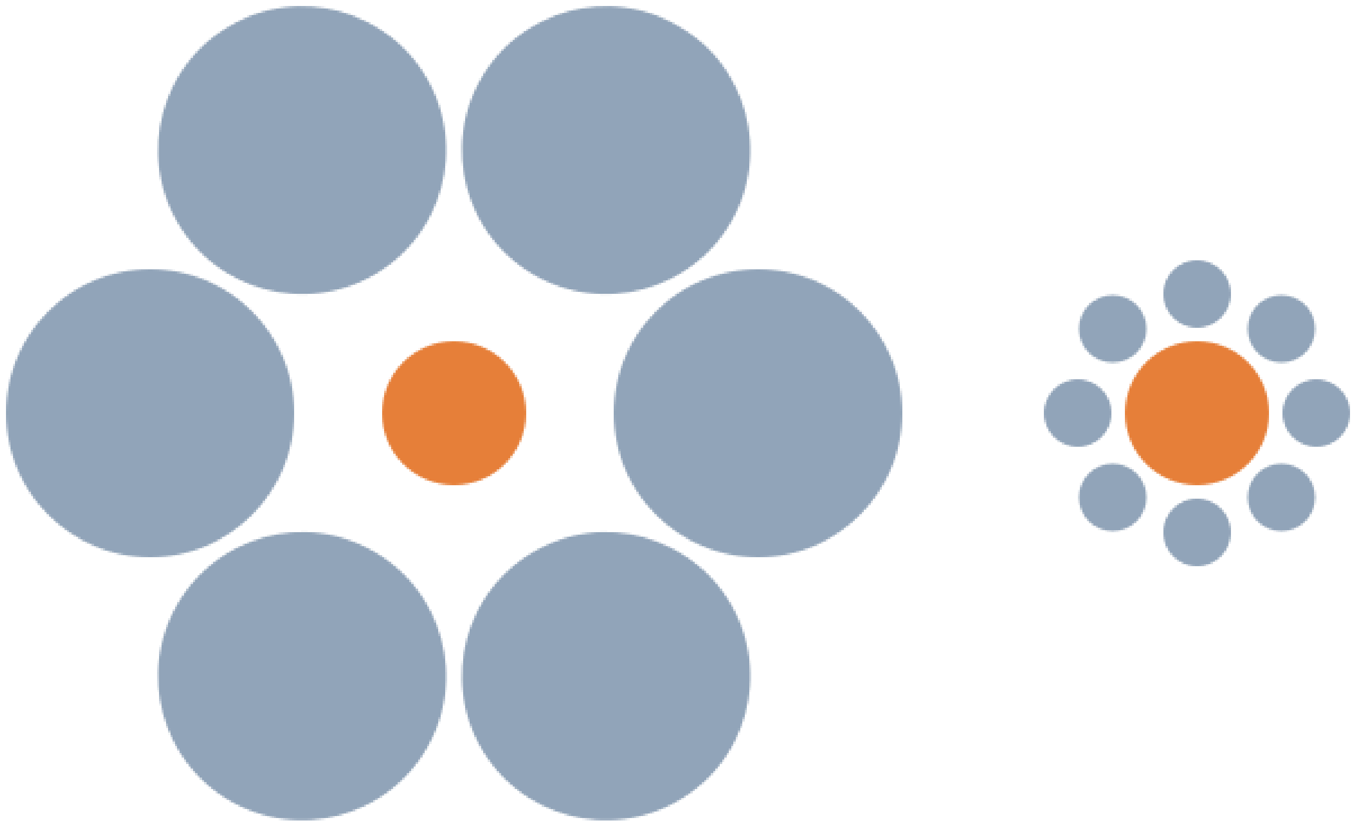
- 1 channel is based on the Y channel of CIE XYZ.
- The other two channels are substractions between the XYZ channels.



Contextual Effects

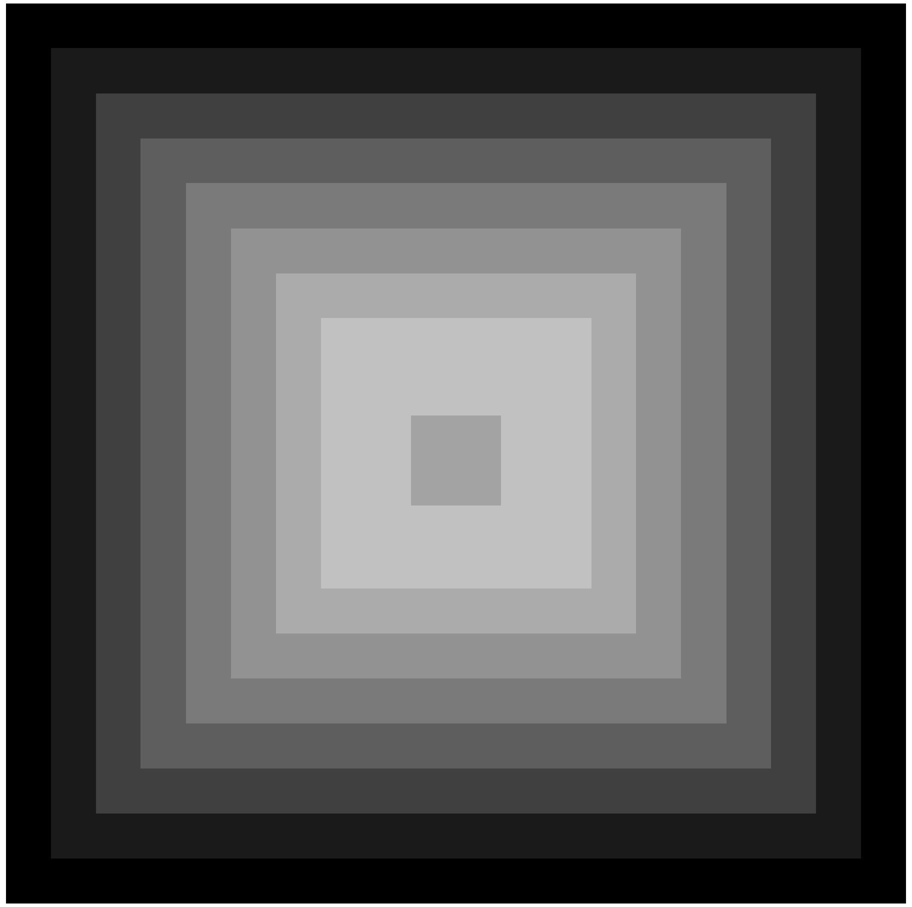
(Human vision is not just a point-wise thing)

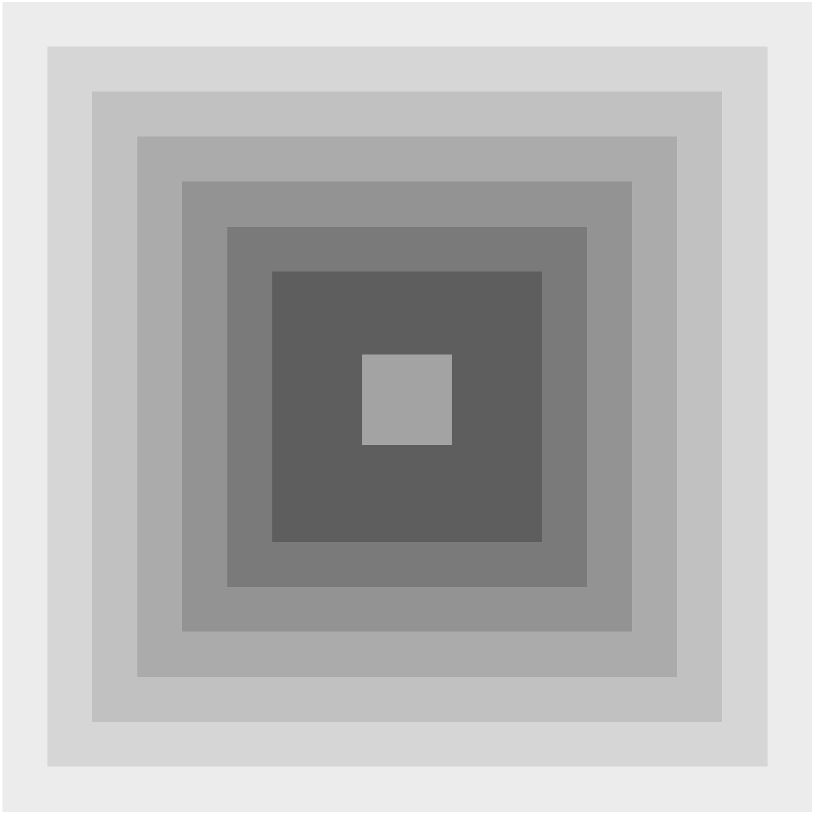
Ebbinghaus illusion or Titchener circles



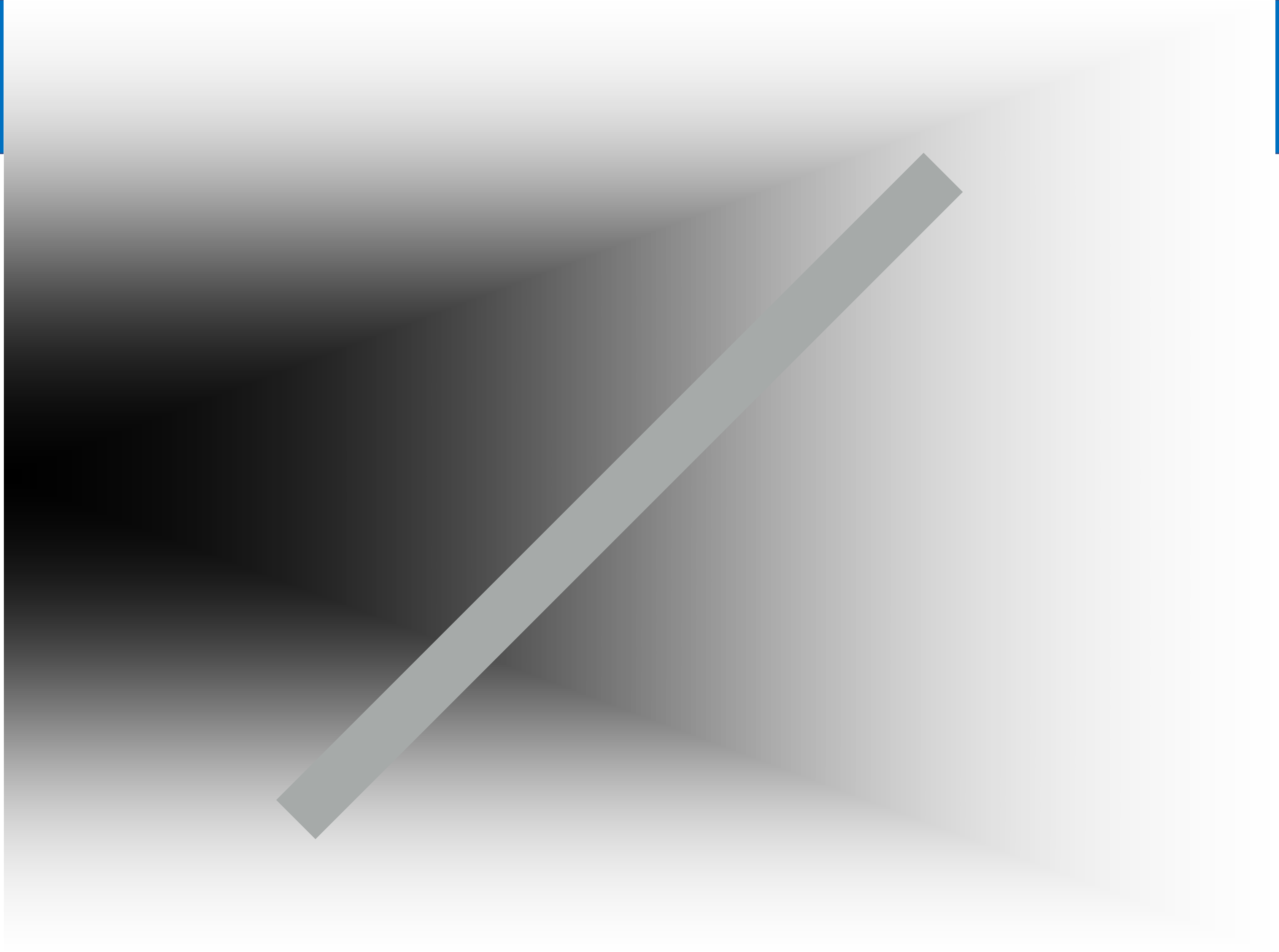
Ebbinghaus illusion or Titchener circles

