University of Warsaw
Faculty of Mathematics, Informatics and Mechanics

Data analysis and visualization (DAV)

Lecture 06

Łukasz P. Kozłowski

Warsaw, 2025

l.kozlowski@mimuw.edu.pl

Data analysis and visualization (DAV)

Lecture 06
Illusions

Łukasz P. Kozłowski

Warsaw, 2025

l.kozlowski@mimuw.edu.pl

The importance of data visualization

> *...make **both** calculations **and** graphs. Both sorts of output should be studied; each will contribute to understanding.*
>
> *— F. J. Anscombe, 1973*
>
> *(and echoed in nearly all talks about data visualization...)*

The importance of data visualization

Anscombe's quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Consider four datasets

The importance of data visualization

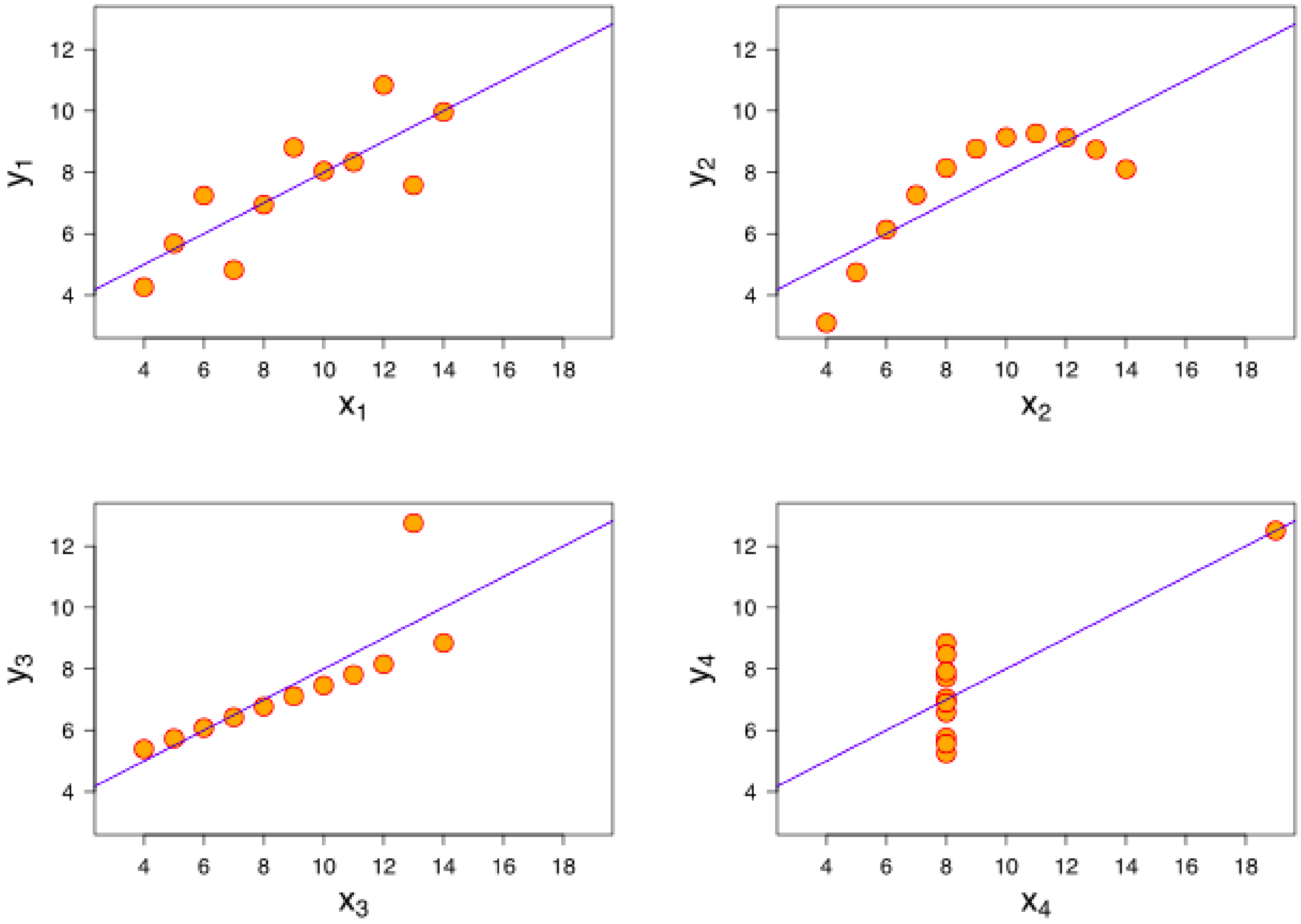
Anscombe's quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71

Property	Value	Accuracy
Mean of *x*	9	exact
Sample variance of *x* : σ^2	11	exact
Mean of *y*	7.50	to 2 decimal places
Sample variance of *y* : σ^2	4.125	±0.003
Correlation between *x* and *y*	0.816	to 3 decimal places
Linear regression line	$y = 3.00 + 0.500x$	to 2 and 3 decimal places, respectively
Coefficient of determination of the linear regression : R^2	0.67	to 2 decimal places

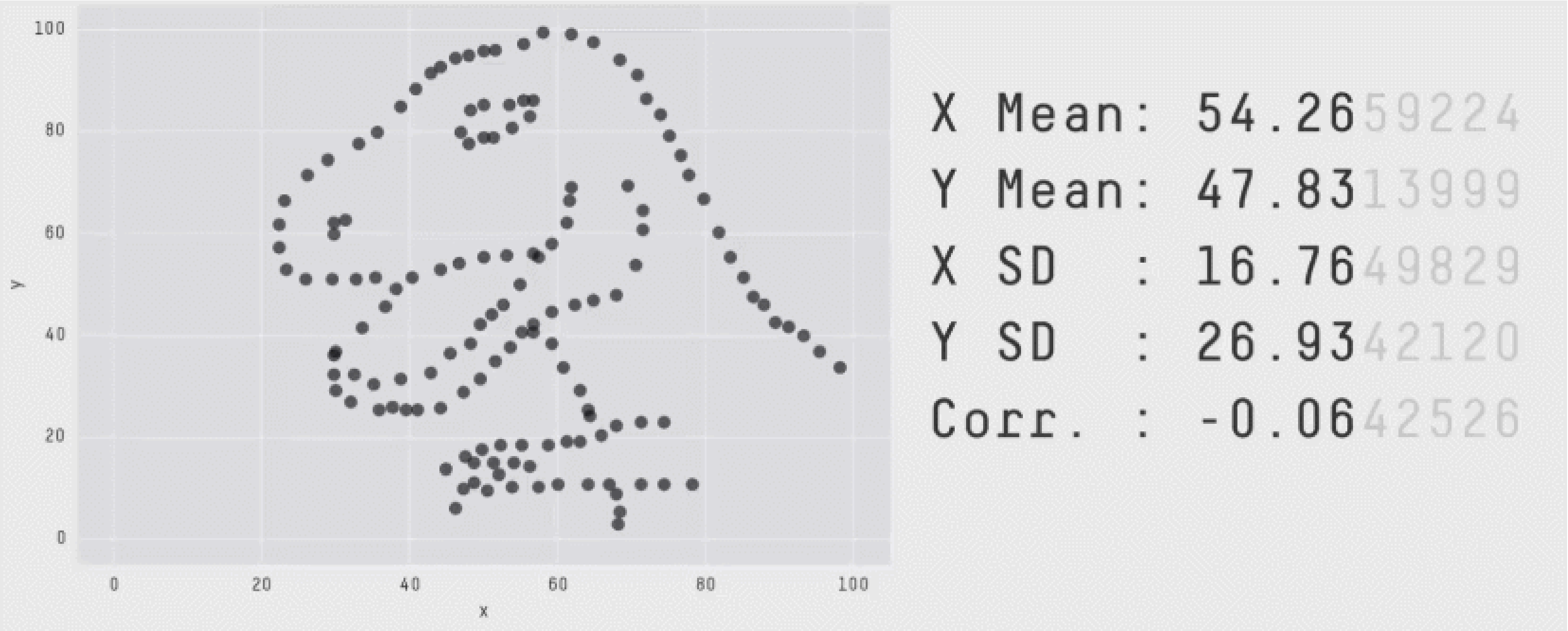
All four datasets are identical when examined using simple summary statistics

The importance of data visualization



Yet, all four datasets vary considerably when graphed

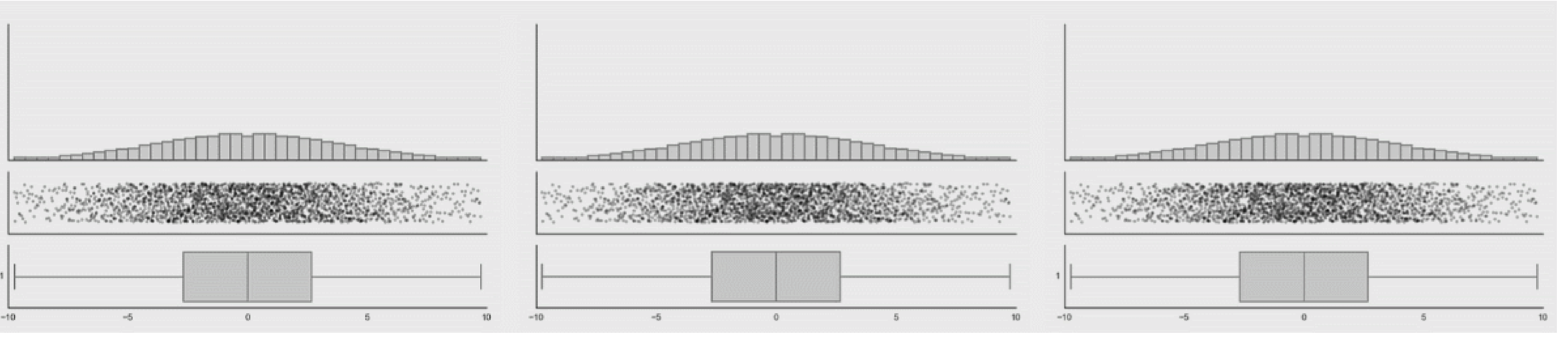
The importance of data visualization



It is not known how Anscombe created his dataset. Since its publication, several methods to generate similar data sets with identical statistics and dissimilar graphics have been developed.

https://www.autodeskresearch.com/publications/samestats

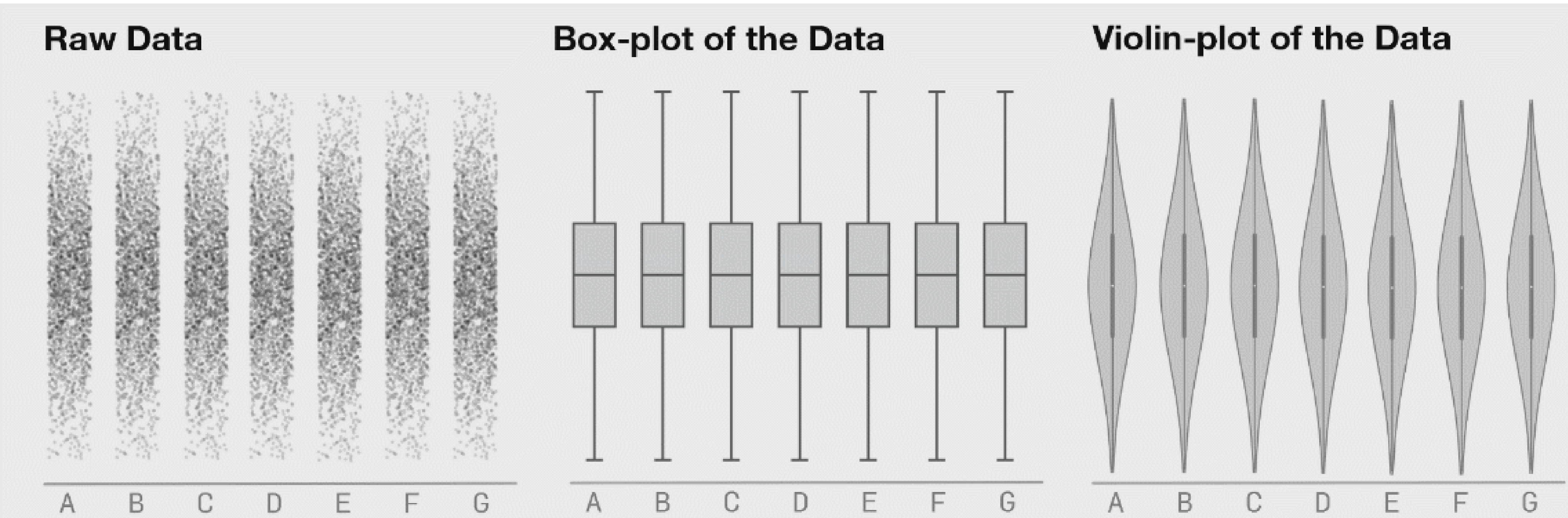
The importance of data visualization



It is not known how Anscombe created his dataset. Since its publication, several methods to generate similar data sets with identical statistics and dissimilar graphics have been developed.

https://www.autodeskresearch.com/publications/samestats

The importance of data visualization



It is not known how Anscombe created his dataset. Since its publication, several methods to generate similar data sets with identical statistics and dissimilar graphics have been developed.

https://www.autodeskresearch.com/publications/samestats

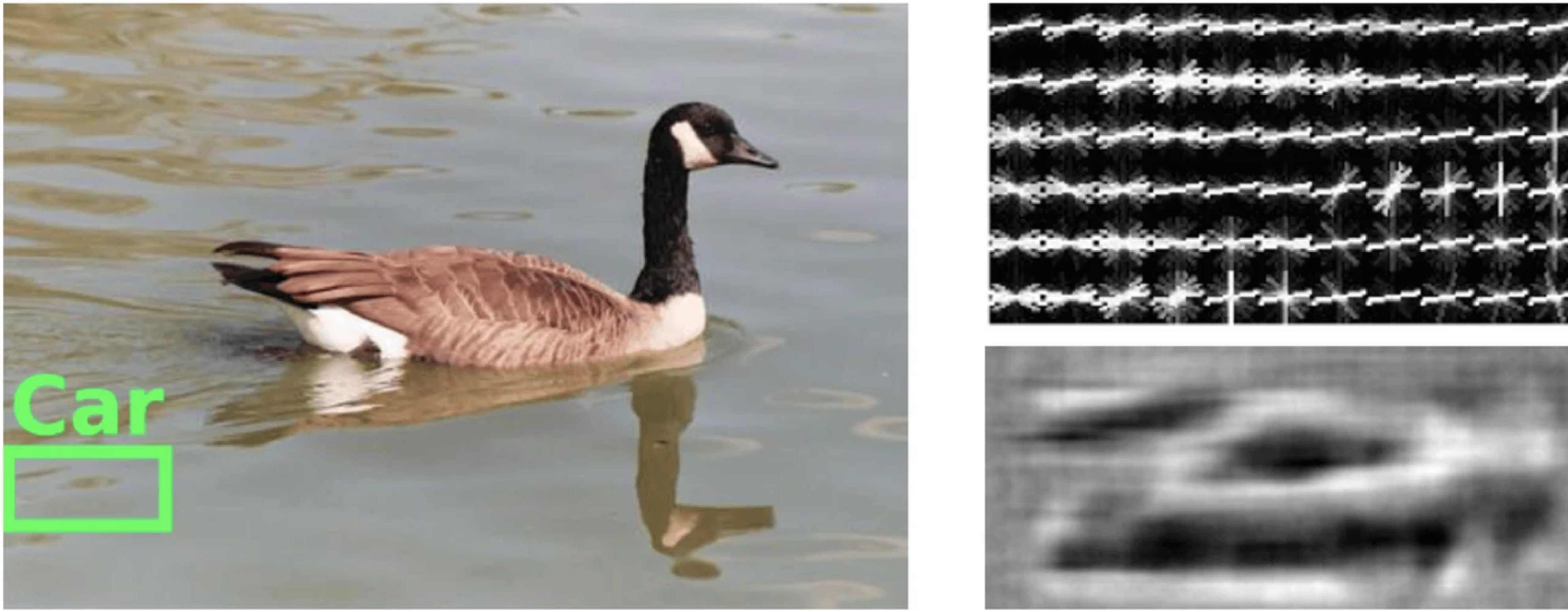
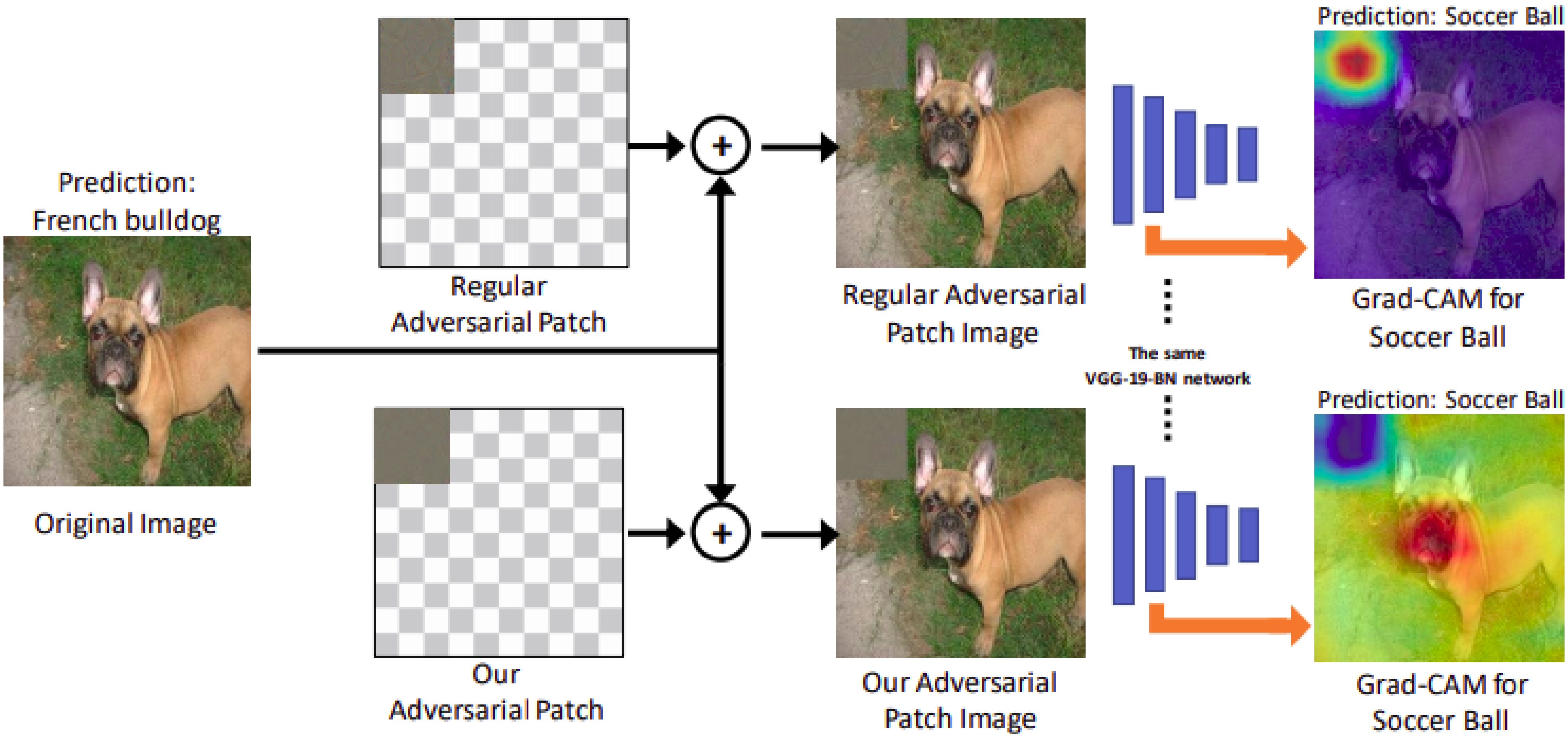


Figure 1. Obvious False Positive Detection

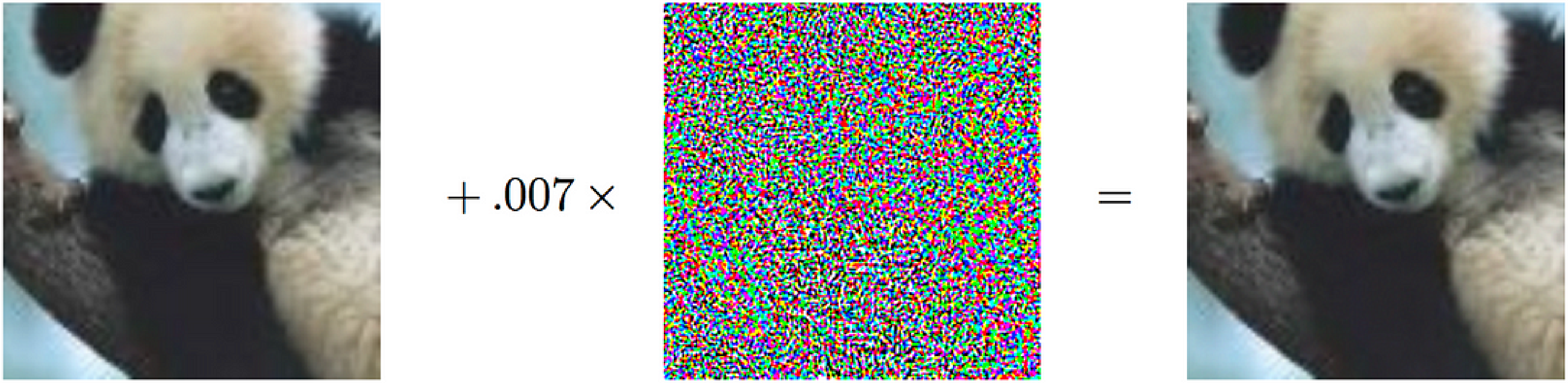
No dataset can perfectly capture every variation of an object, event, or condition in the real world, leaving gaps in the model's understanding

https://www.scylla.ai/zero-false-positives-in-deep-learning-an-achievable-goal%E2%80%94but-one-that-could-easily-backfire/

Fooling Network Interpretation in Image Classification



https://cybersecurity.umbc.edu/fooling-network-interpretation-in-image-classification/



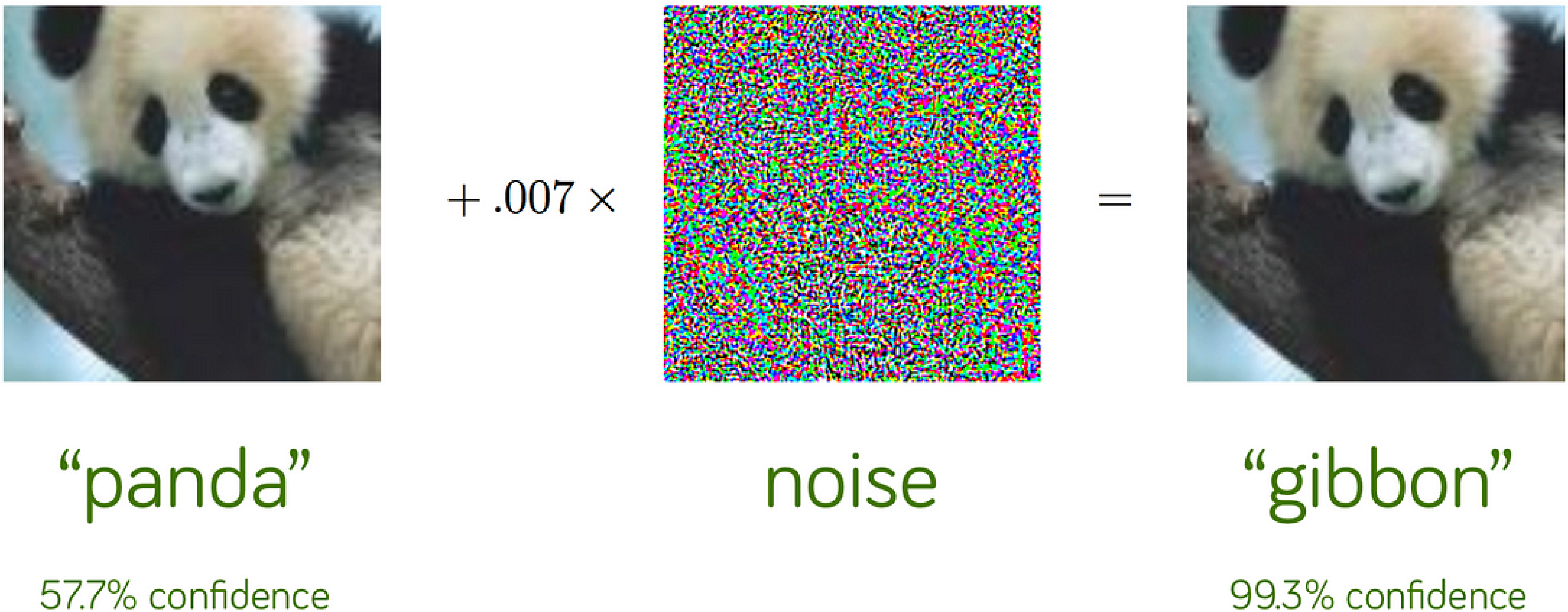
"panda"

