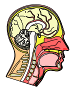


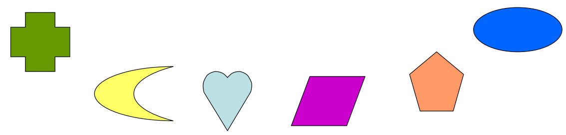

4. Recognizing things

Who goes there?

Aim

How does the brain recognize familiar patterns?

How does this affect the design of multi-media systems?

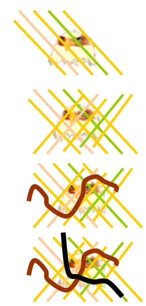


Processing perspective (recap)

- An *analyzer* is a neural assembly that processes sensory input and "fires" if a certain condition is satisfied (e.g. a line at a given orientation)
  - *Low level* analyzers feed their output to higher level analyzers (e.g. textons)
  - *Same level* analyzers interact to excite or inhibit each other
  - *Higher level* analyzers exert downward effects to modify lower ones (e.g. to suppress, enhance or prime - context and expectation effects)

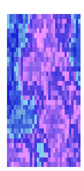
Interference is norma

- Visual system is designed to recognize shapes despite "noise" (e.g. a cover of leaves)
- Given incomplete data, it **must** *hypothesize* what is there
- What seems to us as just a "perception" is actually *the brain's most likely hypothesis of what is there*



Link: Illusory figures

Why minimalism?



What or who is it?

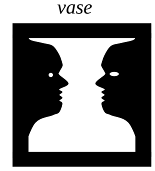
- Why see objects under minimal cues?
- Natural visual fields (e.g. a jungle or a busy road) offer *many potential shapes*
- **Some are real** and some are not: we can see animals and faces in clouds - or under cover

The ability to see illusory figures is also the ability to see camouflaged or covered objects in a normal environment

All input is ambiguous

- In complex sensory fields, ***ambiguity is normal***
  - Incoming sense data can be validly processed multiple ways
  - Ambiguity is simply where the alternatives are equally likely

vase



people talking

Application: Support perception by reducing ambiguity

Try to see alternative views in the same field at once!

Recognition = resolving ambiguity

- Object recognition involves *resolving ambiguity* on various processing levels
- We take these ambiguities to be:
  1. Border contrasts (last lesson)
  2. Figure vs ground
  3. Framing
  4. Composition
  5. Feature detection
  6. Classification

The 6 steps of ***object recognition***

 7

Link: Figure-ground

2. Figure/ground ambiguity

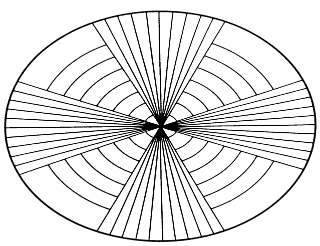
- Either side of a border contrast could be "object"
- Once a decision is made, the visual system makes:
  - *Ground* to look *continuous*, even if covered, *less defined*, and perhaps *without boundary*
  - *Figure* to look as a *hard surface* with *definable boundaries* and *shape*

For a shape to be recognized it must first become "figure"

 8

Example

- If the *Maltese cross is* figure, it *stands out*, the circles seem to continue around it
- If *the circles* is figure, they stand out, and the radiating lines seem to continue around



You can see it one way or the other ***- but not both***

 9

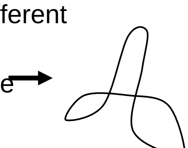
Application: Support F/G choice

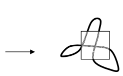
- Backgrounds should be:
  - Fuzzy or out of focus
  - Continuous, and consistent across the field
  - Repetitive simple patterns of undefined surface
  - Not attracting attention
- Foreground should be :
  - Defined boundary
  - Clear shape
  - Hard surface

 10

3. Framing ambiguity

- Which set of edges or lines to analyze?
- There are many ways to "frame" a visual field, and each may give different objects
- For example, you do not see the 4 in this shape unless you frame it, or it is framed for you

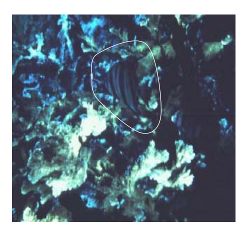




 11

The framing problem

- In a 1,000 x 1,000 pixel field, how many ways to frame a 10x10 object? 990 x 990 = 980,100 ways. For a 20x20 pixel object 980 x 980 ~ 960,400 ways. etc
- For *all size frames*, from 10 to 1,000, the possibilities are *enormous*
- Framing ambiguity presents a huge processing problem



How can we form a shape until we know the frame?

