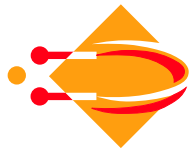
Images and Vision



Introduction to

Images

Photons

The human visual system

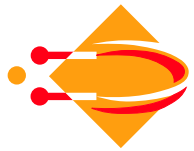
Digital representation

Mads Nielsen, ass. prof.

Martin Lillholm, ph.d. stud.

Torben Vaarby Laursen, teaching assistant

Overview



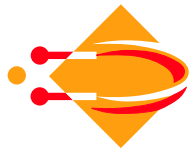
Physics: Electromagnetic waves, photons, energy, measurements.

Anatomy: The eye, retina, visual pathway, visual cortex.

Psychophysics: edges, grouping, illusory contours, the law of Prägnanz.

Computer Science: sampling, quantization, filtering, histograms.

Electromagnetic waves



Electromagnetic waves are governed by Maxwell's equations.

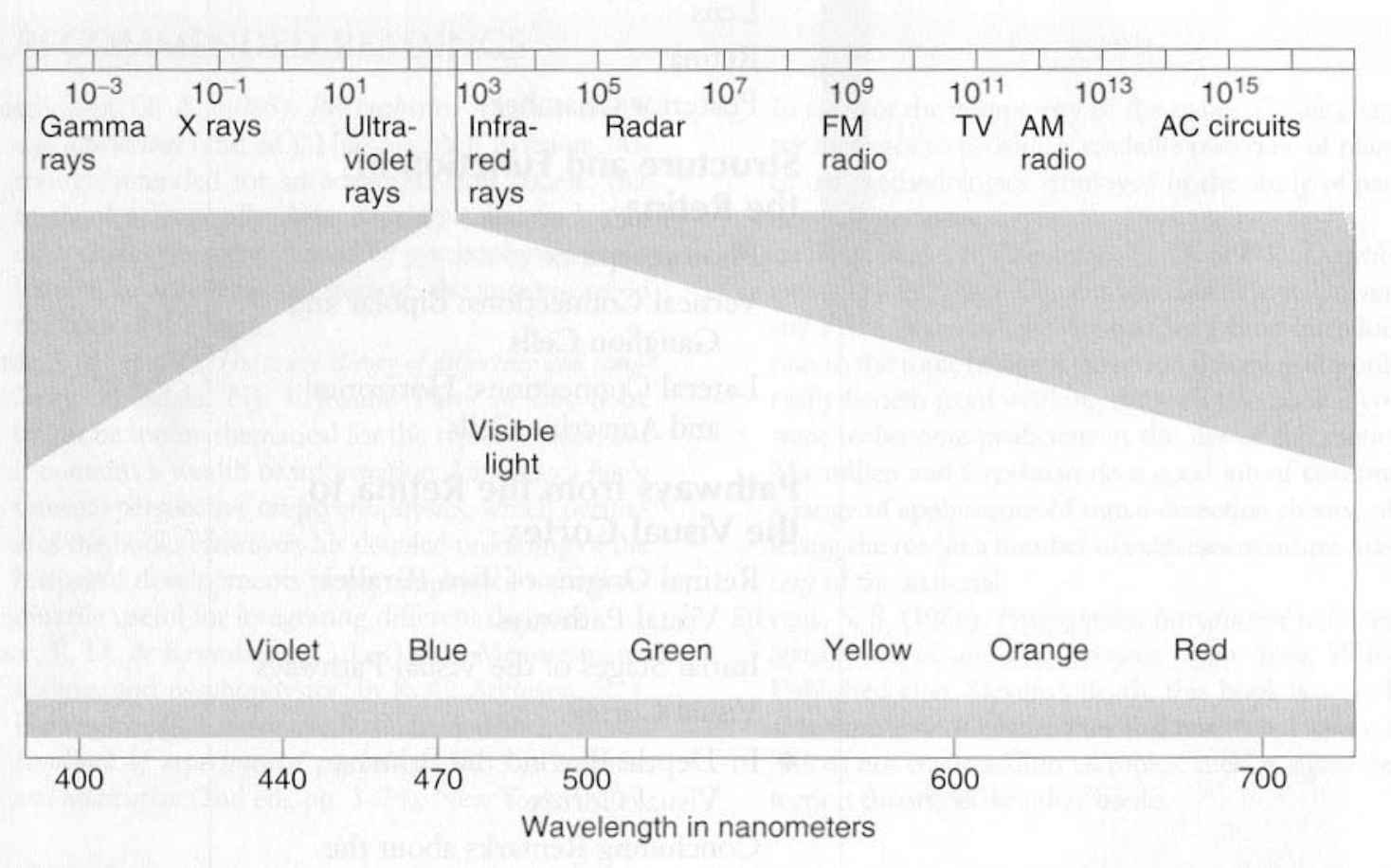
No medium required!

Solutions are waves:	$U = A\sin(kx - \omega T + \phi)$	(V)
Speed:	$c = \frac{\omega}{k}$	(m/s)
Frequency:	$f = \frac{\omega}{2\pi}$	(s^{-1} = Hz)
Wave length:	$\lambda = \frac{c}{f} = \frac{2\pi}{k}$	(m)

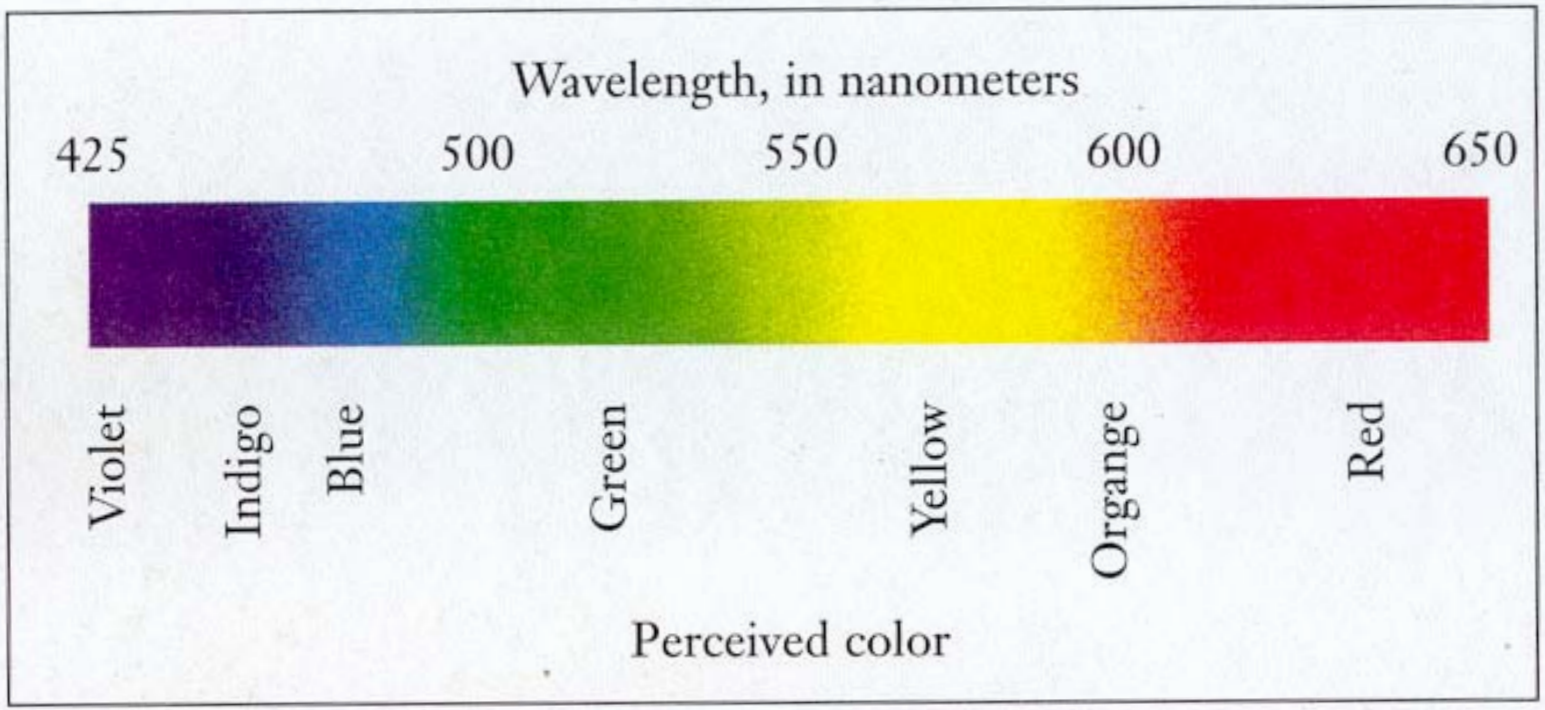
Visible light is 400 to 700 nm.

Speed in vacuum is 3×10^8 m/s.

Electromagnetic wave spectrum

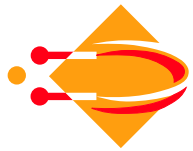


Light spectrum



From *Psychology: From Research to Applications* (Color Plate 4C) by Dennis P. Saccuzzo, 1987, Boston: Allyn and Bacon.

Photons



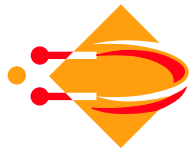
A short electromagnetic wave package is called a photon.

This may be interpreted as a particle.

It is characterized solely by its energy $E = hf$
(h is Planck's constant),
or its frequency f,
or its wavelength λ.

Since speed is different in different media, λ changes as a function of media.

Light



Light consists of a current of photons.

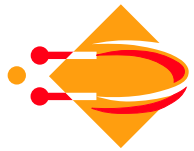
The distribution of wave lengths in the current yields the spectrum of the light.

The amount of energy per second yields the effect of the light.

The intensity and colour of light is a human interpretation.

Light is measured by essentially finding the sum of the energy of the photons in some time-space-frequency window.

