Objects, images and words

IAT 814
Week 10 Lecture B
Lyn Bartram
12.11.2009

SCHOOL OF INTERACTIVE
ARTS + TECHNOLOGY

Object recognition

- Object: identifiable part of the visual world (not feature, but assembly of critical features)
  - Cognitively group visual attributes
  - Powerful way to organise data

- How do we recognise (identify) objects?
  - Image based theories
  - Structural 3D theories

- Object perception: Object displays

Image Based vs. Structure Theories

- Template theories based on 2D image processing
- The mind as a “huge movie reel” (R.Taylor)
- Image memory
- In structural theories we extract the structure of a scene in terms of 3D primitives
- Reconstruction of 3D world

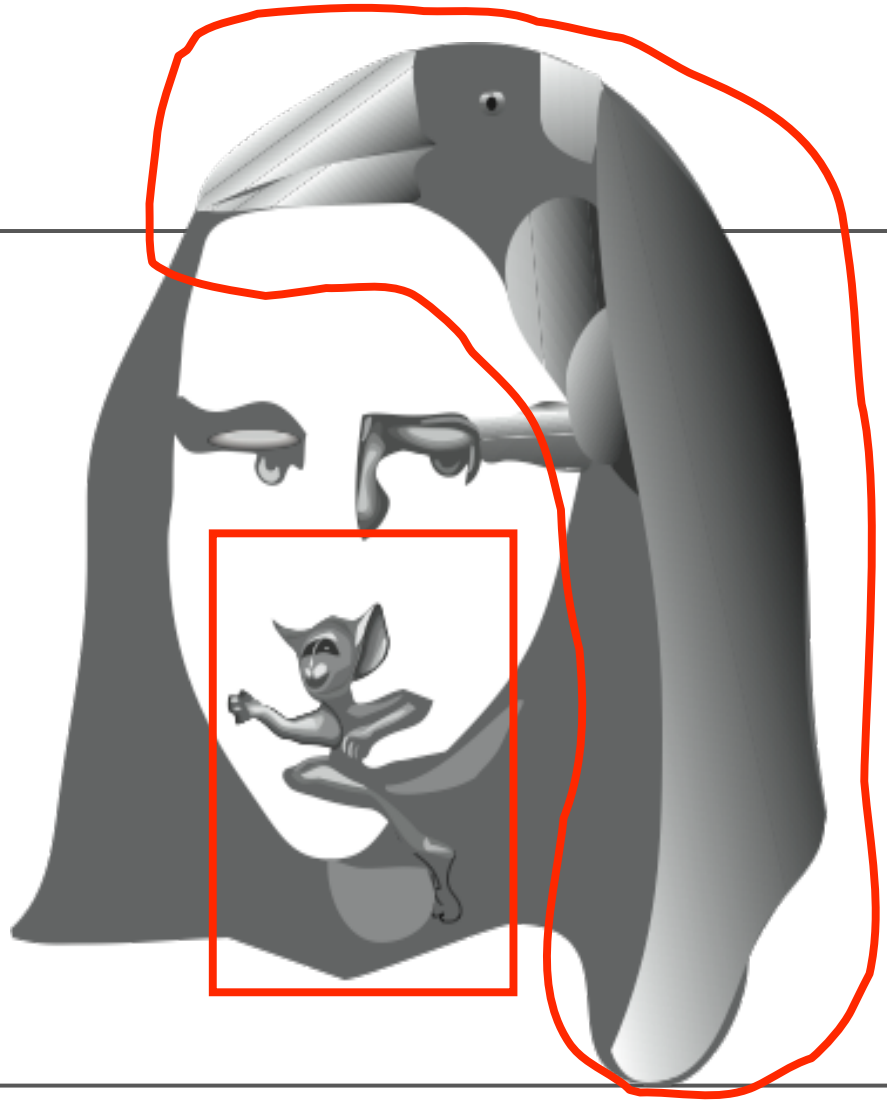
- Probable that both kinds of processes occur
- Faces seem to be special in object perception

Template theories

A template with simple morphing operations

- Most image recognition is size-resistant but scale can matter

- 4-6 degrees is optimal for object perception



Properties of Image Recognition

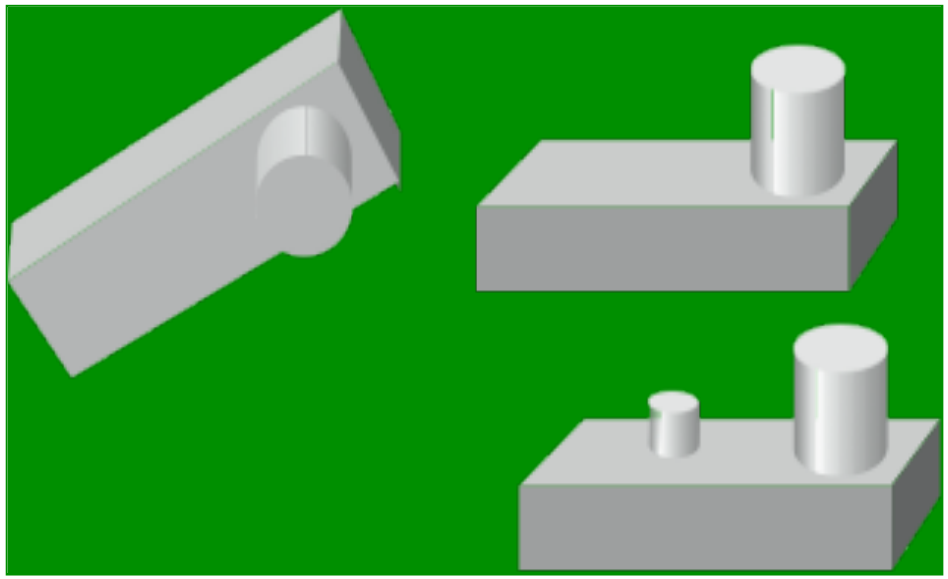
- Remarkable image *recognition* memory
  - 90% recognition accuracy for whether a picture was in a set of 2560 seen before (Standing, ‘70)
- Up to 5 images for second in object identification
  - RSVP
- May store *canonical* views, or multiple views:
  - monkey brains have nerve cells that respond to particular face orientations.

Applications of Image Recognition

- Icons and iconic representations
  - Trigger LTM concepts
  - But recall discriminability issues!

- Applications in image interfaces and databases
  - Search priming
  - Rapid serial visual processing (RSVP) : quick presentation of images

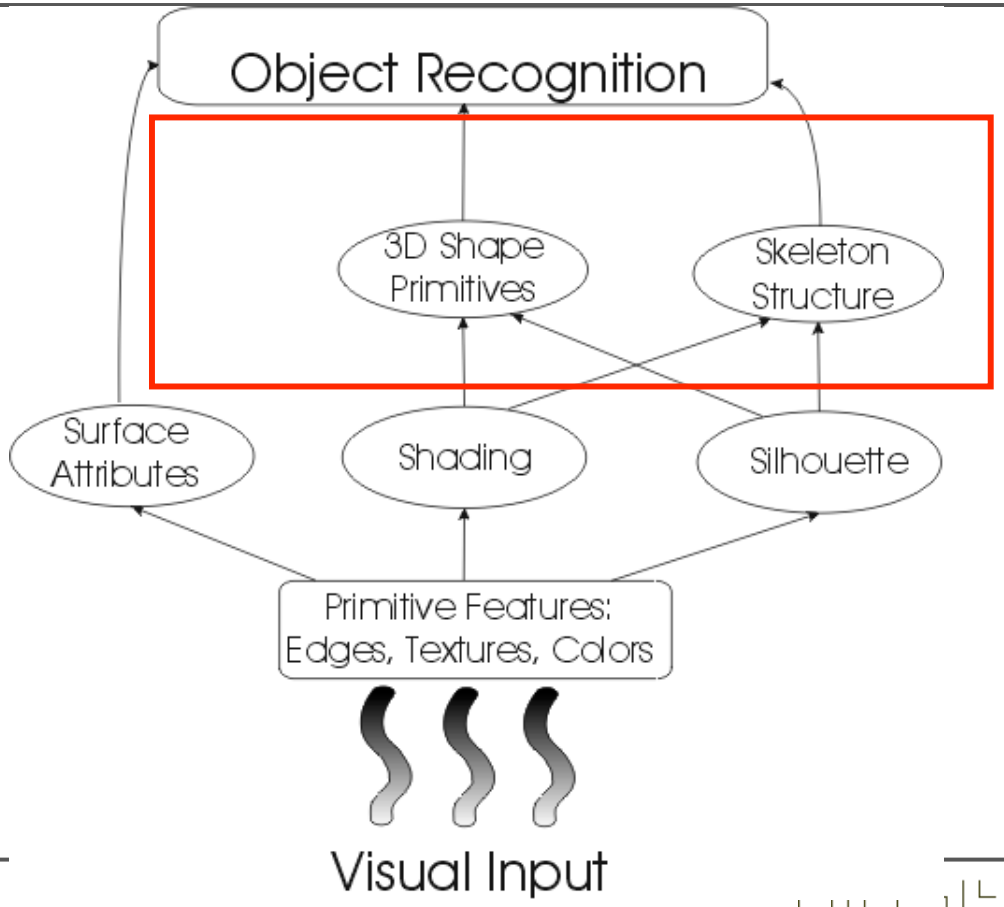
Perception of object structure



We recognise new orientations of familiar objects

Processing stages for recognition

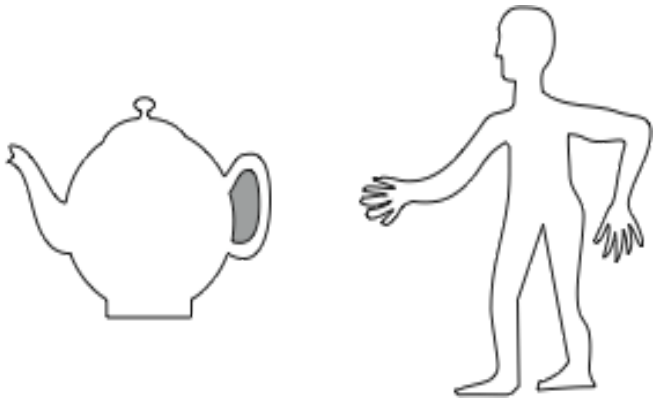
- Biederman's neural network of structural perception