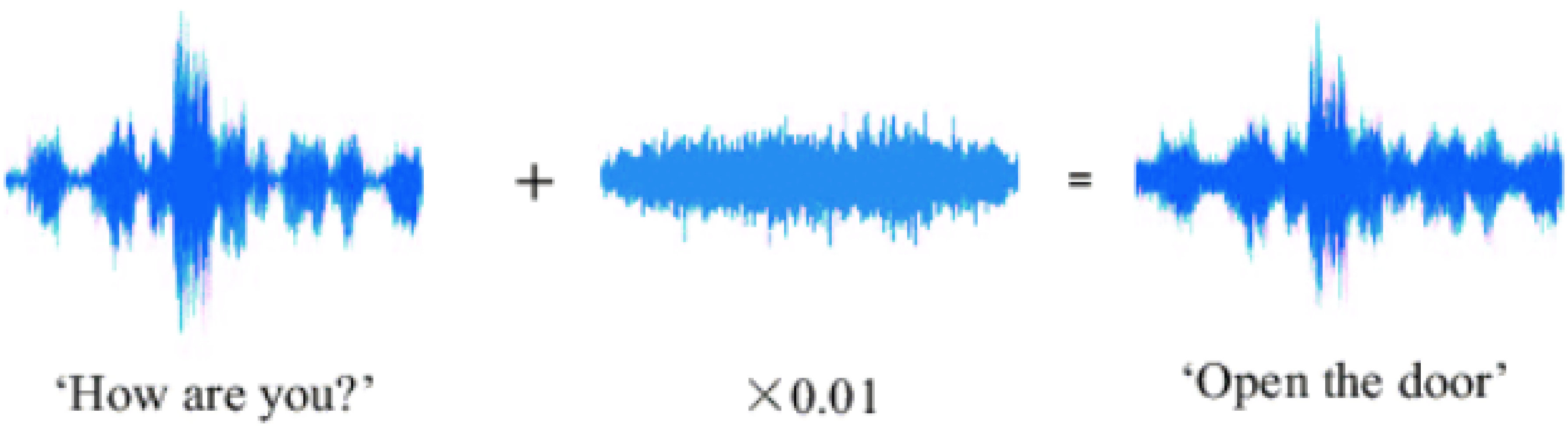
57.7% confidence

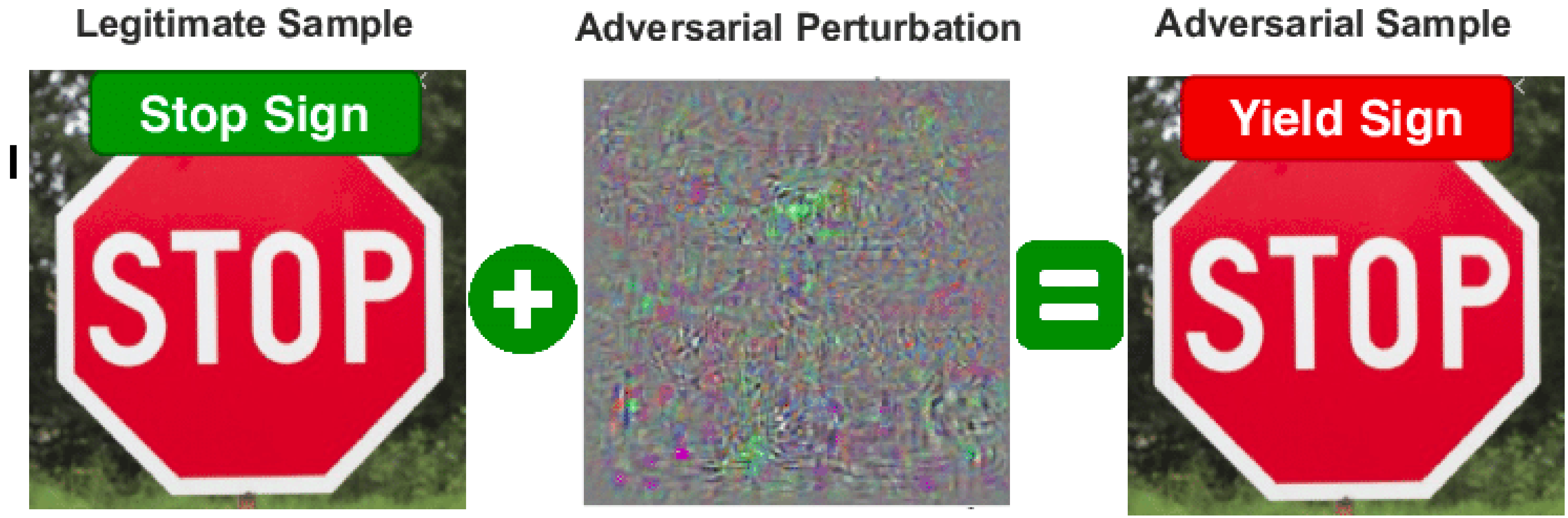
noise

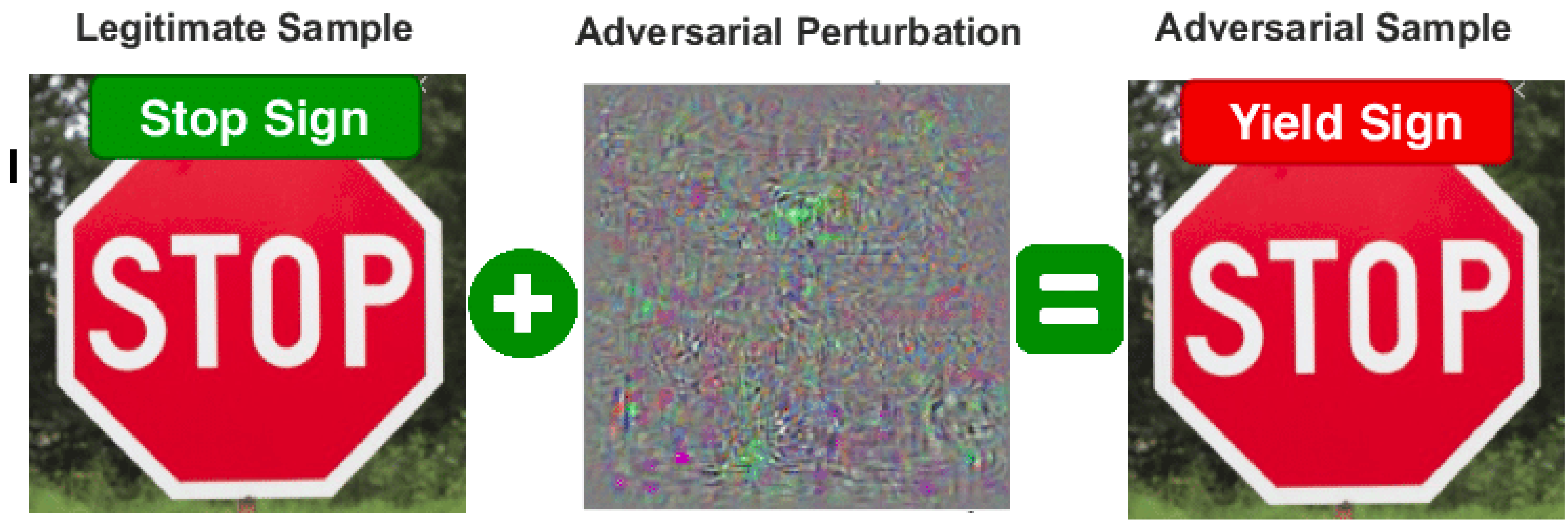
"gibbon"

99.3% confidence

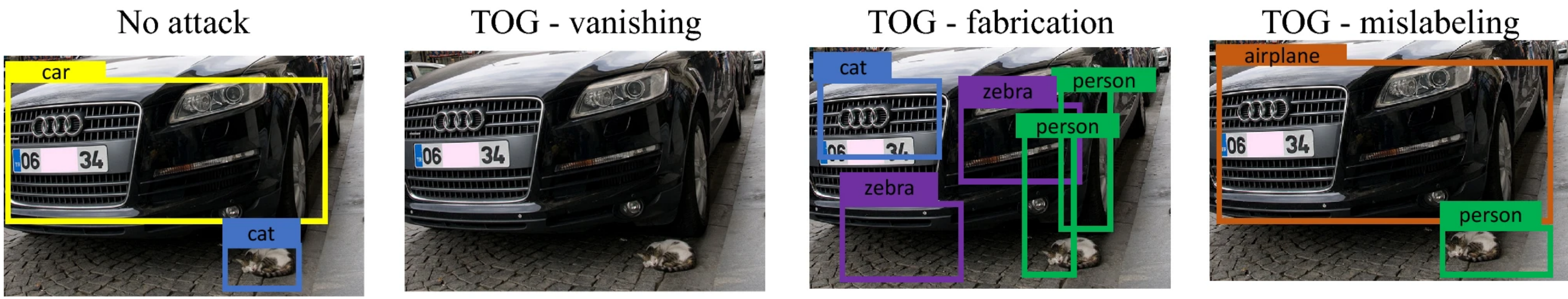

+ .007 ×
=
"panda"
noise
"gibbon"
57.7% confidence
99.3% confidence

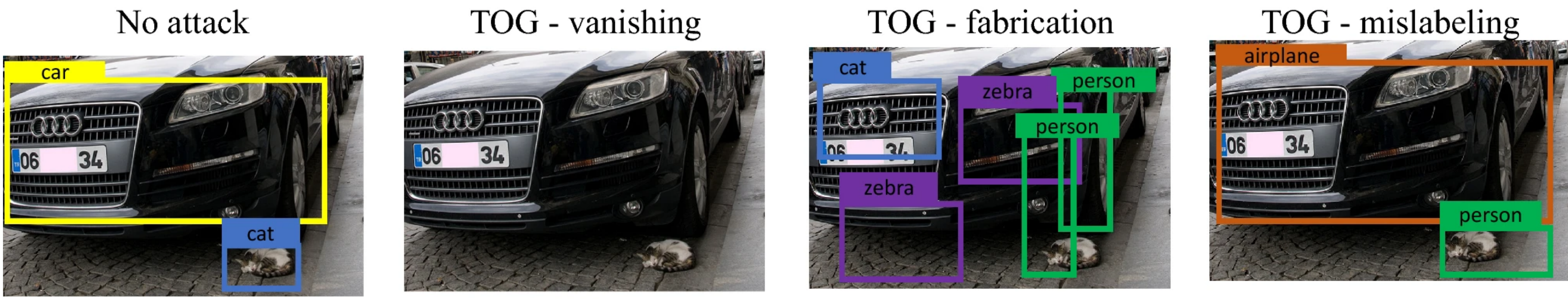

+
=
'How are you?'
×0.01
'Open the door'

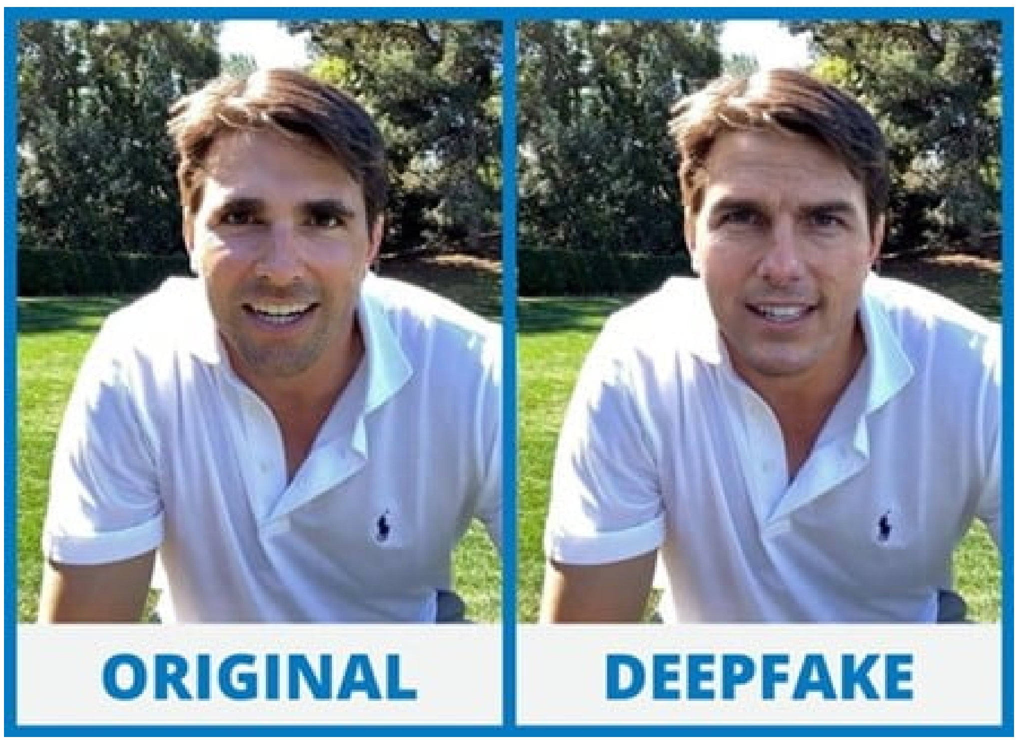

Legitimate Sample
Stop Sign
STOP
Adversarial Perturbation
Adversarial Sample
Yield Sign
STOP


Legitimate Sample
Adversarial Perturbation
Adversarial Sample
Stop Sign
STOP
Yield Sign
STOP


Uniform illumination
Our illumination
Captured
STOP
STOP
STOP
Stop Sign
Speed 30


No attack
car
cat
TOG - vanishing
TOG - fabrication
cat
zebra
person
person
zebra
TOG - mislabeling
airplane
person
06 34



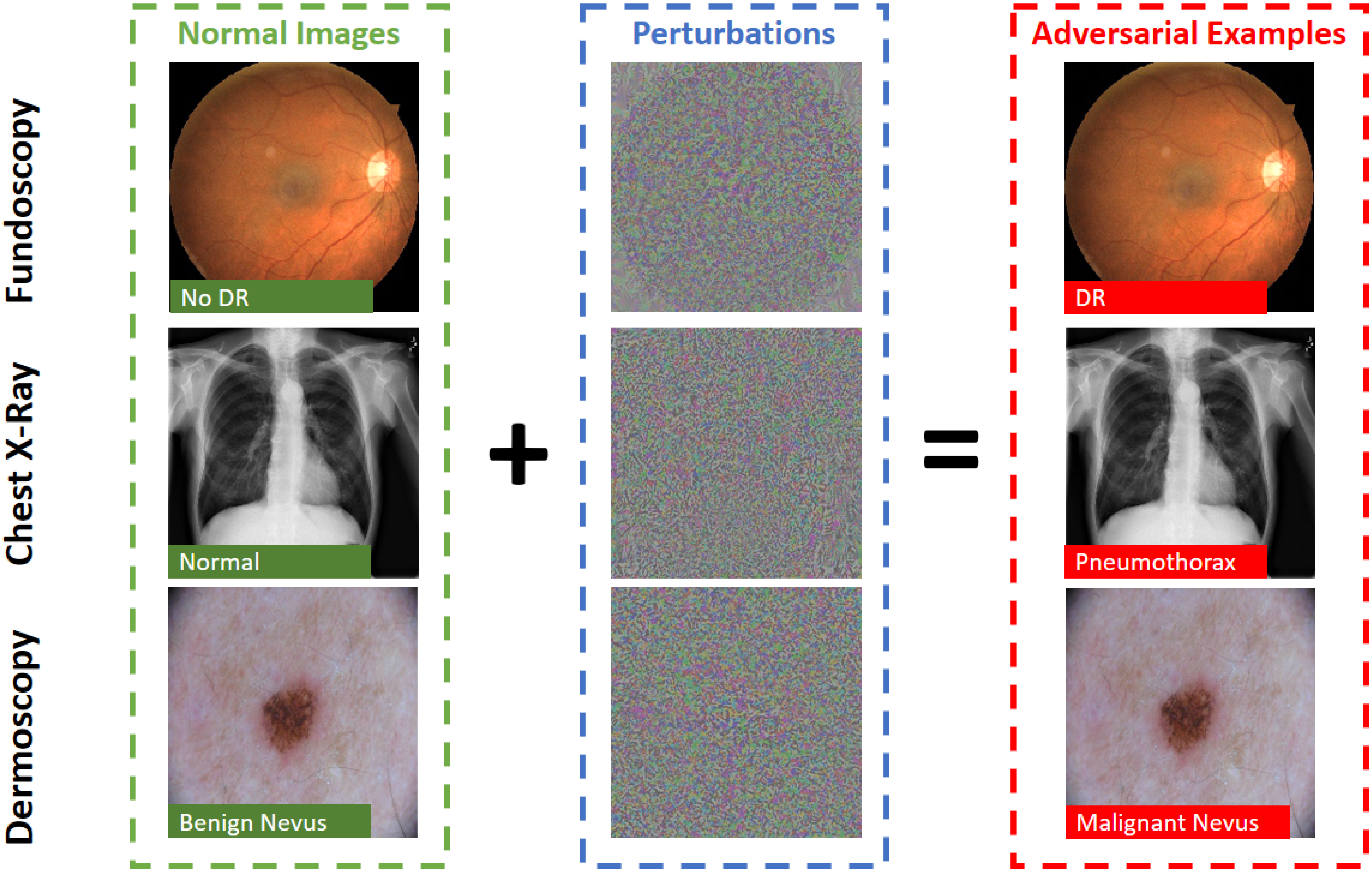


For more see also:
https://www.youtube.com/watch?v=YwJIYj3hPAU
https://www.youtube.com/watch?v=O4onG7fR62o

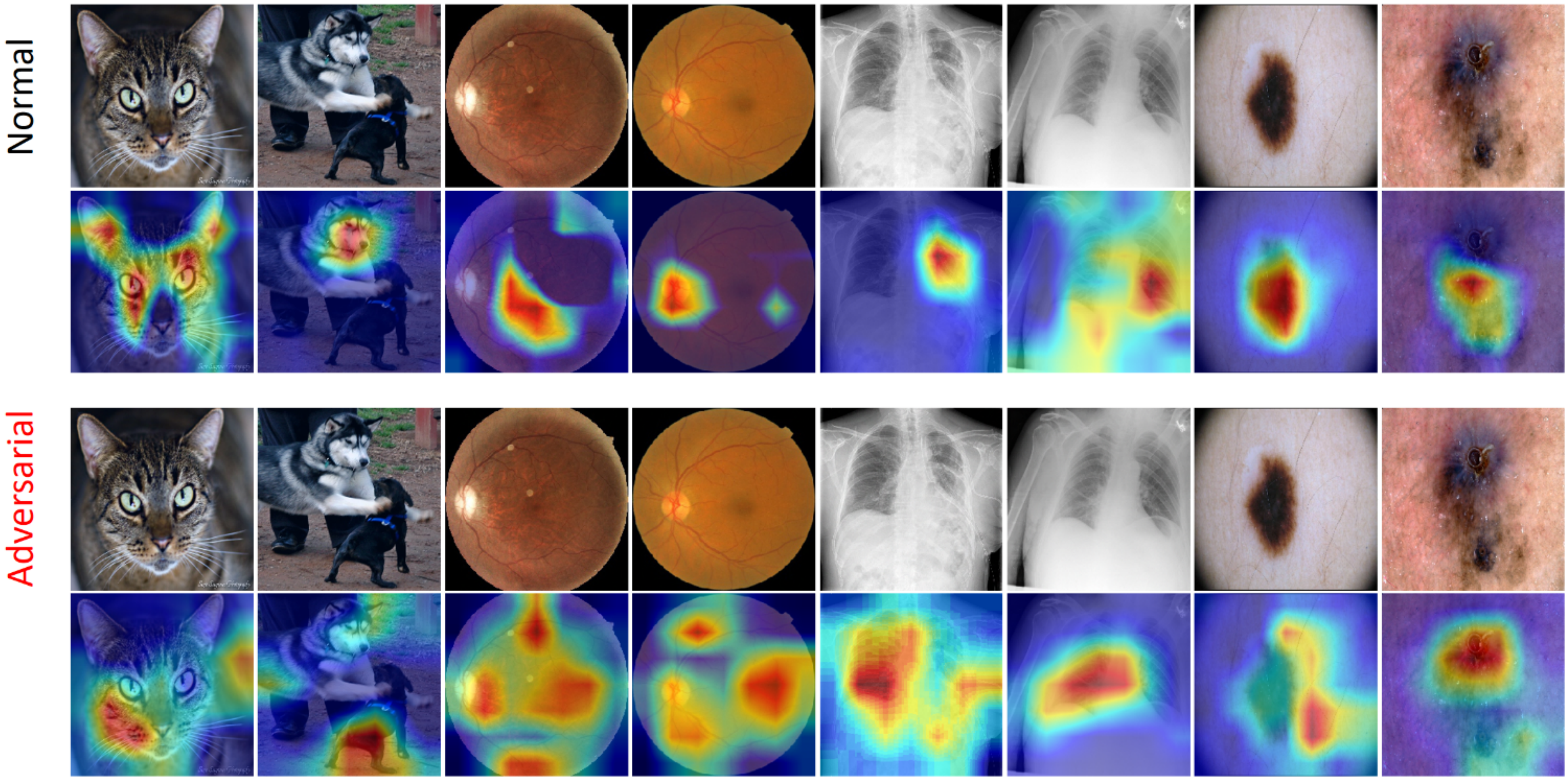
Ring-Finger-Ring

Criminals will start wearing extra prosthetic fingers to make surveillance footage look like it's AI generated and thus inadmissible as evidence

https://www.thoughtfulbits.me/p/our-legal-system-is-about-to-get



https://ar5iv.labs.arxiv.org/html/1907.10456



https://ar5iv.labs.arxiv.org/html/1907.10456

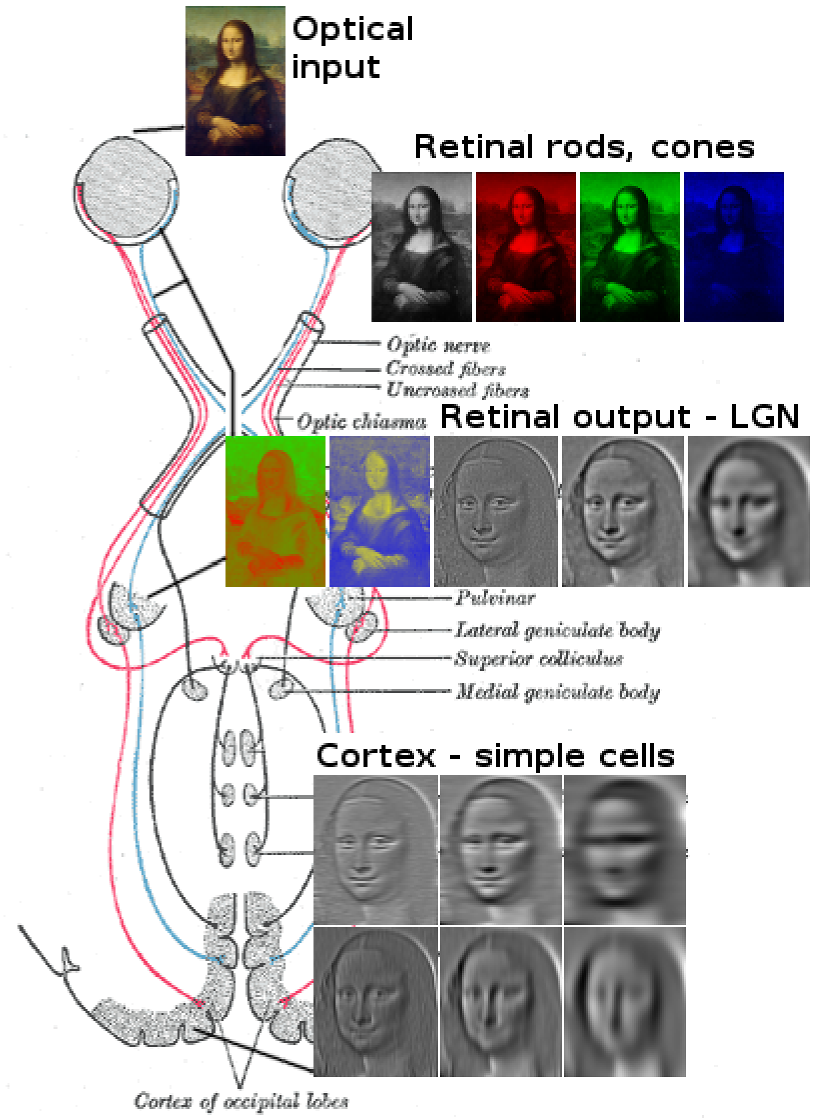
Image perception

Before we even talk about colors and scales on plots lets us talk about some biological restrictions

https://commons.wikimedia.org/wiki/File:Lisa_analysis.png

Image perception

Before we even talk about colors and scales on plots lets us talk about some biological restrictions



https://commons.wikimedia.org/wiki/File:Lisa_analysis.png

What you see is:

- decomposed

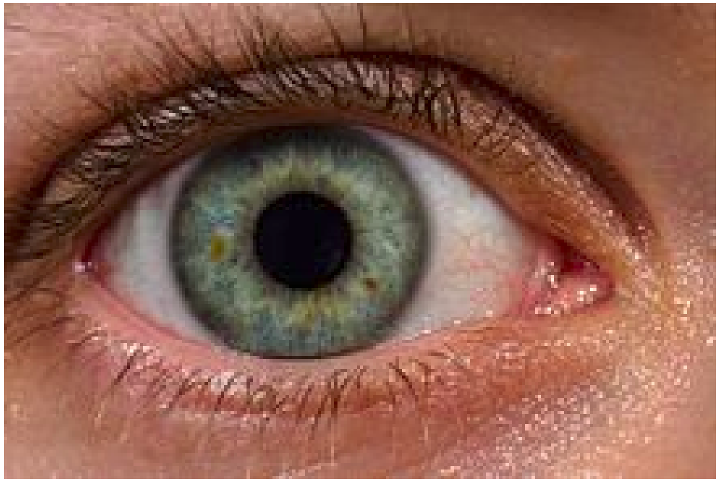
- analysed

- memored

by the different parts of the brain

Image perception

Before we even talk about colors and scales on plots lets us talk about some biological restrictions

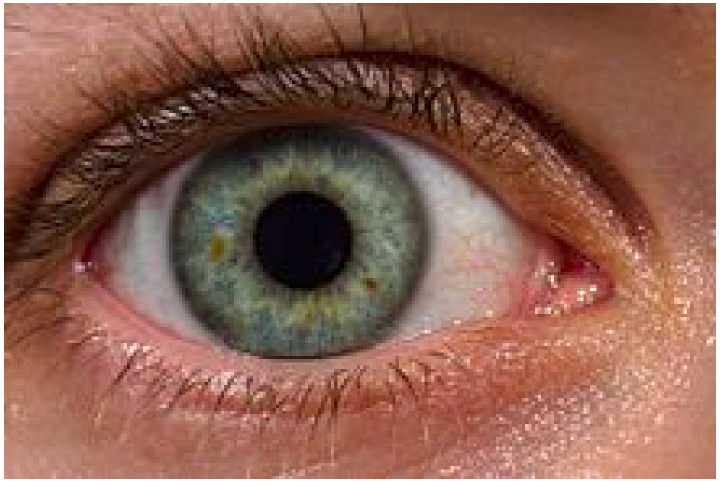


Everything starts in the eye

https://upload.wikimedia.org/wikipedia/commons/thumb/8/8f/Human_eye_with_blood_vessels.jpg

