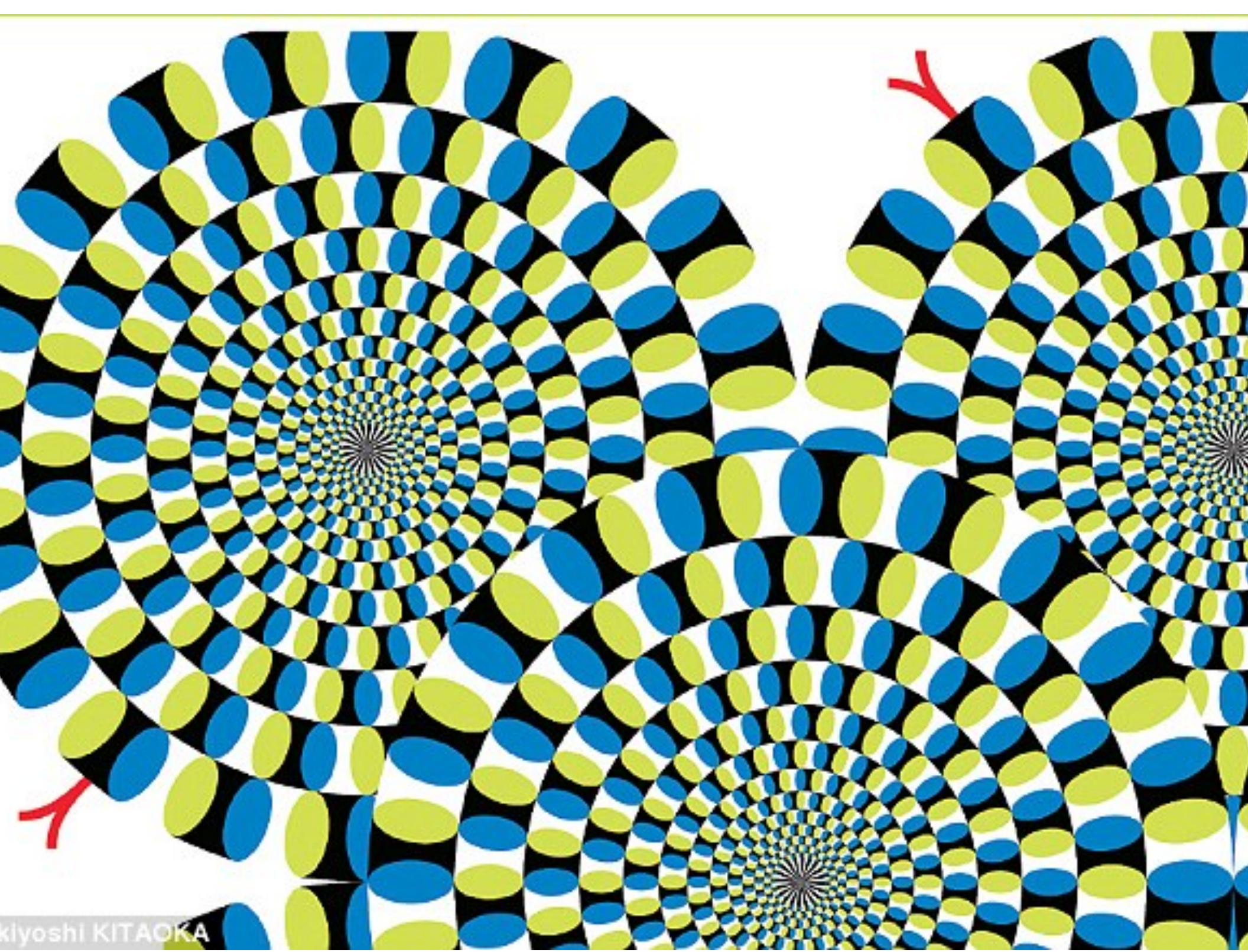

kiyoshi KITAOKA

Human Perception and Visual Cognition

Or

What You See is Maybe Not What You Were Supposed to Get

Seeing

=

Sensing
Selecting
Perceiving

(Huxley)

Seeing

=

Sensing

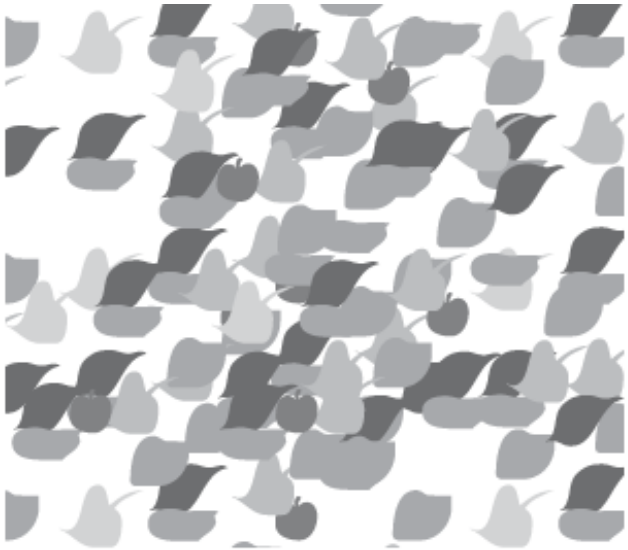
Selecting

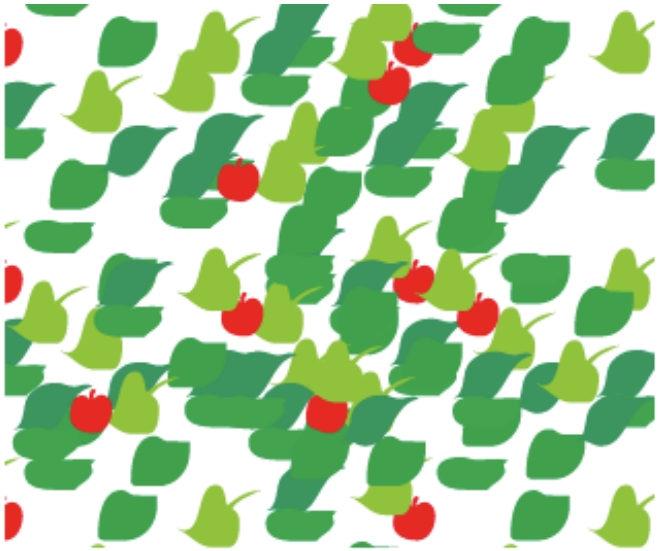
Perceiving

(Huxley)

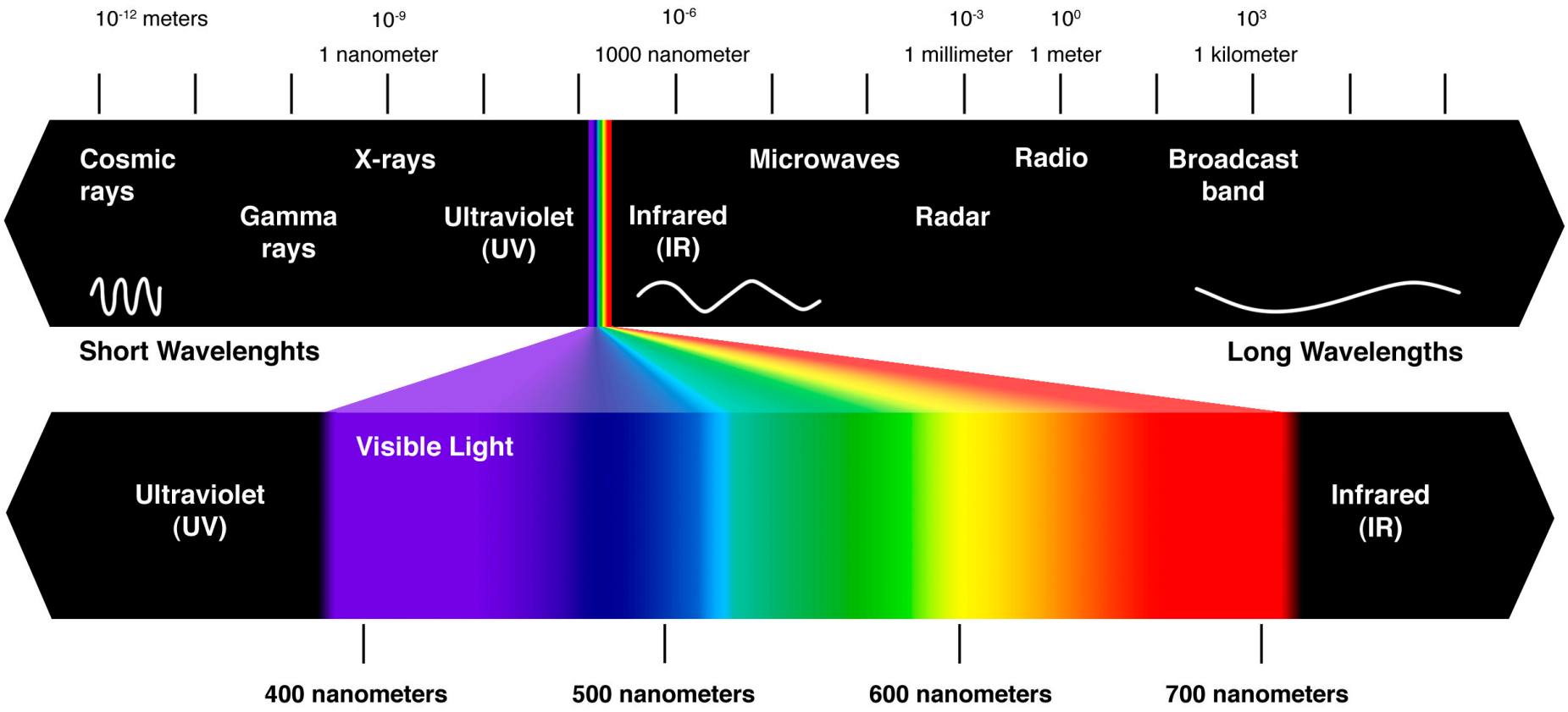
What's vision is for

- Survival (Camouflage) and Navigation (Where to go).
- Tool use (The material type and shpe)
- Food Seeking
  (Where they are, if they are good to be consume)

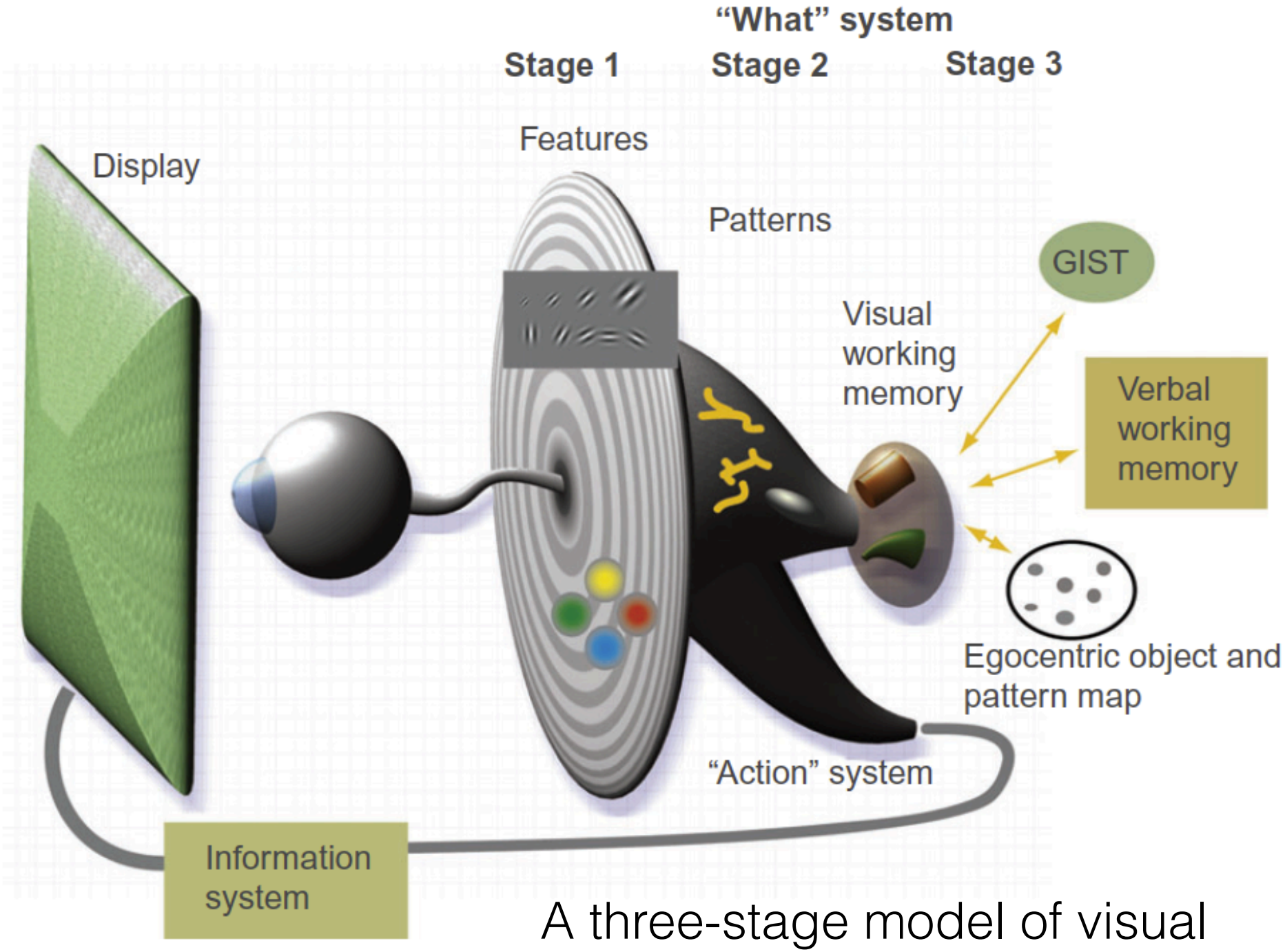




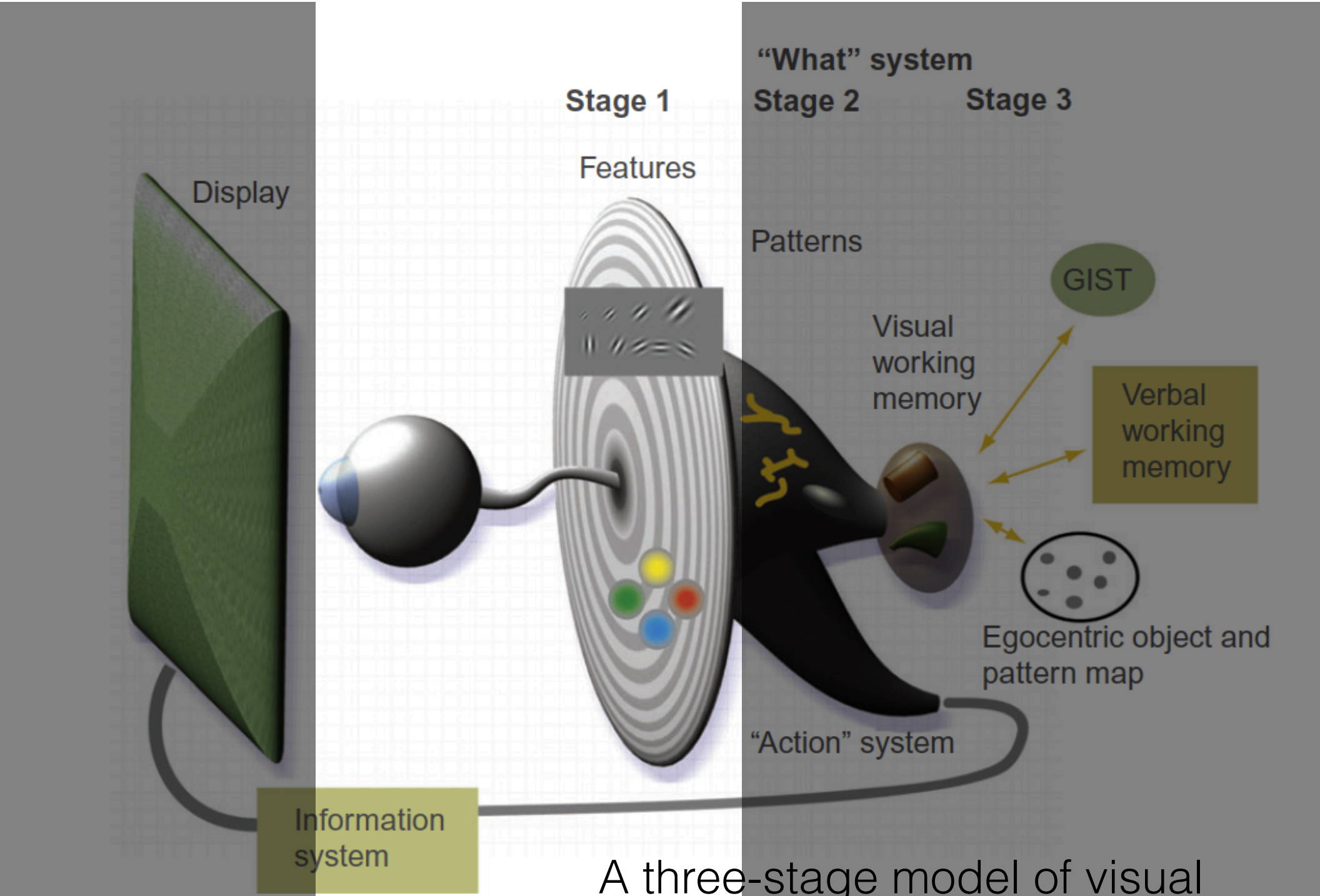
Light



- The visible light spectrum is a tiny part of a much larger spectrum of electromagnetic radiation.



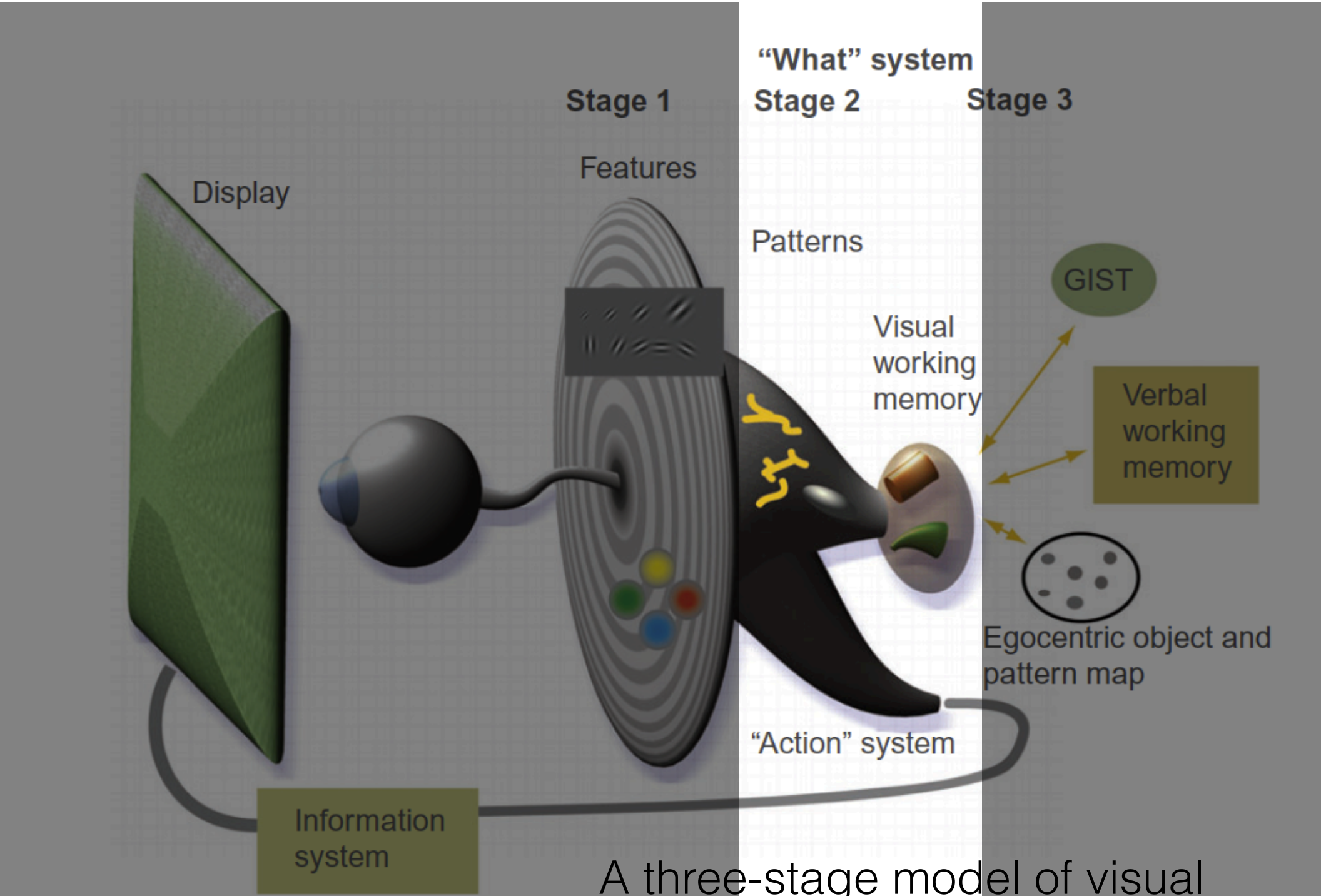
A three-stage model of visual information processing.



A three-stage model of visual information processing.

Stage 1 - Low-level, Parallel

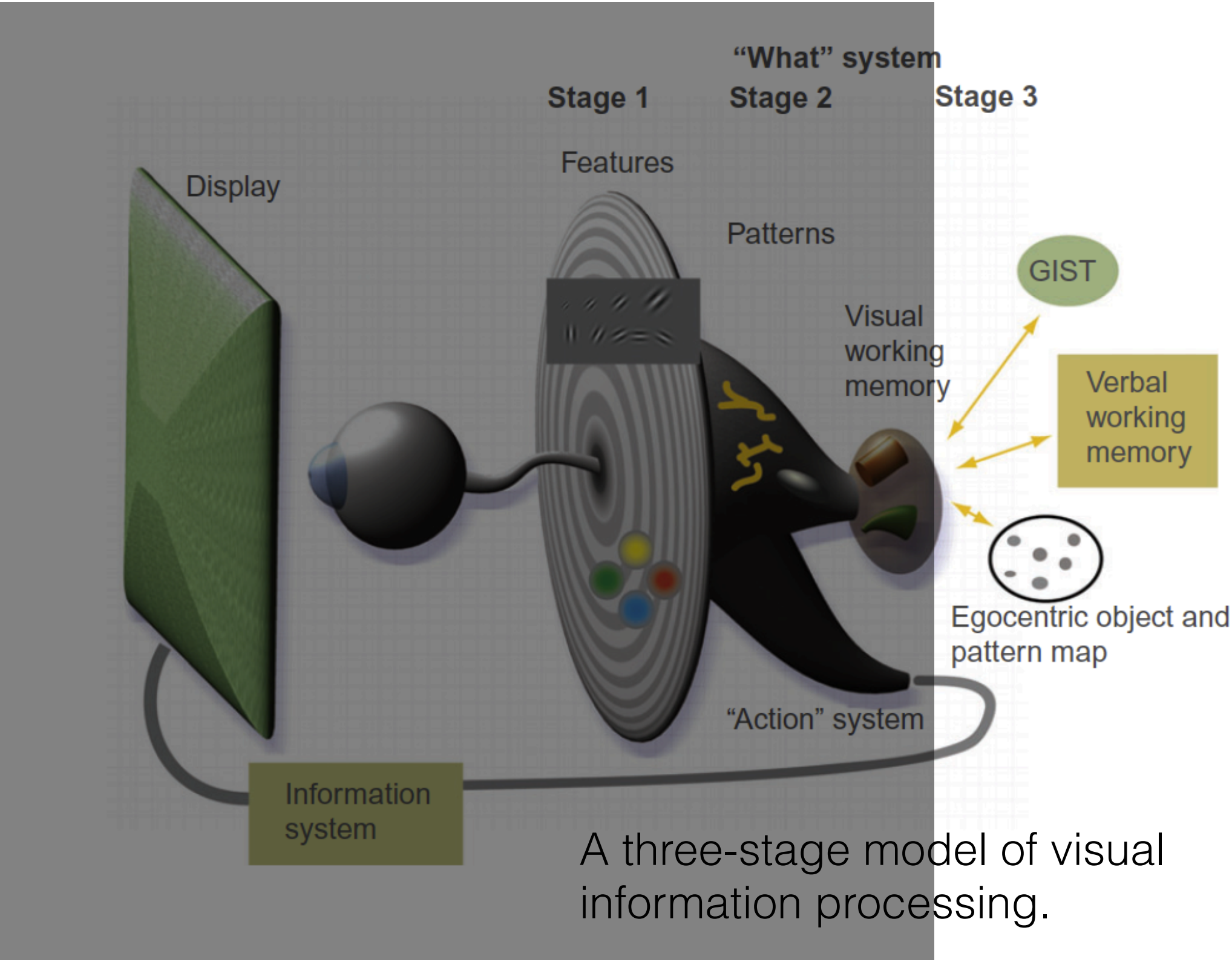
- Neurons in eye & brain responsible for different kinds of information:
- Extraction of features, orientation, colour, texture, and movement patterns
- Rapid parallel processing
- Occurs “automatically”
- Information is transitory, briefly held in iconic store
- Bottom-up data-driven model of processing
- Often called “pre-attentive” processing



A three-stage model of visual information processing.

Stage 2 - Low-level, Serial

- Slower serial processing
- Top-down attention being critical to the formation of objects and patterns pulled out from the feature maps.
- A small number (one to three) patterns becoming “bound” and held for a second or two under top-down attentional processes
- Different pathways for object recognition and visually guided hand motion (the perception and action channels).



A three-stage model of visual information processing.

Stage 3 - high level

- Highest level of perception in which the objects are held in visual working memory by the demands of active attention.
- Only a few objects can be held at a time.
    - They are constructed from the available patterns that may provide answers to the visual query and from information stored in long-term memory related to the task.

- It seems linear, but it has a lot of back and forth.

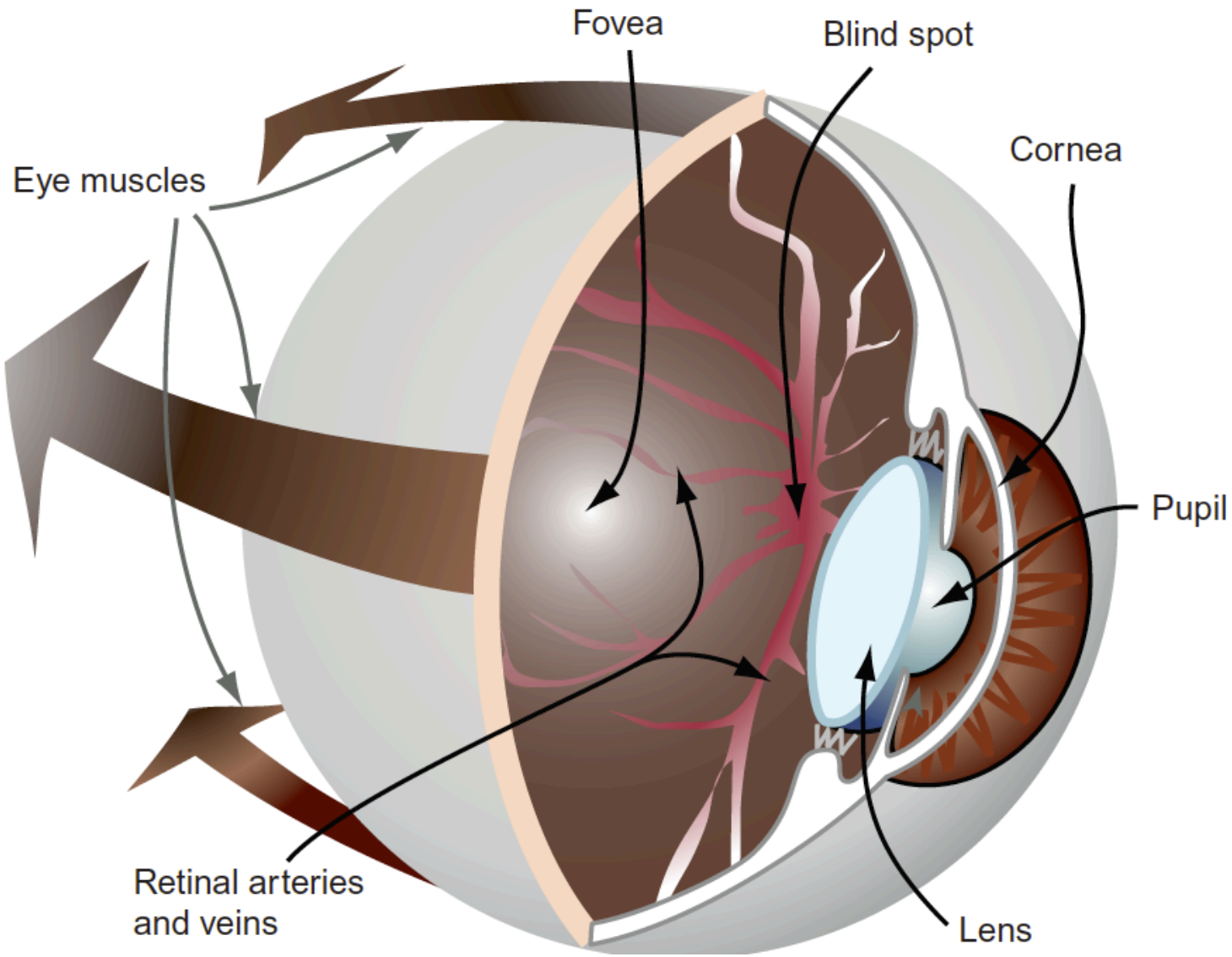
- The entire system is being constantly tuned from top to bottom based on our expectations and on what will be most useful to us.