Surround



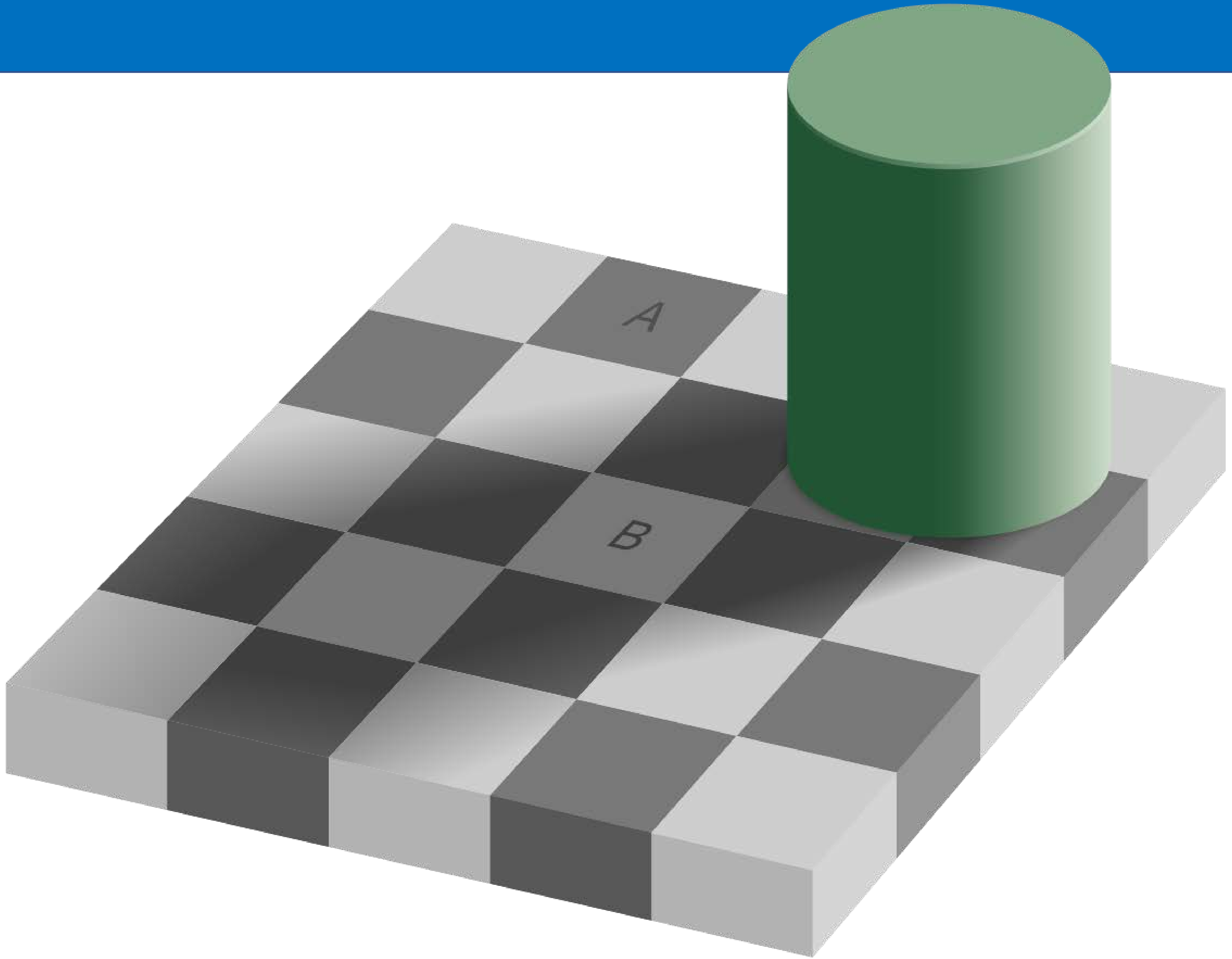


Surround



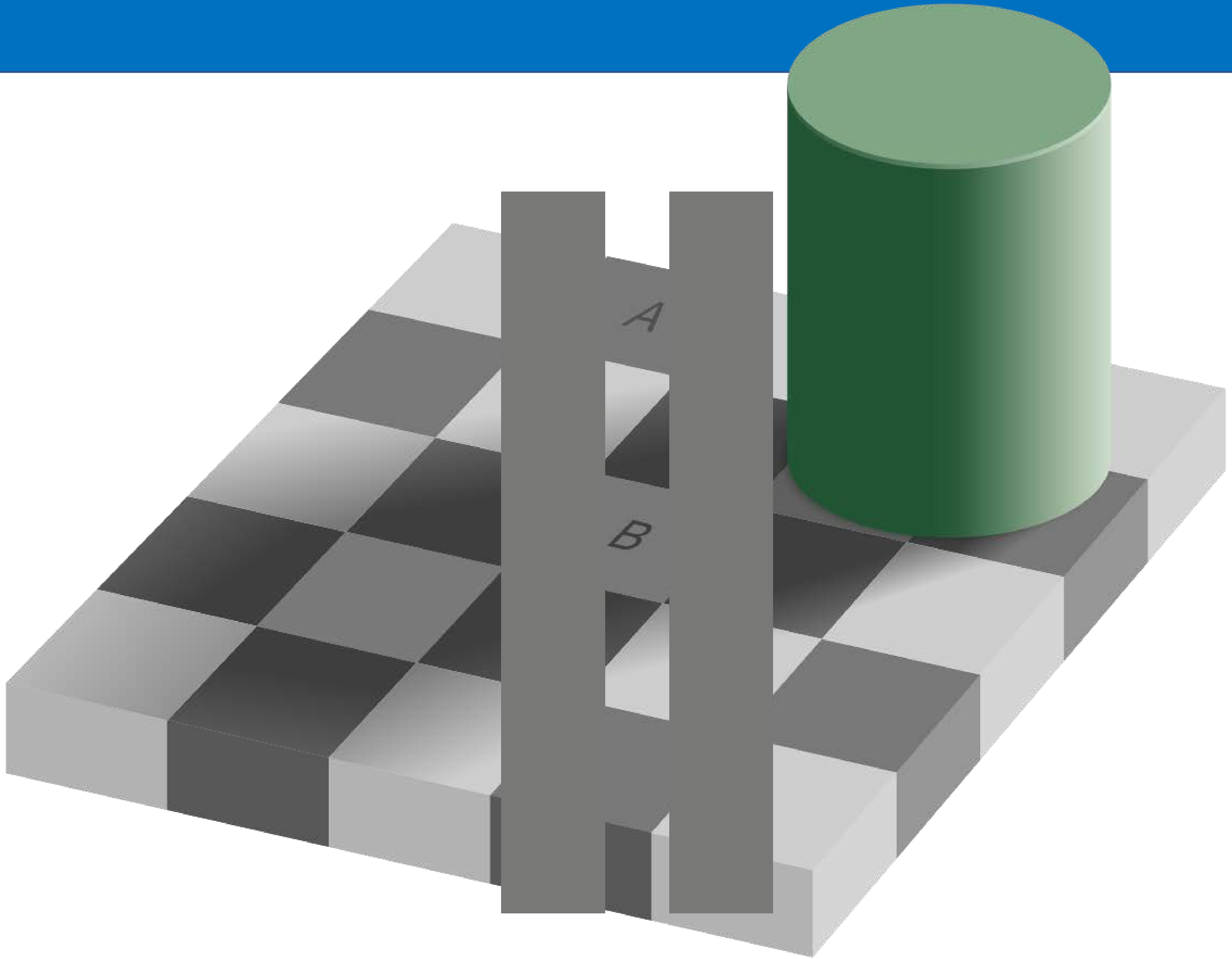
Surround

Surround



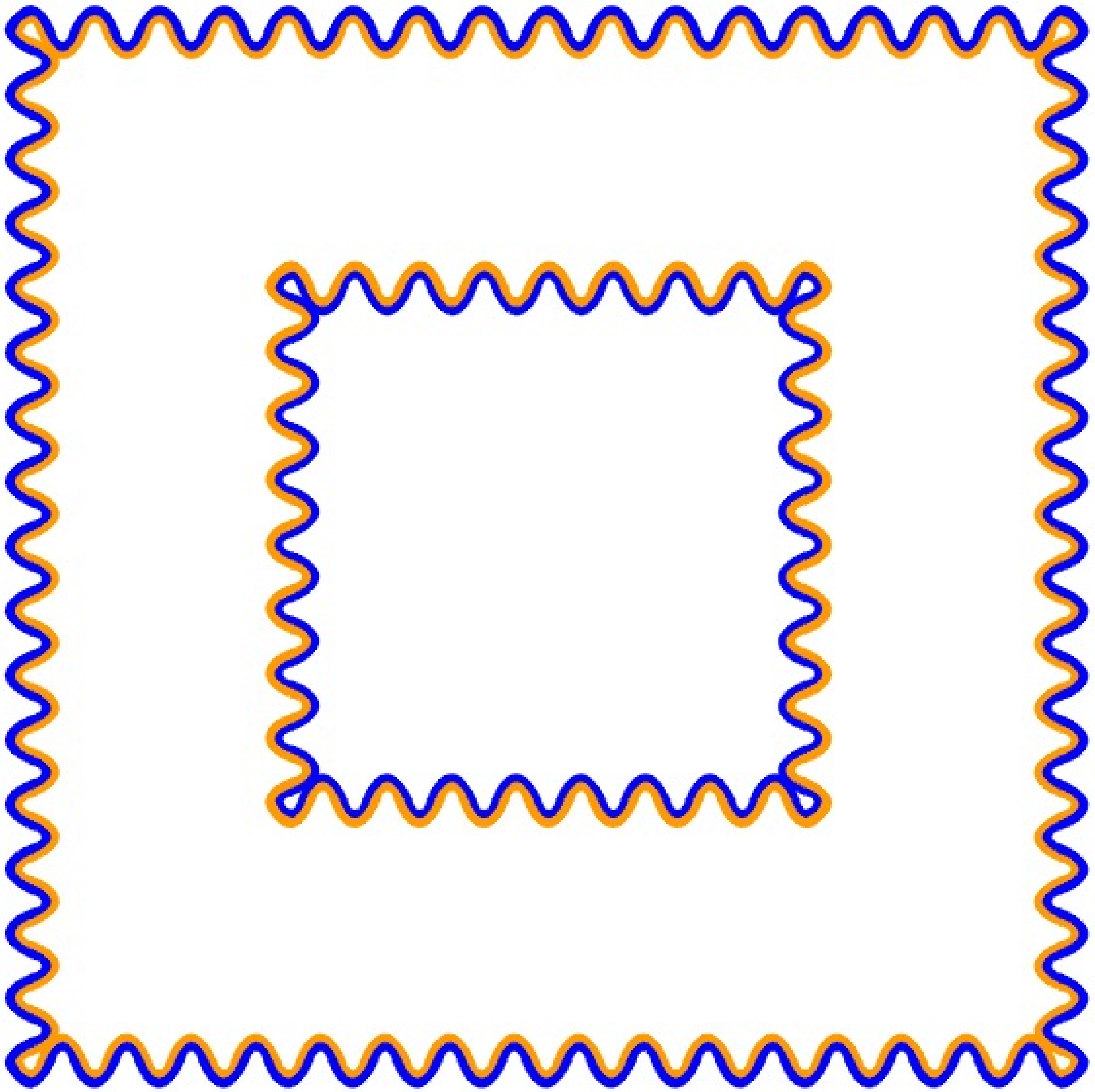
Adelson illusion

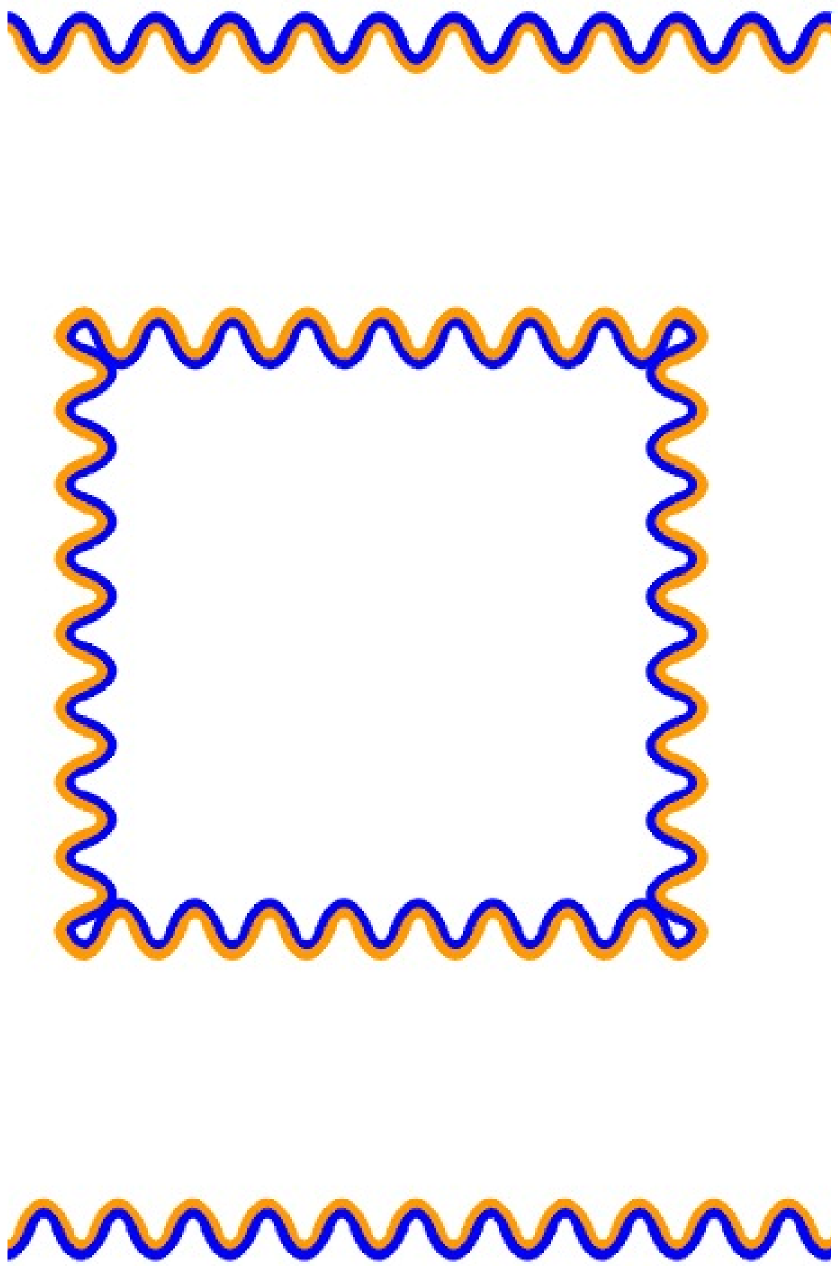
Surround

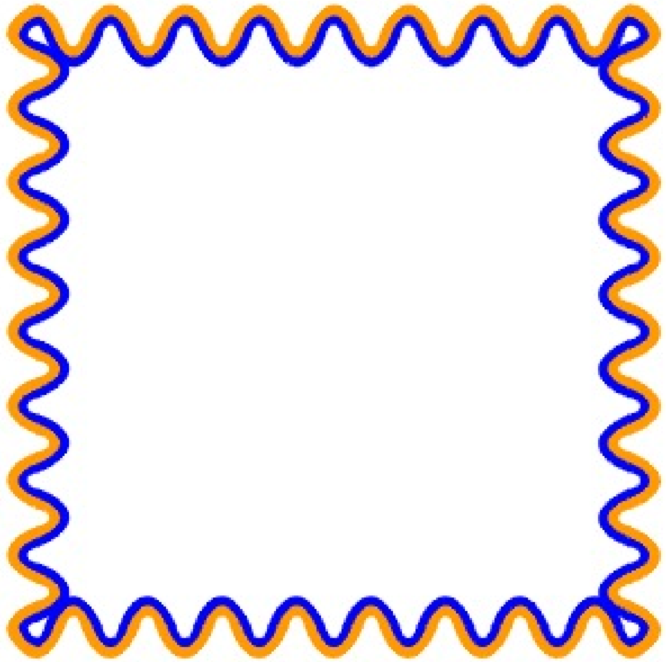


Adelson illusion

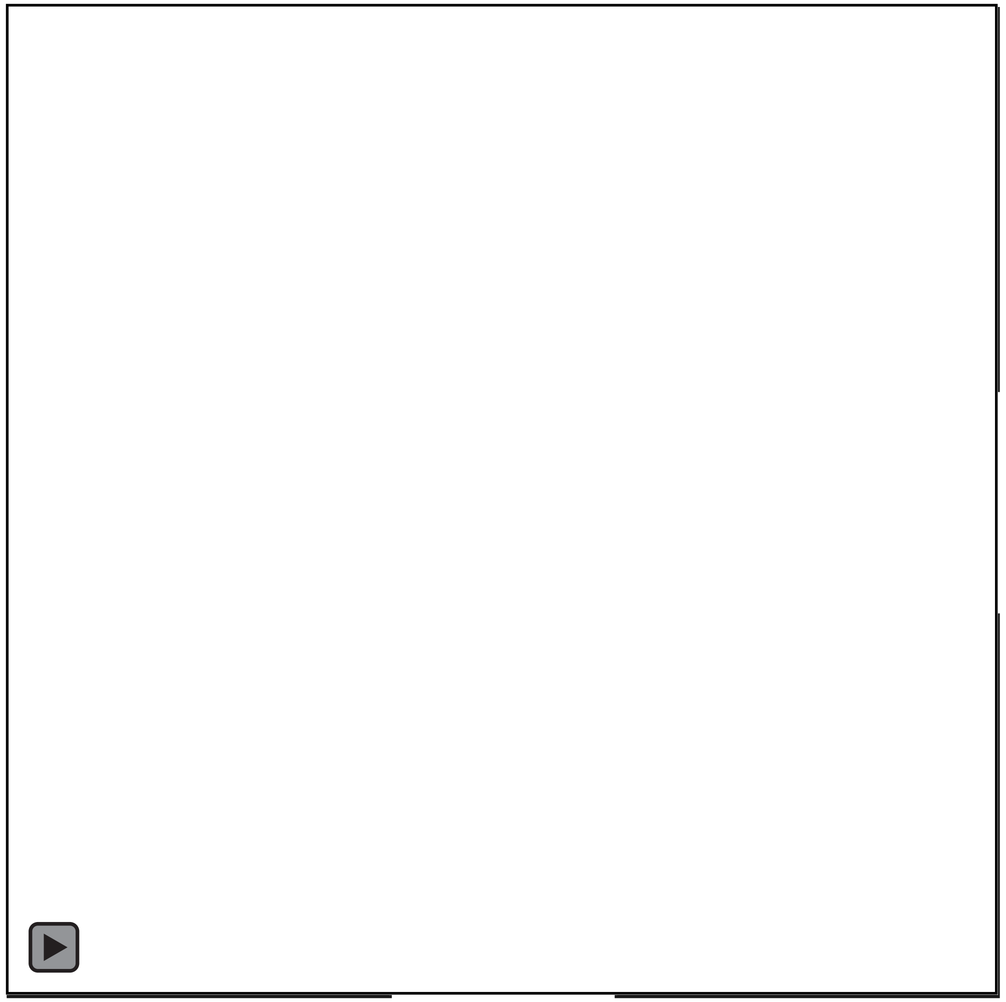
Color Context







Lorenceau & Alais (2001). Form constraints in motion binding. Nature Neuroscience



Lorenceau & Alais (2001). Form constraints in motion binding. Nature Neuroscience

Contextual effects

The perception of a point is space is altered by the surround.