Anatomy



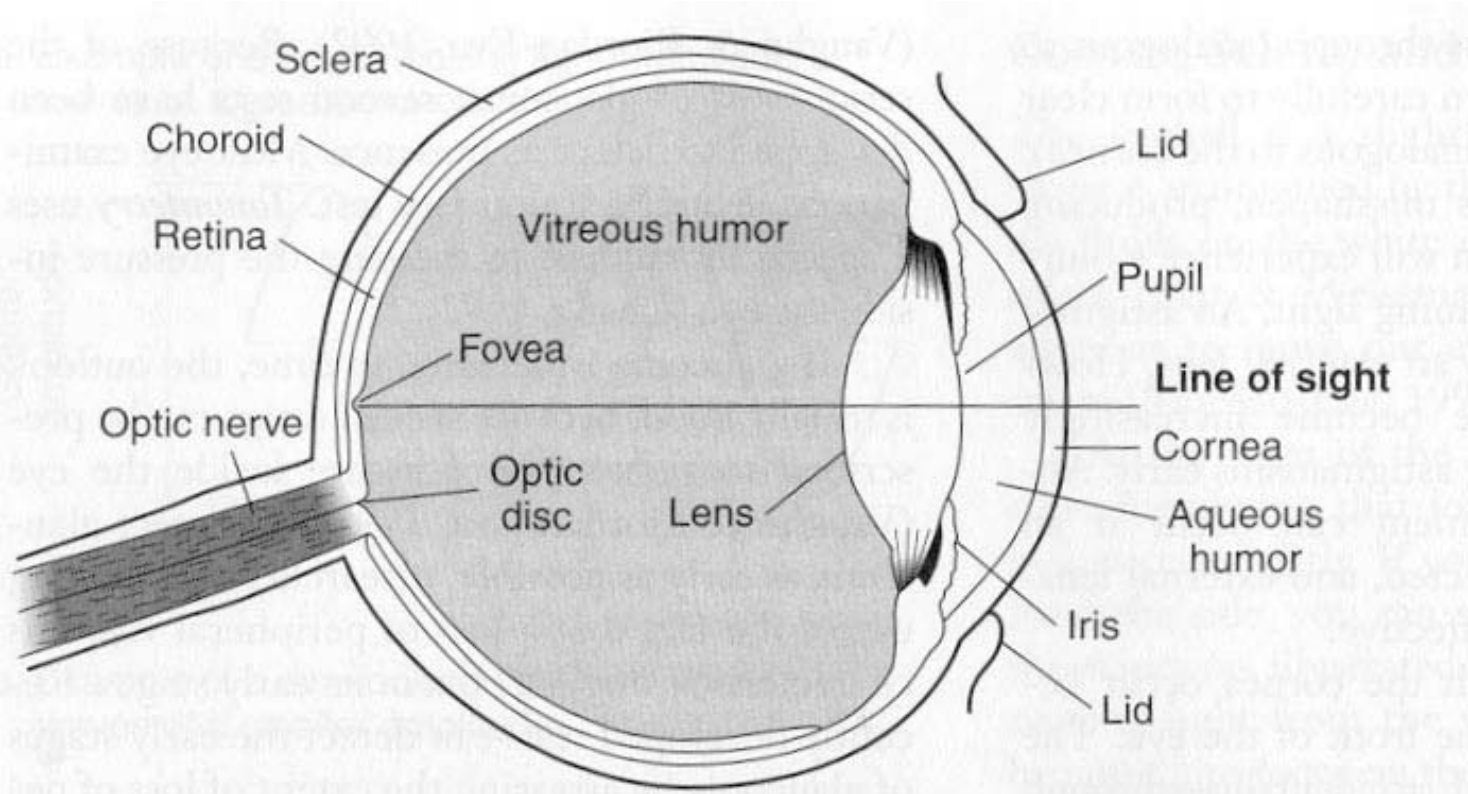
The eye: Focuses the light pattern on the retina.

The retina: Transforms light into electrical signals.

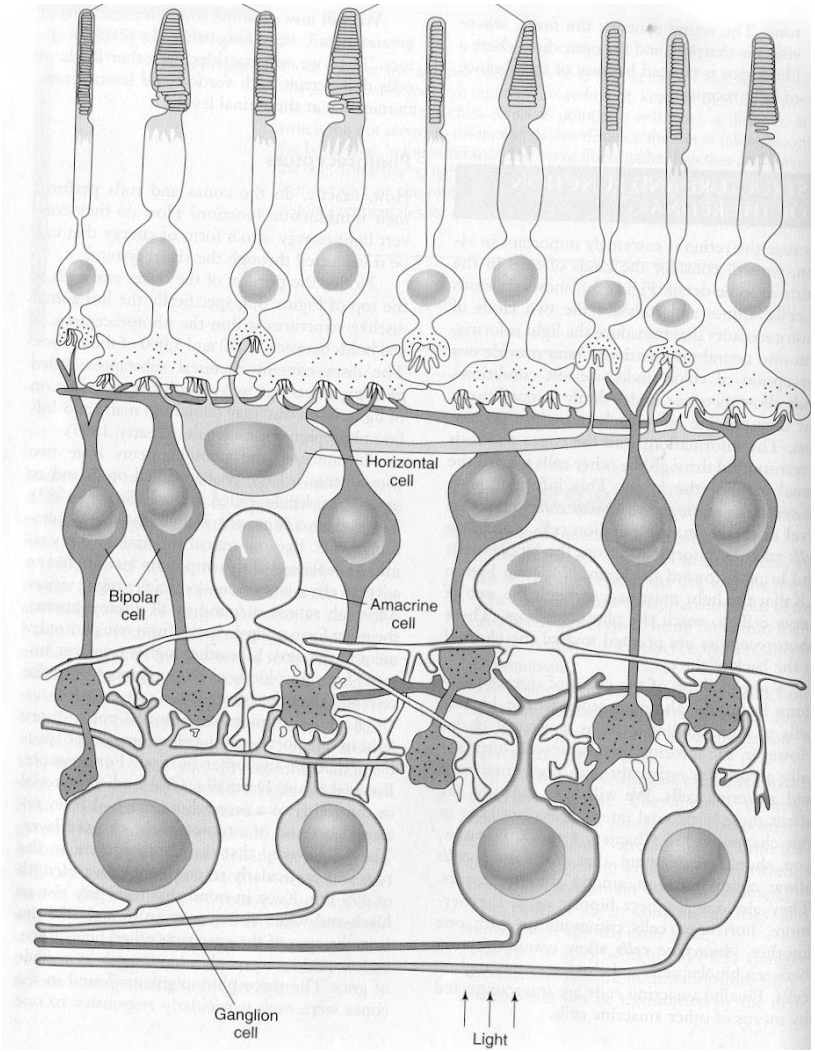
The LGN: Relay station to visual cortex.

The visual cortex: performs first visual "syntactical" interpretation.

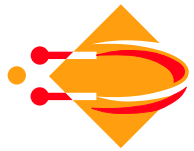
The eye


Sclera
Choroid
Retina
Fovea
Optic nerve
Optic disc
Lens
Vitreous humor
Lid
Pupil
Line of sight
Cornea
Aqueous humor
Iris
Lid

The retina


Horizontal cell
Bipolar cell
Amacrine cell
Ganglion cell
Light

Rods and cones



Characteristic	Cones	Rods
Vision	Colour	Black and white
Number/eye	5×10^6	10^8
Distribution	Most in fovea	Not in fovea
Cells per ganglion	Few	Many
Sensitivity	Poor	Excellent
Acuity	Excellent	Poor
Disc shedding	Evening	Morning
Dark adap.	Rapid	Slow