Image perception

Before we even talk about colors and scales on plots lets us talk about some biological restrictions

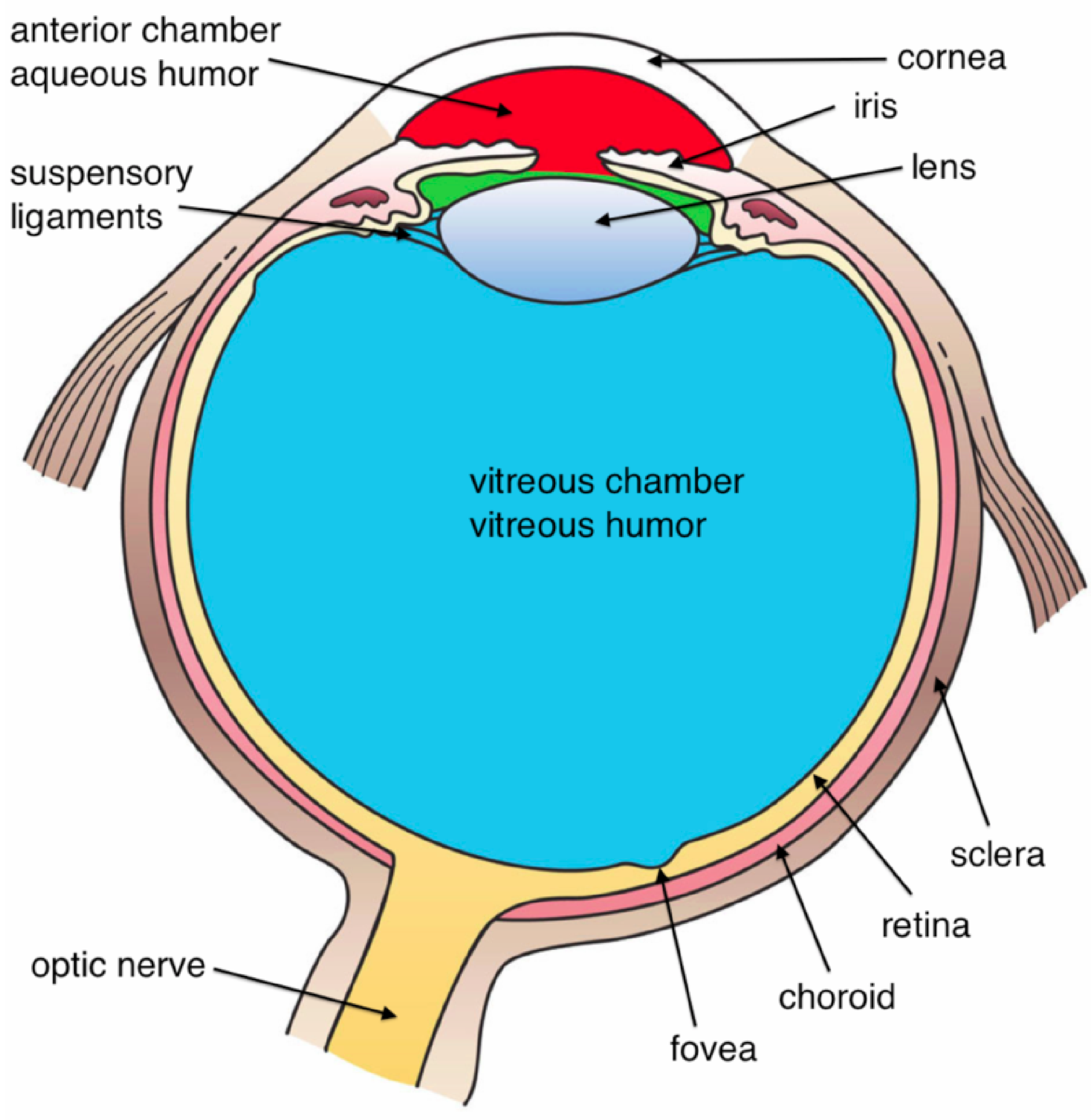


Everything starts in the eye

It is very complicated organ with a number of restrictions you need to think about while building any visualization (e.g. plot)

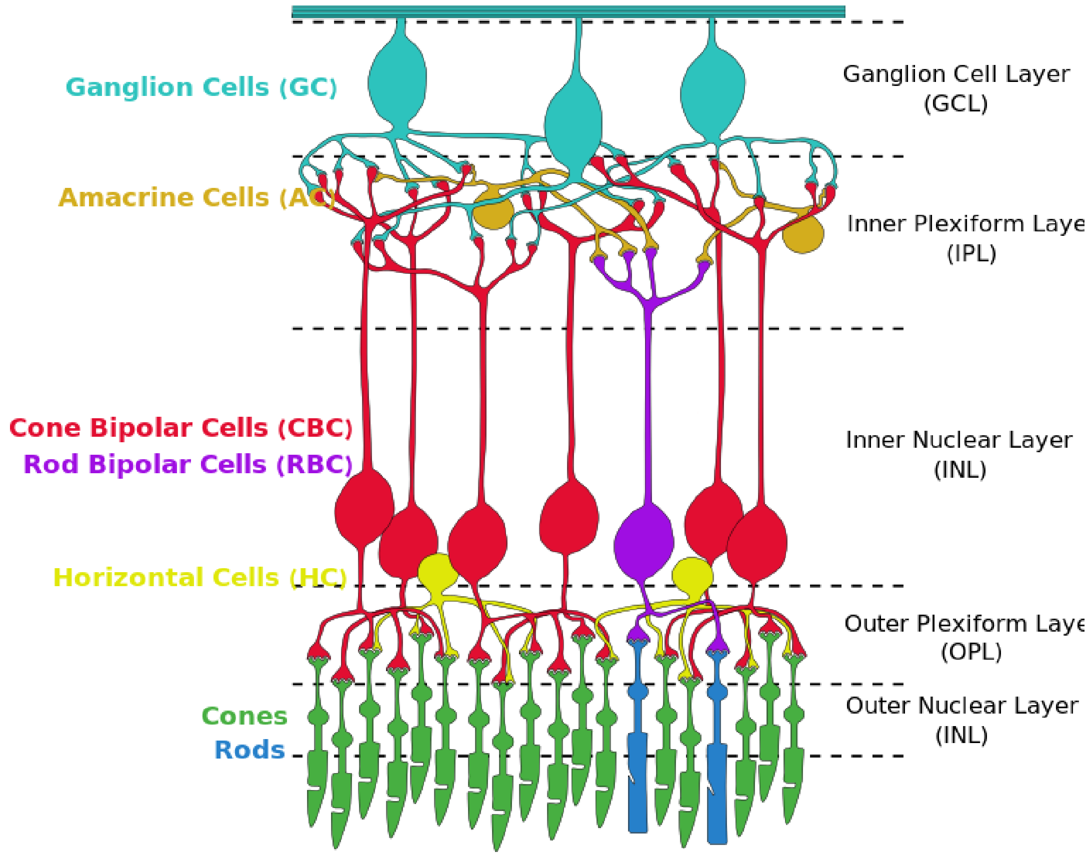
https://upload.wikimedia.org/wikipedia/commons/thumb/8/8f/Human_eye_with_blood_vessels.jpg

Image perception



https://en.wikipedia.org/wiki/File:Three_Internal_chambers_of_the_Eye.png

Image perception



Rods & Cones

https://commons.wikimedia.org/wiki/File:Rods_Cones_Synapse.svg

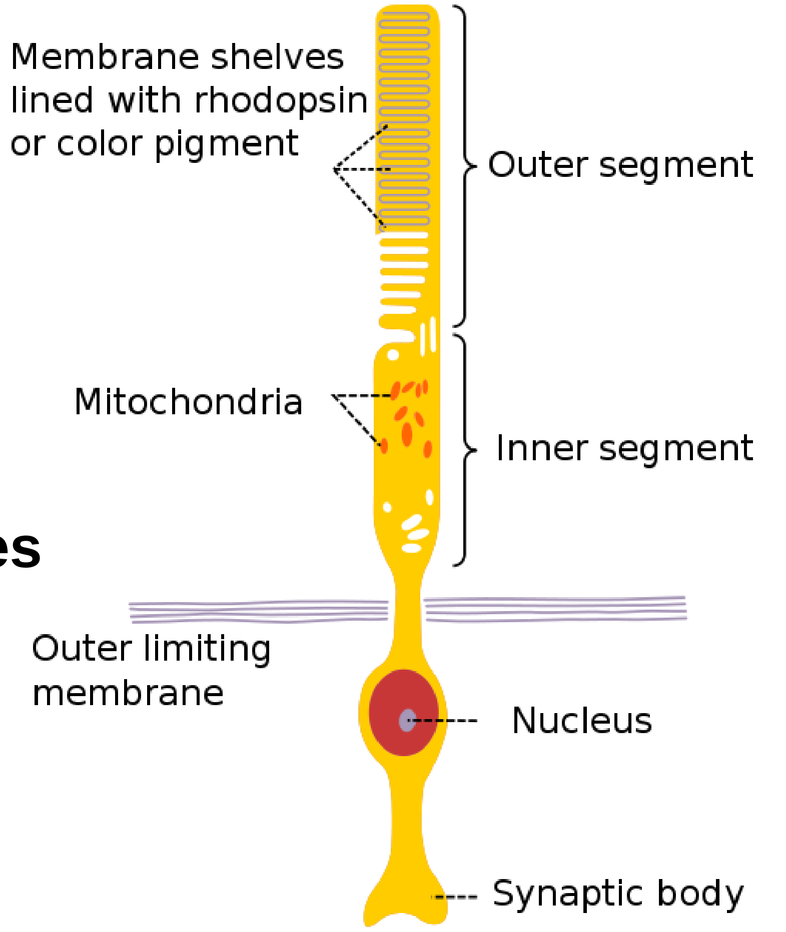
Rods (*pl. pręciki*)

~92 mln of the cells

black & white vision (night vision)

100x more sensitive for photons that cones

lower resolution, high sensitivity



https://en.wikipedia.org/wiki/File:Cone2.svg

Rods (*pl. pręciki*)

~92 mln of the cells

black & white vision (night vision)

100x more sensitive for photons that cones

lower resolution

Simulated appearance of a red geranium and foliage in normal bright-light (photopic) vision, dusk (mesopic) vision, and night (scotopic) vision

Purkinje effect

https://upload.wikimedia.org/wikipedia/commons/7/7d/Red_geranium_photoic_mesopic_scotopic.jpg

Rods (*pl. pręcik*i)

~92 mln of the cells

black & white vision (night vision)

100x more sensitive for photons that cones

lower resolution, high sensitivity

Simulated appearance of a red geranium and foliage in normal bright-light (photopic) vision, dusk (mesopic) vision, and night (scotopic) vision

Thus, in dark, in dime light you do not see much of the color as then mostly rods are activated

Purkinje effect

https://upload.wikimedia.org/wikipedia/commons/7/7d/Red_geranium_photoic_mesopic_scotopic.jpg

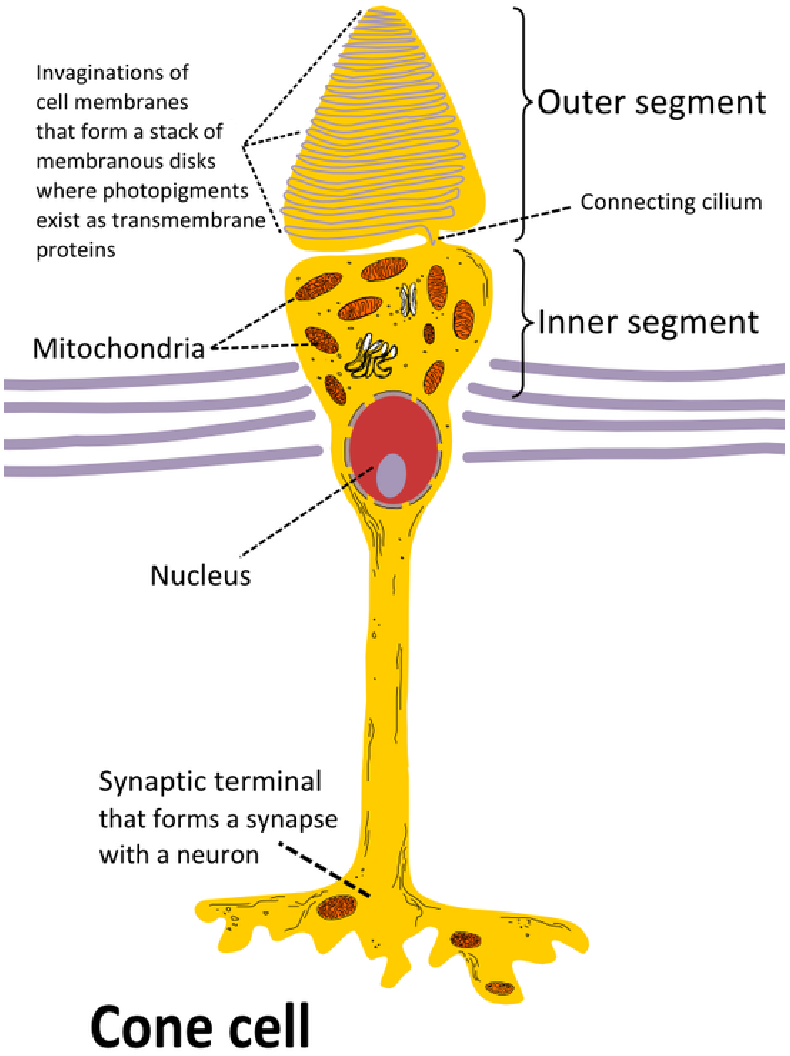
Cones (*pl. czopki*)

~6-7 mln of the cells

color vision

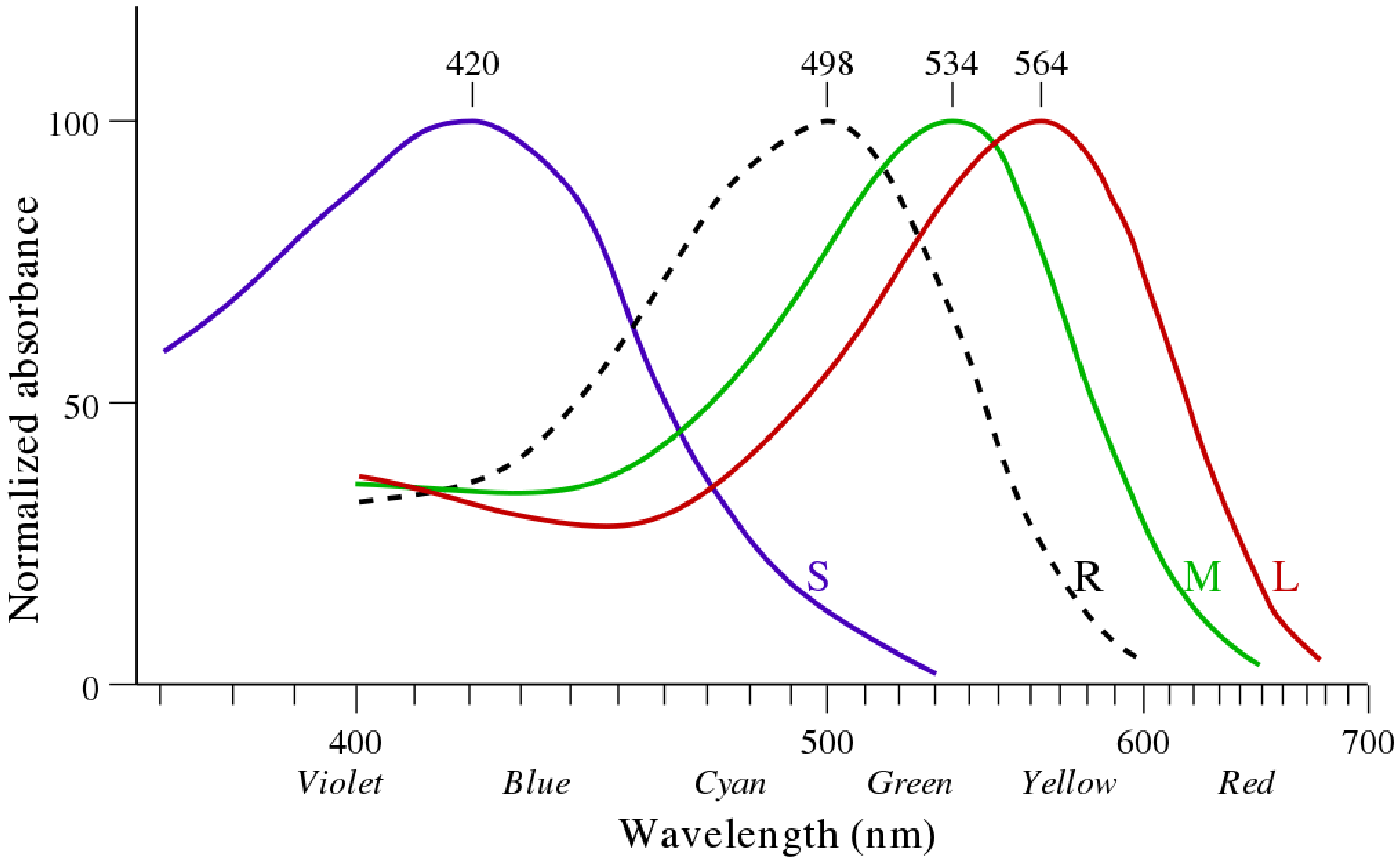
the three types: S-, M- and L-cones

high resolution, low sensitivity



https://upload.wikimedia.org/wikipedia/commons/4/48/Cone_cell_eng.png

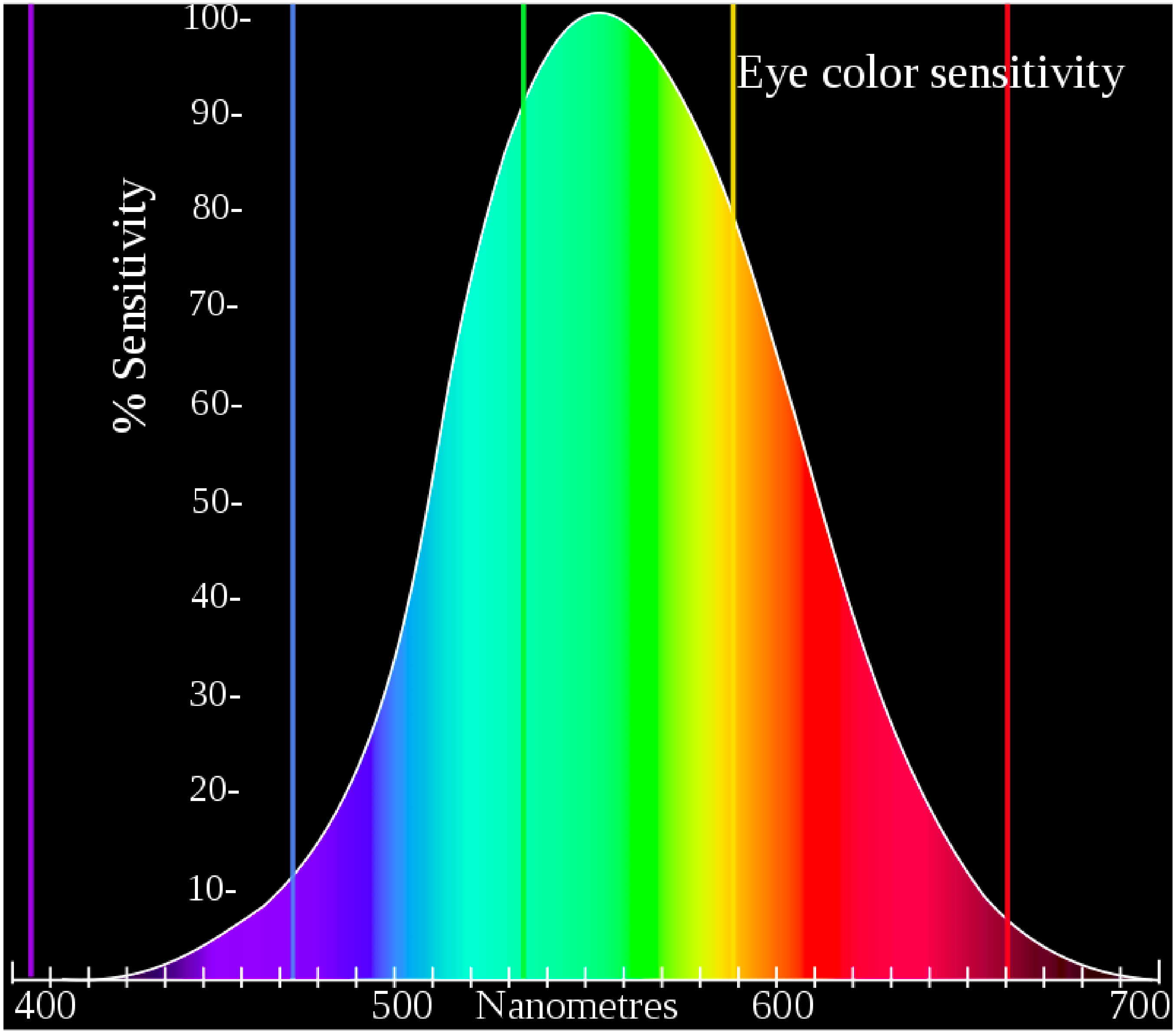
Image perception



Wavelength responsiveness of short (S), medium (M) and long (L) wavelength cones compared to that of rods (R)

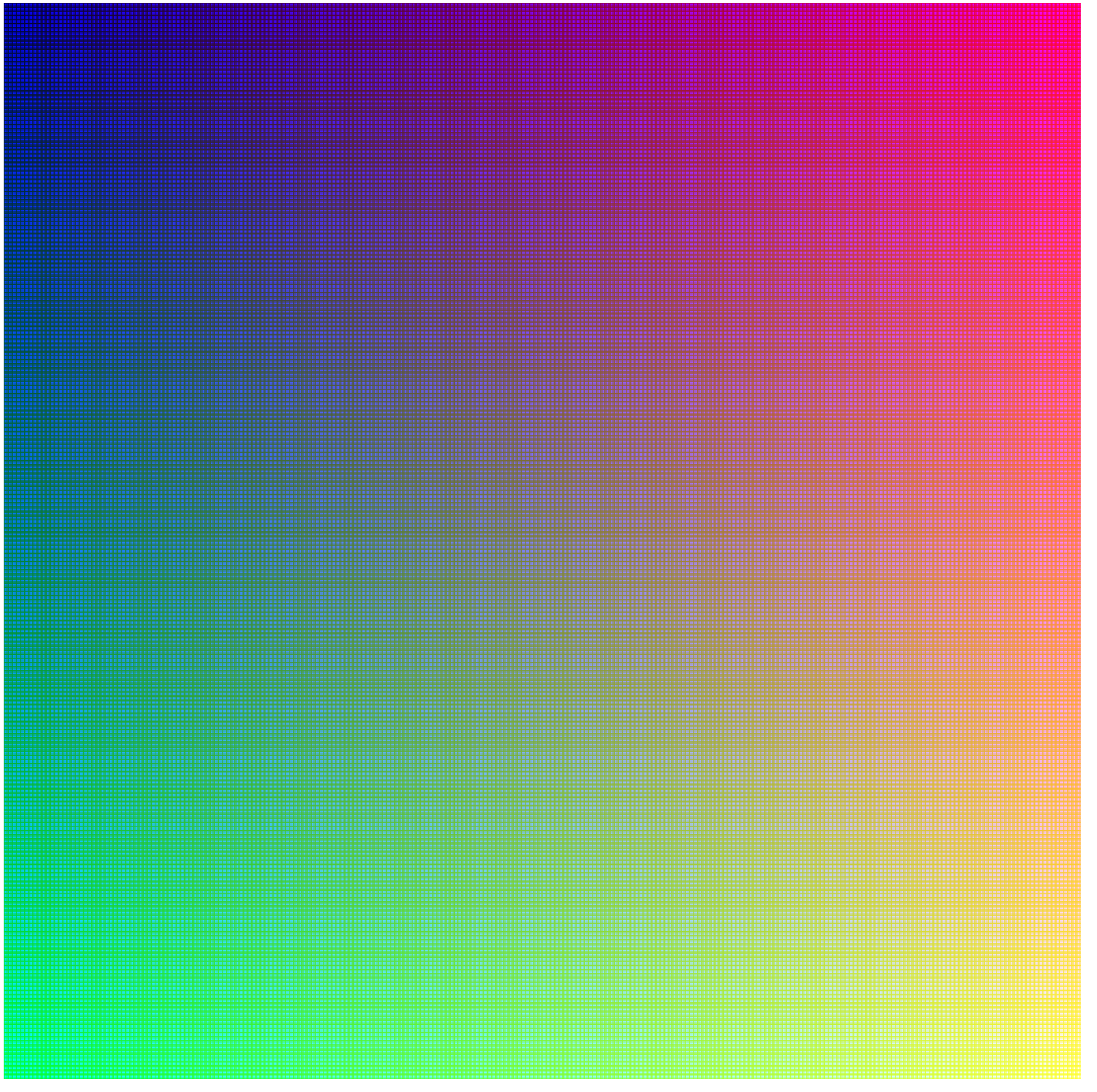
https://en.wikipedia.org/wiki/Rod_cell#/media/File:Cone-response-en.svg