Using simplified stimuli we can begin to understand these contextual effects.

There are multiple levels of explanation for such effects.

Adaptation Effects

Adaptation effects

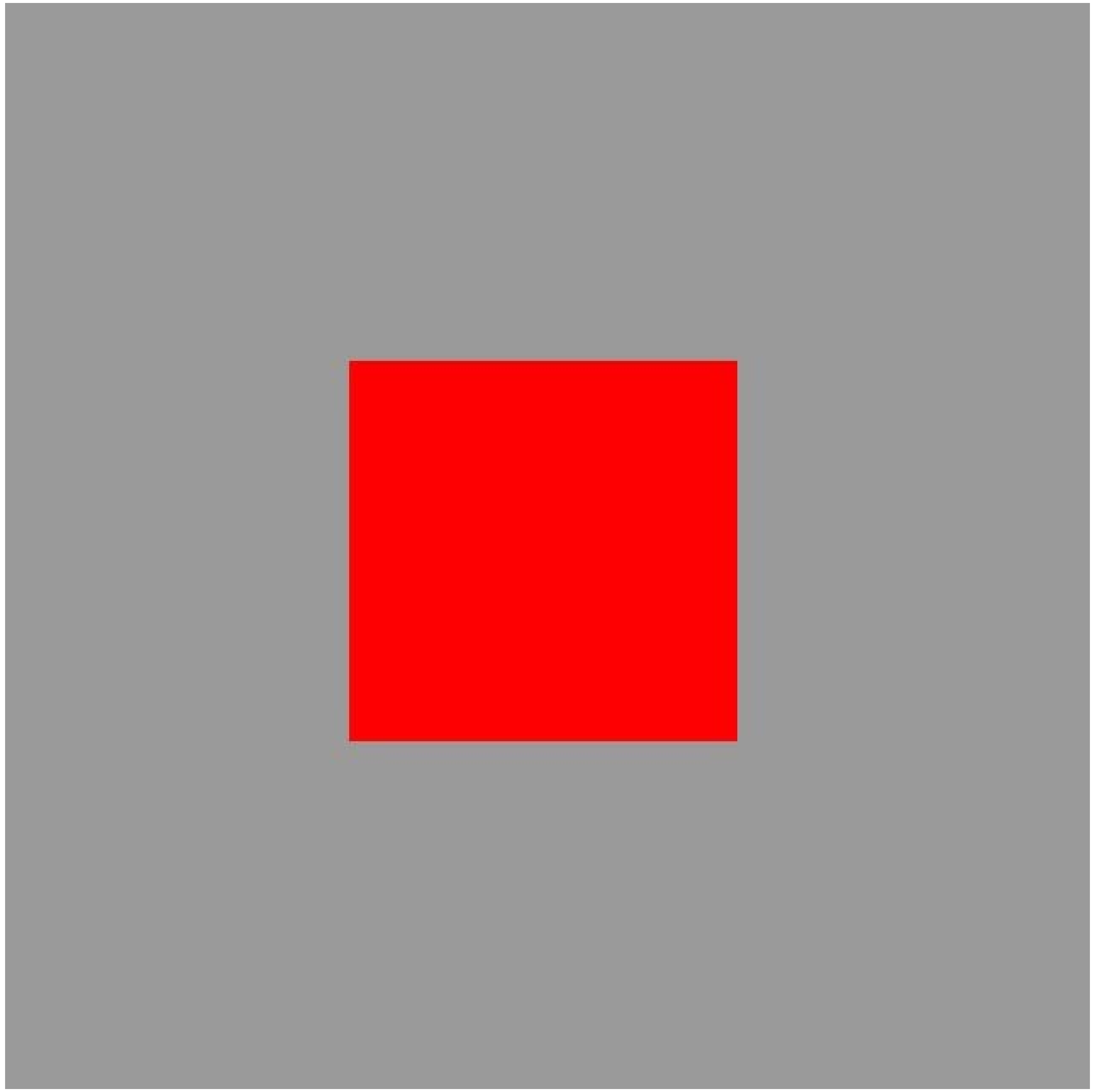
The perception at any point in time is affected by previous stimulation.

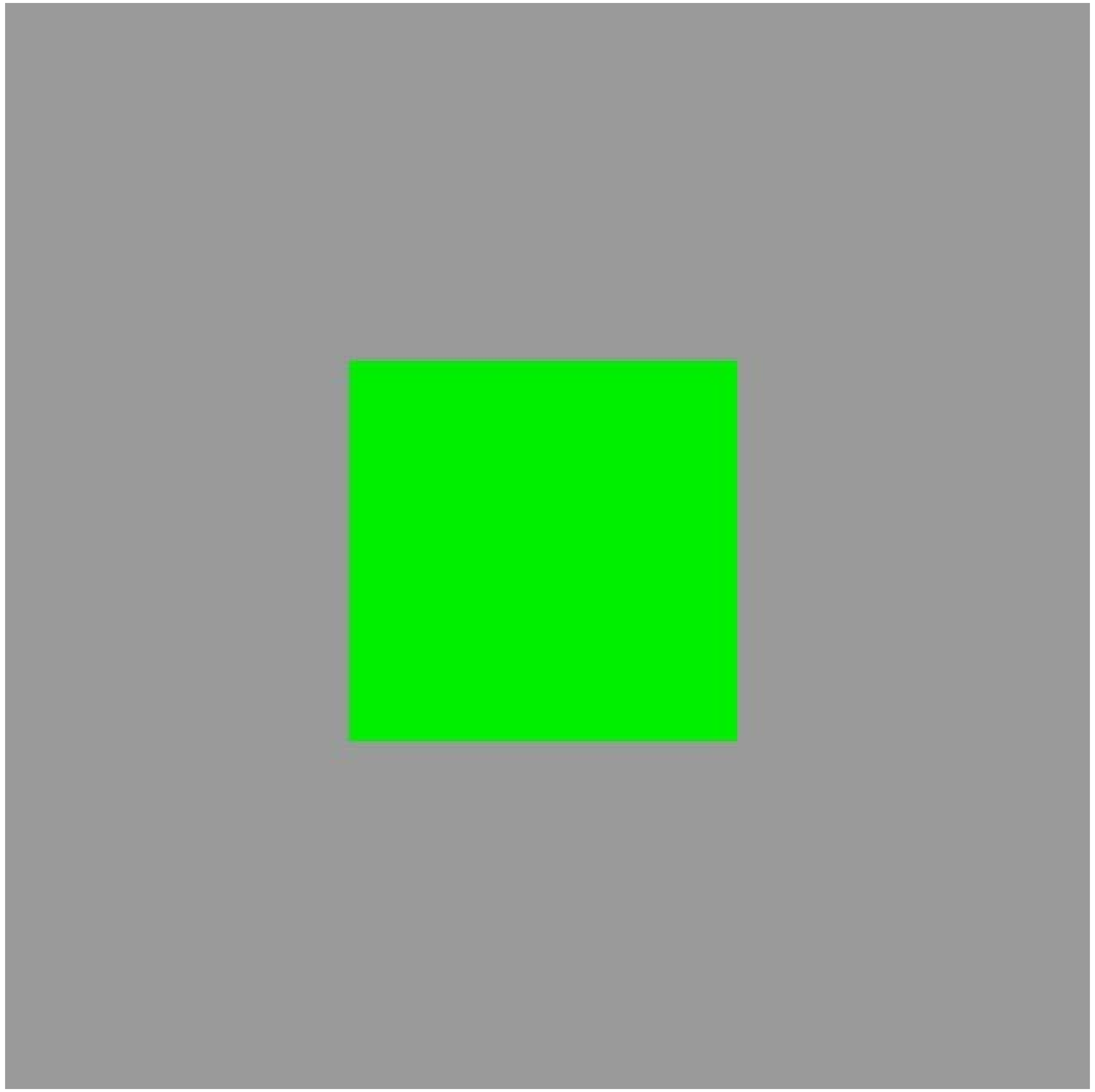
Again, using simplified stimuli we can begin to understand these adaptation effects.

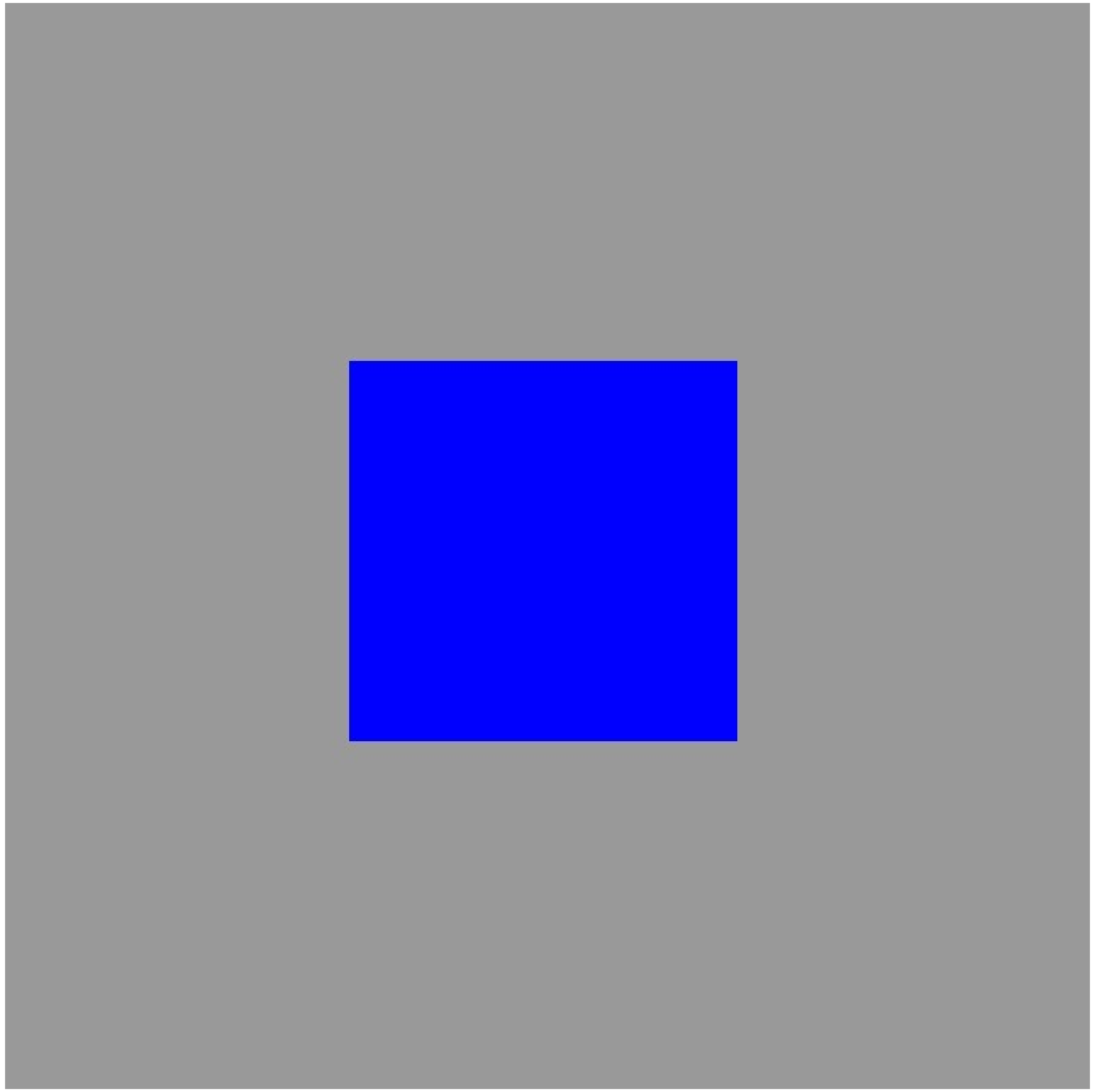
Most adaptation phenomena are explained by a gain mechanism which reduces the response of sensors tuned to the adapting stimulus.