- The generic name for this is

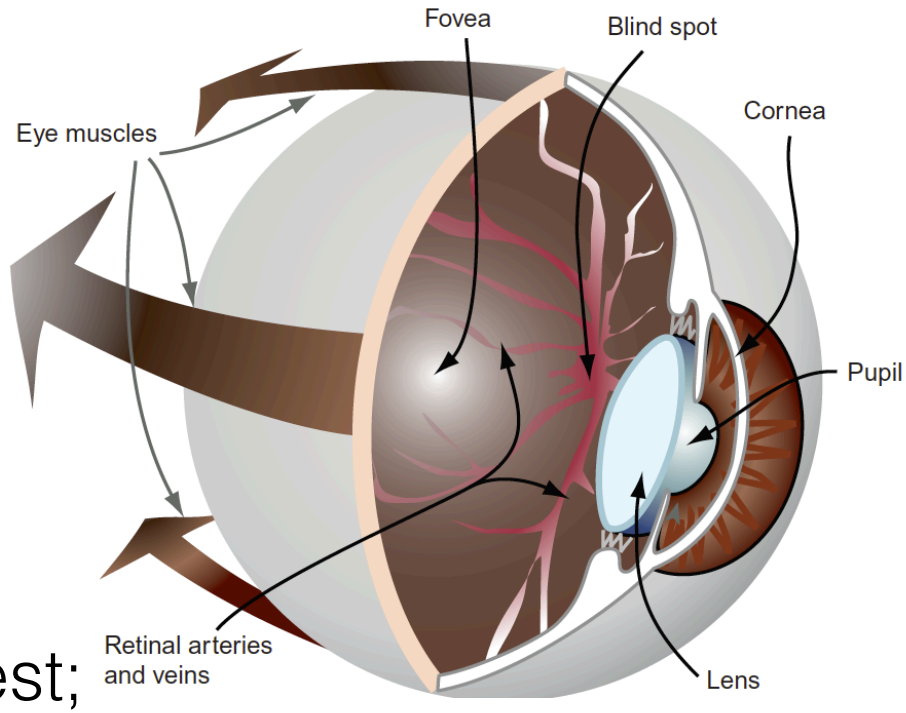
Attention

- A multifaceted pervasive set of processes involving the entire visual system.

Stage 1: Low Level: Visual Features

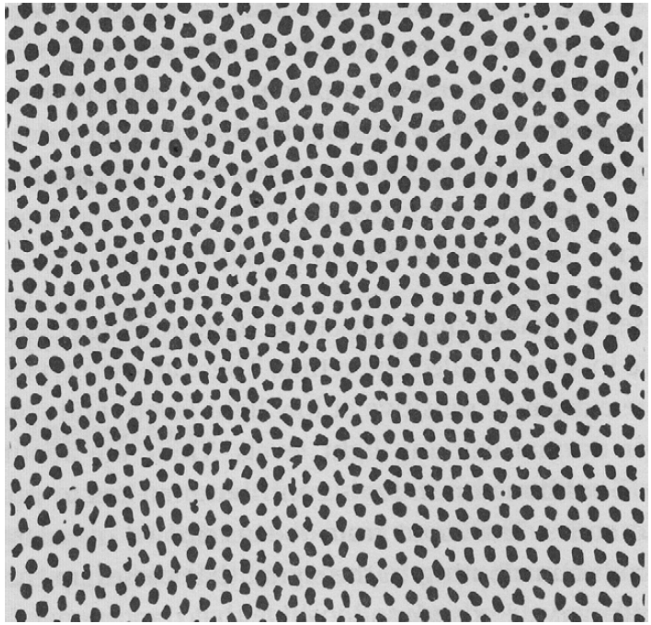


The Eye



- Important features include:
- The fovea, where vision is sharpest;
- the pupil, a round aperture through which light enters the eye;
- the two principal optical elements, the lens and the cornea;
- and the large eye muscles that control eye movements.
- This blind spot is caused by the absence of receptors where the retinal arteries enter the eyeball.

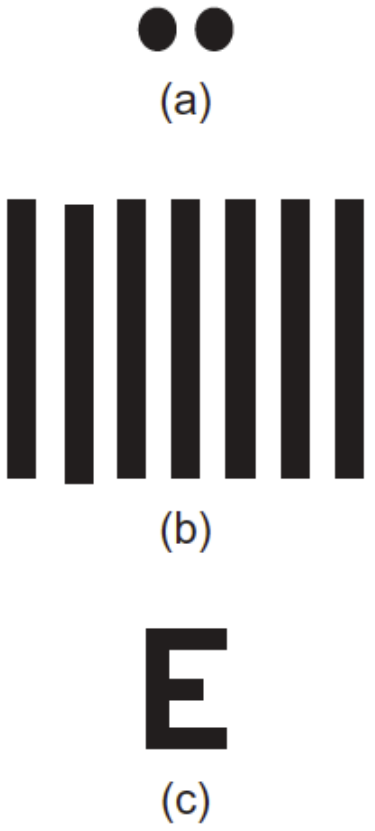
Receptors / Fovea



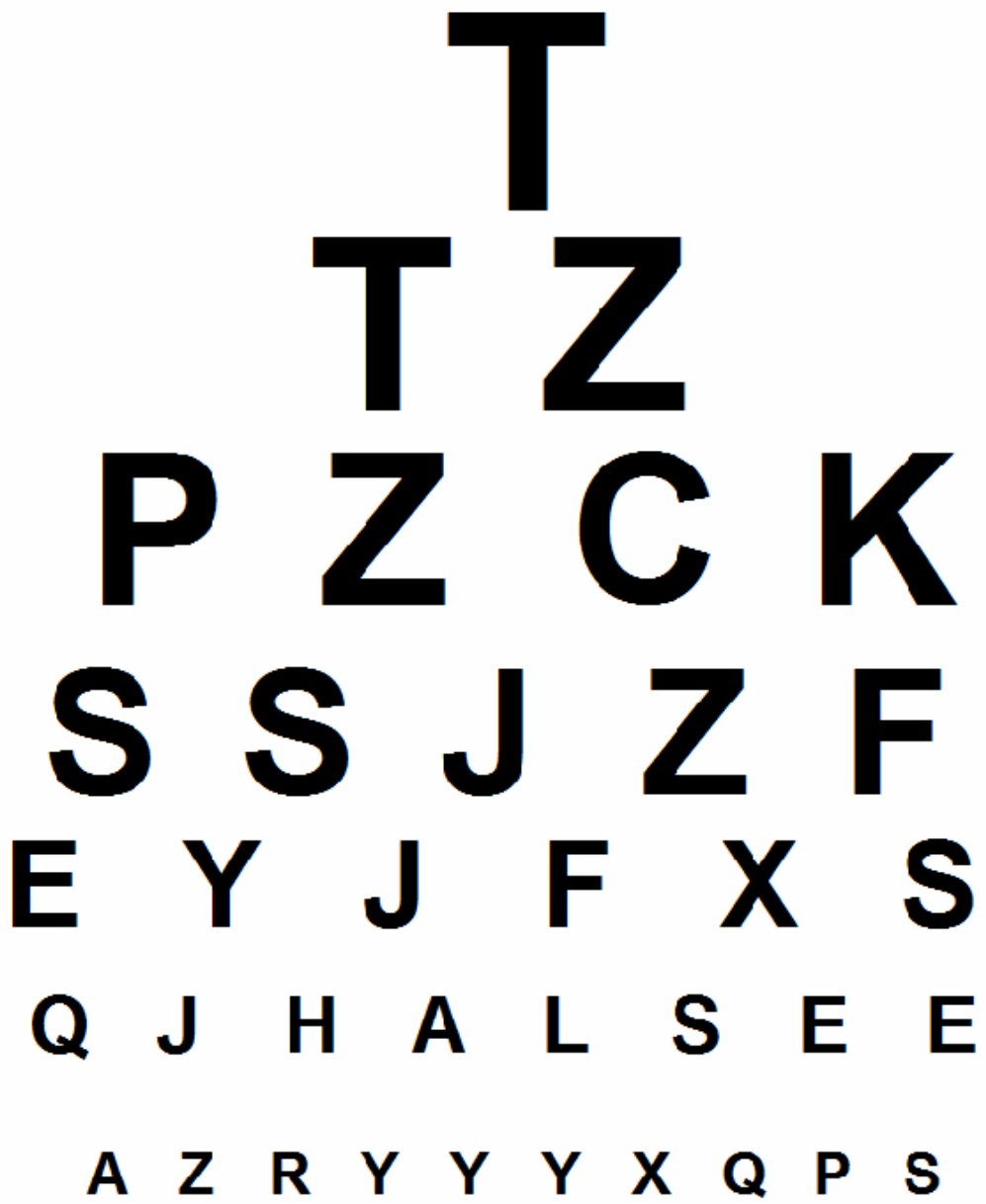
- The receptor mosaic in the fovea: 100,000 cones packed into this central small area
- Visual area: 1.5 to 2 degrees (The thumb test)

Basic Visual Acuities

- (a) Point acuity (1 minute of arc = 1/60 degree): The ability to resolve two distinct point targets.
- (b) Grating acuity (1 to 2 minutes of arc): The ability to distinguish a pattern of bright and dark bars from a uniform gray patch.
- (c) Letter acuity (5 minutes of arc): The ability to resolve letters.

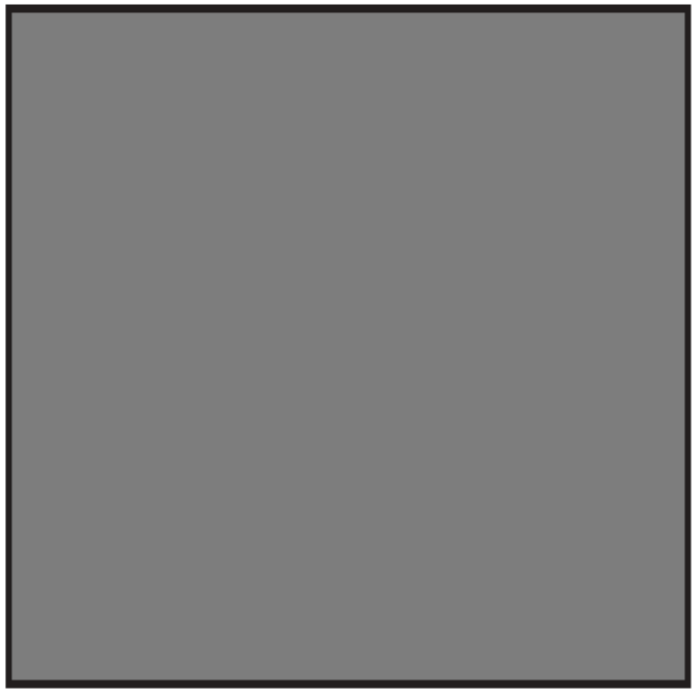


Basic Visual Acuities

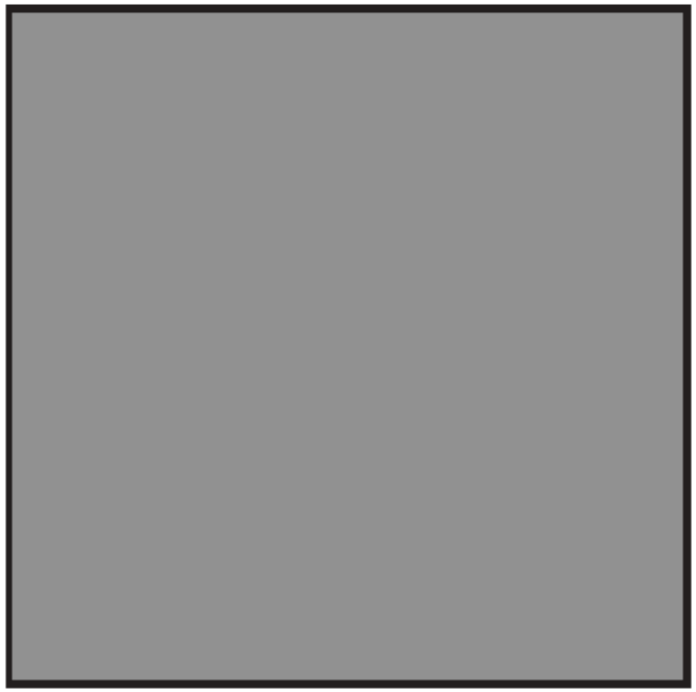


Detecting differences

- Which is brighter?



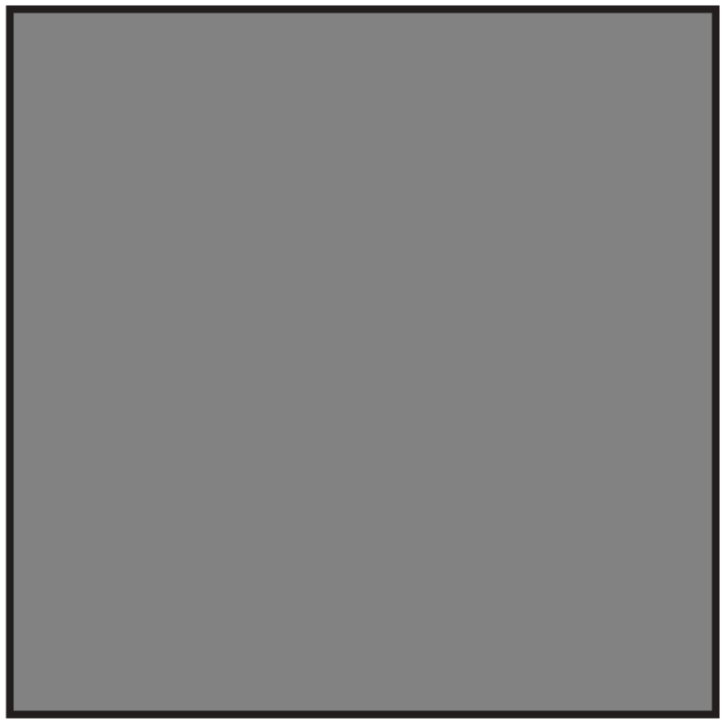
128, 128, 128



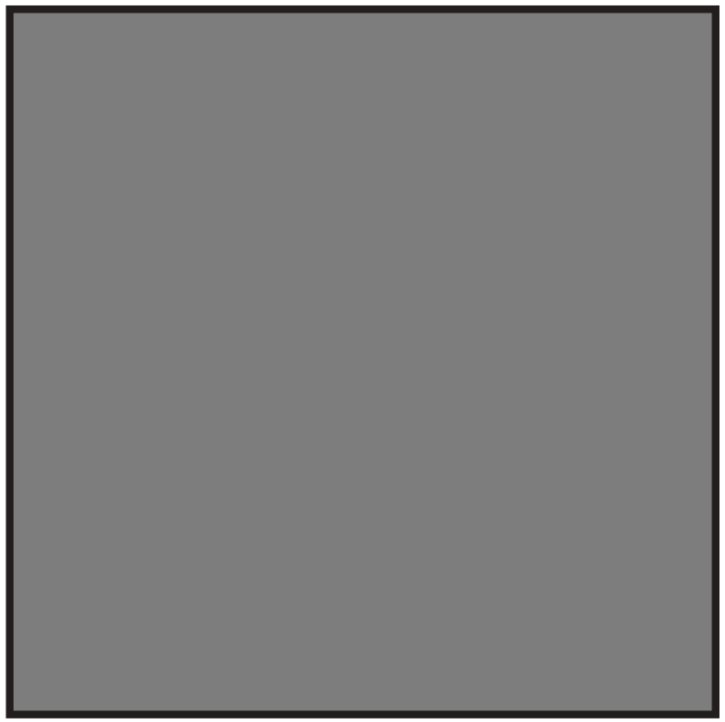
144, 144, 144

Detecting differences

- Which is brighter?



134, 134, 134



128, 128, 128

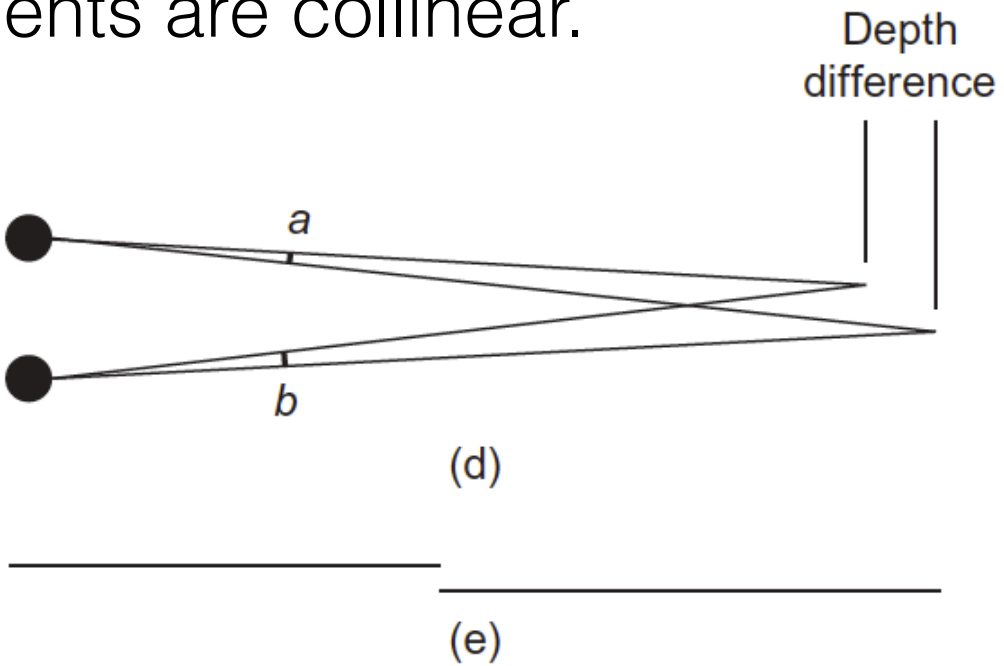
Just Noticeable Difference

- The smallest detectable difference between equally spaced levels of a stimulus
- Ratios are more important than magnitude
- Most continuous variation perceived in distinct steps

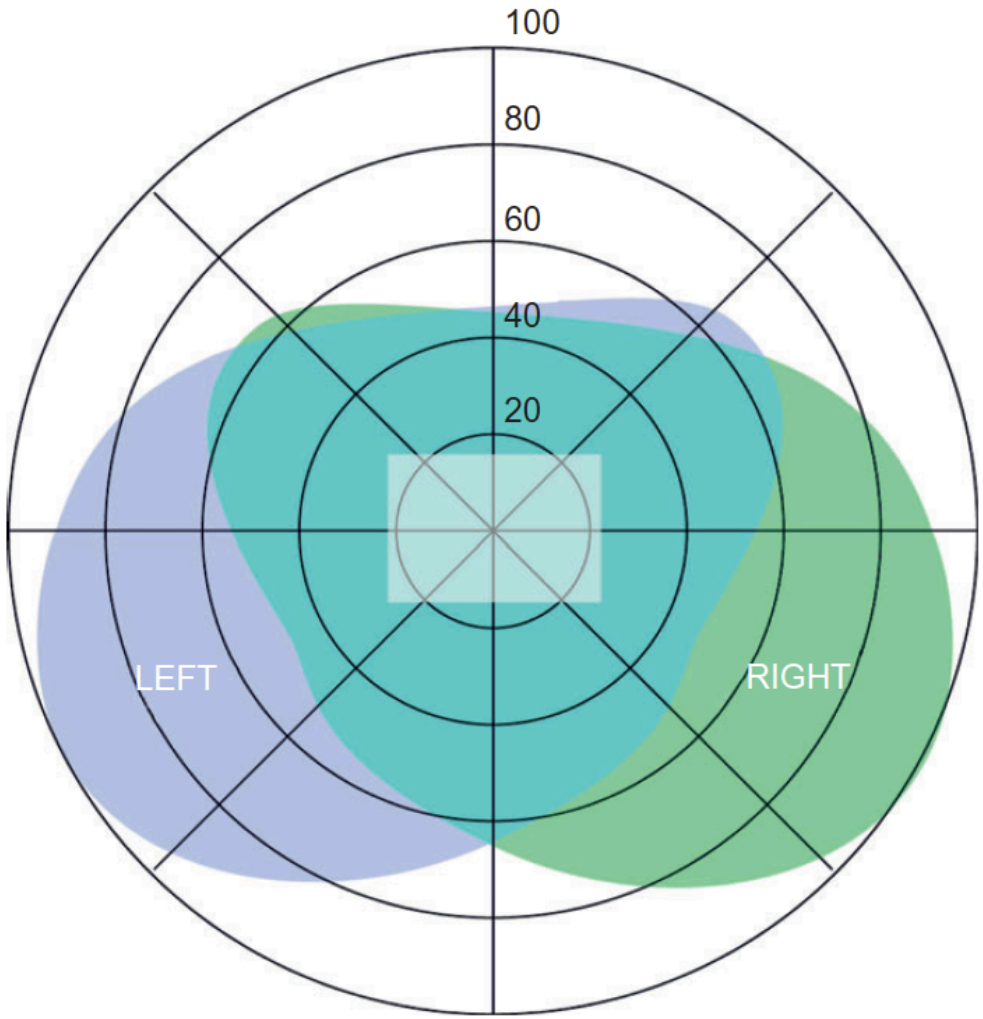
Super Visual Acuities

- (d) Stereo acuity (10 seconds of arc): The ability to resolve depth. The acuity is measured as the difference between two angles (a and b).

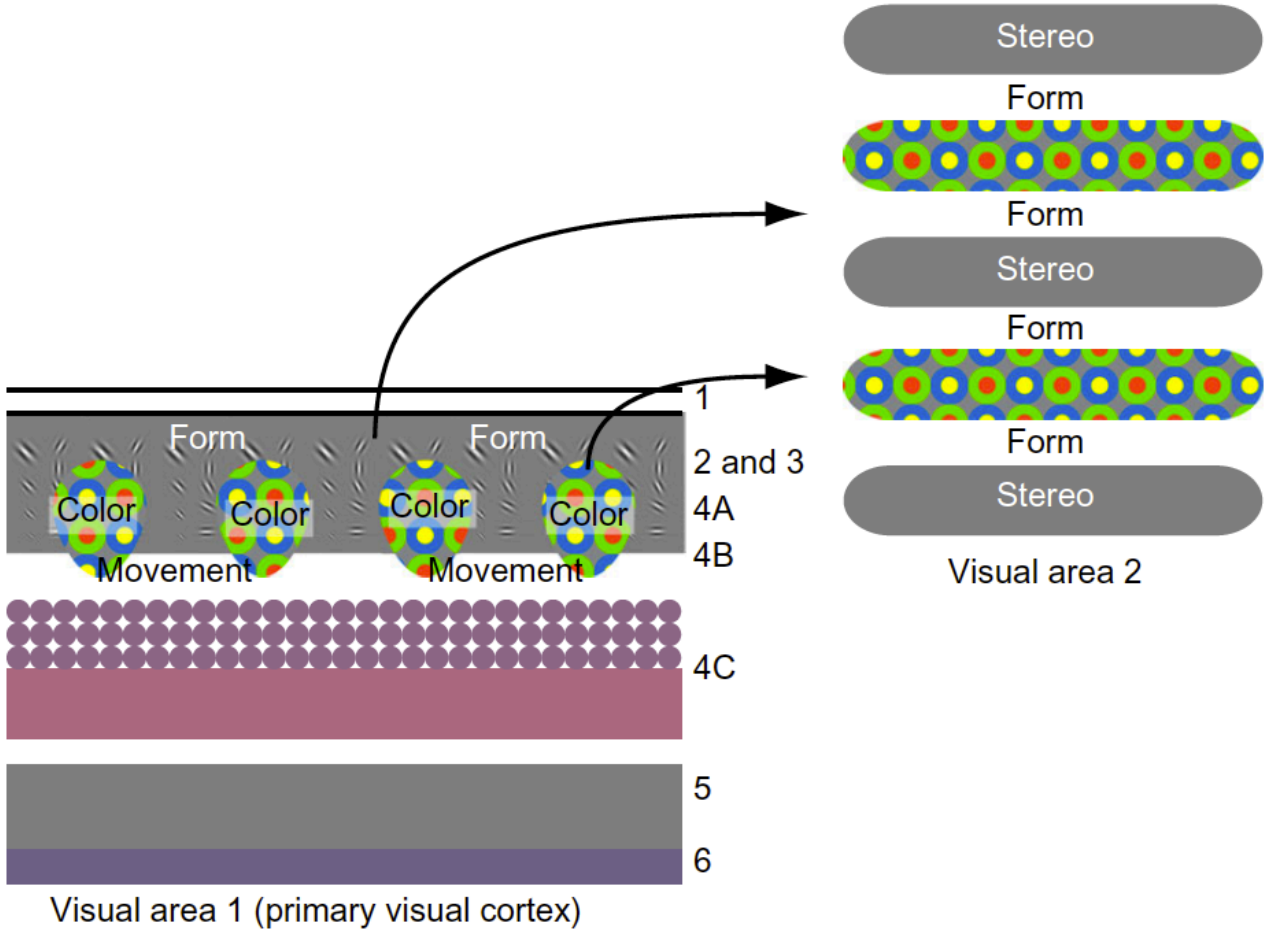
- (e) Vernier acuity (10 seconds of arc): The ability to see if two line segments are collinear.



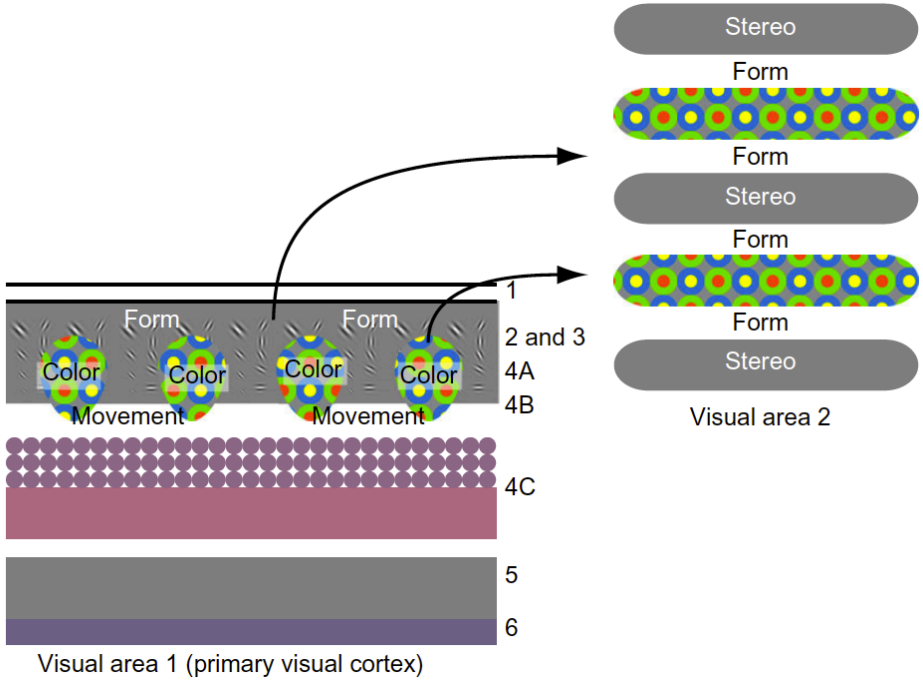
Visual Field



Feature Maps and Channels



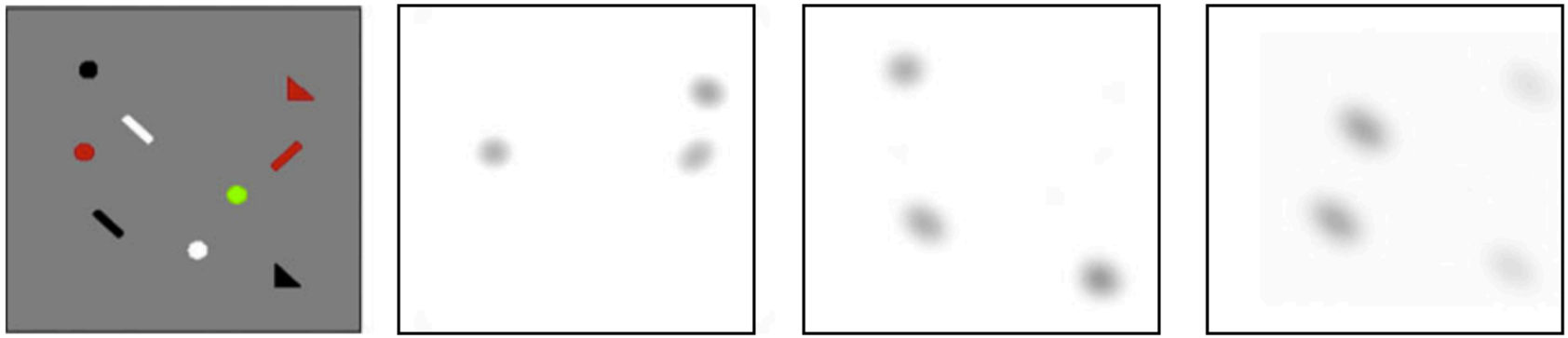
The Elements of Form



- Colour, motion, and the elements of form (orientation and size) are processed separately and therefore are easy to visually separate.

- These different properties are said to have different channels, meaning that information expressed in one channel, the colour of a symbol, does not interfere with information expressed in another, the orientation of a symbol

Feature Maps and Channels



- Because different kinds of visual properties are processed separately
- There is a map for redness, a map for greenness, a map for vertical orientation, a map for horizontal orientation, a map for motion, and so on.