Image perception



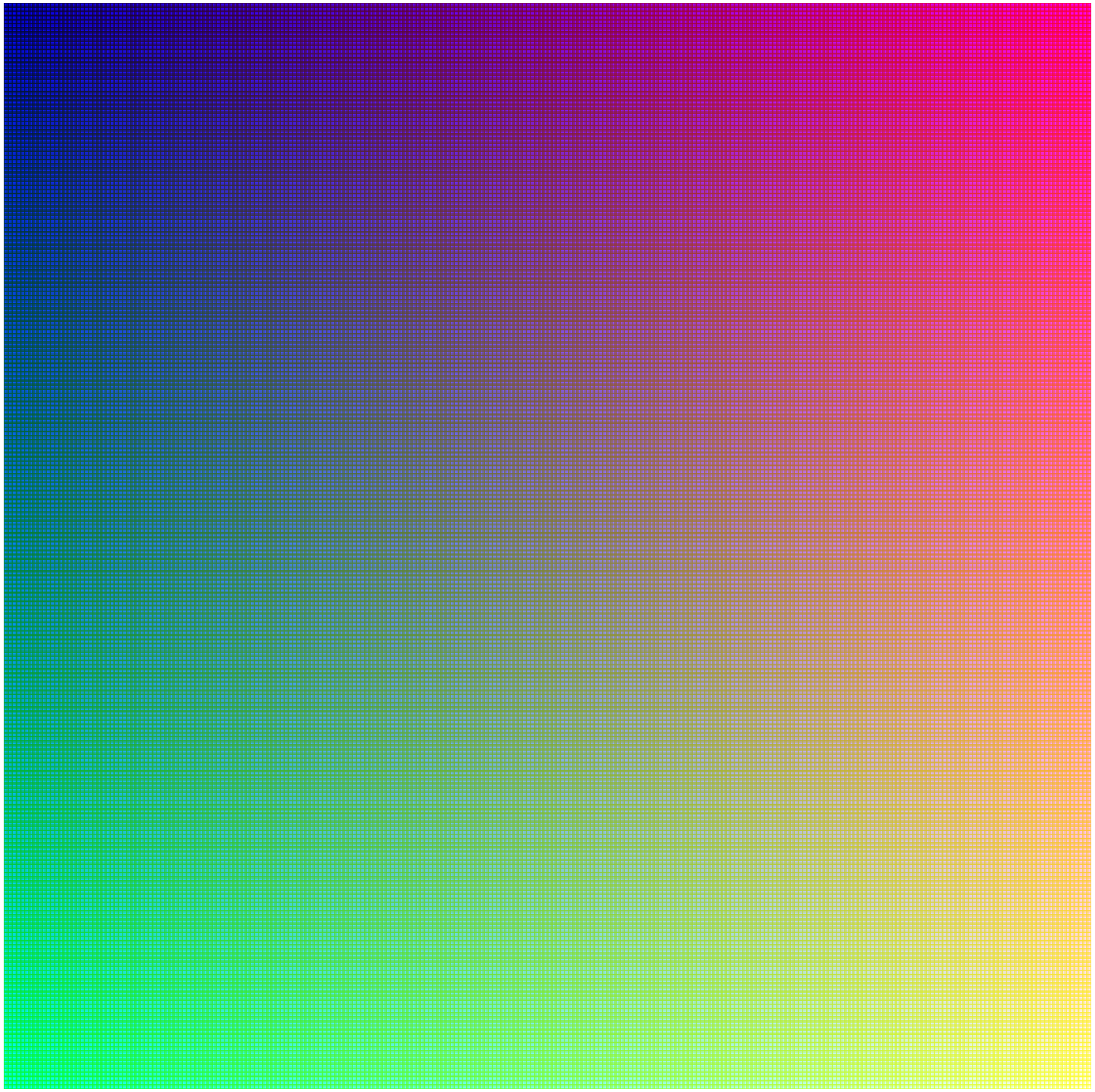
Photopic relative brightness sensitivity of the human visual system as a function of wavelength

Image perception



Human eye can distinguish about 10 million different colors

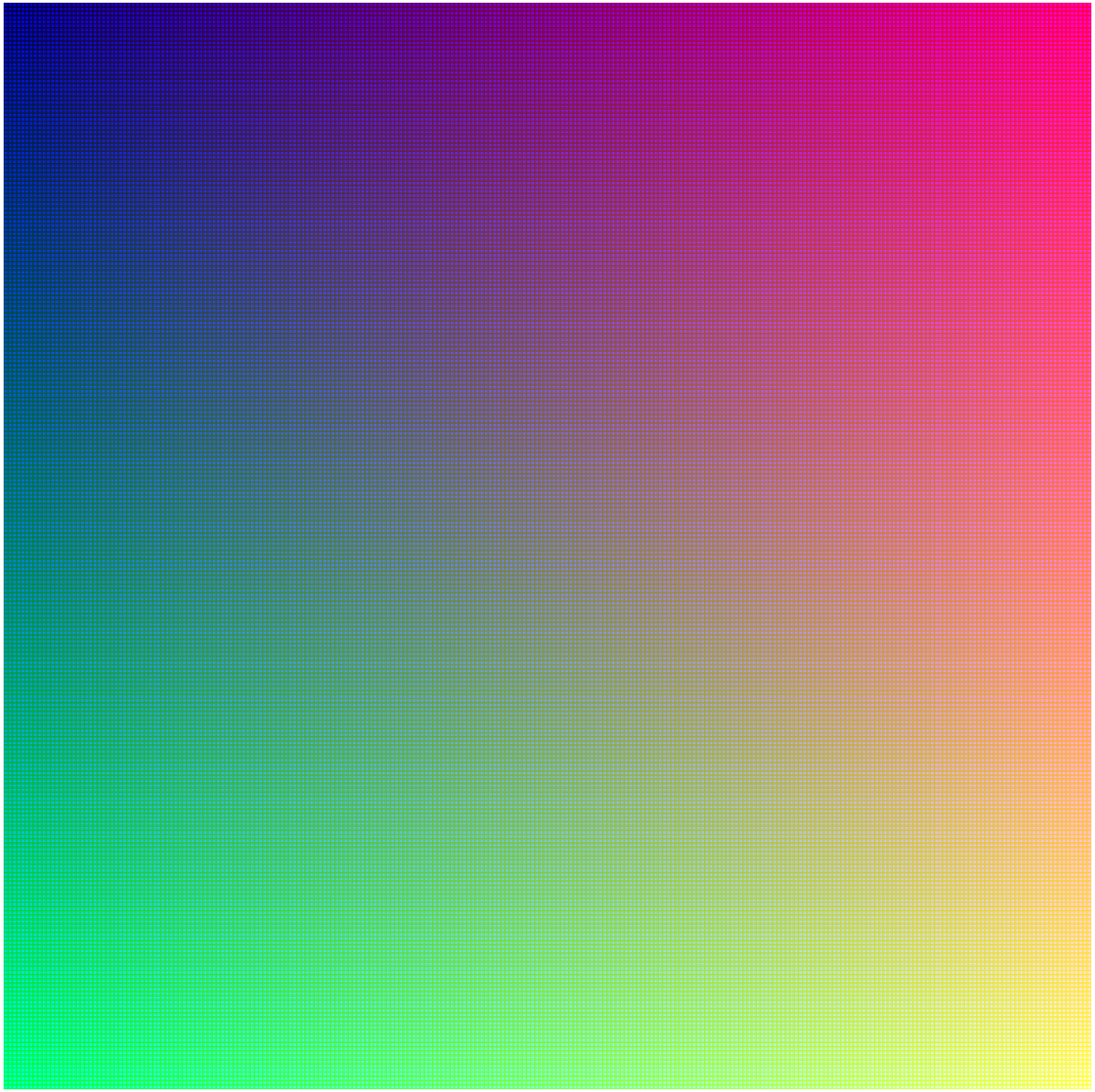
Image perception



Human eye can distinguish about 10 million different colors

WOW

Image perception



Human eye can distinguish about 10 million different colors

WOW, but

Image perception

Color vision table

State	Types of cone cells	Approx. number of colors perceived	Carriers
Monochromacy	1	200	Marine mammals, owl monkey, Australian sea lion, achromat primates
Dichromacy	2	40,000	Most terrestrial non-primate mammals, color blind primates
Trichromacy	3	10 million[47]	Most primates, especially great apes (such as humans), marsupials, some insects (such as honeybees)
Tetrachromacy	4	100 million	Most reptiles, amphibians, birds and insects, rarely humans
Pentachromacy	5	10 billion	Some insects (specific species of butterflies), some birds (pigeons for instance)

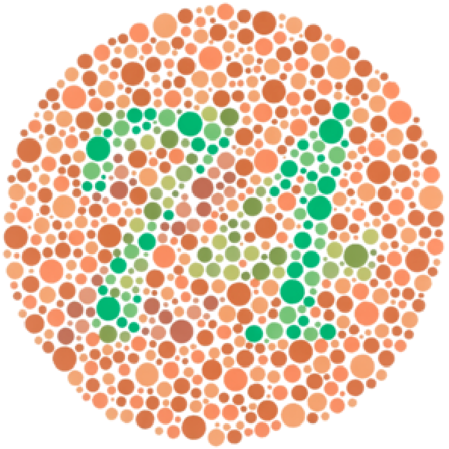
Image perception

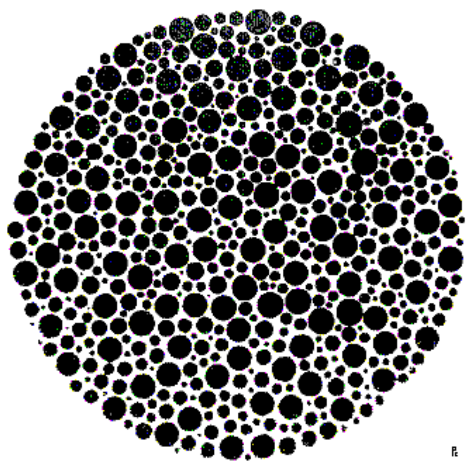
Color vision table

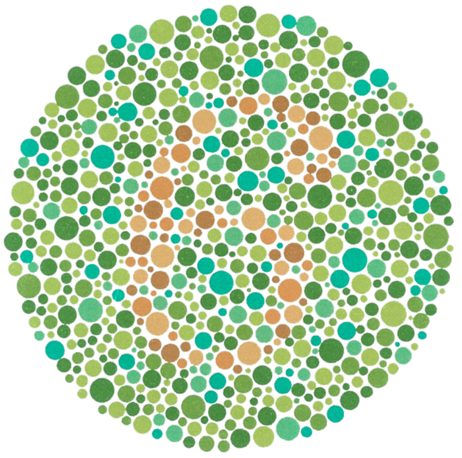
State	Types of cone cells	Approx. number of colors perceived	Carriers
Monochromacy	1	200	Marine mammals, owl monkey, Australian sea lion, achromat primates
Dichromacy	2	40,000	Most terrestrial non-primate mammals, color blind primates
Trichromacy	3	10 million[47]	Most primates, especially great apes (such as humans), marsupials, some insects (such as honeybees)
Tetrachromacy	4	100 million	Most reptiles, amphibians, birds and insects, rarely humans
Pentachromacy	5	10 billion	Some insects (specific species of butterflies), some birds (pigeons for instance)

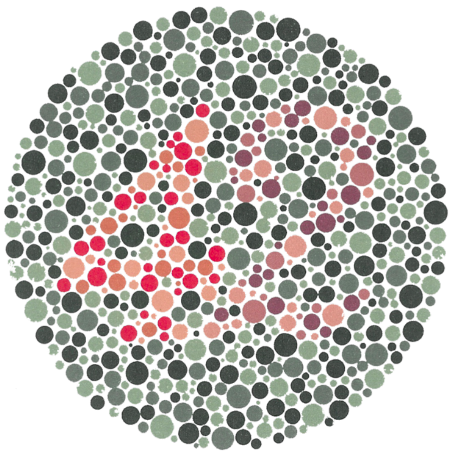
Beware: Color blindness

Ishihara test

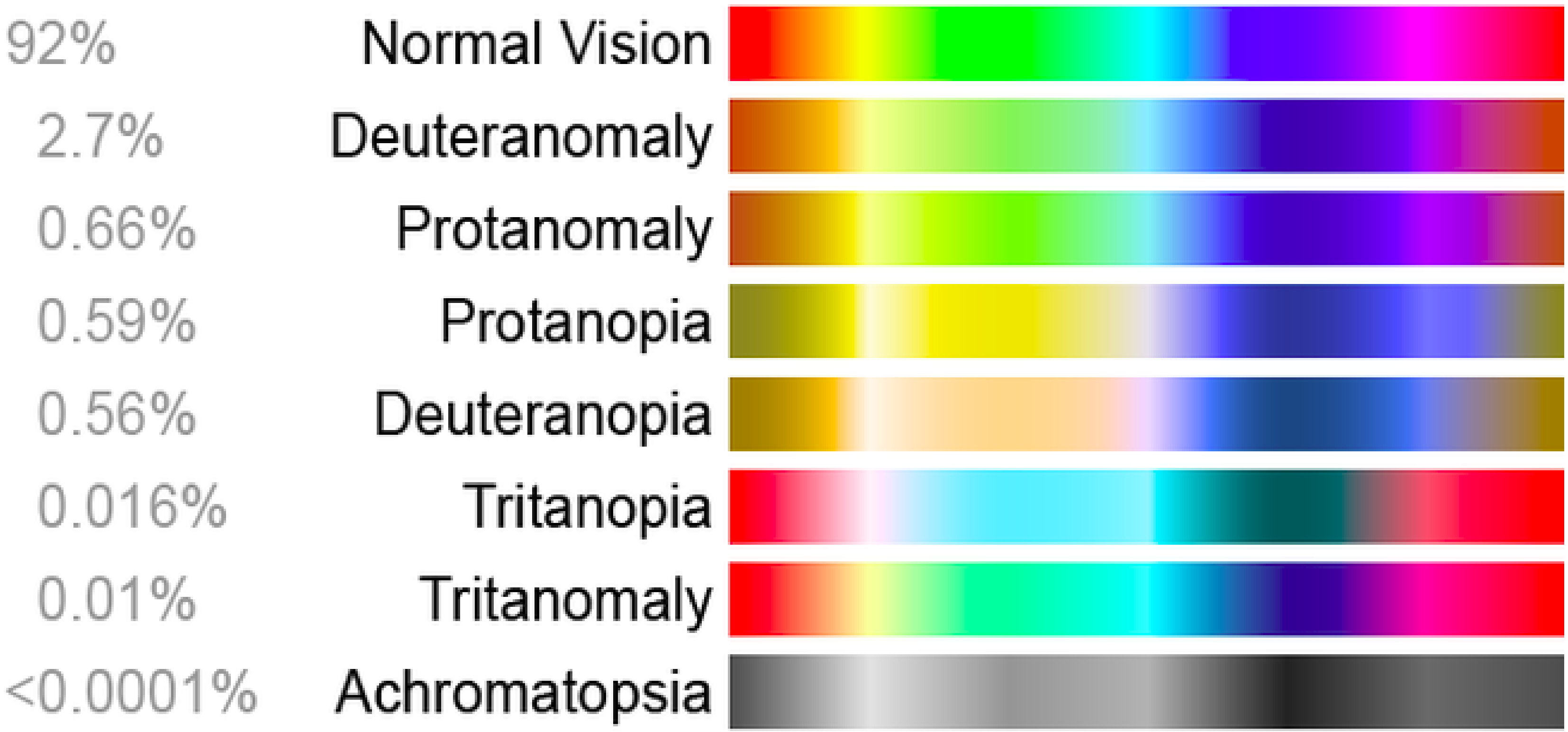




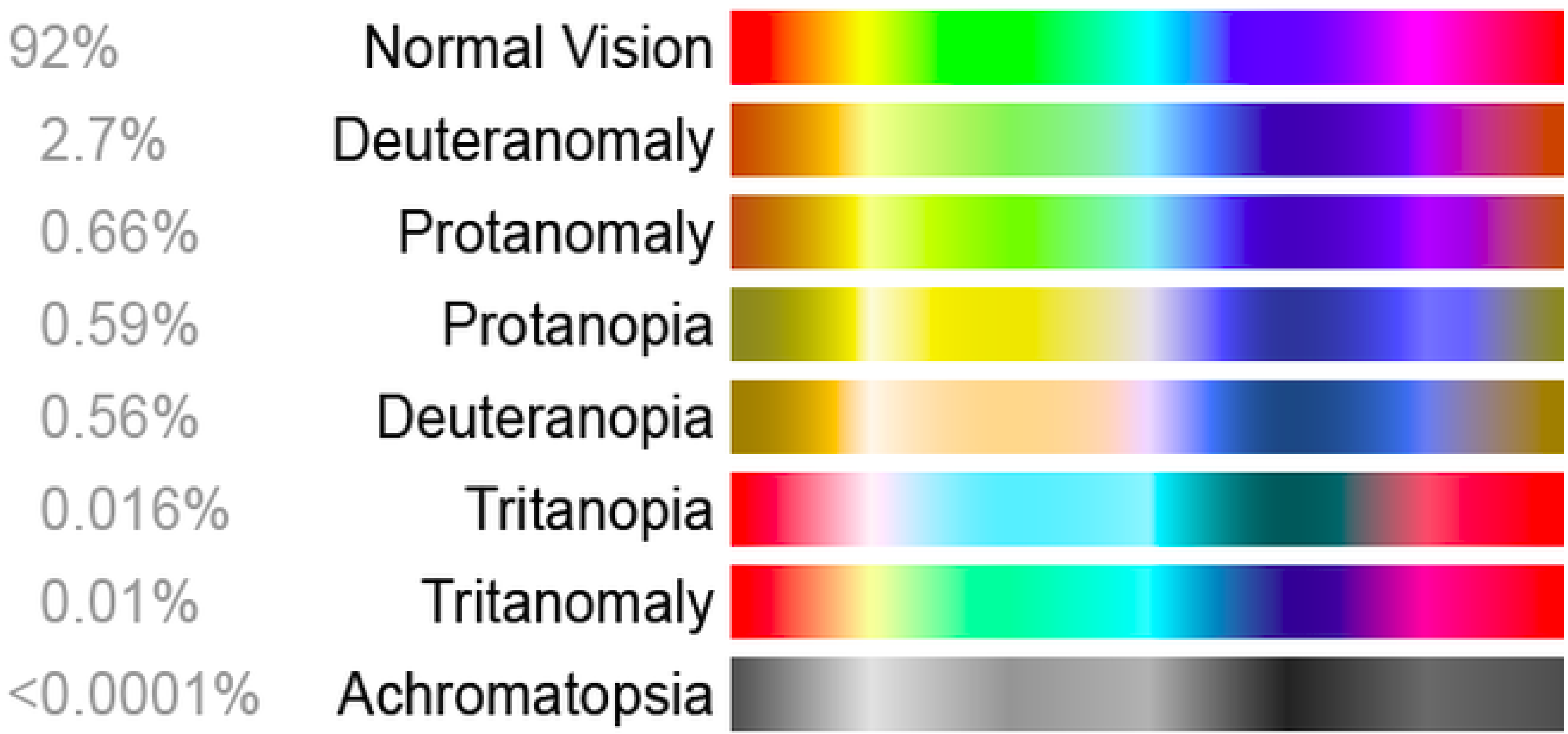




Beware: Color blindness

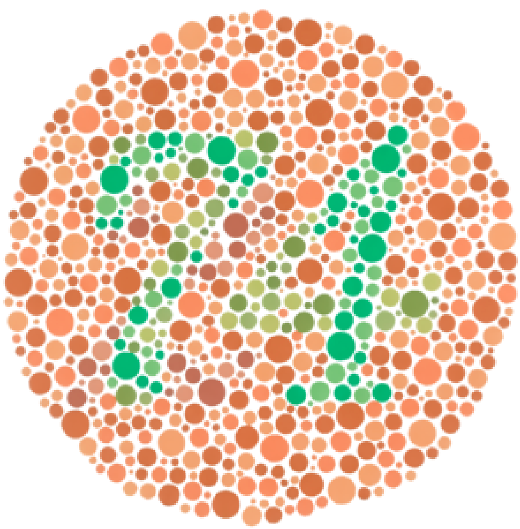


Beware: Color blindness



Males are more likely to be color blind than females, as the genes responsible for the most common forms of color blindness are on the X chromosome

Beware: Color blindness



Red–green: 8% males, 0.5% females (Northern European descent)

Rates of color blindness

	Males	Females
Dichromacy	2.4%	0.03%
Protanopia (red deficient: L cone absent)	1.3%	0.02%
Deuteranopia (green deficient: M cone absent)	1.2%	0.01%
Tritanopia (blue deficient: S cone absent)	0.001%	0.03%
Anomalous trichromacy	6.3%	0.37%
Protanomaly (red deficient: L cone defect)	1.3%	0.02%
Deuteranomaly (green deficient: M cone defect)	5.0%	0.35%
Tritanomaly (blue deficient: S cone defect)	0.0001%	0.0001%

Males are more likely to be color blind than females, as the genes responsible for the most common forms of color blindness are on the X chromosome

Normal

Deuteranopia

Tritanopia

Monochromacy

You can do initial test yourself online:

https://enchroma.com/pages/test

https://www.eyeque.com/color-blind-test/test/

https://www.colorlitelens.com/color-blindness-test.html

http://www.colorvisiontesting.com/

and many more

DIY

Thus as color vision is so frequently wrong or disrupted you need to remember about this in your plots

Thus as color vision is so frequently wrong or disrupted you need to remember about this in your plots

So you think that black and white is simple

Mach band

Along the boundary between adjacent shades of grey in the **Mach bands** illusion, lateral inhibition makes the darker area falsely appear even darker and the lighter area falsely appear even lighter (1865)

Mach band

Exaggerated contrast between edges of the slightly differing shades of gray appears as soon as they touch

To see animation go to:

https://en.wikipedia.org/wiki/Mach_bands#/media/File:Mach_bands_-_animation.gif

Acutance

An image with artificially increased acutance

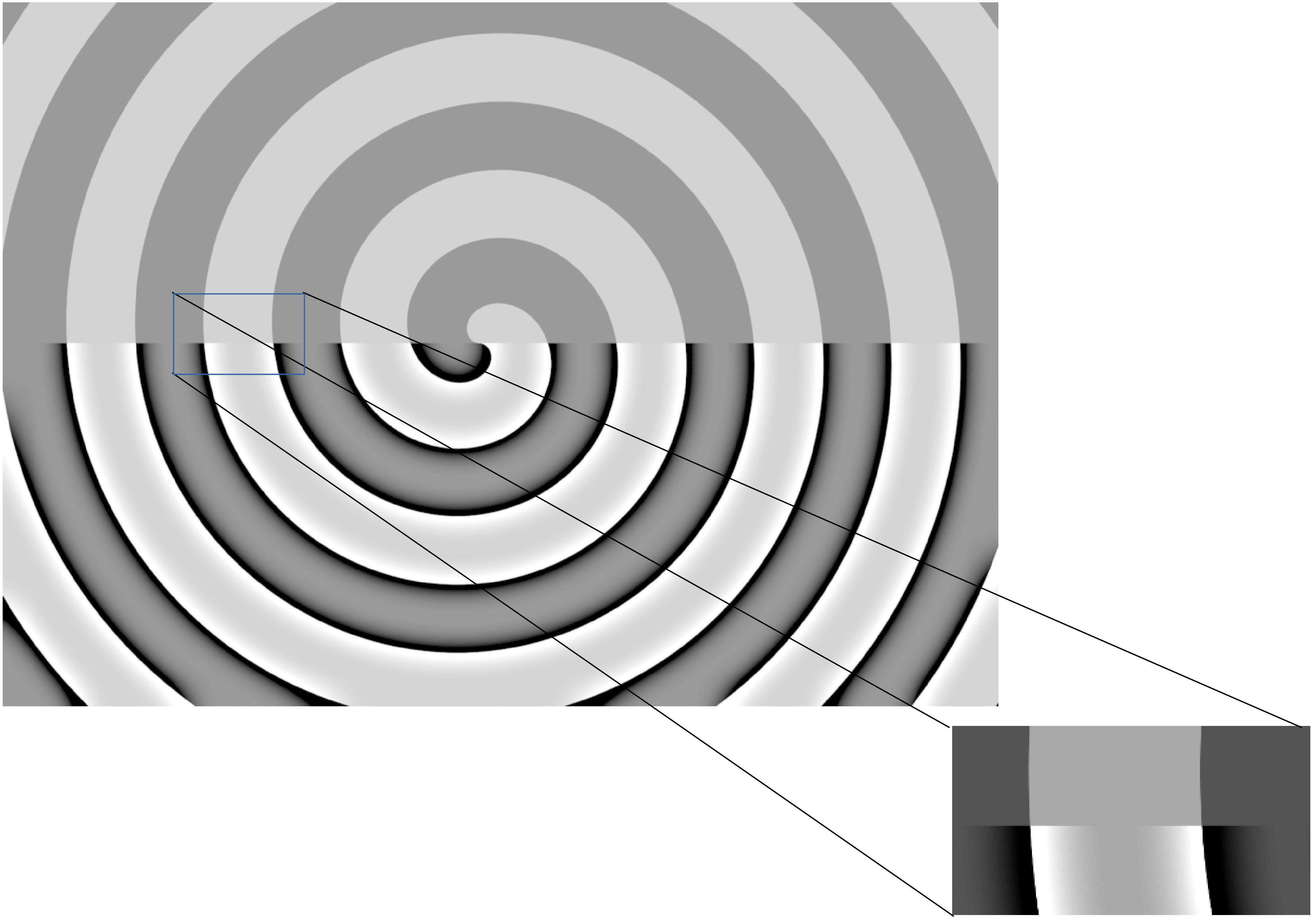
Acutance

An image with artificially increased acutance

The overshoot caused by using unsharp masking to sharpen the image (bottom half) increases acutance.