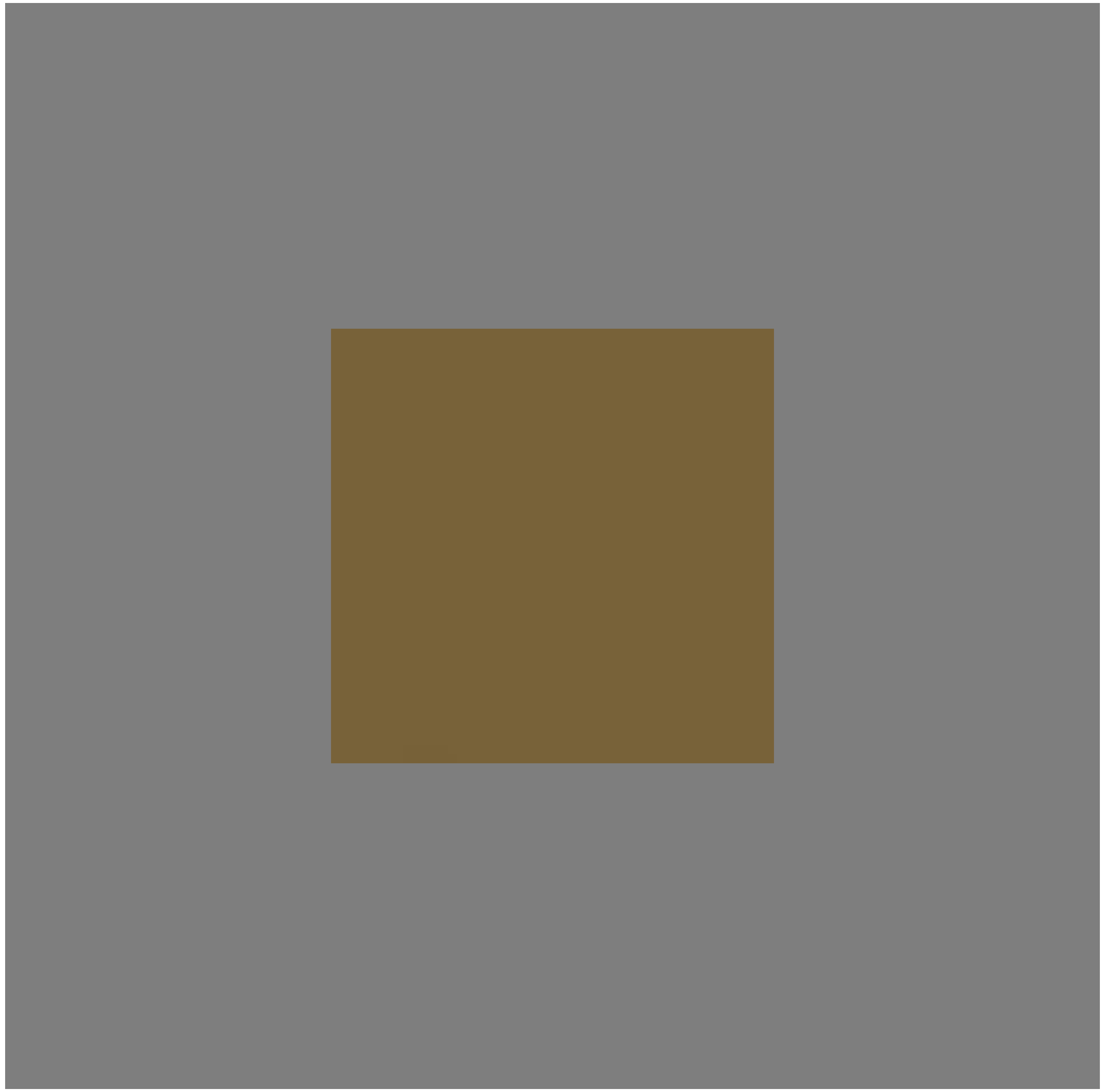
Human color constancy

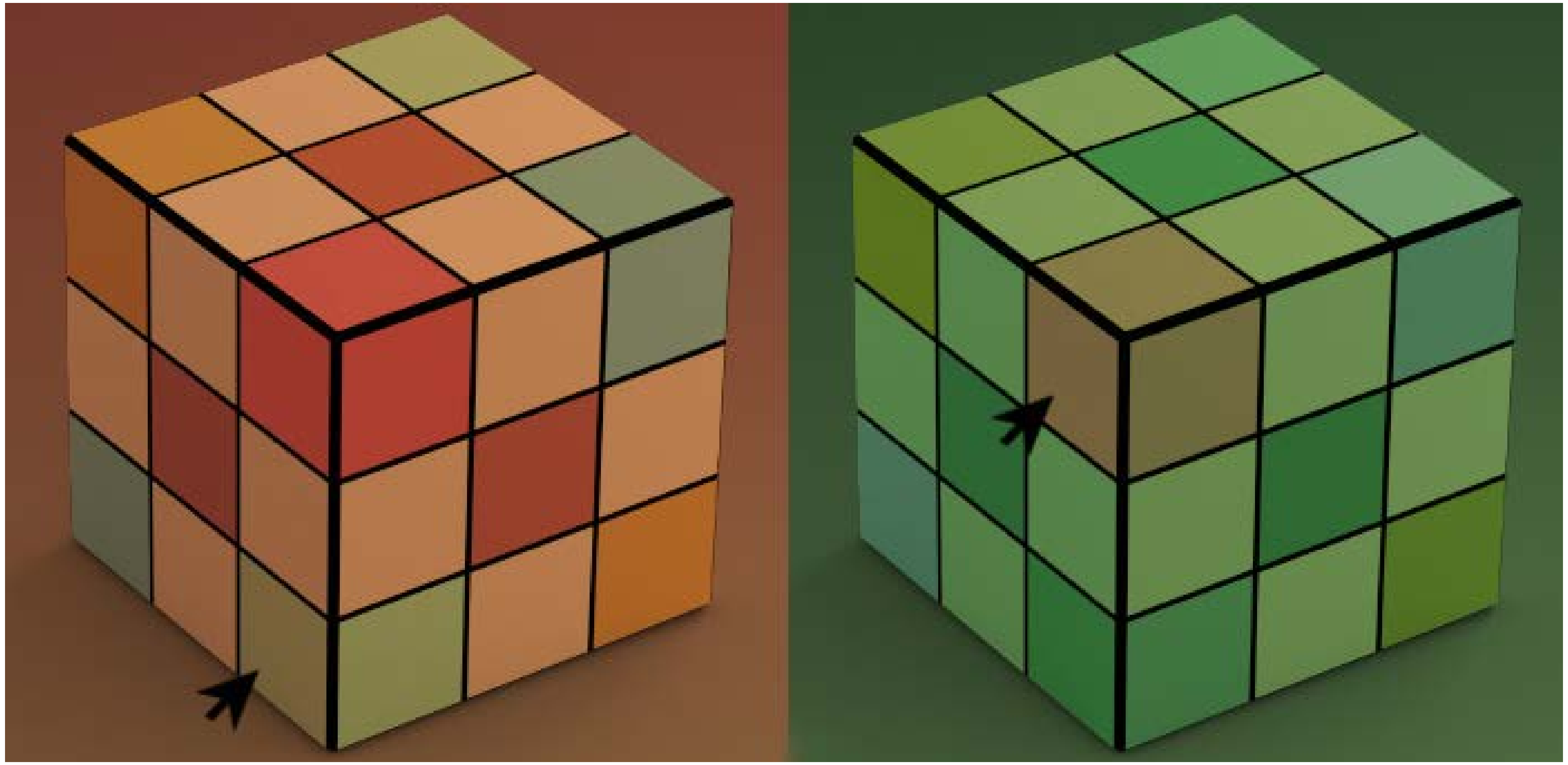
(a case of adaptation)