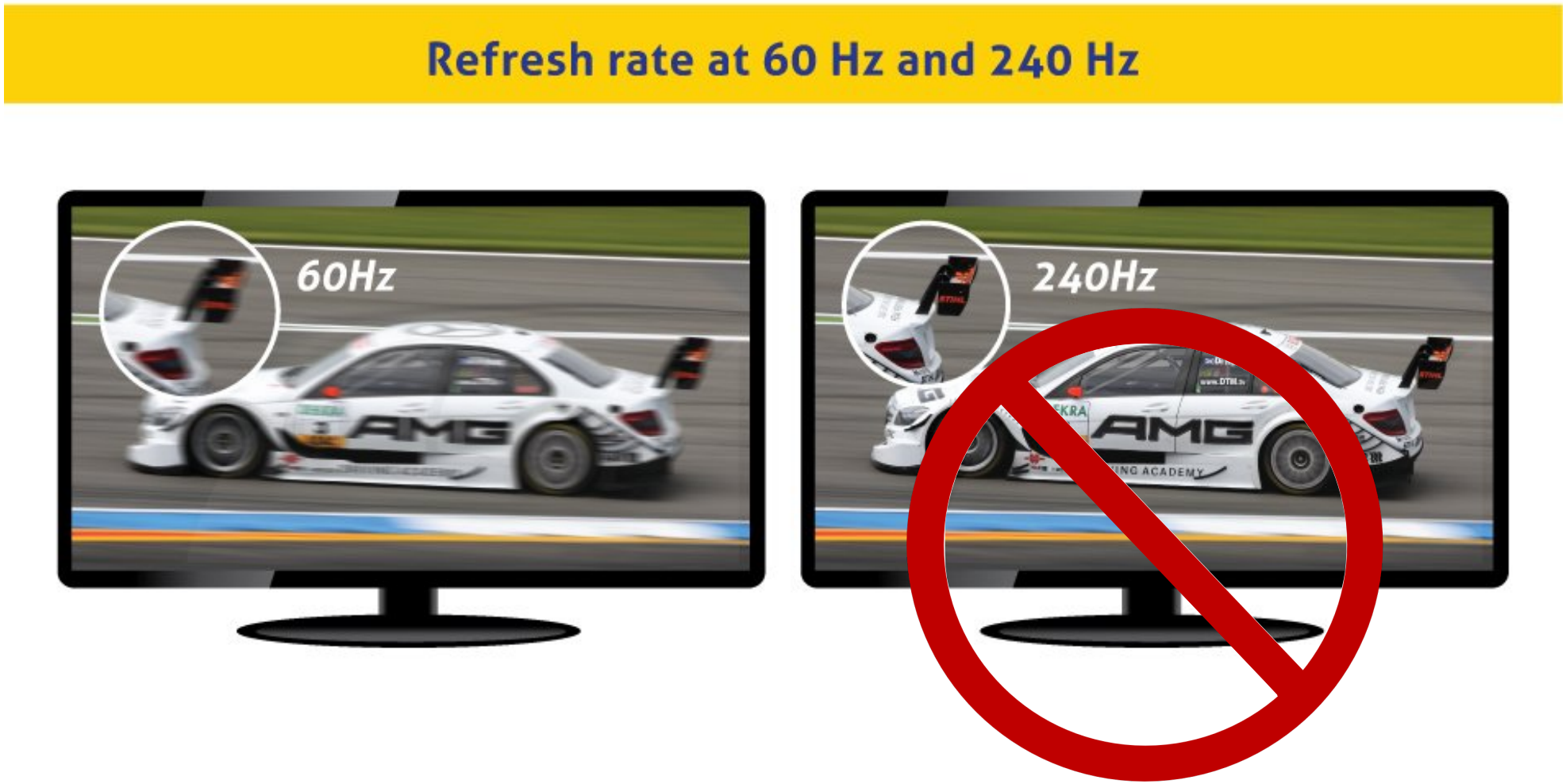
Flicker - Refresh rate

How many “frames” we see in one second?

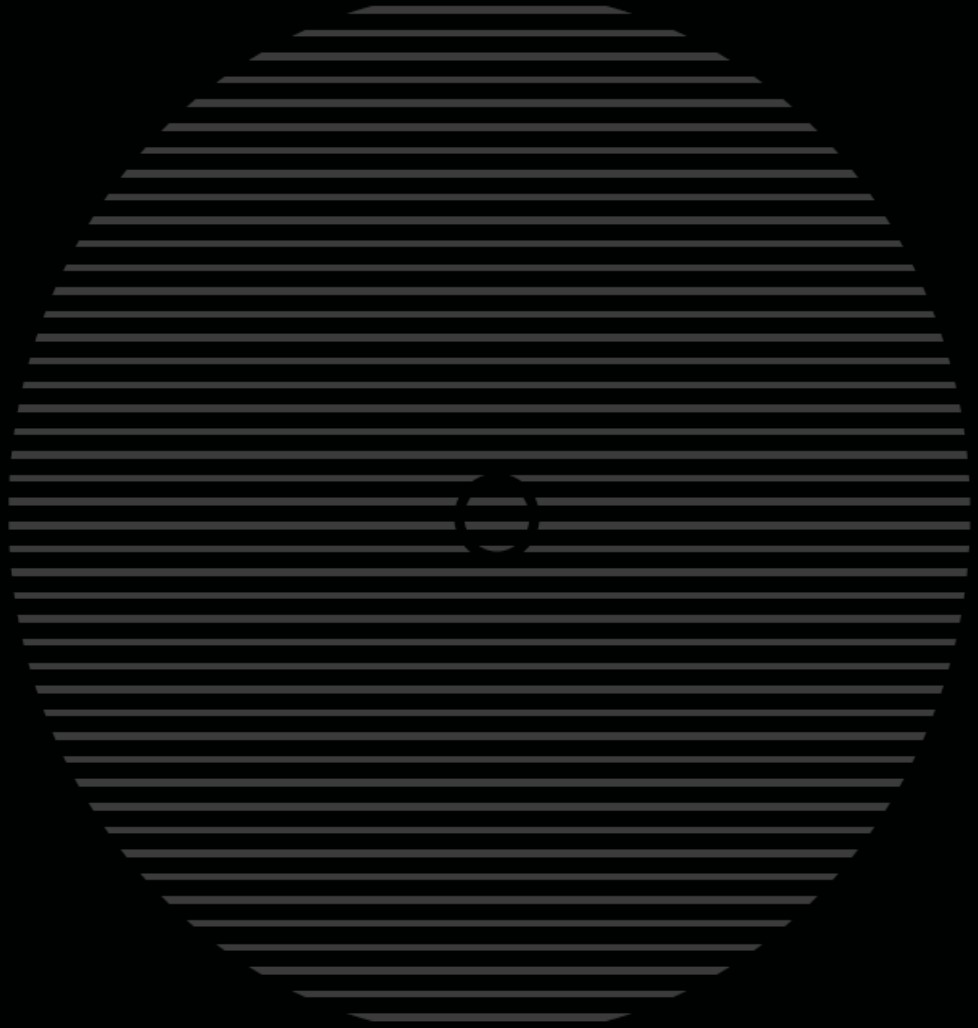
The “reversing wagon wheel”:
https://www.youtube.com/watch?v=6XwgbHjRo30

Flicker - Refresh rate

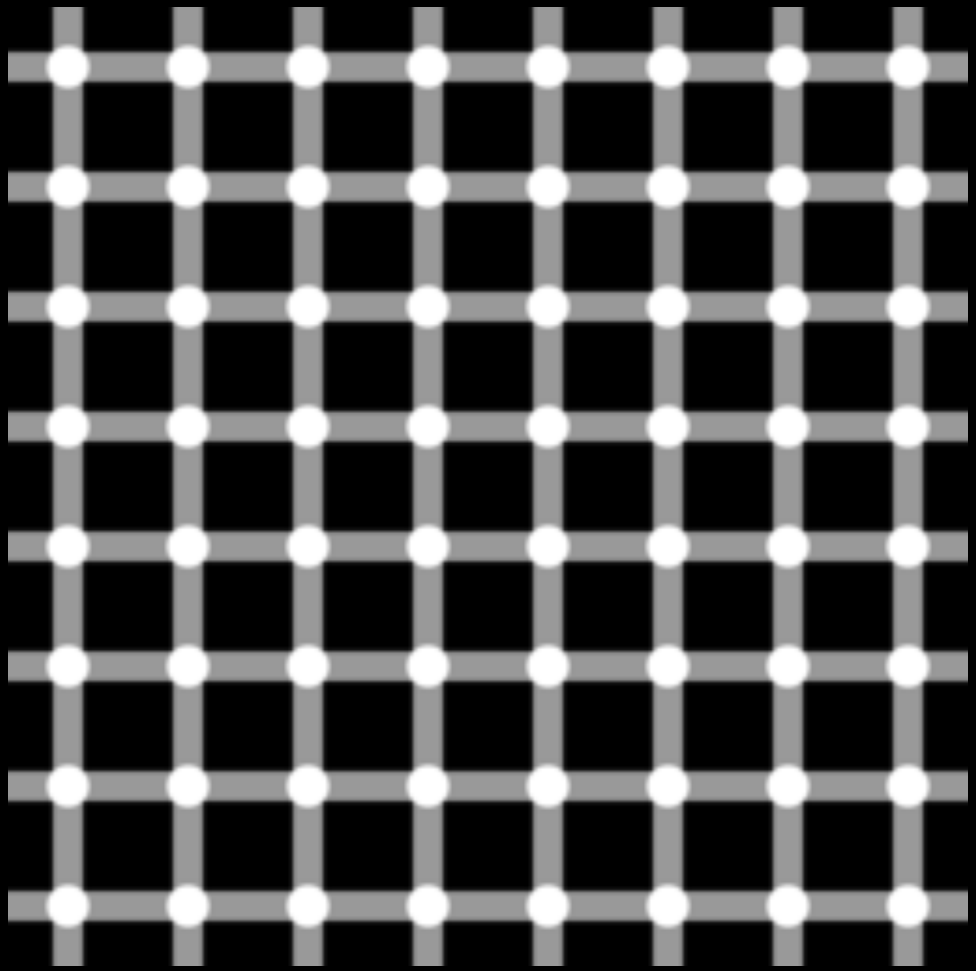
~ 50-Hz



Visual Stress



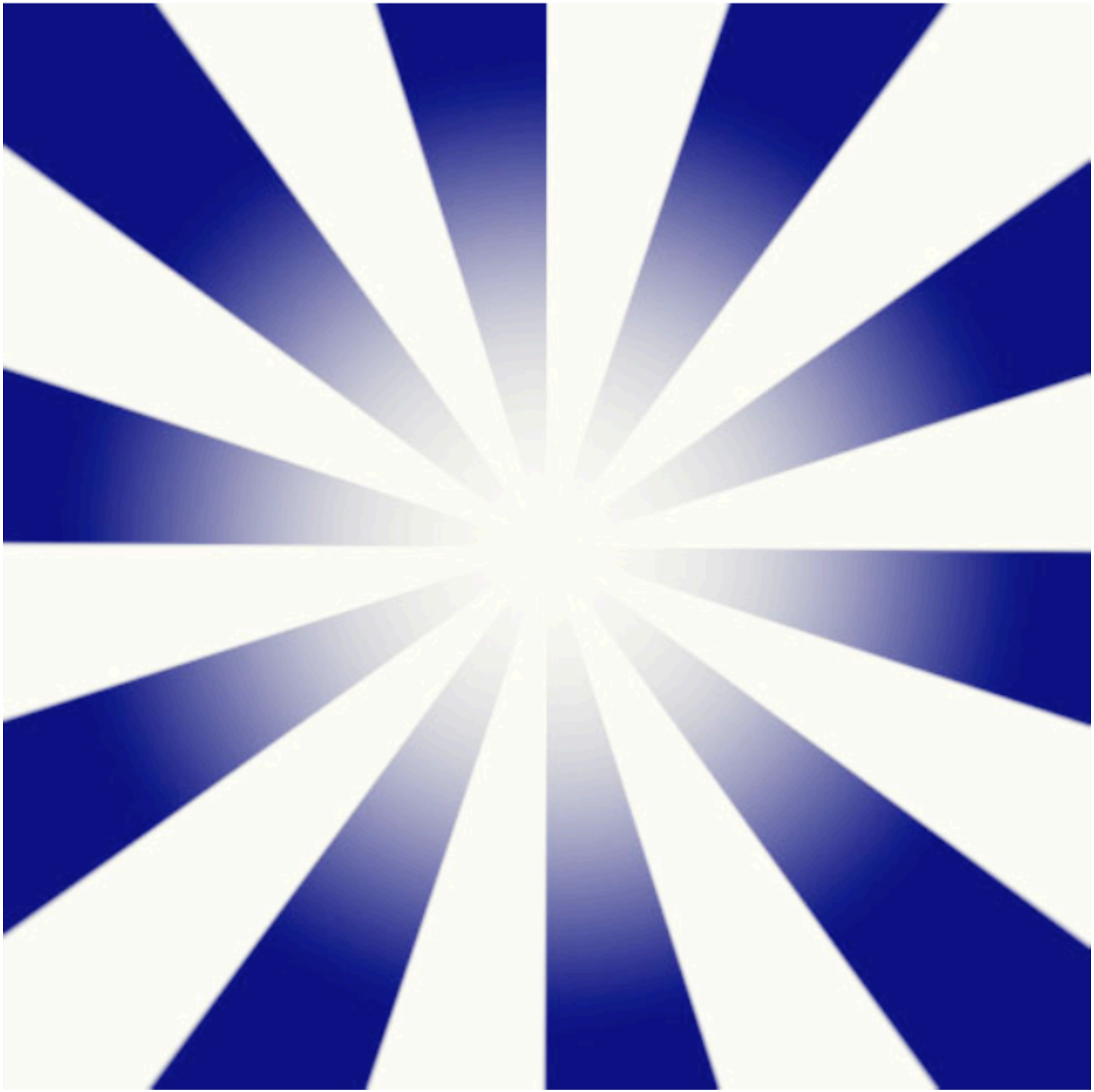
Optical Illusions



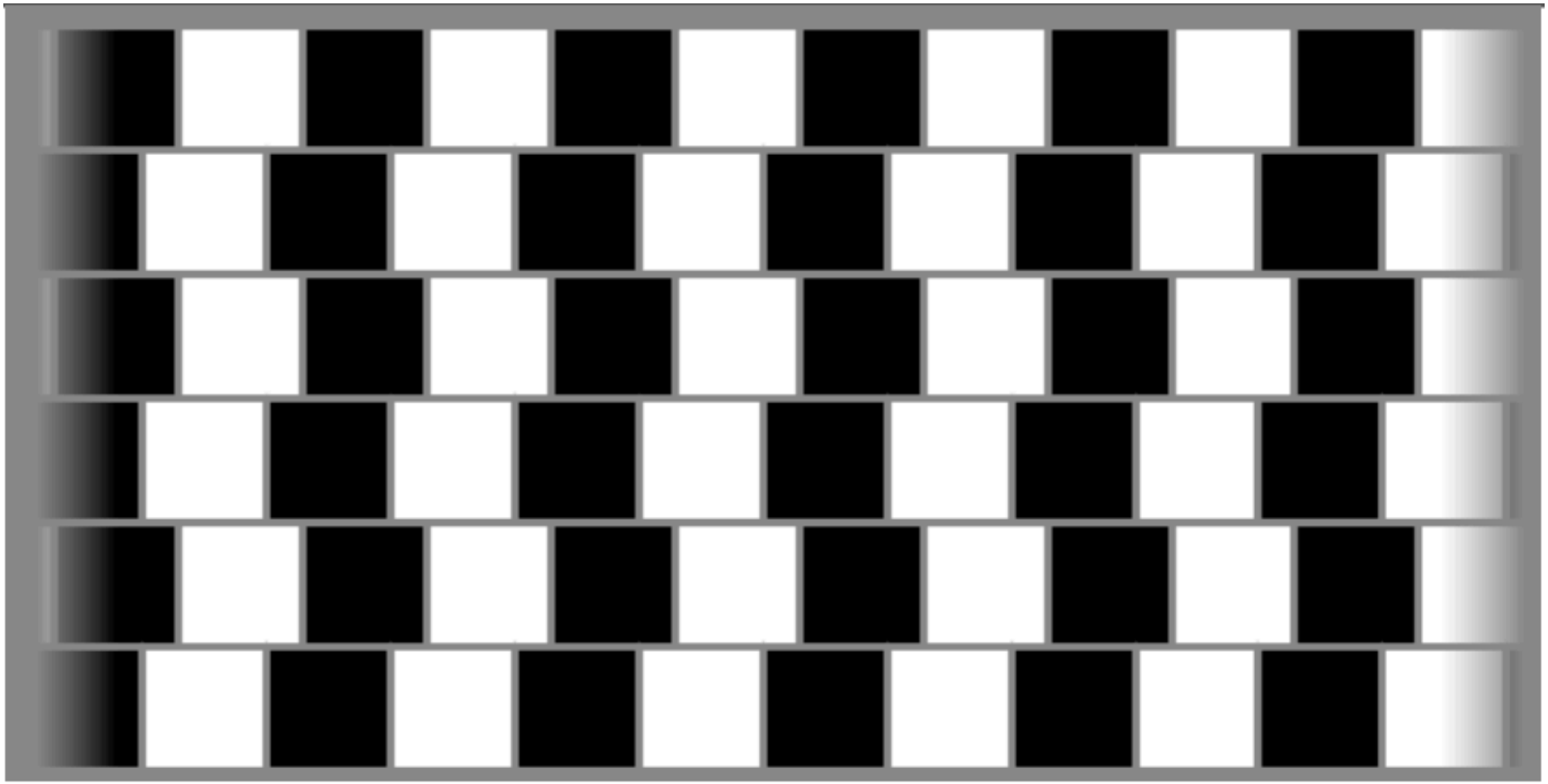
Hermann Grid

Dynamic Luminance

Changes in apparent brightness with quick changes in viewing distance

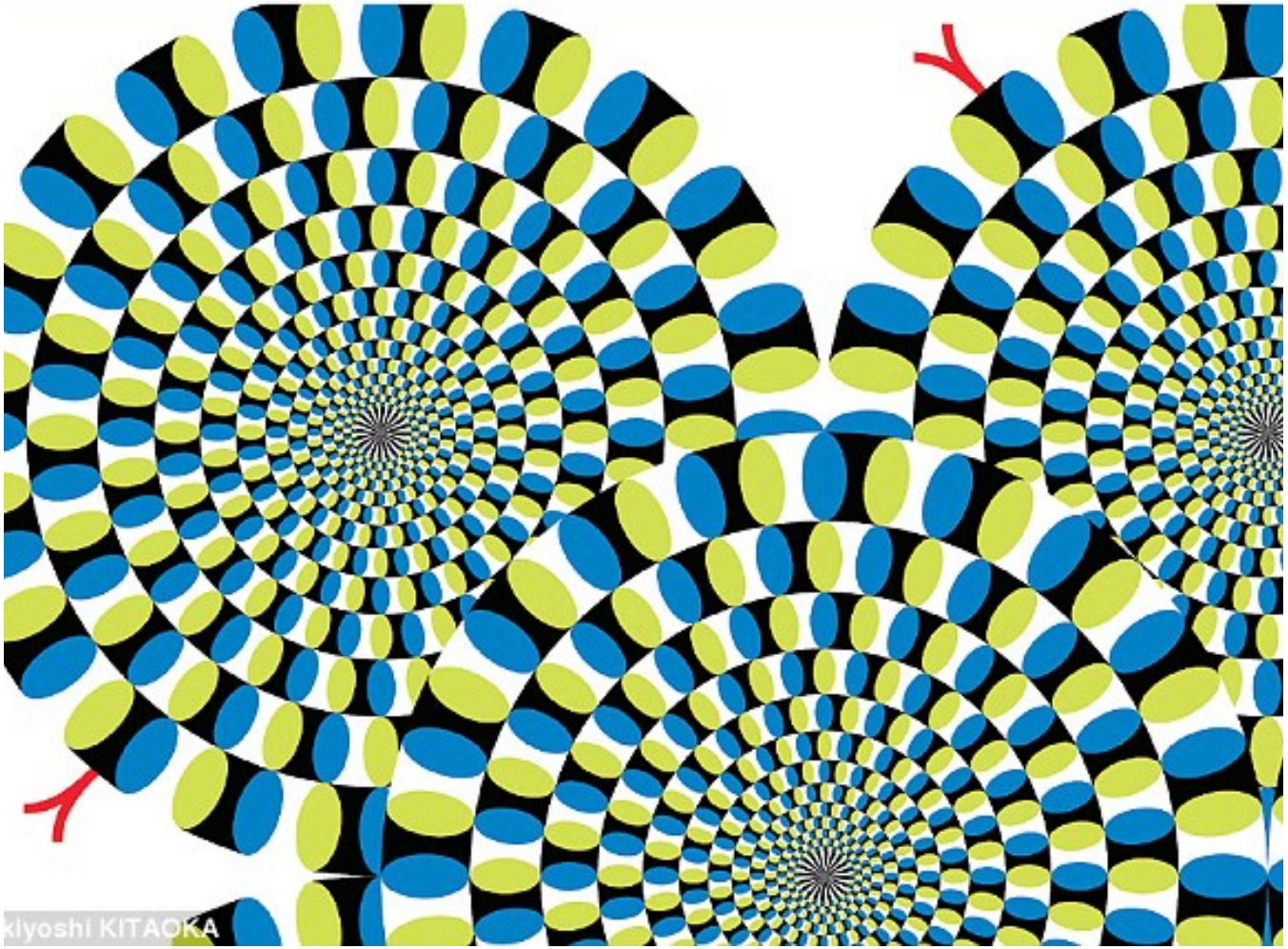


The Café Wall Illusion



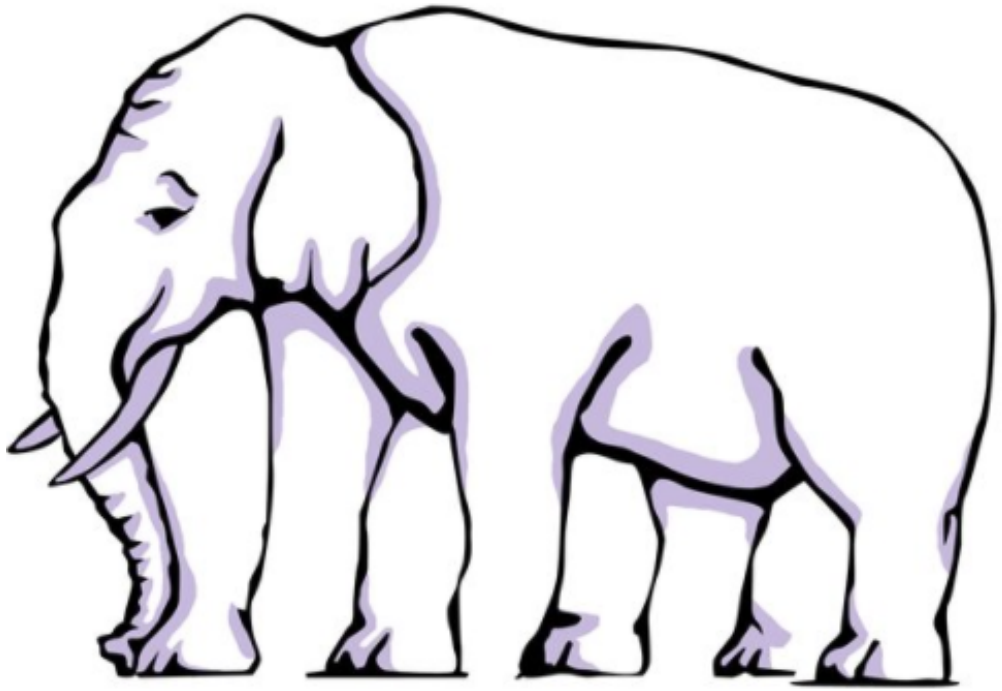
The tiles are actually evenly rectangular

The Snake FX



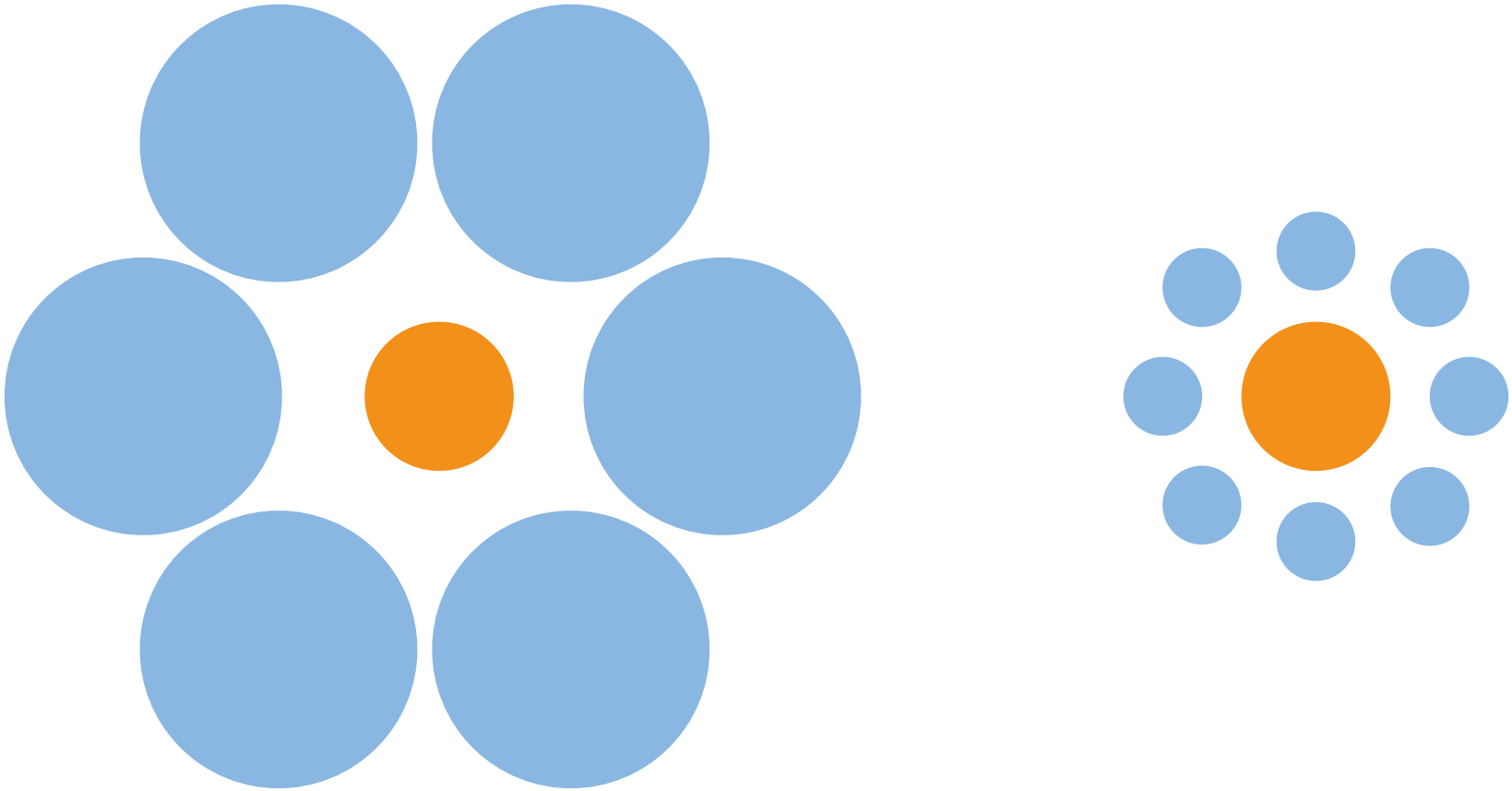
It seems to be moving, but it is not.

Elephant

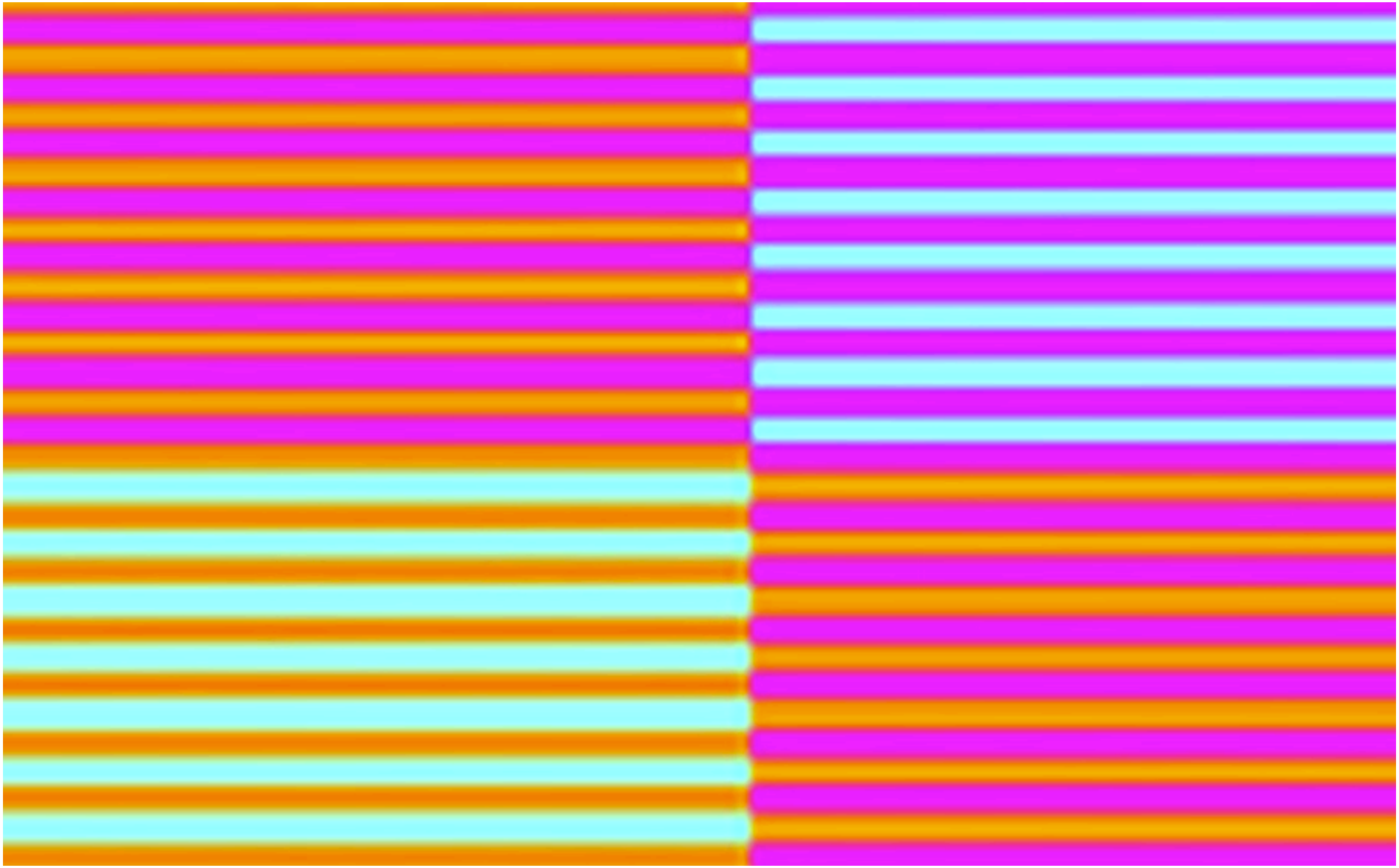


How many legs?

Young woman or old lady?



Which orange circle is bigger?



How many colours do you see?

Preattentive Processing

- How does human visual system analyze images?
- Some things seem to be done preattentively, without the need for focused attention
- Generally less than 200-250 ms (eye movements take 200 ms)
- Seems to be done in parallel by low-level vision system
- A limited set of visual properties are processed preattentively (without need for focusing attention).

How Many 3's?

12817687561389765469845069856049828267629809858458224509856458945098450980943585909103020990595959577256467505067890456788457898098216776548763649085609129496863

How Many 3's?