 12

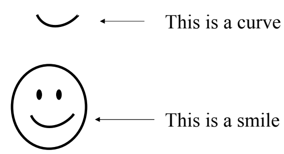
Application: Use framing

- Support framing by drawing a frame around key objects to define the *processing boundary*
- When reducing image size (e.g. a screen clip)
  - First *crop* the image (section what's important)
  - Then *reduce image size* (number of pixels used)

Link: Gestalt theory

4. Composition ambiguity

- How are lines "formed" into a figure?
- *Gestalt theory:* Elements create a *whole form* (gestalt) that affects how its *parts* are perceived:



Reductionism

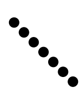
- Is the idea that perception can be *reduced* to base sensations, which add up to create the perception
- ***Reductionism*** is only partially successful
- ***Added "elements" interact*** to change *the perceived elements themselves*
- When element interactions cause still further changes, etc, this gives emergent properties

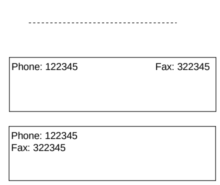
Gestalt "Laws of Organization"

- Define what elements are formed as shapes
- Still used by artists and designers
- *Proximity*
- *Continuity*
- *Similarity*
- *Closure*
- *Simplicity*

Proximity

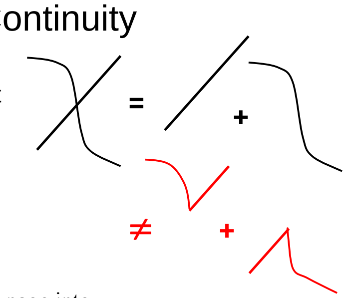
- Dots *close together* are formed into (seen as) a line
- Application:
  - Dotted lines are as effective as solid lines
  - *Put like things together* (in same place) - group similar things



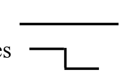


Continuity

- The shape is seen as a straight line crossing a curved one
- Rather than two bent segments



Application: Organize page into continuous not discontinuous lines



Similarity

- We compose or form this pattern as vertical columns (of similar letters)
- Rather than horizontal rows (of oxoxoxo)

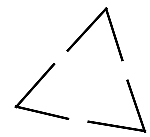
OXOXOXO
OXOXOXO
OXOXOXO
OXOXOXO
OXOXOXO

Application: Similar things will be connected e.g. same color/font headers will be connected to a pattern/structure

 19

Closure

- We see (compose) a triangle, not three "v" shapes
- Missing parts are assumed accidental or covered



Application: Use background colors and patterns that continue to give stability and context

 20

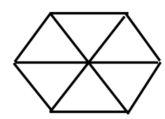
Simplicity - the information reduction principle

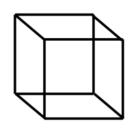
- Incorporates most of the Gestalt principles
- *We form (see) what <u>most simply accounts for all available information</u>*
- *OR - We form (see) the interpretation that gives the biggest reduction in information*
- e.g.
  - a line = three values (length, position and orientation) + processing of "lineness" (*vector graphics*)
  - OR the coordinates of all the dots in the line + processing at the dot level (*bitmap graphics*)

 21

Example

- A is simpler as a *flat figure*, so we see it that way
- B is simpler as a *3D cube figure*, so we see that
- *<u>Both</u> could be seen as 3D cubes or as flat figures*
- The ambiguity is automatically resolved by our visual system

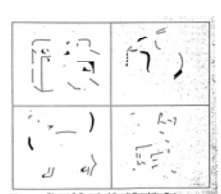

A


B

 22

Example - figure completion

- We "form" figures according to lines
  - proximity
  - continuity
  - similarity
  - closure
- *Ambiguity does not prevent perception!*



What do you see here?