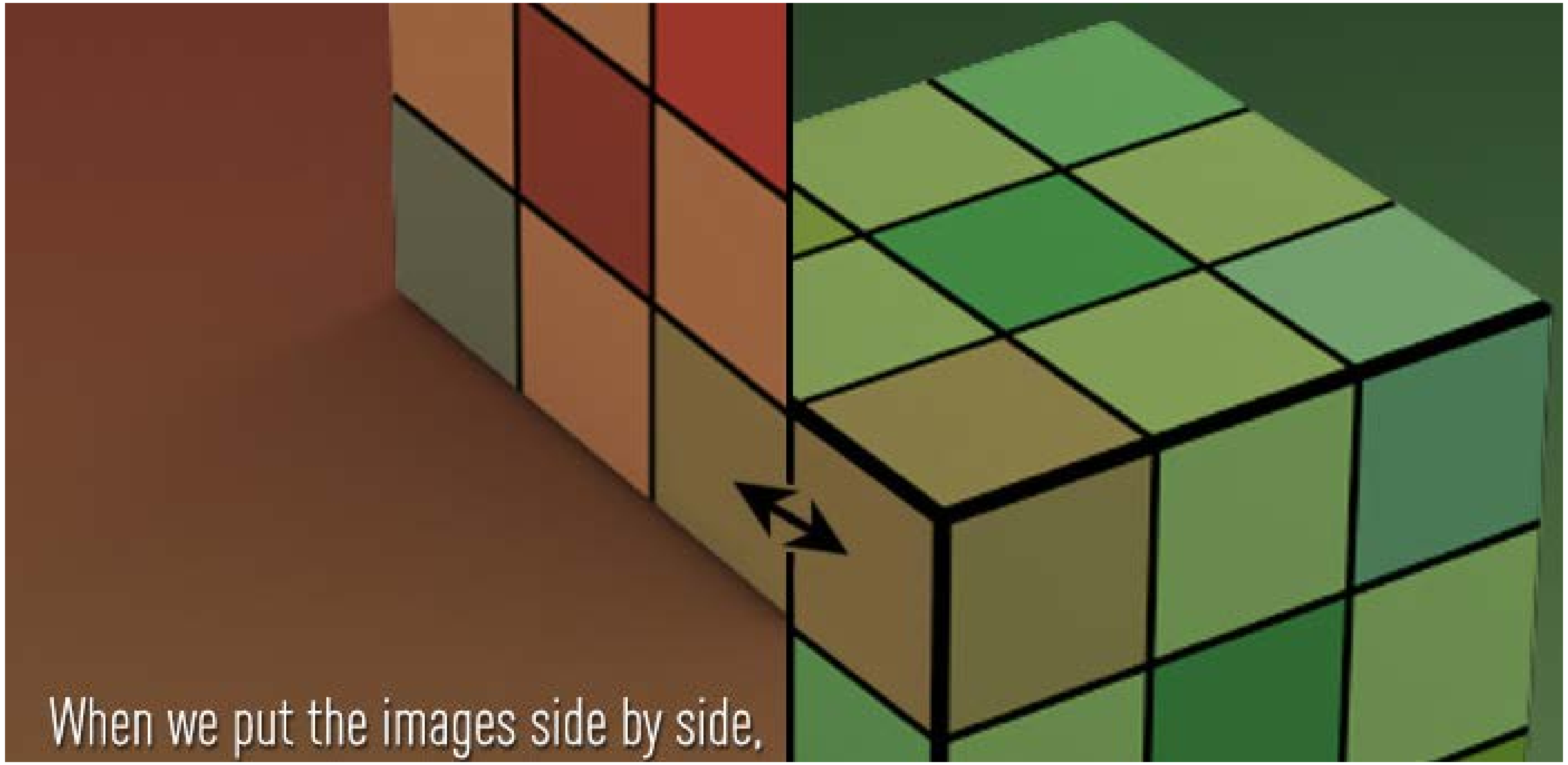


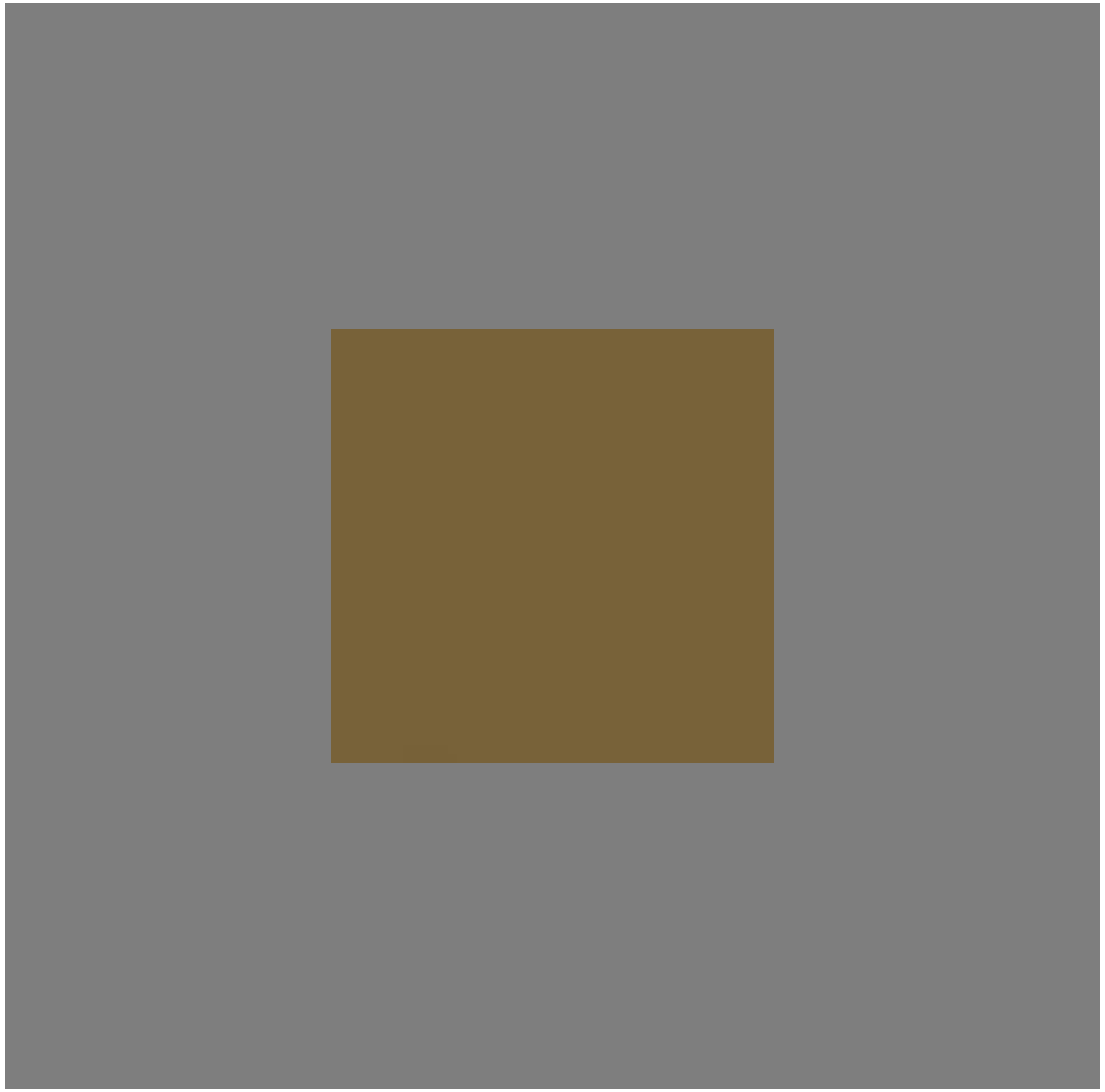


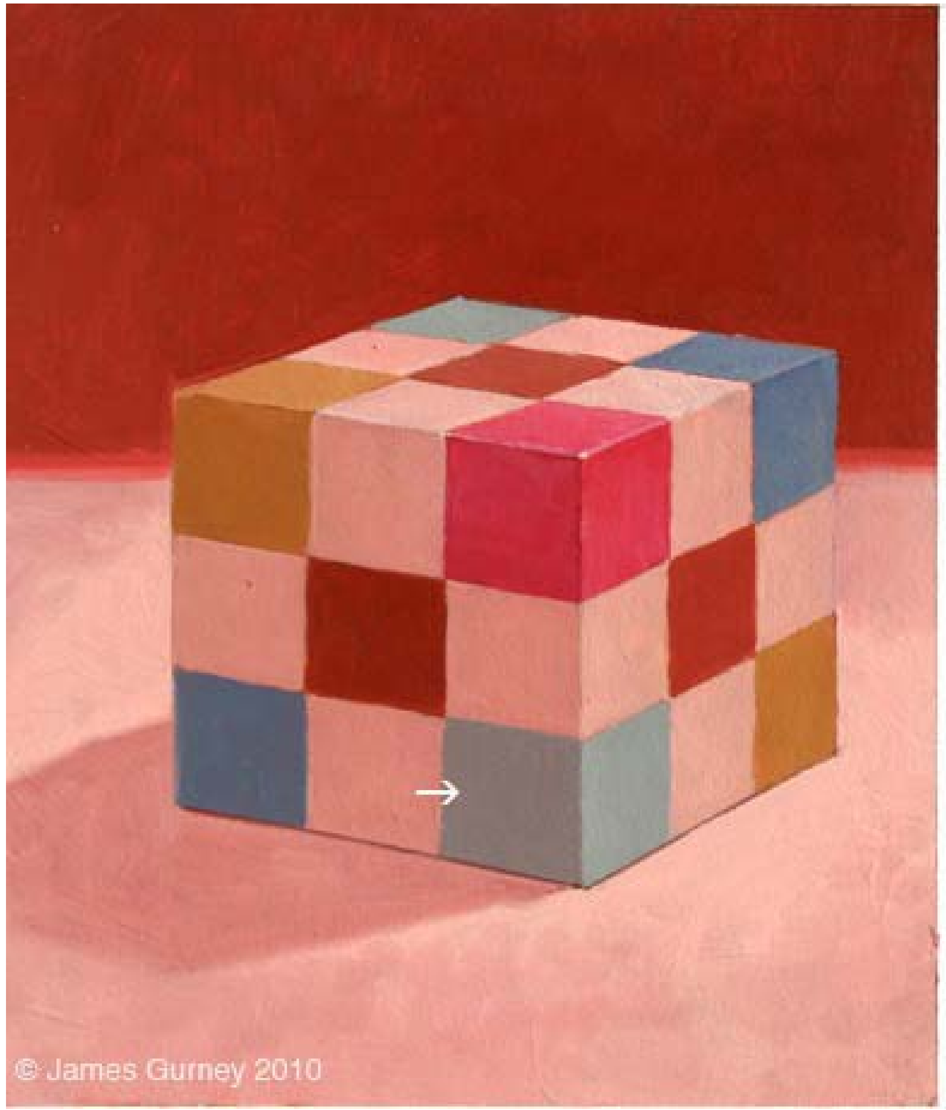




National Geographic


When we put the images side by side,




© James Gurney 2010

Modelling human colour constancy

Von Kries Law:

- Human color constancy can be modeled by a gain mechanism in each color channel

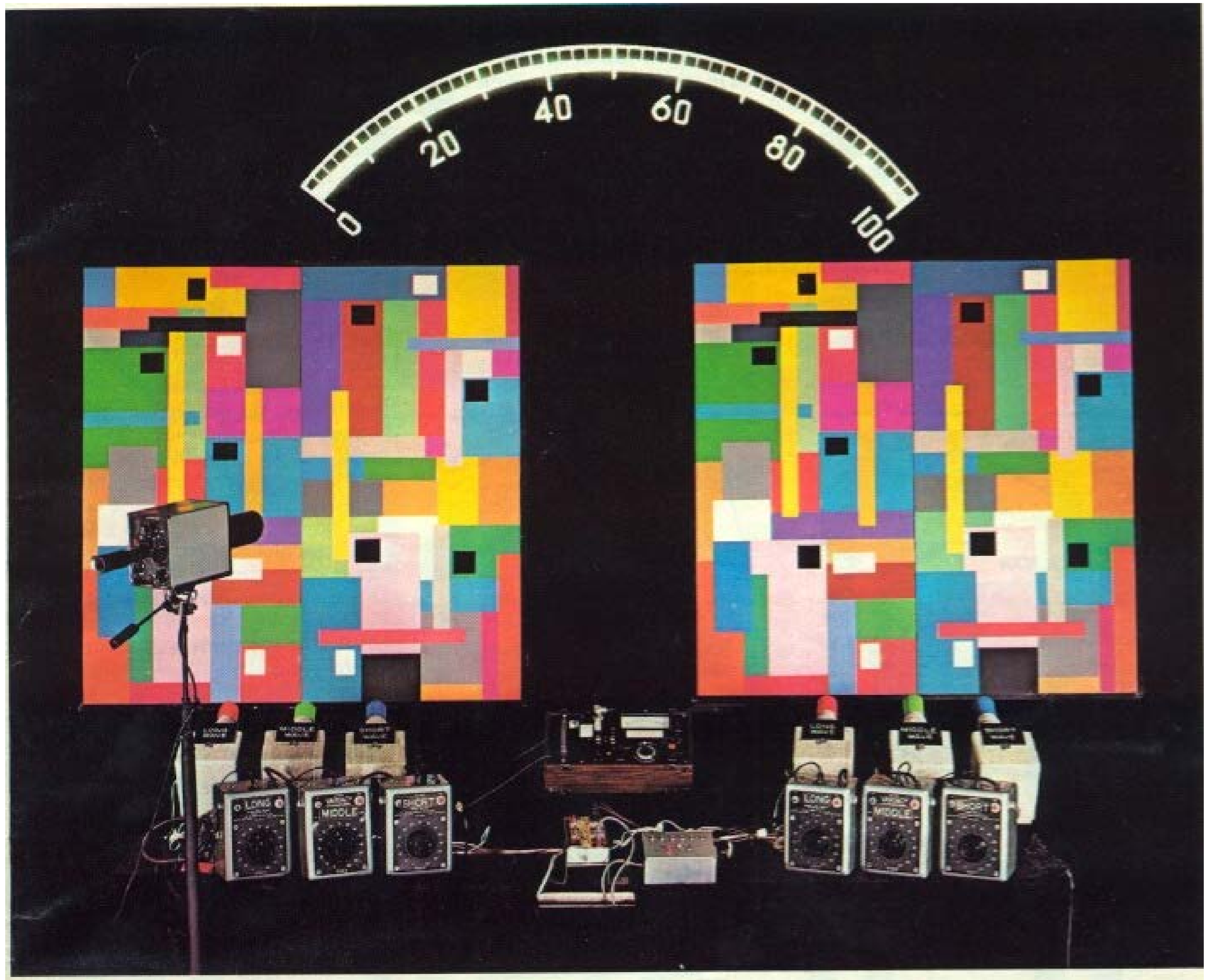
$$R' = k_r R, \quad G' = k_g G, \quad B' = k_b B,$$

Modelling human colour constancy

<u>The Retinex Theory of Colour Vision:</u>

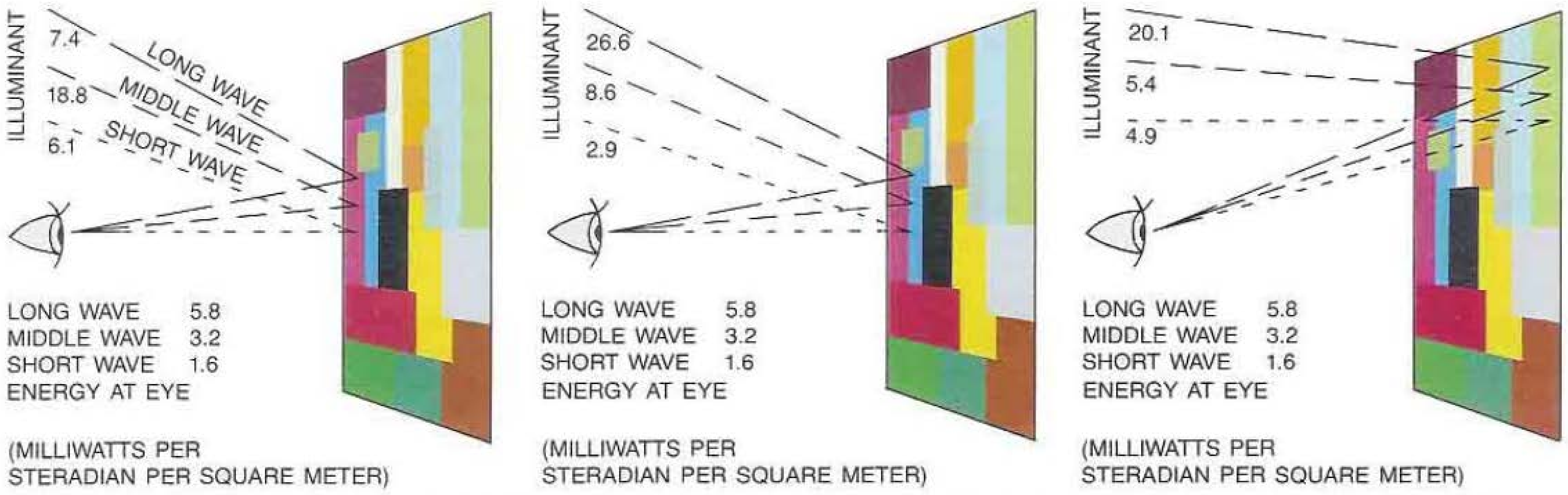
- Portmanteau formed from "retina" and "cortex".

Modelling human colour constancy

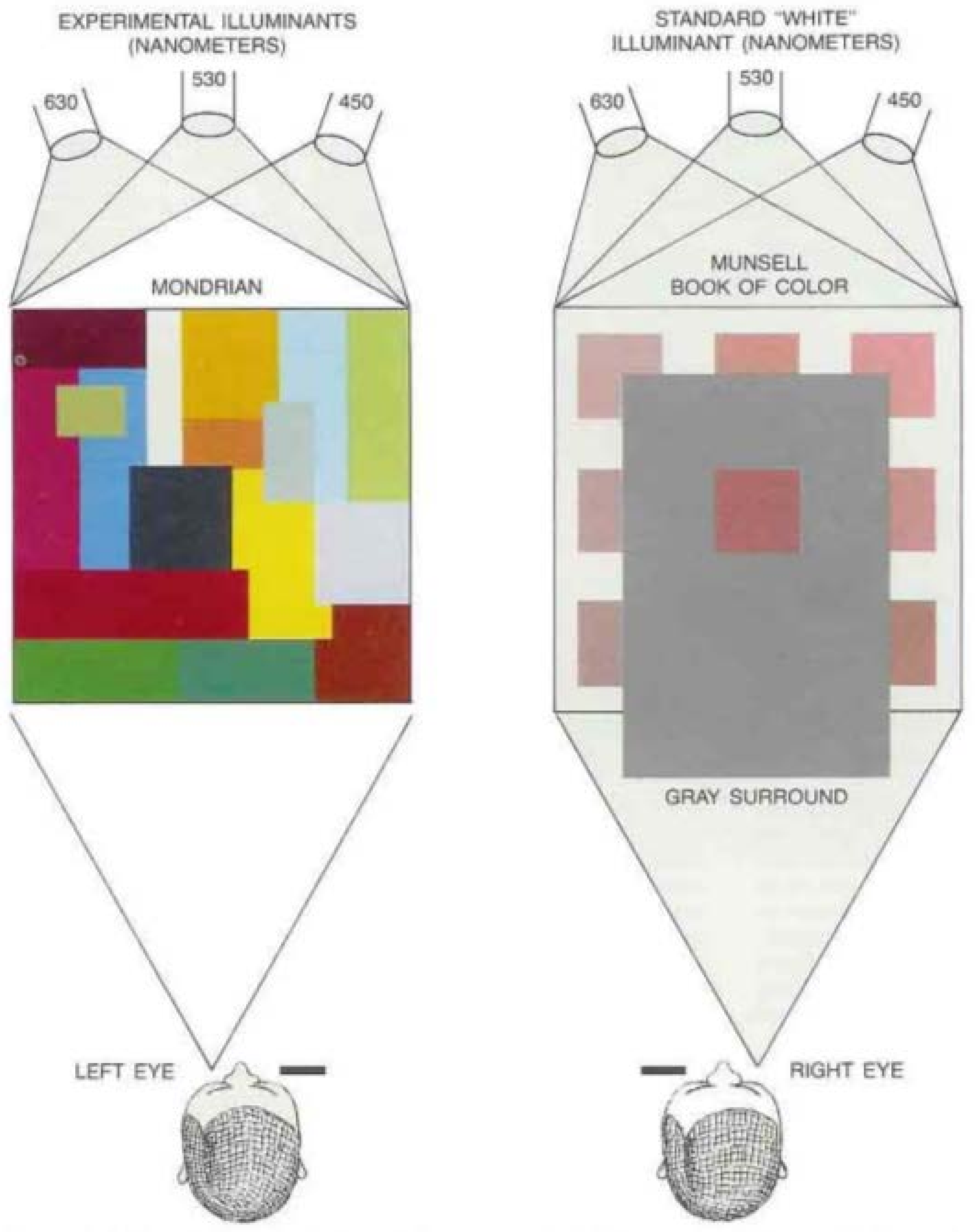


"COLOR MONDRIAN" EXPERIMENT employs two identical displays of sheets of colored paper mounted on boards four and a half feet square. The colored papers have a matte finish to minimize specular reflection. Each "Mondrian" is illuminated with its own set of three projector illuminators equipped with band-pass filters and independent brightness controls so that the long-wave ("red"), middle-wave ("green") and short-wave ("blue") illumination can be mixed in any desired ratio. A telescopic photometer can be pointed at any area to measure the flux, one wave band at a time, coming to the eye from that area. The photometer reading is projected onto the scale above the two displays. In a typical experiment the illuminators can be adjusted so that the white area in the Mondrian at the left and the green area (or some other area) in the Mondrian at the right are both sending the same triplet of radiant energies to the eye. The actual radiant-energy fluxes cannot be re-created here because of the limitations of color reproduction. Under actual viewing conditions white area continues to look white and green area continues to look green even though the eye is receiving the same flux triplet from both areas.