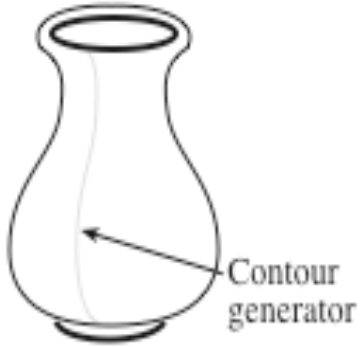


geons

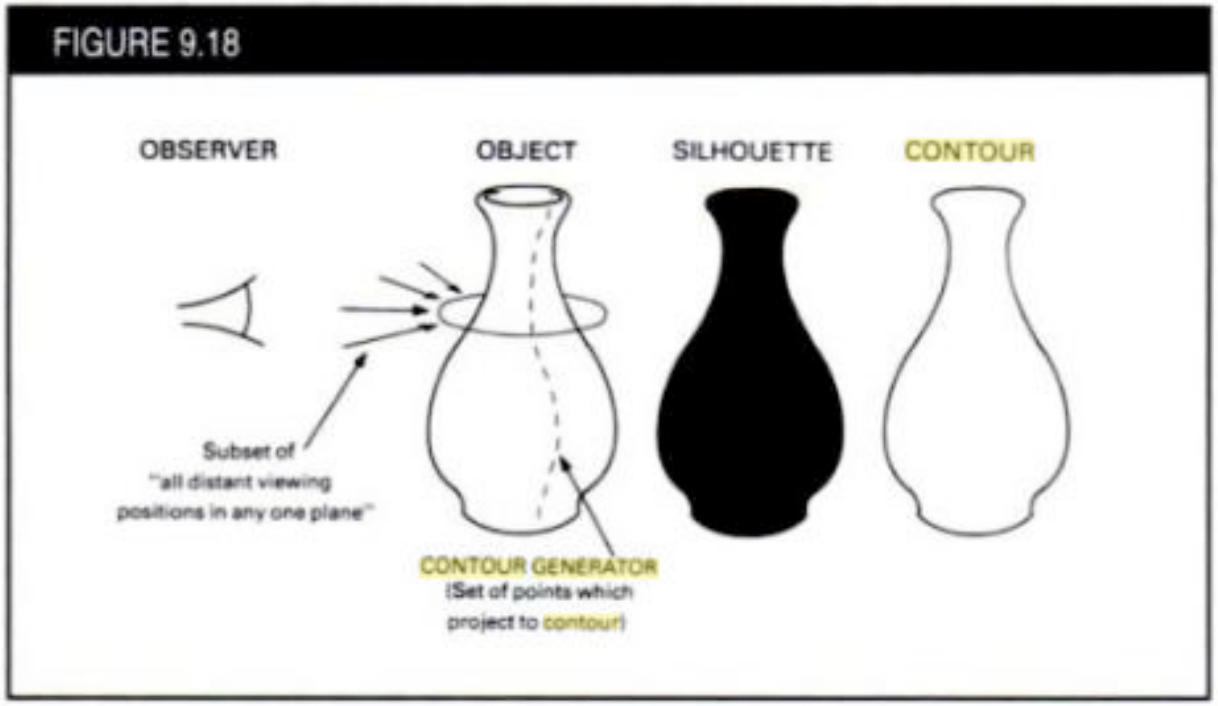
Silhouettes

- Especially important in object perception
  - Cave drawings
  - Children's drawings
- *Canonical* silhouettes
- Contour generator: constraint that determines how silhouette is interpreted as structure (Marr)





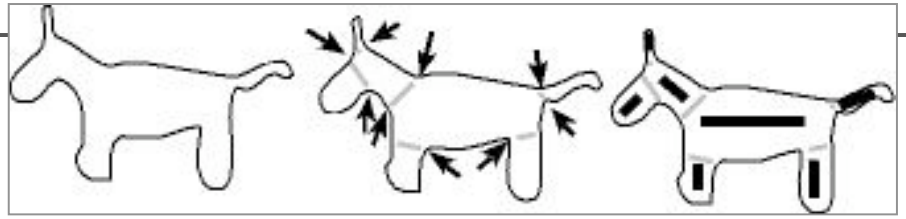
Marr and contours



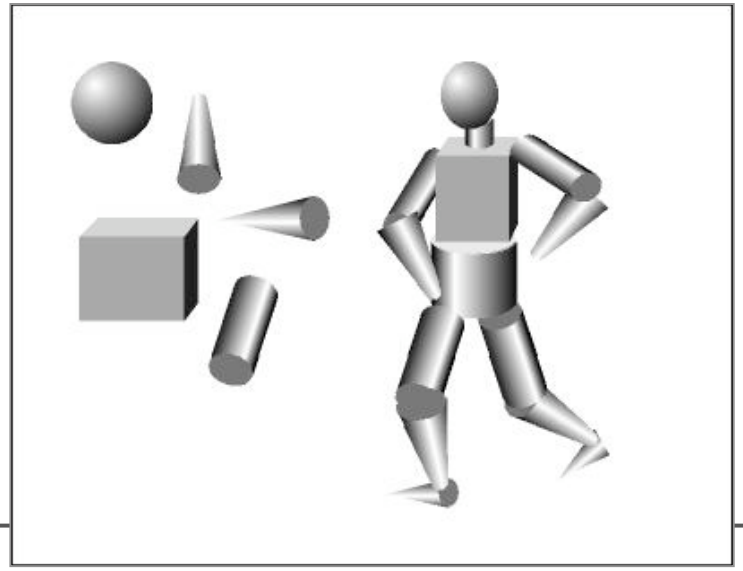
An object, its silhouette, and its contour. The set of points that projects to the contour (the contour generator) is shown. For this figure, all three assumptions (see text) hold for all distant viewing positions in any one plane. Adapted from Marr (1977) and Marr (1982) with permission of The Royal Society.

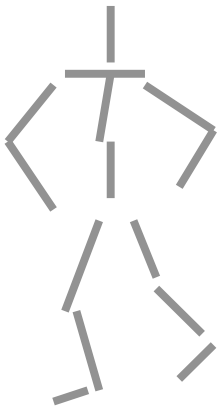
Image from Visual Perception: Physiology, Psychology, & Ecology. Vicki Bruce, Patrick R. Green, Mark A. Georgeson

Geon Theory



Contour information and concave sections define subparts (Marr and Nishihara)