12817687561389765469845069856049828267629809858458224509856458945098450980943585909103020990595959577256467505067890456788457898098216776548763649085609129496 86

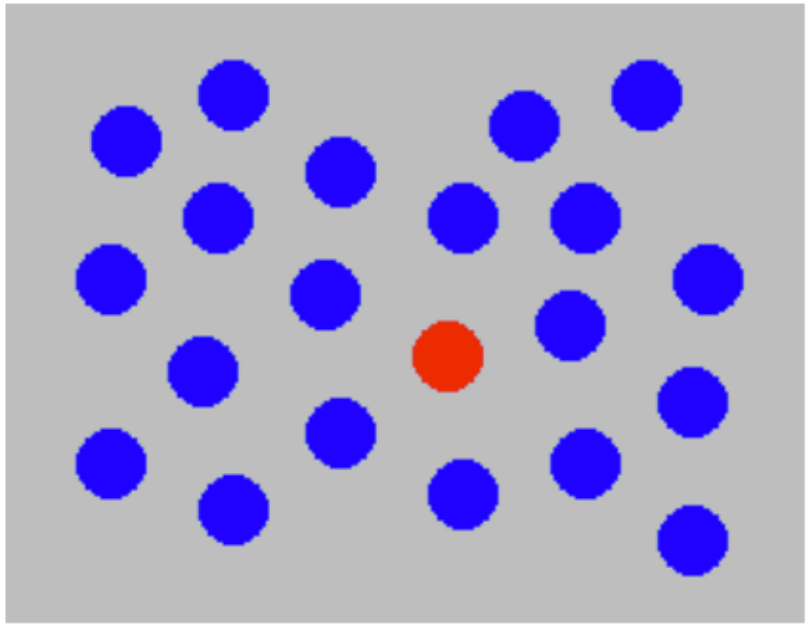
What we can do with this?

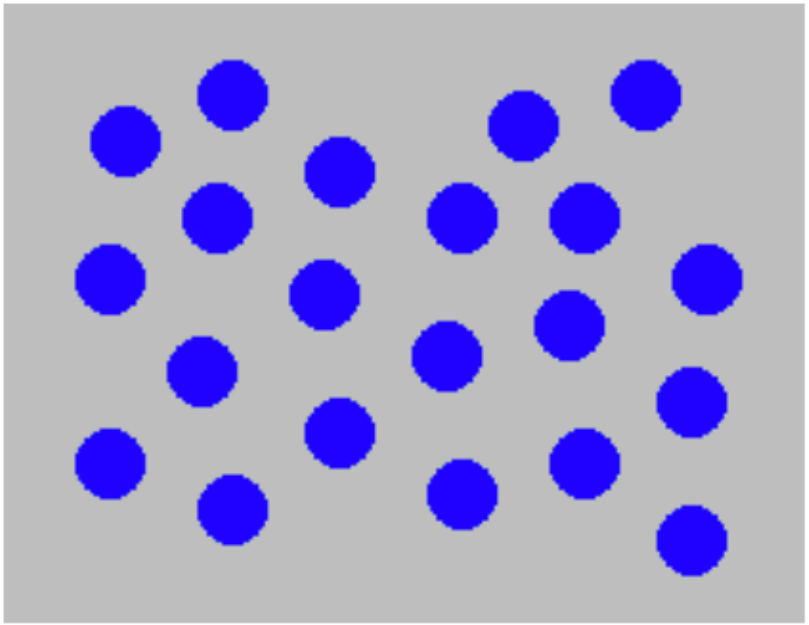
- **Target detection**
  - Is something there?
- **Boundary detection**
  - Can the elements be grouped?
  - What associates them?
- **Counting**
  - How many elements of a certain type are present?

See thing “at a glance”

Colour selection

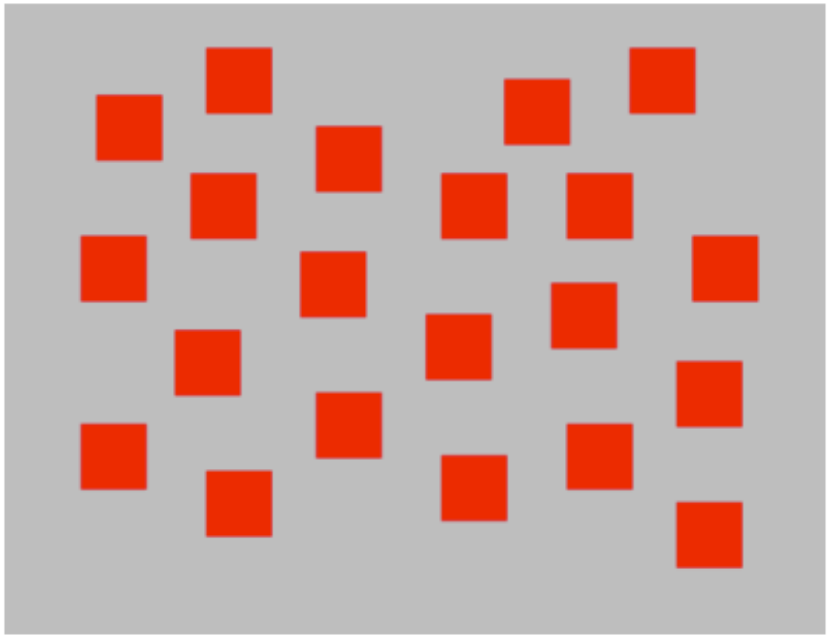
- Where is the red circle? Left or right?
- Put your hand up as soon as you see it.

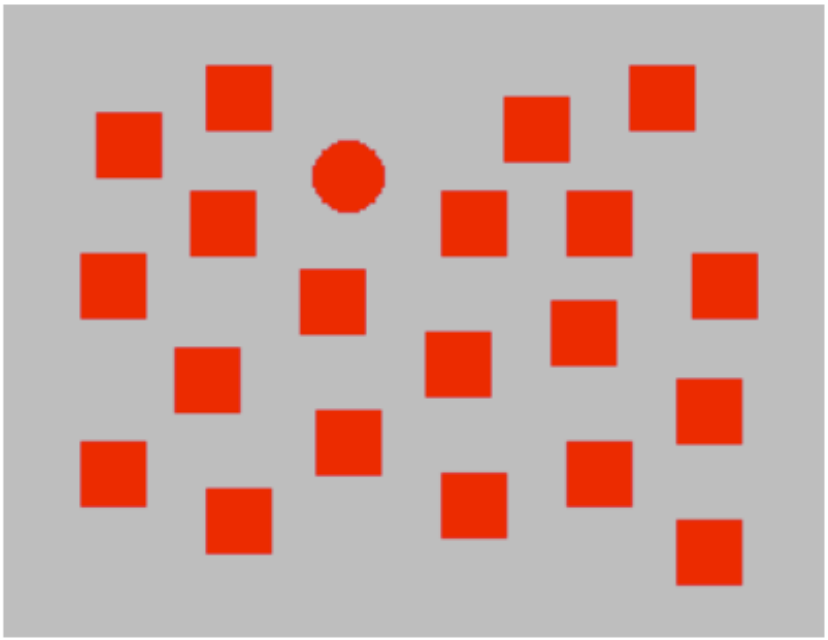




Shape selection

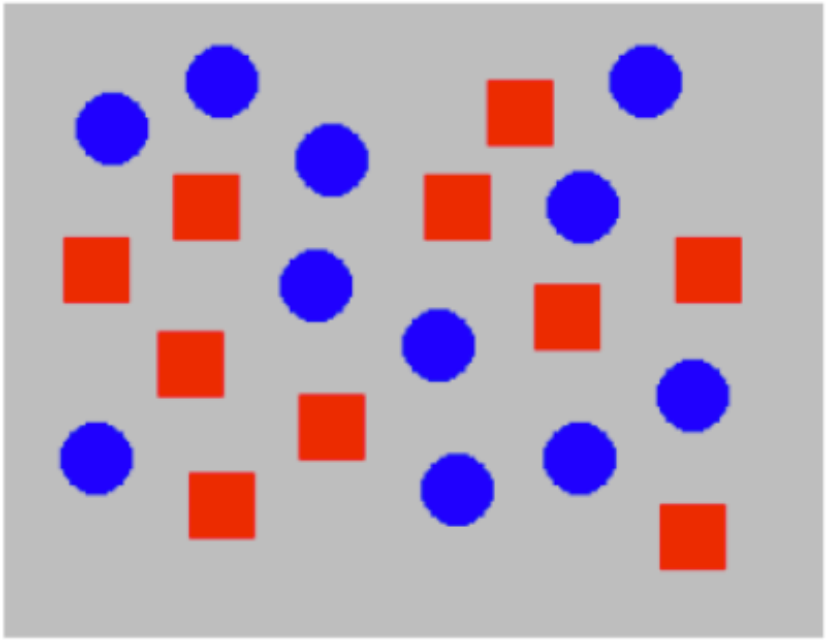
- Where is the red circle? Left or right?
- Put your hand up as soon as you see it.

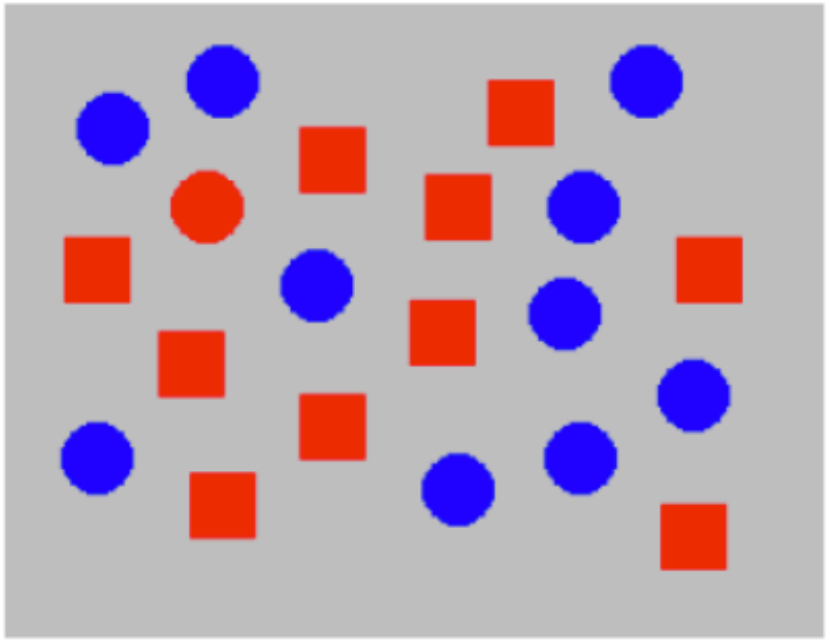




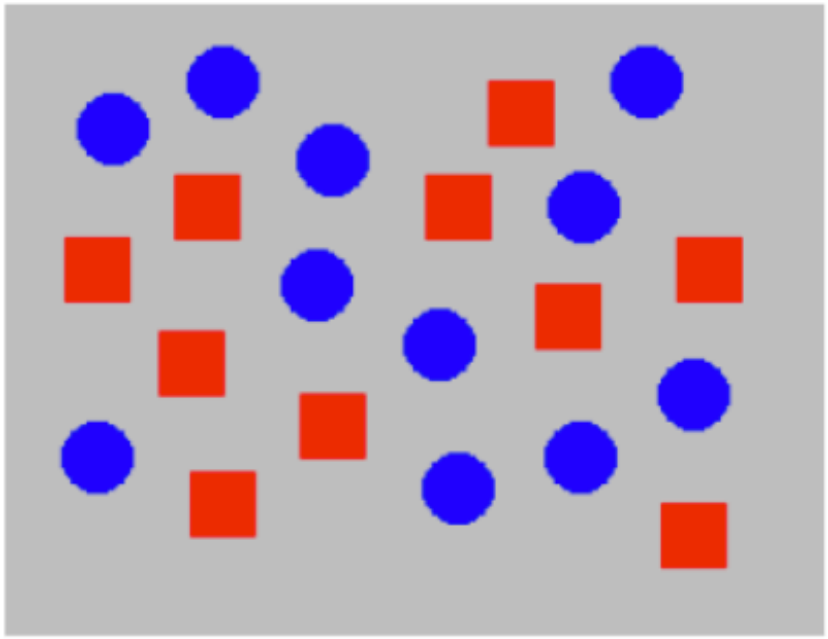
Hue and Shape

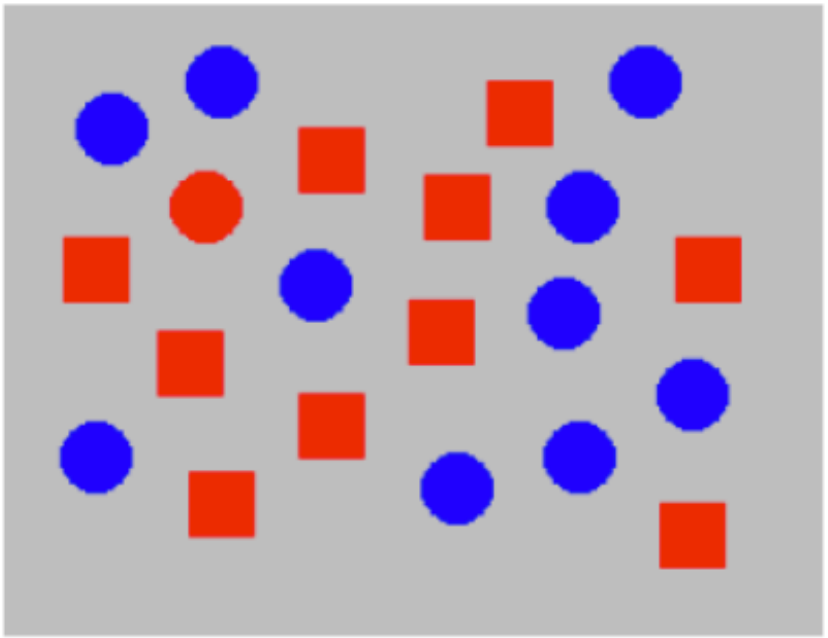
- Where is the red circle? Left or right?
- Put your hand up as soon as you see it.





Hue and Shape

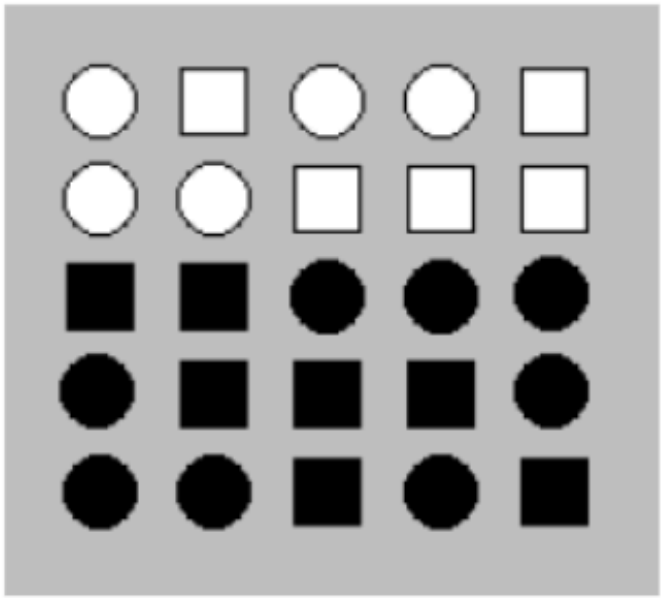


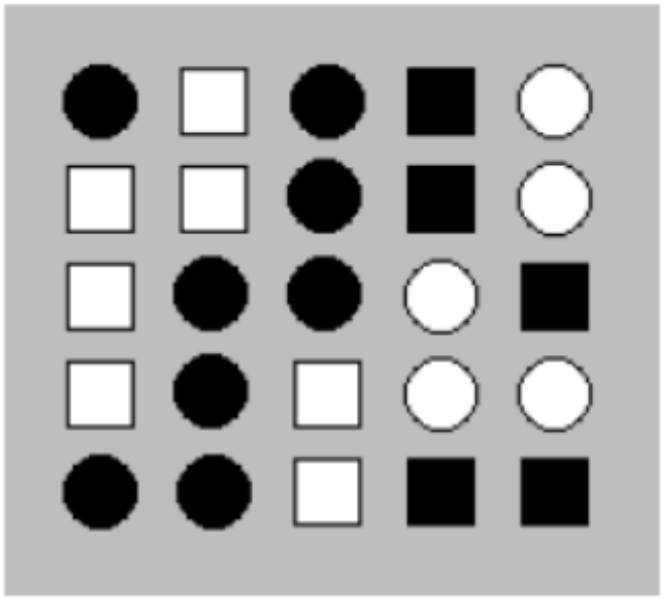


- **Cannot be done preattentively**
- Must perform a sequential search.
- Conjunction of features (shape and hue) causes it.

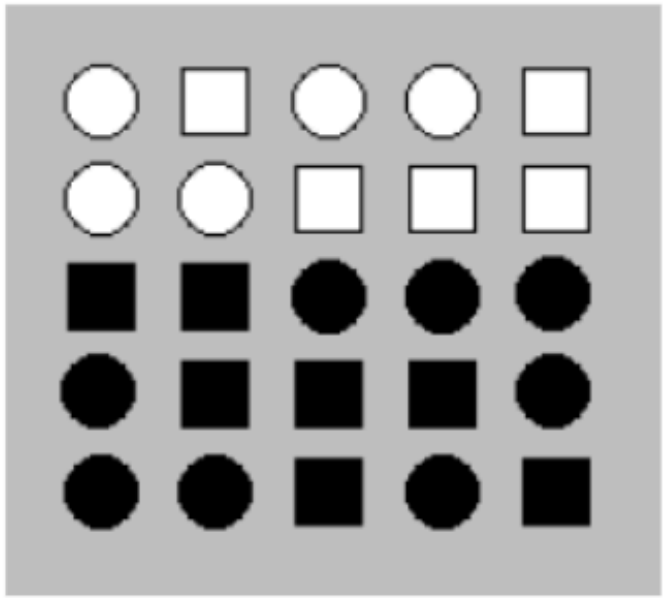
Boundaries matter: Fill and Shape

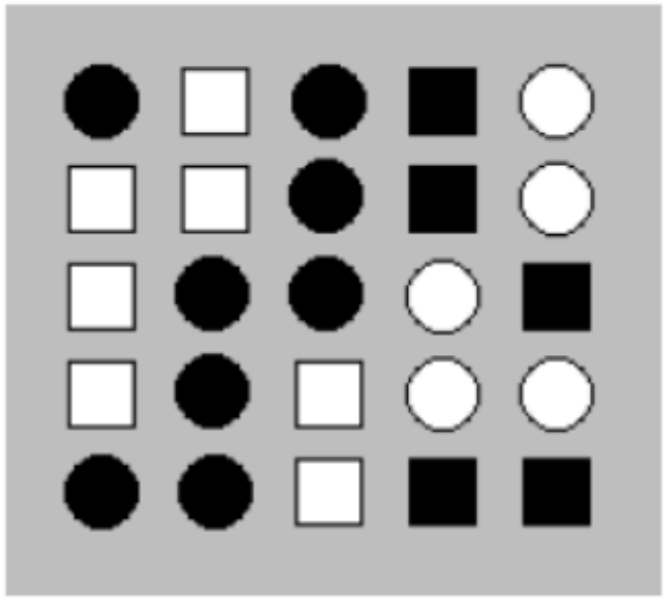
- Where is /are the white circles?
- Put your hand up as soon as you see it.





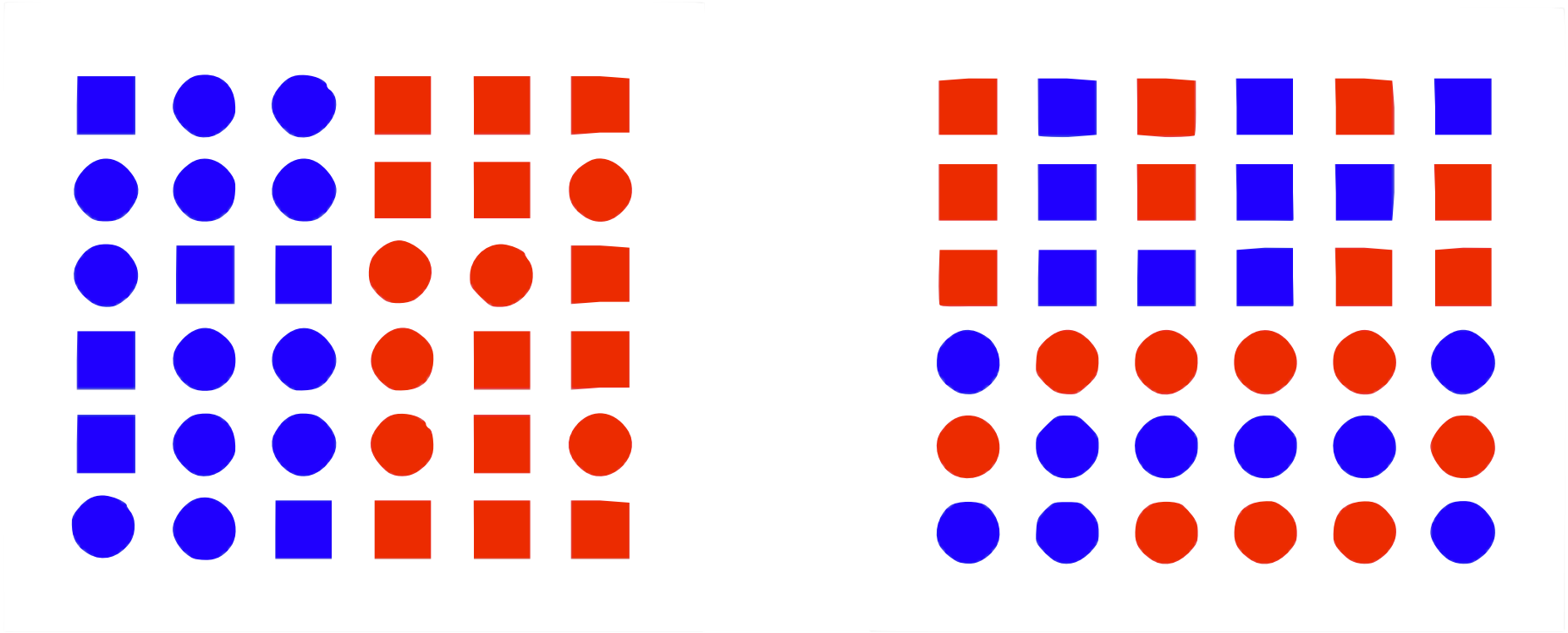
Boundaries matter: Fill and Shape





- Left can be done preattentively since each group contains one unique feature
- Right cannot since the two features are mixed (fill and shape)

Hue versus Shape



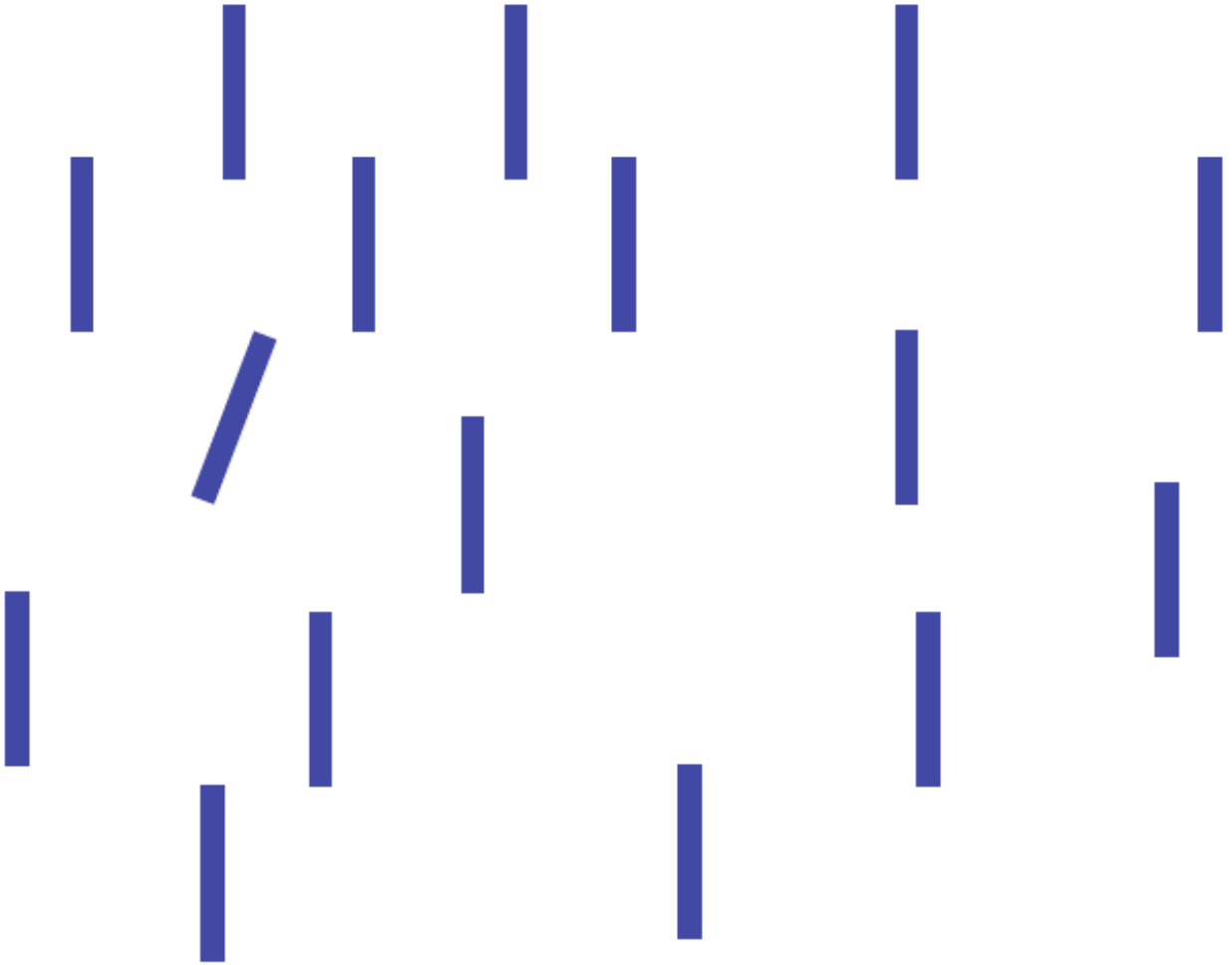
- Left: Boundary detected preattentively based on hue regardless of shape
- Right: Cannot do mixed color shapes preattentively

Try for yourself

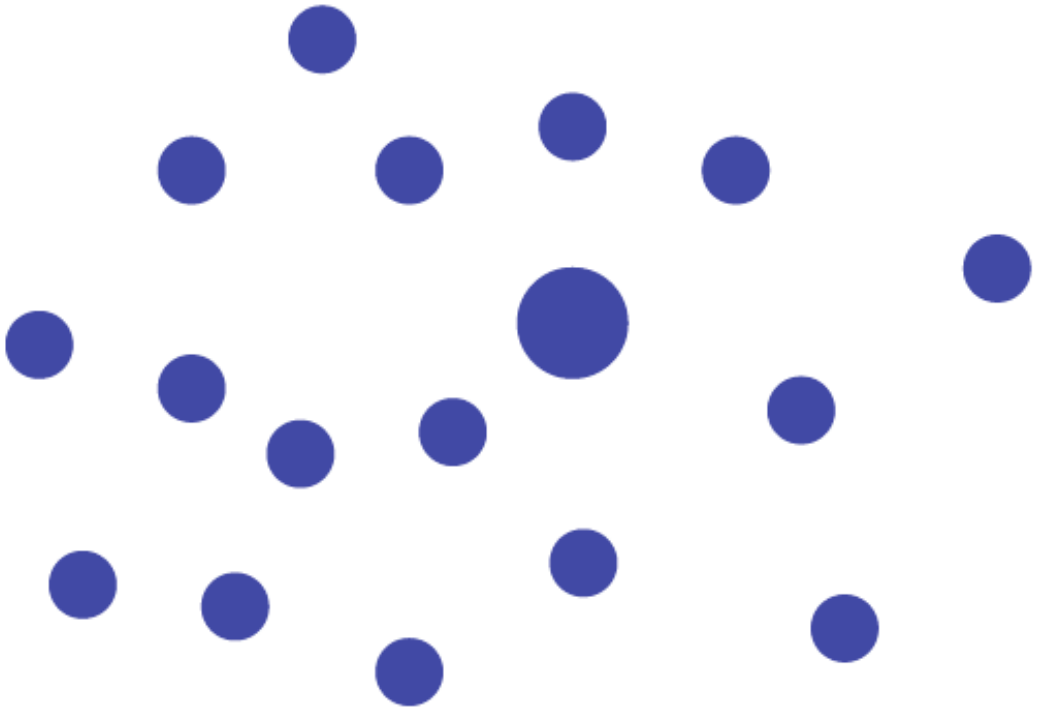
- Healey's applet: https://www.csc2.ncsu.edu/faculty/healey/PP/