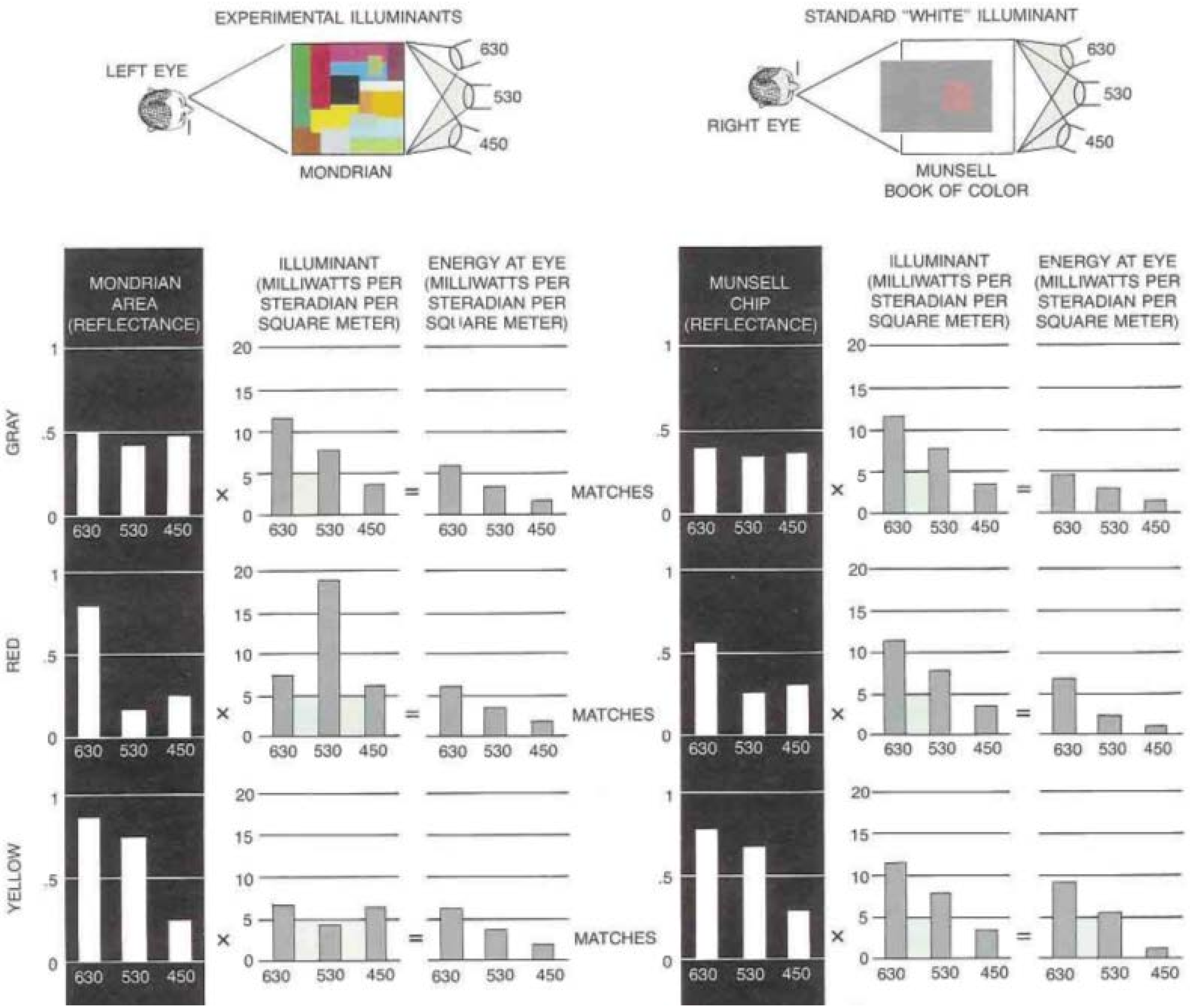
Modelling human colour constancy



Modelling human colour constancy



Modelling human colour constancy



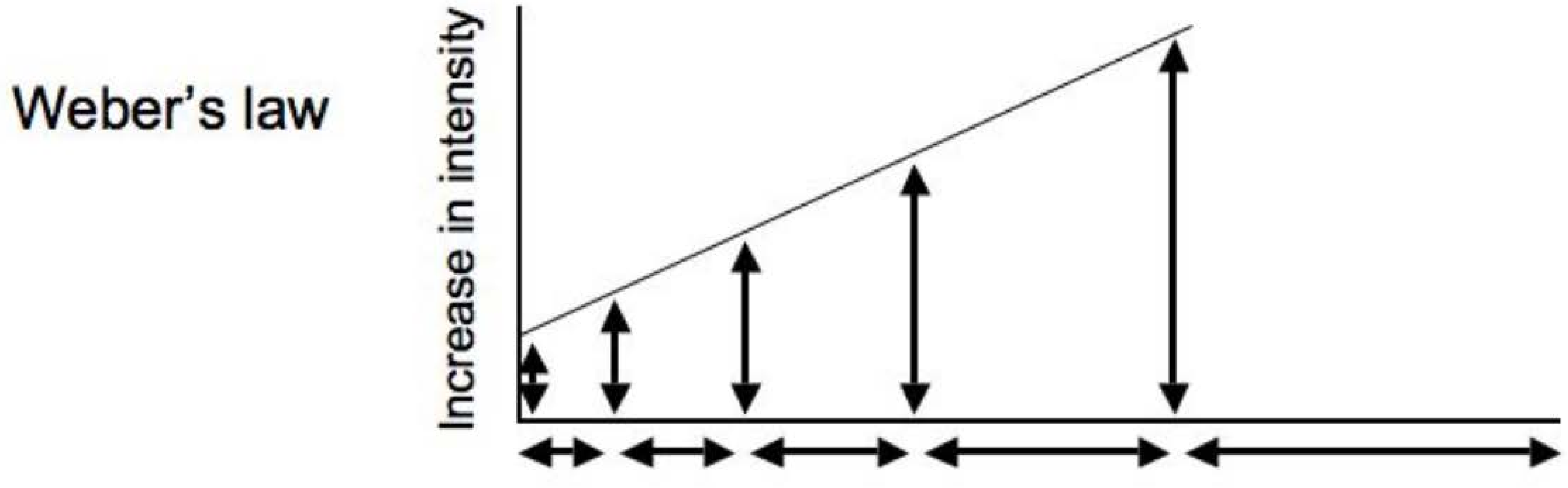
Brightness perception

Weber’s law

The smallest change in stimuli that can be perceived (dS, known as Just Noticiable Difference (JND)) is proportional to the initial intensity *S*. Mathematically,