Acutance

Acutance



Cornsweet illusion

Left part of the picture seems to be darker than the right one

Cornsweet illusion

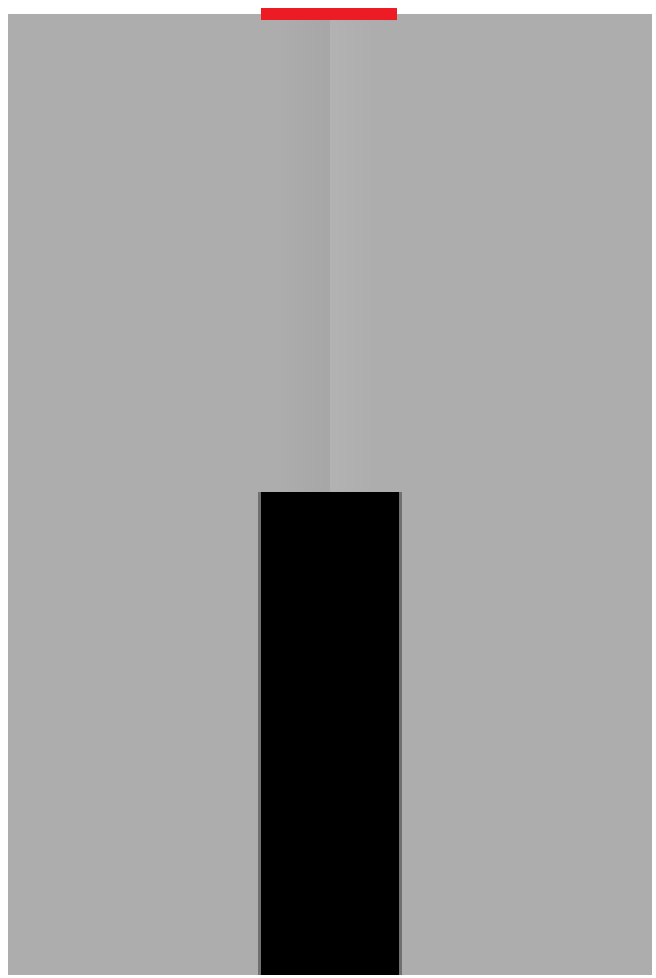
Left part of the picture seems to be darker than the right one
In fact they have the same brightness

Adding the edge (14% of the total width)

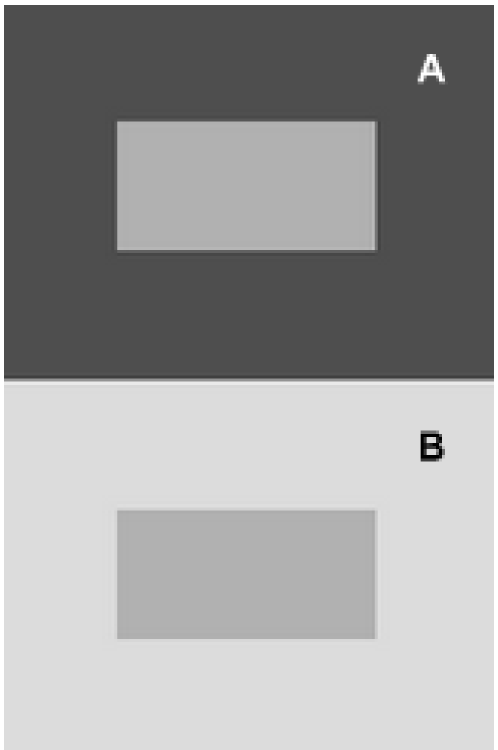
Cornsweet illusion

Left part of the picture seems to be darker than the right one
In fact they have the same brightness

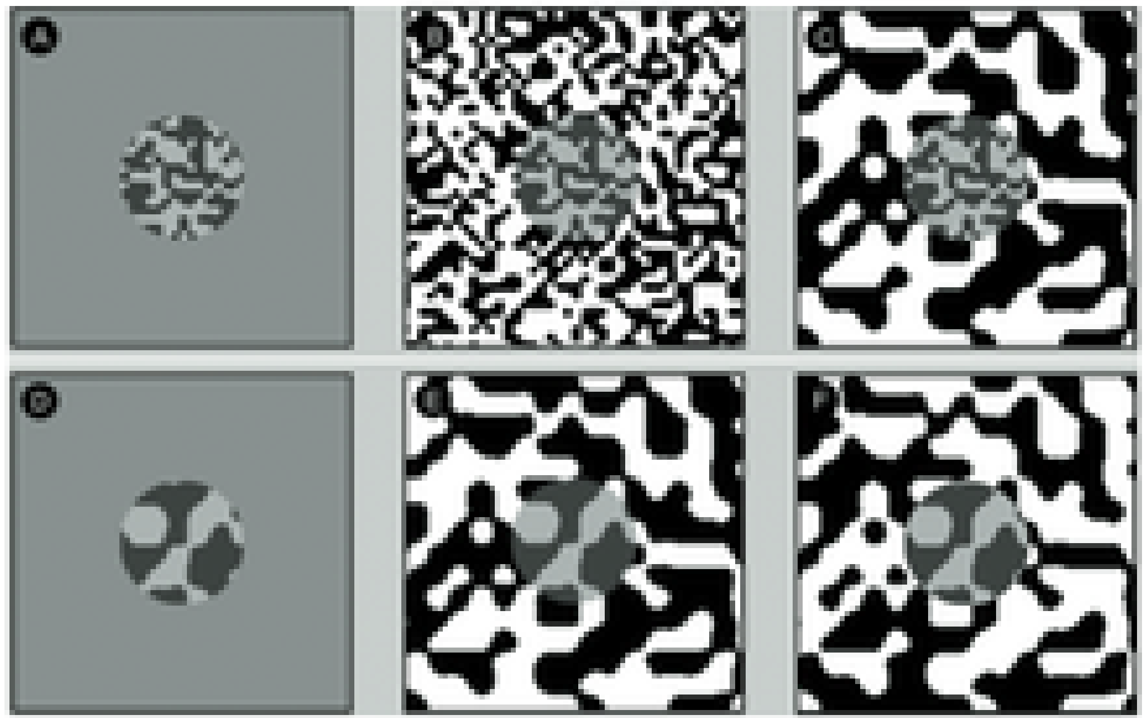
Adding the edge (14% of the total width)



Chubb illusion



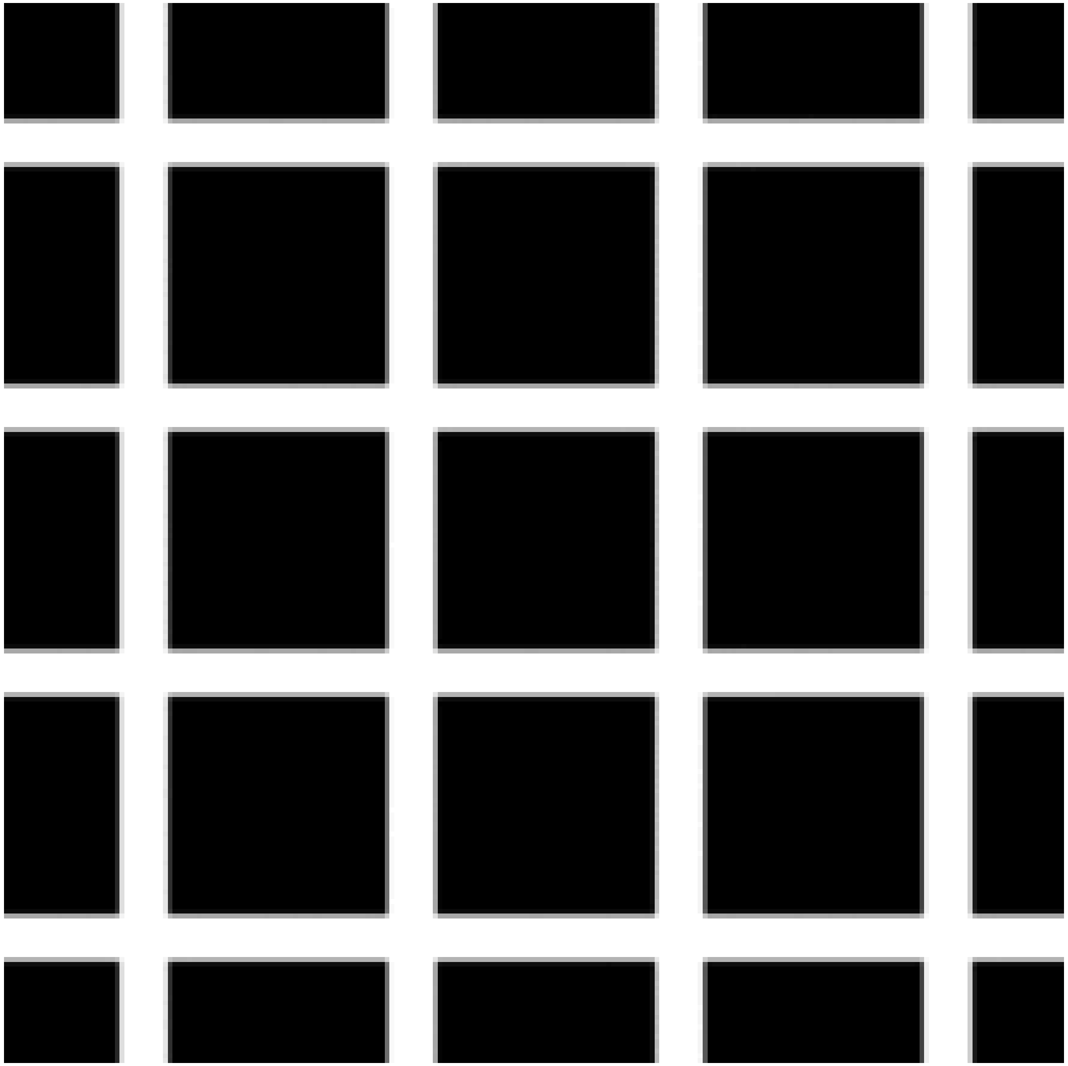
simultaneous contrast



The center areas of two rectangular fields are identical, but appear different because the background fields are different

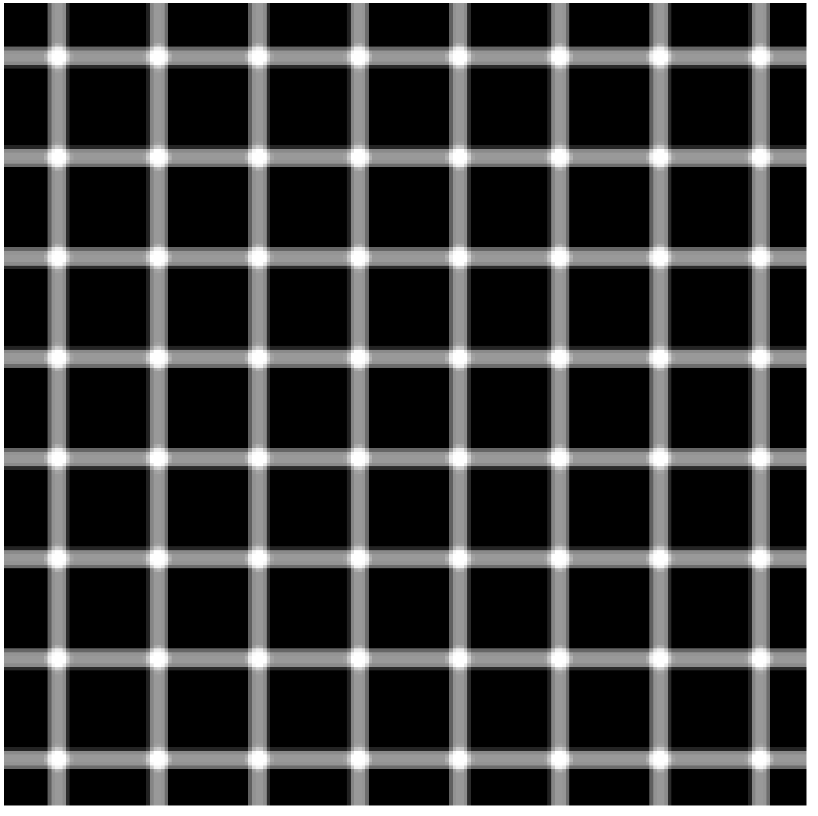
Grid illusion

Dark blobs appear at the intersections

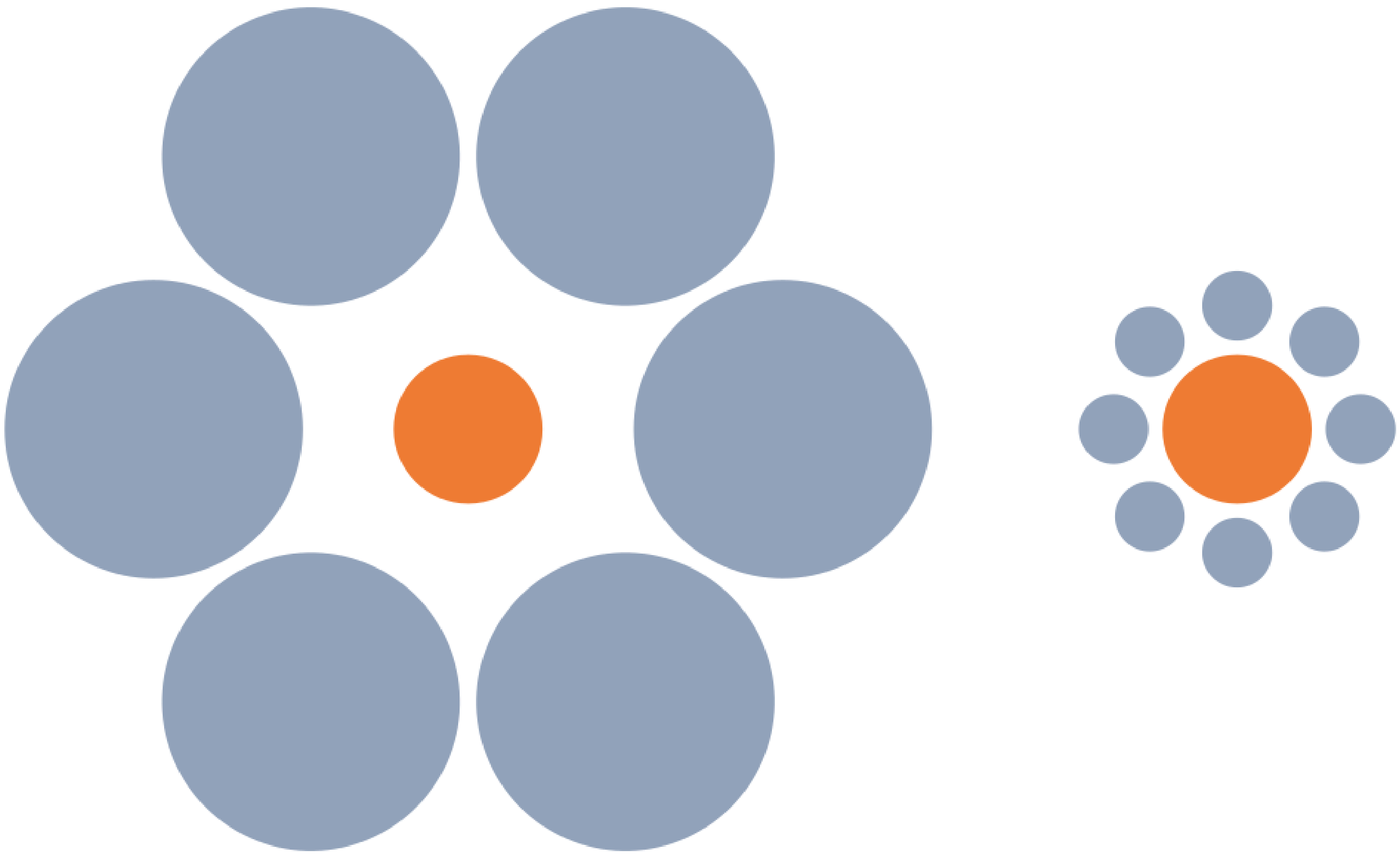


Grid illusion

Dark dots seem to appear and disappear at intersections.



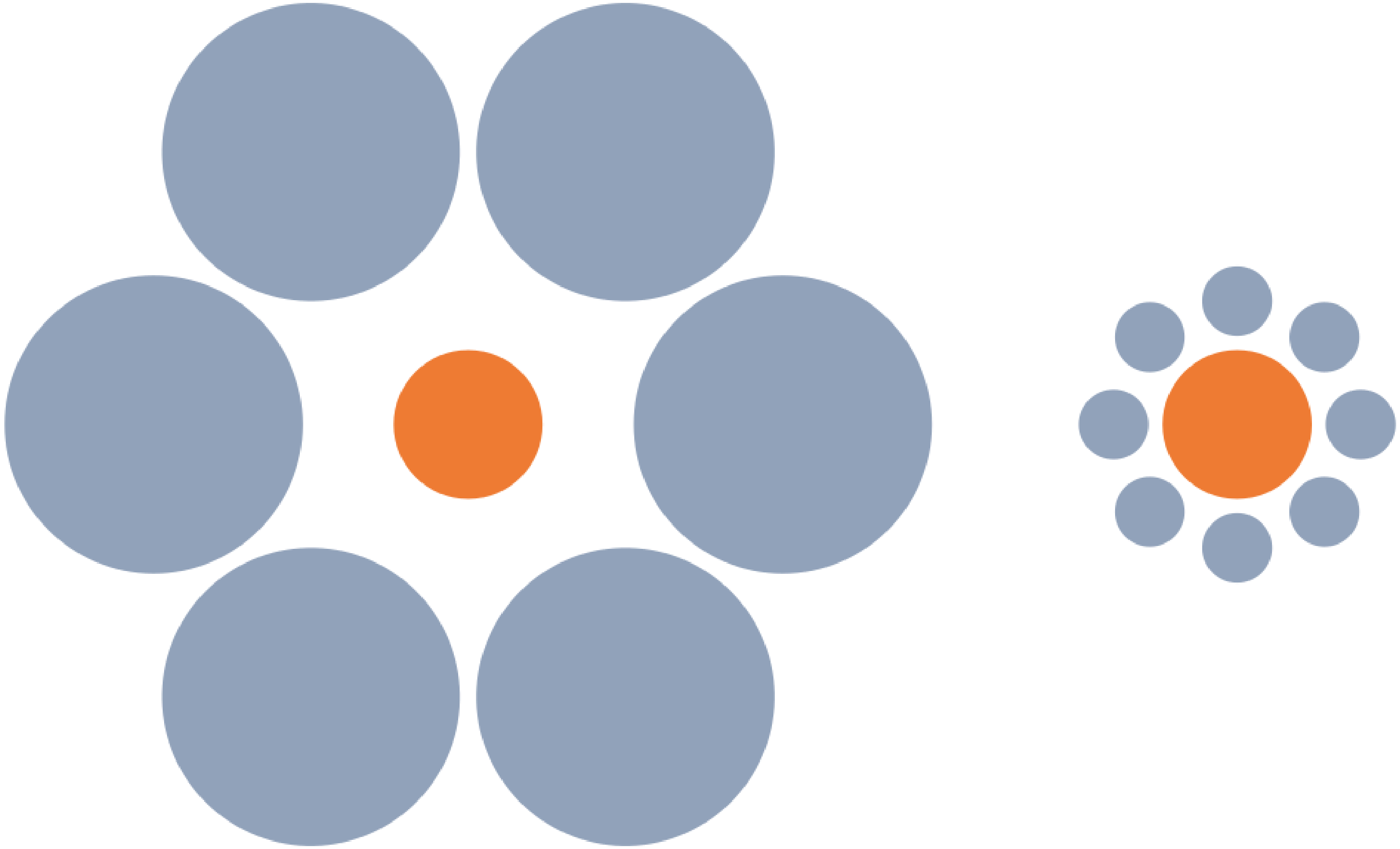
Ebbingshaus illusion



Ebbingshaus illusion

The two orange circles are exactly the same size

However, the one on the right appears larger



Ebbingshaus illusion

The three blue crosses are exactly the same size

however, the one on the left (fig. 1) tends to appear larger

fig. 1

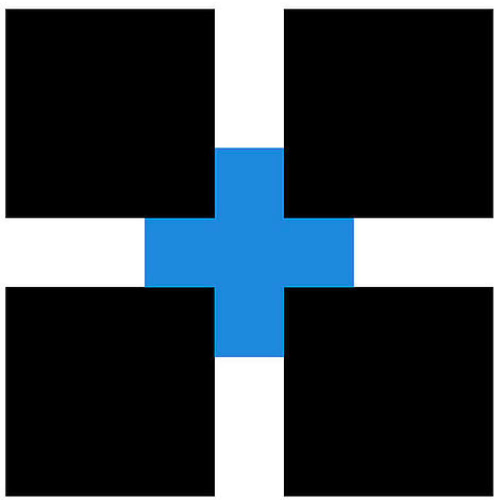


fig. 2

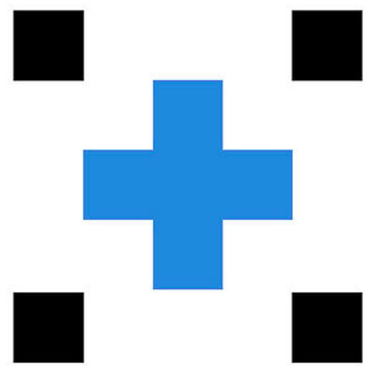
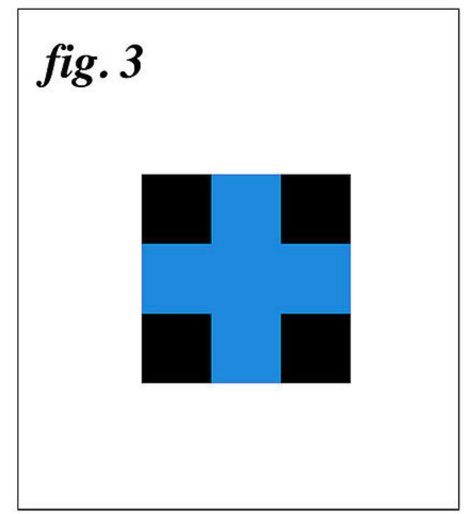


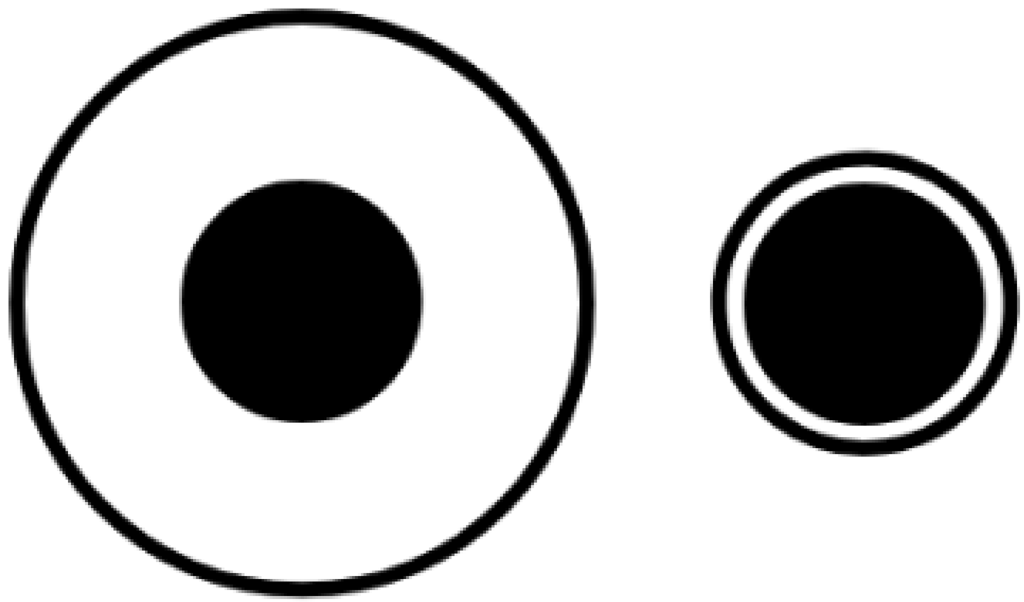
fig. 3



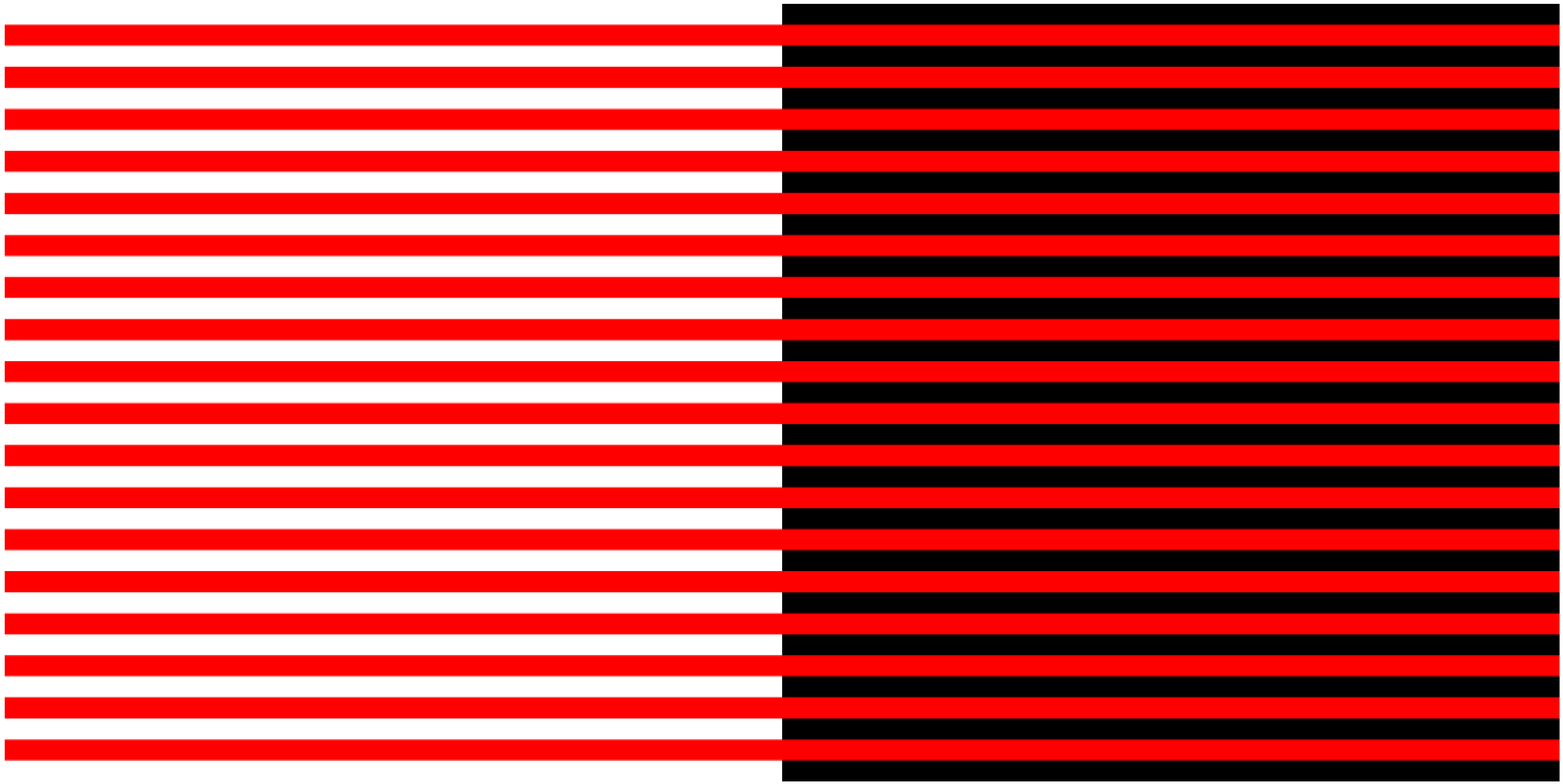
Delboeuf illusion (b&w version)