Preattentive Features: Orientation

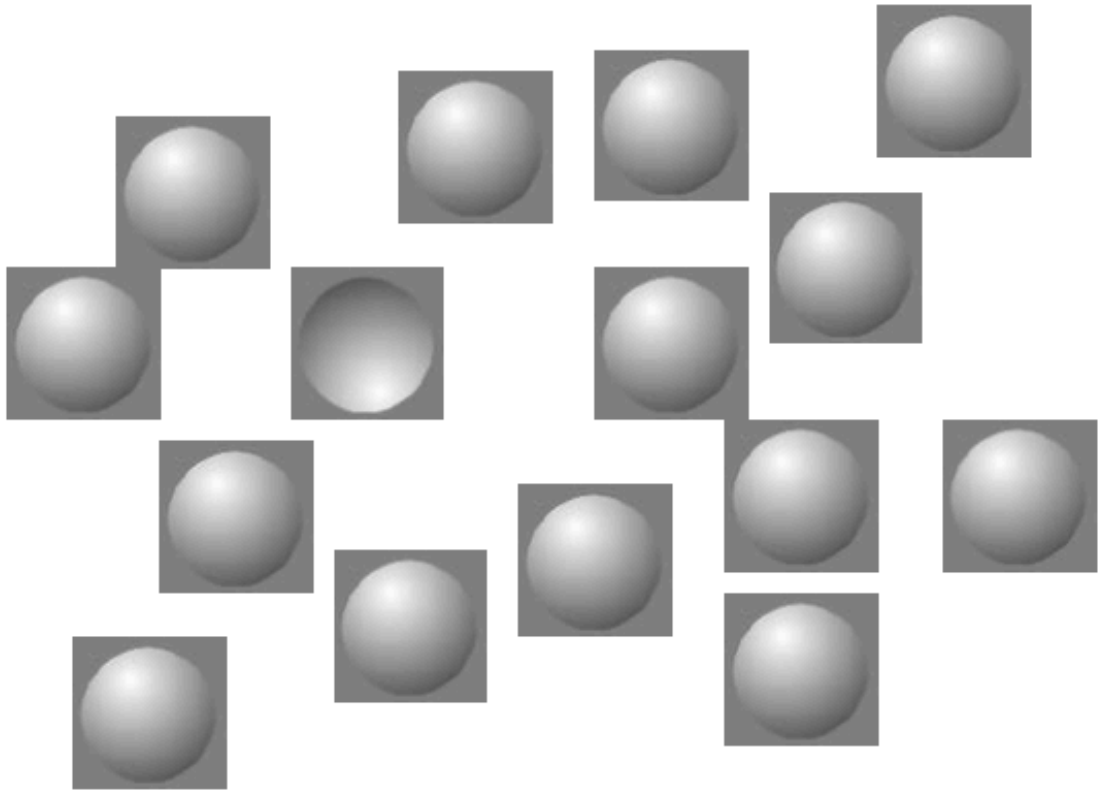
- Certain visual forms lend themselves to preattentive processing. A variety of forms seem to work, like: Orientation



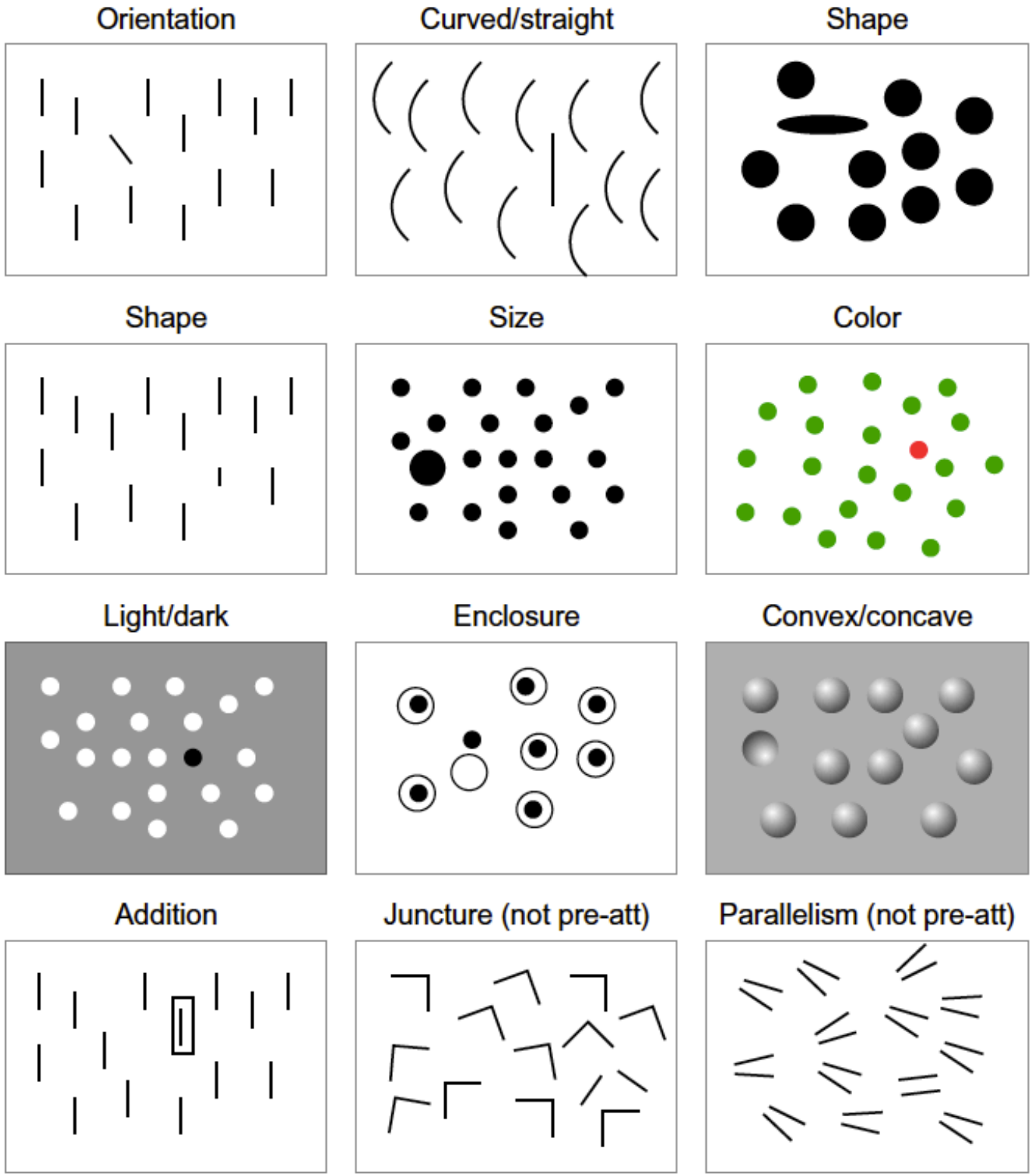
Preattentive Features: Size



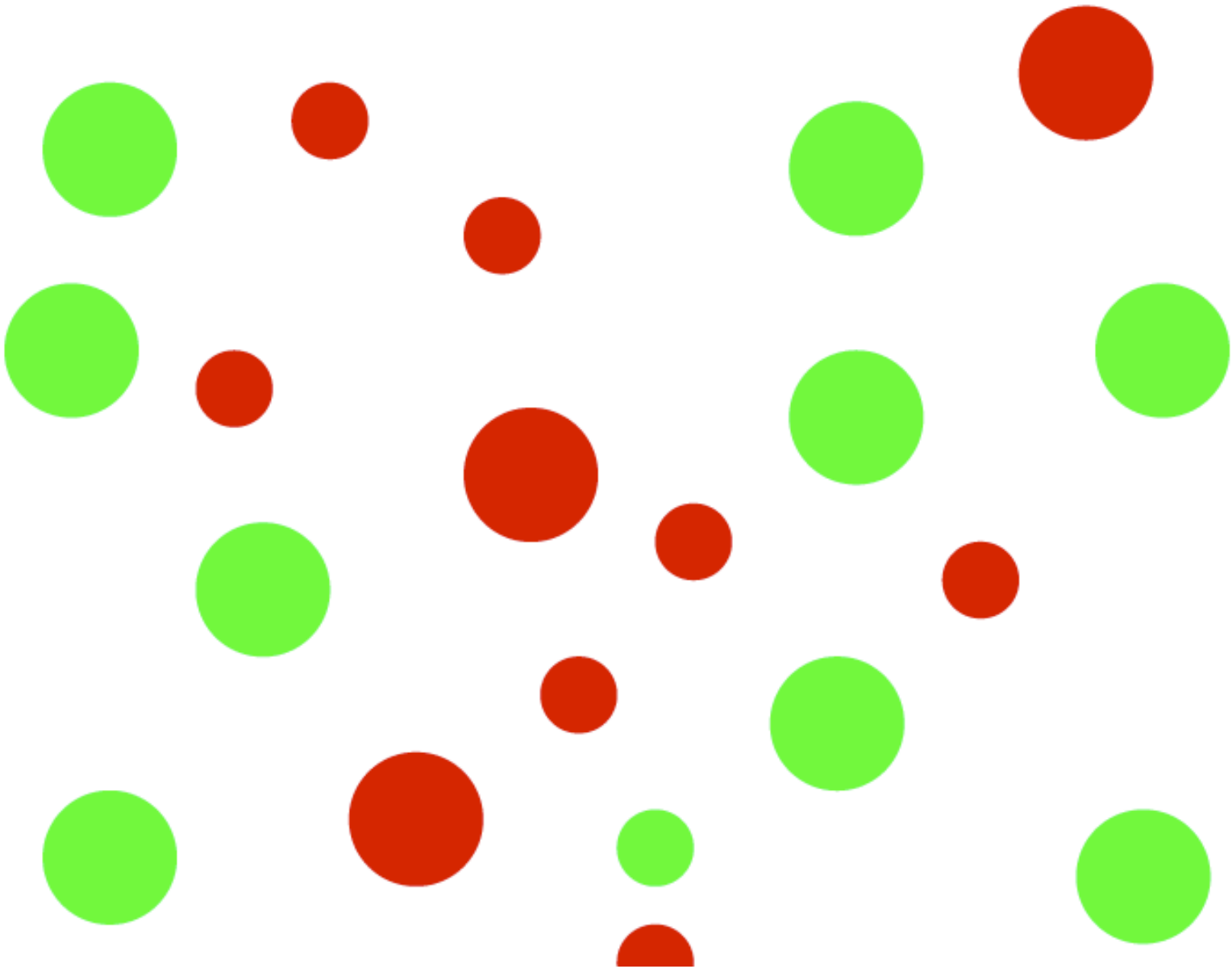
Preattentive Features: Simple Shading



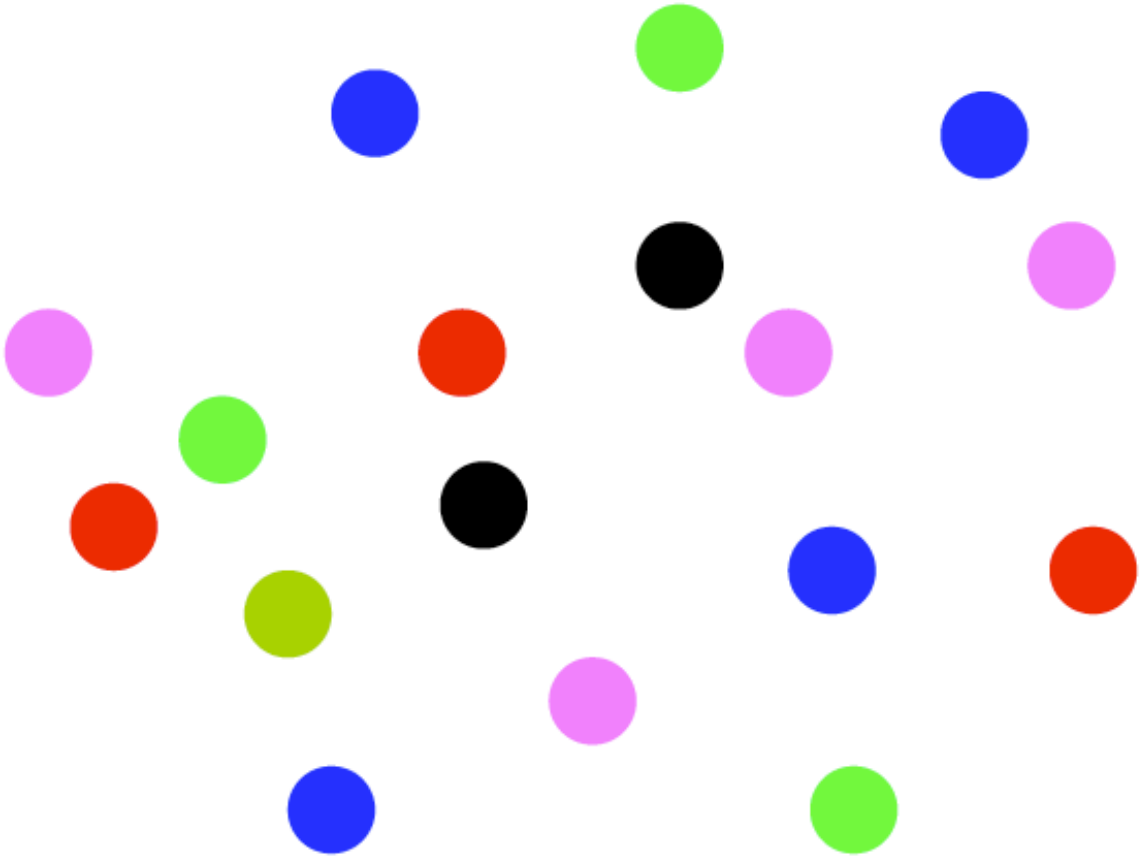
Preattentive Features: Many others



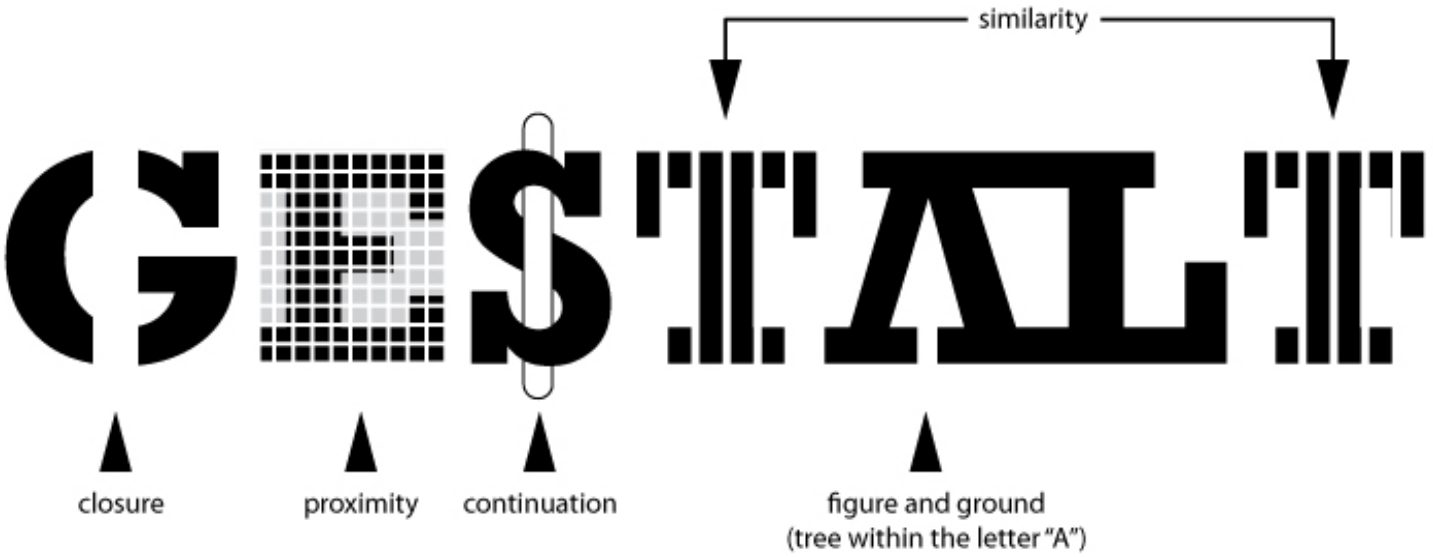
Conjunction does not pop out



Surrounded colours do not pop out

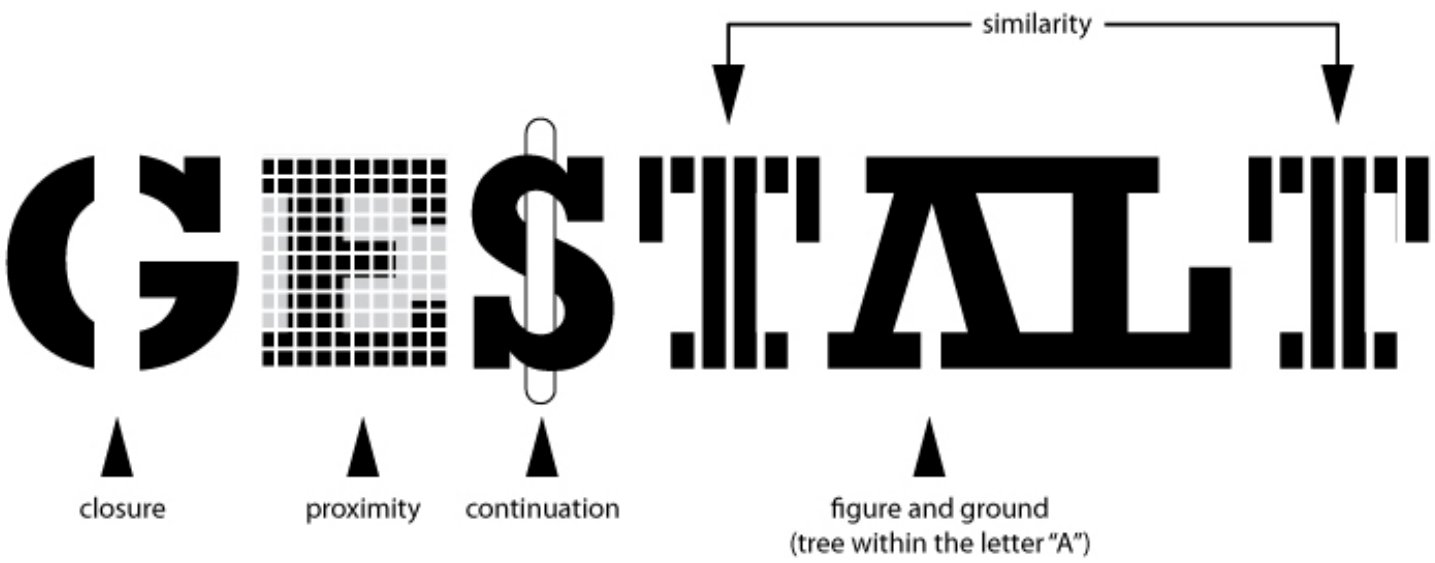


Stage 2:
Intermediate Level:
Patterns recognition



Stage 2: medium Level

- Slow serial processing
- Involves working and long-term memory
- Top-down processing
- Limited resources


similarity
closure
proximity
continuation
figure and ground
(tree within the letter "A")

The Gestalt laws

- 1. Proximity
- 2. Similarity
- 3. Connectedness
- 4. Continuity
- 5. Symmetry
- 6. Closure
- 7. Common fate
- 8. Figure-ground

Proximity

- Things that are close together are perceptually grouped together

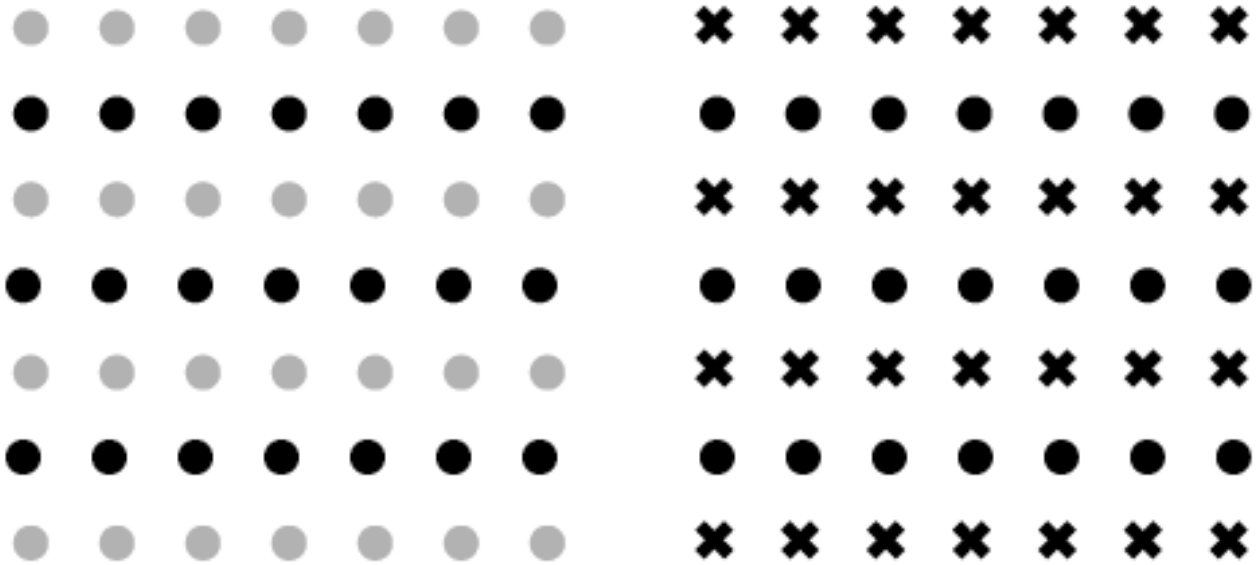
Rows

Columns

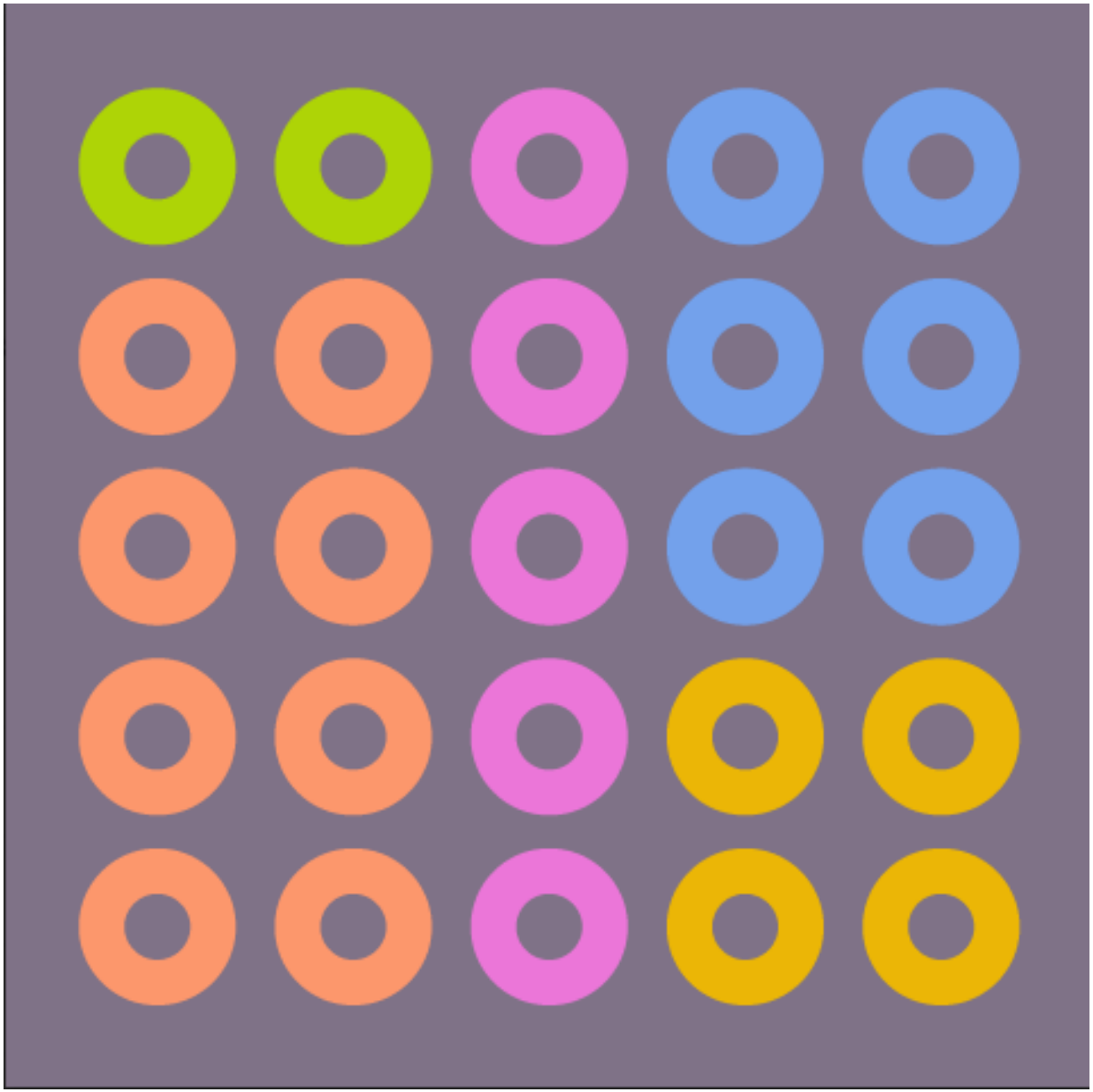
Proximity

Similarity

- The shapes of individual pattern elements can also determine how they are grouped
- Similar elements tend to be grouped together

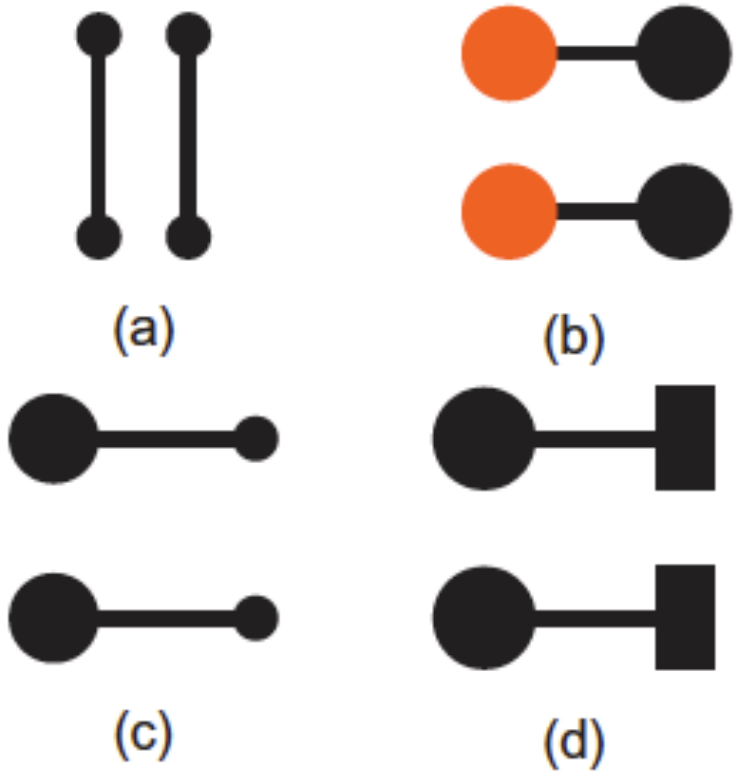


Similarity

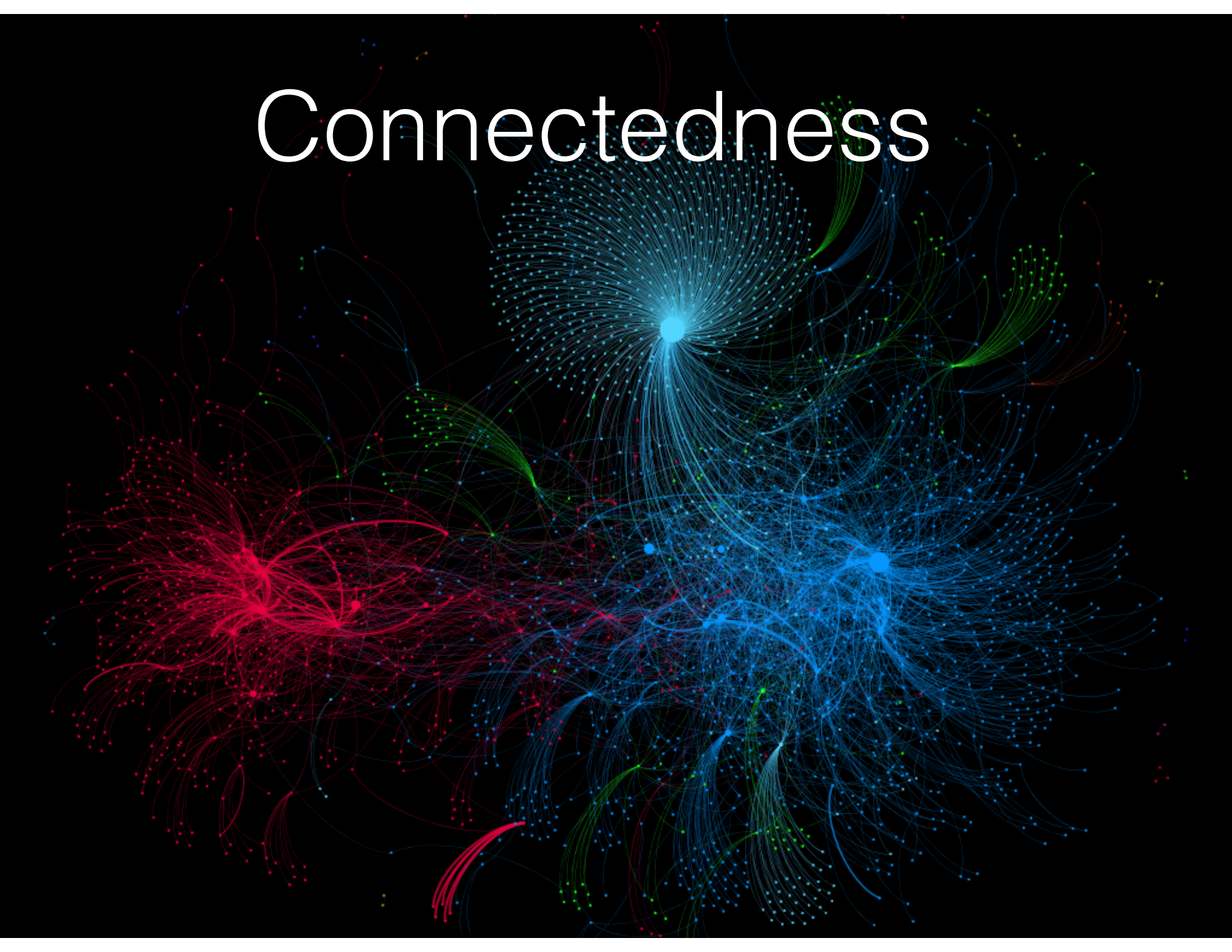


Connectedness

- Connectedness can over-rule proximity, colour, size or shape

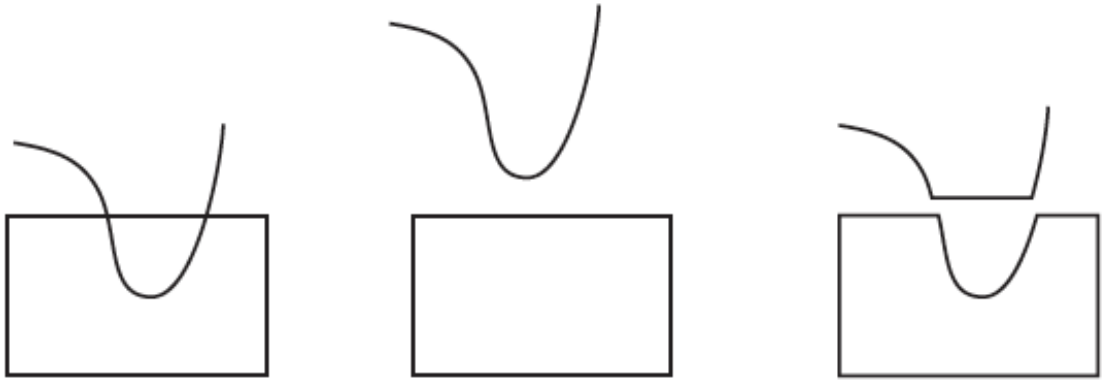


Connectedness



Continuity

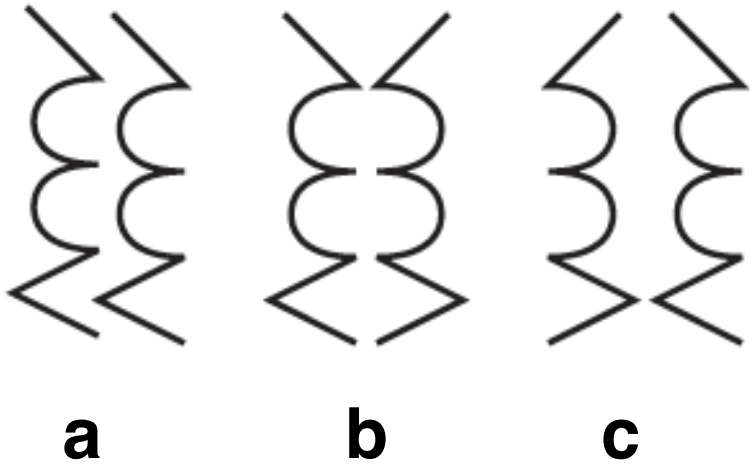
- The Gestalt principle of continuity states that we are more likely to construct visual entities out of objects that are smooth and continuous, rather than those that contain abrupt changes in direction.