 23

Application: Harmony and contrast

- *Harmony:* When perceptual analyses combine to support a single impression (usually good, especially for backgrounds)
- *Confusion*: When perceptual analyses contradict (not usually good)
- *Contrast:* When *one* element is unexpected:
  - What stands out gets attention focus
  - Used in humor – the punchline
  - May be good or bad

 24

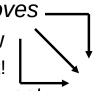
Link: Saccades, Iconic image

Scanning

- Larger figures cannot be formed in one "view"
- We *scan* larger pictures at about 4 views/second
  - Scan to points of interest (boundaries, faces)
  - Westerners scan left to right (reading)
- Each "iconic" view is almost complete, like a photo, but fades quickly
- Even in focused vision, *saccadic* movements occur to refresh the image

Application: Attention flow

- Manage attention flow
  - Assume focus begins at top left *then moves*
  - Draw attention in a logical and easy flow
    - E.g. don't draw first attention to bottom right!
  - Main menus are better at the left or top - not bottom, right or middle
  - Align text on *left* - right alignment is optional
  - Draw attention to points the user wants to go to
  - Put advertisements or routine messages at top or right (so they can be ignored)



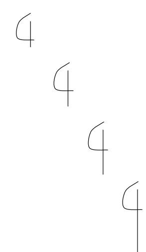
4. Feature ambiguity

- Which aspects of a pattern are important?
- People do not analyze *all* the pixels (as a computer does), but compare *features*
- *Feature:* Some aspect of an object that distinguishes it
- Features can be complex *relations between elements*, e.g. roundness

Link: Computer Vision

Feature example

- Handwriting recognition depends on the *relations* between lines and curves
- A 4 becomes a 9 by lengthening upright - transition from one pattern to another is continuous
- The 4 vs 9 classification depends on line curve *ratio* - not pixel by pixel analysis



Link: Pattern recognition

Pattern recognition

- Objects have many features
- Compare object to
  - a *fixed template* - a dog is like my dog
  - a *general class* - a dog has "dogness"
- Classes also have *defining* features
- *What are key features for a class*?
  - e.g. is size a key feature of dogness? Is a dog as big as a horse still a dog?

Computer pattern recognition

- Computers that use *fixed templates* and process at the *pixel (dot) level* struggle to recognize handwritten numbers and letters because:
  - Humans use class features not templates
  - Complex class features are based on higher order *pixel relations* e.g. handwriting
  - Computers can use *human experts to define key class features*

Application: Support key features

- Identify key recognition features for the object
  - Exaggerate or enhance them
  - Improve contrast
  - Use color to draw attention to them
  - Add key features if necessary
  - May omit non-key features
- *Design icons around <u>key features</u>*

5. Classification ambiguity

- Based on key features, objects are *classified*
- Once classified an object is then *recognized* (e.g. as a dog)
- Prior knowledge about that <u>class</u> then carries forward (e.g. that dogs can bite)
- Classes can expand, contract, or form sub-classes
  - Animals bite
  - Aggressive dogs bite
  - Little dogs just nip you
- If not classified an object is "new"

Link: Danger recognition

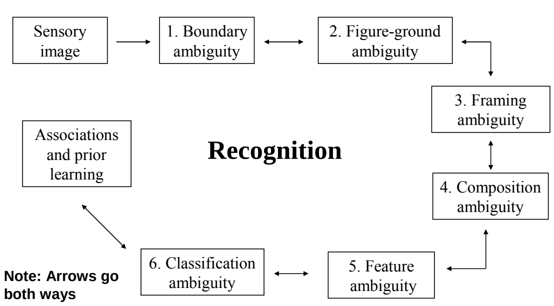
Non-recognition

- An object of *unknown class* usually generates **<u>arousal,</u>** either *fear* or *curiosity*
- Most animals respond to something new first with flight (*fear*), then with approach and investigation (*curiosity*)
- Most categories are formed in youth - adults only *rarely* form new classes
- Adults usually classify new objects into known class categories

Class associations

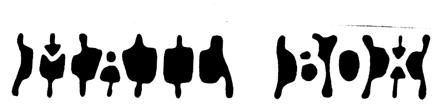
- Class associations carry forward prior learning with similar objects
- Previous associations can be
  - *Positive* - dogs are friendly
  - *Negative* - dogs are dangerous
- Can also be
  - *Appropriate* (positive learning transfer)
  - *Inappropriate* (negative learning transfer)
- *Application*: Check the associations you evoke are as you expect i.e. positive and appropriate