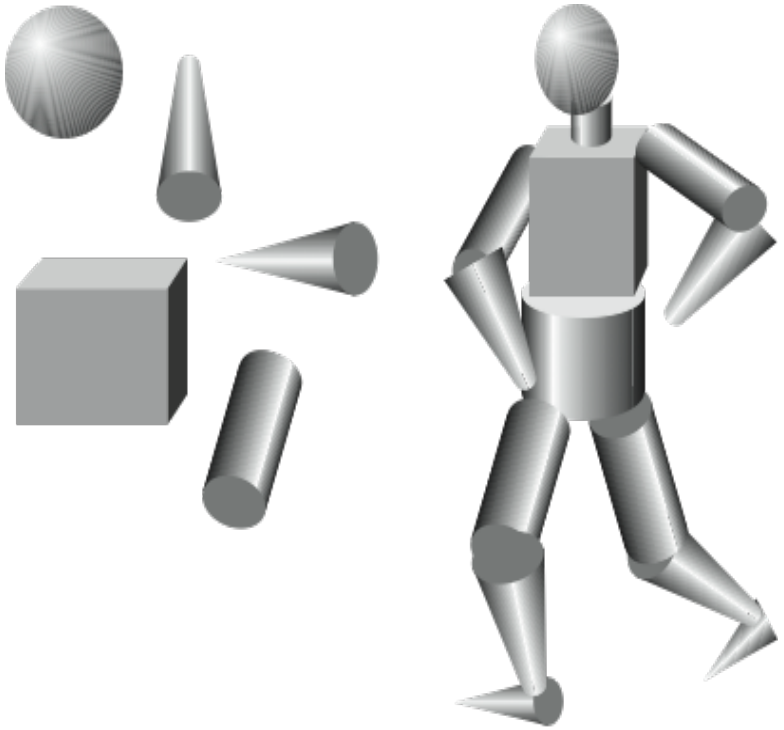


3D Primitives
"Geons"
Structural skeleton

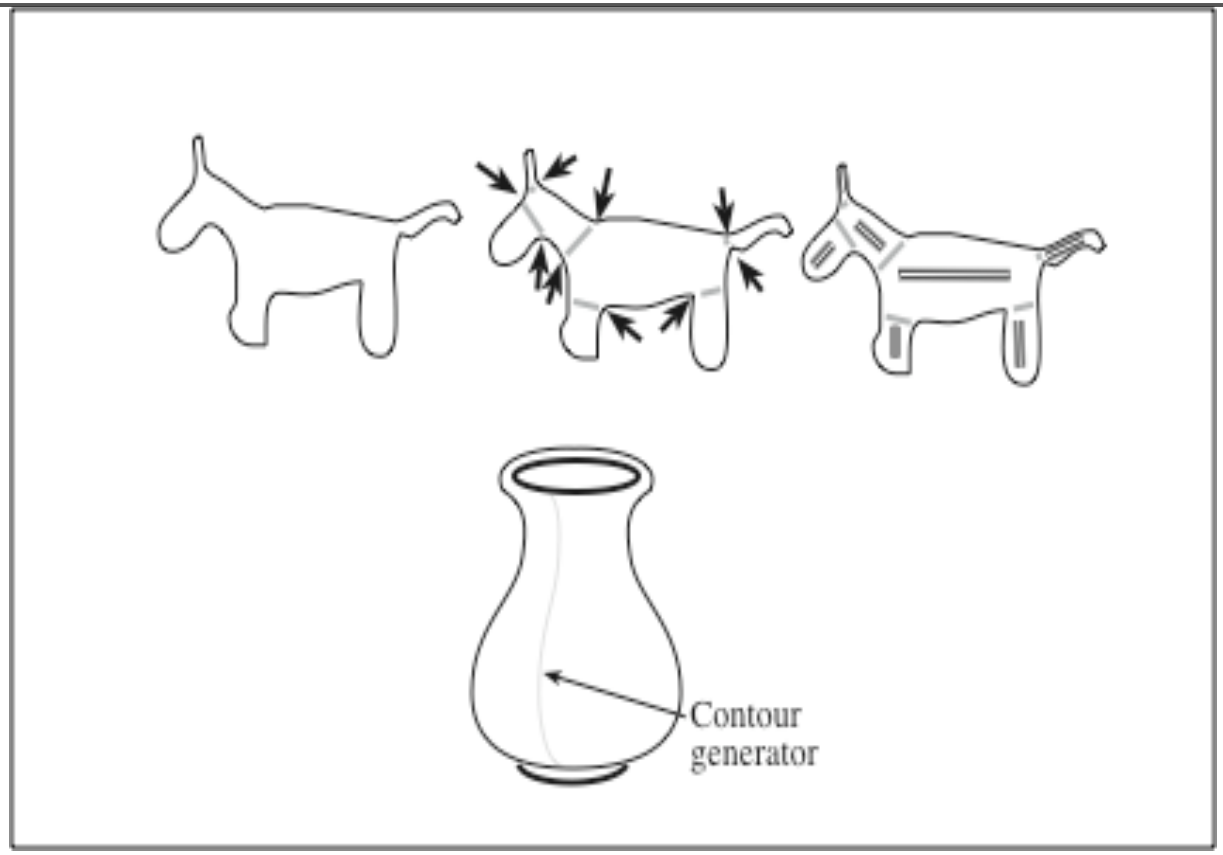
Shape from shading
is also primitive

Geon Theory

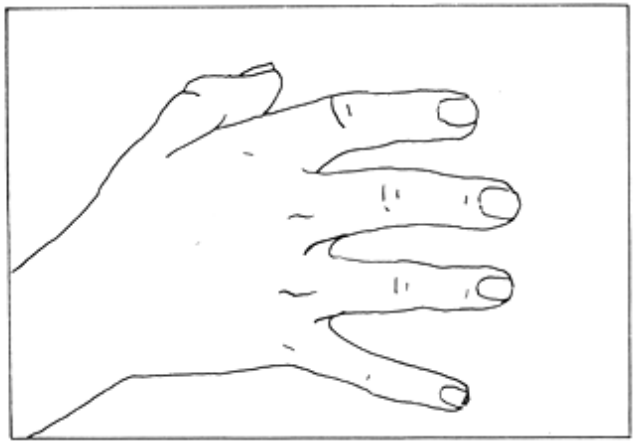
- The whole is a sum of a set of basic primitive geometrical elements (*geons*)
- The way they are connected is also encoded
- Biedermann's *structured object recognition theory*



How Does the Brain Find Geons?



Sometimes simple is better than real



People are faster at recognising streamlined representations of common objects than detailed representations (Ryan and Schwarz, 1956)

Object identification (Ryan and Schwarz)

2 2 3 1



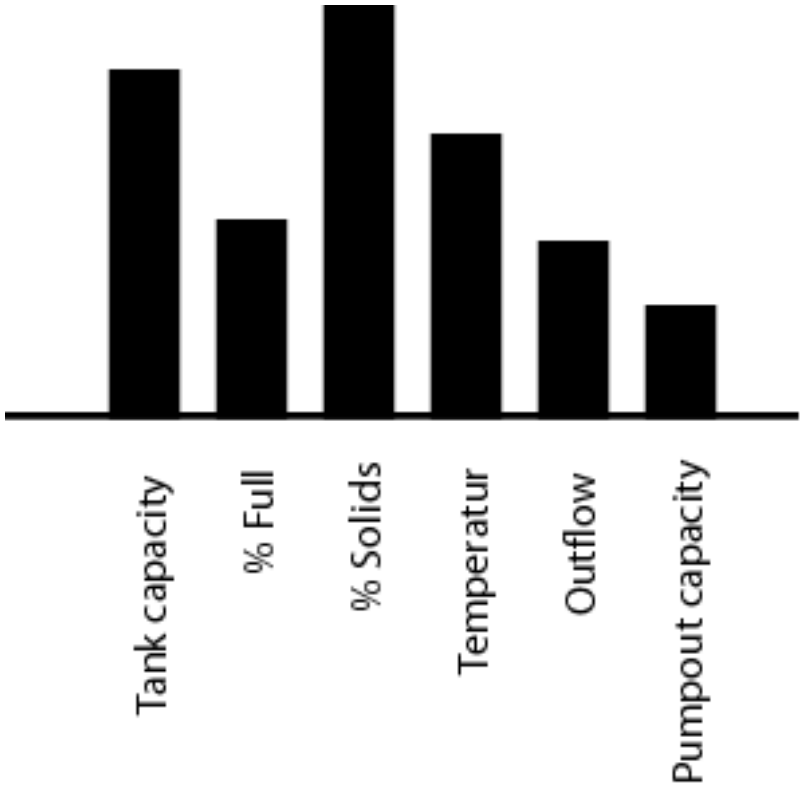

- Cartoon was easiest
- Detailed line drawing was hardest!

- Implication: diagrams can be more effective than photographs
  - Visual “noise” impedes

The Object Display (Wickens)

- Use complex objects to “fuse” variables
- Map entities to object parts
- Map structure to object structure
- Can be metaphorical – an engine + fuel tank
- Map attributes to object attributes
  - color, size, motion etc.