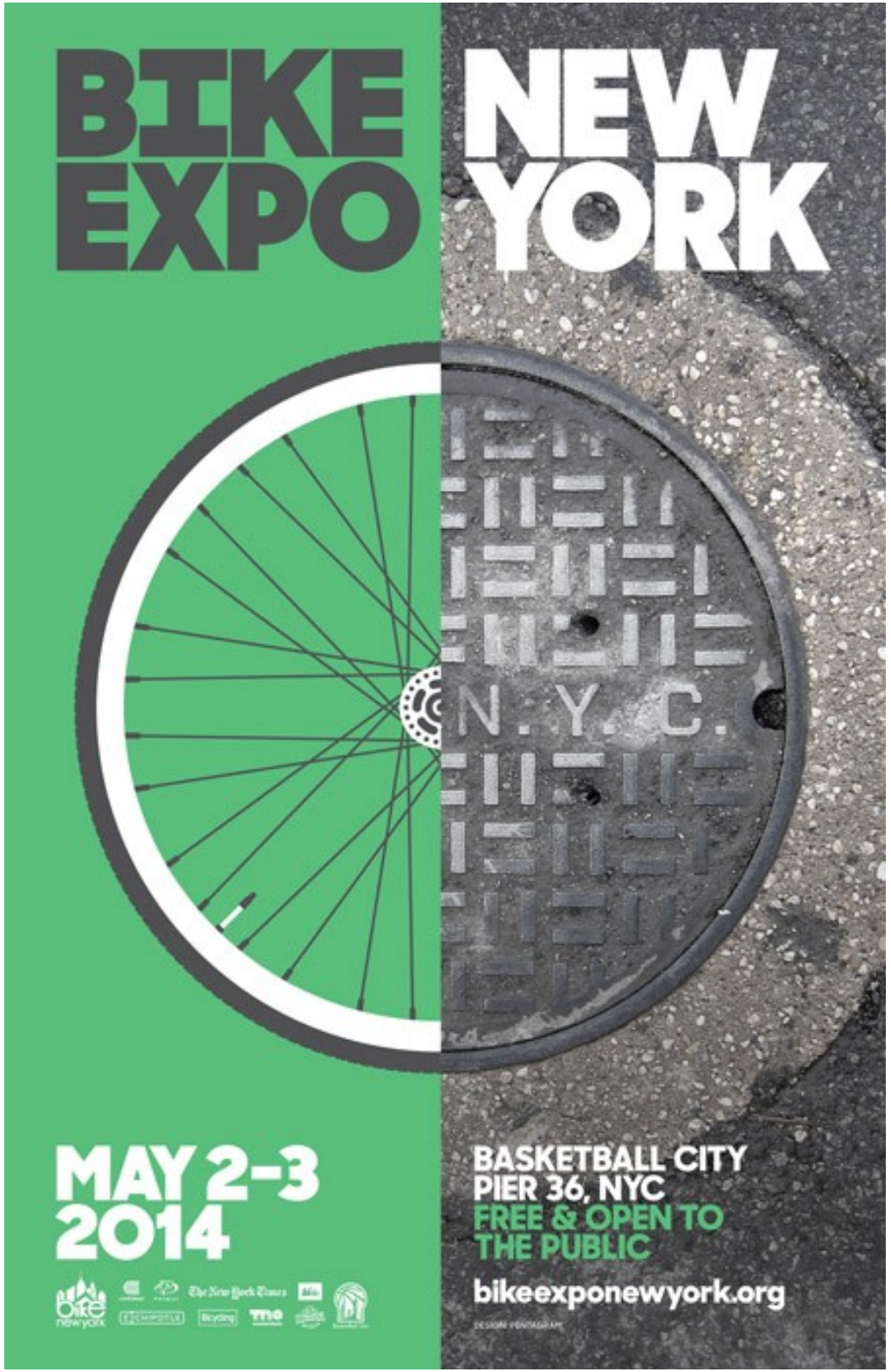
Continuity

Symmetry

- Symmetry creates visual whole
- Powerful organizing principle
- b and c are seen as figures/objects, where a is a pair of parallel lines



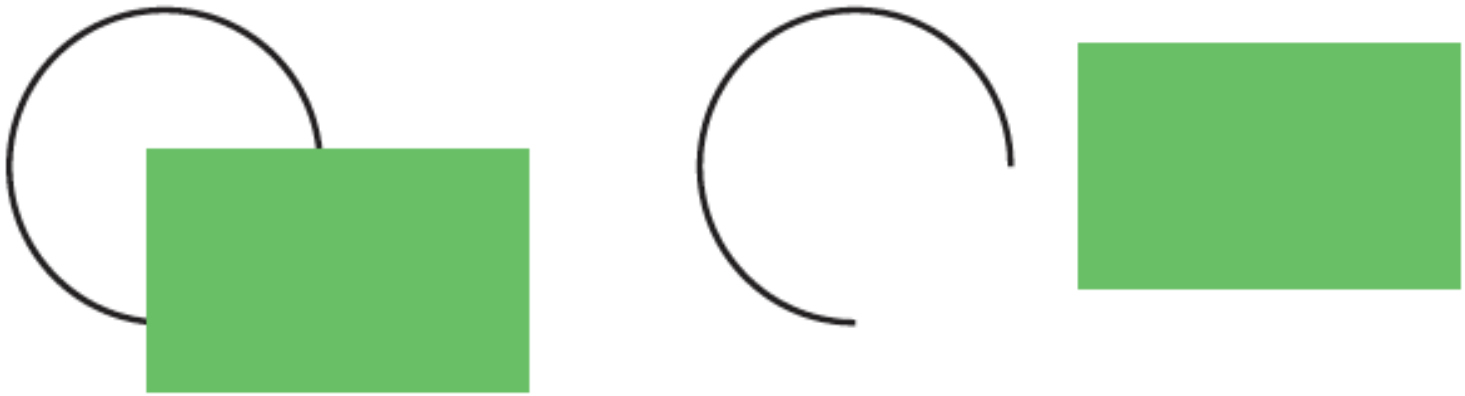
Symmetry


BIKE EXPO
NEW YORK
N.Y.C.
MAY 2-3 2014
BASKETBALL CITY
PIER 36, NYC
FREE & OPEN TO THE PUBLIC
bikeexponewyork.org

Symmetry

Closure

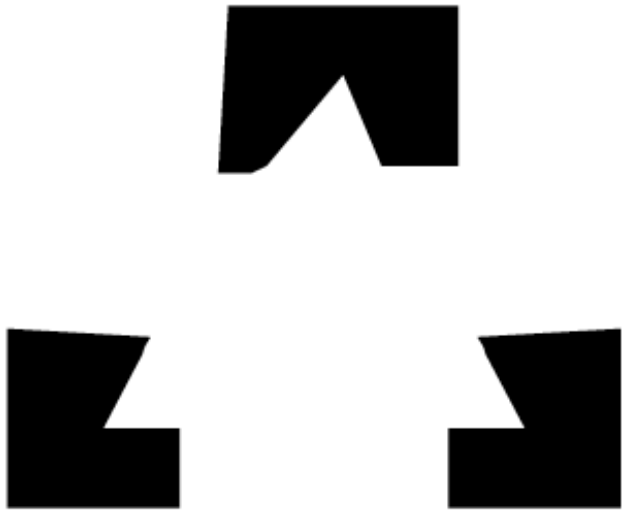
- **Over-rules proximity !**
- A closed contour tends to be seen as an object
- The Gestalt psychologists argued that there is a perceptual tendency to close contours that have gaps



(a) (b)

Closure

- We construct an object from pieces



Closure

Closure

Common fate

- When lines or shapes “move” in the same direction, they are perceived to be in some relationship.

Common fate

Common fate

Figure-ground

- We separate a dominant shape (a 'figure' with a definite contour) from what our current concerns relegate to 'background' (or 'ground')
- Symmetry, white space, and closed contour contribute to perception of figure.
- The perception of figure as opposed to ground can be thought of as the fundamental perceptual act of identifying objects.

Figure-ground

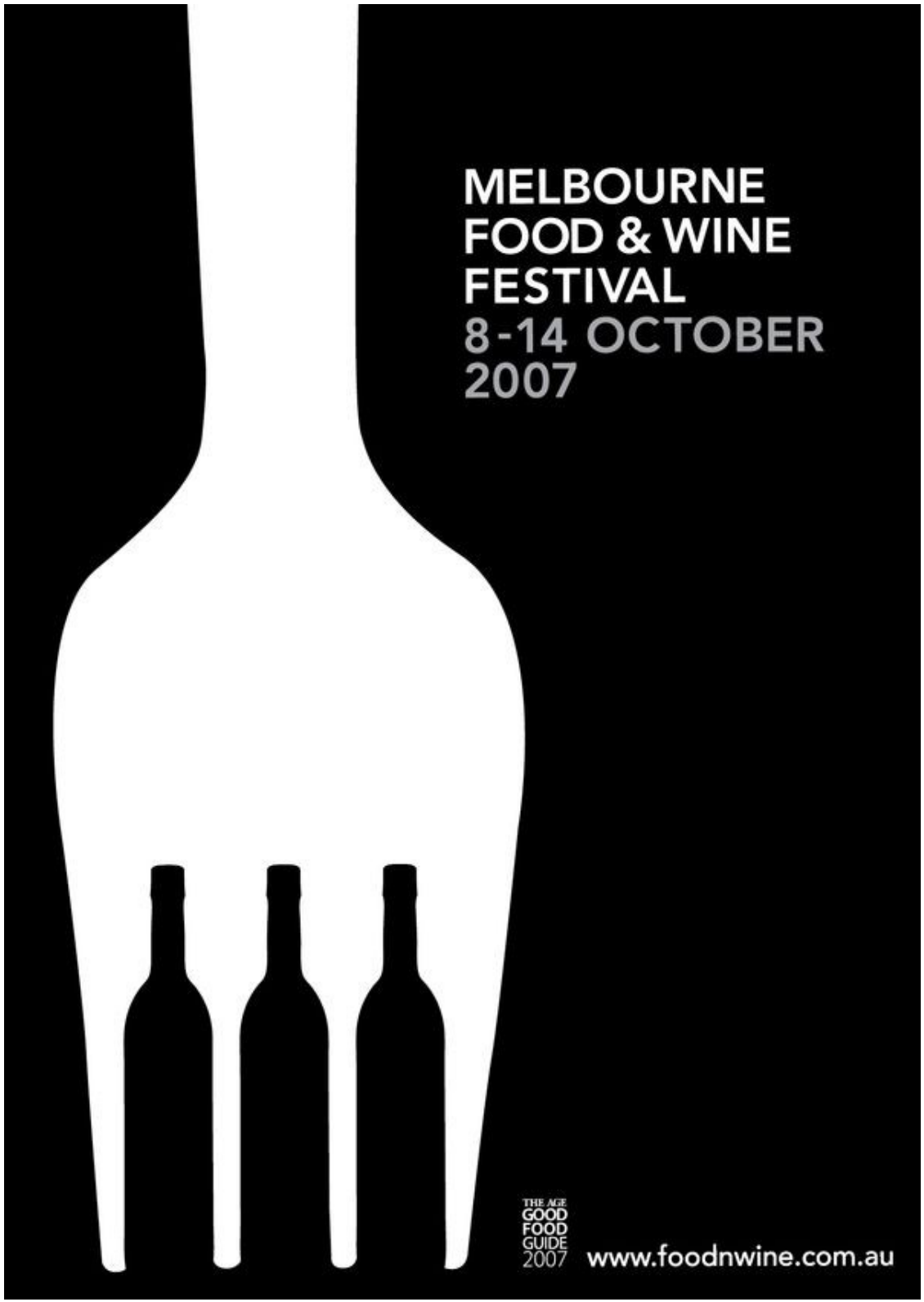
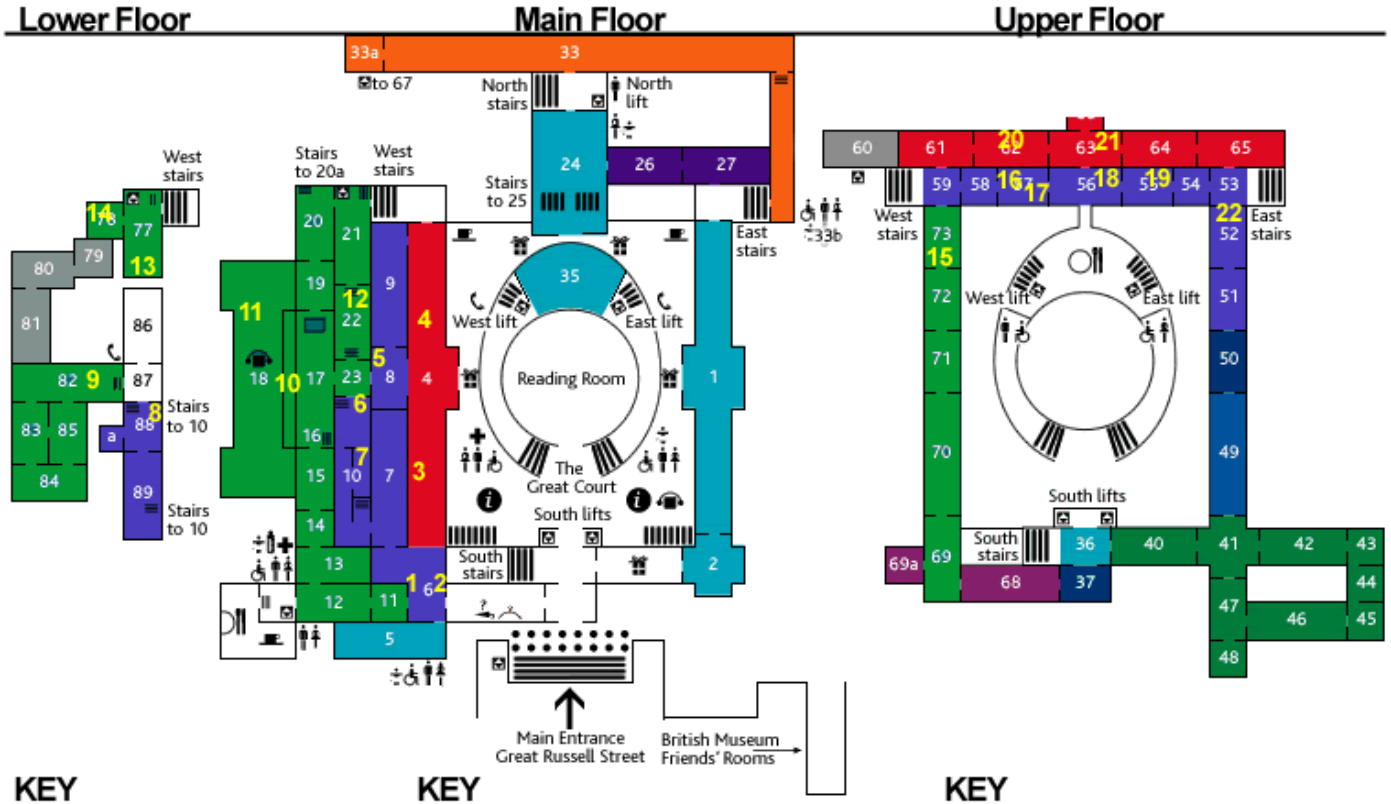


Figure-ground

Stage 3: “High” Level: Finding information



Finding Information

- Visual search is one of the basic things we do everyday.
- There is a kind of mental inner scan, within a fixation, where a few visual patterns are tested for query-resolving properties

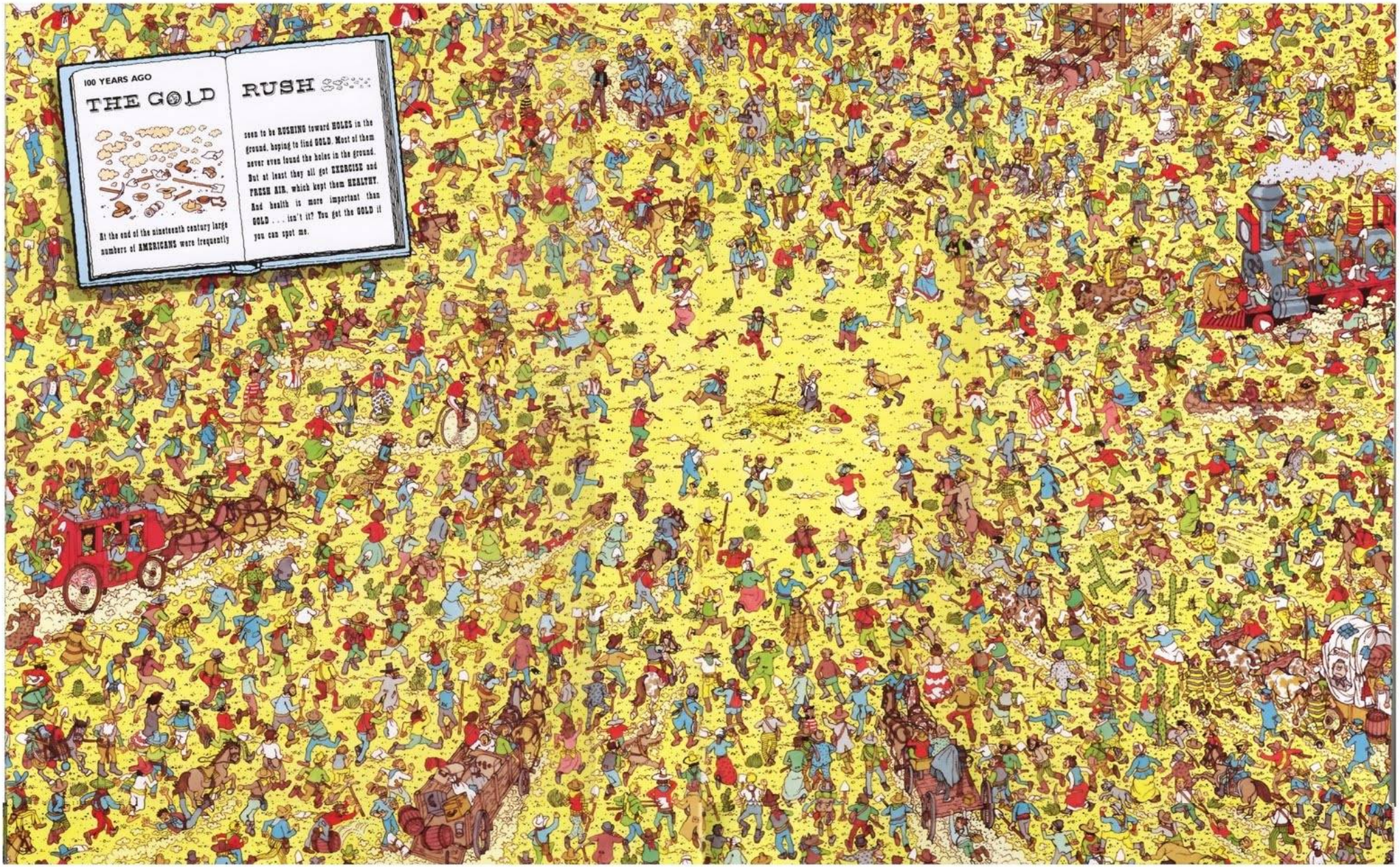
Finding Information

- Involves reformulating part of the problem so that the solution can be found through a visual pattern search.
- Understanding what makes a pattern easy to find is critical in determining how efficiently the query will be executed.
- In understanding how visual queries are resolved we gain a deeper understanding of how best to visually communicate information.

Finding Information

- Three things determine what is easily findable
- 1. A priori salience: Some patterns excite more neural activity in the feature maps than others.
- 2. Top-down salience modification. Depending on what we are looking for, topdown mechanisms retune the feature maps to increase their sensitivity to certain features.
- 3. Scene gist. If a type of scene is well known, then the eye movement strategies will be automatically primed for activation.

Finding Information



Eye Movements

- Moving our eyes causes different parts of the visual environment to be imaged on the high-resolution fovea where we can see detail.
- E.g.: When you read a book, your eye makes between 2 and 5 jerky movements per second (saccades).
- Each of these movements can be thought of as a basic act of visual search.

Saccadic movements

- The eye moves rapidly from fixation to fixation.
- Stationary for: 200 ~ 400 ms
- Saccade takes between 20 and 180 ms and depends on the angle moved.
- Add an extra half a second or more if you need to turn you head.

Saccadic movements

- Saccadic eye movements are said to be ballistic.
- once the brain decides to switch attention and make an eye movement, the muscle signals for accelerating and decelerating the eye are preprogrammed.
- The movement cannot be adjusted in mid saccade
- During a saccadic eye movement, we are less sensitive to visual input

Saccadic eye movements

- Another implication of saccadic suppression is that the brain is usually processing a rapid sequence of discrete images.
- increasingly being exploited in television advertising, in which more than one cut per second of video has become commonplace.

Smooth-pursuit movements

- The eye has the ability to lock onto and object and track it.
- This ability also enables us to make head and body movements while maintaining fixation on an object of interest.
- E.g: A Tennis match;
- A car passing on the street

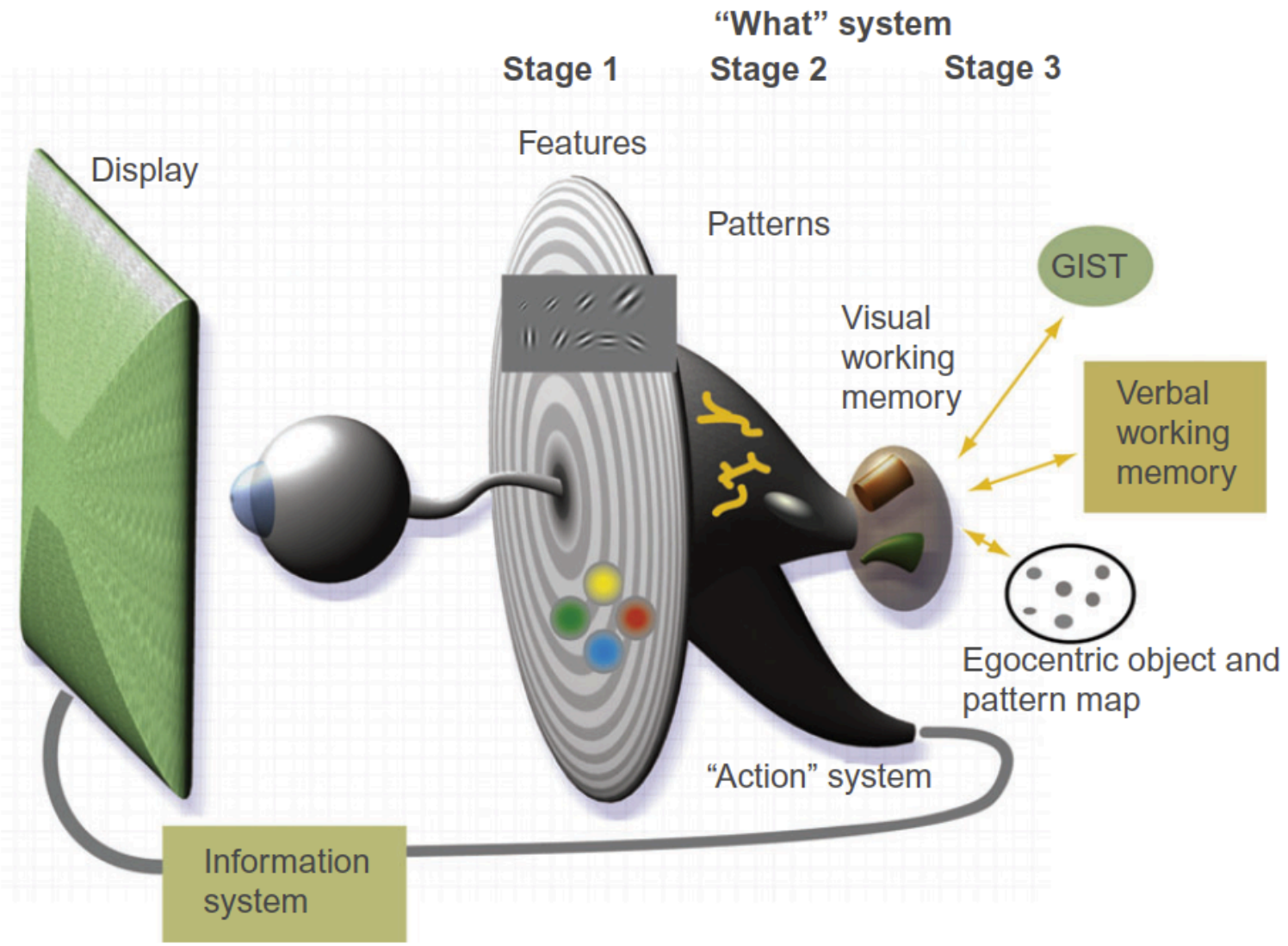
Convergent movements

- When an object moves toward us, our eyes converge.
- When it moves away, they diverge.
- Convergent movements can be either saccadic or smooth.
- E.g.: When something is throw against you.

Accommodation

- When the eye moves to a new target at a different distance, it must refocus, or accommodate, so that the target is clearly imaged on the retina.
- An accommodation response typically takes about 200 ms
- As we get old, however, the ability to accommodate declines and refocusing the eyes must be accomplished by the use eyeglasses.

Recap


"What" system
Stage 1
Stage 2
Stage 3
Features
Patterns
Display
GIST
Visual working memory
Verbal working memory
Egocentric object and pattern map
"Action" system
Information system


Attention

Visual attention

"Everyone knows what attention is. It is the taking possession by the mind in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought...It implies withdrawal from some things in order to deal effectively with others, and is a condition which has a real opposite in the confused, dazed, scatterbrained state."

William James (1890, p. 403)

Visual attention

- Attention is the cognitive process of selectively concentrating on one aspect of the environment while ignoring other things.
    - Comprises a range of cognitive and perceptual processes
- Interactive and iterative process between Top-down and bottom-up:
    - Top down :“cognitive”,voluntary
    - Bottom-up :“perceptual”, involuntary

A Game Challenge

Two teams: white and black shirts

- Each team passes a ball
- Count how many times the white-shirt team passes
- You need to be very focused to get it right.

Selective attention test:
https://www.youtube.com/watch?v=vJG698U2Mvo&list=PLB228A1652CD49370&index=1

Inattentional blindness

Phenomena of inability to perceive features in a visual scene if they are not being attended to.

50% of the participants don't realize that something is changed in the scene if they are ready focus on one thing.

- Are there only some kinds of things we see when we are not attending?

- What is the relationship between attention and perception?

- How much, if anything, of our (visual) world do we perceive when we are not attending to it?

Inattentional blindness

What are the practical implications of this for everyday life?

The door study:
https://www.youtube.com/watch?list=PLC0A3CAC7B3A0E288&v=FWSxSQsspiQ

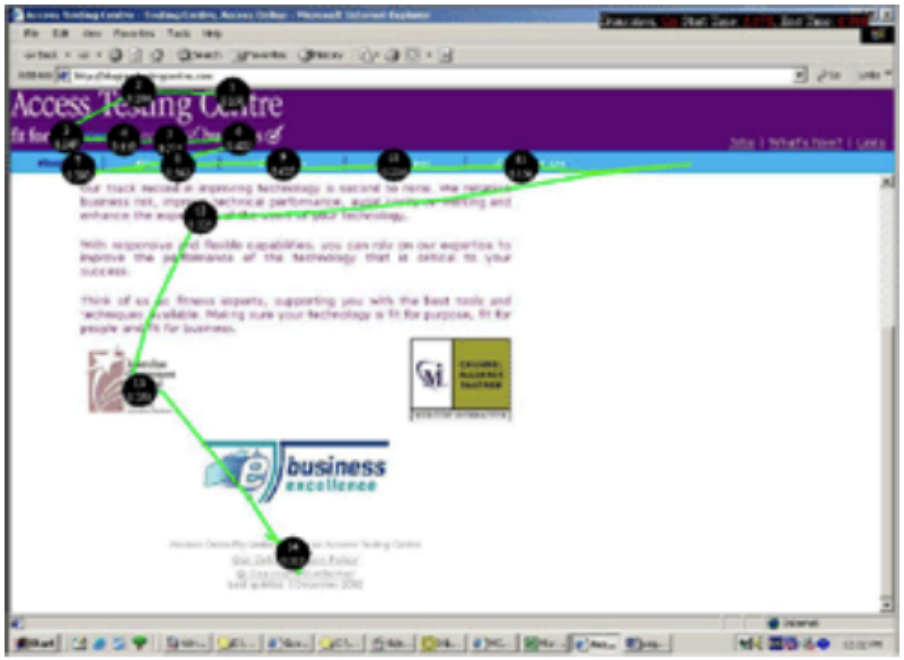
And just in case you are not convinced ..