$$dS = S \cdot K$$

S is the reference stimulus
K is a constany

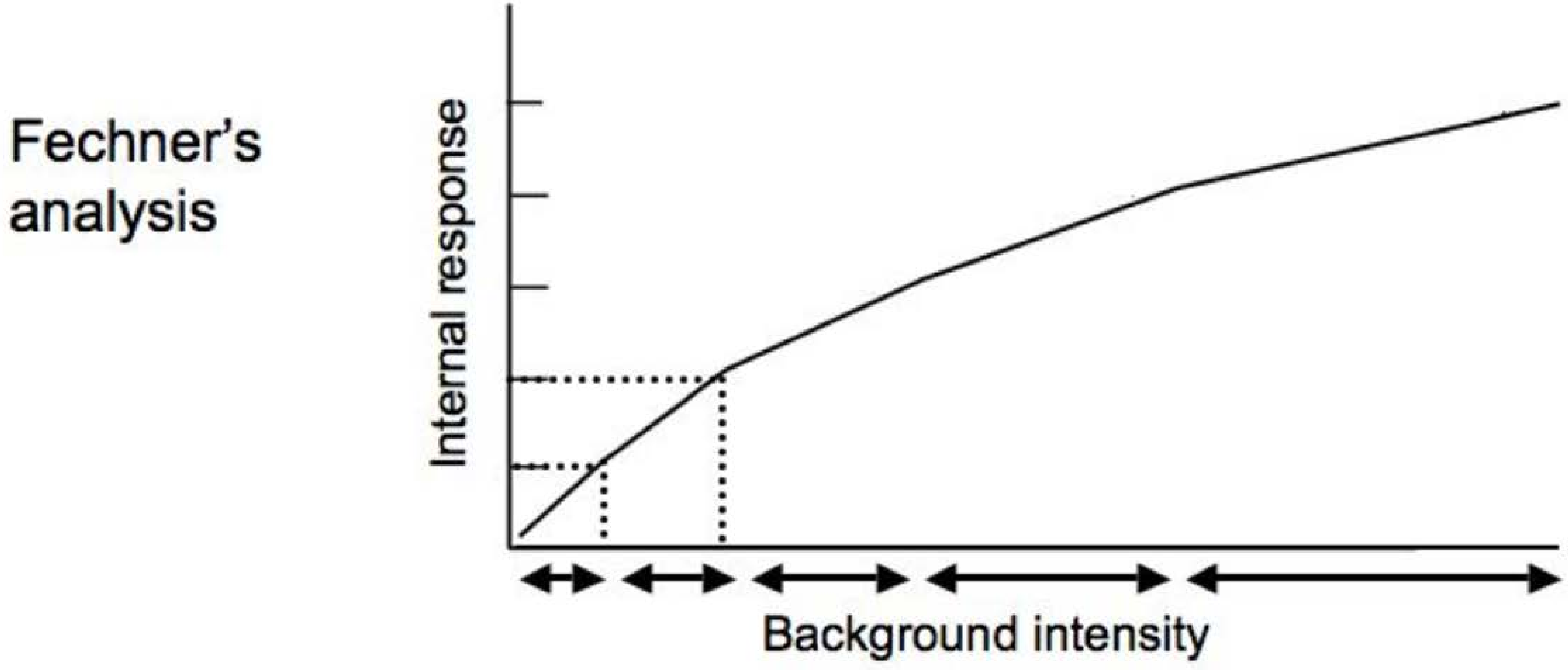


Fechner’s law

Perceived brightness is proportional to the logarithm of the actual intensity

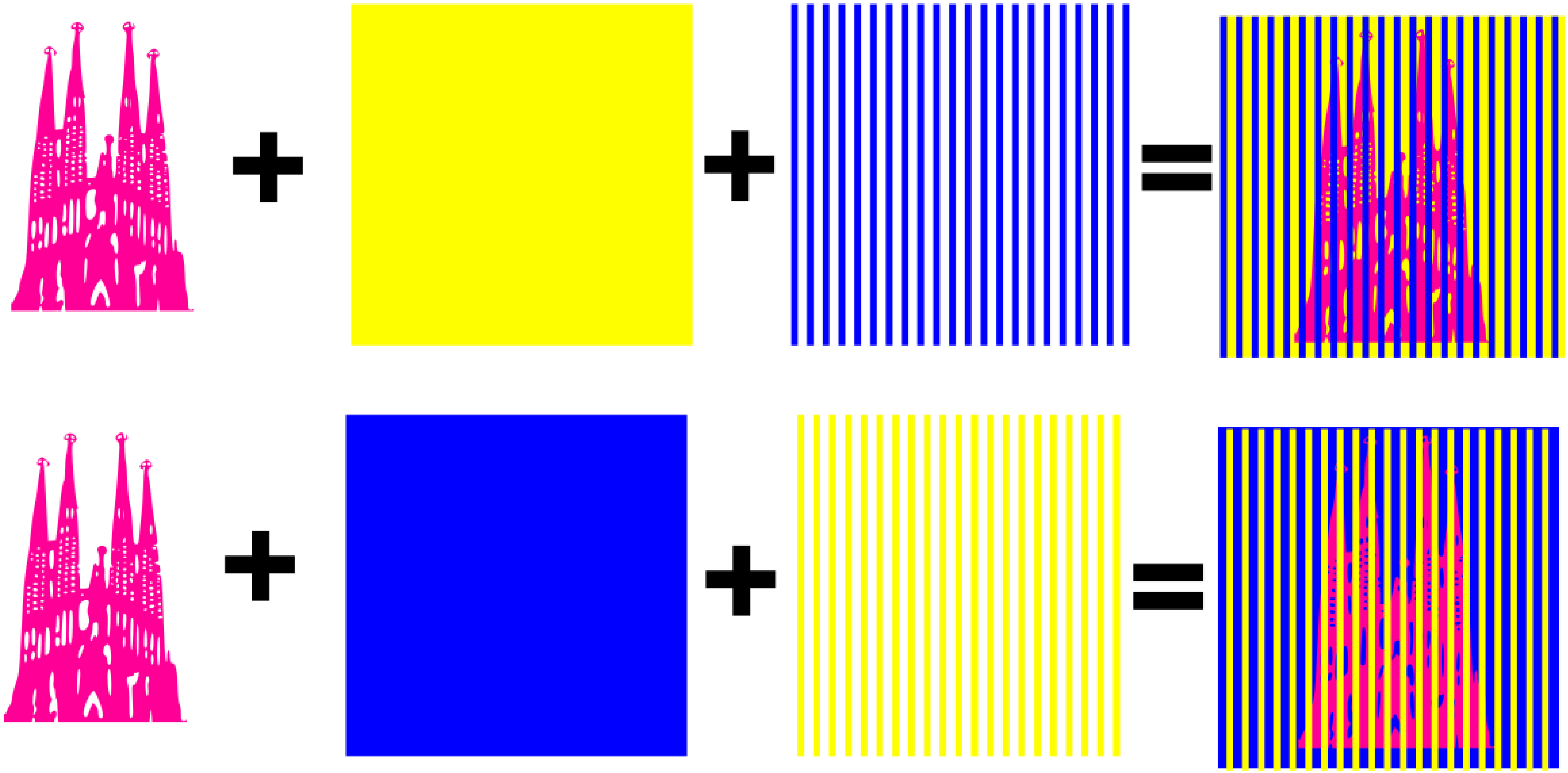
$$p = K \ln(S)$$

S is the reference stimulus
K is a constany

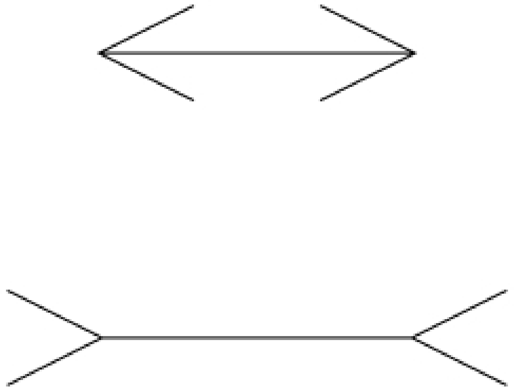


Visual illusions and CNNs

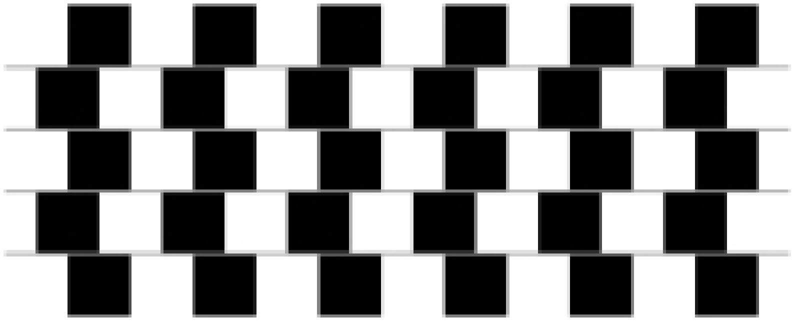
Visual Illusions



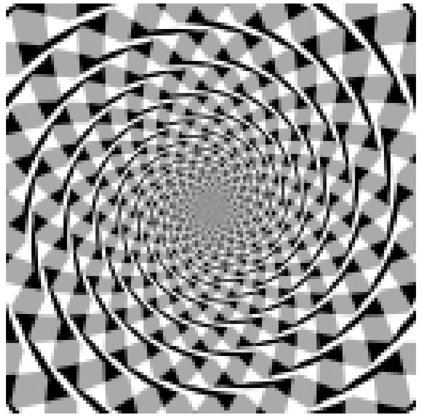
Visual Illusions



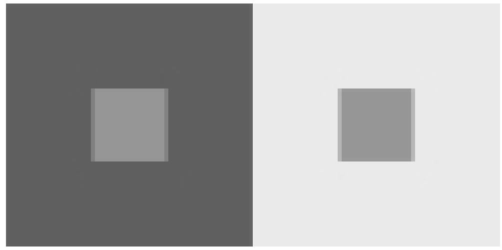
Müller-Lyer illusion



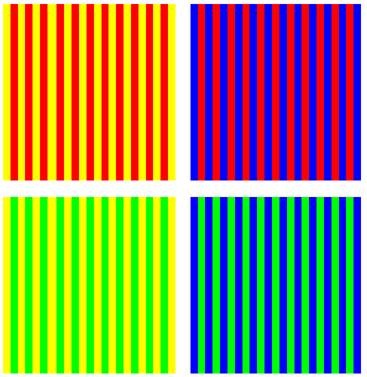
Cafe Walls illusion



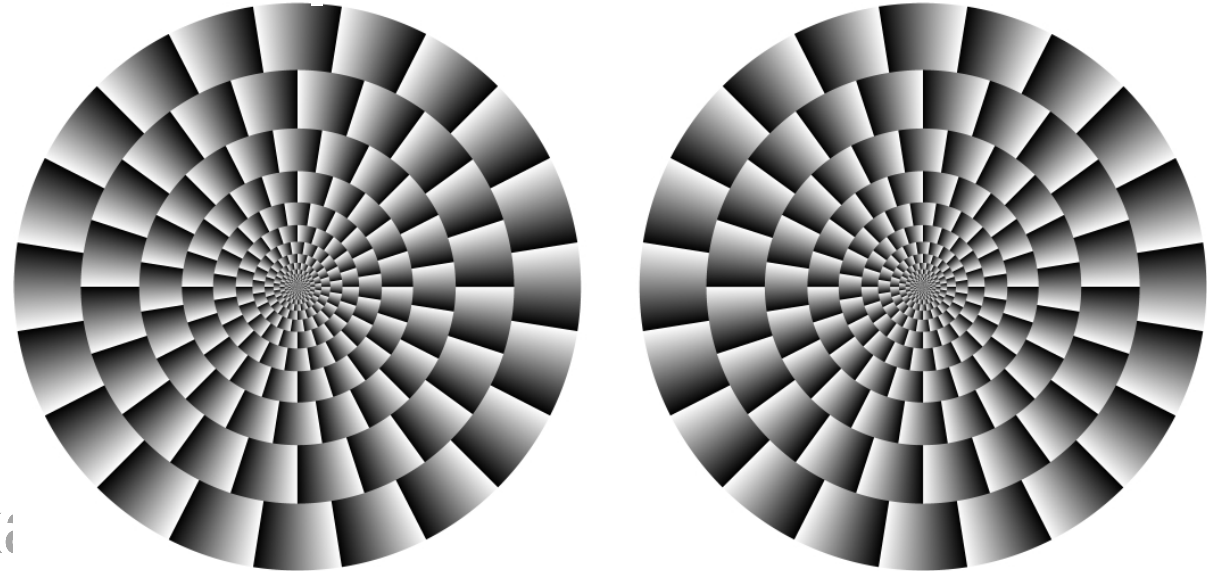
Fraser’s spiral illusion



Brightness contrast



Color assimilation



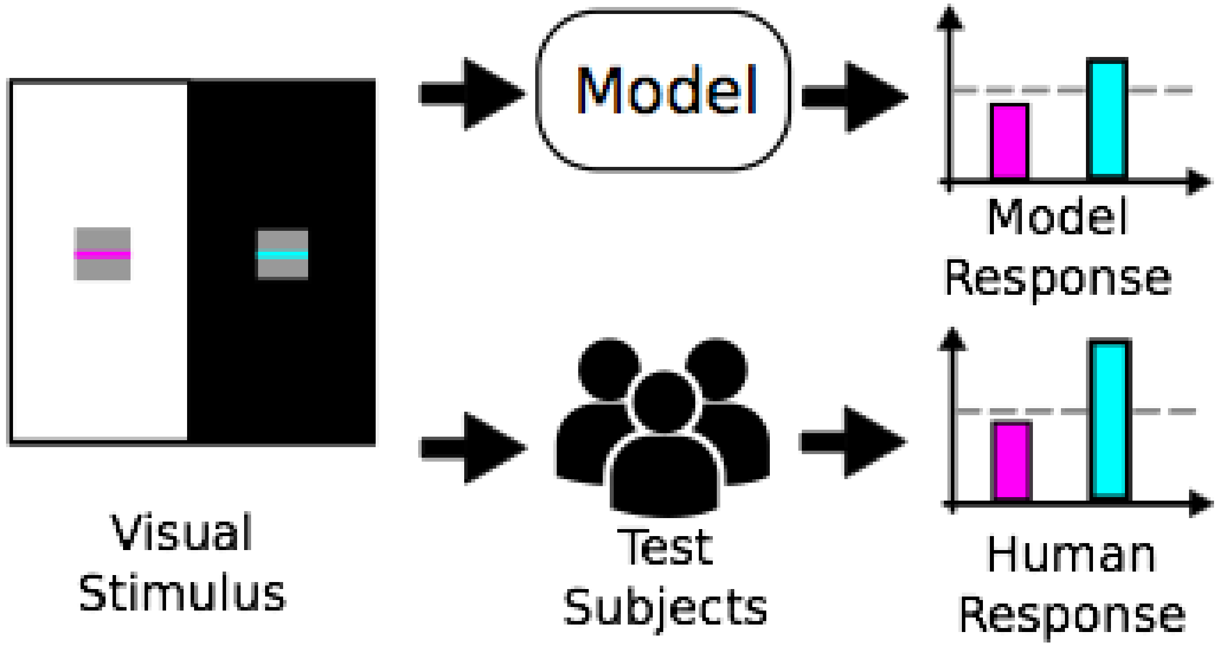
Fraser-Wilcox illusion

‘A Catalogue of illusions’ from Prof. Akitaoka

http://www.psy.ritsumei.ac.jp/~akitaoka/cataloge.html

Why are Visual illusions important in vision science?

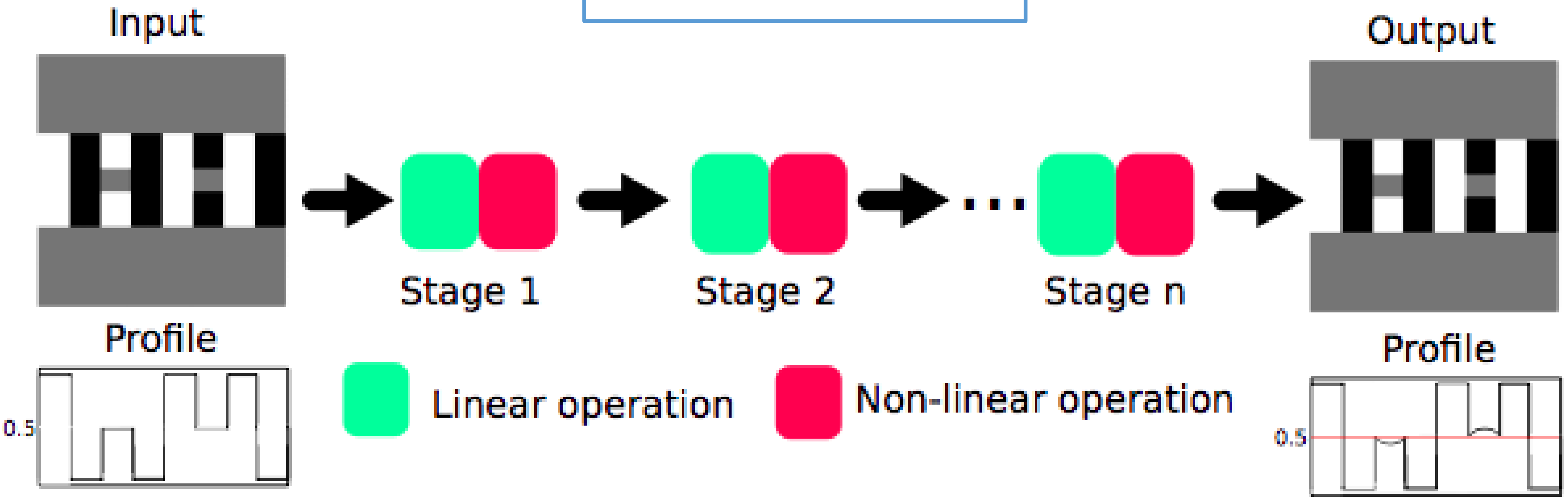
- Reveal differences between perception and reality.
- These *perception errors* are key to understand how vision works.
- Good vision models should reproduce human perception of visual illusions.



Cascade of Linear + Non Linear Operations

Perception is explained in many models of vision science and neuroscience as a **cascade** of modules composed by a **linear operation followed by a nonlinearity**.

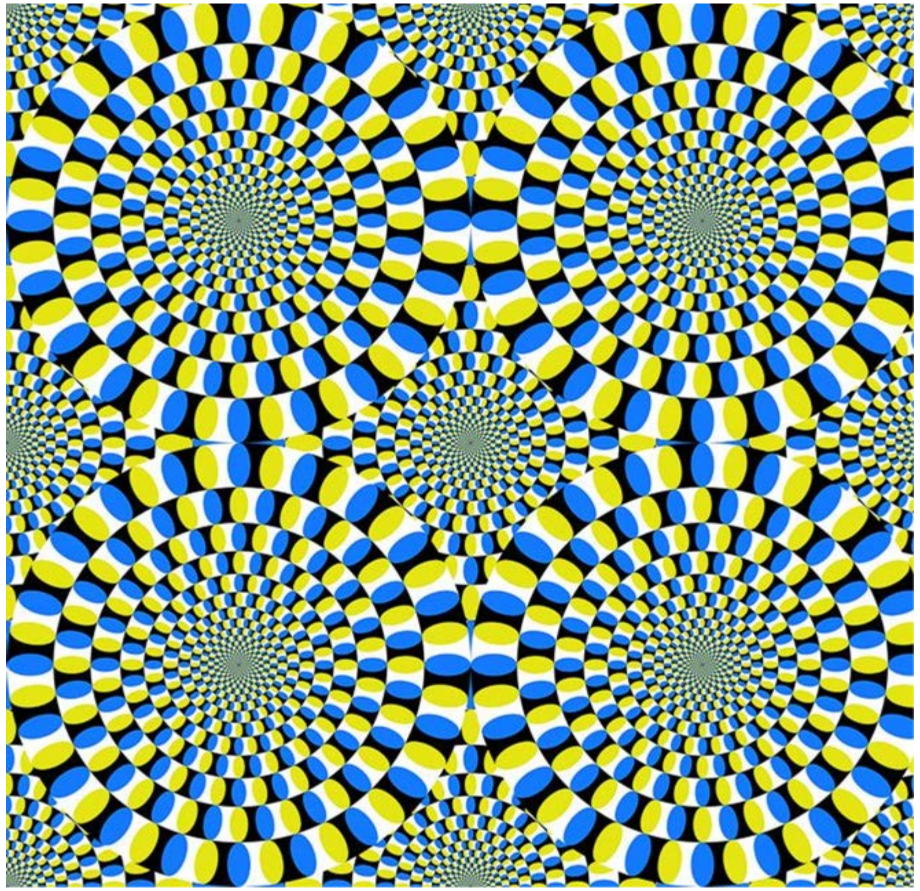
$$x^{k+1} = \boxed{f_{k+1}\left(L_{k+1}x^k, \theta_{k+1}\right)}$$



Other works linking CNNs and Visual Illusions

- *E. Watanabe et al. Illusory motion reproduced by deep neural networks trained for prediction. Frontiers in psychology, 9:345, 2018. 2*

- Kim, B., Reif, E., Wattenberg, M. and Bengio, S., 2019. Do Neural Networks Show Gestalt Phenomena? An Exploration of the Law of Closure. arXiv preprint arXiv:1903.01069.
- Sun, E.D. and Dekel, R., 2019. ImageNet-trained deep neural network exhibits illusion-like response to the Scintillating Grid. arXiv preprint arXiv:1907.09019.
- Ward, E.J., 2019. Exploring perceptual illusions in deep neural networks. bioRxiv, p.687905.
- Anonymous (ICLR 2020 submission). The function of contextual illusions.https://openreview.net/forum?id=H1gB4RVKvB
- Jacob, G., Pramod, R. T., Katti, H., Arun, S. P. Do deep neural networks see the way we do?