Object Display



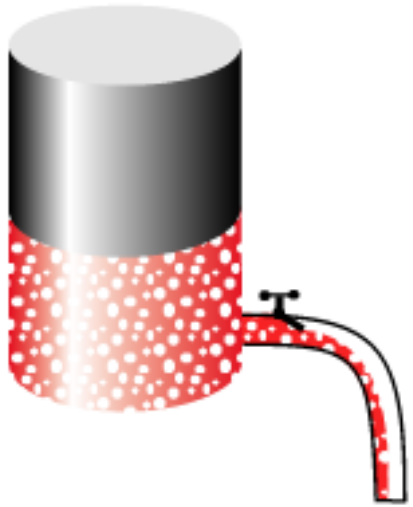
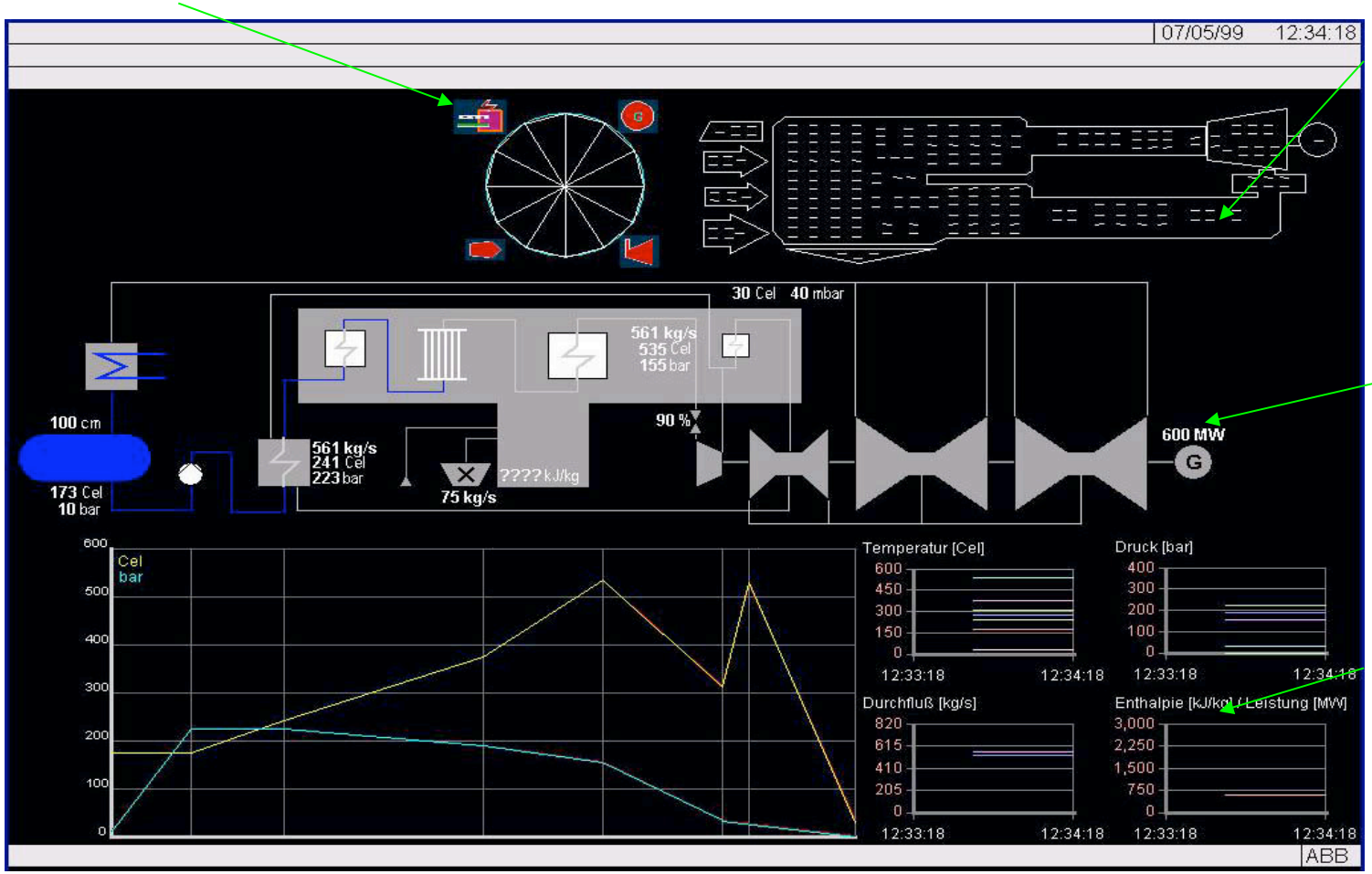


ABB “mimic” displays

Polar Star



Mass Data Display

Plant mimic

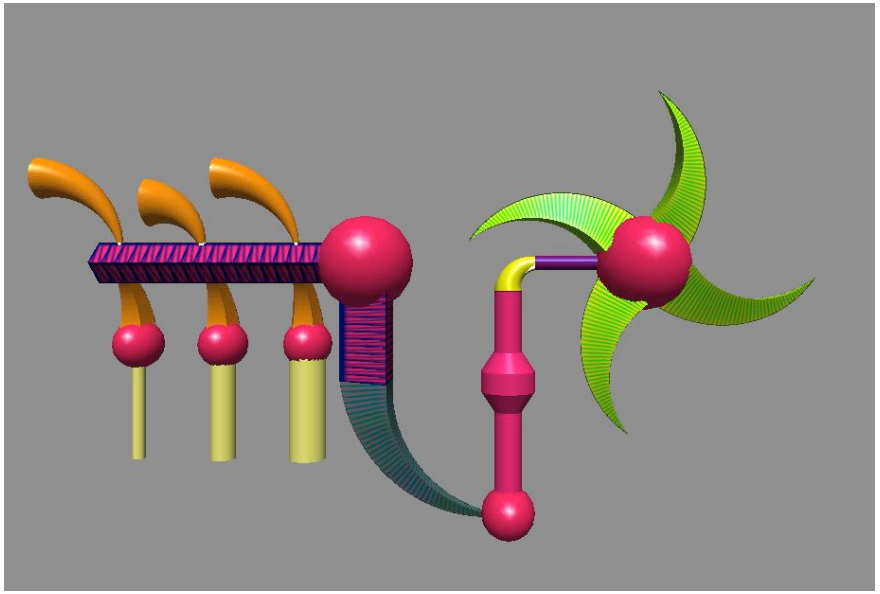
Plant graphs

Object-based display: Chernoff faces

Figure 7.10 Chernoff Faces. Different data variables are mapped to the sizes and shapes of different facial features.

The Geon Diagram Pourang Irani

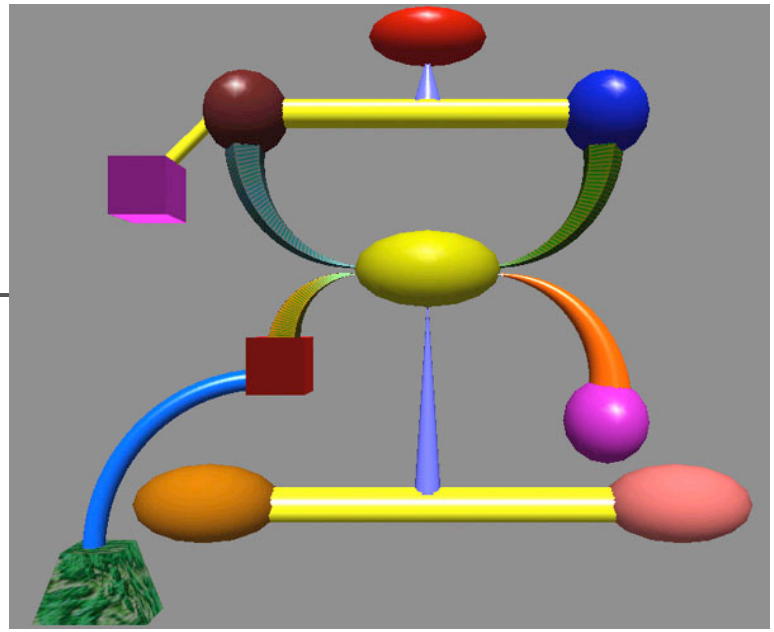
- 3D shape primitives for architecture - entities and relationships
- Surface texture and color for attributes



Pattern finding & Recognition

13% errors: 4.3 sec sub-structure

22% memory errors

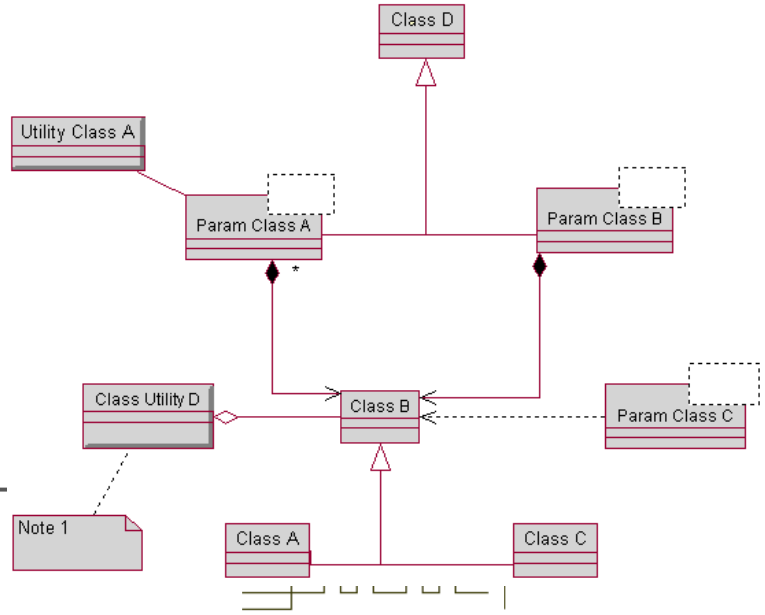


Evaluated geon vs. UML diagrams

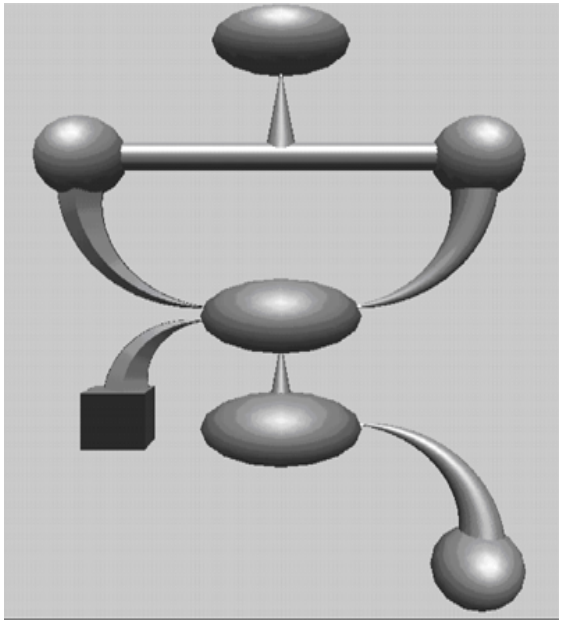
Task was rapid Identification of substructures