The Monkey Business Illusion:
https://www.youtube.com/watch?v=IGQmdoK_ZfY&list=PLB228A1652CD49370&index=2

Main models in attention research

- **Selective attention (focused):**
  - Whether we become aware of sensory information
  - Non-random selection
- **Divided Attention (multitasking)**
  - Attention can be split between multiple tasks
  - Allocation approach
  - Is what some of you are doing right now
- **Control and Automaticity**

Selective Attention



- "Top-down, consciously driven "spotlight" set: focus of attention

Divided attention: “multitasking”

- "Top-down, consciously driven “spotlight” set: focus of attention
  - "How many foci can we maintain?
  - What kinds of tasks demand more resources than others?
- "Bottom-up, stimulus driven “demand” events
  - "Involuntary response to perceptual cue
  - "Flashing light or alarm bell
- "How much can we attend to? No established capacity

Automatic and controlled processing

Automatic processing (automaticity):

- Highly familiar and learned tasks
- Does not require conscious attention
- Occur without intention
- Not available for conscious inspection
- Well practiced responses
- Unaffected by capacity
- Fast
- Difficult to modify (Behaviour)
- E.g.: Driving a car and listening to the radio
    - Reading and (not) listening to your partner

Controlled attention:

- Requires conscious attention
- Takes resources
- Limited capacity
- Not well practiced
- Slow

E.g:

- Driving on the other side of the road
- Reading unfamiliar/rare words
- Listening to Luciano's lecture

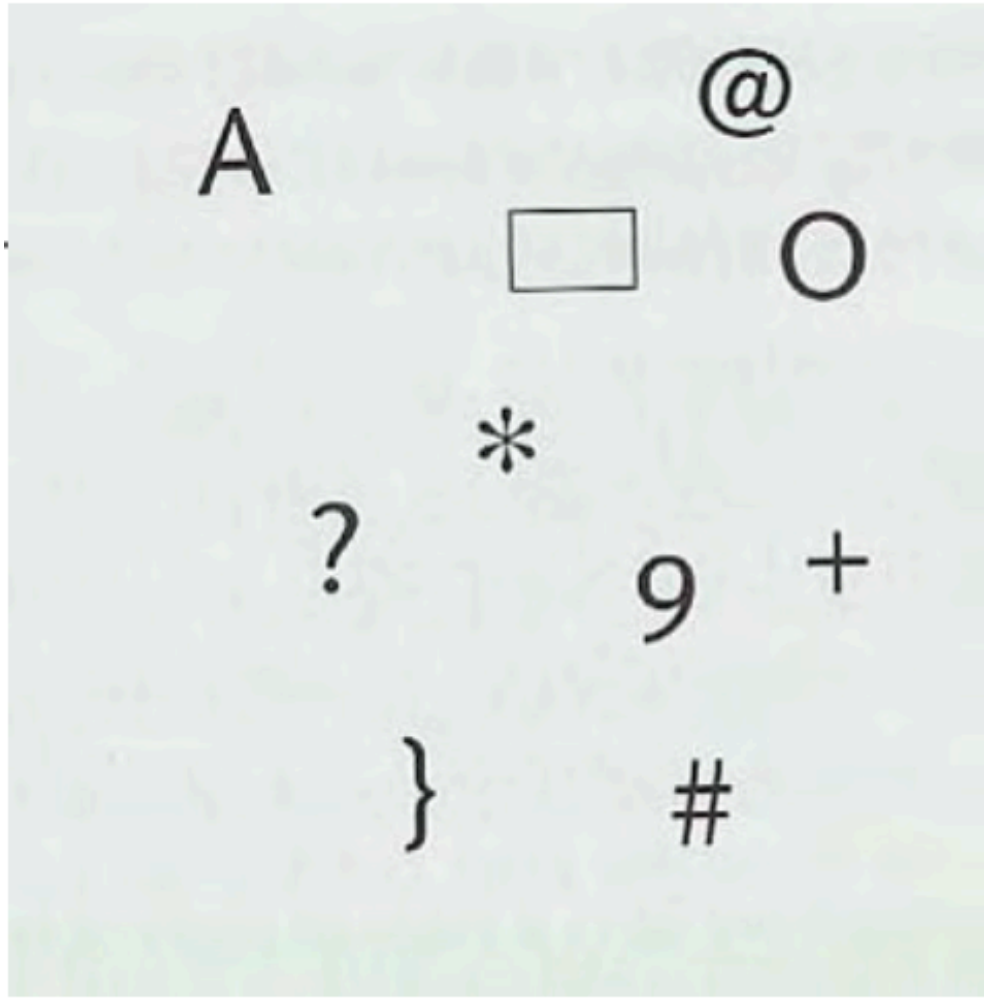
Conscious attention and restricted awareness

- Can we be conscious of things without attending to them?
- Are there only some kinds of things we see when we are no attending?
- "If we don't have a highly salient cue of some kind, we will miss changes in the world
  - Motion
  - Sound
  - Sensation

How much can we hold in our working memory?

Buffer

- How many of these symbols can you remember after a glimpse 1/10 second long?

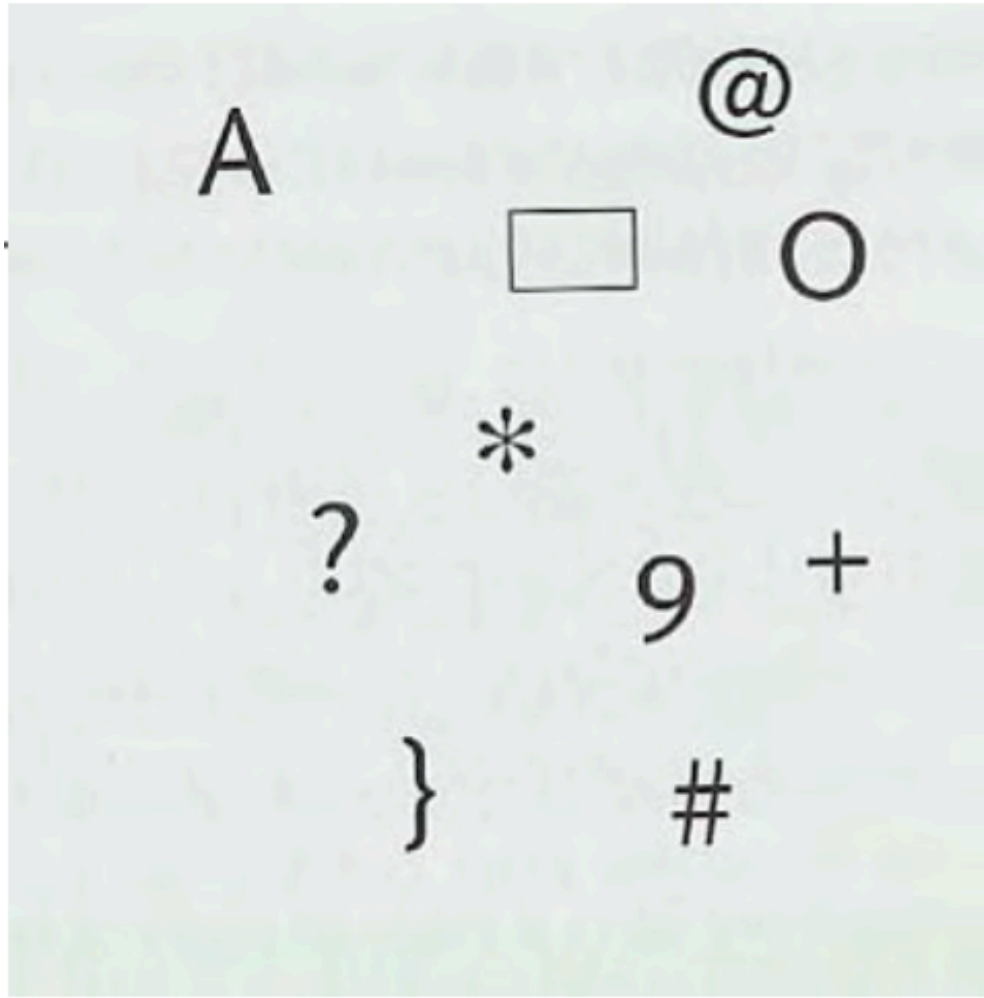


Buffer

- How many of these symbols can you remember after a glimpse 1/10 second long?

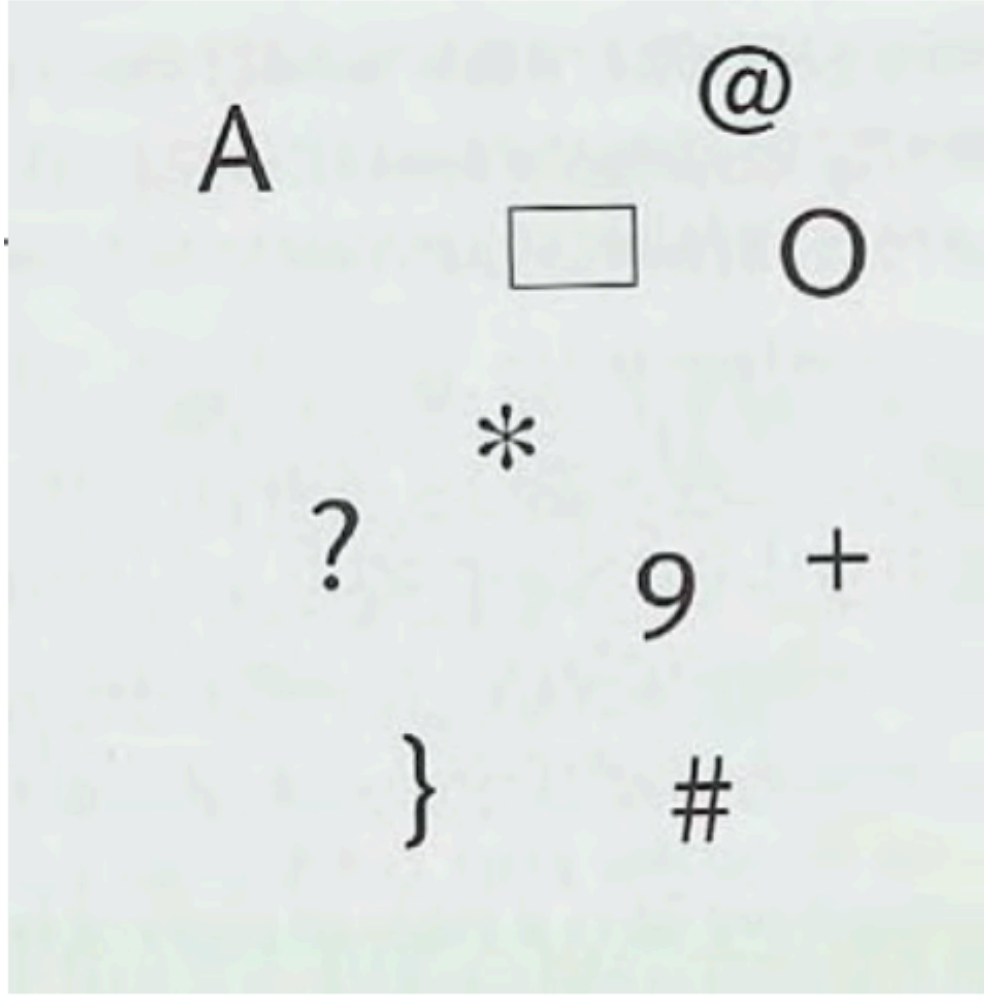
Iconic Buffer

- How many of these symbols can you remember after a glimpse 1/10 second long?
- Typically 3 -7 only
- Short-term (iconic) memory
  - Highly transient
- Visual working memory for symbolic analysis



What this tells us

- We are not aware of everything in the entire scene at the same time.
- We constantly shift our gaze and our attention to look at different parts of the scene and examine them in detail.
- We think we see the scene in detail, but we don't.



What this tells us

- We are constantly making rapid eye movements, known as saccades, as we scan a scene.
    - Vision is suppressed during saccades.
- People fail to notice large changes in the scene if the change occurs during a saccade. (McConkie, Grimes, Ballard and others).
- People also fail to notice large changes in the scene if they occur during a brief disruption (e.g. short blank period).
- This is known as change blindness. (Rensink et al., 1996)

Change blindness

- Different from inattentional blindness you don't notice when things around you are altered to be drastically different than they were a moment ago.
- Because we remember so little, we miss changes in the image
- Saccades (eyes movement)
- To see an object change, it is necessary to attend to it

Change blindness

- For example, a cut between scenes, with a change in camera angle, can also induce change blindness.
    - Movie Perception (1:15): https://www.youtube.com/watch?v=wBoMjORwA-4&list=PLB228A1652CD49370&index=4
    - Movie Perception conversation (2:00): https://www.youtube.com/watch?v=6JONMYxaZ_s&list=PLB228A1652CD49370&index=3
    - Only 1 in 10 people detected a change.
- Change blindness occurs even for objects that are the center of attention:
- Gradual change (rock) (1:20): https://www.youtube.com/watch?v=1nL5ulsWMYc&list=PLB228A1652CD49370&index=5
    - Only 33% of 40 people noticed the main change.
- Disruptions in real life
    - Intuitions about perception (2:30): https://www.youtube.com/watch?v=5YPiVSdh-RY&list=PLC0A3CAC7B3A0E288&index=2

What causes change blindness?

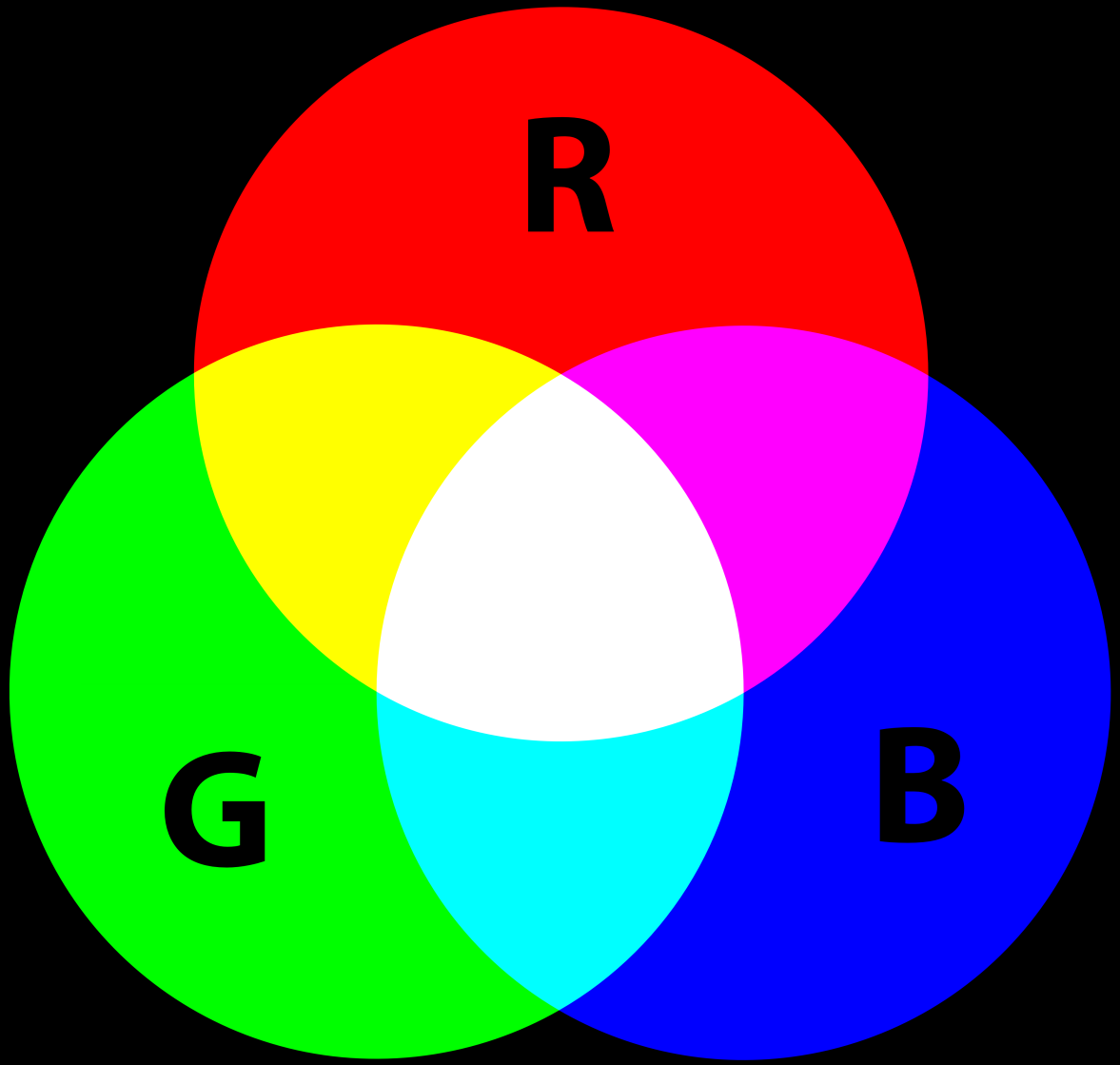
- Don't see the entire scene in detail - only the region attended to.
- We constantly shift our eyes to see other parts of the scene in detail.
    - Only attended regions get into short term memory.
- We must serially scan the picture, item by item, to find the one that is changing.
- **Attention is not enough**
    - We must intentionally process the details in order to detect the changes.

Factors affecting attention

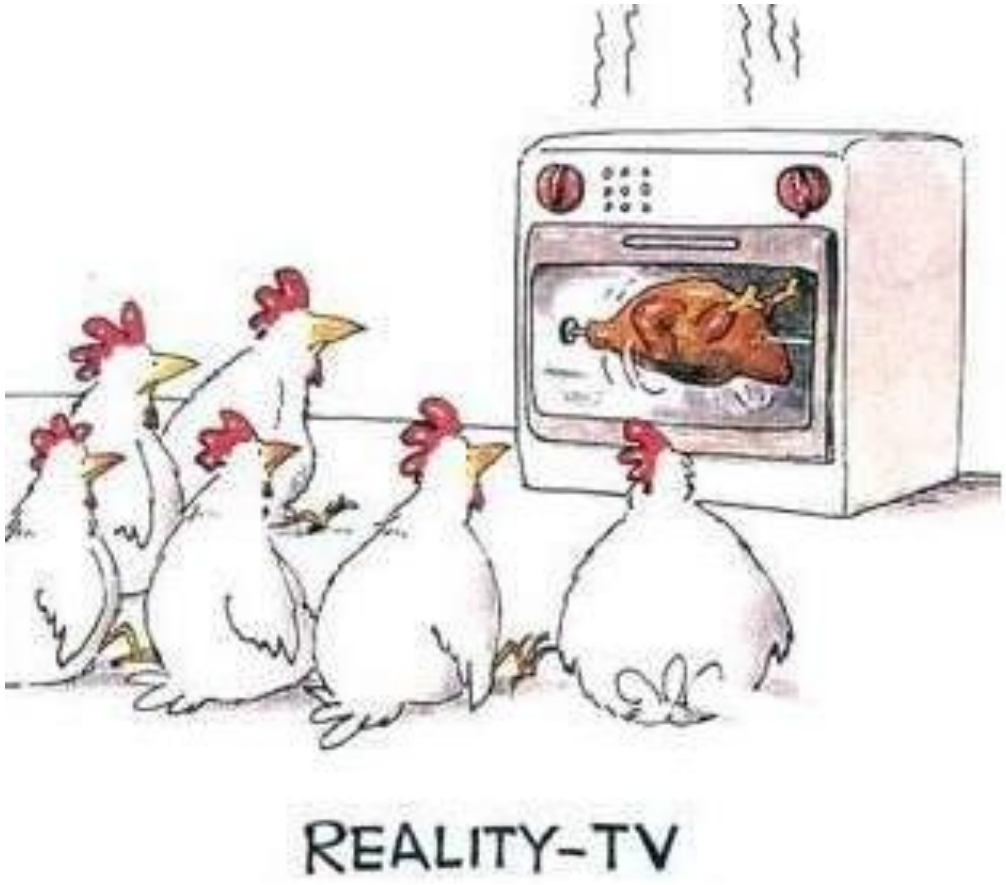
- Stress
  - Environmental stressors
  - Noise, heat, light
- Physical conditions
  - Fatigue, impairment (e.g. blurred vision)
- Psychological factors
  - Fear, anger, boredom, excitement
- Current level of demand
  - It’s not an unlimited resource!


Colours

Trichromacy Theory

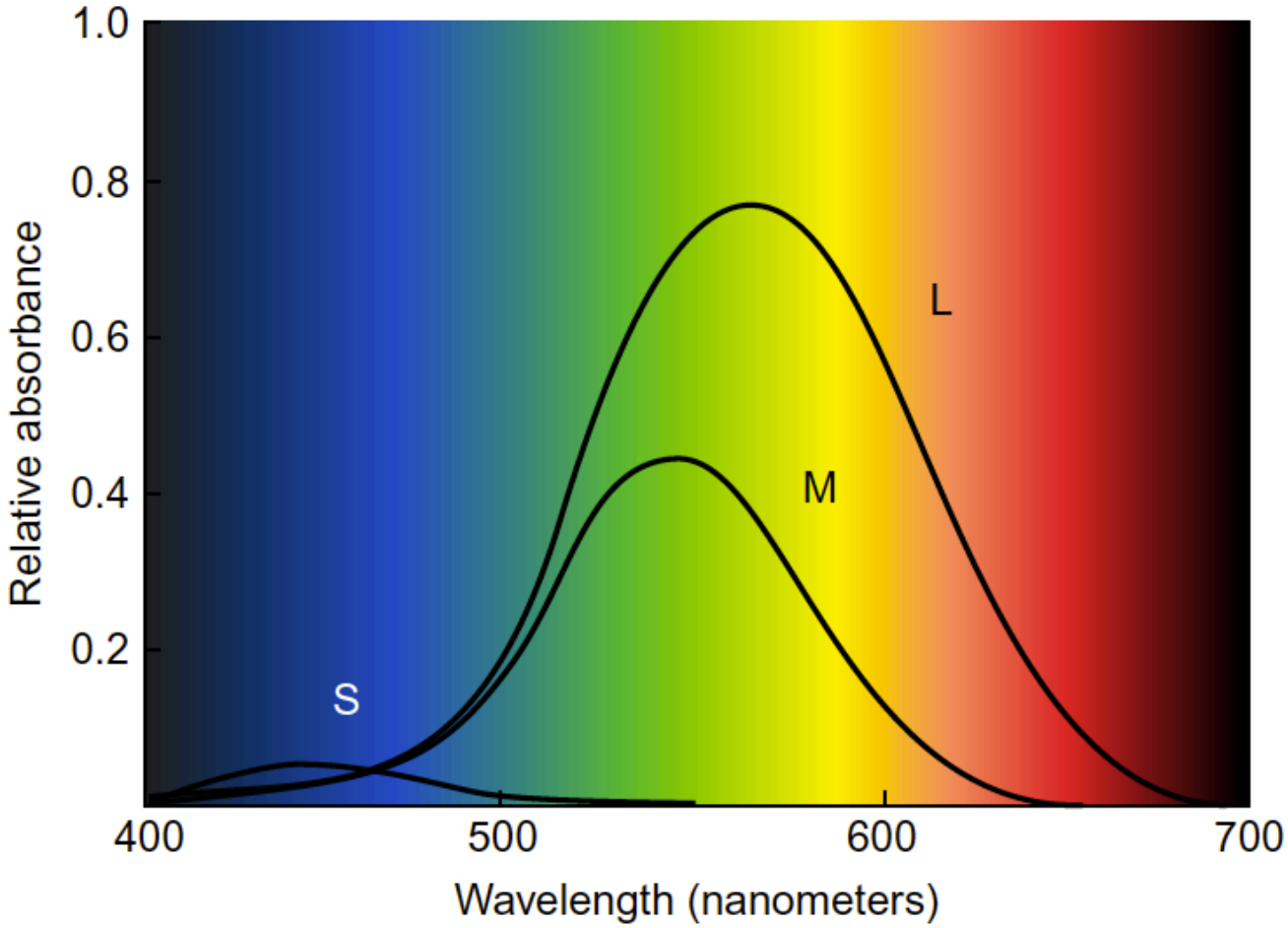


Trichromacy Theory



- Some birds, such as chickens, have as many as 12 different kinds of color-sensitive cells.

Light spectrum

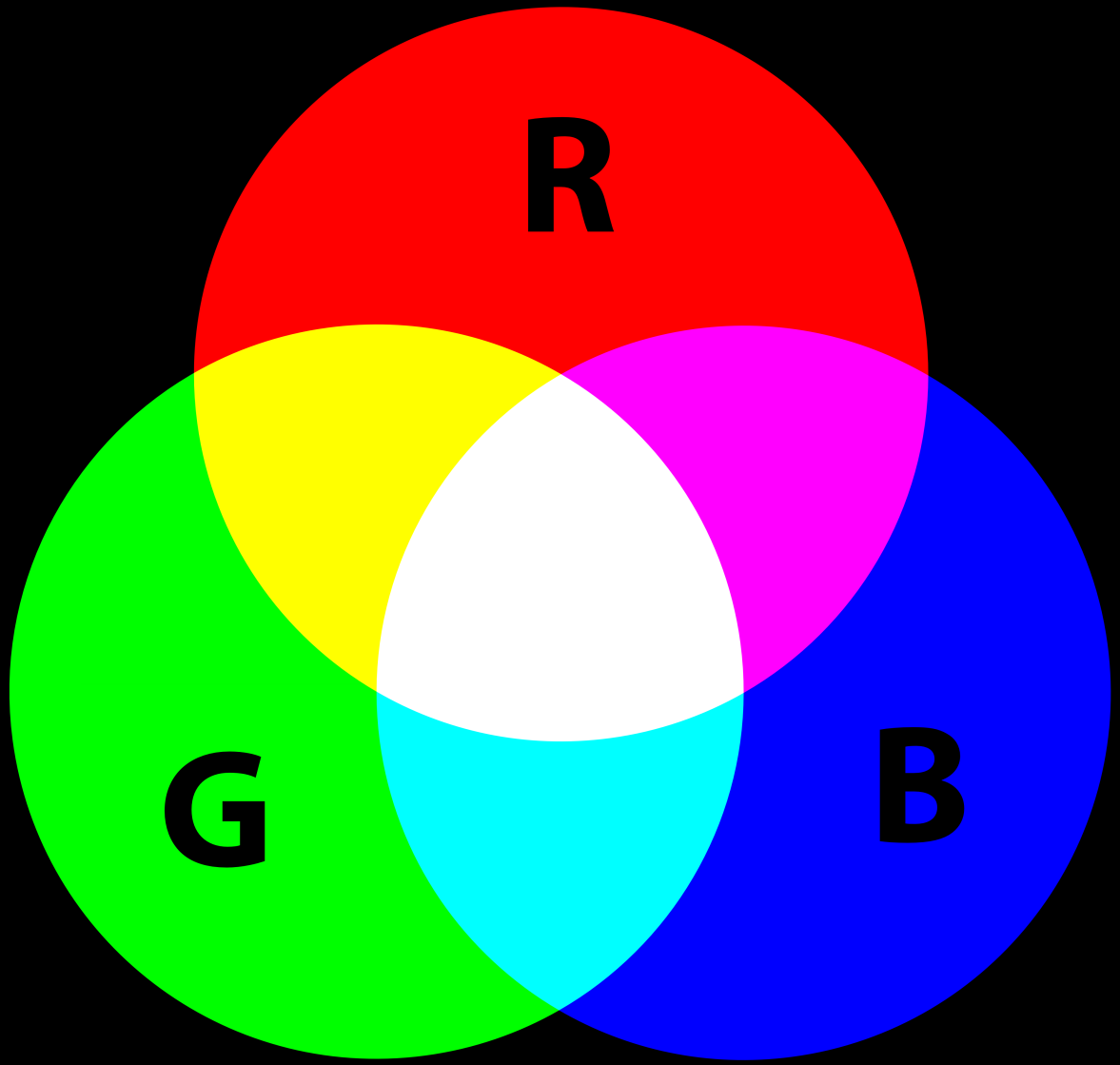


- Cone sensitivity functions. The colours are only rough approximations to spectrum hues. Abbreviations: S, short-wavelength cone sensitivity; M, medium wavelength cone sensitivity; L, long-wavelength cone sensitivity.

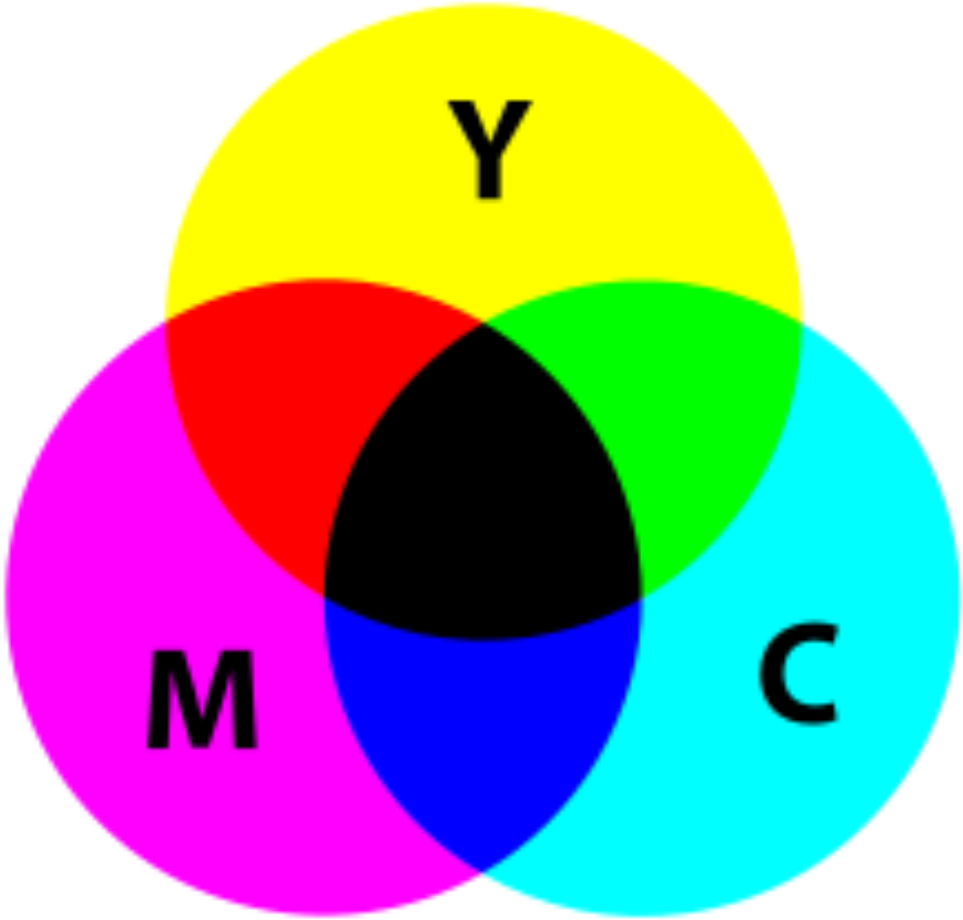
The short-wavelength receptor absorbs light in the blue part of the spectrum and is much less sensitive, which is another reason why we should not show detailed information such as text in pure blue on a black background.