Rods and cones II

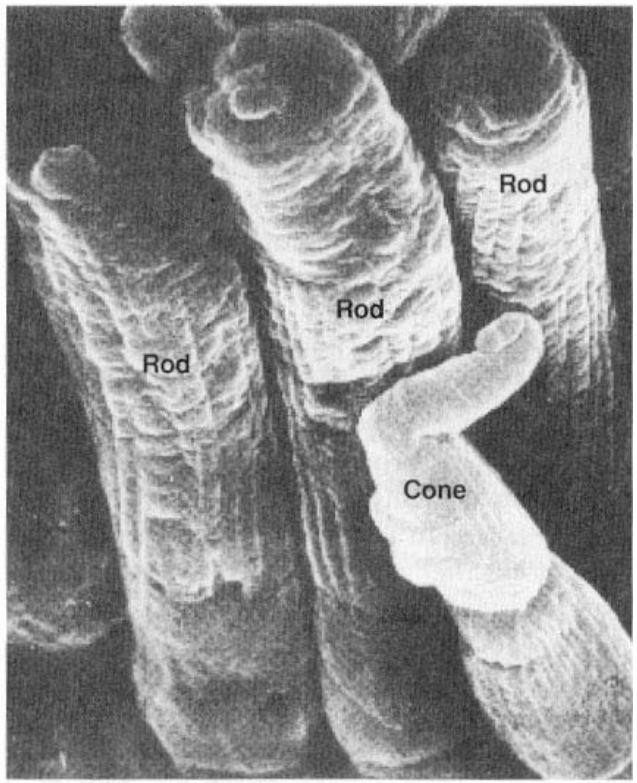
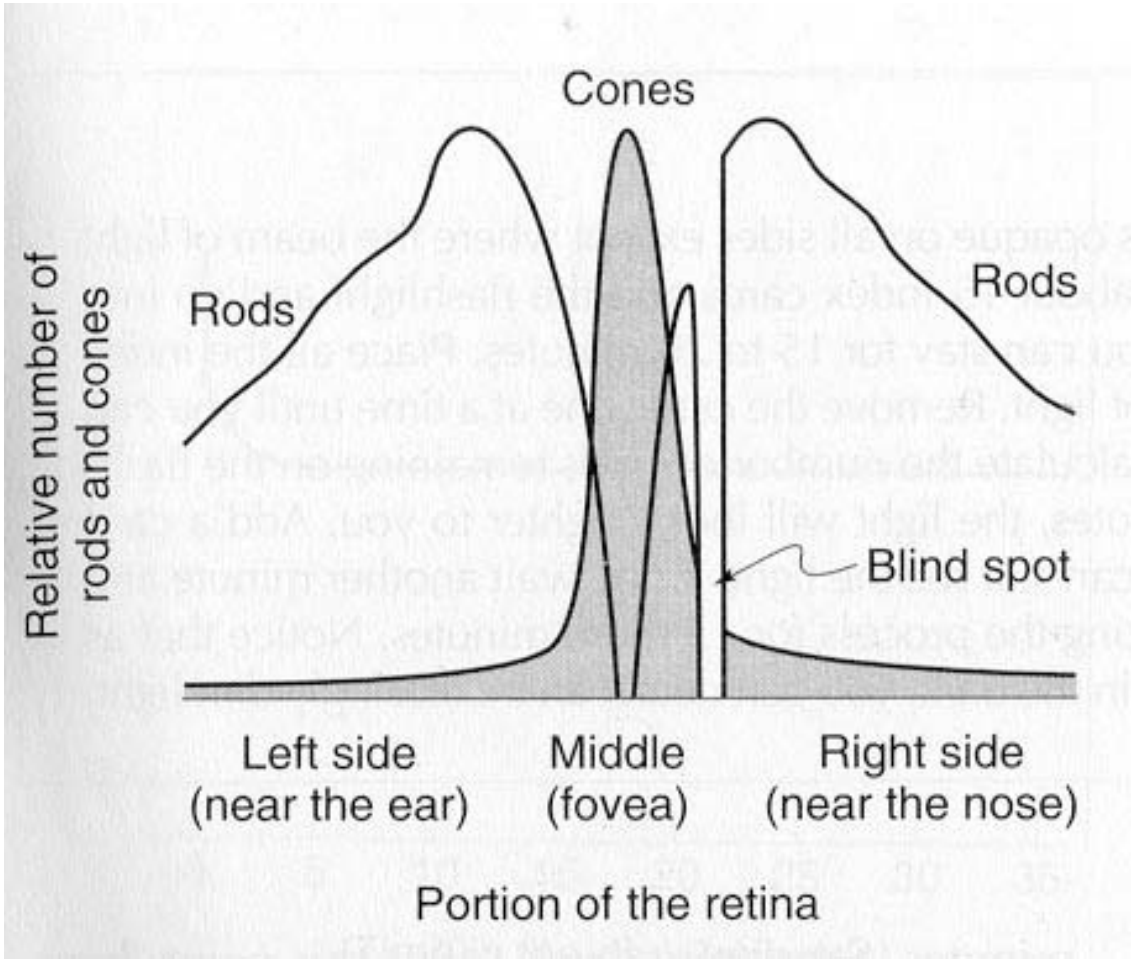
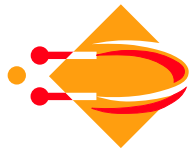


FIGURE 3.9 An electron microscopic view of rods and cones. (Photo courtesy of Dr. Frank Werblin)



Ganglion cells



The portion of the retina that via the bipolar cells stimulates a single ganglion cell, is called its receptive field.

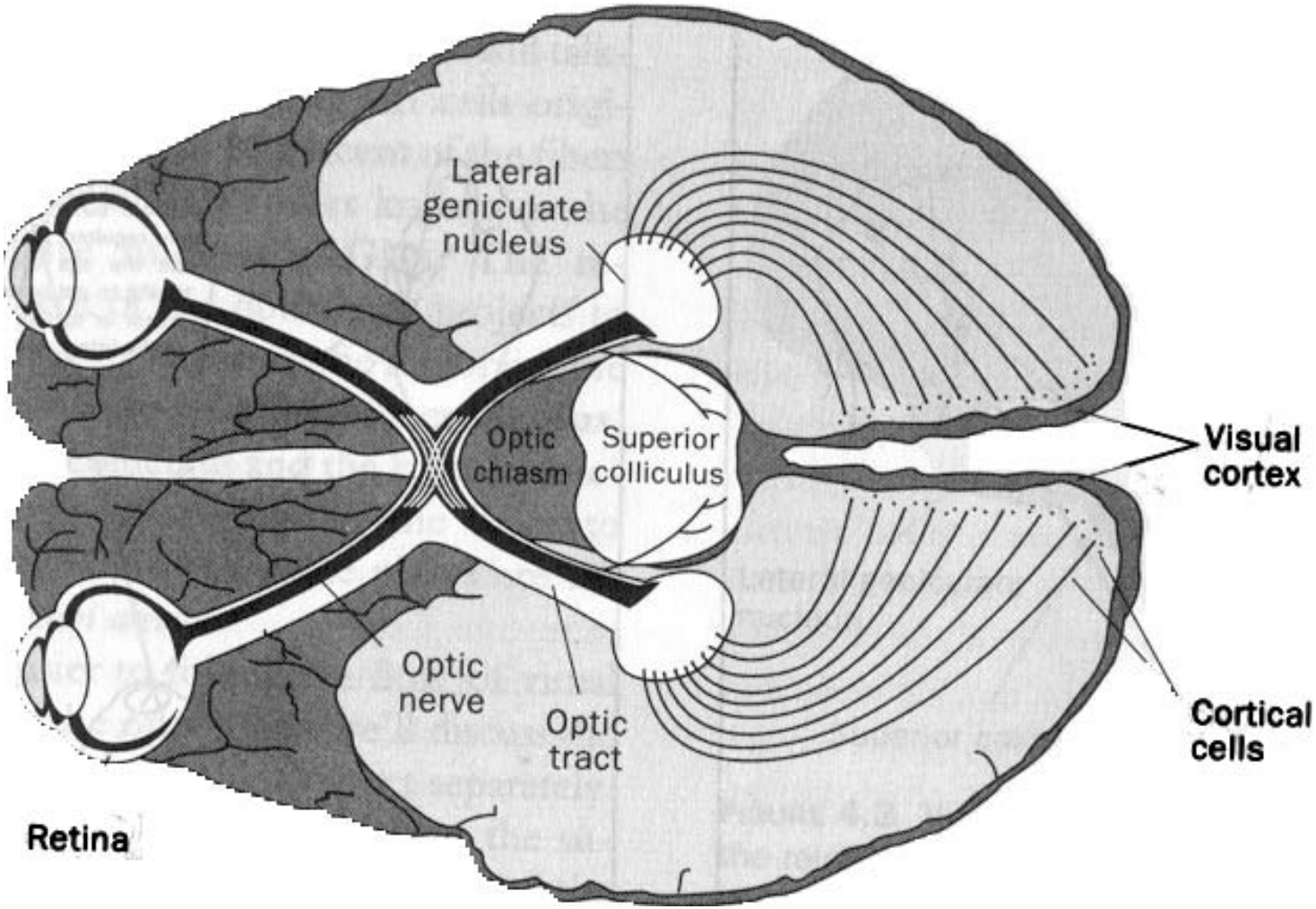
The receptive fields are small in fovea $\Leftrightarrow$ high resolution.

The receptive fields are large off the fovea.

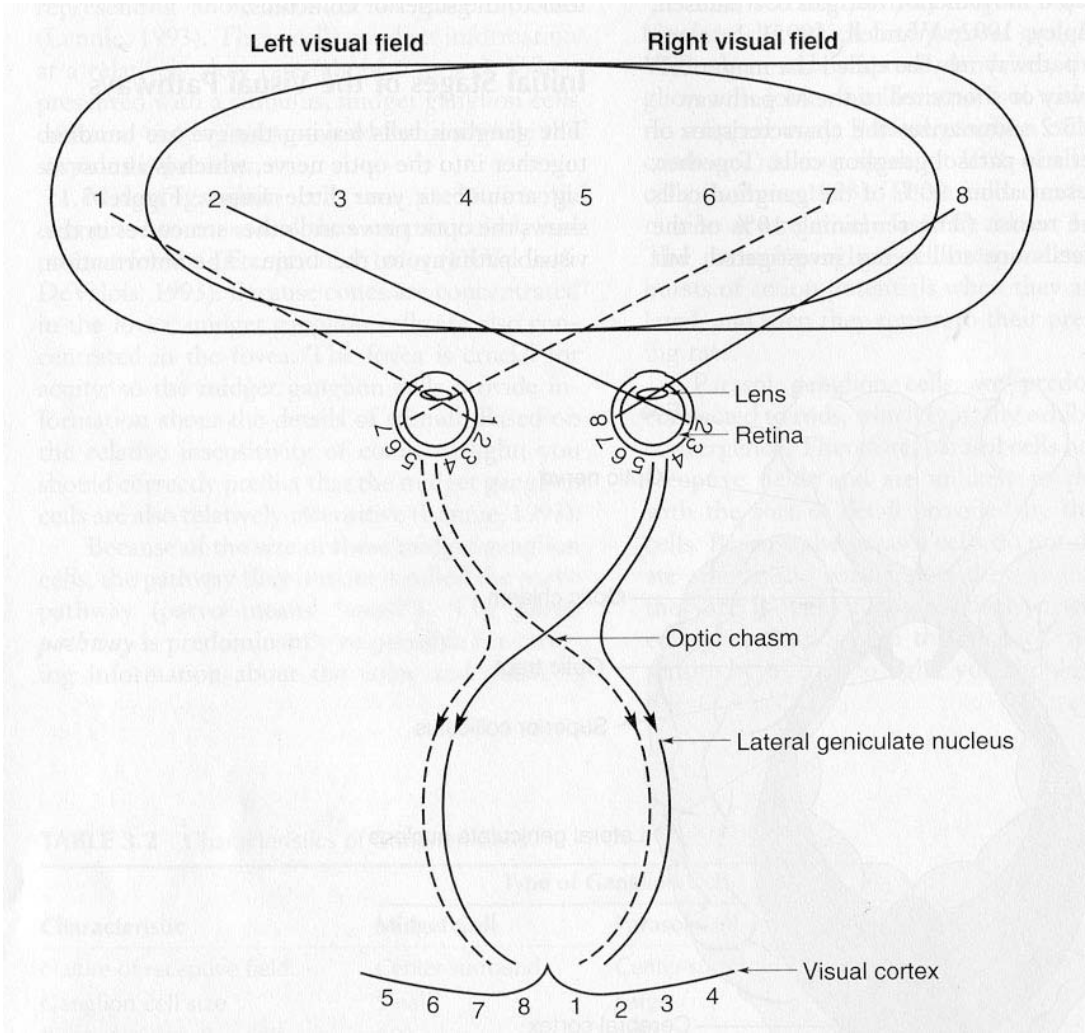
The size s may be modeled as $s = \log \frac{r}{r_0}$
r is the distance from fovea center
r_0 is the radius of the fovea