26% errors 7.1 sec sub-structure

42% memory errors

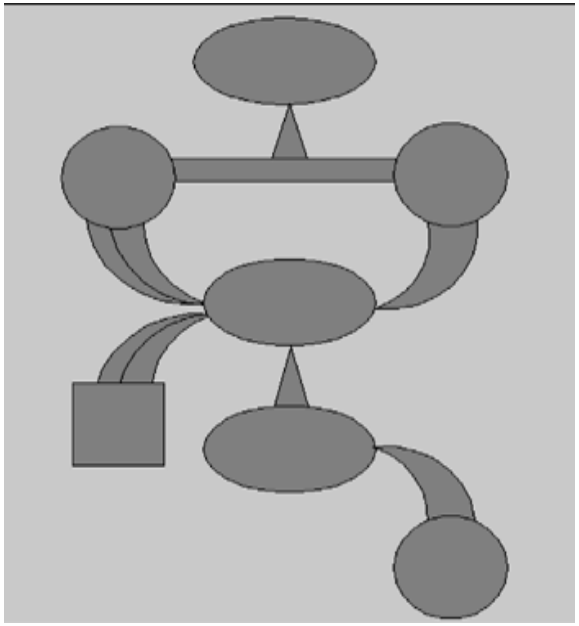


3D versus 2D



11.4% errors 3.7 sec
sub-structure

20% memory
errors



21% errors 5.1 sec
sub-structure

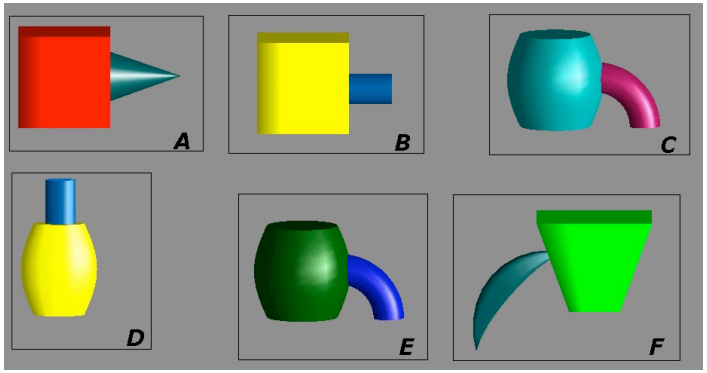
34% memory
errors

Semantics

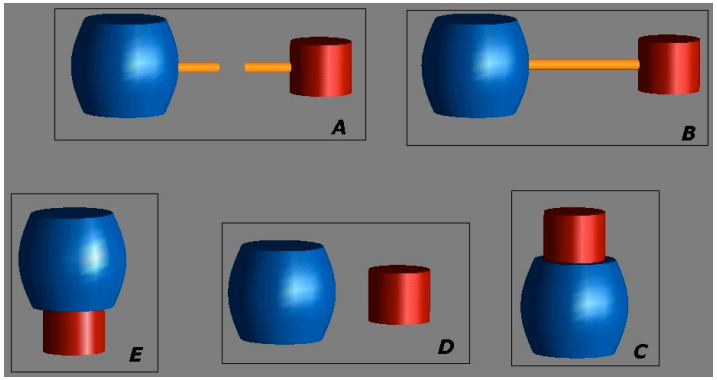
- On top of - relies on
- Underneath – support, foundation
- Inside – containment (private code)
- Attachment points (external interfaces; part_of relationships) more than topology
  - mid
  - Upper
  - Lower

