Computer Vision
INM460/IN3060

Lecture 1
Introduction and the human visual system

Dr Giacomo Tarroni
Slides credits: Giacomo Tarroni, Sepehr Jalali

Lecture capture

Lecture capture is on: avoid area this area if not you don't want to be recorded!

Timetable

Activity	Start time	Finish time	Room	Week commencing	Activity given by
Lecture	3 pm	4:50 pm	C314	20/01/’20 – 17/02/’20, 02/03/’20 – 30/03/’20	G. Tarroni (& guests)
Lab (UG)	5 pm	5:50 pm	EG01	Same as above	K. Ryan E. Jimenez-Ruiz
Lab (PG)	5 pm	5:50 pm	EG06	Same as above	A. Galkin O. Galkina

Contacts

- My contacts:
  - For admin-related enquiries: giacomo.tarroni@city.ac.uk
  - For questions on the lecture materials: please come to my office during the drop-in hours: Wednesdays 2 – 4pm, A302B

- Teaching support:
  - Kevin Ryan (kevin.ryan@city.ac.uk)
  - Alex Galkin (oleksandr.galkin@city.ac.uk)
  - Olga Galkina (olga.galkina@city.ac.uk)
  - Ernesto Jimenez-Ruiz (ernesto.jimenez-ruiz@city.ac.uk)

Materials

- All the materials (lecture slides, labs tutorials, and extra resources) will be available in Moodle (IN3060/INM460 Computer Vision)

- No books are compulsory, but you can find some in the Reading List (https://city.rl.talis.com/index.html) for Computer Vision. Some examples:
    - "Computer vision: algorithms and applications" by Richard Szeliski
    - "Pattern recognition and machine learning" by Christopher M. Bishop
    - Other more practical books on deep learning for CV

- However, CV is an extremely fast-moving field. If you want to do this in your career:
    - Look for online courses on specific subjects (e.g. Stanford's http://cs231n.stanford.edu/)
    - Remain updated on most cited online papers on arXiv

Our expectations of you

- Attend all lectures and all lab sessions: attendance is recorded for the lectures. Register during the first 30 mins on https://moodle.city.ac.uk/mod/attendance/index.php?id=35000
- Arrive on time
- Please no mobile phones during lectures or labs
- Laptops are of course allowed (but please no chatting, reading news, watching Netflix, …)
- Prepare by reading slide handouts and doing lab tutorials **as the module progresses**
- Participate:
  - During lectures: by discussing/asking questions (stop me at any time!)
  - At home: by working out in-class examples and lab tutorials
- If you want to implement deep learning algorithms, find a pc with a CUDA compatible NVIDIA GPU (https://developer.nvidia.com/cuda-gpus): City should have some resources

Assessment

- Marks will be based on:
  - 50% coursework (programming task)
  - 50% final examination (written test)

- UG and PG students will have different coursework and final exam

- Use of sample code from the Internet in your coursework is encouraged but **must** be referenced

- The content of lectures 2-10 can be in the final exam (potential exceptions will be clearly stated)

- Dates and details to follow

Coding

- Lab tutorials will be mainly based on Matlab
- Matlab is installed already on lab machines
- If you plan on using your own computer:
  - Install Matlab 2018a (or more recent) ASAP. Go to https://www.city.ac.uk/current-students/it-support, click on “Resources and Facilities”, log in, and click on “SPSS, Matlab, NVivo, OxMetrics”, click on Matlab and follow the instructions from there
  - Make sure to install the Computer Vision System and Neural Networks toolboxes
  - Install AlexNet, ResNet, VGG16, VGG19, GoogLeNet add-ons (this should happen by default with the previous toolboxes)
- Familiarise with Matlab’s Computer Vision Toolbox (http://uk.mathworks.com/products/computer-vision/): visit to access help pages, tutorial, forums, etc.
- Code samples will be available both on lecture materials and lab notes on Moodle

Why Matlab?

- The most popular languages/frameworks for computer vision are Matlab and Python. Why we will teach CV using Matlab:
  - Matlab is easy to program
  - It is an integrated development environment (IDE): straightforward interrogation of variables in memory, execution of code during debugging
  - It has many built-in features (like plotting) that are indispensable
  - It has an extensive help section (with examples and tutorials)
  - Well maintained between different releases
- However, Python is
  - Faster
  - Free
  - Better geared for deep learning applications
- You are highly encouraged to learn Python and familiarise with deep learning frameworks, including TensorFlow by Google and/or PyTorch by Facebook. Check courses on Udemy or books in the Reading List

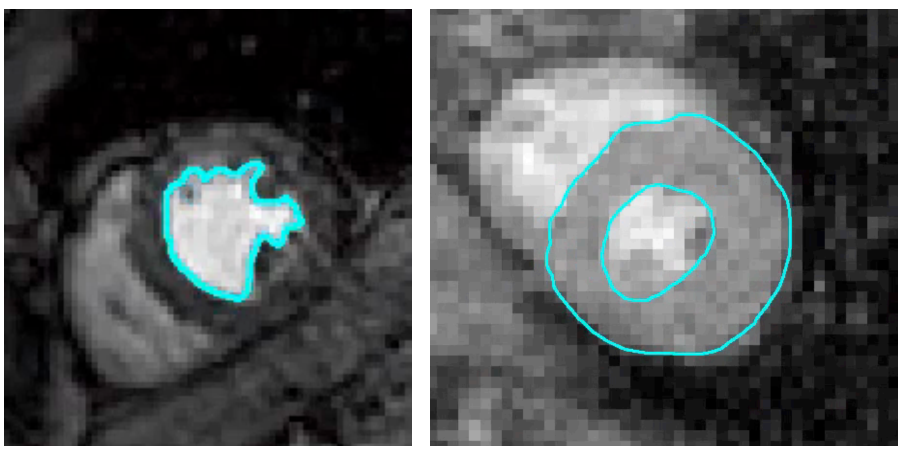
Preprequisites

- Math:
  - CV make is a math-heavy subject
  - Effective knowledge of linear algebra (e.g. lines and planes in space, representations and operations between matrices) and basic calculus (e.g. derivatives, differential equations) will be taken for granted
  - On Moodle, you will find some materials (e.g. Math Primer) for a recap of the main concepts as well as a test
- Programming:
  - Lab tutorials will be in Matlab
  - The first lab tutorial will be a recap of the basic notation and usage of Matlab language
- Machine learning (ML):
  - Modern CV is based on ML algorithms
  - No previous knowledge of ML methods is expected, but basic understanding of ML concepts (e.g. classification, regression, training/testing, generalization, overfitting) will be only briefly introduced

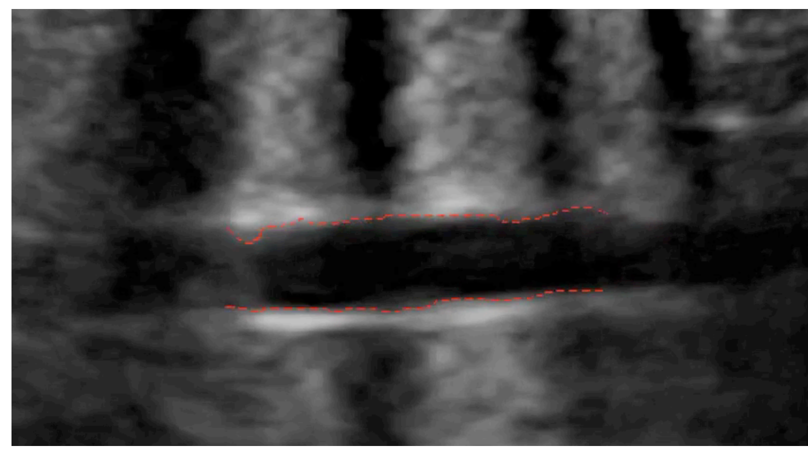
My background

Actually, not a computer scientist!

- PhD in Bioengineering from University of Bologna (Italy):
  Automated analysis of first-pass myocardial perfusion sequences in cardiac MRI
- Post-doc at University of Padova (Italy):
  Automated analysis of the aorta from fetal ultrasound images



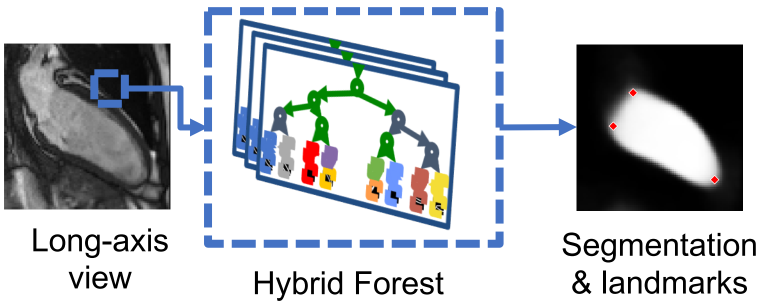
Cardiac perfusion MRI



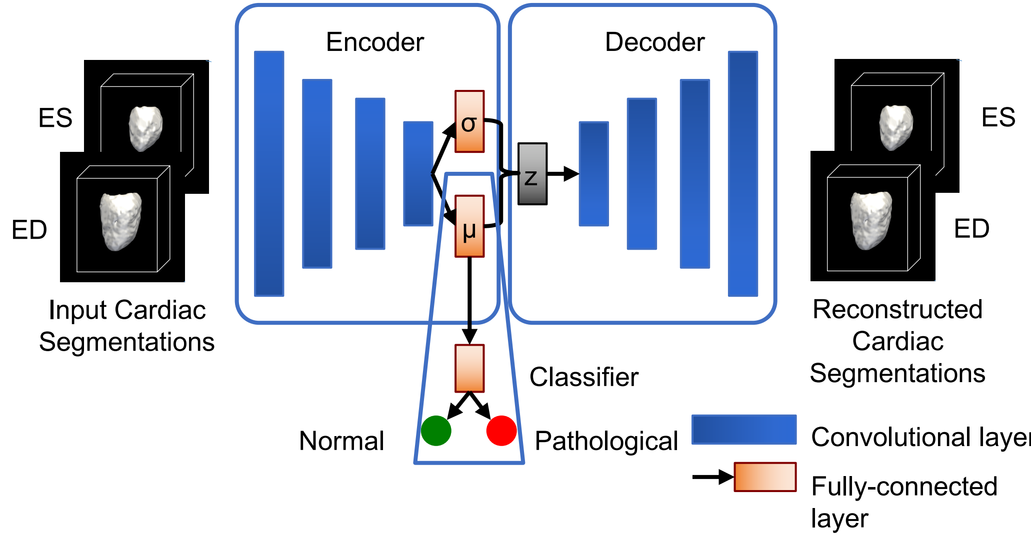
Fetal ultrasound

My background

- Marie-Curie Fellow at Imperial College London (U.K.): Developed automated machine learning techniques for cardiac MRI
- Lecturer in Artificial Intelligence at City, University of London (U.K.): Same as above, and more!



Segmentation & landmark detection for cardiac MRI



Variational autoencoder neural network for interpretable shape classification

Individual projects

- Individual projects are available. Some topics:
  - Detection of structures in medical images
  - Detection of anomalies/artefacts in medical images
  - Image translation between modalities

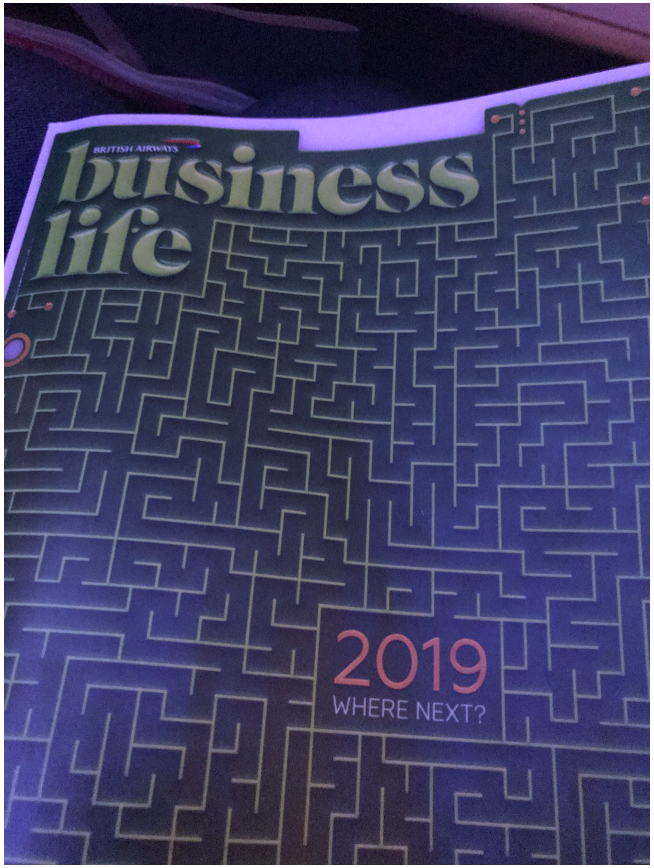
- Sepehr Jalali (the previous year’s lecturer) is also taking up students for individual projects:
  - Contact him if interested: Sepehr.Jalali.1@city.ac.uk
  - You can find his introductory lecture for last year here: https://www.youtube.com/watch?v=DhUusoMsnMI

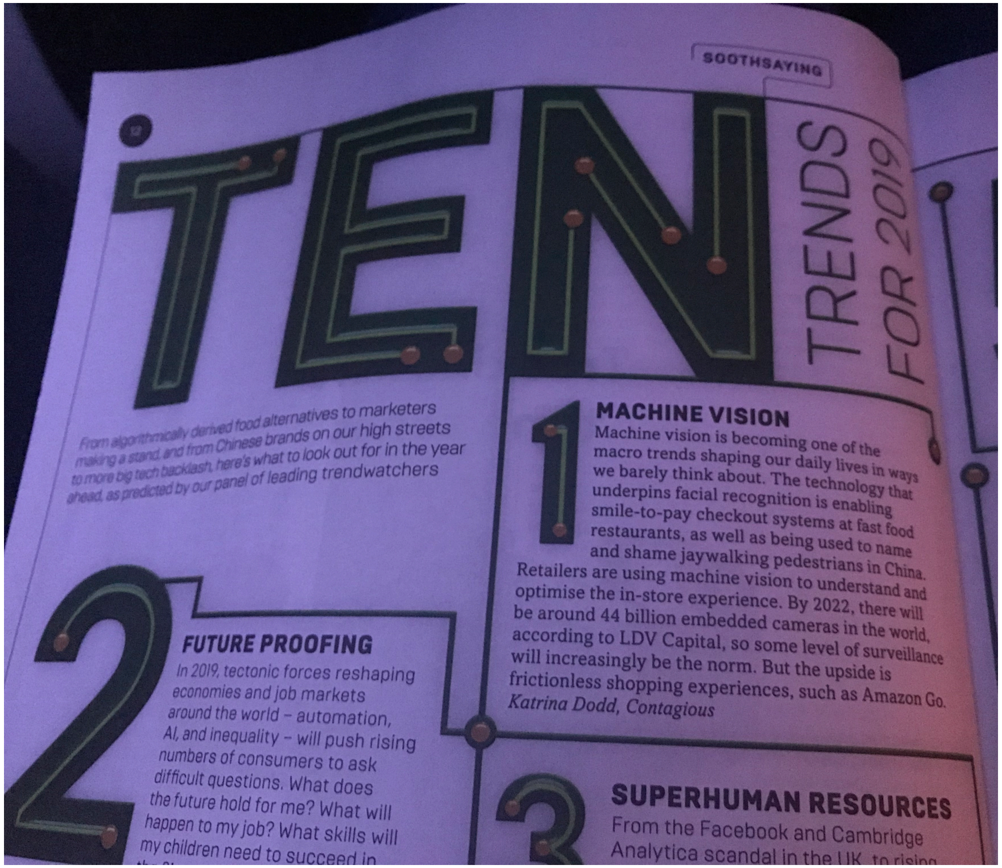
Computer vision

Computer vision (CV) is a scientific discipline that studies how computers can efficiently perceive, process, and understand **information** from visual data such as images and videos

Computer vision

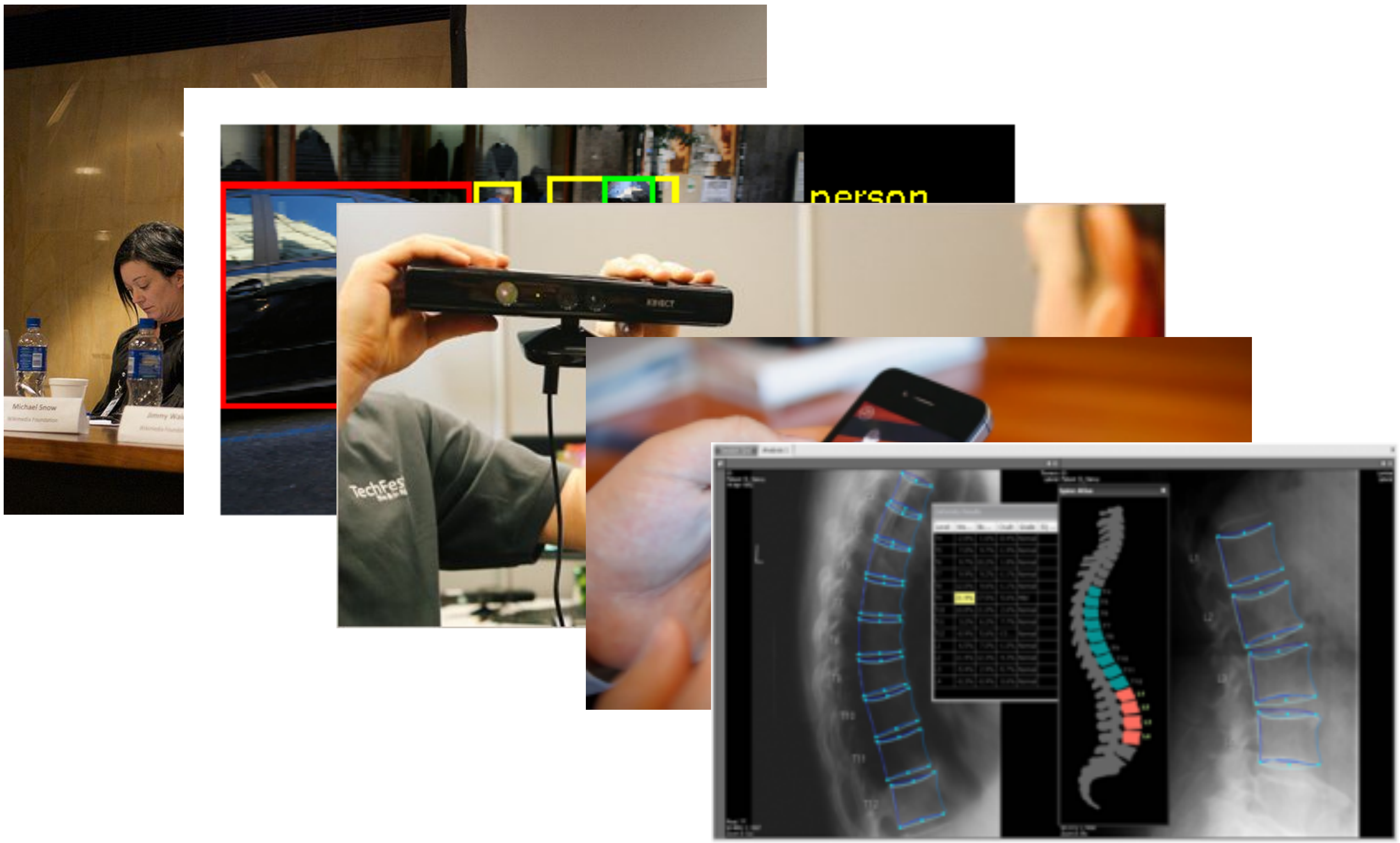
Very sought-after skill in the market at the moment