Midget ganglions (80%) are small, slow, low sensitivity, colour
Parasol ganglions (10%) are large, fast, high sensitivity, B/W

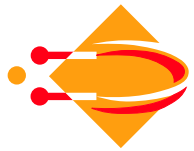
Visual pathway



Retinal mapping



Visual cortex



Primary visual cortex:

- Simple cells: lines and edges
- Complex cells: motion
- End-stopped cells of both kinds

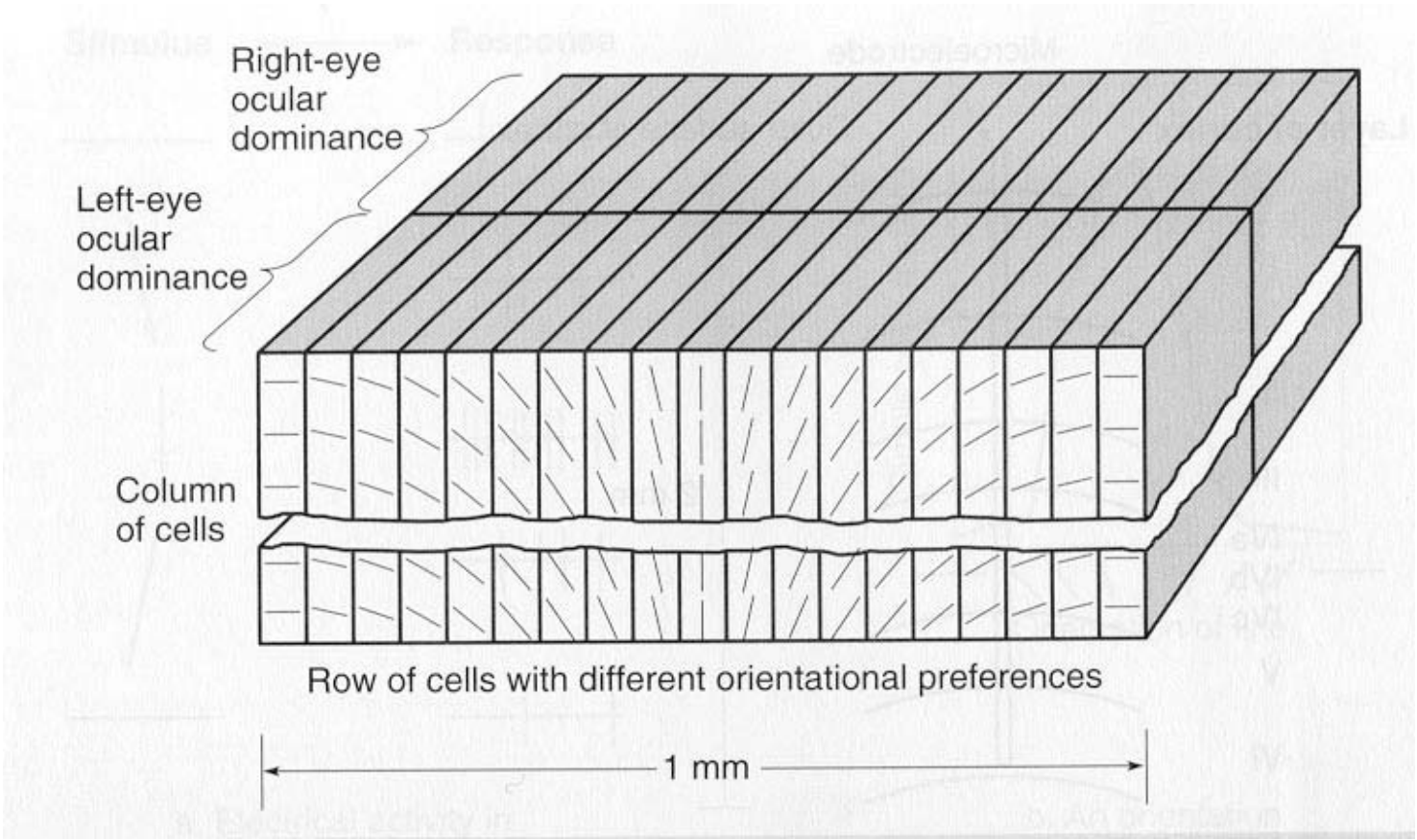
Secondary visual cortex:

- Colour perception
- Motion perception

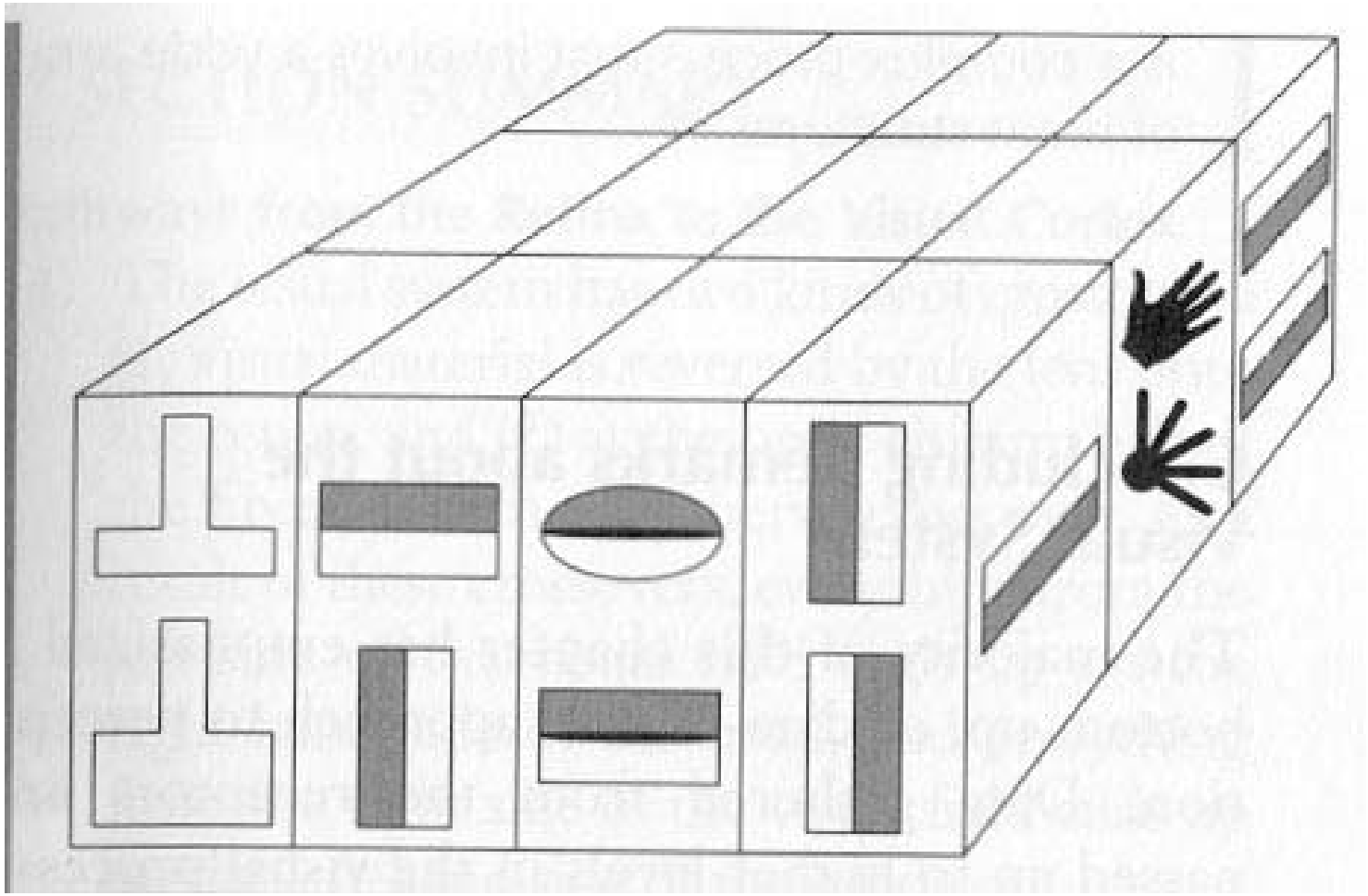
Inferior temporal cortex:

- Object recognition

Primary visual cortex

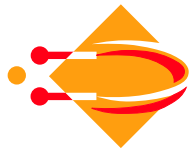


Inferior temporal cortex



Psychophysics

Adaption and edges



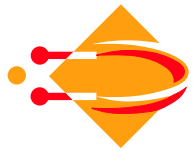
The eye adapts to the stimulus.

A constant scene disappears as sensation.

Micro movements make edges move slightly.

Only edges are left in a constant scene.