Natural semantics

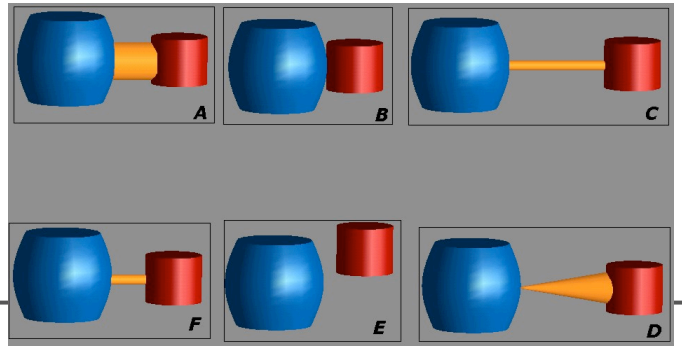
- Instances



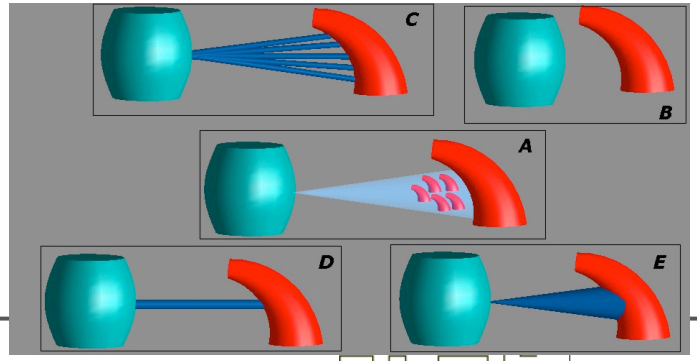
Dependency



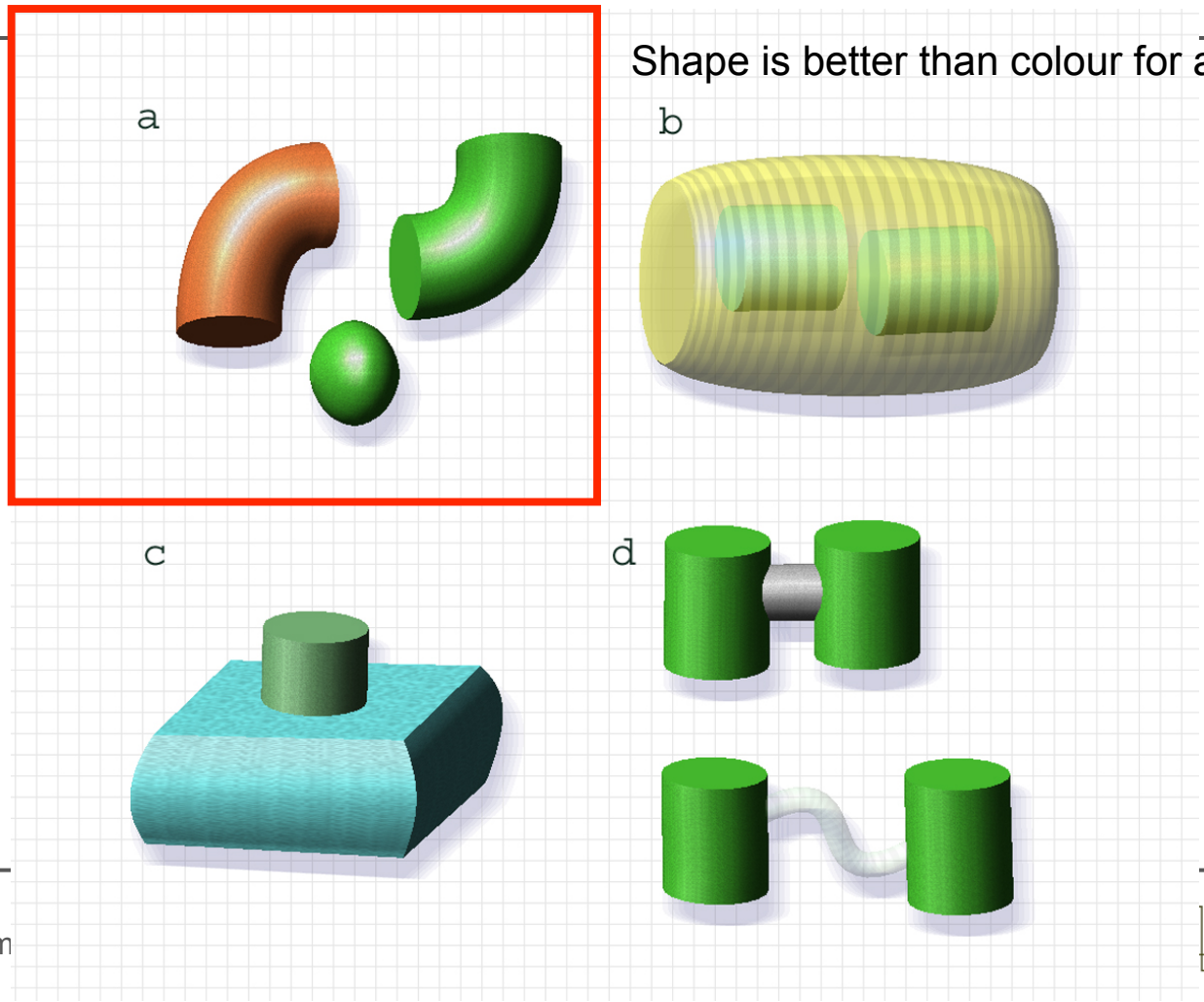
Strength of Relationship

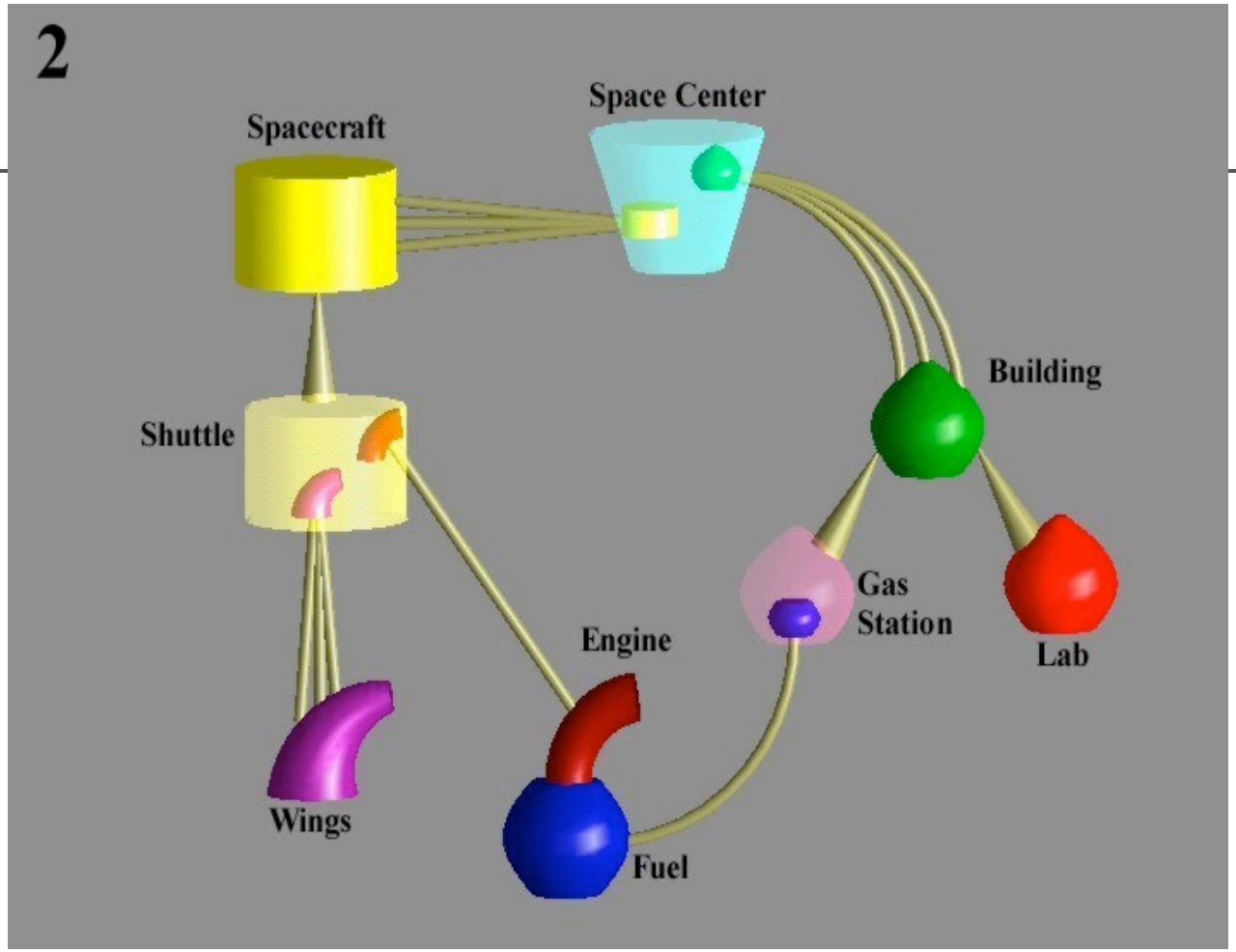


Multiplicity

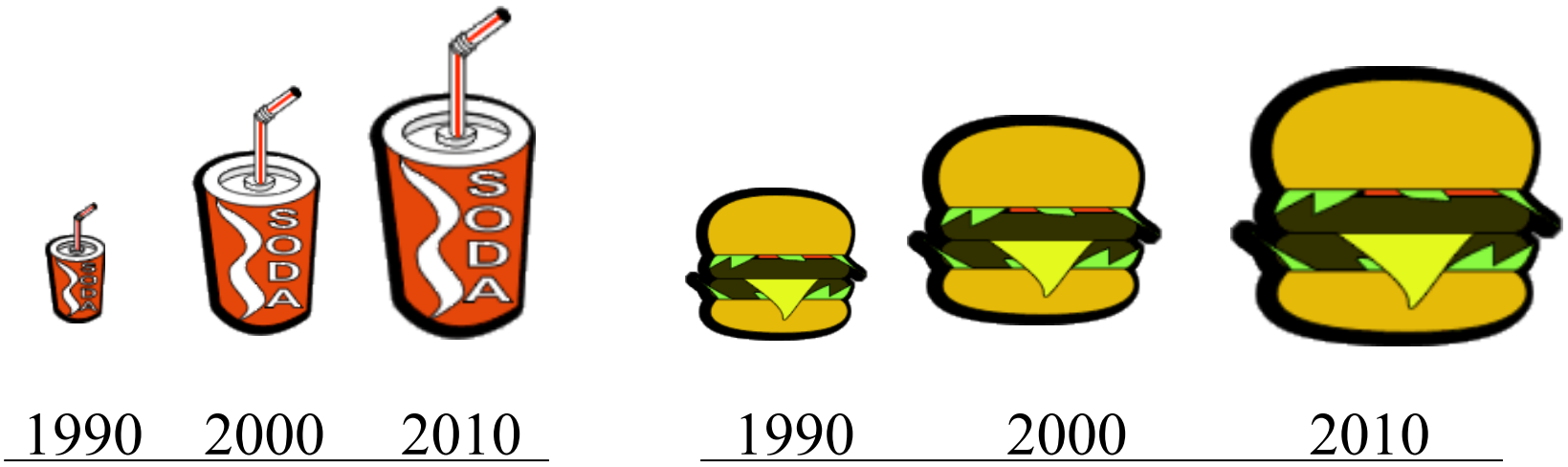


Shape is better than colour for association




2
Space Center
Spacecraft
Building
Shuttle
Gas
Station
Lab
Engine
Wings
Fuel

Infographics or Chart Junk (Tufte) ?



- It's not *all* chart junk

2 ½ D design

- Use 3D objects to represent entities
- Layout to make structure clear in 2D.
- Use canonical views

Surfaces

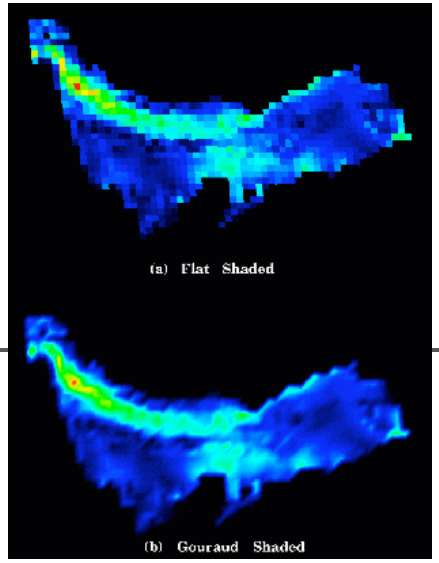
- Continuous surfaces with no discrete shapes or clear boundaries

- Continuous surfaces used in visualizations:
  - digital elevation maps
  - Physical property maps
  - Functional maps

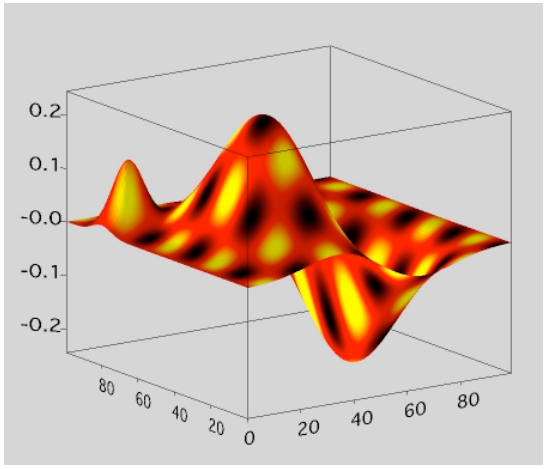
Surface Shape perception

- Univariate maps
- 2D scalar fields
- **Spatial cues** from
  - Surface shading models
  - Surface texture
  - Contours

- Traditional methods
  - Contours (Cartography)
  - Pseudocoloring (chapter 4)

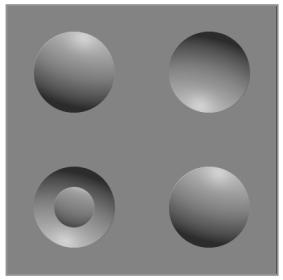


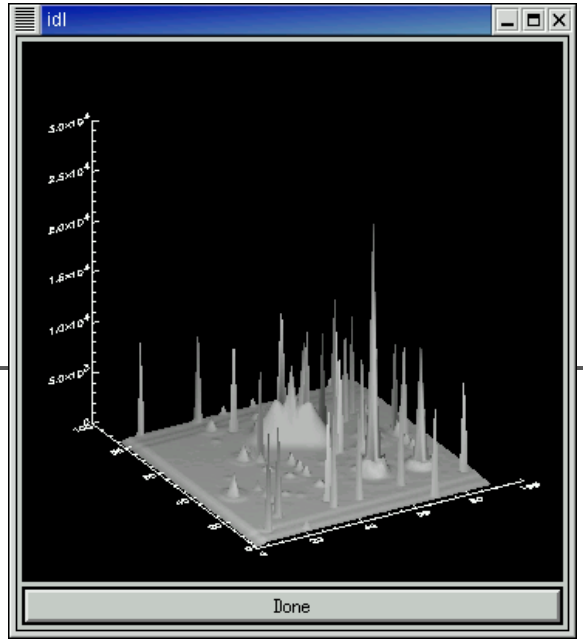
Shading models [www.gri.msstate.edu/.../docs/1995/spmag_06.html]