Computer vision

Some examples:



What you will learn in this module

Lectures:

- Human visual system
- Digital representation of an image
- Image transformations and filtering
- Edge detection
- Image segmentation
- Feature extraction and description
- Image matching through interest point detection
- Image classification (including: SVMs, NNs, CNNs)
- Object detection
- Instance segmentation
- …

What you will learn in this module

Lab tutorials:

- Learn how to implement the techniques described during the lectures
- Hands-on activities

Overview of today's lecture

- Why do we study biologically inspired vision?
- Human brain
- Human visual system
  - Neurons
  - Simple/complex cells
  - Structure
  - Illusions and limitations
- Perception
  - Colour perception
  - Attention
  - Object recognition
  - Gestalt Laws
  - Depth perception
  - Motion perception

Common CV task

Image classification

- Identify the class of the object shown in the image

Classification

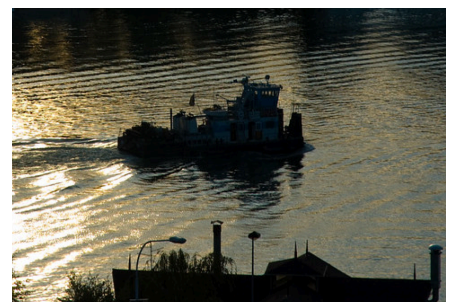
Class label: "Cat"

Image classification

Humans are very good at this task:





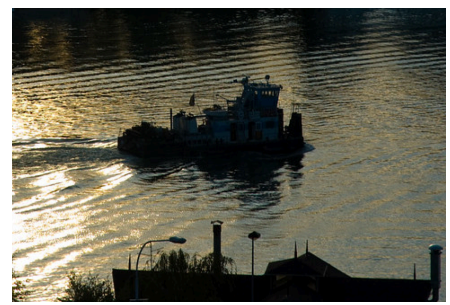
"…boats"

Image classification

However, these images look extremely different from each other





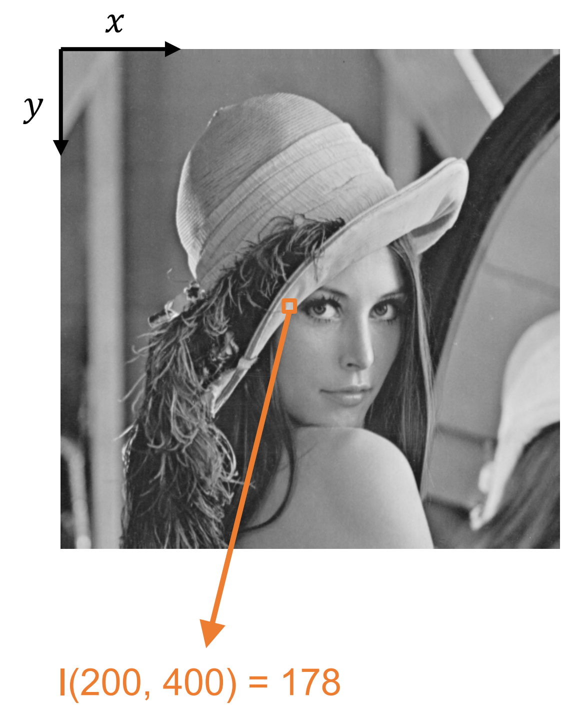
Human visual system enables a detection invariant to differences in viewpoints, color, scale, background, occlusions, cropping, shapes, rotation, …

Image representation

- A digital image I represents a scene with a 2D matrix of pixel values at (x, y) locations
- A 12MP image (with colour) has 4000x3000x3: 36,000,000 values
- If each value is represented with 8 bits (e.g. “24 bit RGB”), the range for each value is 0÷255
- Total number of possible combinations: $256^{36,000,000}$: image space is absurdly big!

Consequences:

- Images are difficult to analyze (e.g. classify)
- Different lighting or camera properties can cause two images of the same scene to have drastically different pixel values



Why biologically inspired vision?

Because it is (usually) great: learning how humans perceive images can enable the design of robust computer algorithms

- How much do we know about human visual system?
- What are the pros and cons of human visual system?
- What are the variations in the images we want to be robust against?
- How can we effectively create a biologically-inspired computer vision algorithm?

Studying the brain

- **Neuroscience**
  The scientific study of the nervous system (traditionally, biology-based)

- **Psychophysics**
  The study of the relationship between perception (psycho) and stimulus (physics)
  - Qualitative Methods: basic description of what a person perceives
  - Quantitative Methods:
    - Absolute Threshold: adjusted continuously until observer detects it
    - Difference Threshold: smallest detectable difference in two stimuli
    - Magnitude Estimation: compare the stimulus to a standard
    - Reaction Times: time from presentation of stimulus to observer's response is measured

Visual Neuroscience labs here at City University (Health Sciences):
https://www.city.ac.uk/about/schools/health-sciences/research/centre-for-applied-vision-research/visual-neuroscience-and-psychophysics

Neuroscience

	Structure	Function
• **Accidents/ Lesions** ◦ By accidents/removing parts of the brain we were able to learn what different parts of the brain do	✗	✓
• **Electroencephalogram (EEG)** ◦ An EEG machine measures electrical currents produced by brain cells. The currents generate the so-called “brain waves”, e.g.: • Beta waves (high frequency): state of engagement • Delta waves (low frequency): state of deep relaxation, sleep	✗	✓

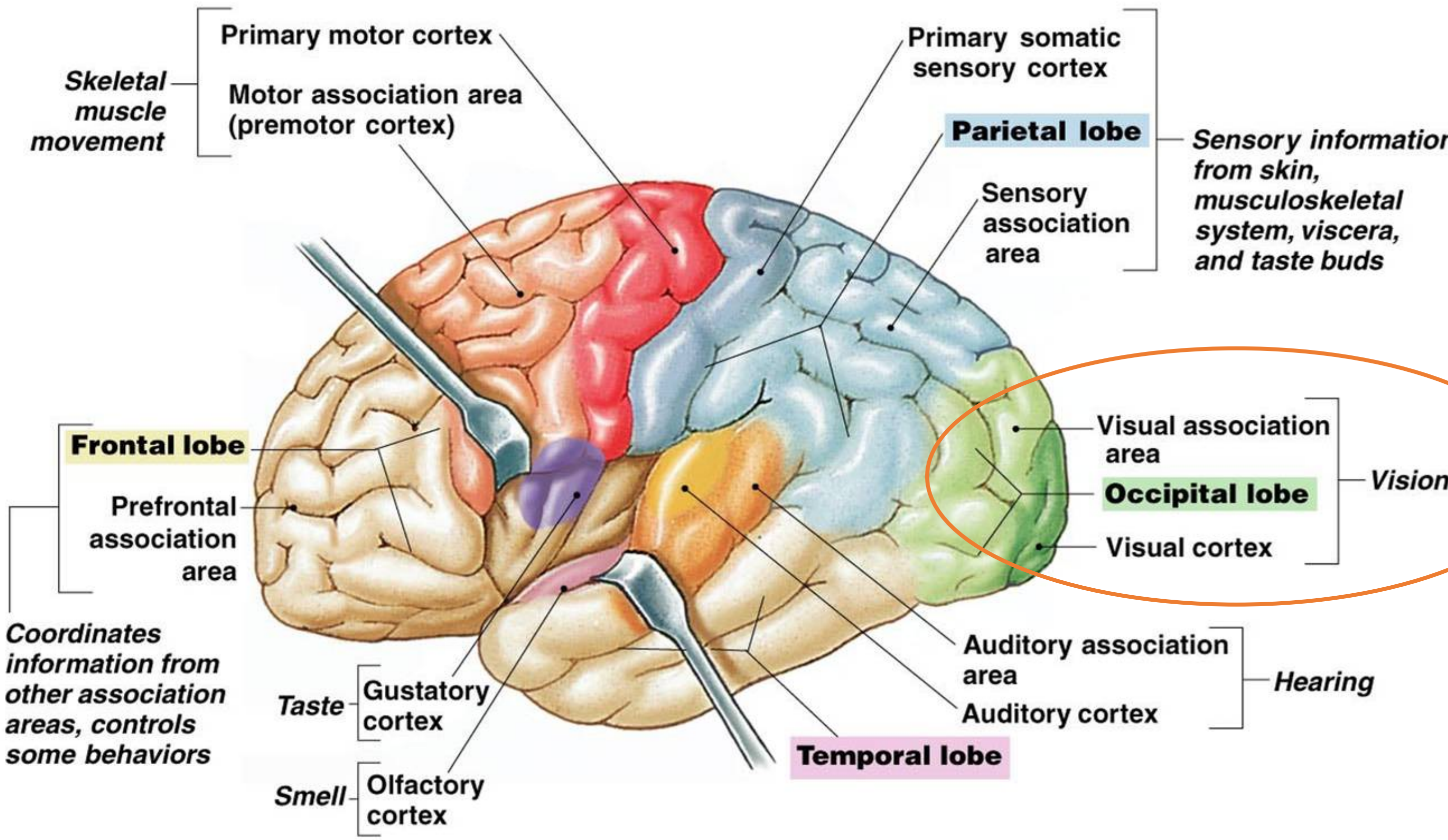
Neuroscience

	Structure	Function
Computerized Axial Tomography (CAT)	✓	✗
Magnetic Resonance Imaging (MRI)	✓	✓
Positron Emission Tomography (PET)	✗	✓

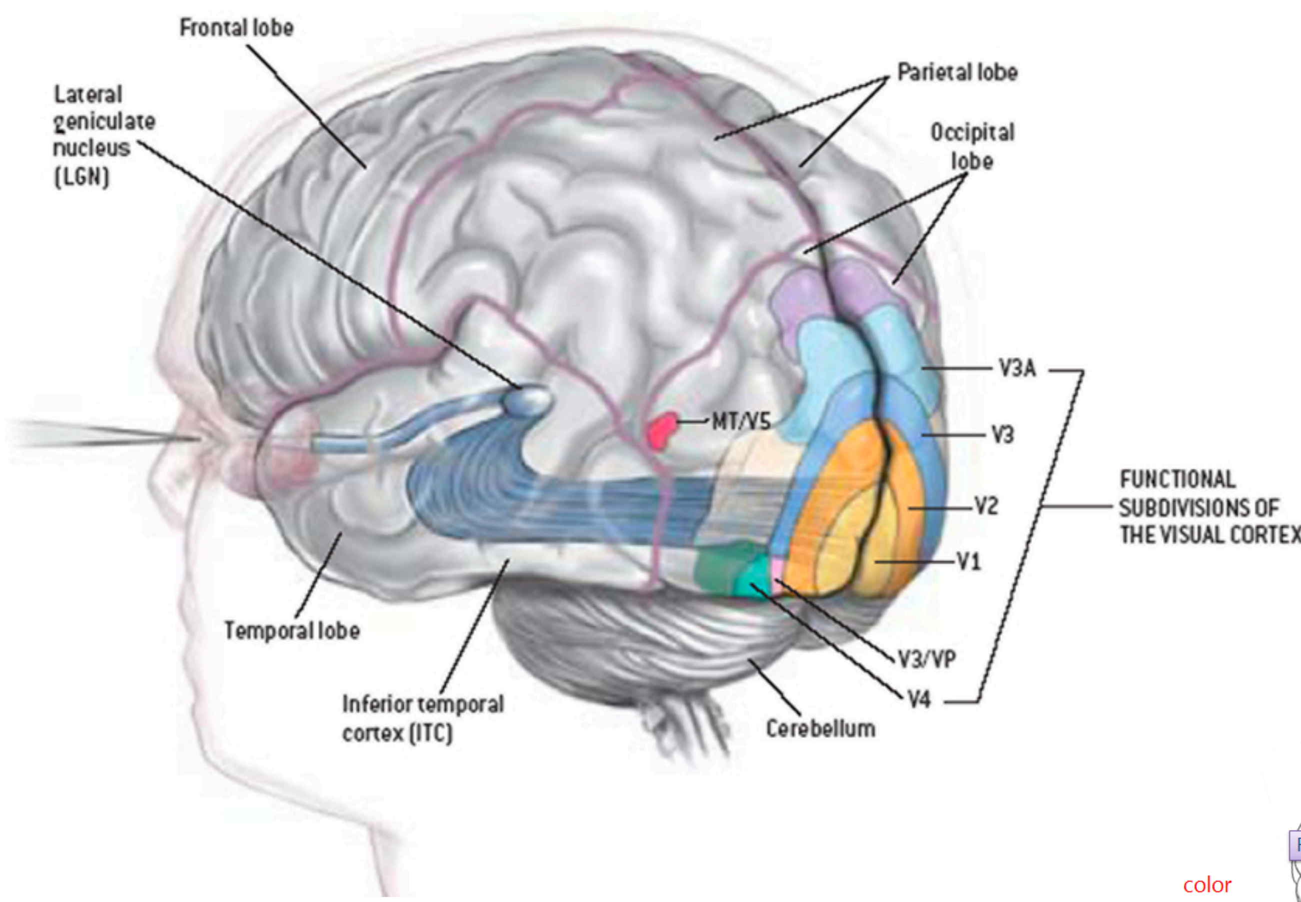
- **Computerized Axial Tomography (CAT)**
    - A CAT scan is essentially a sophisticated, 3-dimensional x-ray scan
- **Magnetic Resonance Imaging (MRI)**
    - It uses a strong magnetic field and radio-frequency waves to excite ^{1}H nuclei in the body, and generates an image by measuring how they return to equilibrium
    - It can be used to detect blood flow to specific brain regions: functional MRI
- **Positron Emission Tomography (PET)**
    - The patient usually swallows a substance (e.g. lab-altered glucose) and the PET scan shows what brain regions are gathering (i.e. using) that substance

Human brain

Humans are highly visual, with substantial cortical areas dedicated to processing visual information