the two circles are the same size, even though

the left circle seems smaller than the right one



Bezold effect

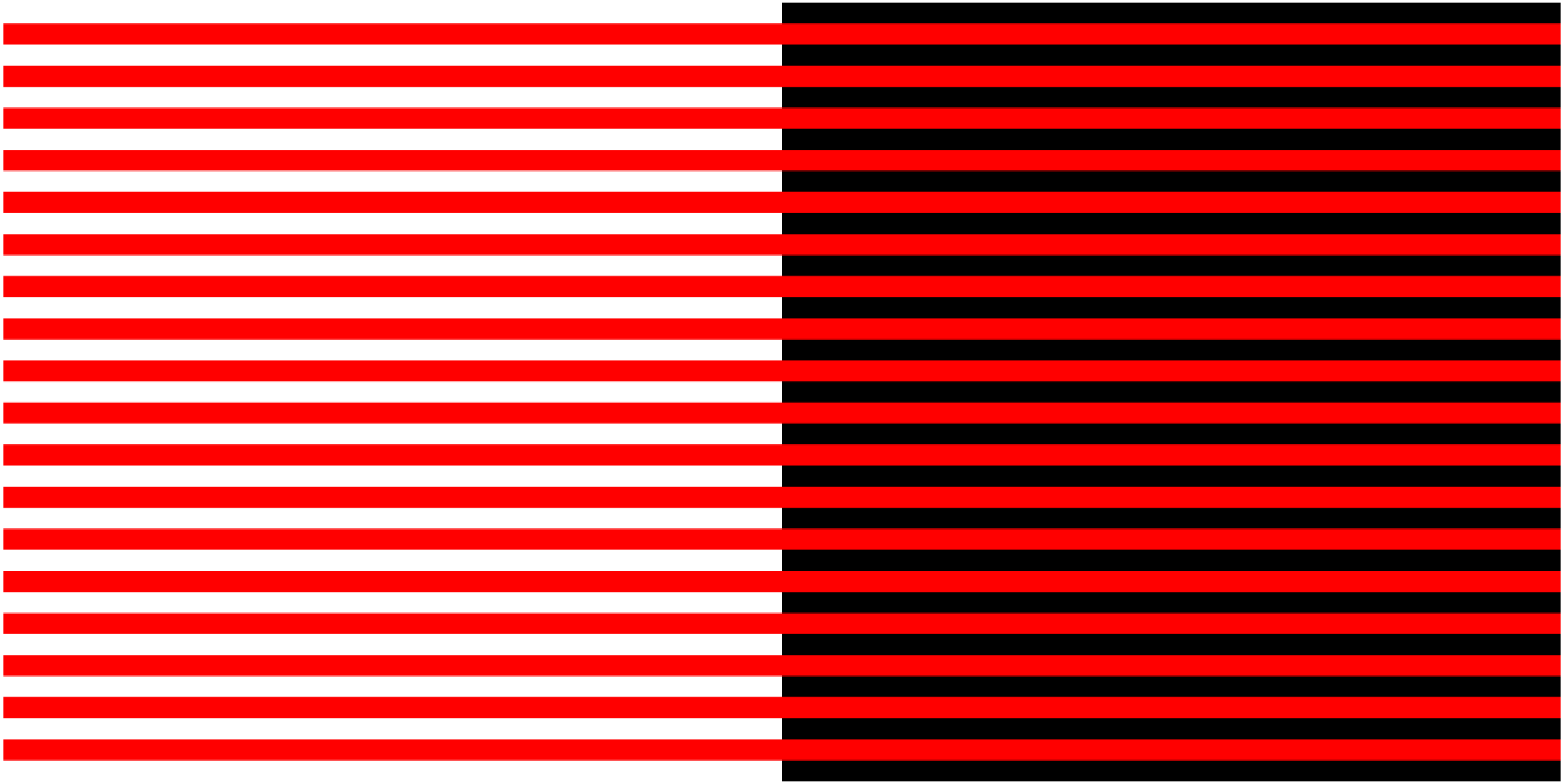


Bezold effect

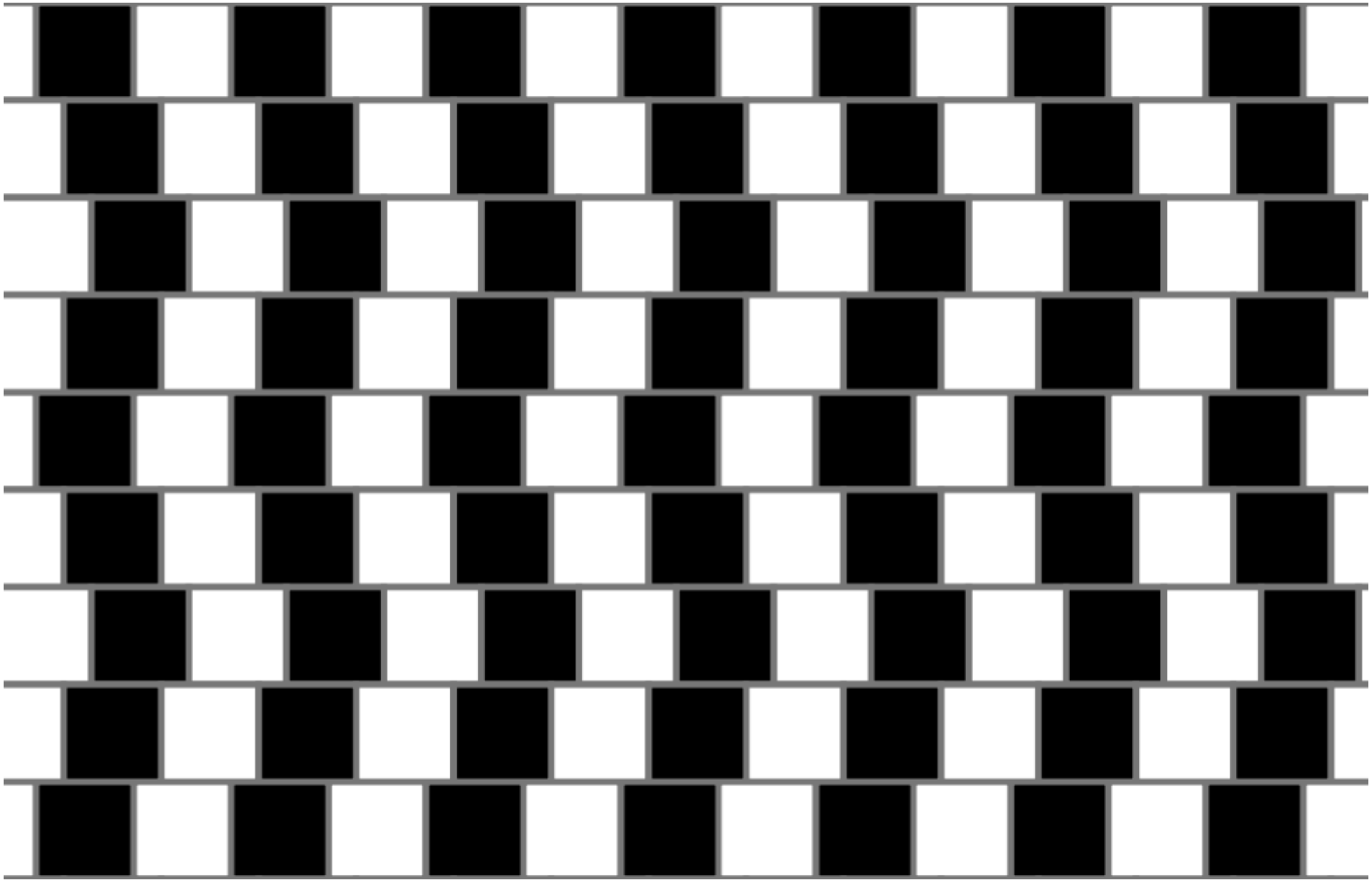
The red seems lighter combined with the white

and

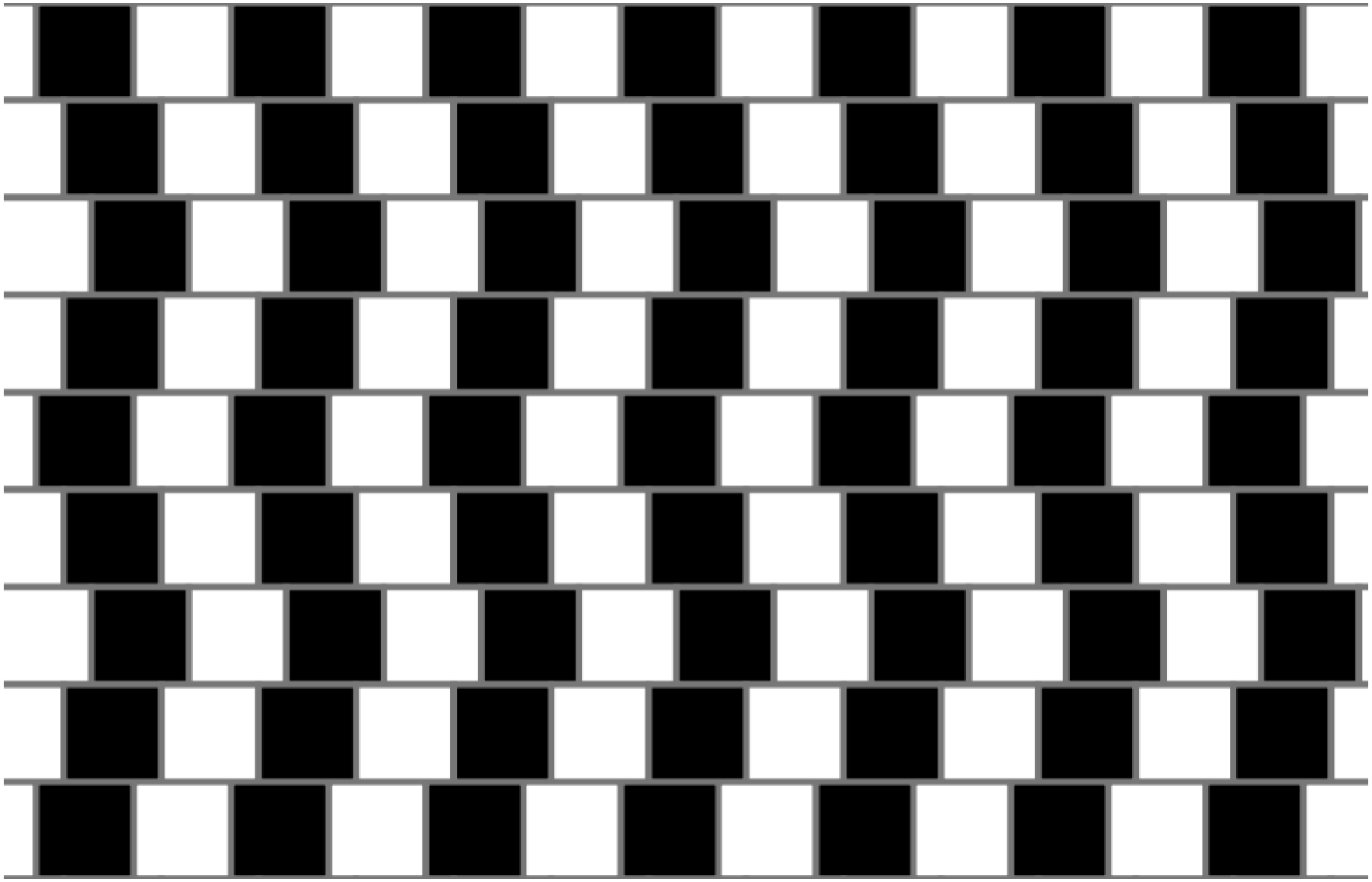
darker combined with the black



Cafe wall illusion

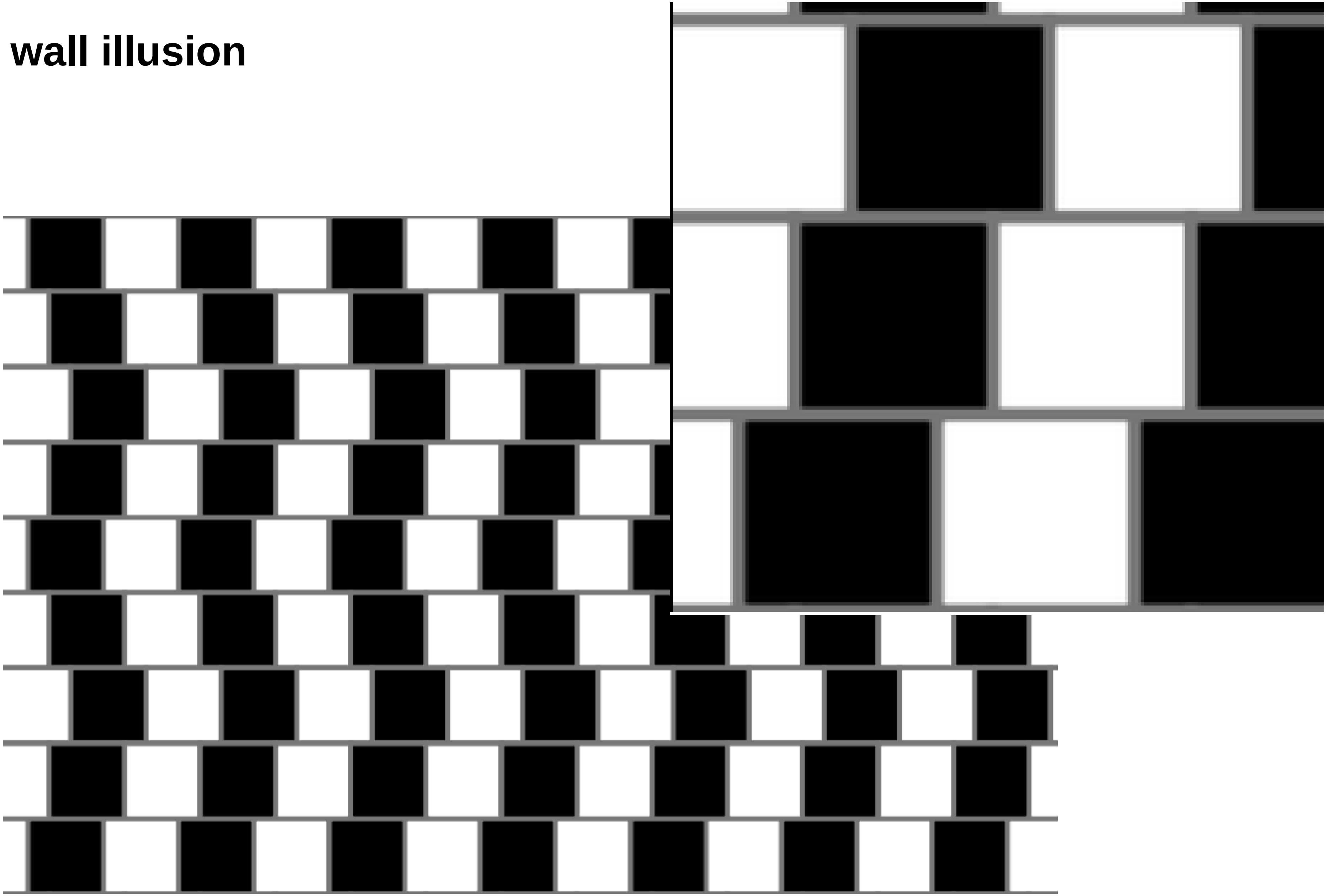


Cafe wall illusion



horizontal lines are parallel, despite appearing to be at different angles to each other

Cafe wall illusion

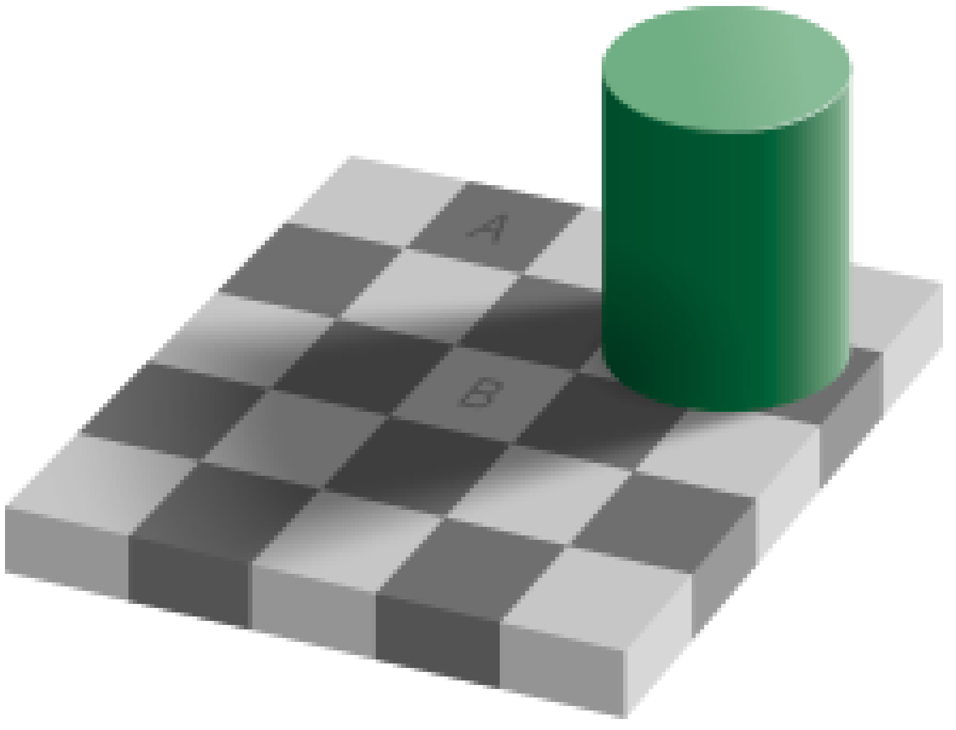


horizontal lines are parallel, despite appearing to be at different angles to each other

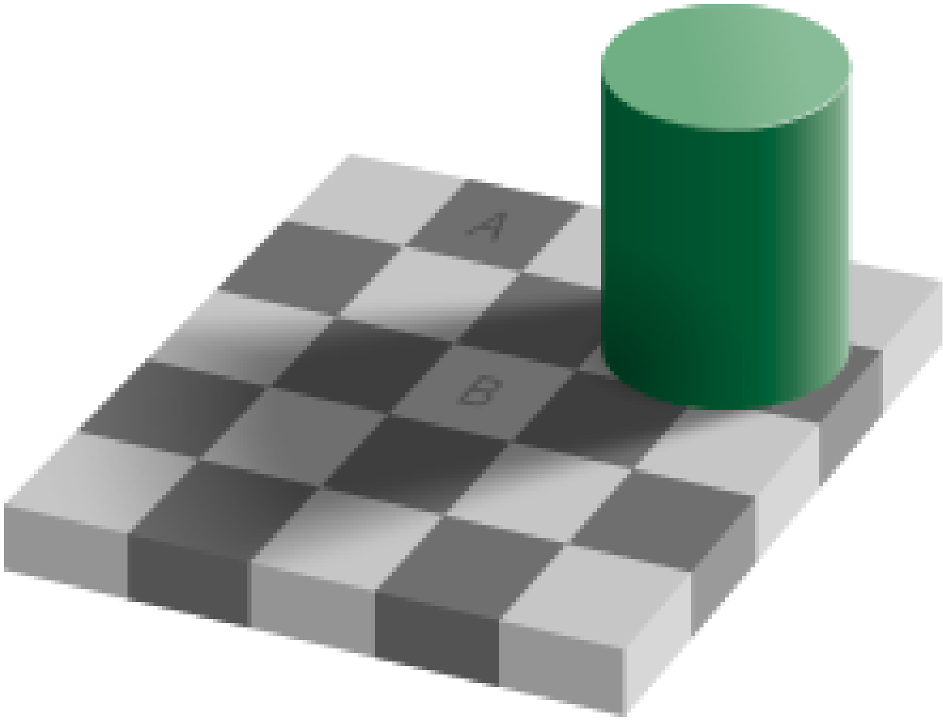
Cafe wall illusion

[Melbourne Docklands]

Checker shadow illusion

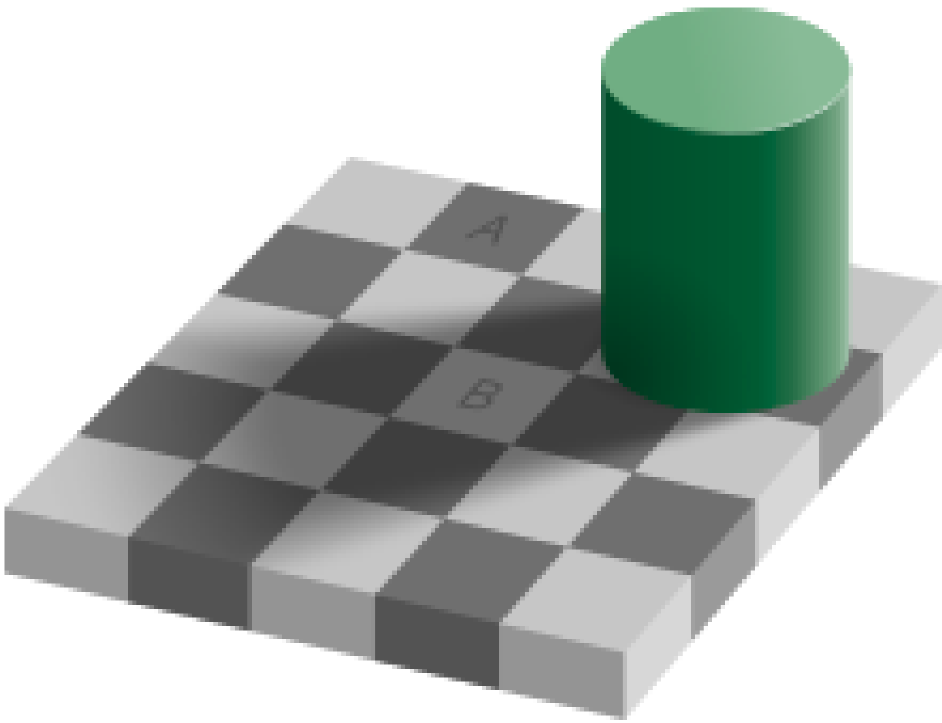


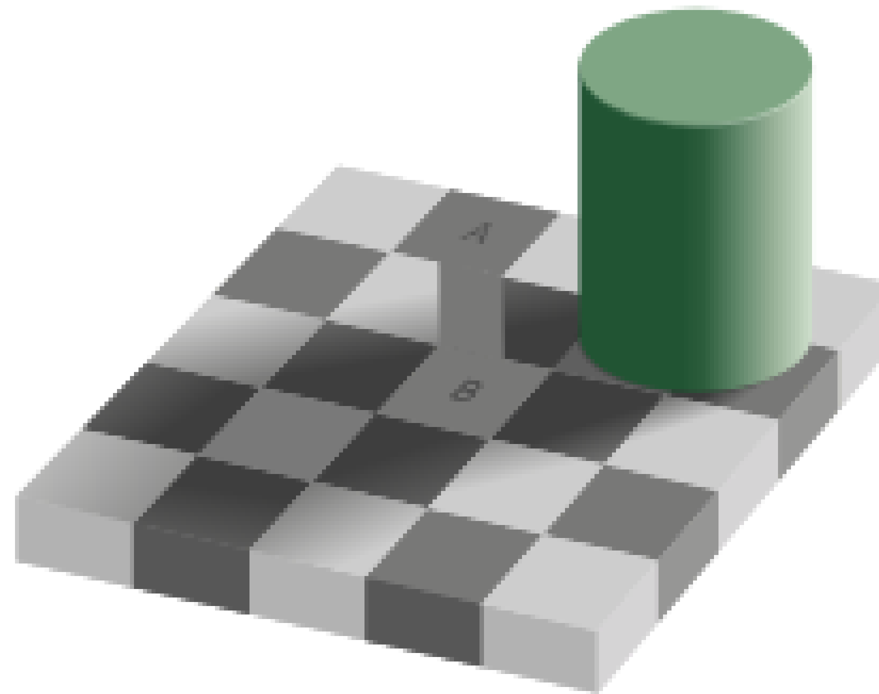
Checker shadow illusion



The squares marked A and B
are the same shade of gray

Checker shadow illusion

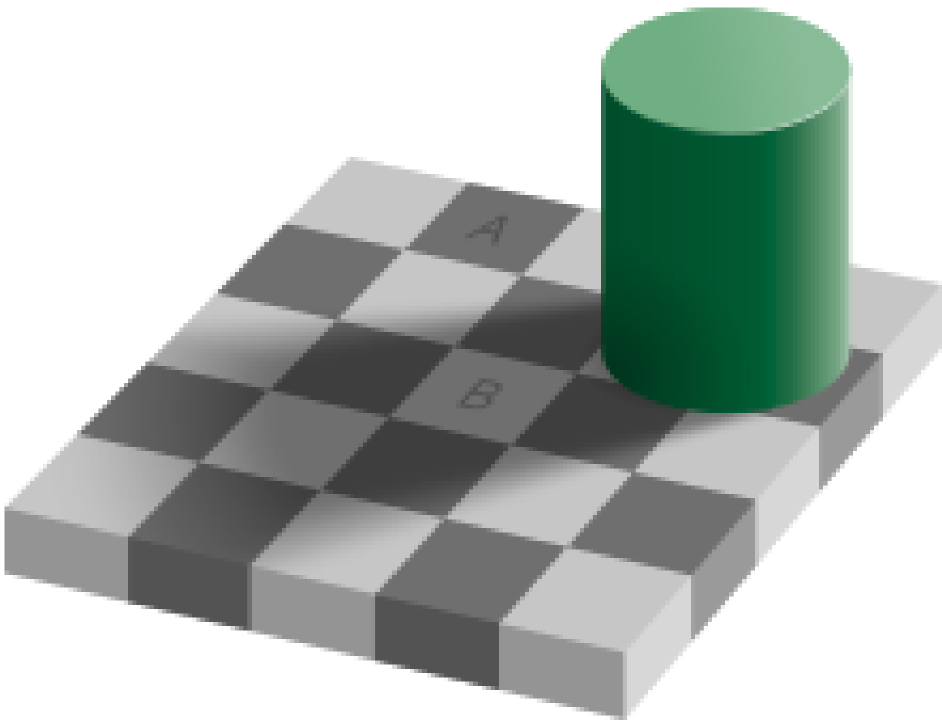




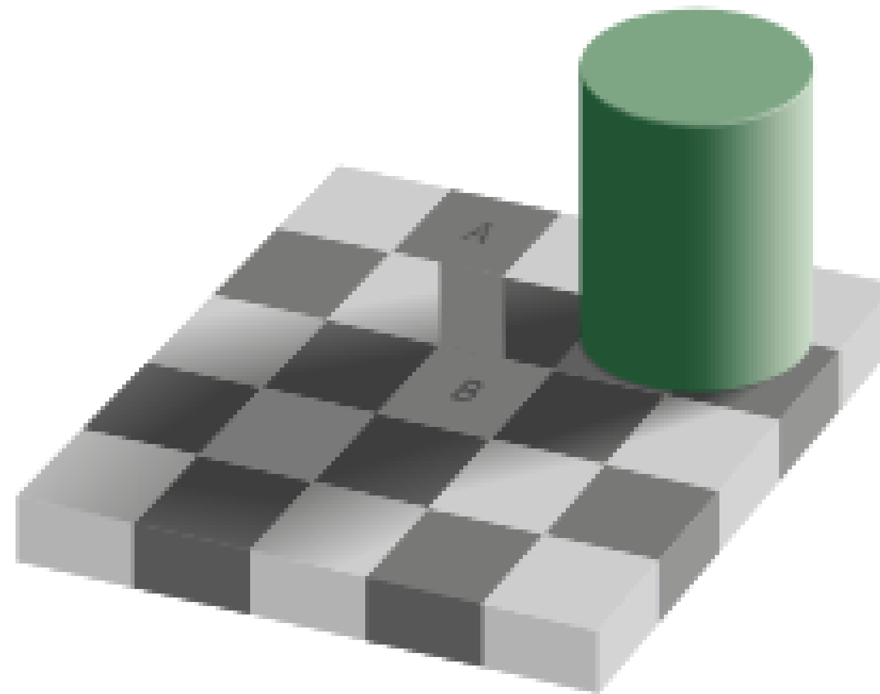
The squares marked A and B
are the same shade of gray