2D scalar field [www.wavemetrics.com/.../3dandvolume/surface.htm]

Shading

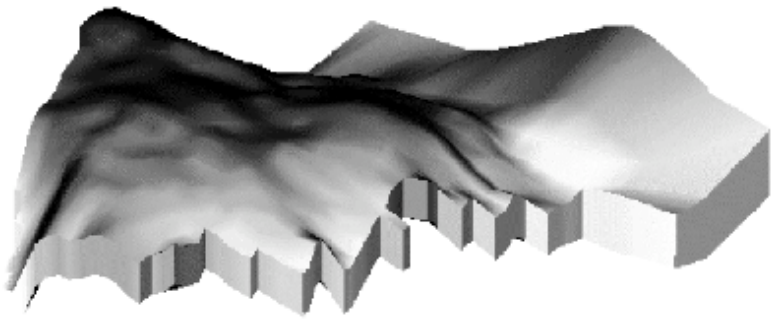




http://mips.as.arizona.edu/MIPS/IDP3/idp3roi.html

- Basic components (from before)
  - Lambertian shading: diffuse reflection
  - Specular shading: glossy highlights
  - Low-contrast texture with linear elements
  - Cast shadows: On itself or another object

- Goal is revealing shape, not realism
  - Visual system assumes a single light source from above
  - Multiple light sources may be confusing
  - Cast shadows inform relative positions



www.scielo.br/img/ fbpe/jbcos/v3n3/04f02.gif

Texture and depth

- Gibson claims that a non-textured surface is just a patch of light
- Shape information comes from texture gradient
- Texture is very important for stereo (see, for example Bair 2007: Vis 2007)
- Untextured polygons produce no internal stereoscopic correspondences
- Stereo correspondences reveal surface shape

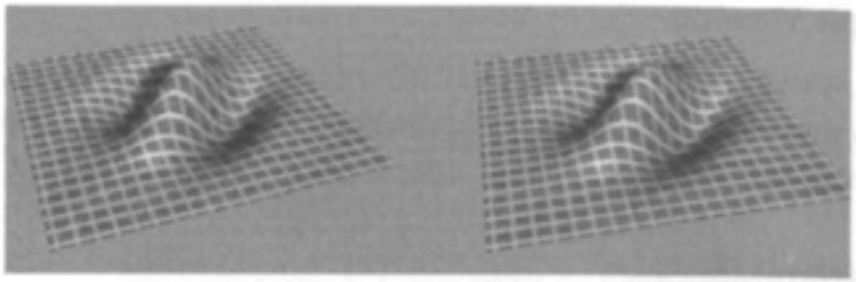
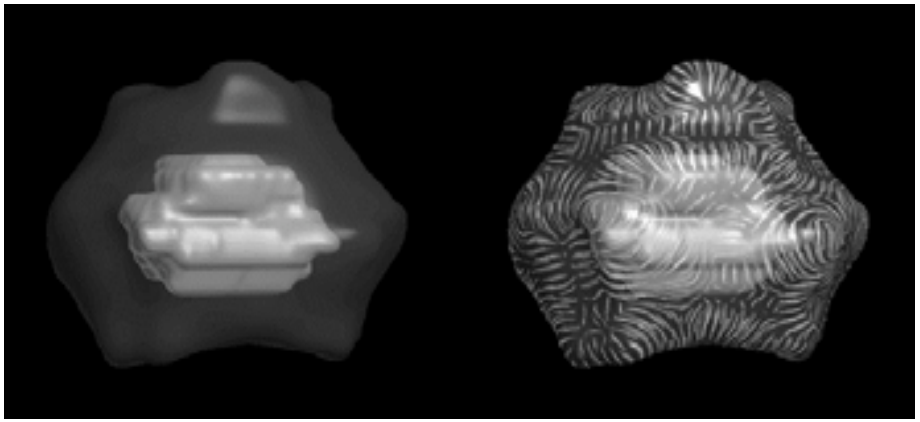


Figure 7.15 A stereo pair showing a textured surface.

Surface texture and transparency

- Without texture, it is “usually” impossible to distinguish one curved transparent surface from another behind it

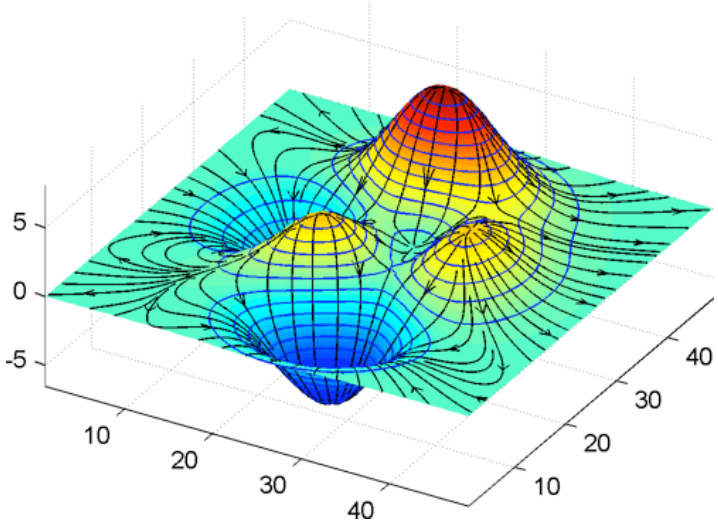


Surface display guidelines

- One light at infinity, from above
  - Lambertian + moderate specular lighting
  - Specular lighting is important to reveal details
  - Specular lighting is local, so enable control over light
- Surfaces should be textured with low-contrast textures that have linear features
  - Linear textures are better than stippled textures for revealng shape
- Cast shadows if they don't interfere: soft edges on the shadows
- Rotation and stereo (and head tracking) helpful

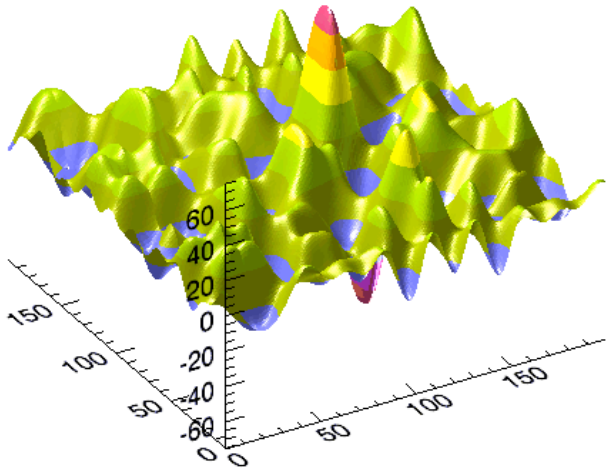
Contours

- A contour is the intersection of a plane with a scalar field
- Internal contours (contour lines)
- External contours (edges or borders of shape)



Contour and shading

- Contours interact with shading to alter perception of similarly shaded surfaces
- Contours are ambiguous wrt *degree and direction of slope*
  - Use shading
- Shading is not good at *gradient* information
  - Use contours



The concrete/abstraction tradeoff

- Objects lend themselves to overly literal interpretation
- More abstract sketches leave more room for interpretation (better support the creative process)
- Information discovery may be better supported by abstractions because of bias to structure
- Object displays must be custom designed

Pictures and Words

- When should we use a visual display?
- What is a visual language?
- Dual coding theory
- How to integrate images and words

Hieroglyphs gave way to more abstract symbols

- Why turn back the clock?

$$x \propto \int_{1}^{\infty} \varpi \oint_{\lambda} \left\| \frac{\prod \lambda_i}{\oint \Psi} \right\|$$