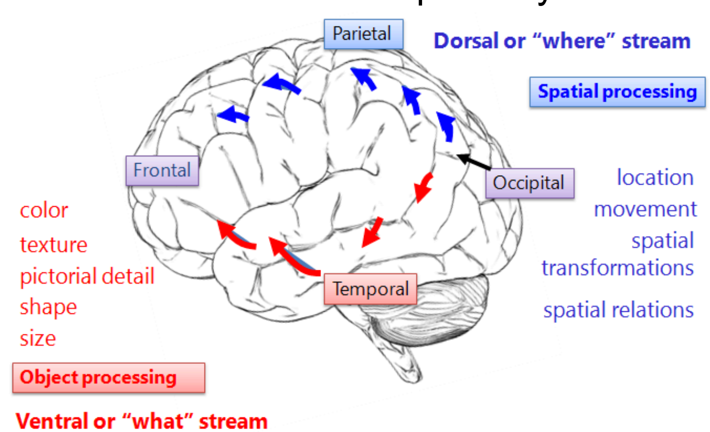
Human visual cortex



Function of different parts of visual cortex

V1	Edges
V2	Surfaces
V4	Color
MT	Motion
IT	Object Recognition

Two information pathways

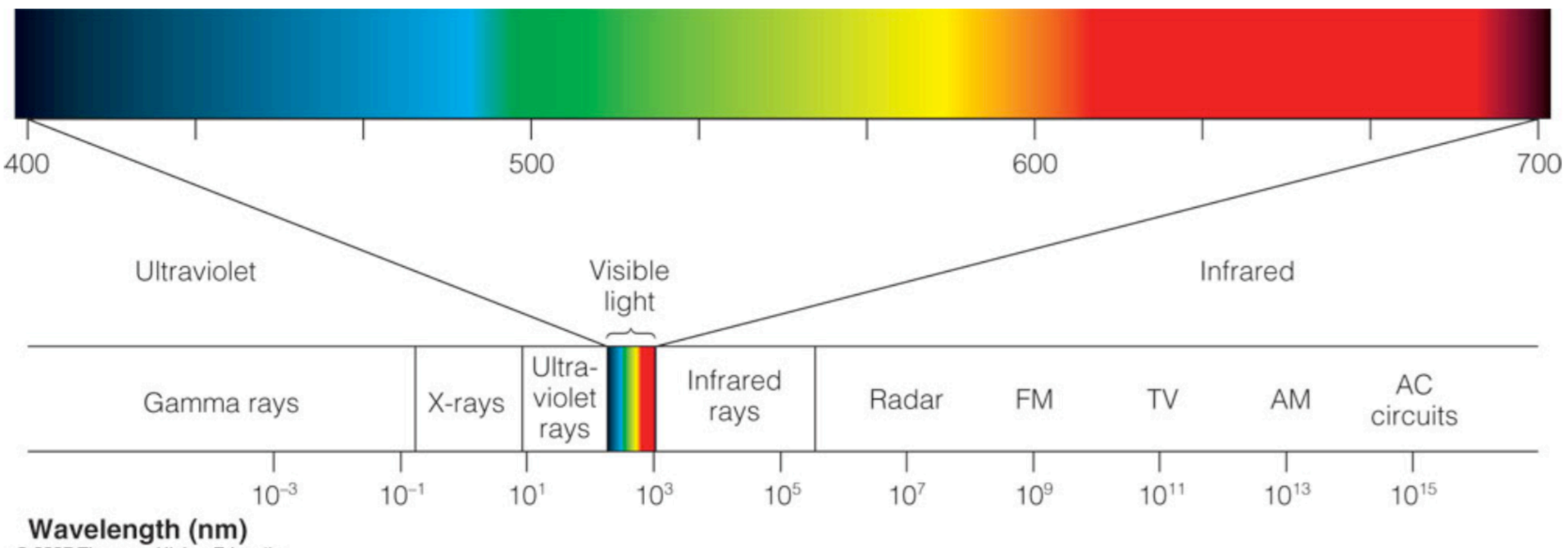


Credit: Bruno Cessac

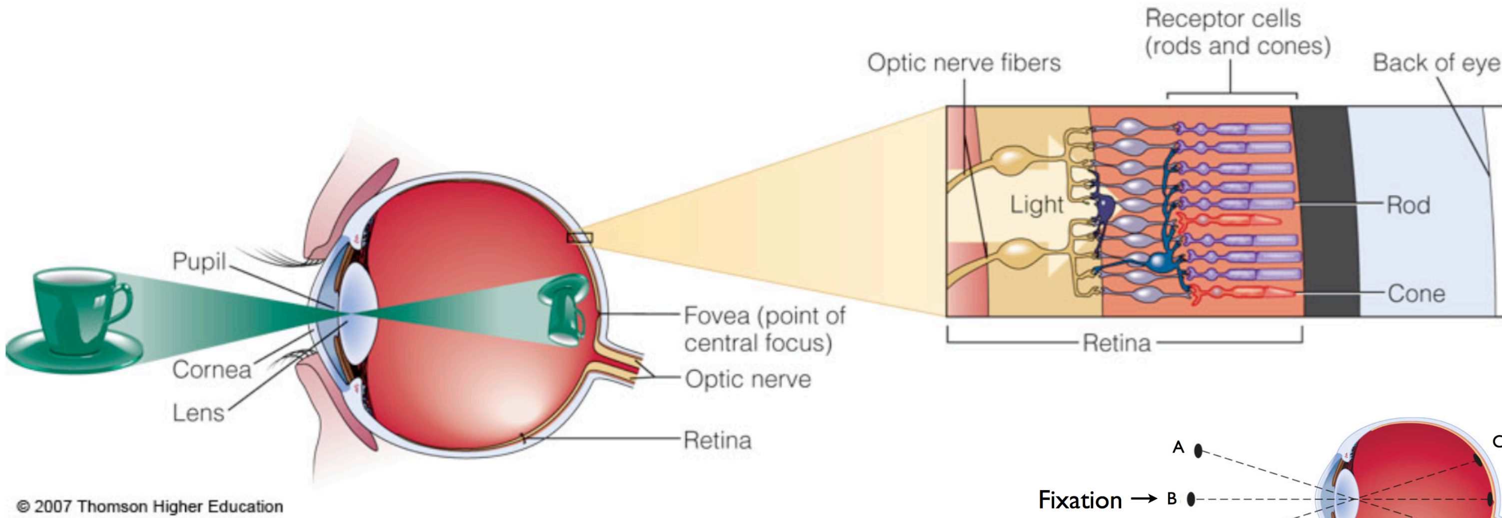
How much do we see?

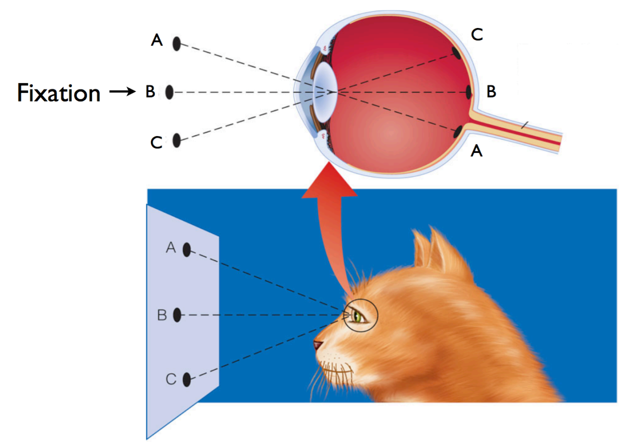
The visible spectrum:

- the portion of the electromagnetic spectrum that is visible to the human eye
- electromagnetic radiation in this range (~390÷700 nm) of wavelengths is called visible light



The eye





- Fovea: portion of retina responsible for sharp central vision (also called foveal vision)
- Receptor cells in the retina: rods and cones
- Blind spot: optic nerve on the retina: https://www.youtube.com/watch?v=IRgwMVRGqAY

Speed of processing

Humans can perform image classification in as little as 150 ms

Test: which picture (left/right) contains an animal?

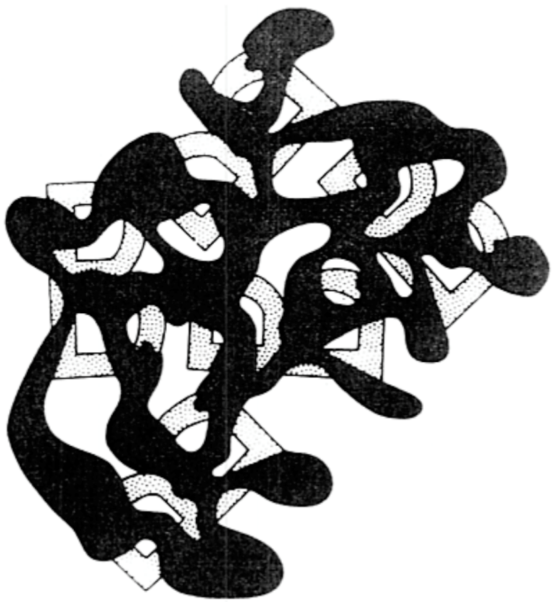
Human visual cortex

Humans effortlessly understand scenes that can be problematic for sophisticated algorithms/computers

Why is our visual system so good?

- Physiology: different brain areas gradually (but quickly) transform the information on our retinas into our perception
- Experience and inference: inference (from lighting, colour, etc.) is influenced by previous experience in a top-down fashion

- This "pipeline" has been driven by evolution
- Understanding it will allows us to build better machines



"… a group of Bs"

Limits of the human visual system

- What exactly are the capabilities of our visual system?
- Do we have a visual memory?
- How about illusions?

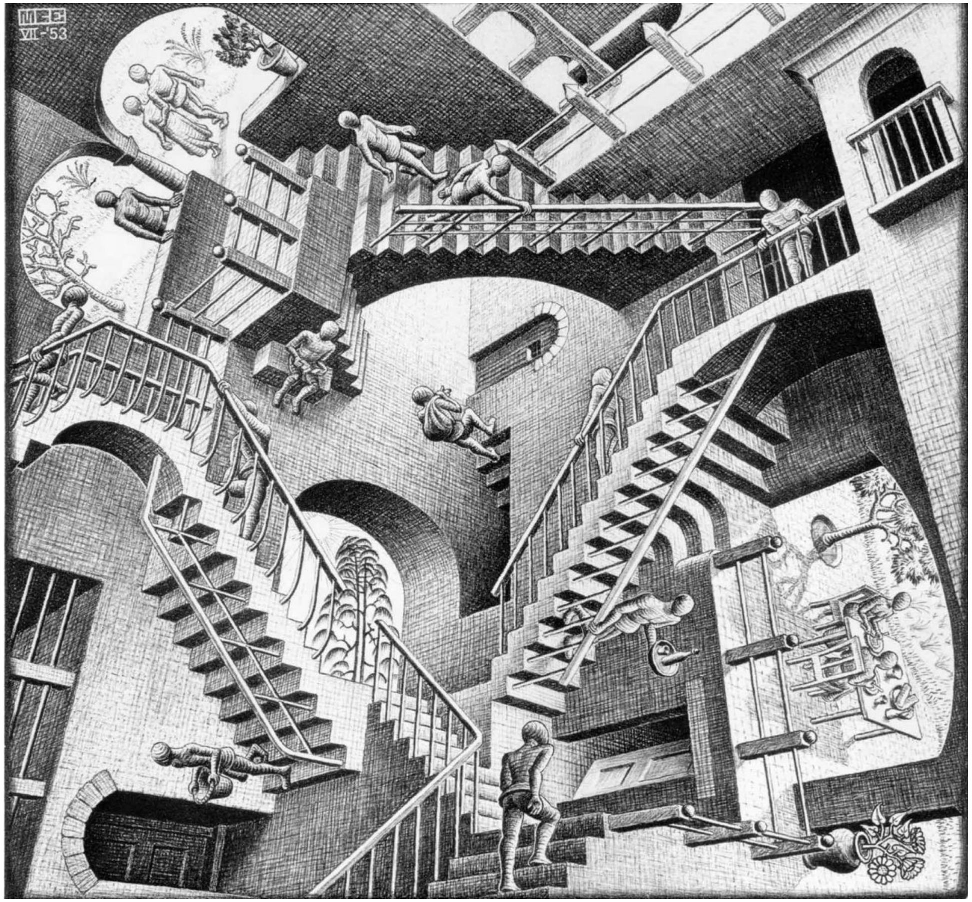
Scene understanding

Experience can have a strong positive influence on scene understanding

“Pintos” by Bev Doolittle

Scene understanding

However, experience can also hamper scene understanding (e.g. in cases of physical implausibility)



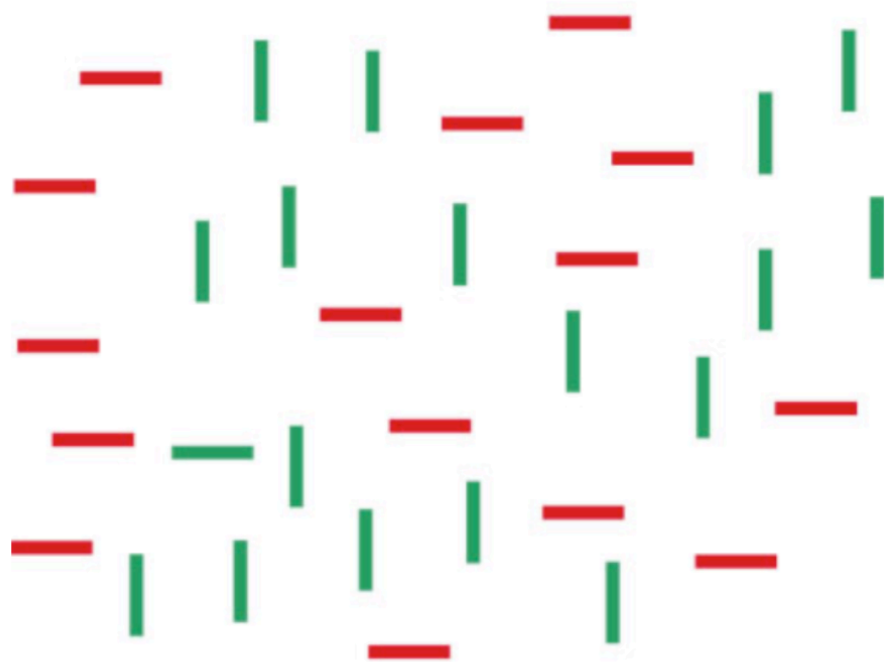
“Relativity” by M. C. Escher

Visual search

Test: where is the green horizontal bar?

Visual search

Test: where is the green horizontal bar?