Resolution



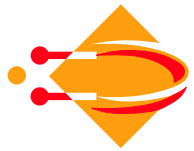
The resolution is limited by:

Focus blurr: accommodation focuses the lens

Retinal resolution: High in fovea, but demands more light

The resolution is measured by a Snellen chart

Groupings



Law of proximity

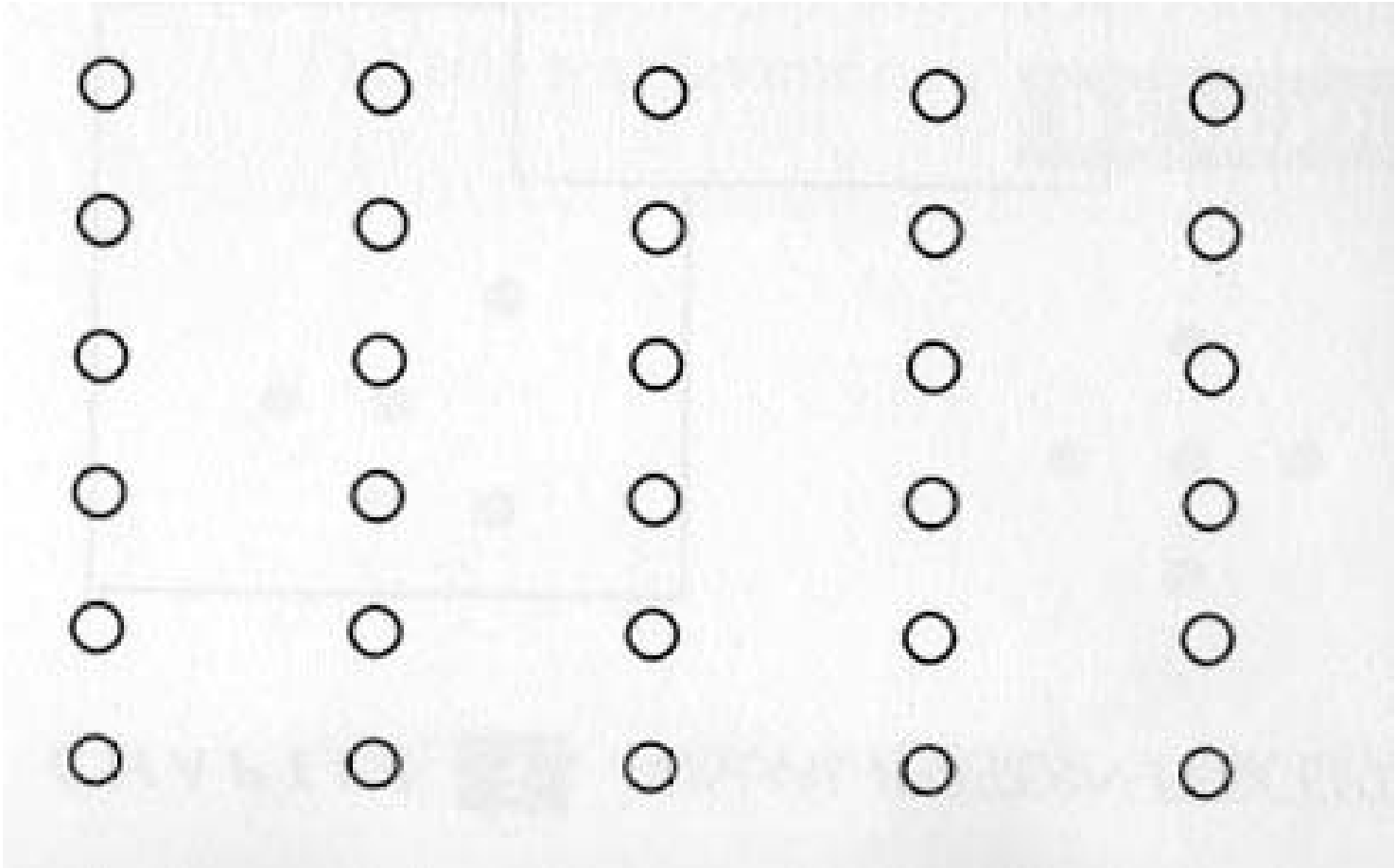
Law of similarity

Law of good continuation

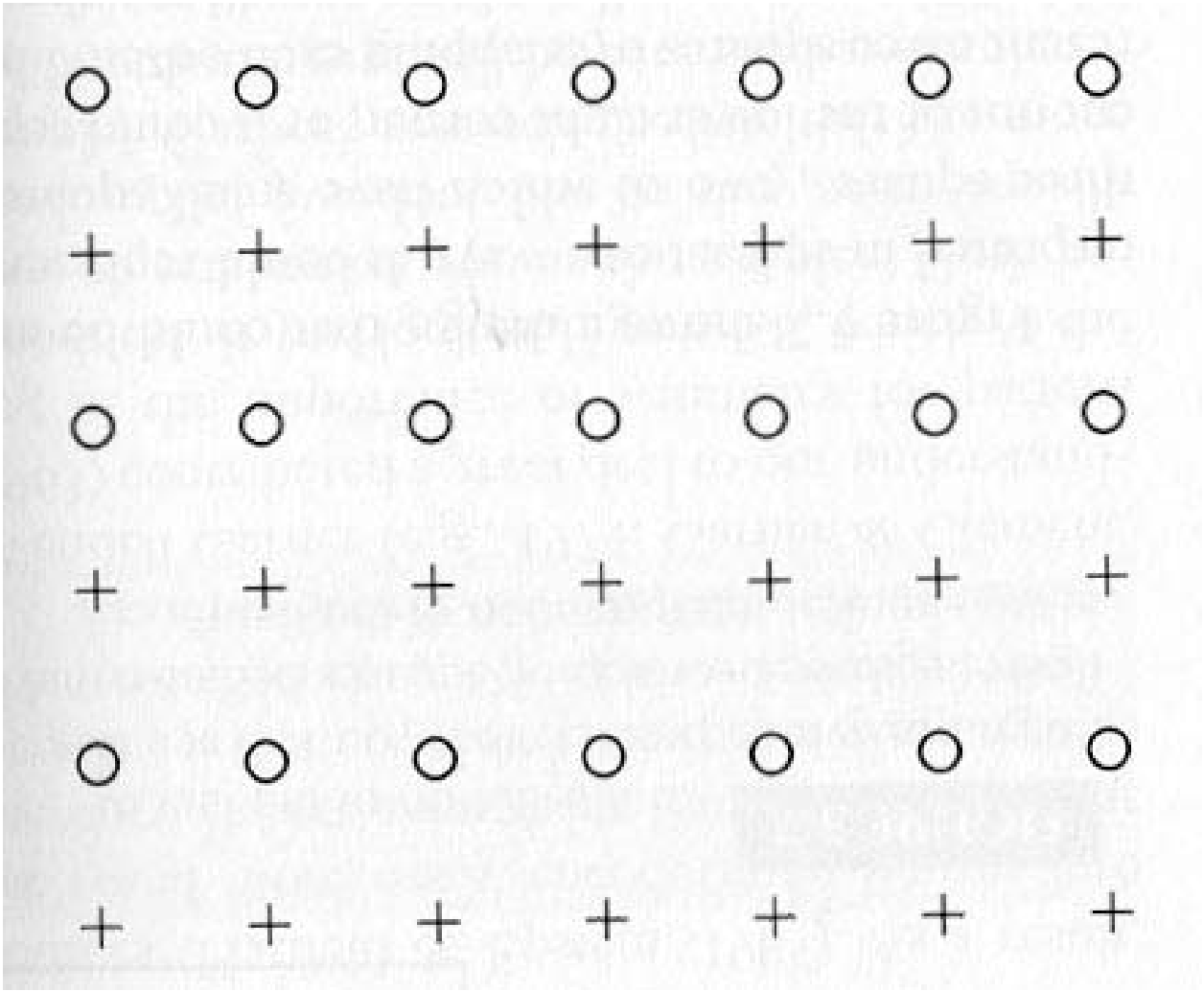
Law of closure

Law of common fate