Review



Top down effects - context

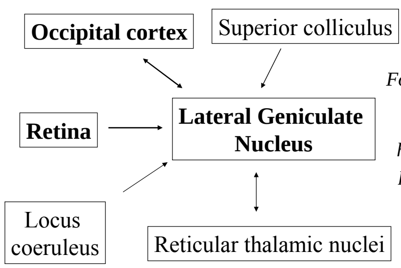
- Figure vs ground is one of the *first* choices of object processing
- *High level processing* (e.g. language) can alter a *low level decision* (e.g. figure vs ground)
  - Example: Is there anything in yours?



Link: Visual system organization

Neural basis

Brain is designed to process *down* as well as *up*



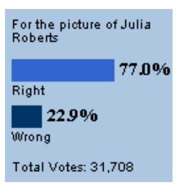
For every projection to the LGN from the retina, there are hundreds from other parts of the nervous system

Example: face recognition

- Still hard for computers due to the huge variation e.g. what gives the "family resemblance" in the faces of a family?
- Babies recognize faces from birth - built-in face analyzers
- People frame and form faces in a tenth of a second, i.e. very *quickly*
- Can recognize faces even after years of absence and aging

Example 1

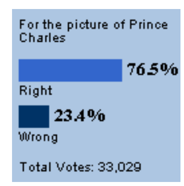
- Who is this? (computers find the problem hard because of all the search possibilities)
- How long did it take you?



Results of online poll

Example 2

- Who is this?
- How long did you take?
- Did you "search" every face you know?



Symmetry- hemispheres compare

- *All faces* have left/right symmetry - even those we have never seen (e.g. strange animals)
- Ability to detect symmetry helps detect enemies and recognize friends
- Requires high level hemisphere comparisons, between left and right visual fields
- *"Beautiful"* people are more symmetric
- Intelligence may be based on this *ability to judge "same" and "different"* (= and ≠)

Recap

- *Pattern recognition* is the resolution of a series of information processing ambiguities
- Resolving each ambiguity increases the higher level information gained, but each choice can be *mistaken,* as illusions illustrate
- "A picture is worth a thousand words" due to all the processing our visual system does
- *Information theory* explains how visual processing reduces sensory information overload

Link: Information theory

Information theory

- *Amount of information* in a signal is the number of possible alternative values it can take.
  e.g. two choices = 1 "bit", 256 choices = 1 byte
- The *encoding scheme* determines how information is *processed into the signal*
  - Computer's use binary encoding, 1 signal = 1 bit
  - ASCII coding is for characters (1 letter = 1 byte)
- *Higher level encoding* can be more efficient
  - Data compression (same information - smaller signal)
  - Morse code is more efficient than ASCII

 43

Encoding – information per signal

Link: code

- One computer bit (a 1 not 0)
- The first letter - "a" (1 byte)
- The first word ("aardvark" - many bytes)
- The first book (Zend-Avesta?)

More choices (processing) means more information

Higher encoding means the *<u>same signal</u>* *carries* *<u>more information</u>!*

 44

Higher level encoding

- Processing *encodes* information at higher levels
  - A square can be a complex set of many dots (pixels) OR
  - A square can be four lines (vectors) at right angles
  - It is *simpler* as four lines
- Processing *reduces* the information that must be handled - Vector graphics are more efficient than bitmaps for CAD work
- Processing *percepts leads to concepts (or ideas)*
- *Application:* Support simplification process at all levels e.g. present ideas and content in a logical *<u>structure</u>*

 45

Higher processing takes time

- Symbol recognition time depends not on the signal form but on *the number of alternatives* e.g
  - 7 is one of ten numbers
  - f is one of 24 letters
  - a word is one of millions

  Longer to recognize
- Recognition is harder (requires more processing) if there are more options
- *Application:* Give the user as few choices as necessary *at one time*