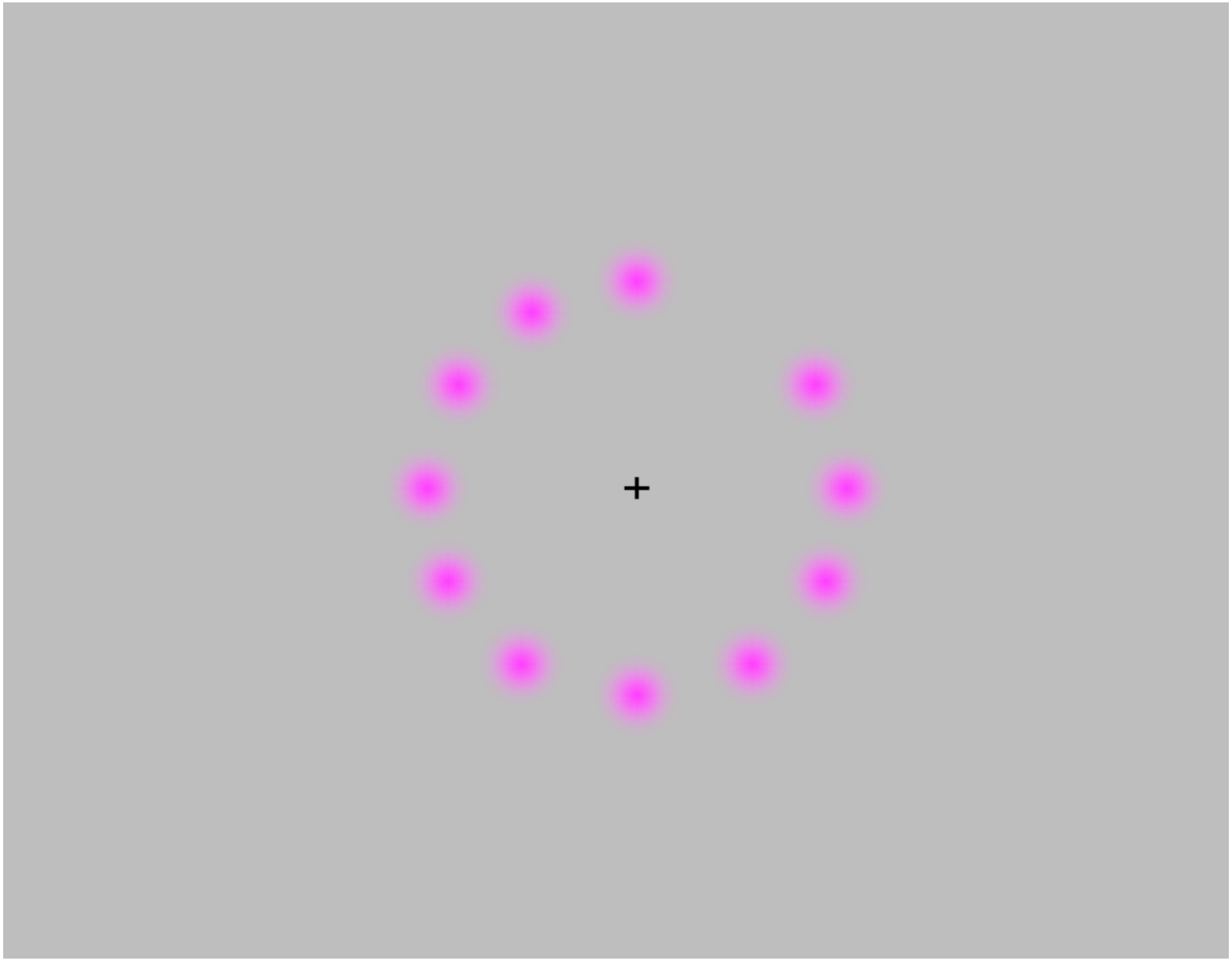


Colour inference can be misleading and slow down image understanding

Visual after-images

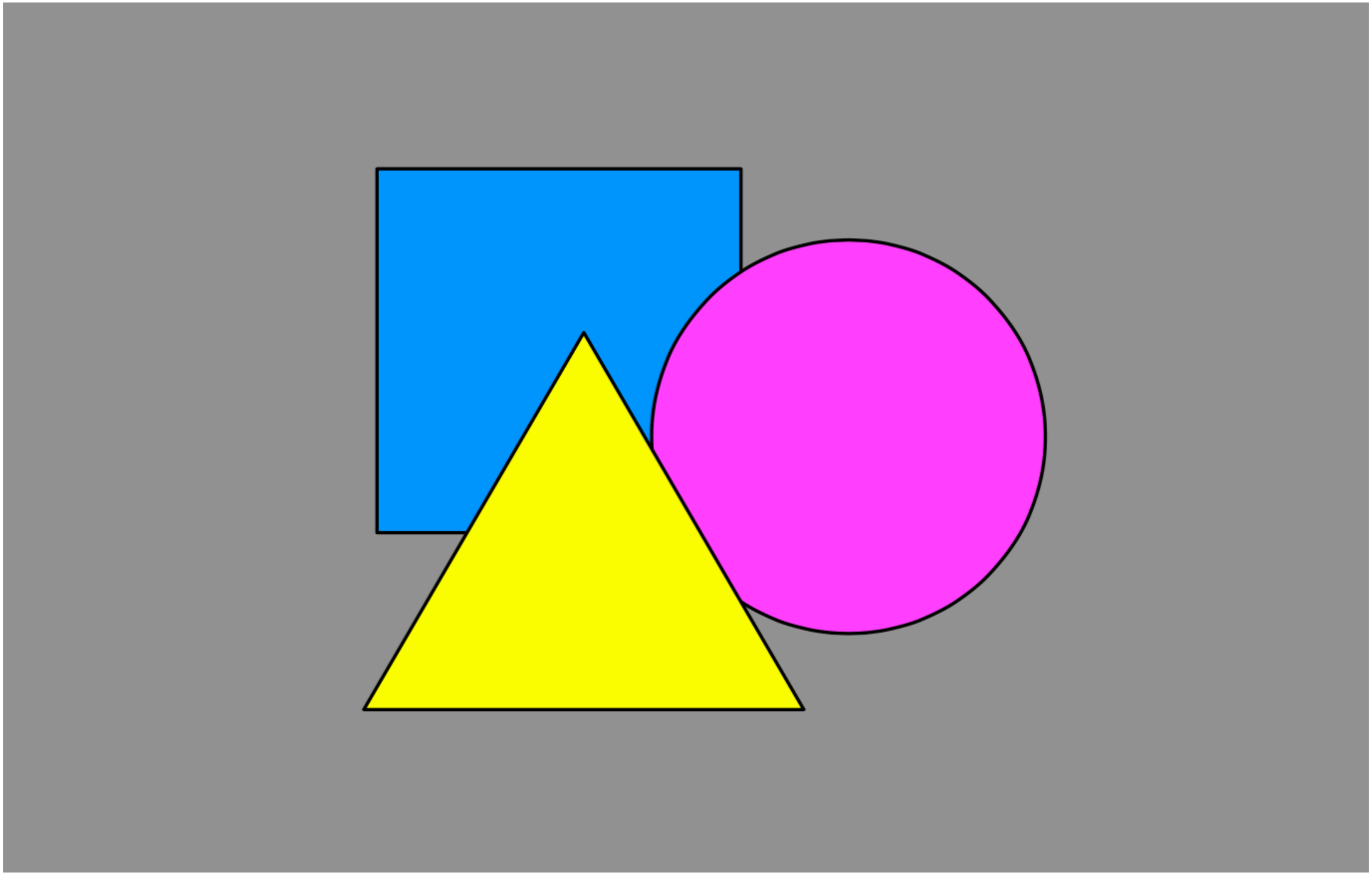
Test: stare for 15s at the arrow

CITY
UNIVERSITY OF LONDON
EST 1894



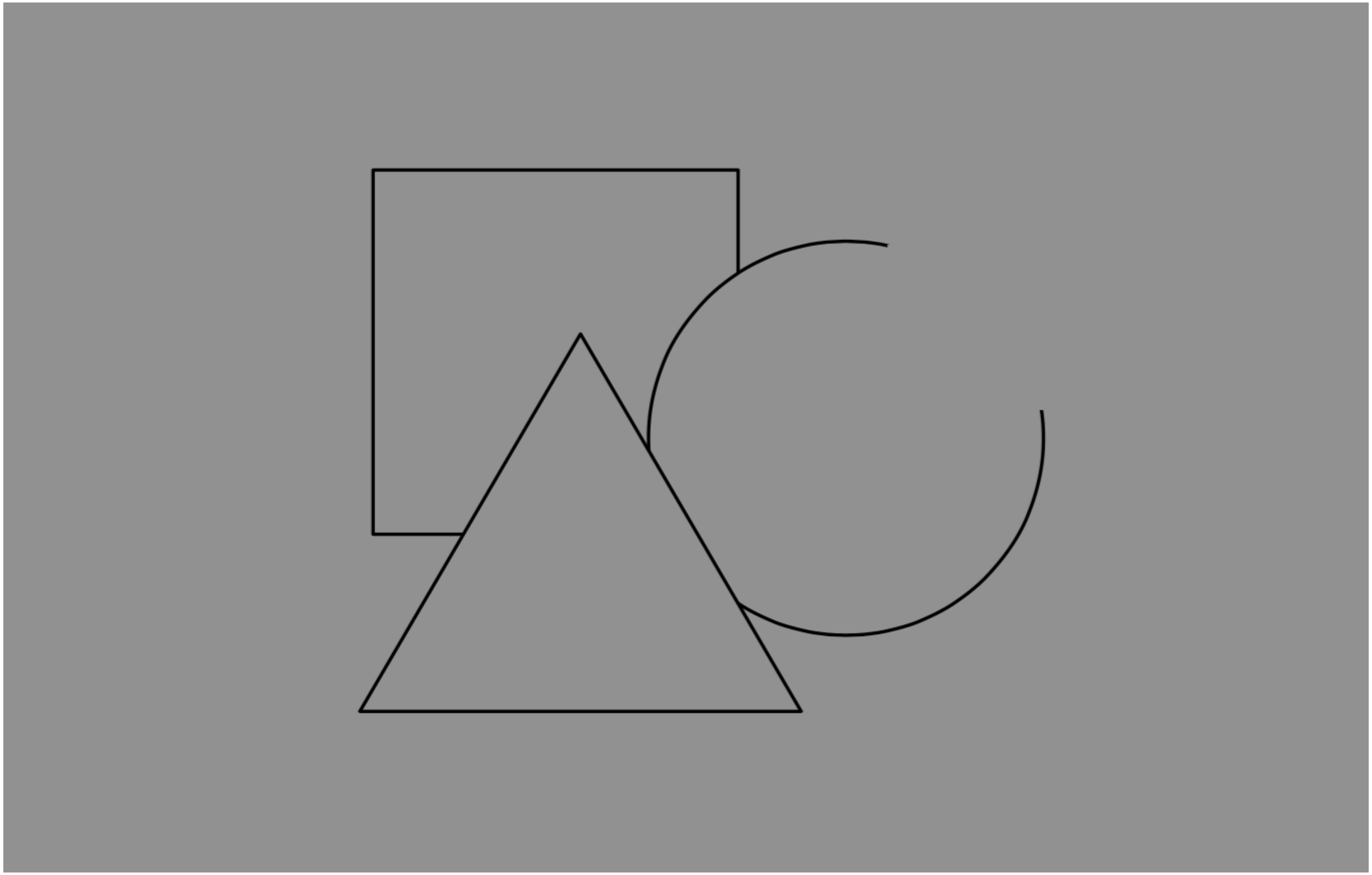
Visual after-images

Test: stare for 15s at the point of intersection between the 3 shapes



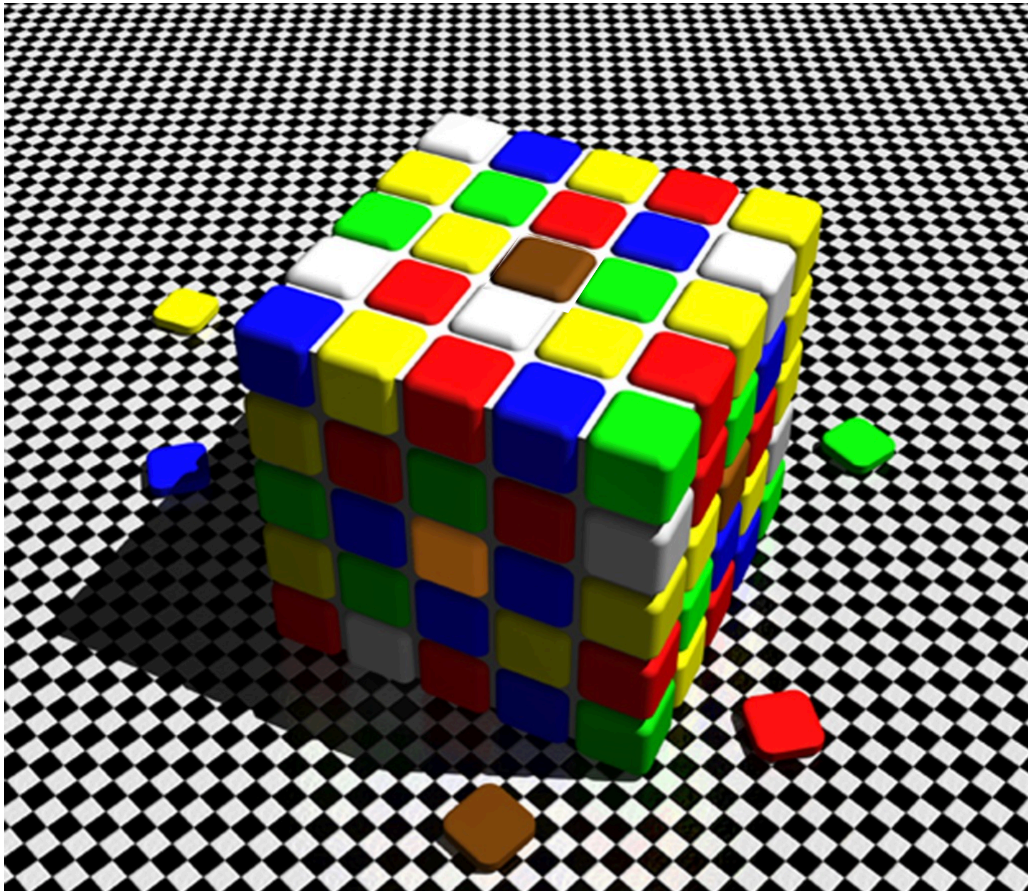
Visual after-images

Test: stare for 15s at the point of intersection between the 3 shapes



Shading inference

Test: how do the colours of the central tiles compare?



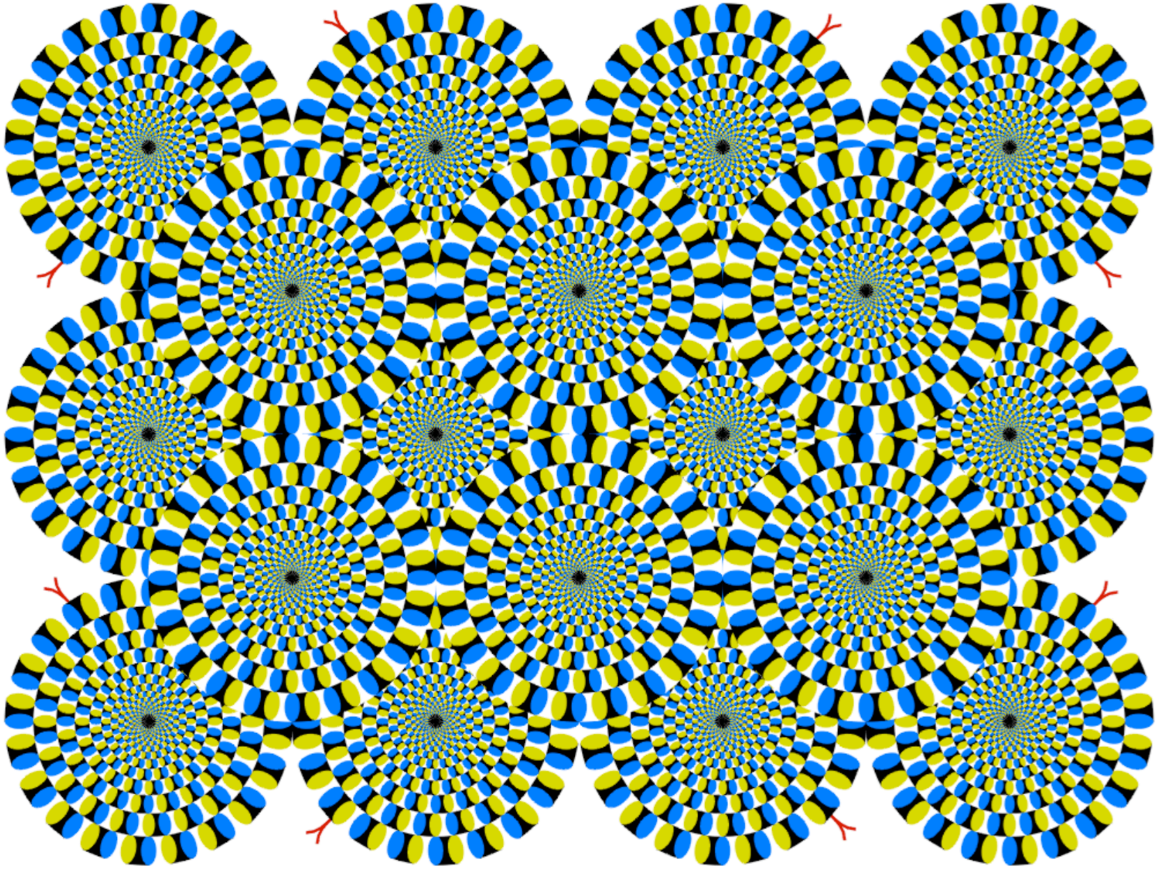
Shading inference

Test: how do the colours of the central tiles compare?

The perception of shade can mislead the perception of colour

Motion illusion

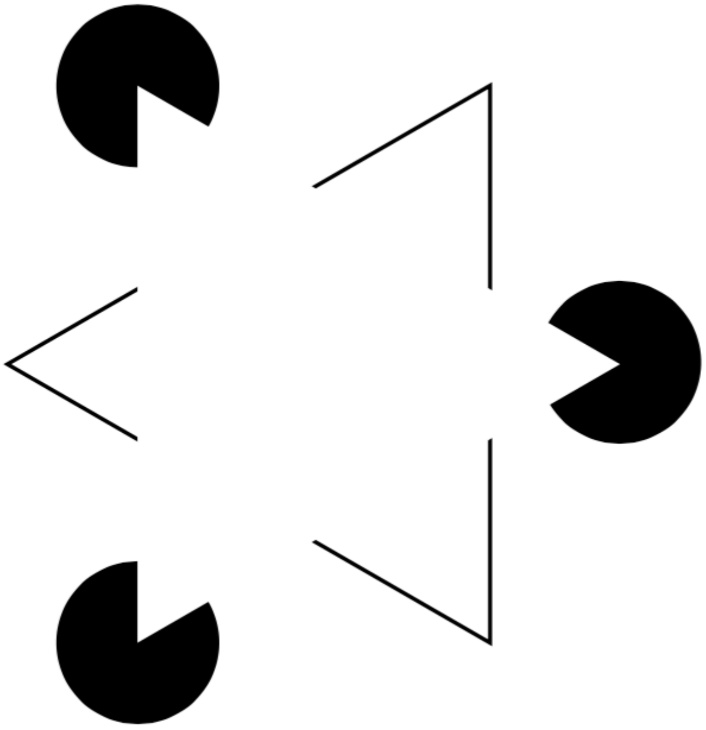
Motion can be perceived from static images