Verification

Checker shadow illusion



The squares marked A and B
are the same shade of gray

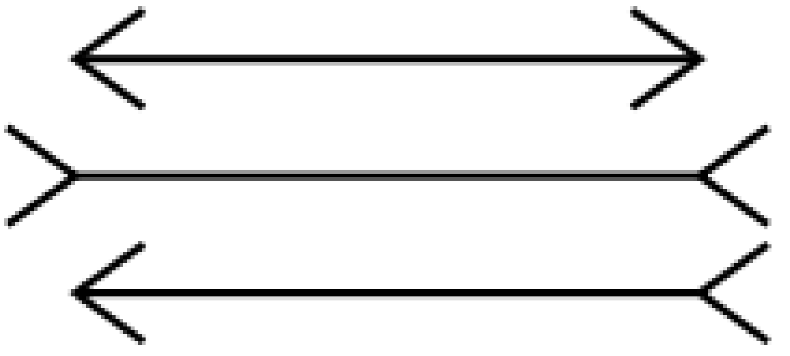


Verification

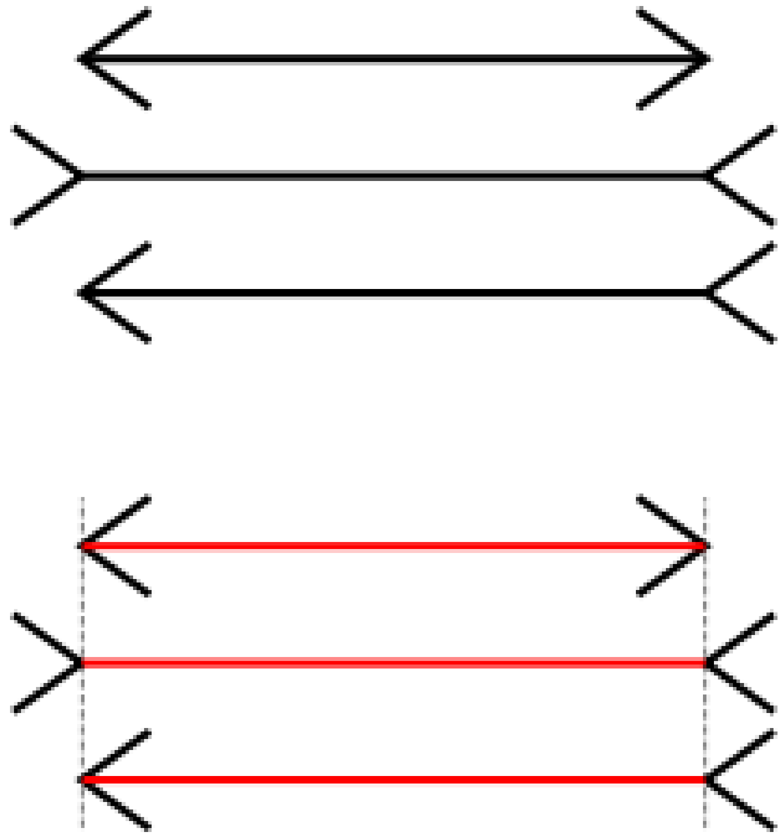
Avoid shades in the plots

Maybe you think they make the plots pretty, but you only add another layer of confusion

Müller-Lyer illusion



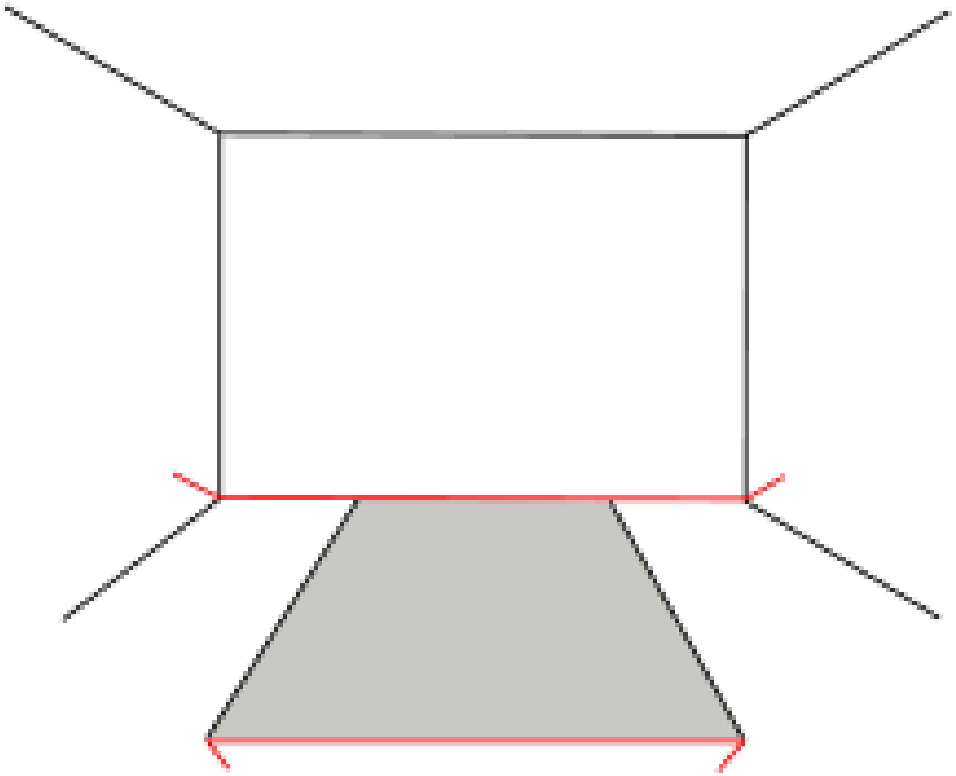
Müller-Lyer illusion



The set on the bottom shows that

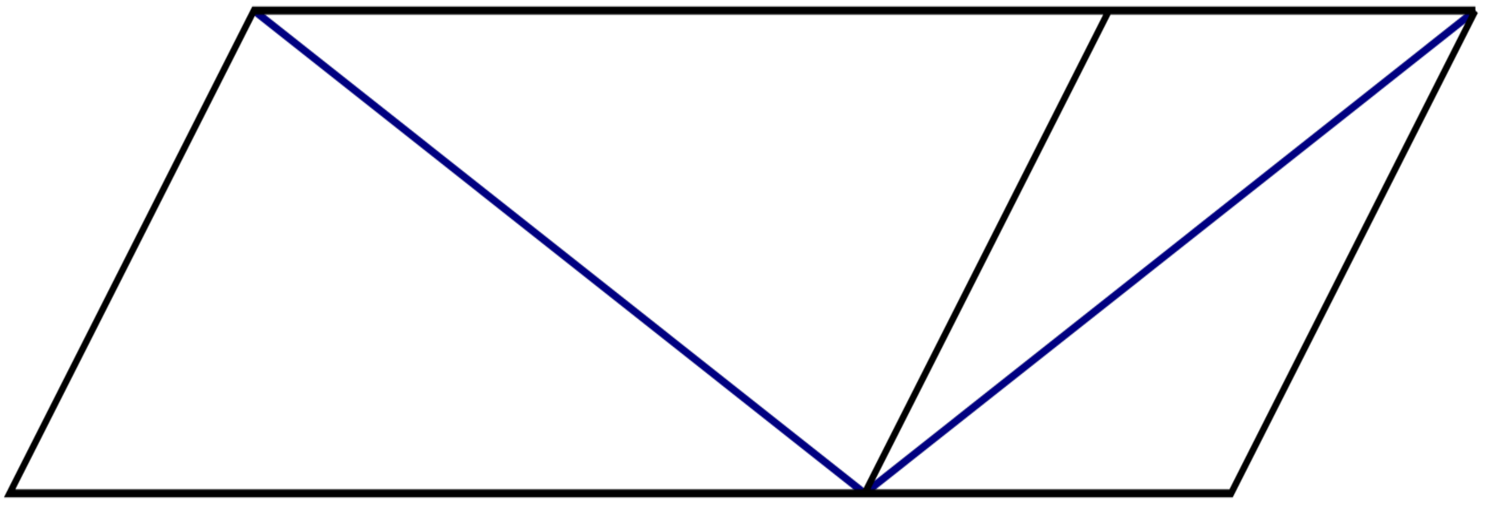
all the shafts of the arrows are of the same length.

Müller-Lyer illusion

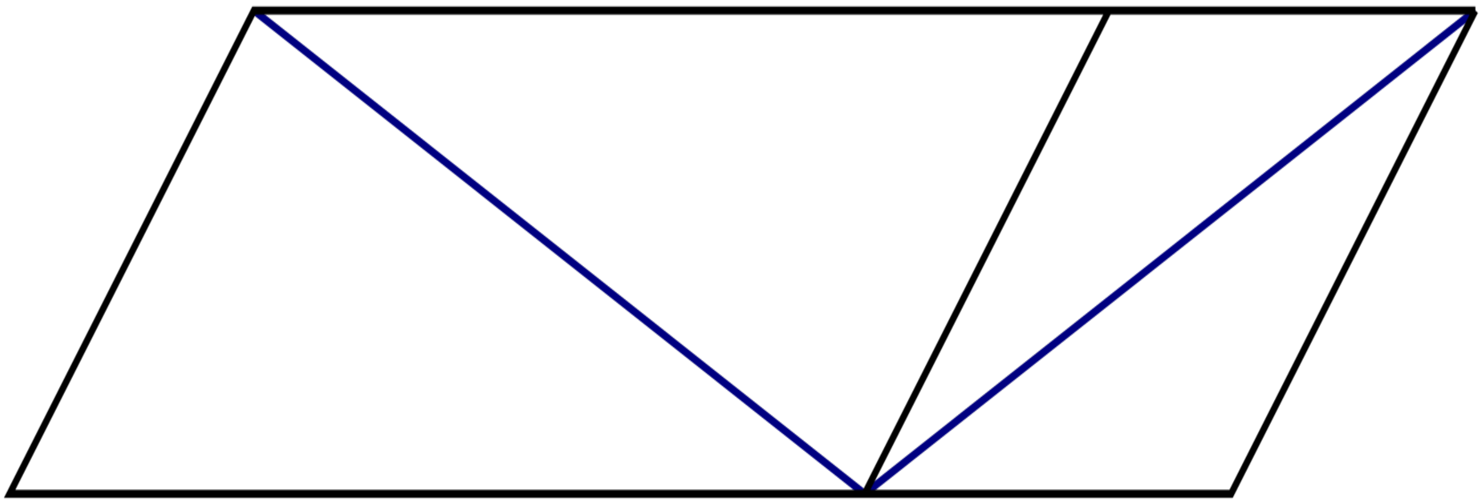


Thus, avoid 3d plots

Sander illusion



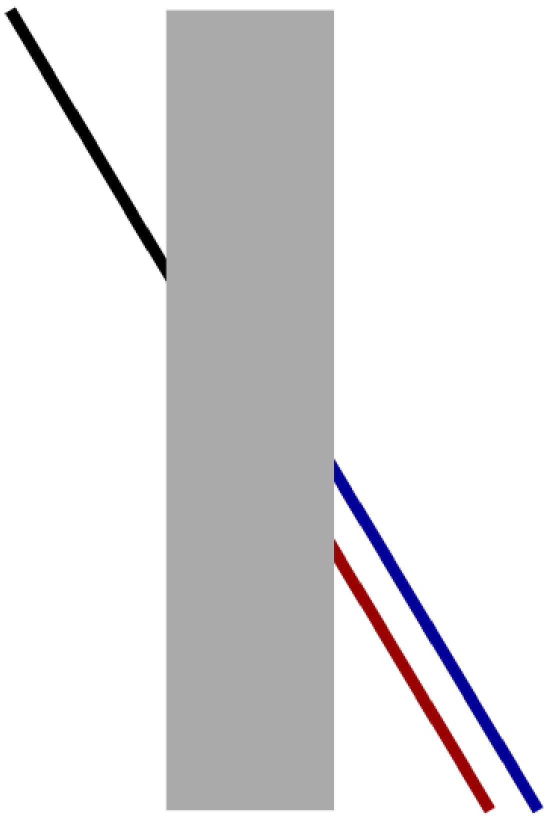
Sander illusion



The diagonal line bisecting the larger, left-hand parallelogram appears to be considerably longer than the diagonal line bisecting the smaller, right-hand parallelogram, but is the same length.

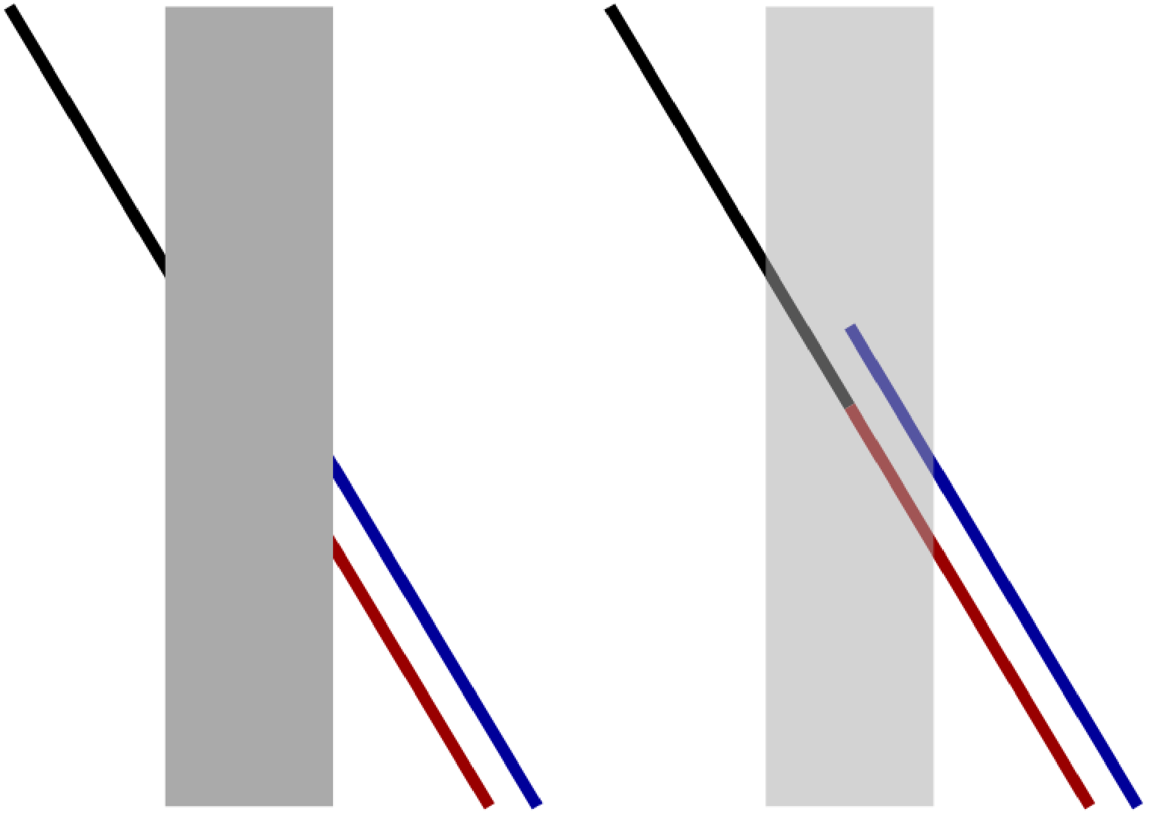
Thus, avoid 3d plots

Poggendorff illusion



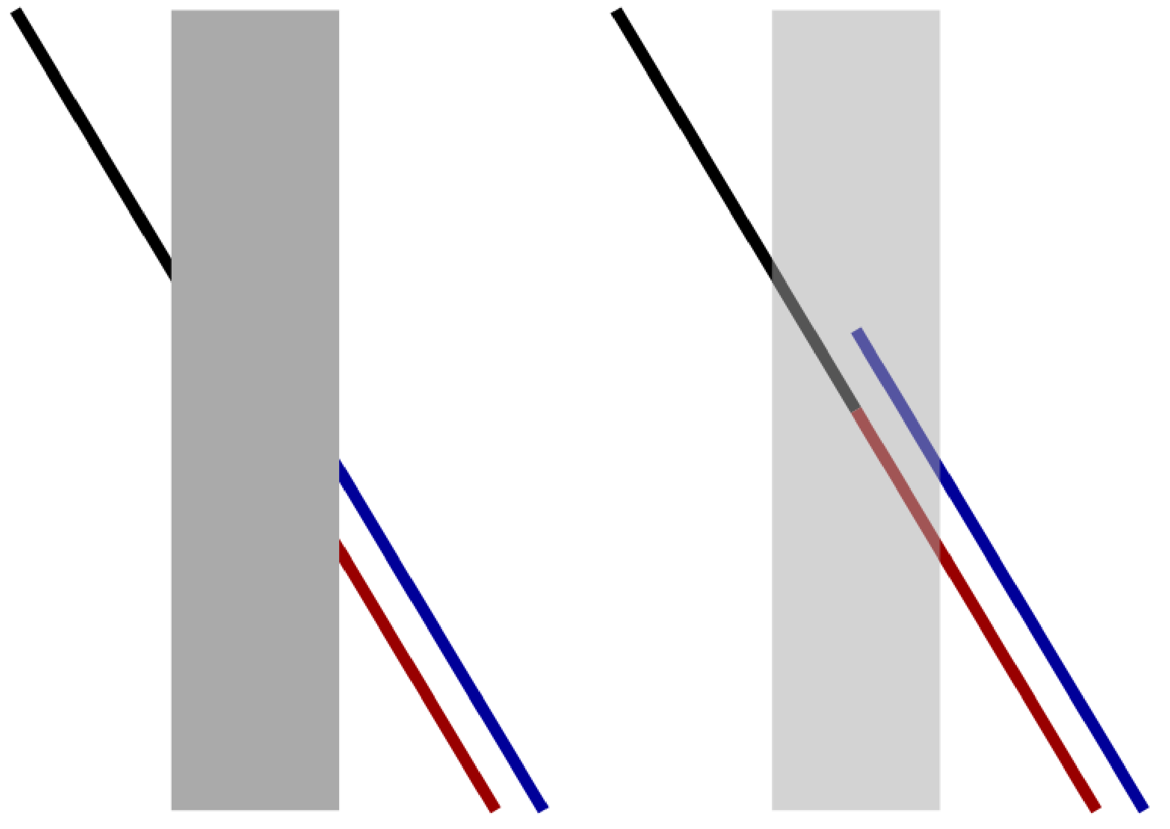
The blue line, rather than the red line,
appears to be a continuation of the black one

Poggendorff illusion



The blue line, rather than the red line,
appears to be a continuation of the black one which is not the case

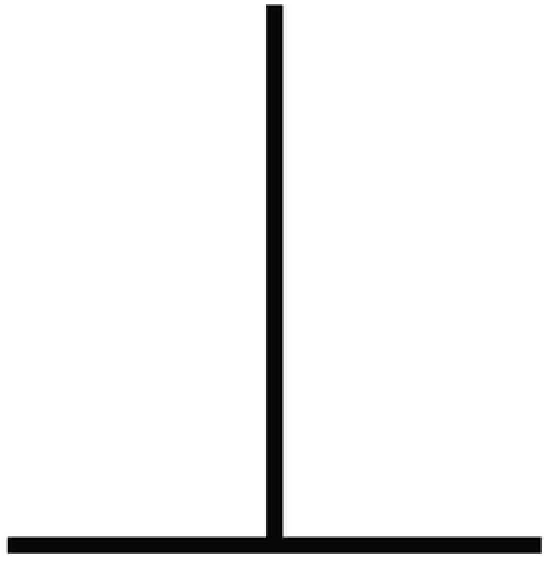
Poggendorff illusion



The blue line, rather than the red line,
appears to be a continuation of the black one which is not the case

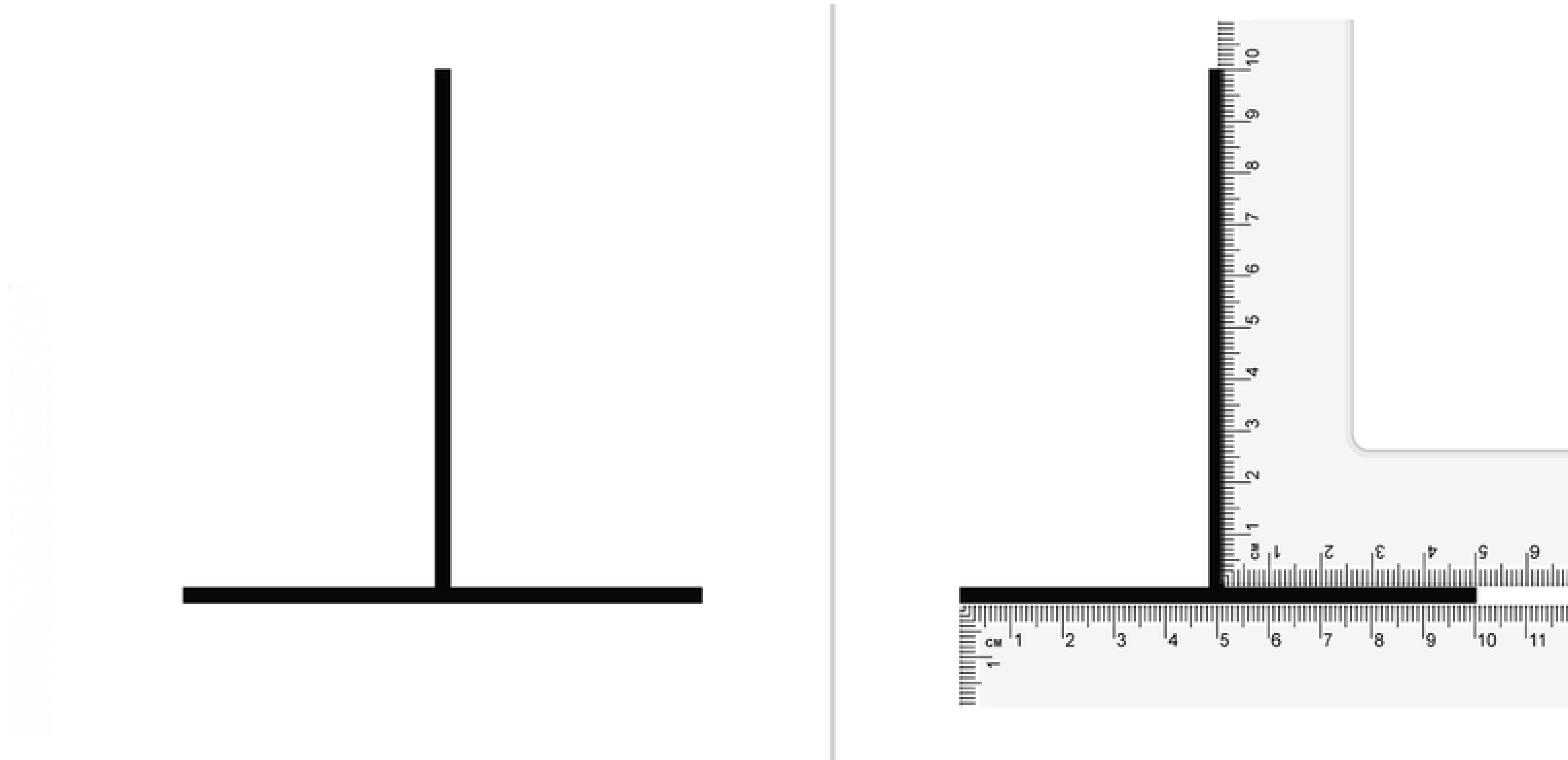
Thus, do not overlap/merge parts of plots

Vertical–horizontal illusion



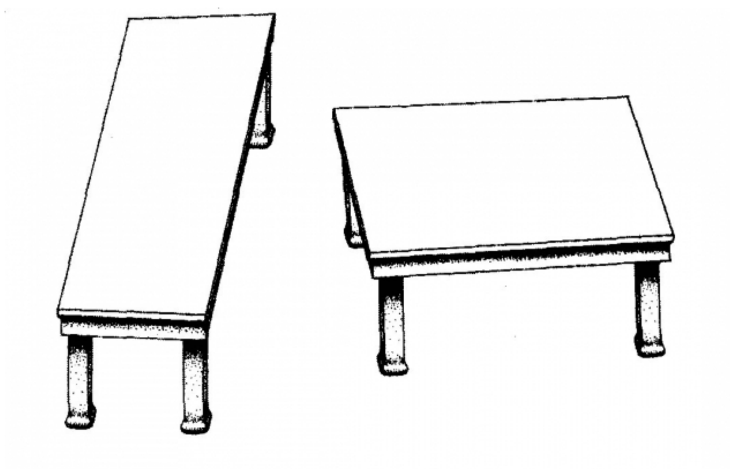
Which is longer?