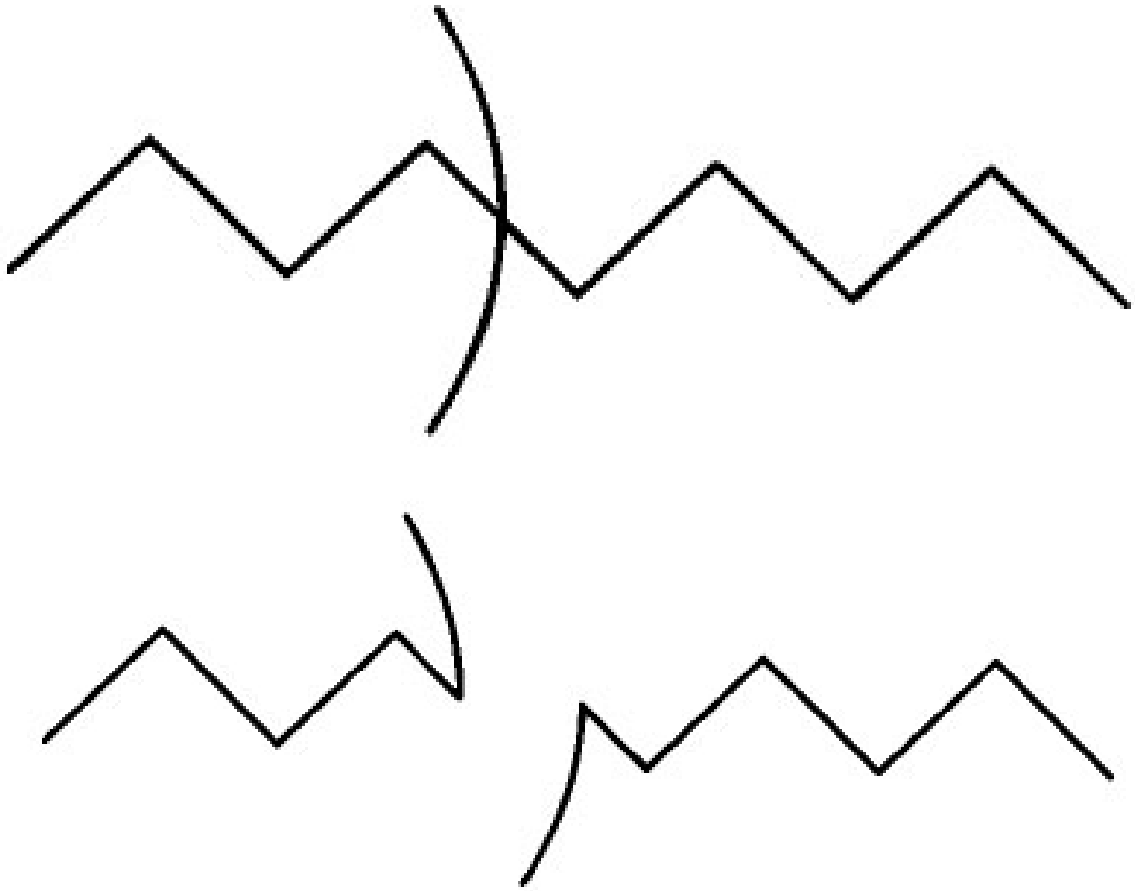
Law of proximity



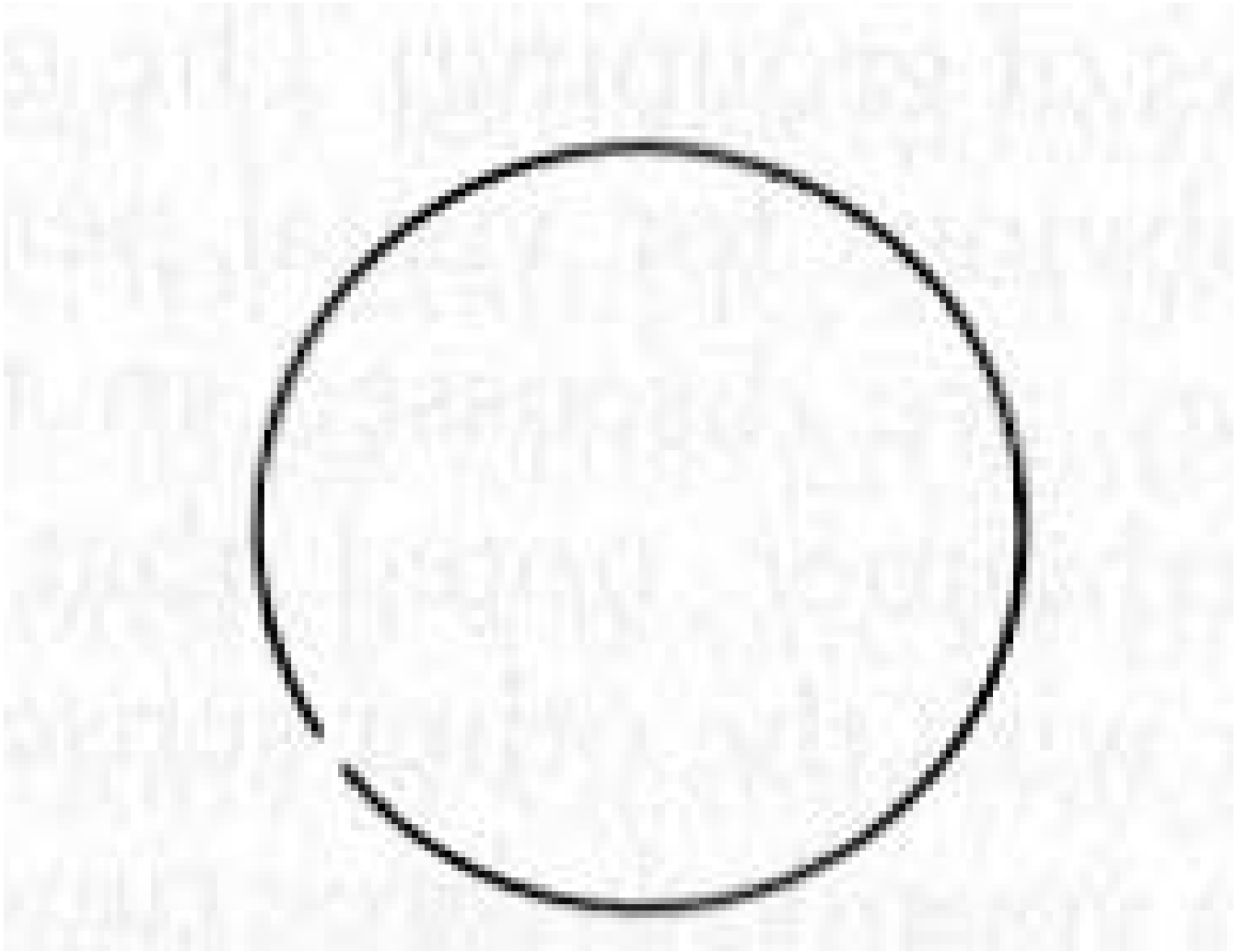
Law of similarity



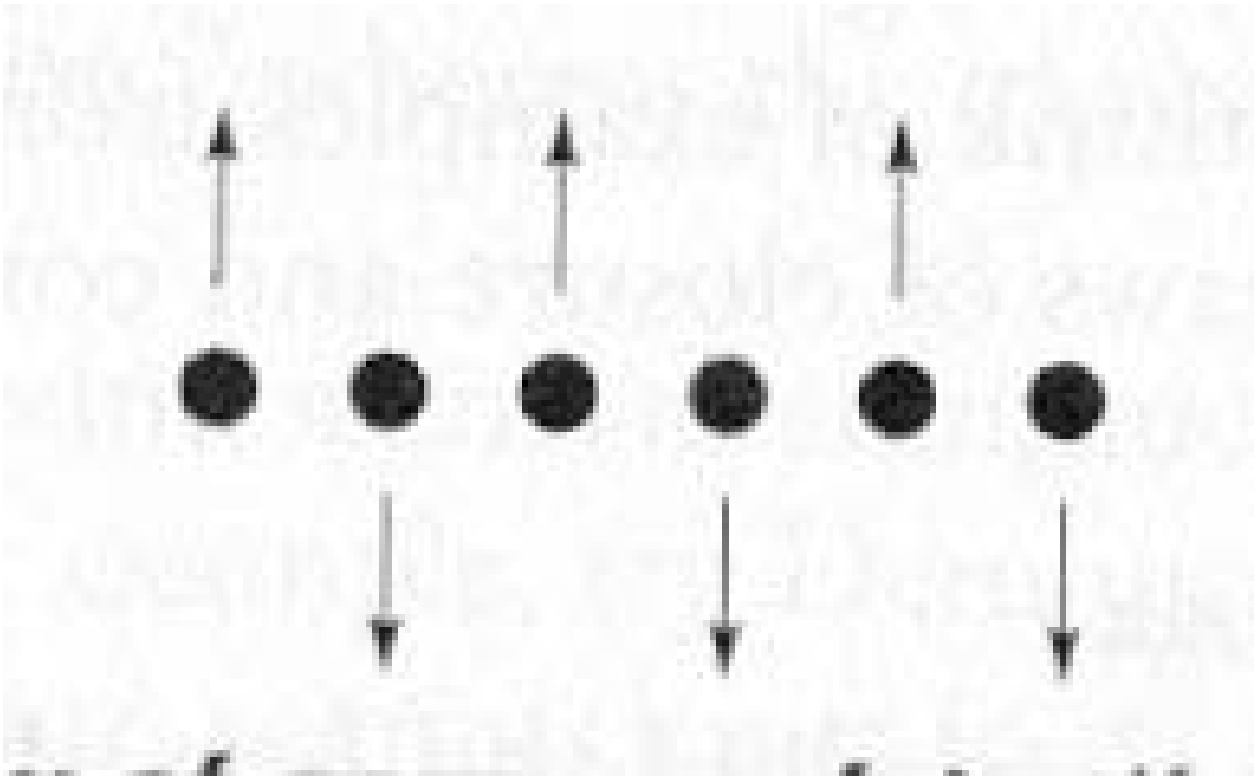
Law of good continuation



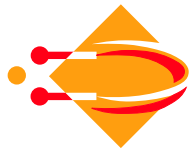
Law of closure



Law of common fate



Law of Prägnanz



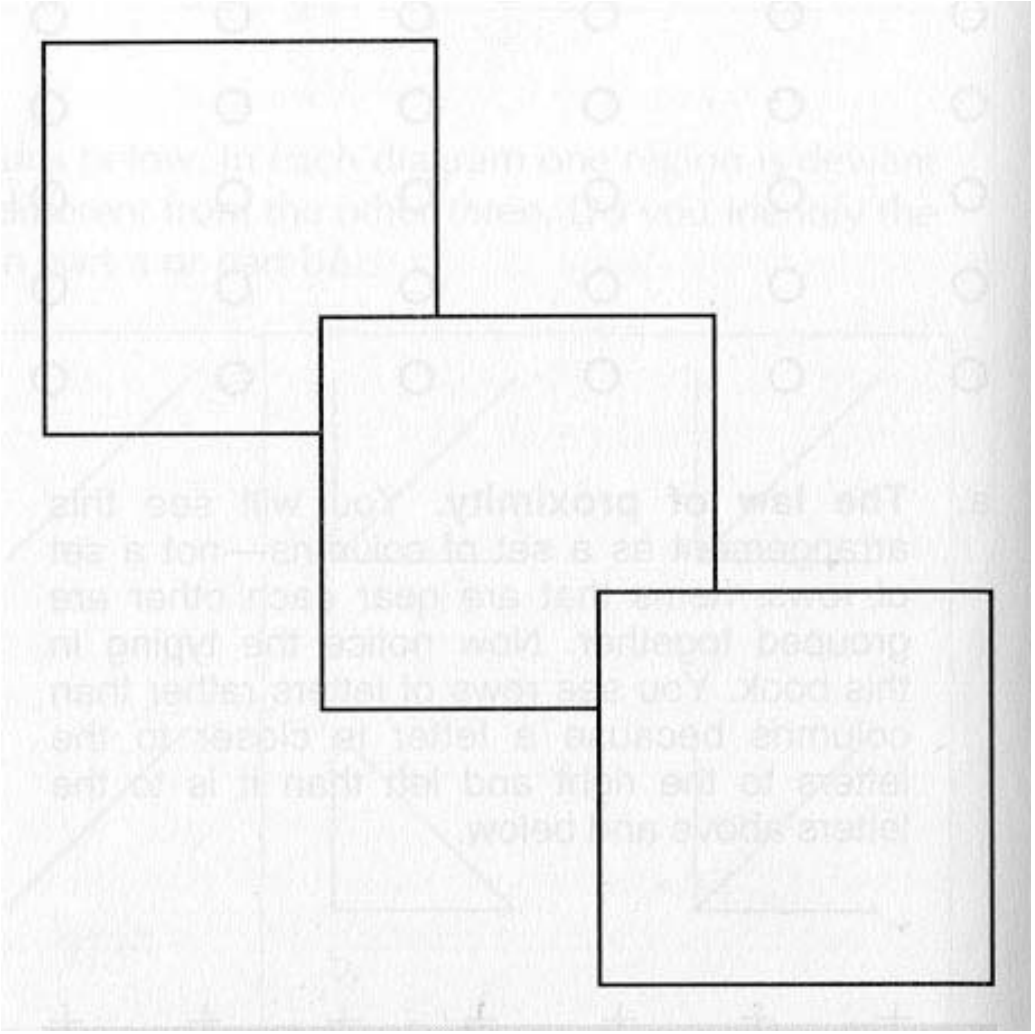
Of several geometrically possible organisations the one will actually occur which possesses the best, simplest, and most stable shape.