“Rotating snakes illusion” by Akiyoshi Kitaoka

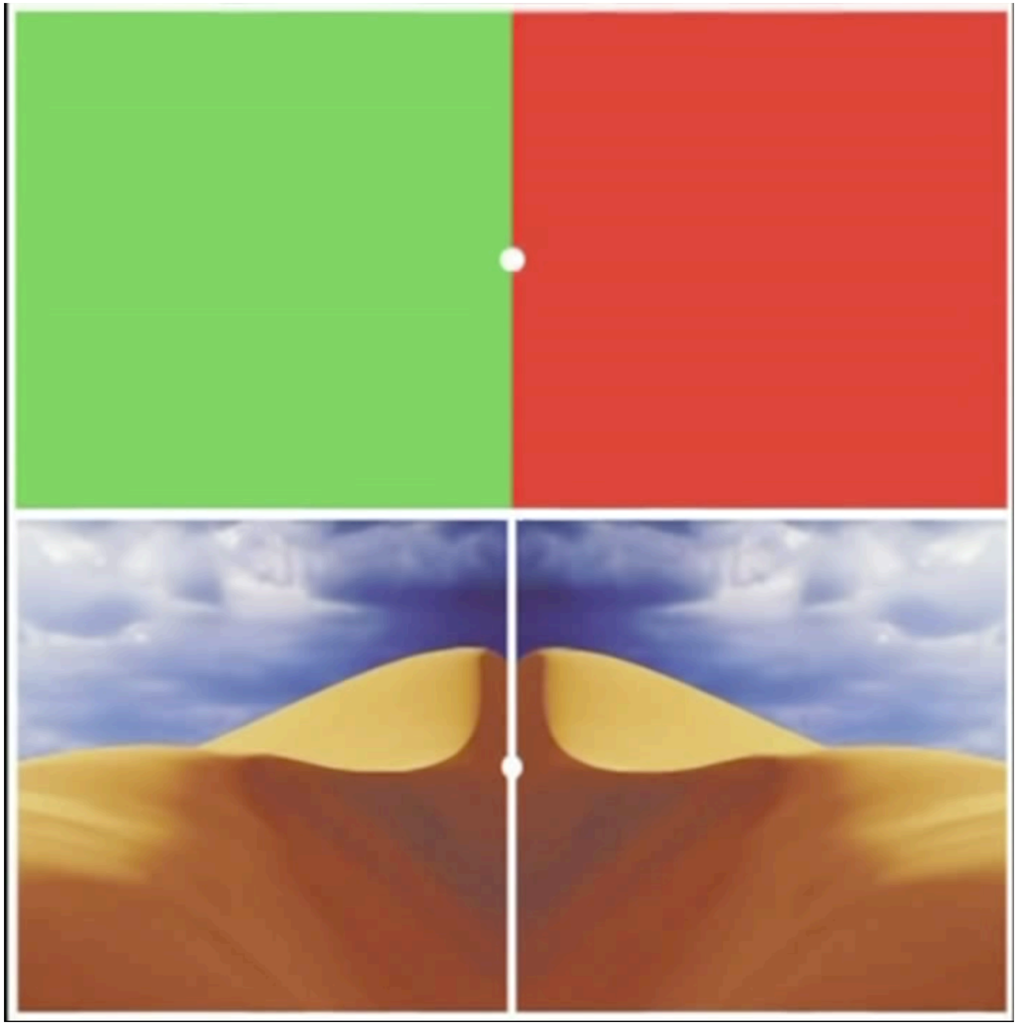
Surface perception

Perception of surfaces can be constructed from open contours



Context

Test: fixate on upper white dot for 20s, then fixate on the lower dot

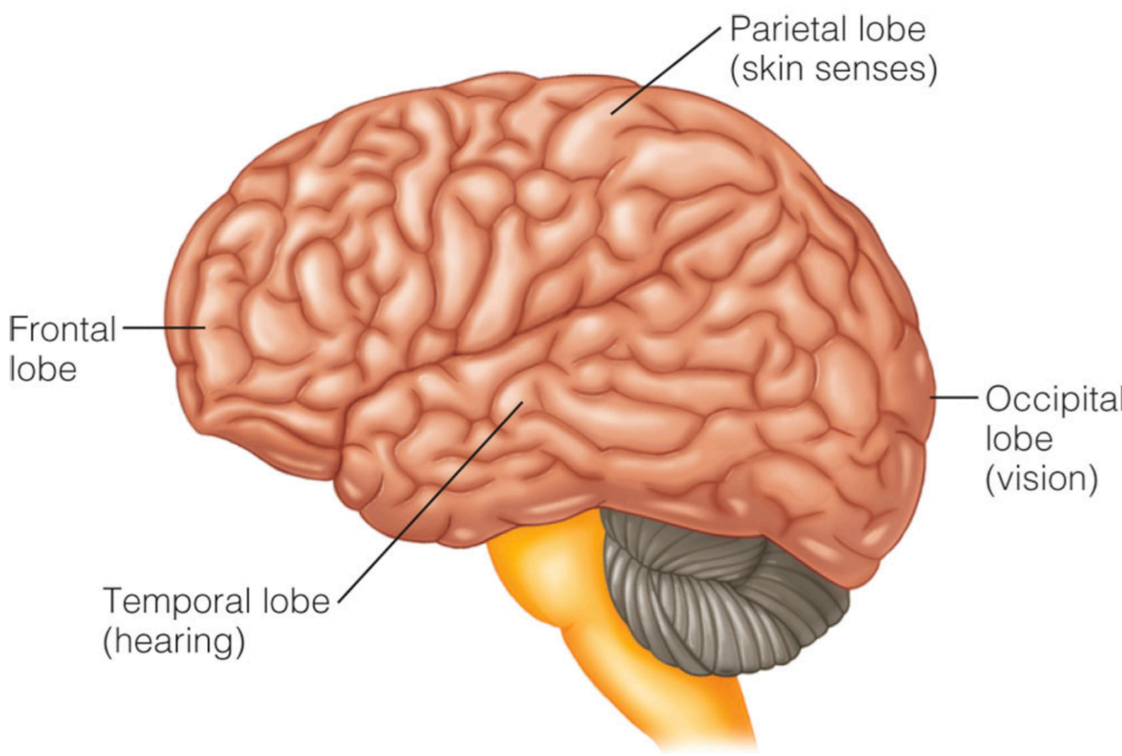


Context

https://www.youtube.com/watch?v=YPIB20a0CQw

Physiology & perception

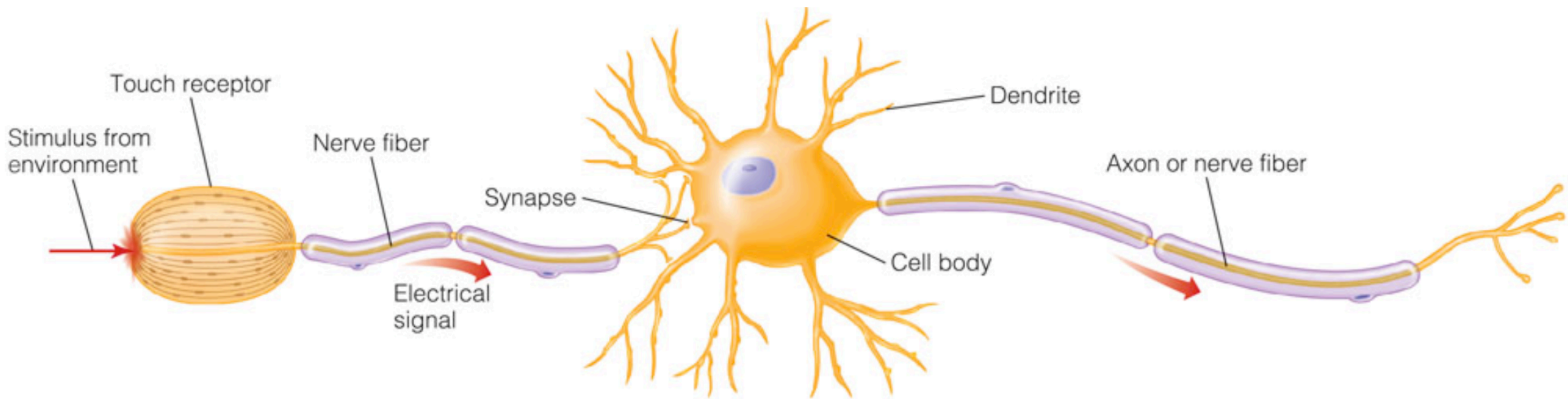
- How are physiological processes involved in perception?
- How is light transformed into electricity in the eye?
- How is what we see determined by the properties of the receptors in our retinas?



Neuron

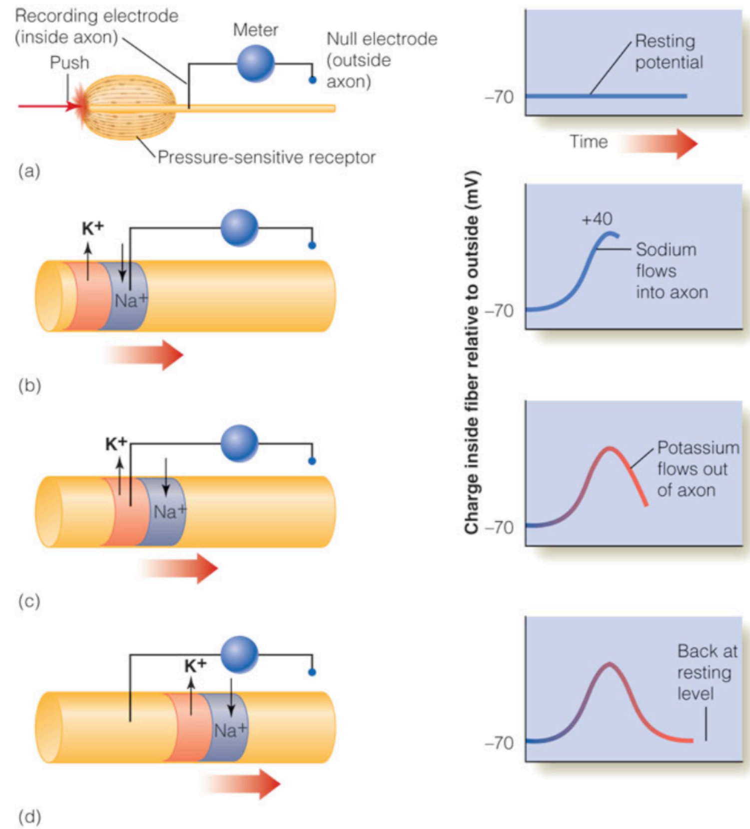
Neurons are the electrically-excitable cells that form the nervous tissue

- Elements of a neuron:
  - Dendrites: receive signals from other neurons (input of the neuron)
  - Axon: sends signal to other neurons (output of the neuron)
  - Cell body: keeps the neuron alive
  - Synapses: conjunctions between axons and dendrites
- Some numbers:
  - A typical neuron fires ~0.5÷10 times every second
  - Each individual neuron can form ~1000 synapses with other neurons
  - There are ~100 billion neurons in the human brain: ~100÷1000 trillion synapses



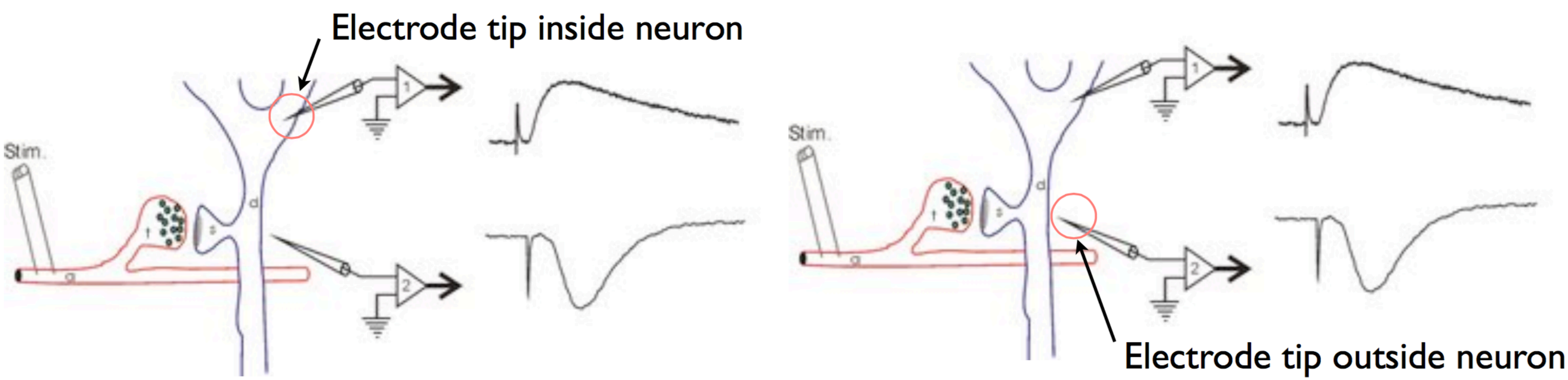
Changes in membrane potential

- Neurons contain and are surrounded by a solution rich in ions (electrically charged particles):
  - Positive: Sodium ions (Na+), Potassium ions (K+)
  - Negative: Chlorine ions (Cl-)
- At rest, equilibrium between electrical force and diffusion determines resting membrane potential: inside of neuron is negative, outside is positive
- Each synapse (for a given neuron) can produce two types of input:
  - Excitatory: positive change to membrane potential (depolarization)
  - Inhibitory: negative change to membrane potential (hyperpolarization)
- If summation of inputs determines depolarization, current is generated through the axon called action potential



Recording neural signals

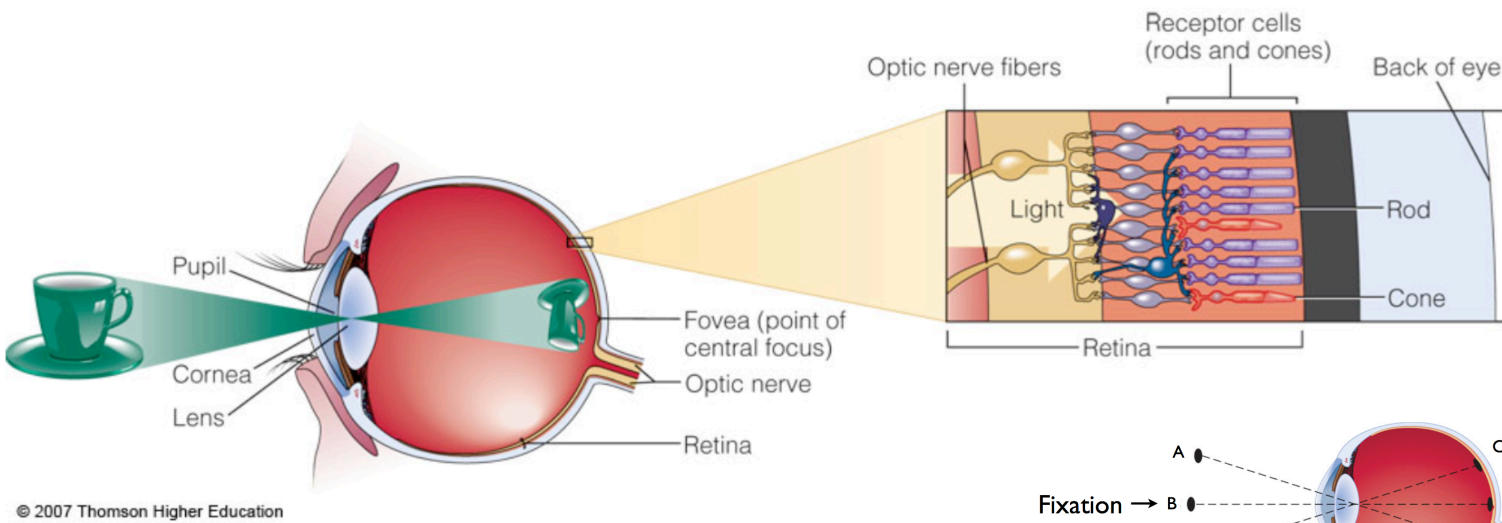
- Intra-cellular recordings provide access to the membrane potential of a neuron to observe action potentials and sub-threshold changes
- Extra-cellular recordings provide access to the action potentials of multiple neurons
- Both highly invasive and hurtful for the subject (only used in past animal studies), now more superficial readings

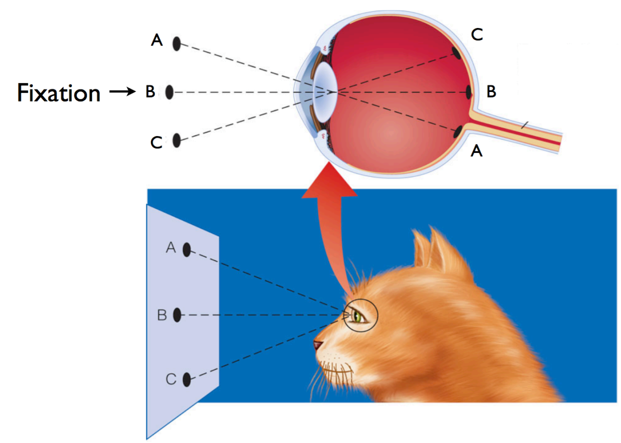


Intra-cellular recording

Extra-cellular recording

The eye

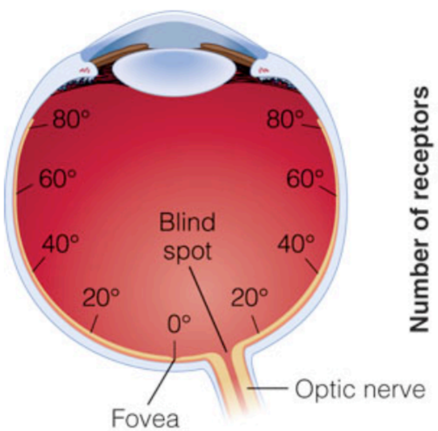


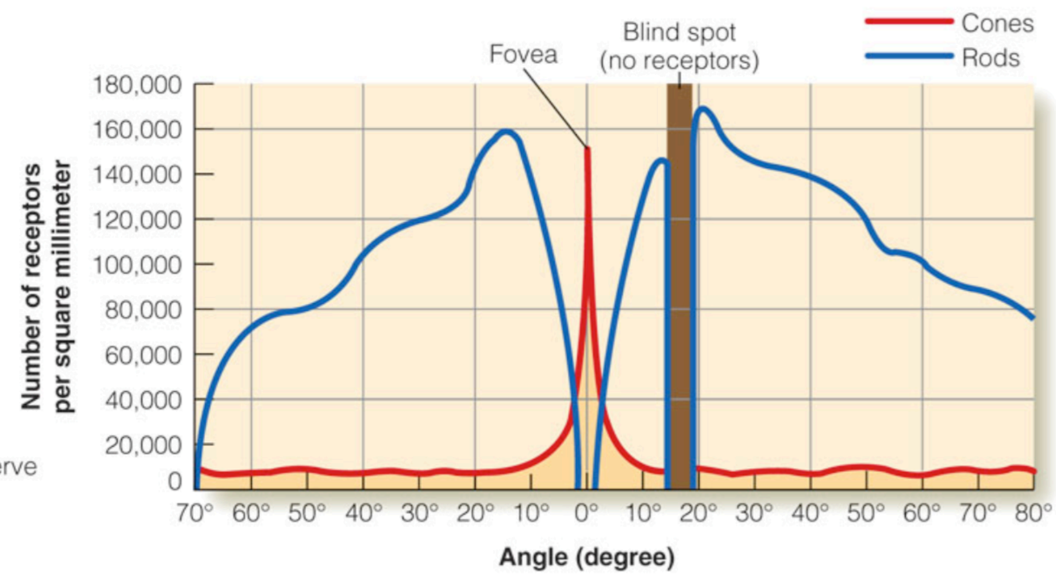


- Fovea: portion of retina responsible for sharp central vision (also called foveal vision)
- Receptor cells in the retina: rods and cones
- Blind spot: optic nerve on the retina: https://www.youtube.com/watch?v=IRgwMVRGqAY

Rods and cones

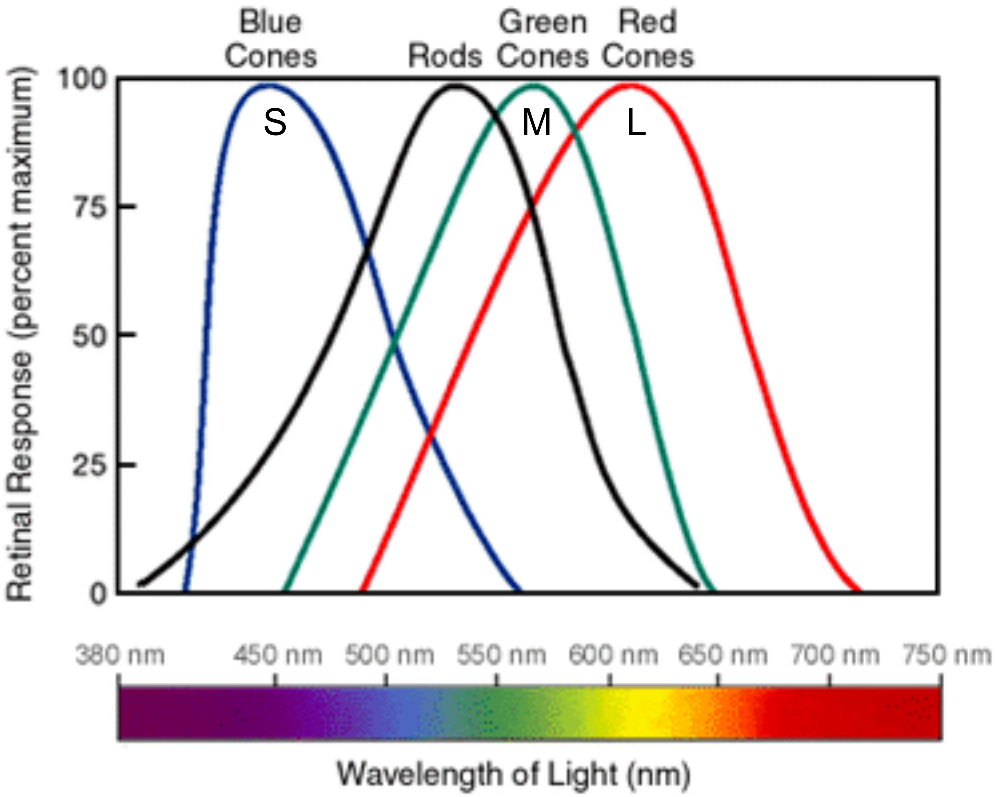
- Rods:
  - Large and cylindrical
  - Only in the peripheral retina and in much greater numbers
  - Essential at very low light levels, no colour vision
- Cones:
  - Small and tapered
  - Found in the fovea and in the peripheral retina
  - Three types, different spectral sensitivity: colour vision