 46

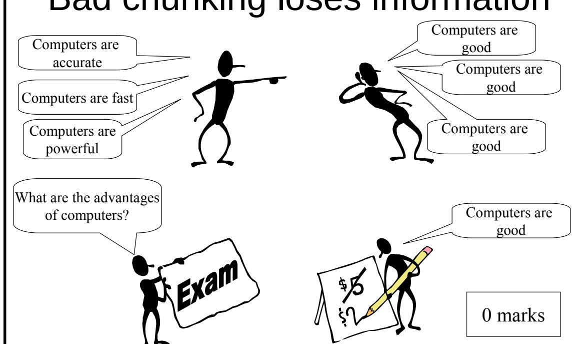
Link: Chunking

Chunking

- Chunking helps encode information at a higher processing level by *showing many things as one thing*
- It improves perception and memory enormously
- Without chunking, people can only reliably manage about seven distinct items at once
- *Application:*
  - Lists/menus/toolbars should only have 7 items in - if there are more, *group them*
  - No more than seven areas on the screen
  - No more than seven choices offered

 47

Bad chunking loses information



 48

Link: Visual information rate

Handling information overload

- Base visual hardware information processing rate is an impressive 600 Mbits/sec
- People use higher processing to improve performance
  - e.g. New drivers have too many things happening at once. Information overload occurs because they cannot handle (process) the input. With experience the same information load seems less - better processing
- Consider the vast information input while driving at speed on a motorway. The experienced driver knows:
  - what is background
  - frames objects quickly
  - what to watch for (key features)
  - classifies situations correctly
  - what associations apply

Experienced drivers know when to slow down and when to speed up, i.e. recognize **situations**

 49

Summary

- Shapes must be "figure" not ground
- Shapes must be framed in the visual field
- Shape composition follows laws of "good organization" or best information reducing process
- Eyes scan key features in object composition
- Key features/attributes are used to classify
- Classification effort depends on number of choices

 50

Summary (ctd)

- Recognition may give positive or negative transfer of previous knowledge
- Non-recognition may give fear or curiosity
- Perception involves information reduction by higher processing
- Only about 7 items can be processed consciously at once - "chunking" solves this problem but can also lose information
- People like symmetry (left/right "sameness")

 51

Application summary

- Be clear what is figure and what is ground
- Choose common color/texture for the same ground
- Shapes with a clear outer boundary are easier to form
- Framing an object makes it easier to recognize
- Viewers will "fill in the gaps" for incomplete figures
- Use proximity, continuity and similarity to organize
- Use key features to represent things with minimal effort (e.g. on buttons)
- Put important things at top left, or at top, or at left
- Manage the *flow* of attention

 52

Applications (ctd)

- Minimize categories when presenting user choices
- Use appropriate symbols to carry forward previous knowledge correctly
- Use familiar things so new users feel comfortable
- For CAD applications, vector graphics are preferred as they use much less space than bitmapped graphics
- Don't present more than 7 items in a single list
- Reduce choices to 7 or less
- Left/right symmetry, as for example in faces, is attractive

 53

Homework – To Read

- Figure-ground
  - http://en.wikipedia.org/wiki/Figure-ground_%28perception%29
- Gestalt theory
  - http://en.wikipedia.org/wiki/Gestalt_psychology
- Pattern recognition
  - http://en.wikipedia.org/wiki/Pattern_recognition_%28psychology%29
- Information theory
  - http://en.wikipedia.org/wiki/Information_theory
- Chunking
  - http://www.answers.com/topic/chunk
  - http://en.wikipedia.org/wiki/Chunking_%28psychology%29

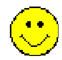
 54

OTHER LINKS

- Illusory Figures
  - http://www.iovs.org/content/11/6/540.full.pdf
- Saccades
  - http://www.answers.com/topic/saccades
- Iconic Image
  - http://www.answers.com/topic/iconic-image
- Computer Vision
  - http://en.wikipedia.org/wiki/Computer_vision
- Danger Recognition
  - http://www.answers.com/topic/danger-recognition
- Visual System Organization
  - http://www.answers.com/topic/visual-system-organization
- Code
  - http://en.wikipedia.org/wiki/Code
- Visual Information Rate
  - http://www.answers.com/topic/information-rate-of-vision

 55

Next: Depth and movement

- What is the difference between
  - this? 
  - And this? 

 56