This is a perceptual version of Occam's Razor: Of several interpretations the simplest is most likely.

This is later formalised in the information theoretical approach Minimal Description Length (MDL) principle: The interpretation that can be communicated in the shortest message is the most probable.

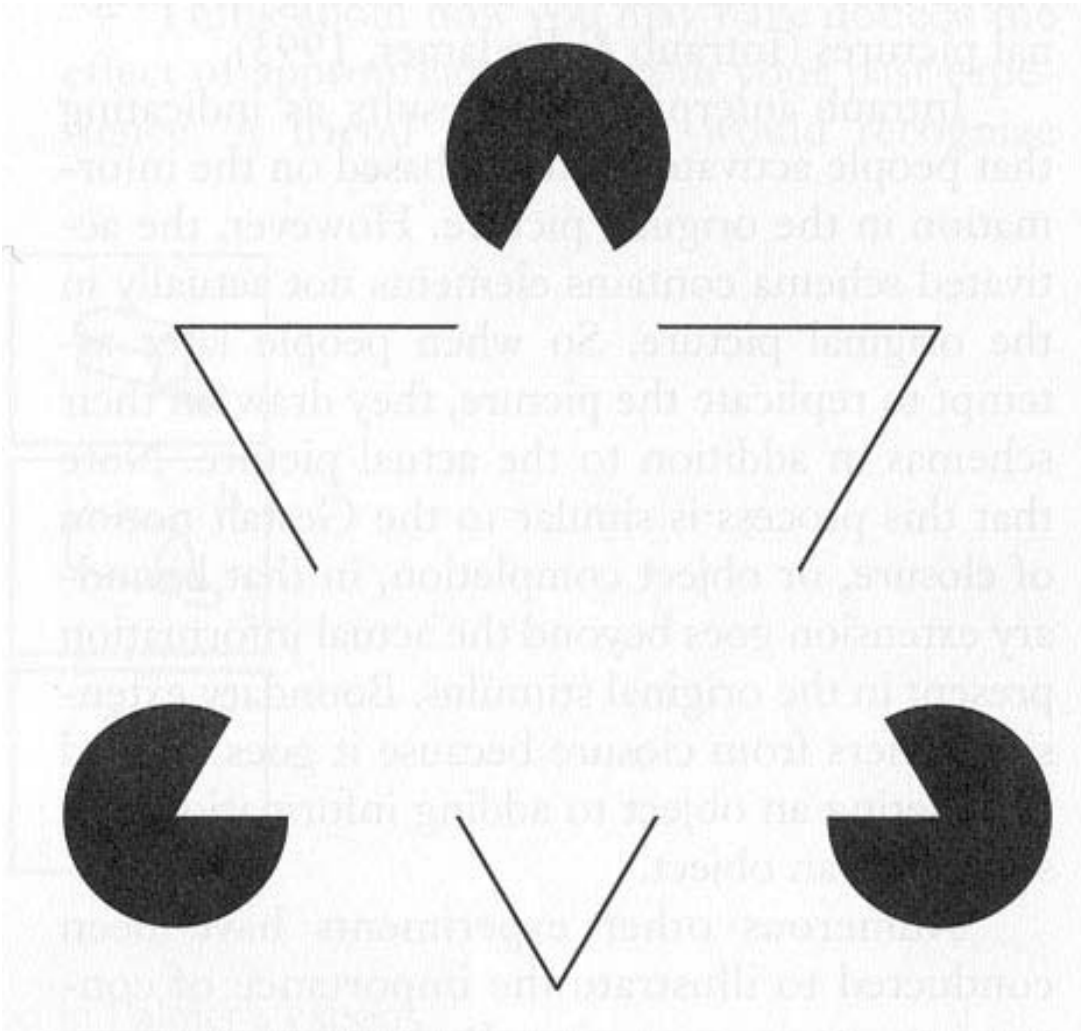
This is uncomputable since it is equivalent to the Kolmogorov complexity.