Trichromatic theory

The relative strengths of the signals detected by the three types of cones are interpreted by the brain as a visible colour



L, M, S based on wavelengths

Spectral sensitivity

Colour blindness

Normal trichromat

Missing M cones: red and green look similar

Colour perception

Test: focus on dot for 20s

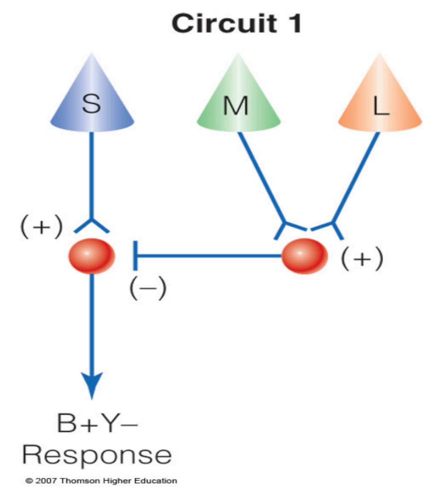
Colour perception

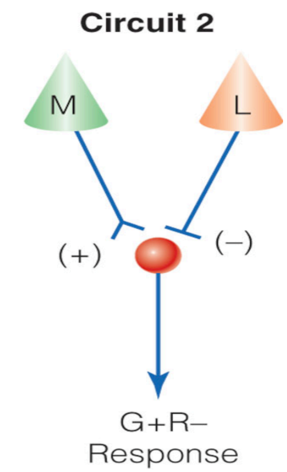
Colour can ne perceived “through” an after-image

Opponent process theory

One member of the colour pair suppresses the other

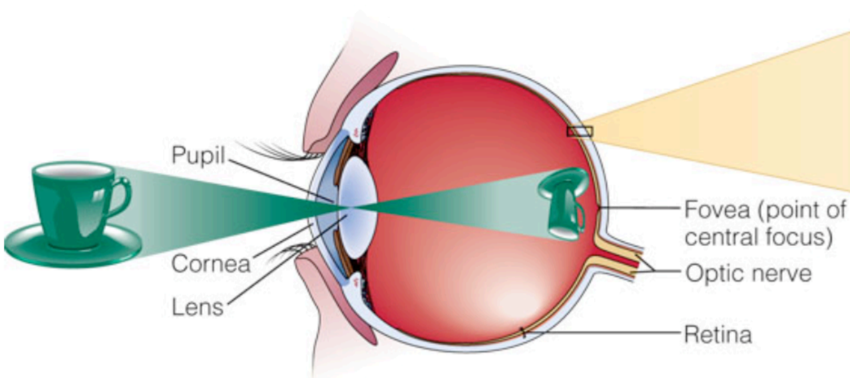
- Colour pairs:
    - blue versus yellow
    - red versus green
    - black versus white
- Researchers performing single-cell recording, found opponent neurons for cones which respond in an excitatory manner to one end of the spectrum and an inhibitory manner to the other
- In reality, trichromatic theory explains vision at receptor level, and opponent process theory explains vision at neural level

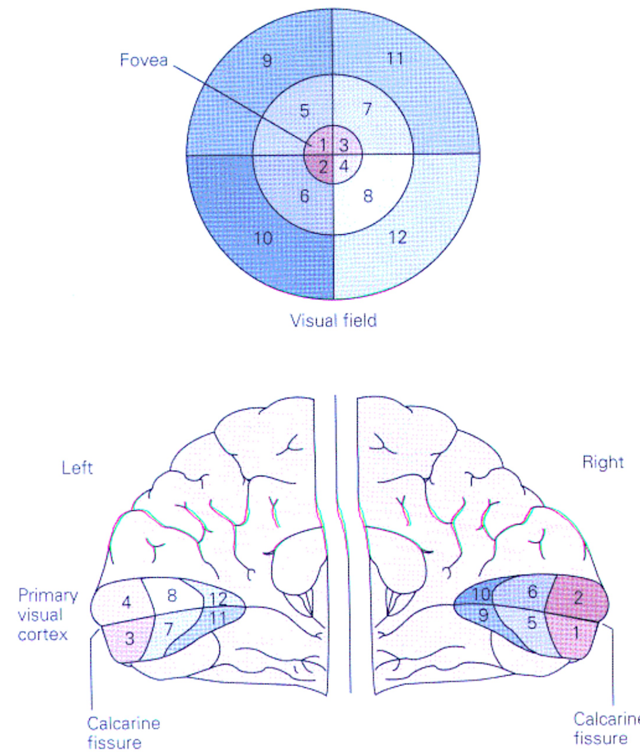




Cortical magnification

- Fovea accounts for 0.01% of retina's surface
- However, signals from fovea account for 8÷10% of the visual cortex
- This provides extra processing power for high-acuity tasks
- Peripheral view mostly sensitive to movement





Cortical magnification

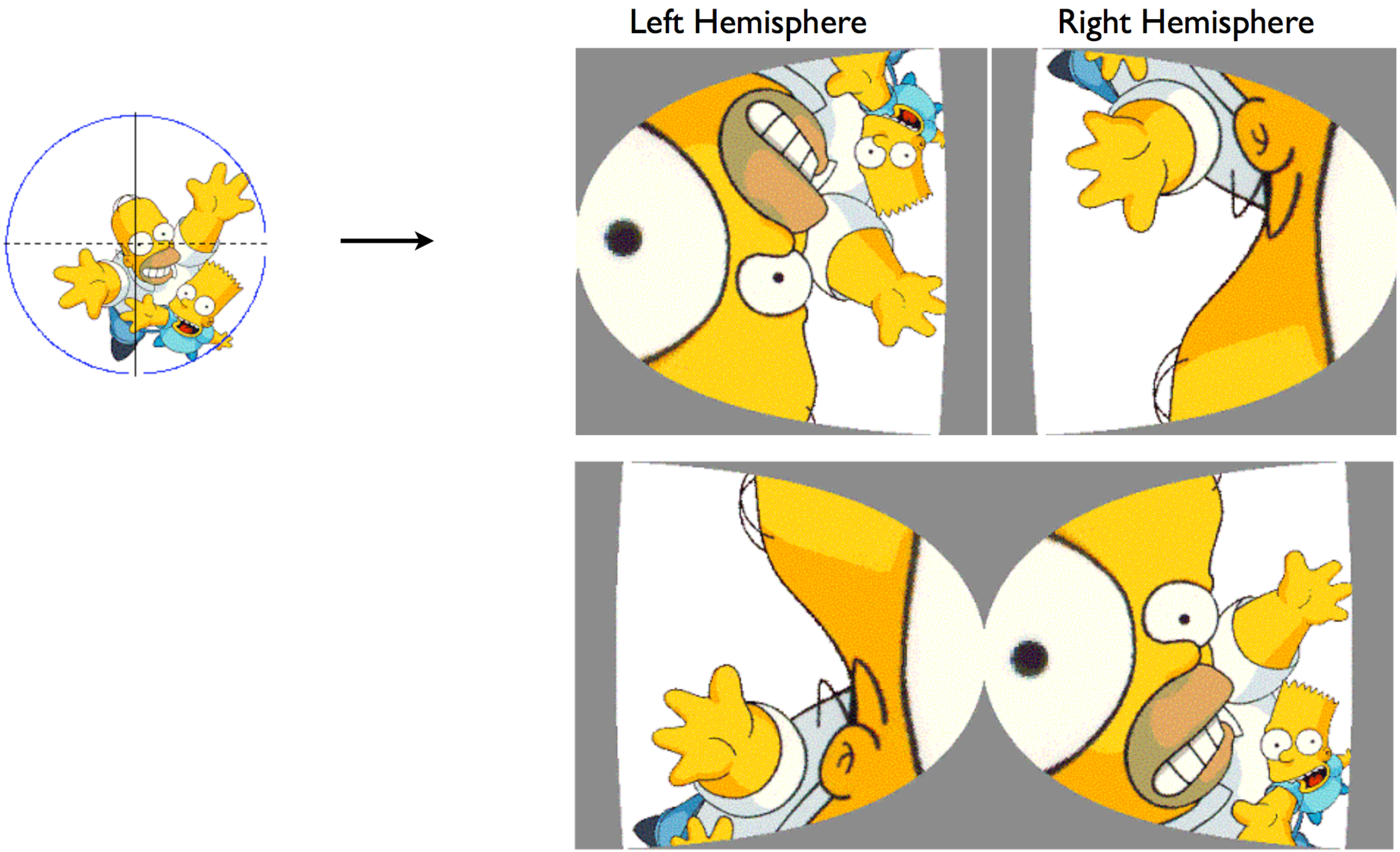
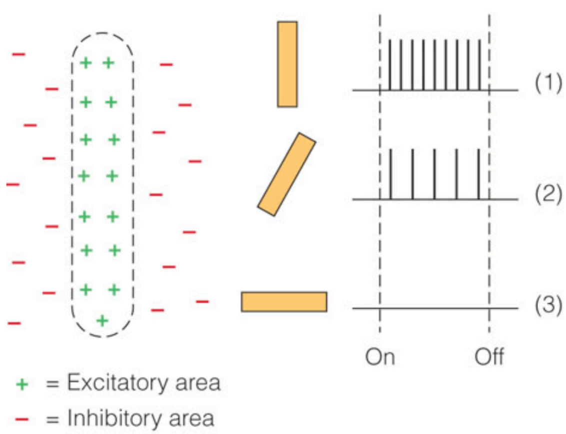


Illustration of how retinal information is mapped onto cortical one

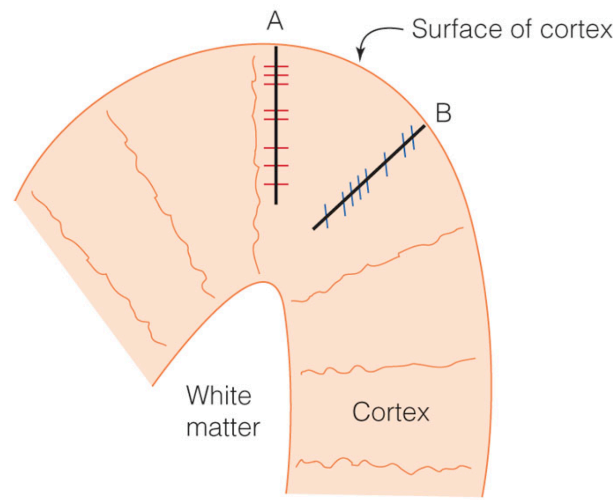
V1: simple cells

- Simple cortical cells (mostly found in primary visual cortex, V1) are neurons organized in columns
- Simple cells of one column fire maximally to the same orientation of stimuli (edges)
- Adjacent columns change preference in an orderly fashion
- 1 mm across the cortex represents entire range of orientations
- Sensitive also to colour

Simple Cells



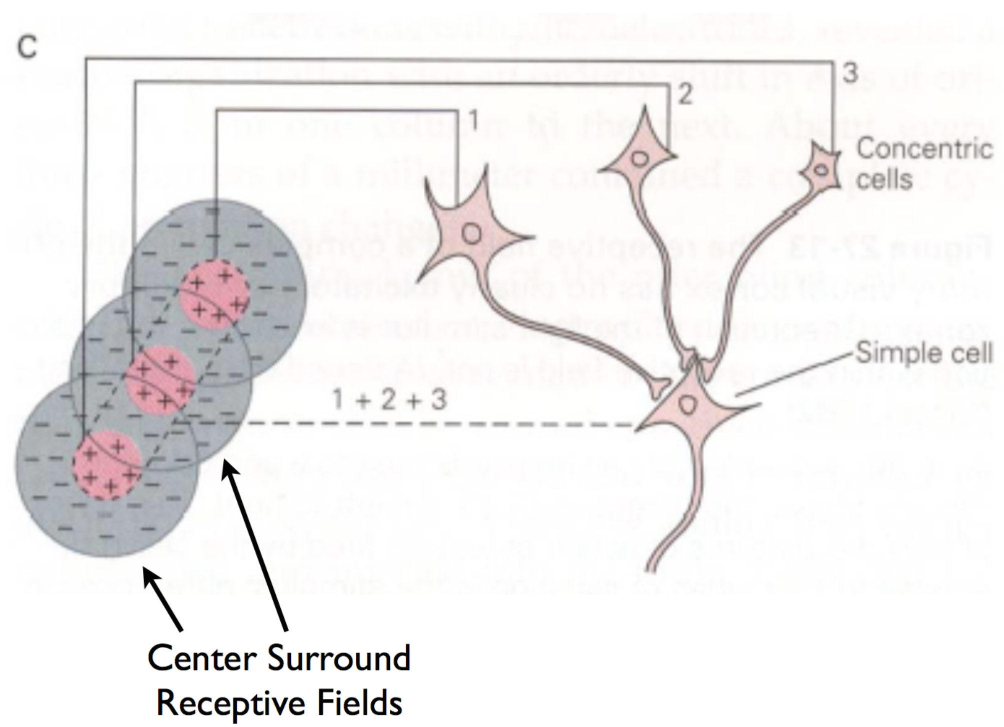
(a)



V1: simple cells

How simple cells work:

- Gather input from concentric cells:
  - Side-by-side receptive fields
  - Respond to spots of light
- Respond best to bar of light oriented along the length of the receptive field

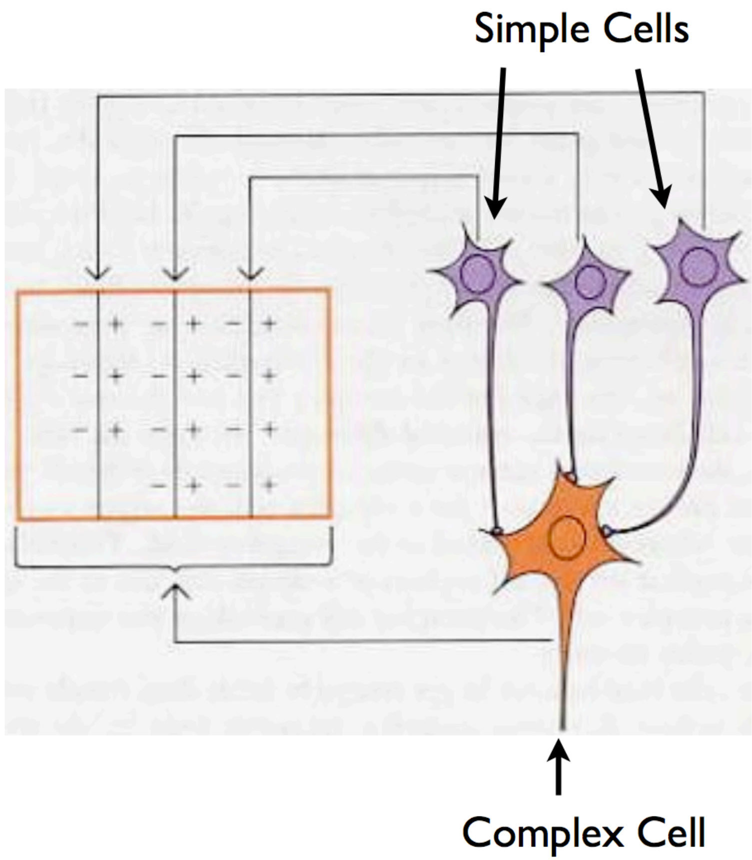


V1: complex cells

Complex cells:

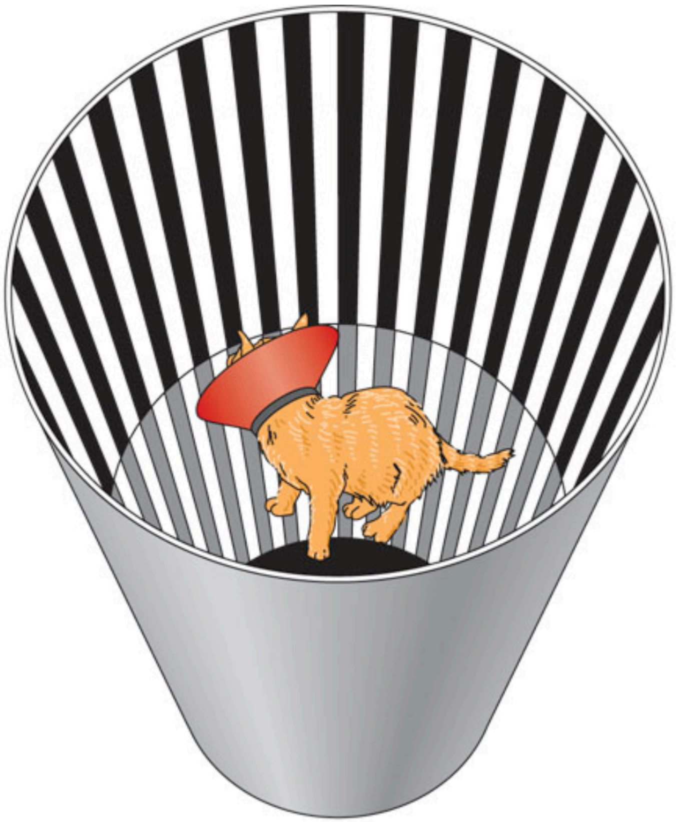
- Gather input from column of simple cells:
  - Response is independent from exact location
  - Respond to patterns of light in a certain orientation within a large receptive field
- Some complex cells respond optimally only to movement in a certain direction

Aperture problem:
http://web.mit.edu/persci/demos/Motion&Form/demos/one-square/one-square.html



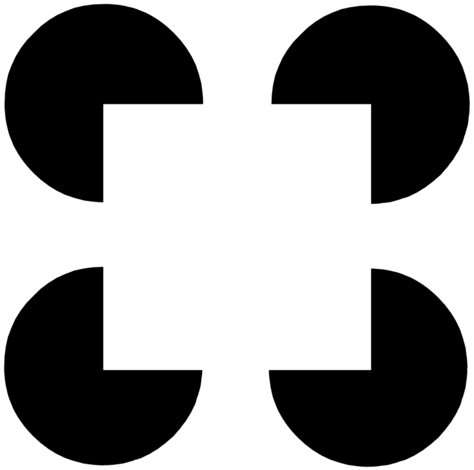
Selective rearing

- Animals reared in specific environment which limits type of stimuli present **develop lack of ability** to see characteristics unavailable in environment (e.g. horizontal edges) once freed
- Learning is fundamental for our brains (and neural networks)