The short-wavelength receptor absorbs light in the blue part of the spectrum and is much less sensitive, which is another reason why we should not show detailed information such as text in pure blue on a black background.

Addictive Colour system



Subtractive Colour system

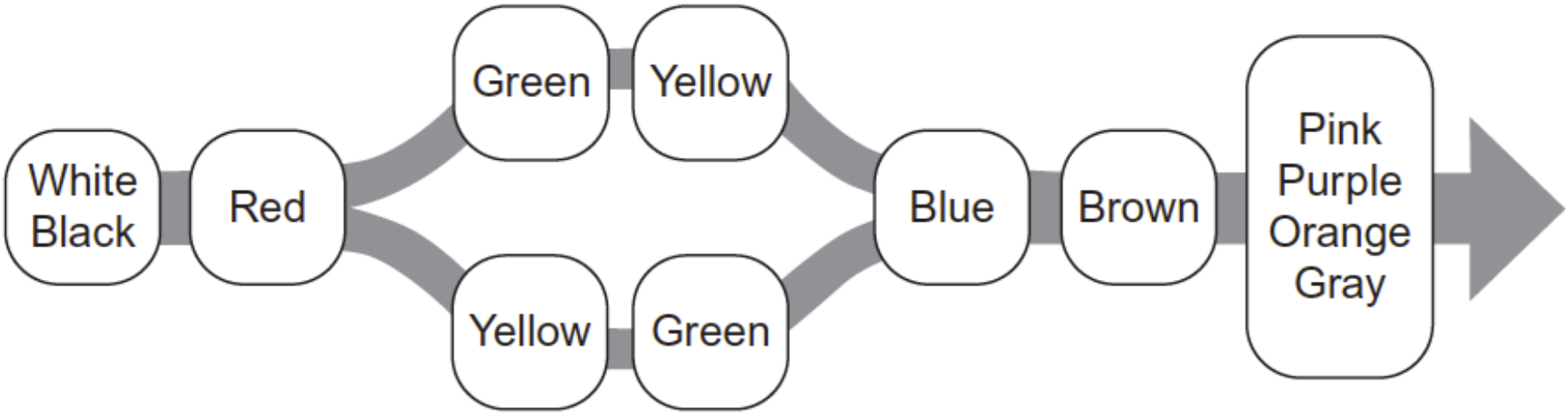


Colour Naming

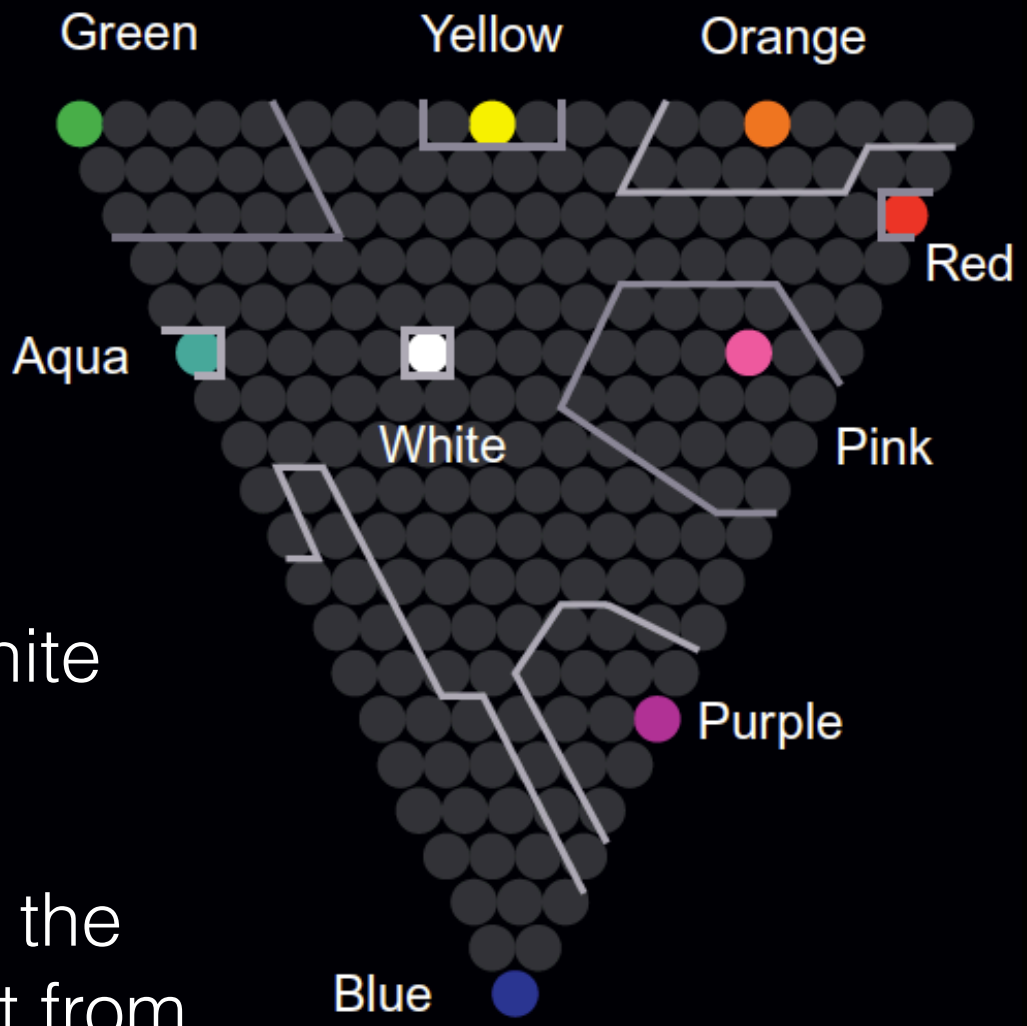
What are the basic colours?

Do we always agree in naming colours?

Cross-Cultural Naming

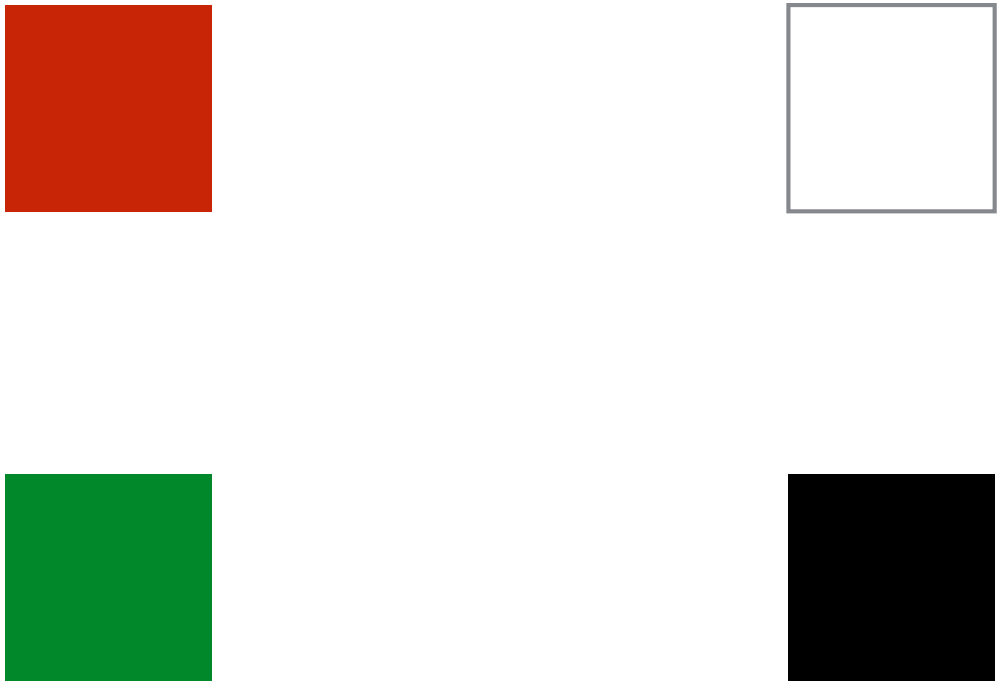


Categorical Colours

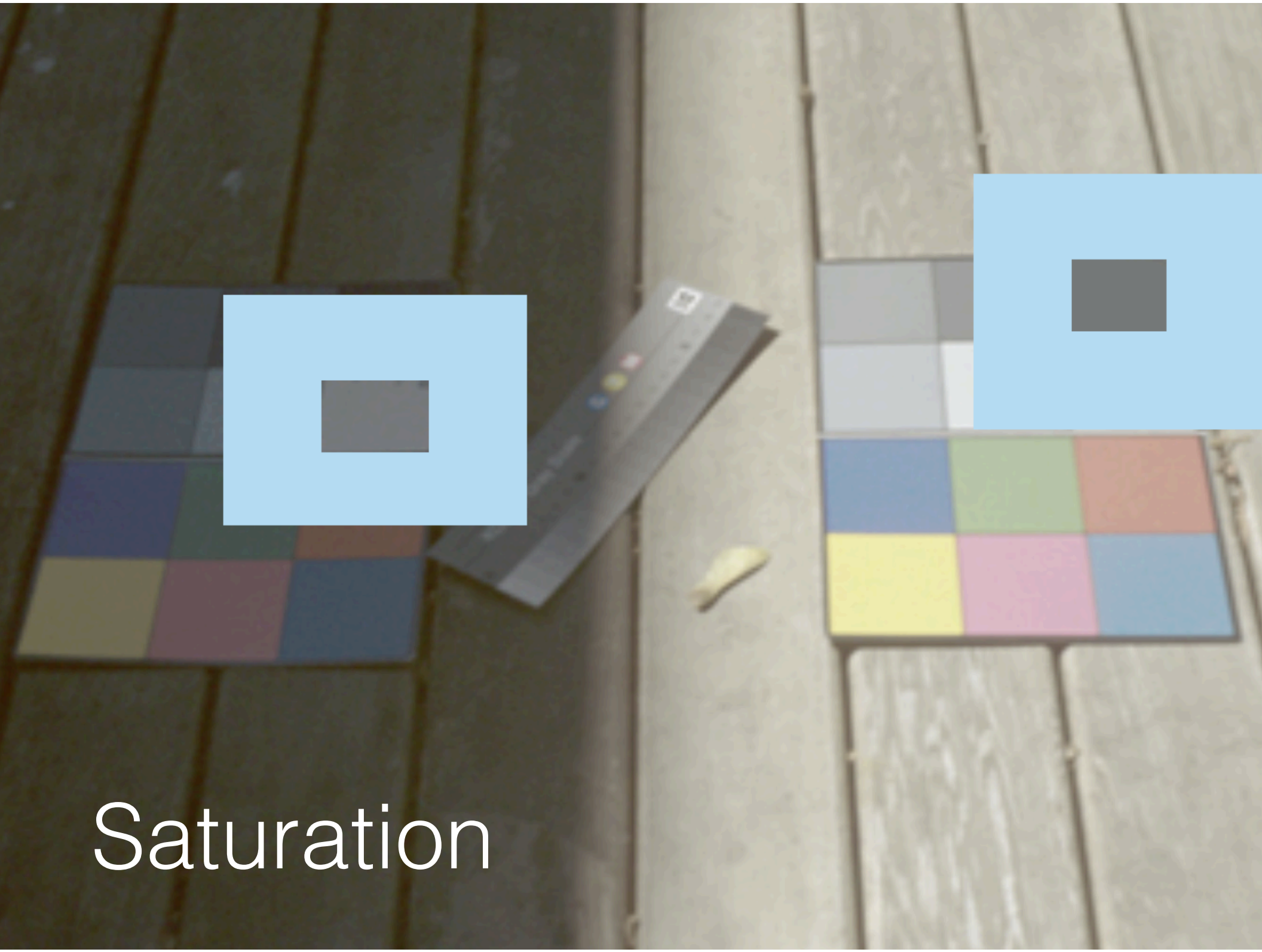


- Only eight colours plus white were consistently named
- A true colour red required the addition of a small amount from the blue monitor primary.

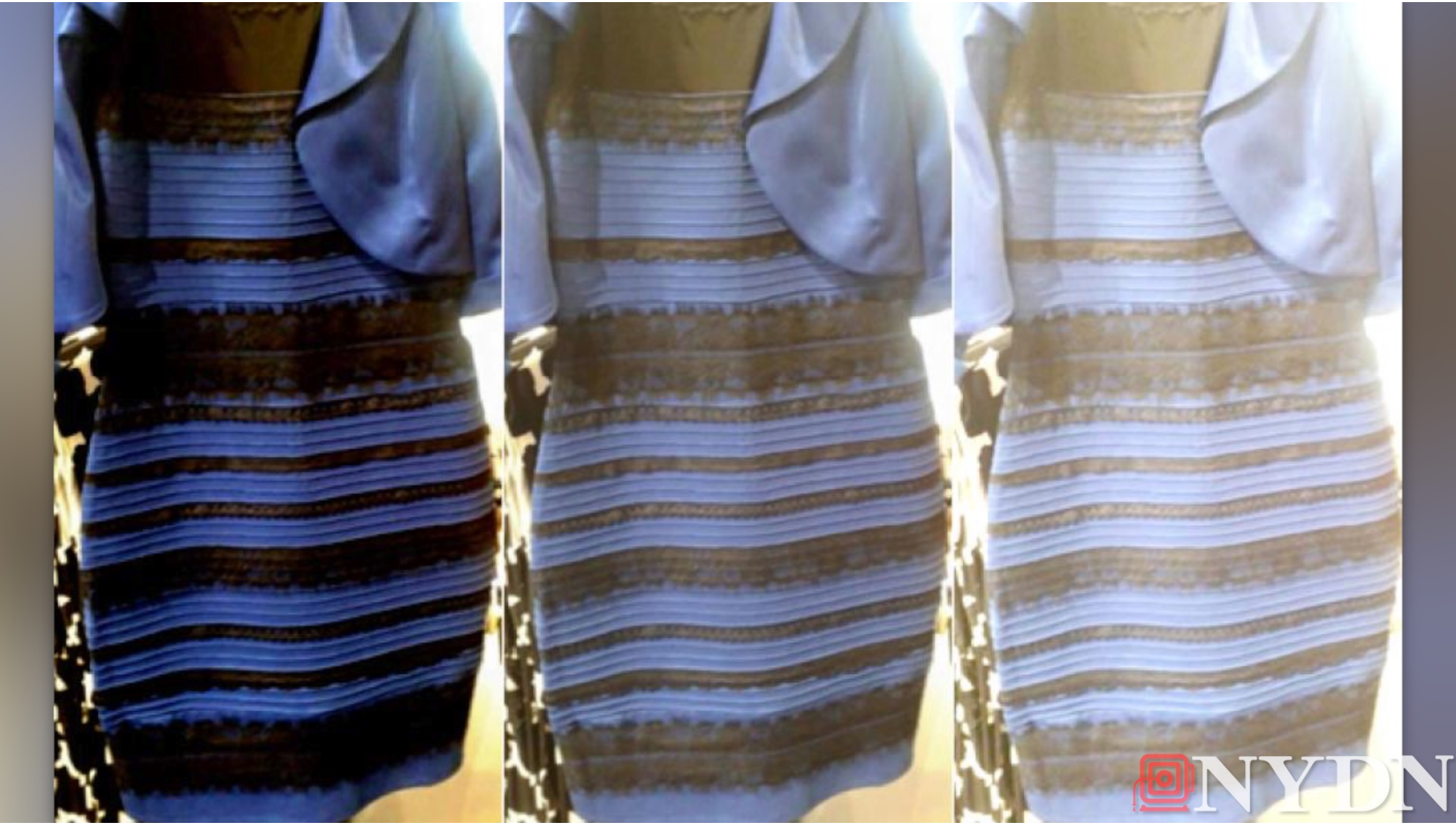
Cultural issues



Colour Appearance

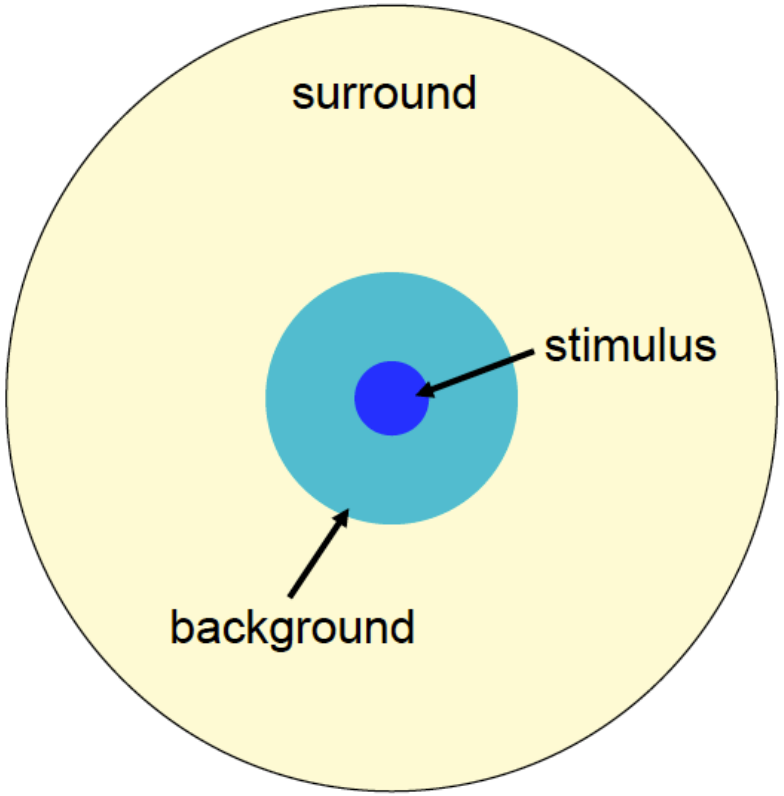

Saturation

Blue or white dress?

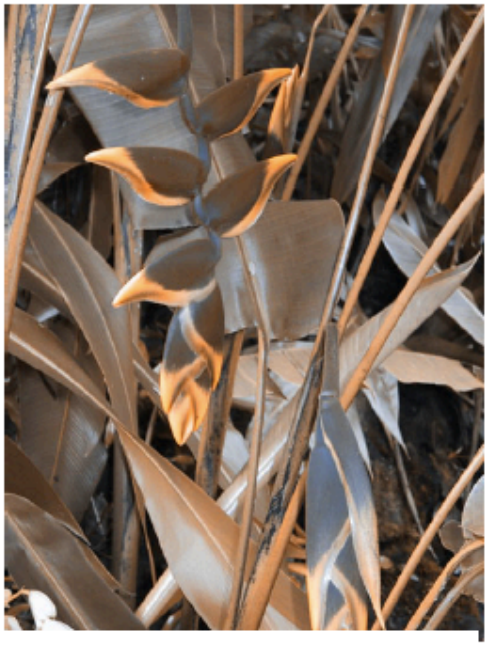


Colour Appearance

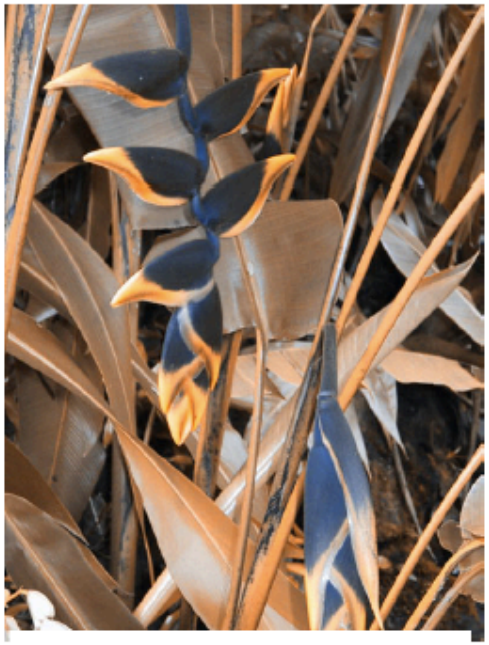
- More than a single colour:
  - Adjacent colours (background)
  - Viewing environment (surround)
- Appearance effects
  - Adaptation
  - Simultaneous contrast
  - Spatial effects
- Color in context



Colour Blindness



Deuteranope

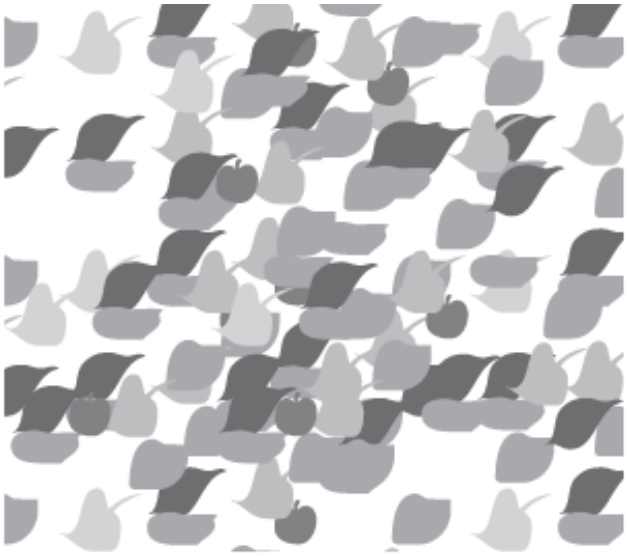


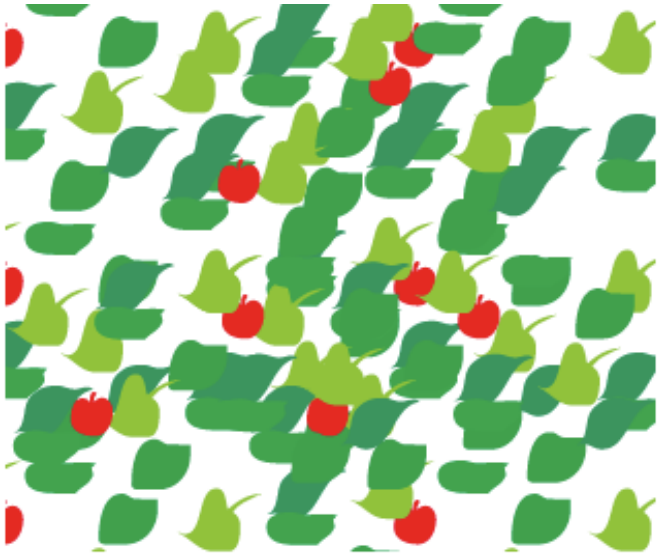
Protanope

Tritanope

- About 10% of the male population and about 1% of the female population have some form of colour vision deficiency.

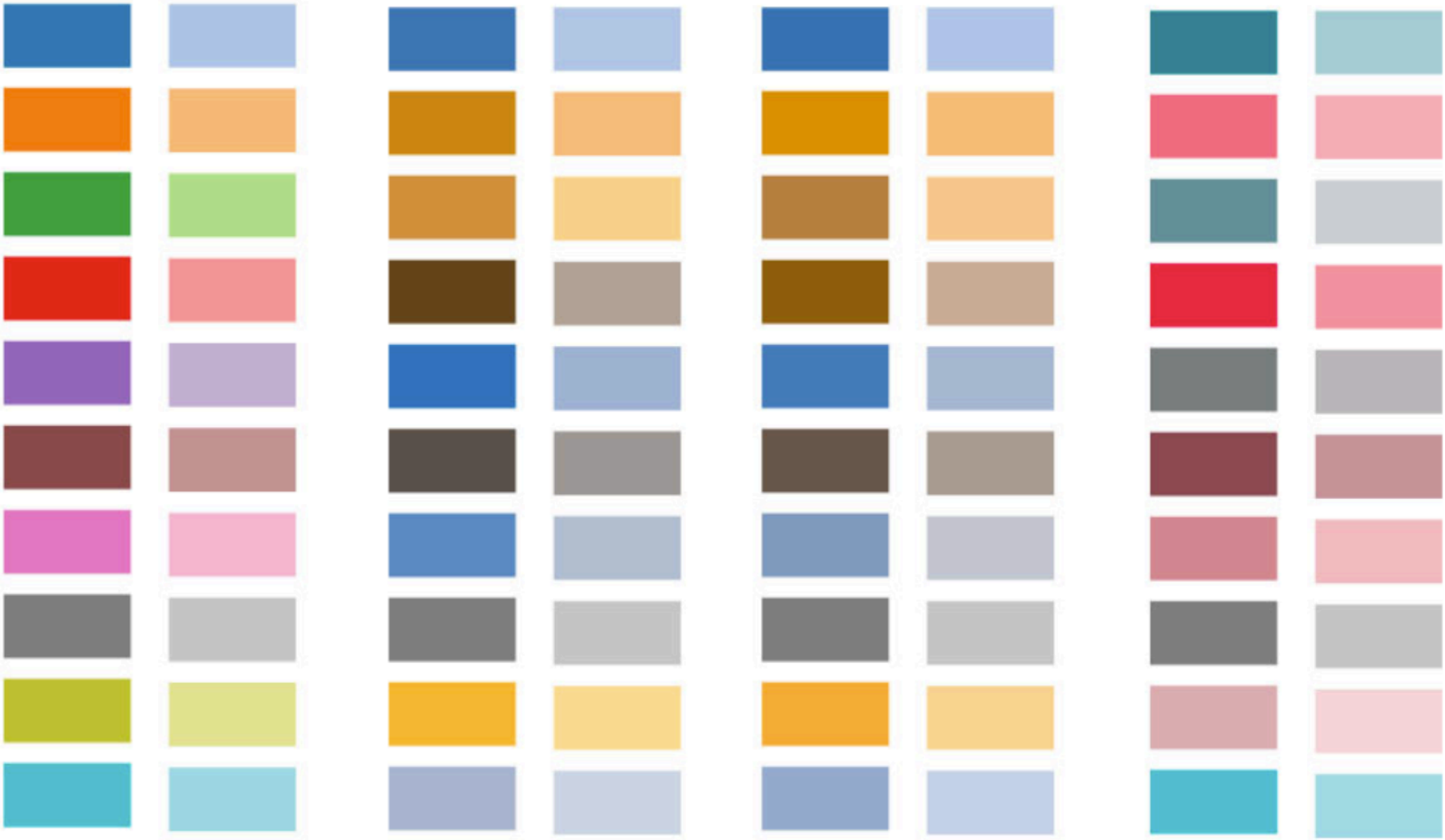
Colour Blindness



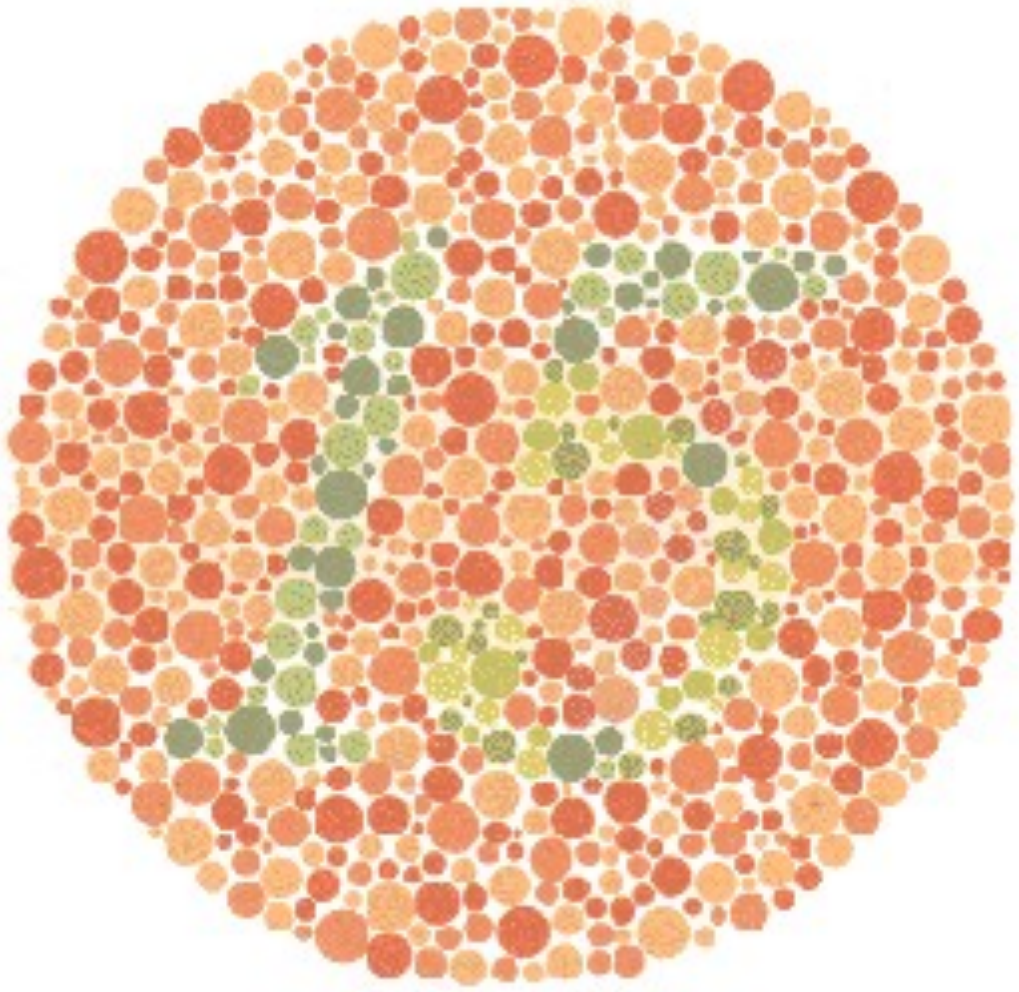


- Finding the cherries is much easier with colour vision.
- Simulates colour vision deficiencies: Photoshop plug-in (View > Proof Setup > Color Deficiency)

Colour Blindness



Colour Blindness



- Sample test: http://enchroma.com/test/instructions/