Law of Prägnanz II

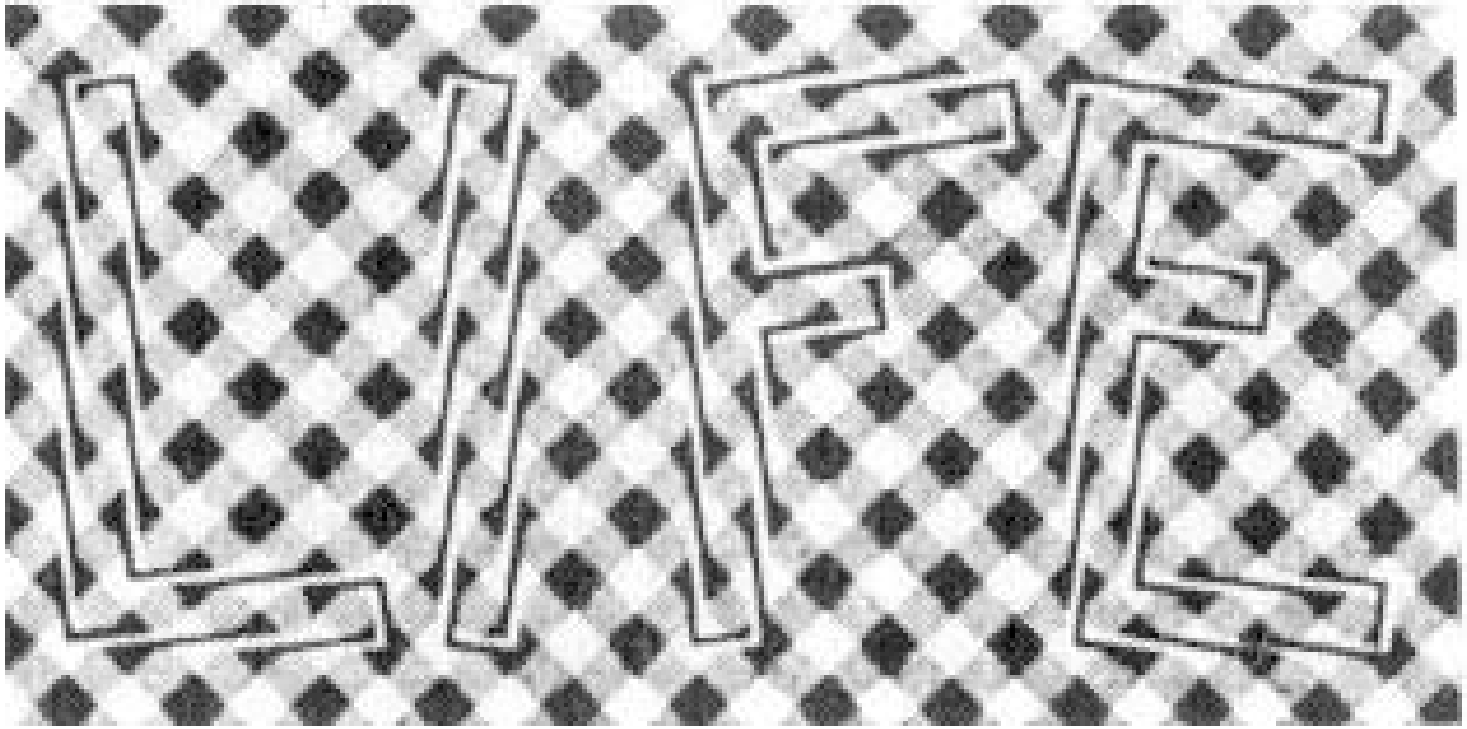


Law of Prägnanz III

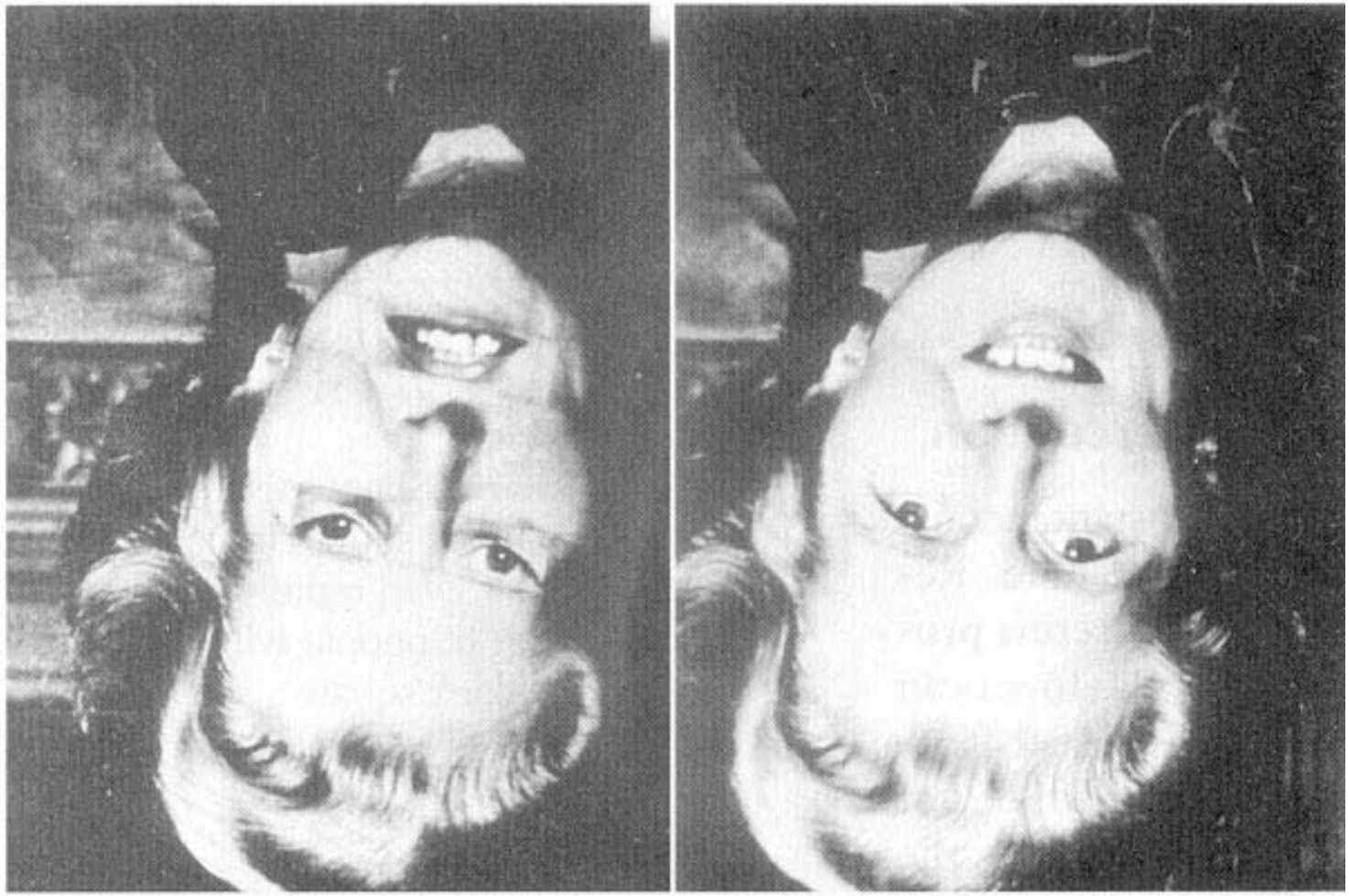
Illusory contours



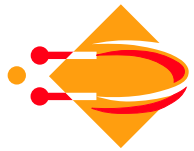
Shape $\leftrightarrow$ context



Orientation



Computer science



Analog video: Synchronization, interlacing

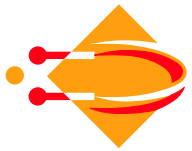
A/D conversion and digital images: CCD cameras

Image formats: Bits and bytes, header, ...

Image contents: Resolution, discretization, noise, histograms

Image analysis: inference, classical problems

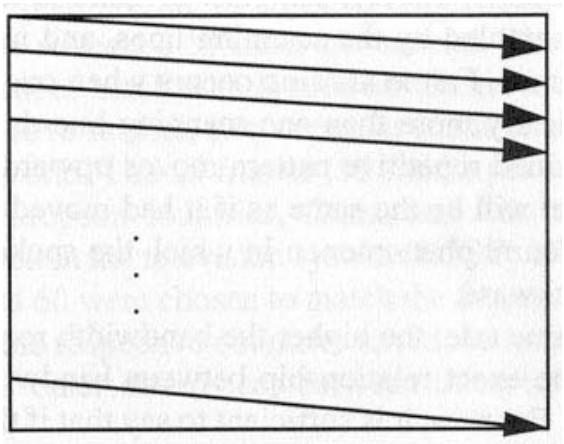
Analog video

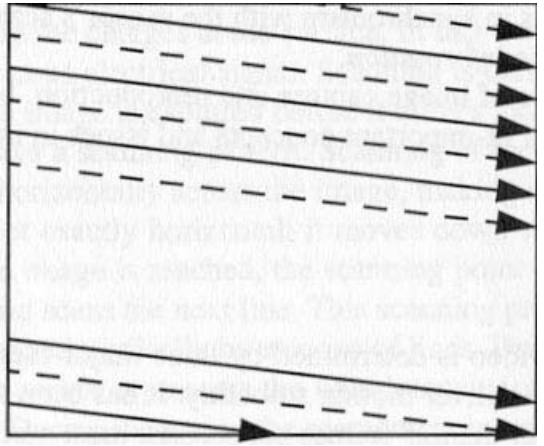


An analog video signal is separated into:

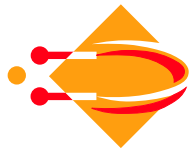
- row information
- horizontal blanc
- vertical blanc

Lines are mixed (interlaced) to improve temporal coherence.





Digital images



Digital images are created as A/D conversions of analog images by spatial sampling and quantization.

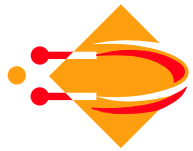
From analog video signals the number of rows is given.

Digital images are recorded by a CCD (Charged Coupled Device) chip. Array of receptive fields: pixels.

Until recently they were read out into a video signal and then later A/D converted.

Now we begin to have direct digital video.

Bits and bytes



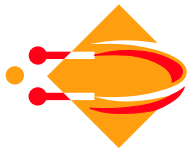
Representation in a computer is in binary numbers:

Decimal	64	32	16	8	4	2	1
Binary	1 000 000	100 000	10 000	1 000	100	10	1

A binary digit is called a BIT

A byte contains 8 bits [0;255[

Image format



Typically an image file contains

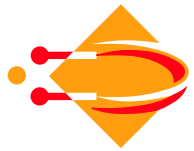
Header:

- Rows
- Columns
- Pixel type
- Time stamp
- LUT
- Compression information
- ...

Body:

- Pixel information

Image contents



The quality of the image is limited by:

- The effective resolution
- The quantization
- The noise level

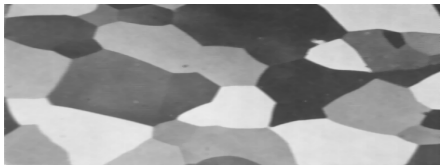
The use of the quantization bins is described through the histogram:

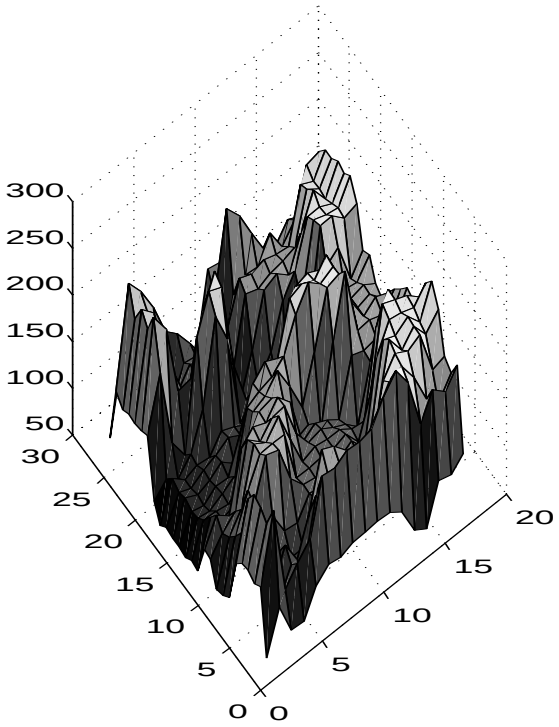
- a table telling: for each pixel value, how many pixels has this value

Histogram = Number of pixels(pixel value)

The histogram contains no geometrical information.

Histograms




300
250
200
150
100
50
30
25
20
15
10
5
0
0
5
10
15
20

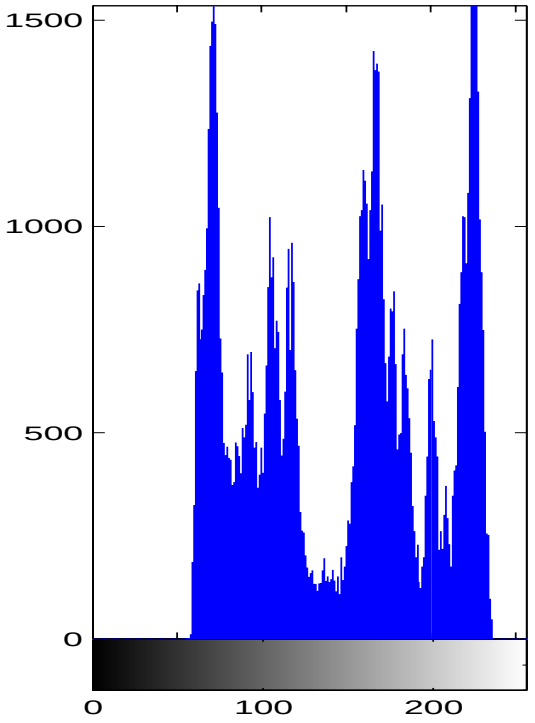
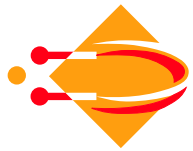

1500
1000
500
0
0
100
200

Image analysis



A semantic description of an image cannot be constructed solely on basis of pixel values.

We need:
Models describing how events express themselves in pixel values
Expectations of what we see

This is formalised either as statistical Bayesian estimation or as information theoretical inference using Minimum Description Length principle.

Models + Expectations $\Rightarrow$ Computational theory

Algorithm does not follow directly!