Beyond V1

- V2: similar functions to V1 but also
  - Responds to illusory contours
  - Signals border ownership
  - Creates representation of surfaces
- V4:
  - Detects geometrical shapes of intermediate complexity
  - It is heavily modulated by attention
- MT:
  - Detects (apparently more complex) motion



V2 neurons respond to illusory edges

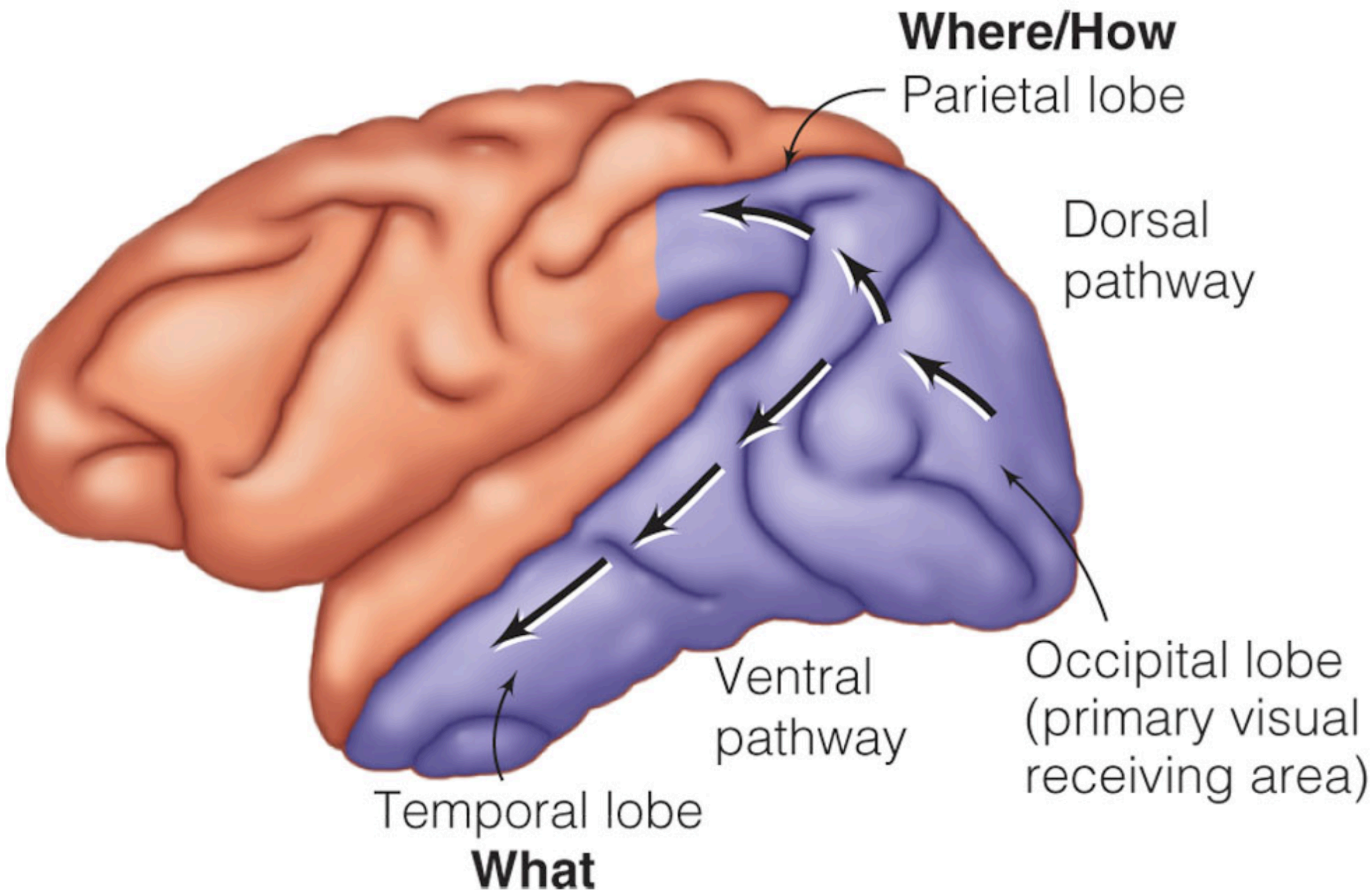
In reality, situation is much more complex:

- Feedback loops (e.g. V1 → V2 → V1)
- By-pass connections (e.g. V1 → V4)
- Redundancy, modulation, …

Lesioning (ablation experiments)

- Procedure:
  - First, an animal is trained to indicate perceptual capacities
  - Second, a specific part of the brain is removed or destroyed
  - Third, the animal is retrained to determine which perceptual abilities remain
  - The results reveal which portions of the brain are responsible for specific behaviours
- Using ablation, part of the parietal lobe was removed from half the monkeys and part of the temporal lobe was removed from the other half
- Removal of temporal lobe tissue resulted in problems with the object discrimination task – “What” pathway
- Removal of parietal lobe tissue resulted in problems with the landmark discrimination task – “Where” pathway

“What” and “where” pathways



The Gestalt approach

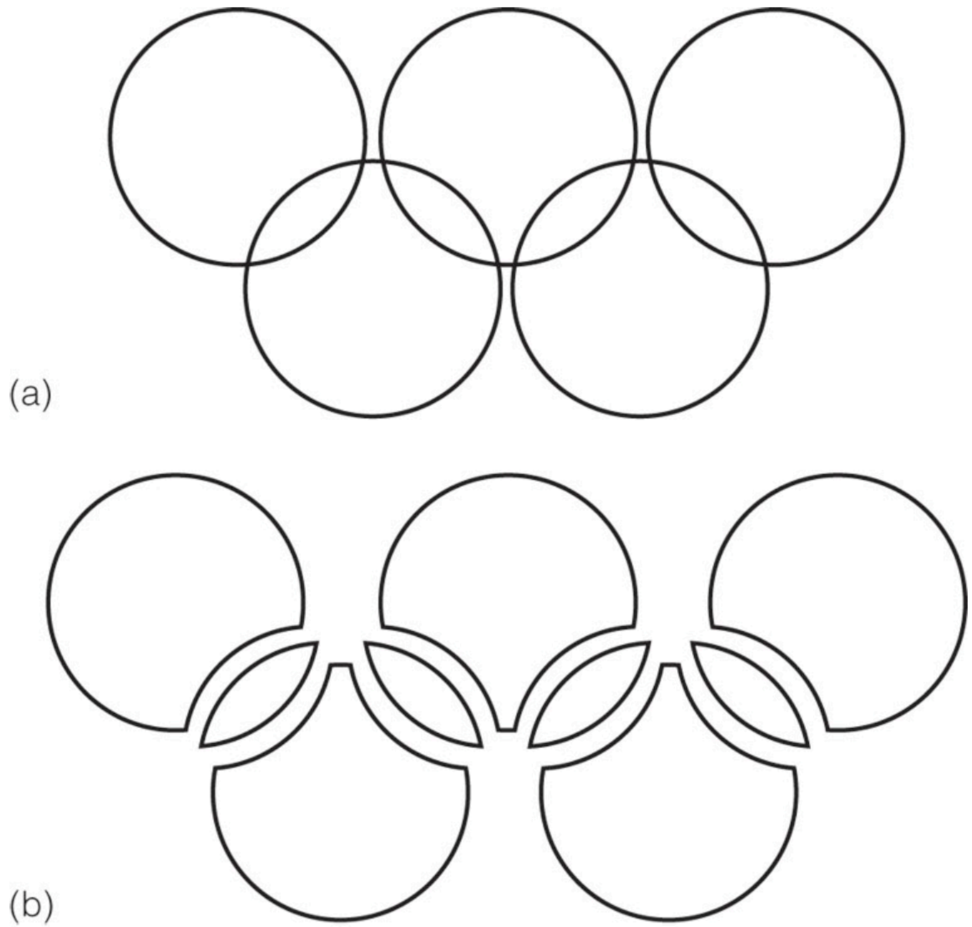
"The whole differs from the sum of its parts"

- Visual perception of a scene consists of a single act, and the attributes of the whole are not deducible from the analysis of the parts in isolation
- Gestalt laws try to explain how we perceive things

Test: what do you see in this image?

Gestalt laws

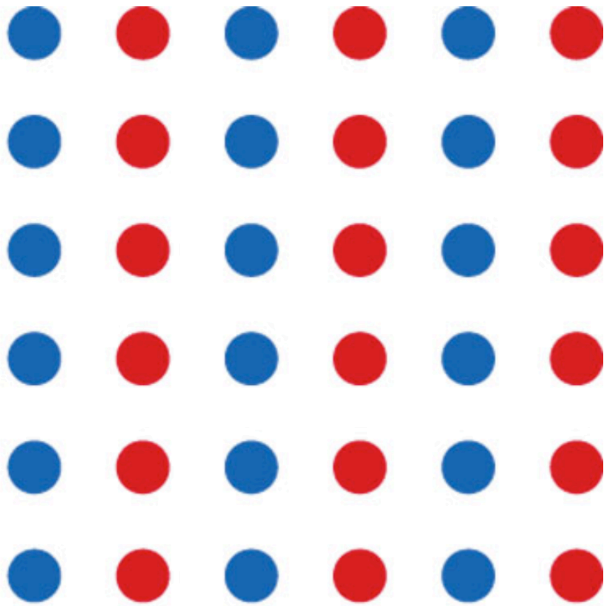
- Pragnanz/Simplicity: every stimulus is seen as simply as possible



Gestalt laws

- Similarity: similar things are grouped together

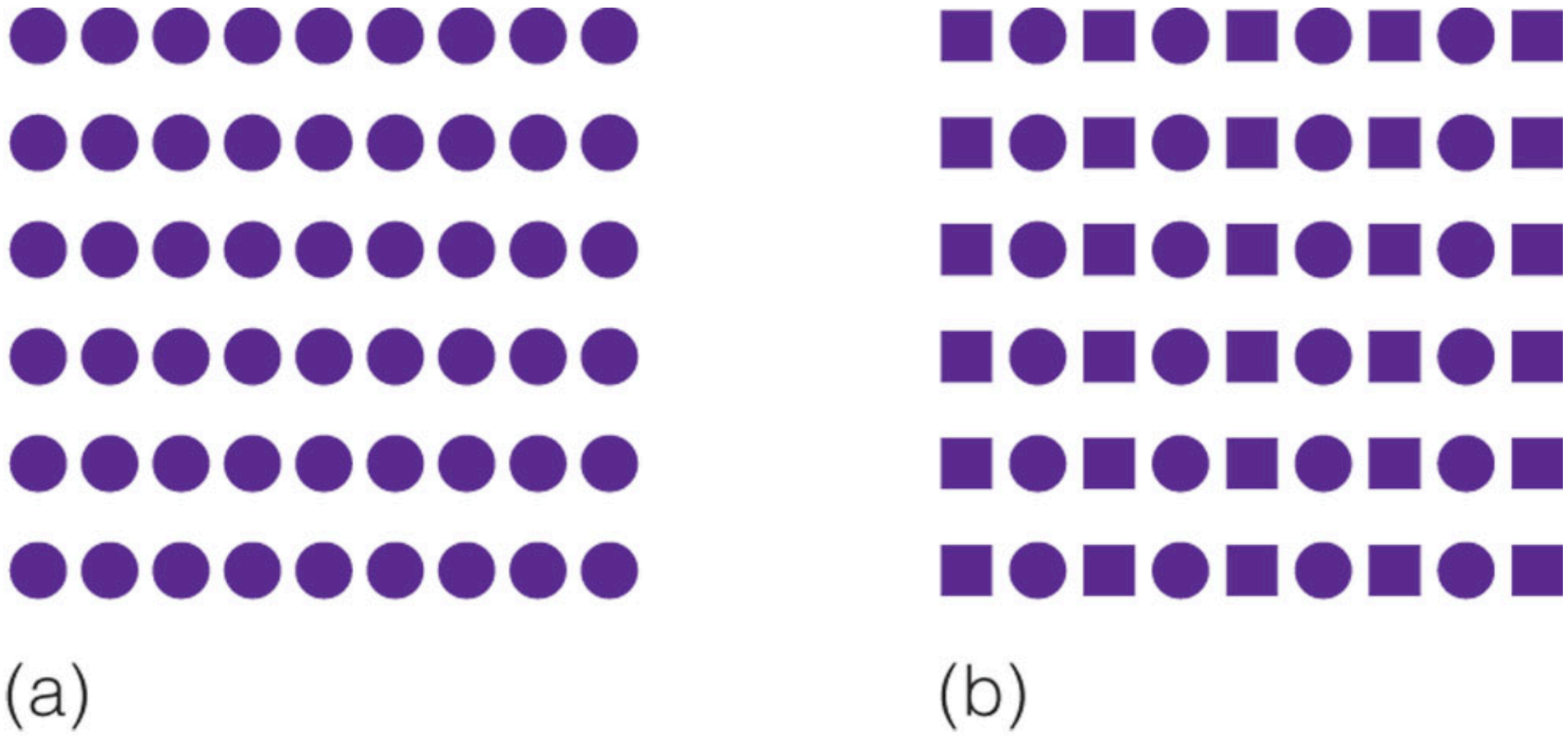
(a)



(b)