Vertical–horizontal illusion



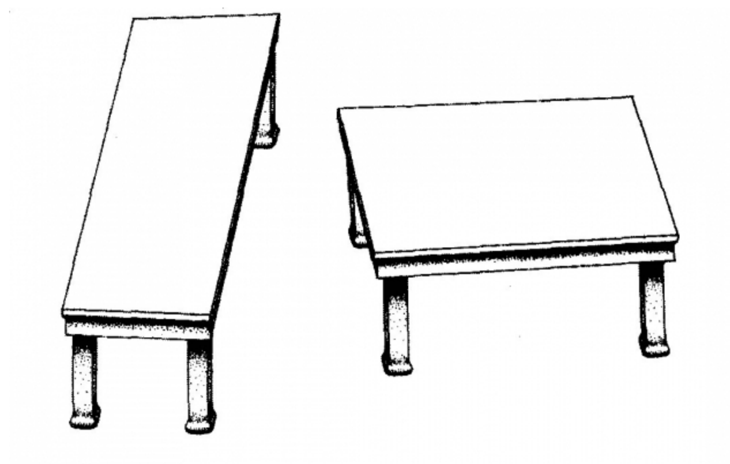
the tendency for observers to overestimate the length of a vertical line relative to a horizontal line of the same length

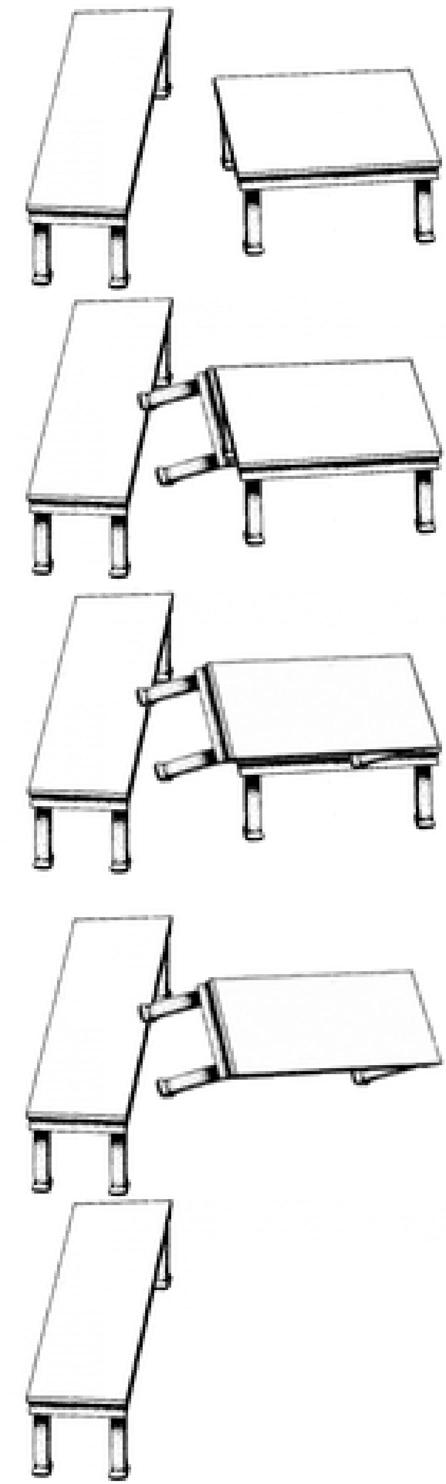
Shepard tabletop illusion



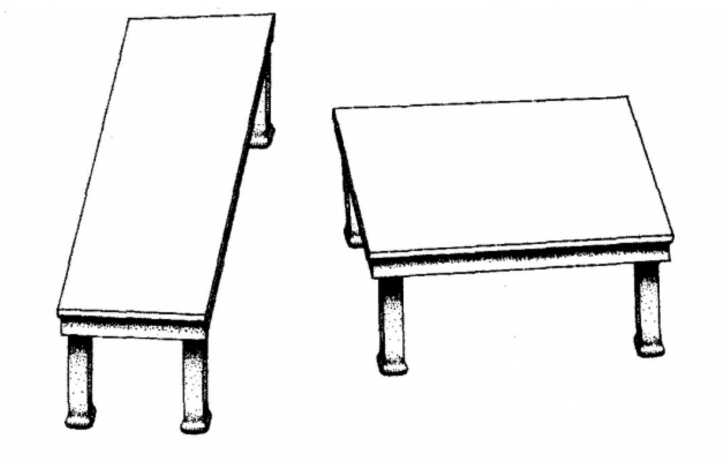
Which is bigger?

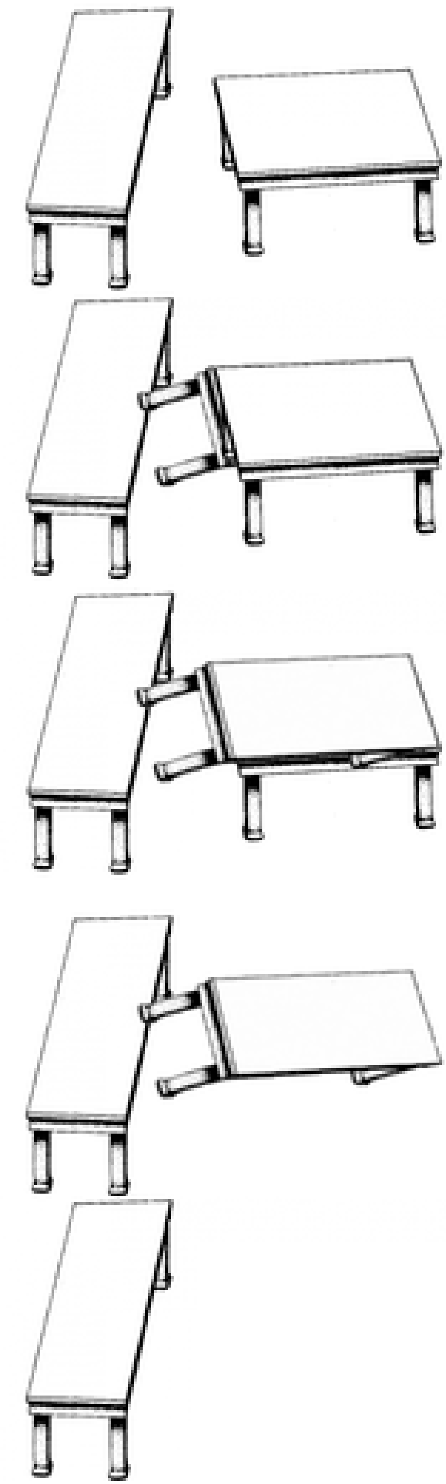
Shepard tabletop illusion





Shepard tabletop illusion



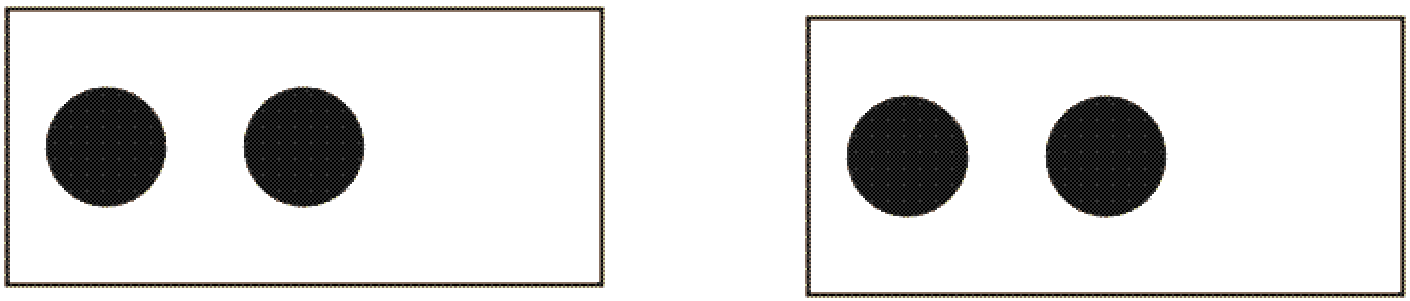


typically creating length miscalculations of 20–25%

Do not mess with 3d or distort shapes

(e.g. bar plots)

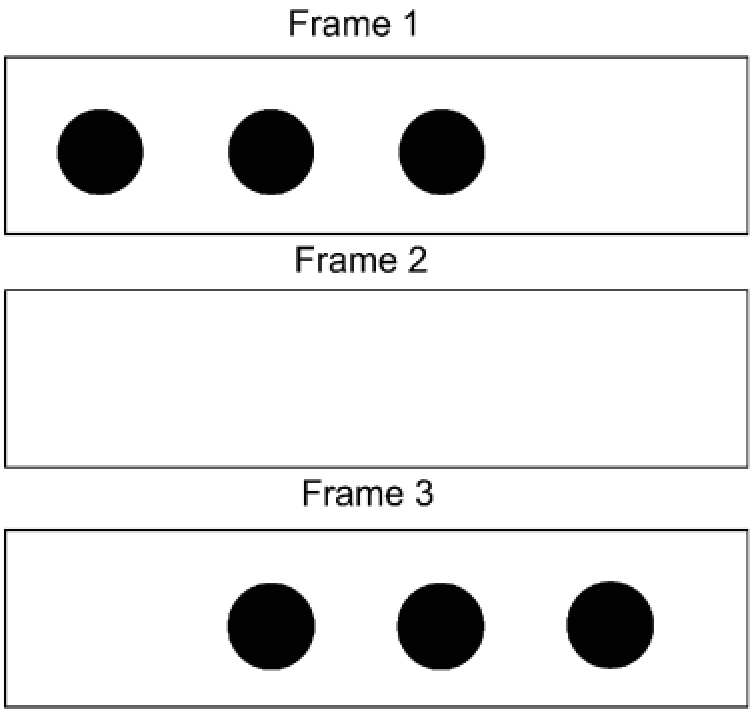
Ternus illusion



https://en.wikipedia.org/wiki/Ternus_illusion

Can be used for animation with sparse data

Ternus illusion



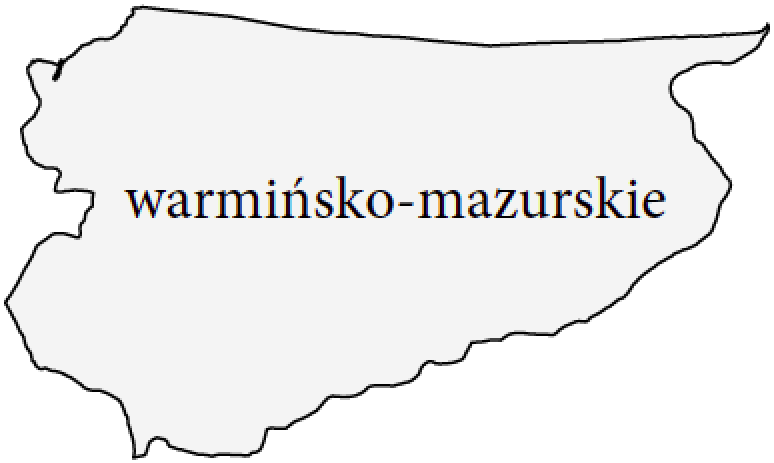
https://en.wikipedia.org/wiki/Ternus_illusion

Can be used for animation with sparse data

Which voideship is bigger?

Which voideship is bigger?

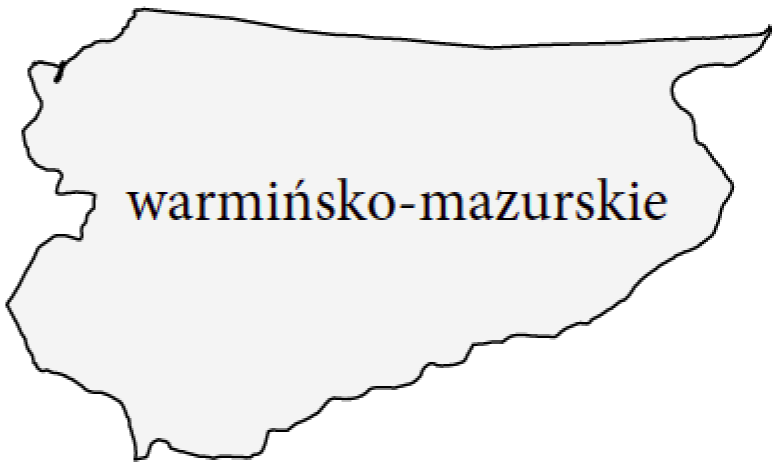
24 173 km^2



25 133 km^2

Which voideship is bigger?

24 173 km^2

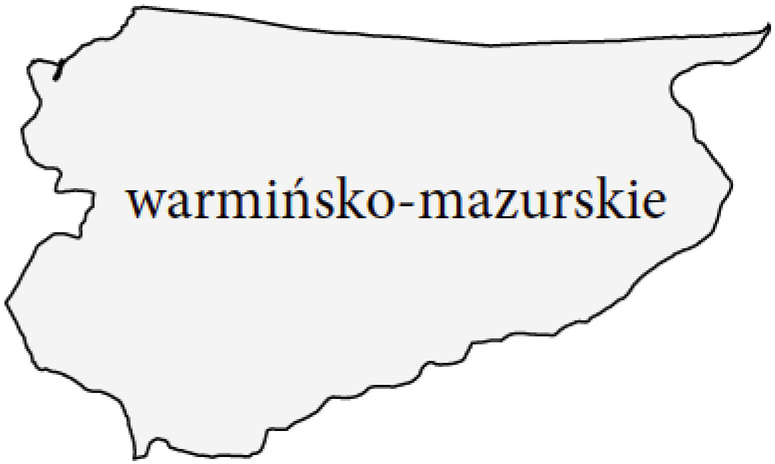


25 133 km^2

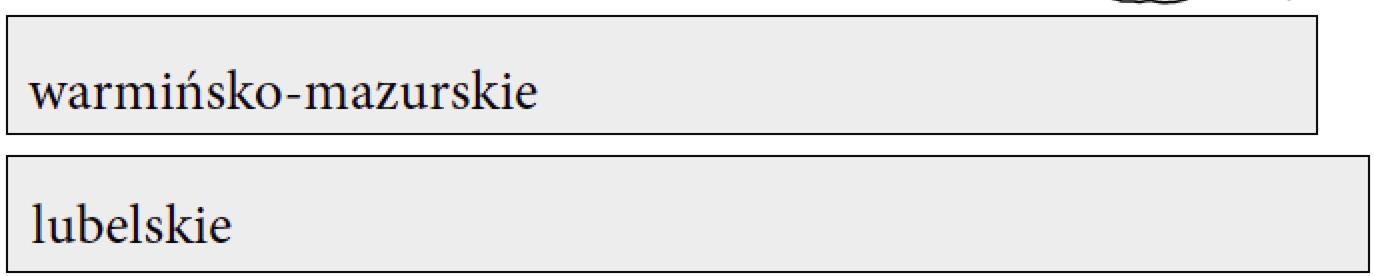
Still not convinced?

Which voideship is bigger?

24 173 km²



25 133 km²



Which voideship is bigger?

24 173 km^2

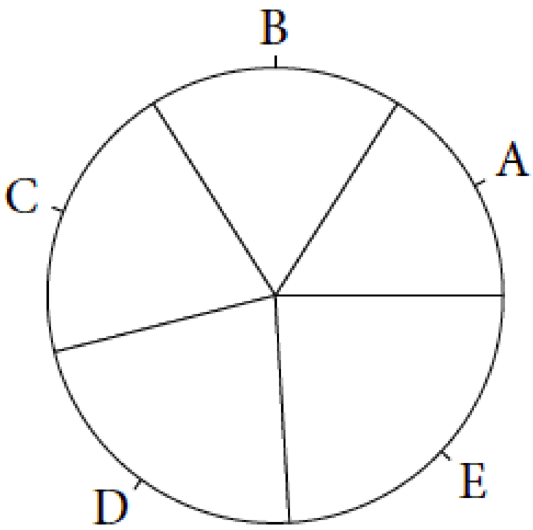
25 133 km^2

Why this misinterpretation happen?

1) area is very poor indicator/metric

2) the differences can be easier decoded by human brain if their presented as the length (of the bar)

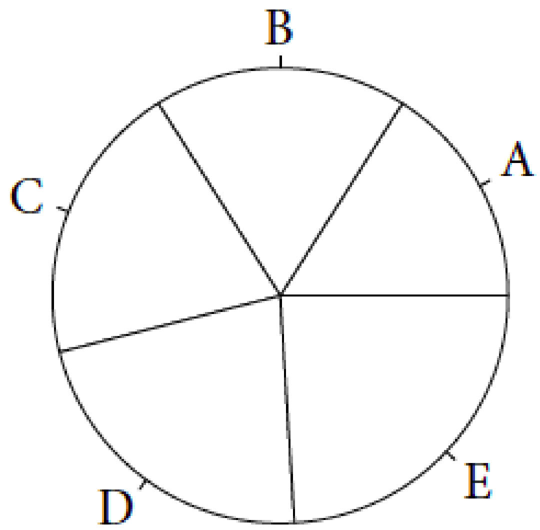
Pie chart use area



Can you order classes?

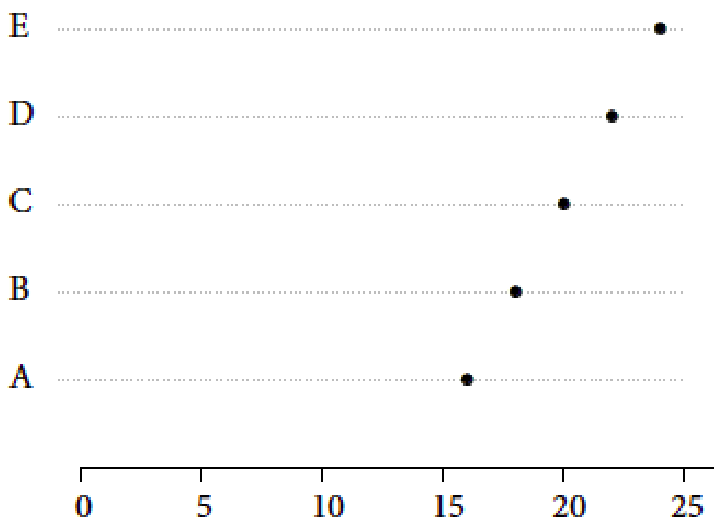
Can you estimate the values (area)?

Pie chart use area



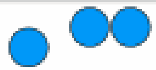
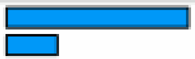
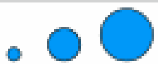
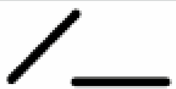
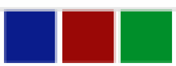
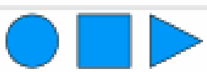
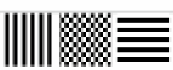
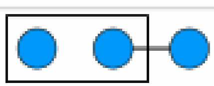
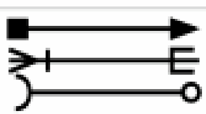
Can you order classes?

Can you estimate the values (area)?



Much better, don't you think so?

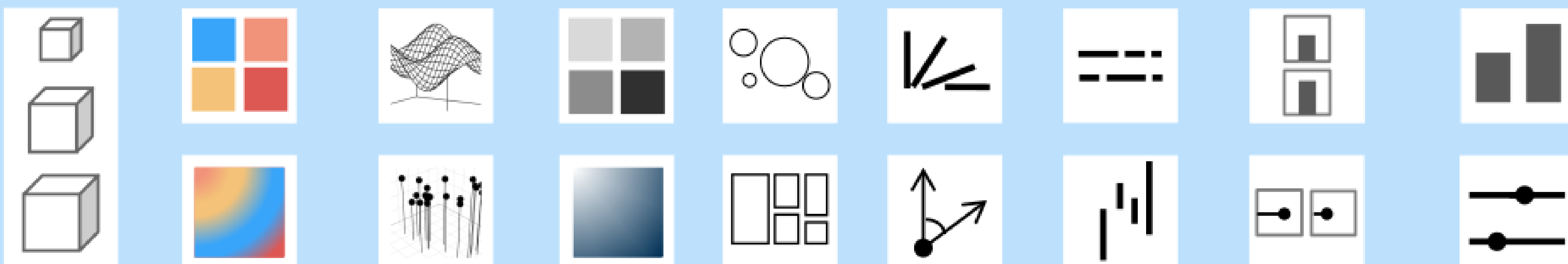
Properties and Best Uses of Visual Encodings

Example	Encoding	Ordered	Useful values	Quantitative	Ordinal	Categorical	Relational
	position, placement	**yes**	**infinite**	Good	Good	Good	Good
1, 2, 3; A, B, C	text labels	**optional** (alphabetical or numbered)	**infinite**	Good	Good	Good	Good
	length	**yes**	**many**	Good	Good		
	size, area	**yes**	**many**	Good	Good		
	angle	**yes**	medium/few	Good	Good		
	pattern density	**yes**	few	Good	Good		
	weight, boldness	**yes**	few		Good		
	saturation, brightness	**yes**	few		Good		
	color	no	few (< 20)			Good	
	shape, icon	no	medium			Good	
	pattern texture	no	medium			Good	
	enclosure, connection	no	**infinite**			Good	Good
	line pattern	no	few				Good
	line endings	no	few				Good
	line weight	**yes**	few		Good		

Effectiveness Ranking

A graph is a representation of data that visually encodes numerical values into attributes such as lines, symbols and colors. The Cleveland-McGill scale can be used to select the most effective attribute(s) for your purpose.

Volume	Color hue	Depth: 3d position	Color intensity	Area	Slope or Angle	Length	Position on unaligned scale	Position on common scale
volume charts	poorly designed heat maps	multivariate density plots	heat maps	bubble charts, mosaic charts	line graphs, pie charts	stacked bar charts, waterfall chart	small multiple plots	dot plots, bar charts, parallel coordinate plots



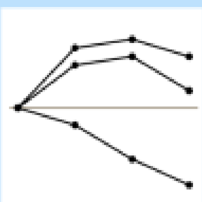
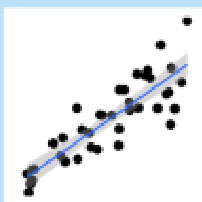
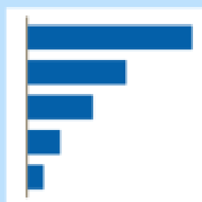
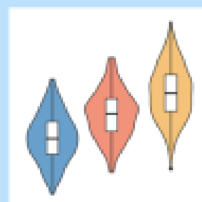
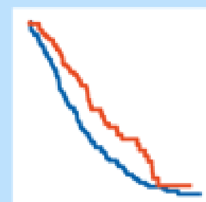
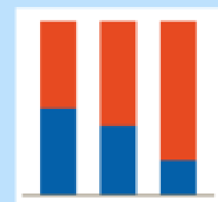
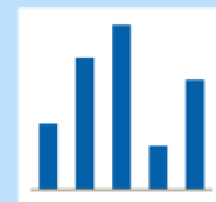
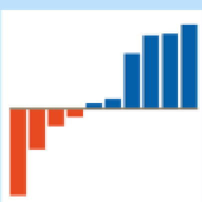
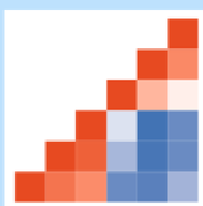
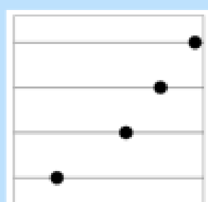
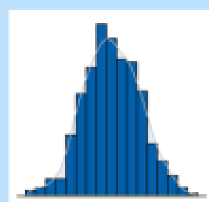
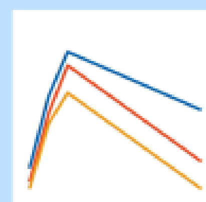
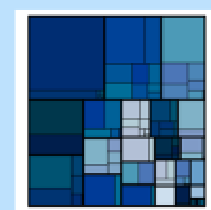
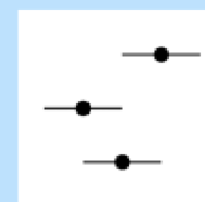
Least accurate → **Most accurate**

Alison Margolskee, Mark Baillie, Baldur Magnusson, Julie Jones, Marc Vandemeulebroecke (Novartis)

Selecting the right base graph

Consider if a standard graph can be used by identifying suitable designs based on the:
(i) **purpose** (i.e. message to be conveyed or question to answer) and (ii) **data** (i.e. variables to display).

Example plots categorized by purpose

Deviation	Correlation	Ranking	Distribution	Evolution	Part-to-whole	Magnitude
Chg. from baseline	Scatter plot	Horizontal bar chart	Boxplot	Kaplan Meier	Stacked bar chart	Vertical bar chart
						
Waterfall	Heat map	Dotplot	Histogram	Line plot	Tree map	Forest plot
						

Alison Margolskee, Mark Baillie, Baldur Magnusson, Julie Jones, Marc Vandemeulebroecke (Novartis)

Thank you for your time
and
See you at the next lecture

Any other
questions & comments

l.kozlowski@mimuw.edu.pl