Choosing three imaging tasks related with HVS

Denoise | Deblur | Restoration

I

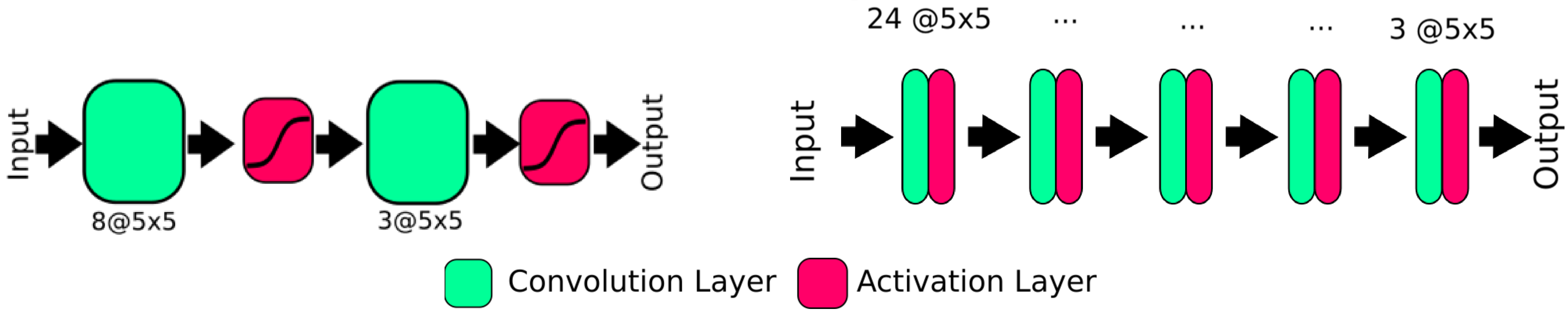
I + noise

I + blur

I + blur + noise

CNN Implementation details

DN-NET, DB-NET, Restore-Net **Deep DN-NET, DB-NET, and Restore-Net**



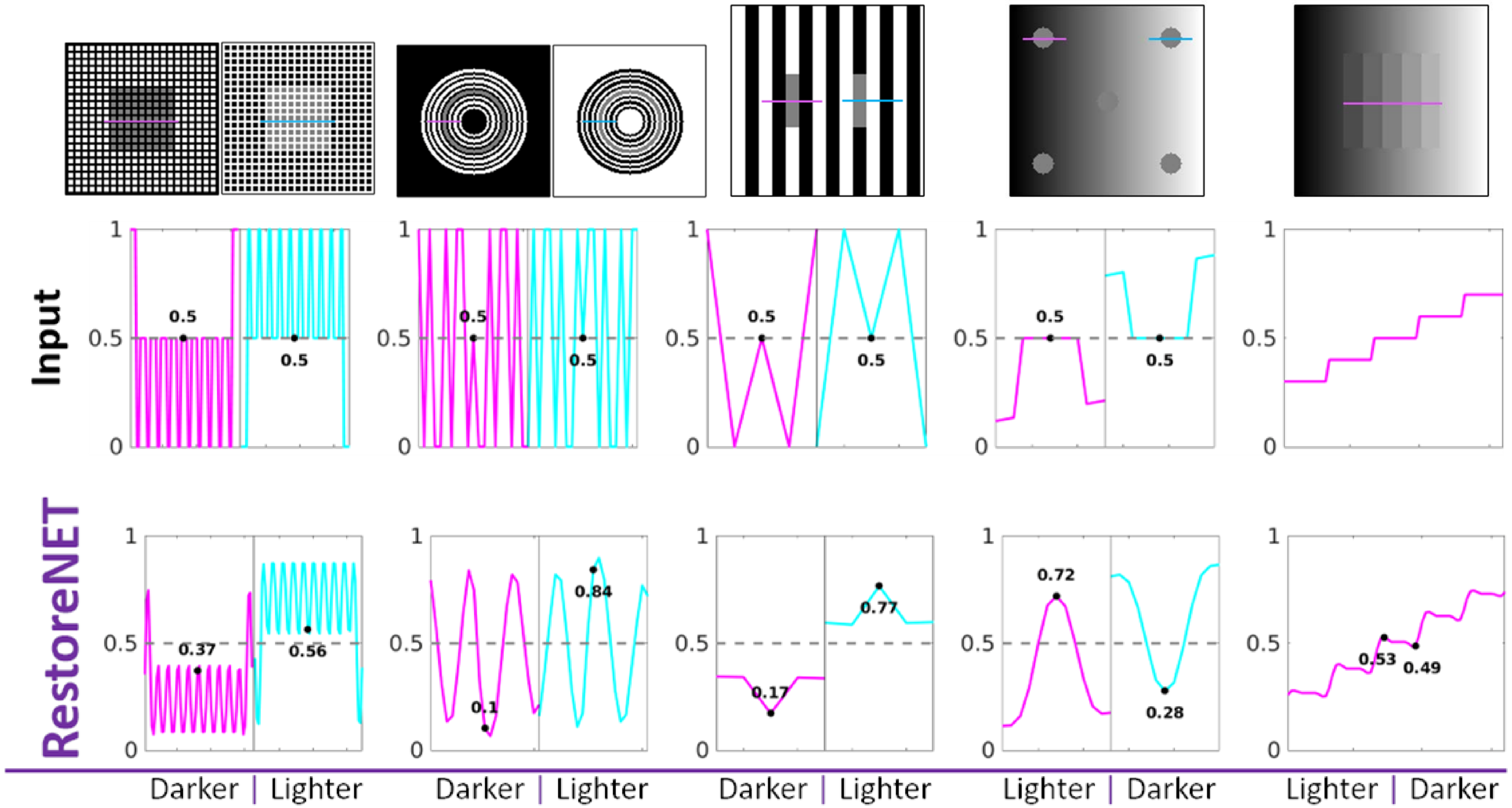
Zhang et al: Denoising deep CNN with state-of-the-art performance

Zhang, K., Zuo, W., Chen, Y., Meng, D. and Zhang, L., 2017. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE Transactions on Image Processing, 26(7), pp.3142-3155.

Dataset: All of them trained on a subset of ImageNet

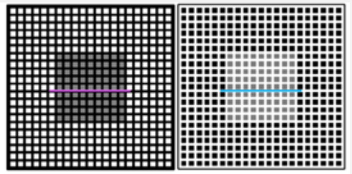
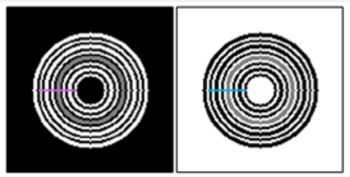
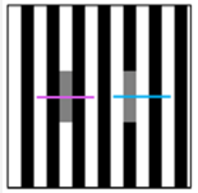
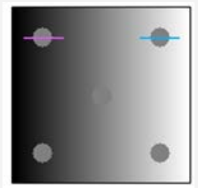
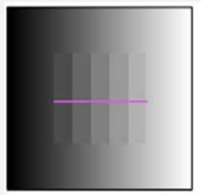
Loss: Mean squared error

Results

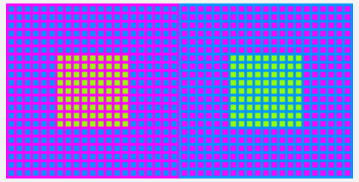
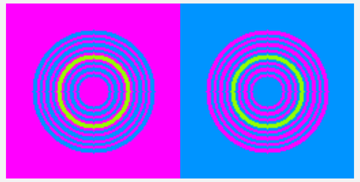
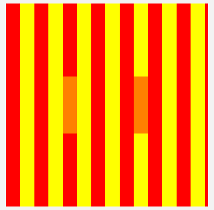
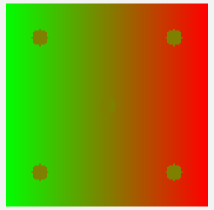


Input
RestoreNET
0.5
0.37
0.56
0.1
0.84
0.17
0.77
0.72
0.28
0.53
0.49
Darker | Lighter
Darker | Lighter
Darker | Lighter
Lighter | Darker
Lighter | Darker

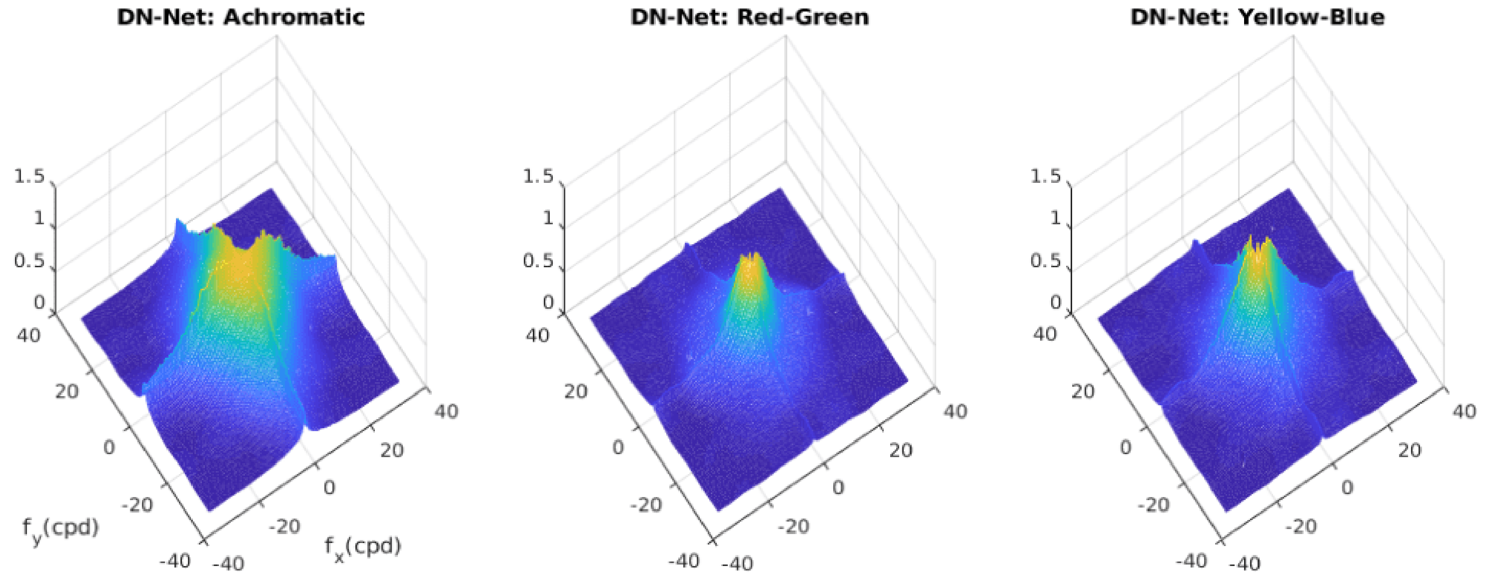
Summary of replication of grayscale VIs

Visual Illusion					
DN-NET	✓	✓	✓	✓	✗
DB-NET	✓	✓	✓	✓	✓
Restore-Net	✓	✓	✓	✓	✓
Deep DN-NET	✓	✓	✓	✗	✗
Deep DB-NET	✗	✓	✓	✓	✓
Deep RestoreNet	✓	✓	✓	✓	✓
Zhang et al.	✓	✗	✗	✓	✗

Summary of replication of color VIs

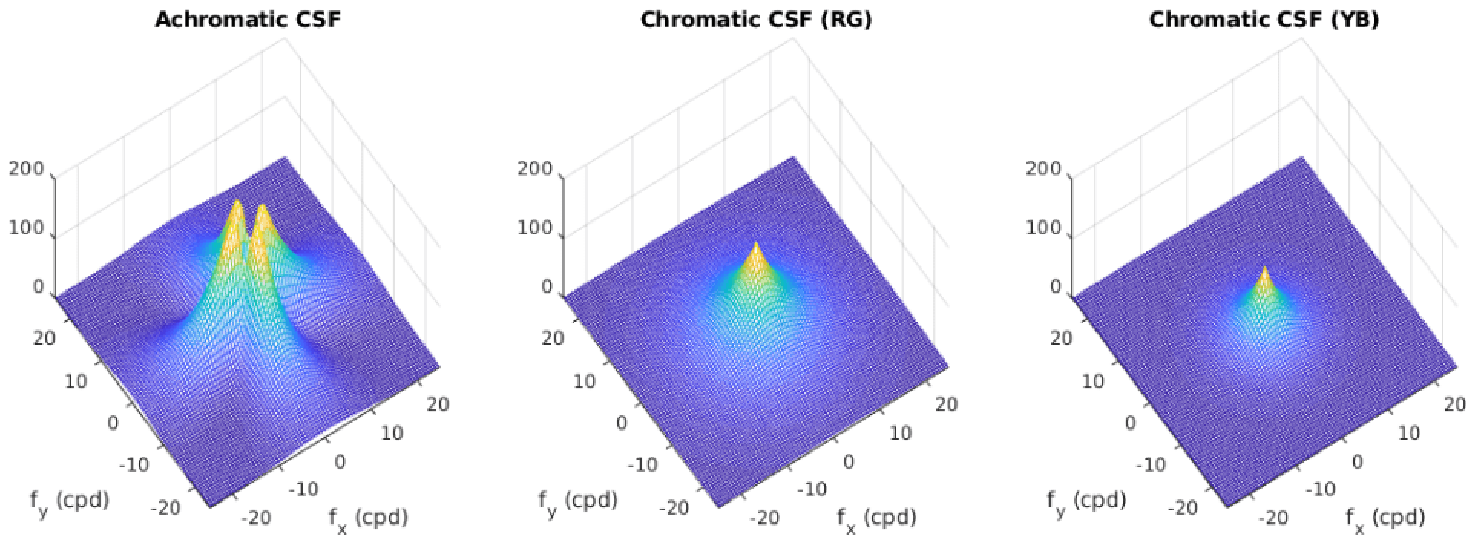
Visual Illusion					
DN-NET	✓	✓	✓	✗	✓
DB-NET	✓	✓	✓	✗	✓
Restore-Net	✓	✓	✓	✗	✓
Deep DN-NET	✓	✓	✓	✗	✓
Deep DB-NET	✓	✓	✓	✓	✓
Deep RestoreNet	✓	✓	✓	✗	✓
Zhang et al.	✓	✗	✗	✓	✗

A comparison with human perception



Contrast sensitivity function of DB-NET

Gomez-Villa, A., Martín, A., Vazquez-Corral, J., Bertalmío, M. and Malo, J., 2019. Visual Illusions Also Deceive Convolutional Neural Networks: Analysis and Implications. arXiv preprint arXiv:1912.01643.



Human contrast sensitivity functions

Mullen, K. T. (1985). The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. The Journal of physiology, 359(1), 381-400.

Watson, A. B., & Malo, J. (2002, September). Video quality measures based on the standard spatial observer. In Proceedings. International Conference on Image Processing(Vol. 3, pp. III-III). IEEE.

Human visual system and perception

Javier Vazquez-Corral
Universitat Pompeu Fabra, Spain

javier.vazquez@upf.edu

Slides credits: Marcelo Bertalmío and David Kane