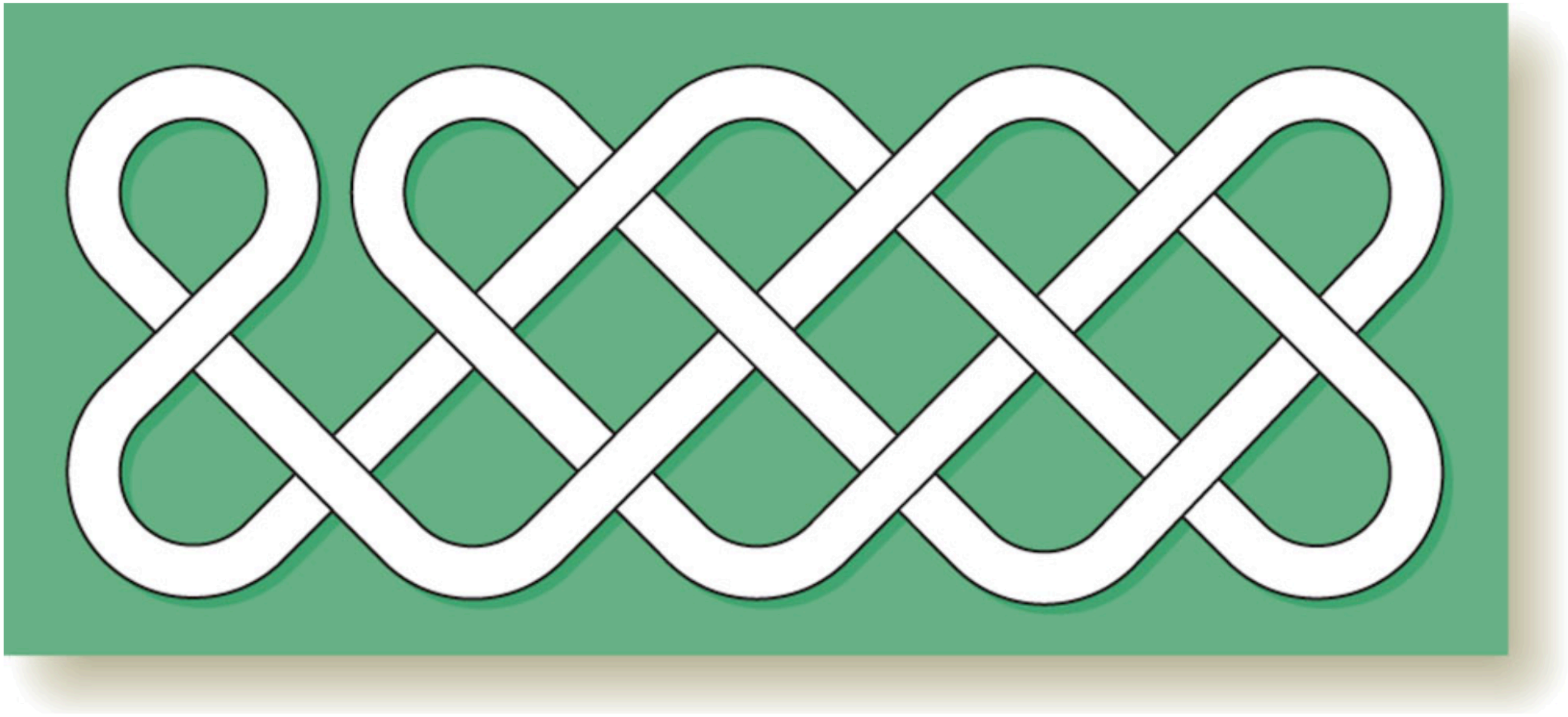
Gestalt laws

- Proximity: things that are near each other are grouped together



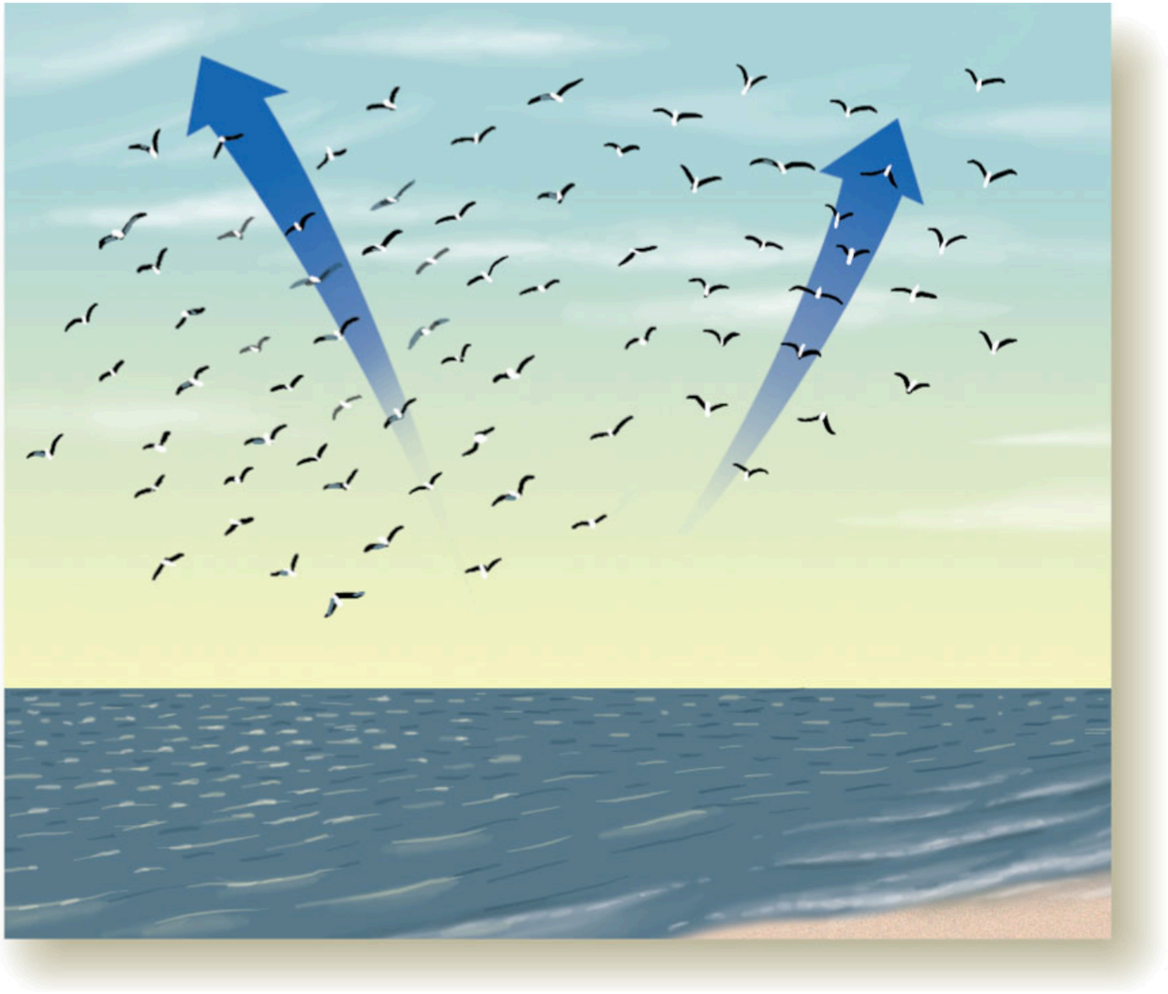
Gestalt laws

- Good continuation:
  - Connected points resulting in straight or smooth curves belong together
  - Lines are seen as following the smoothest path



Gestalt laws

- Common fate: things moving in the same direction are grouped together

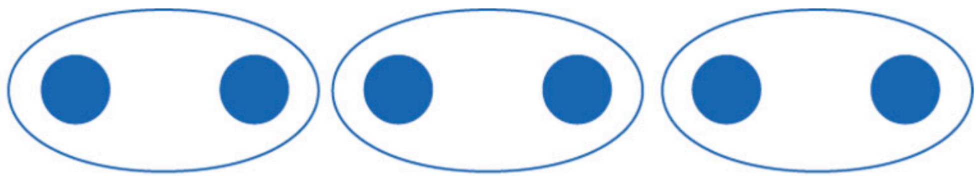


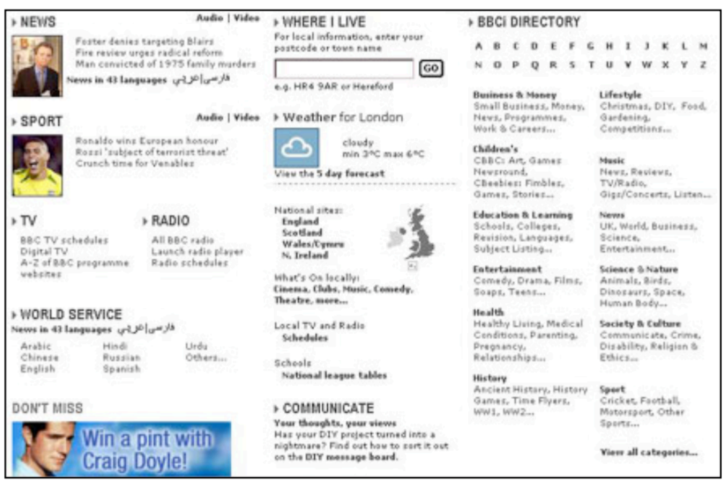
Gestalt laws

- Meaningfulness or familiarity: things form groups if they appear familiar or meaningful (top-down?)

Gestalt laws

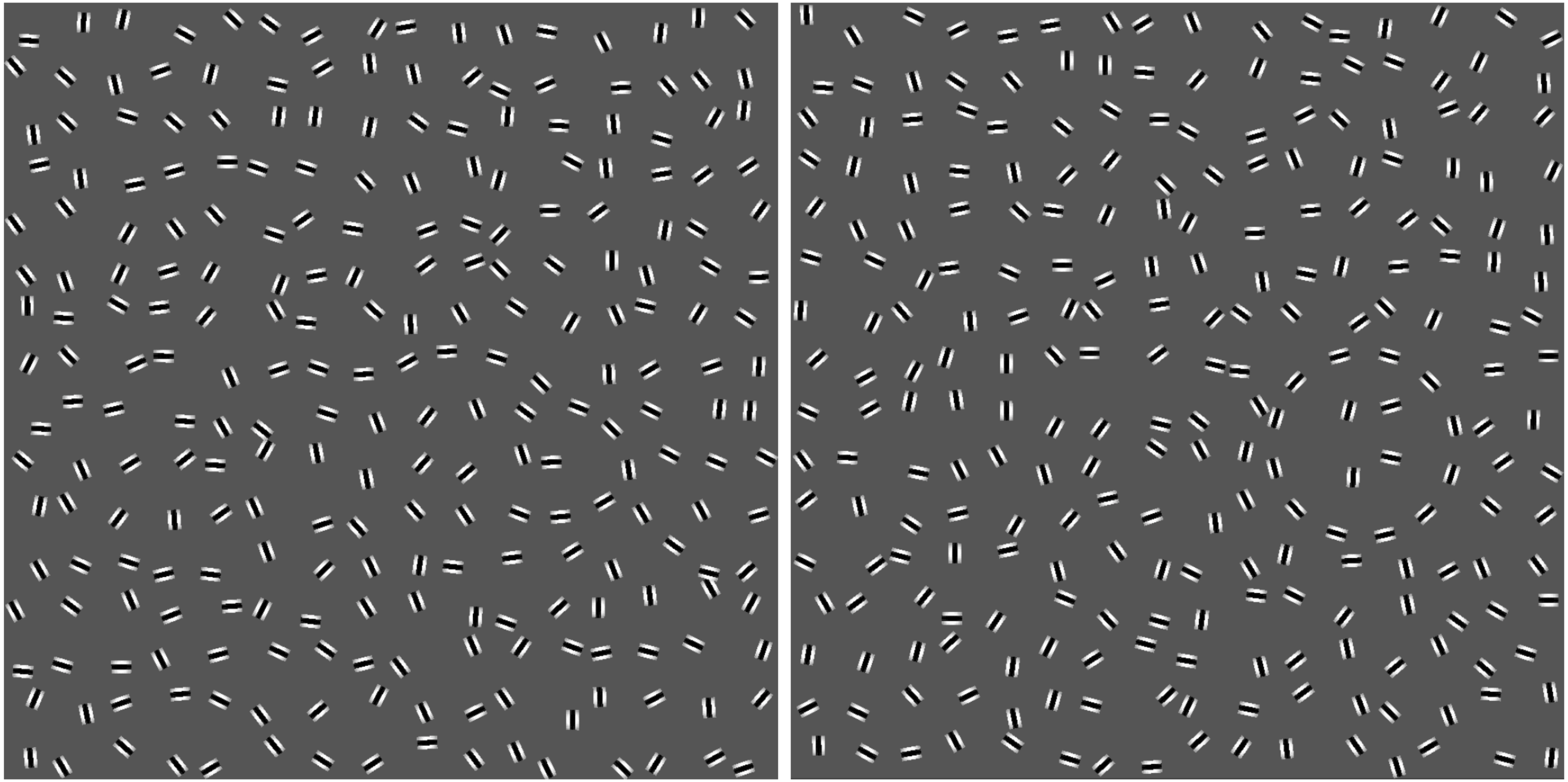
- Common region: elements in the same region tend to be grouped together





Gestalt laws

- Closure: a very strong grouping cue



Gestalt laws

- Uniform connectedness: connected regions of visual properties are perceived as single unit

- Synchrony: elements occurring at the same time are seen as belonging together

Inattentive blindness

- https://www.youtube.com/watch?v=ubNF9QNEQLA
- http://www.youtube.com/watch?v=LBL2SbRj5CA
- https://www.youtube.com/watch?v=wBoMjORwA-4

Visual attention

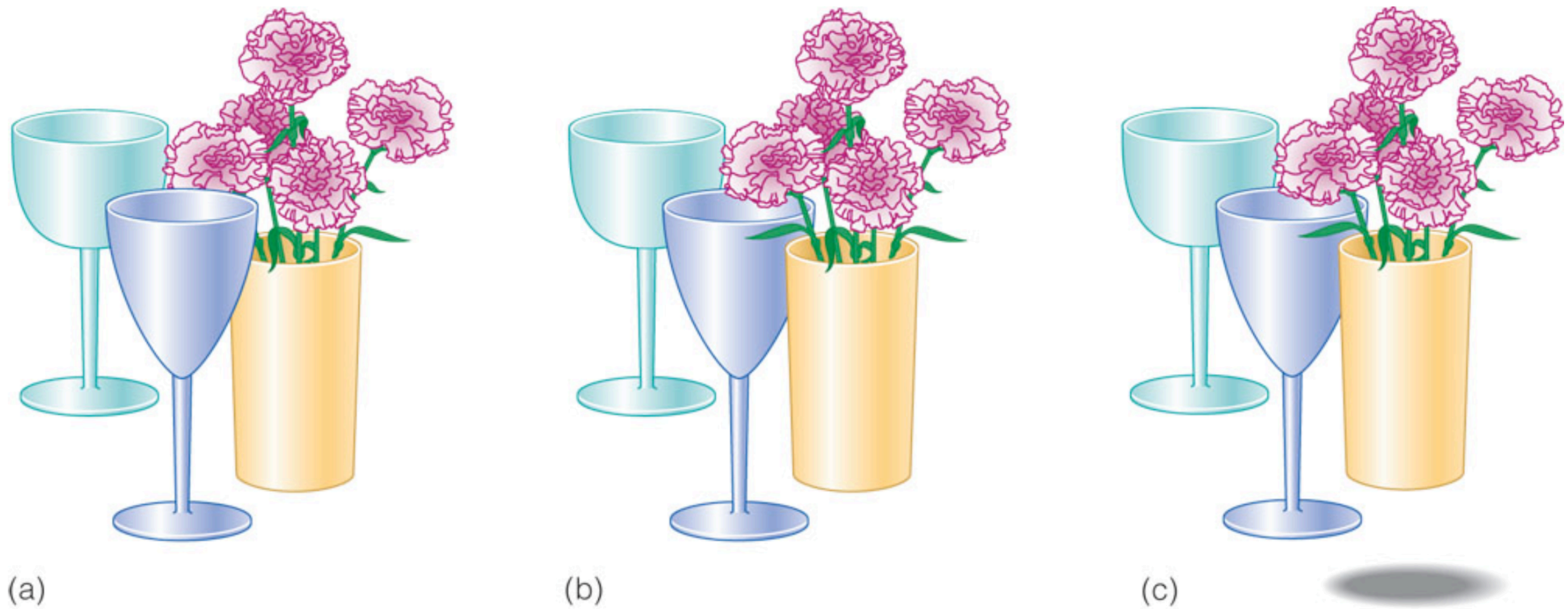
- Why do we pay attention to some parts of a scene but not to others?
- Do we have to pay attention to something to perceive it?
- Does paying attention to an object make the object “stand out”?

Stimulus salience:

- Based on the detection of areas of stimuli that attract attention due to their properties. Colour, contrast, and orientation are relevant properties
- Saliency maps show fixations are related to such properties in the initial scanning process
- Bottom-up process that is unrelated to meaning

Depth perception

Objects are the same in the 3 cases, but occlusion and shadows determine different perceived depths



Depth perception

- Retinal image is two-dimensional so depth cues are required to perceive 3D
- Depth cues can be combined or be put into conflict
  - Oculomotor cues: feedback to the brain from our eye muscles (convergence and accommodation)
  - Monocular depth cues: occlusion, relative height, relative size, perspective convergence, familiar size, atmospheric perspective, texture gradients, shadows, shading, specular reflections
  - Binocular disparity: differences in image location of an object seen by the left and right eye
  - Motion produced cues: motion parallax (observer moves and looks at relative changes on position of static objects), deletion (object disappears behind others) and accretion (object appears behind others)

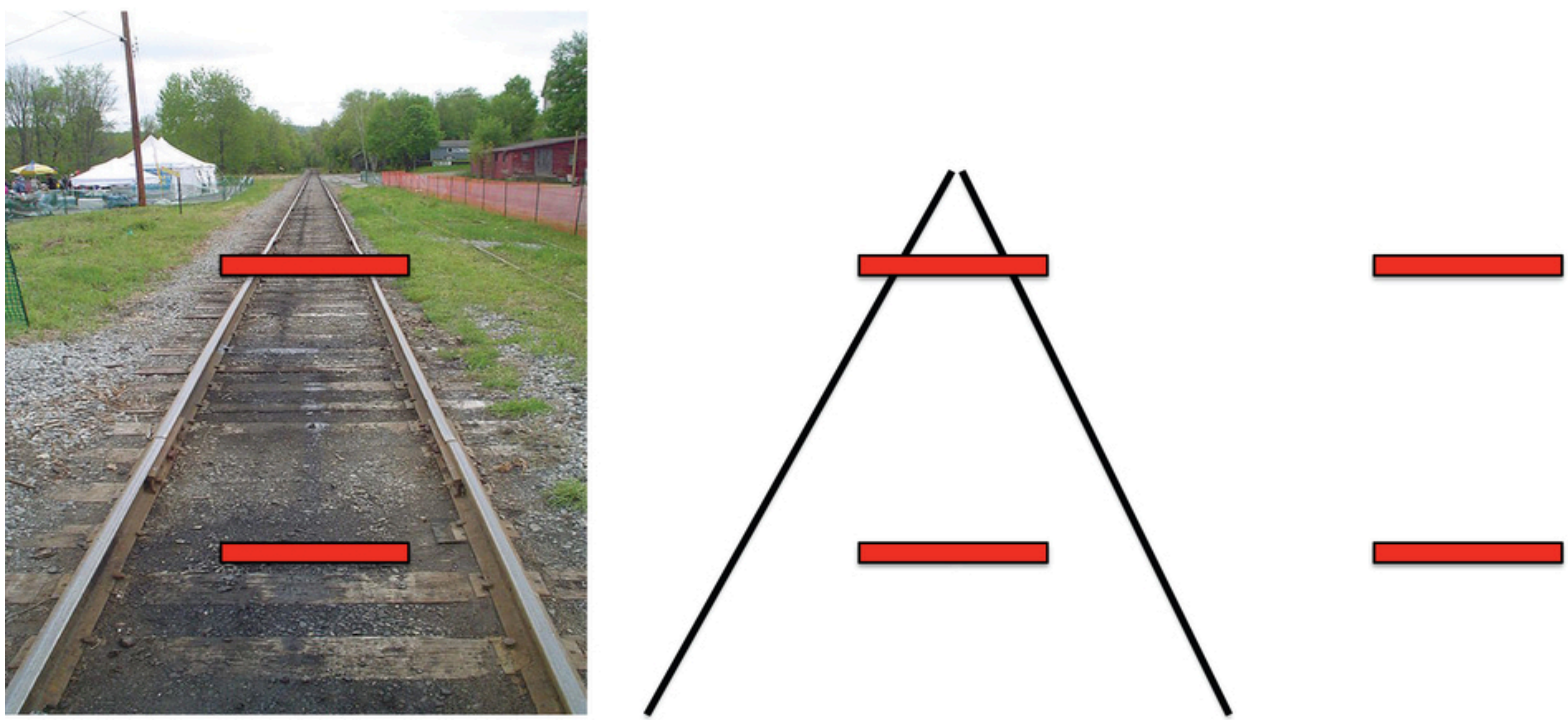
Depth perception

- The two red segments seem of different lengths, but they are equal
- In nature, this would be an unlikely scenario

a. The original scene with all its complex detail:

b. Some of the cues and the relevant information:

c. Relevant information without the cues:



Motion perception

- Why do some animals freeze in place when they sense danger?
- How do films create movement from still pictures?
- When we scan a room, the image of the room moves across the retina, but we perceive the room and the objects as remaining stationary. Why does this occur?

Types of motion perception:

- Real (physical) movement of an object
- Apparent movement (Phi phenomenon)
- Stationary stimuli presented in slightly different locations
- Movies, TV

Motion perception

Motion can be used to organize perception

- Biological motion: https://www.youtube.com/watch?v=f8TFi6qvPbc
- Neurons in areas as early as V1 are tuned to motion
- Neurons in MT are responsible for general motion perception, while neurons in STS are responsible for biological motion
- Neurons in MT integrate information from multiple V1 neurons in order to resolve the aperture problem: http://web.mit.edu/persci/demos/Motion&Form/demos/one-square/one-square.html

What we have learned in this lecture

Human visual system:

- How the eye works
- Different theories for colour perception
- How the signal is transmitted from the eye to the visual cortex
- Different functions of the visual cortex parts
- Simple/complex cells
- Advantages & limitations
- Visual illusions
- Importance of learning for proper visual perception
- Gestalt laws: interpreting how perception is created
- Depth and motion perception cues