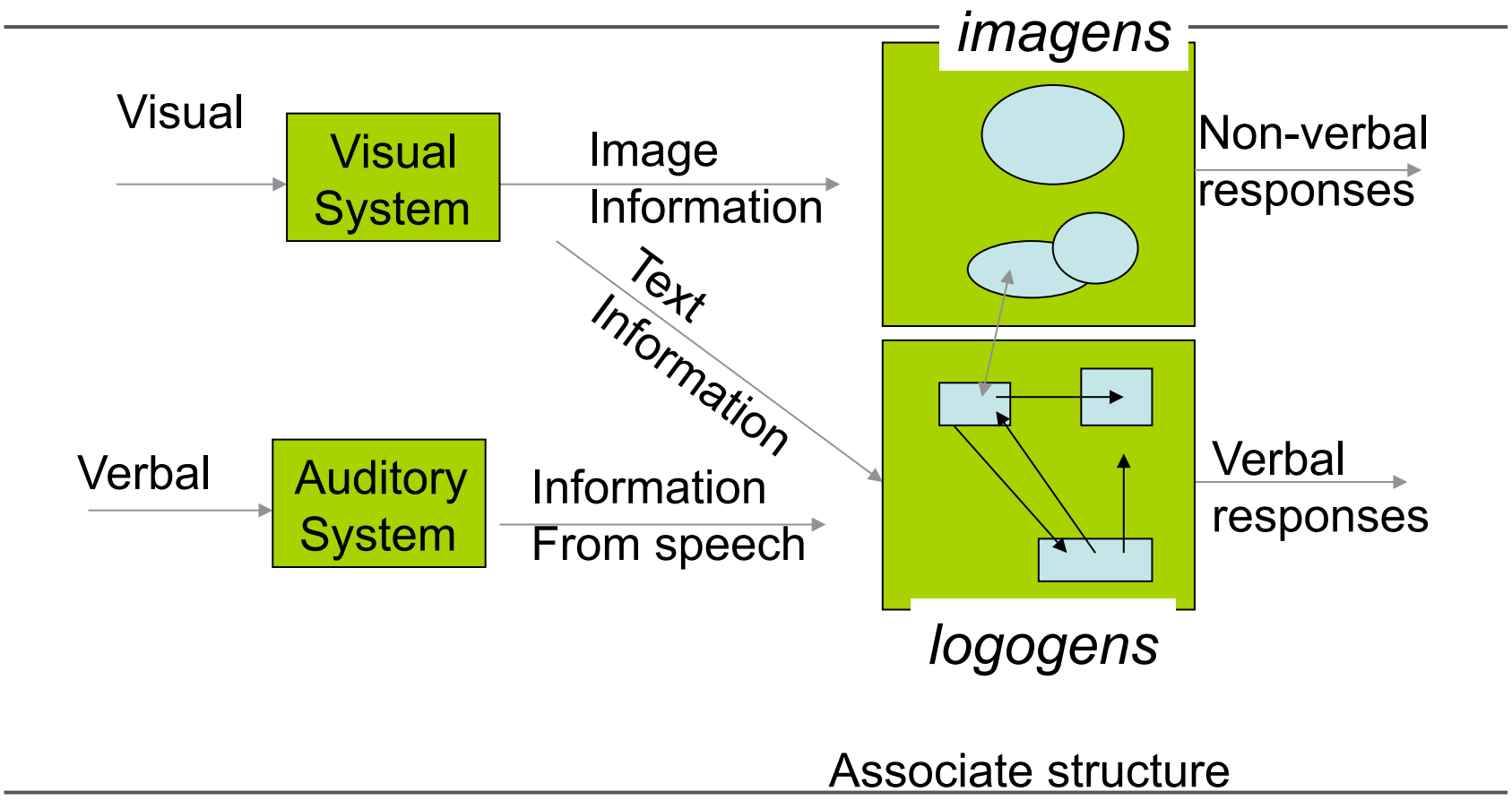
-17K years

Coding words and images

Bertin distinguishes 2 sign systems:

- Auditory information
  - Mathematical symbols, natural language, music
- Visual information processing
  - Graphics
  - Abstract and figurative imagery

Pavio's dual coding theory (1987)



Coding words and images

- (visual) *imagens* consist of:
  - Objects, natural groupings of objects, and parts of objects
  - Spatial layout
- (verbal) *Logogens* store basic information pertaining to language
  - But NOT the sounds of the words
  - Processed by reading, speech, logical thought subsystems

- Logogen and imagen systems are separate but can be strongly linked
  - E.g. “cat”, language-based concepts around cat linked to appearance of cats and their environment

Coding words and images

- Different neural processing centres
  - verbal information (speech areas of temporal cortex)
  - Visual information (the visual cortex)
- “visual thinking” idea is relatively new
- Evidence from mental imaging

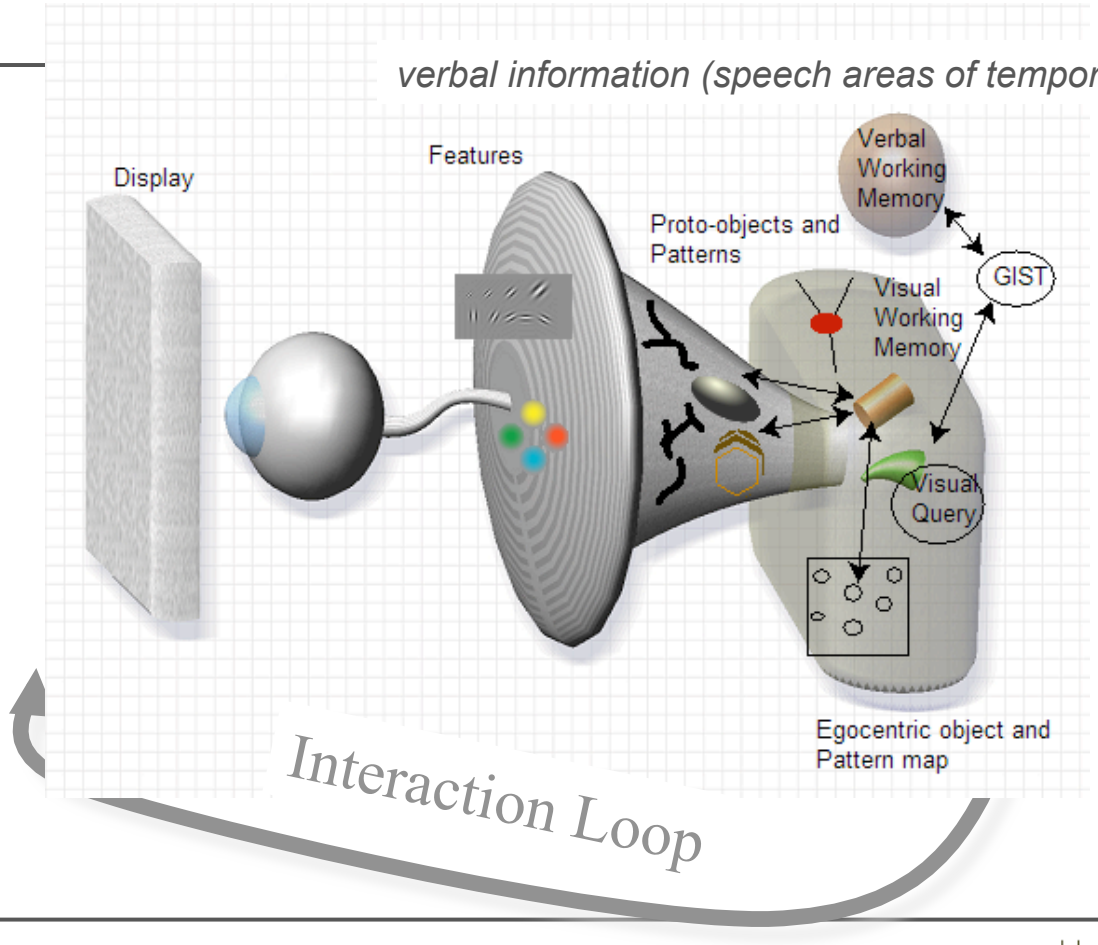
Coding words and images (Kosslyn)

- when people are asked to compare visual attributes such as size and color, they claim to use mental imagery

- People treat objects in mental images as if they have “real” sizes and locations in space

Architecture for visual thinking

10 billion neurons
Parallel, automatic



Visual information (the visual cortex)

Coding words and images

- Seeing a cow and mentally visualising a cow activate/ excite the same neural pathways (at least partly)

- Modern visual memory theory takes the position that *visual object processing* and *visual object recognition* are part of the same process

Coding words and images

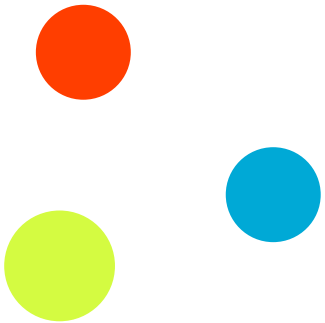
- Visual memory “traces” of objects and scenes are stored as part of the processing mechanism
  - Object does not need to be fully processed for recognition to occur
- This explains why recognition is so much more powerful than recall
  - Easier to recognise than reproduce

Capacity of visual working memory
(Vogal, Woodman, Luck, 2001)

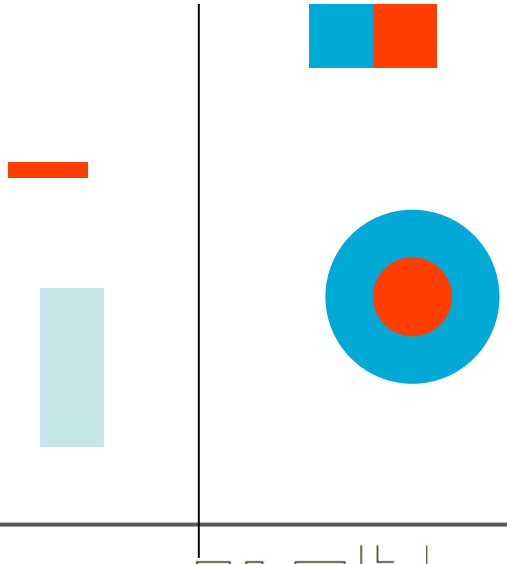
- Task – change detection
- Can see 3.3 objects
- Each object can be complex

1 second

- Change blindness



color, size, texture

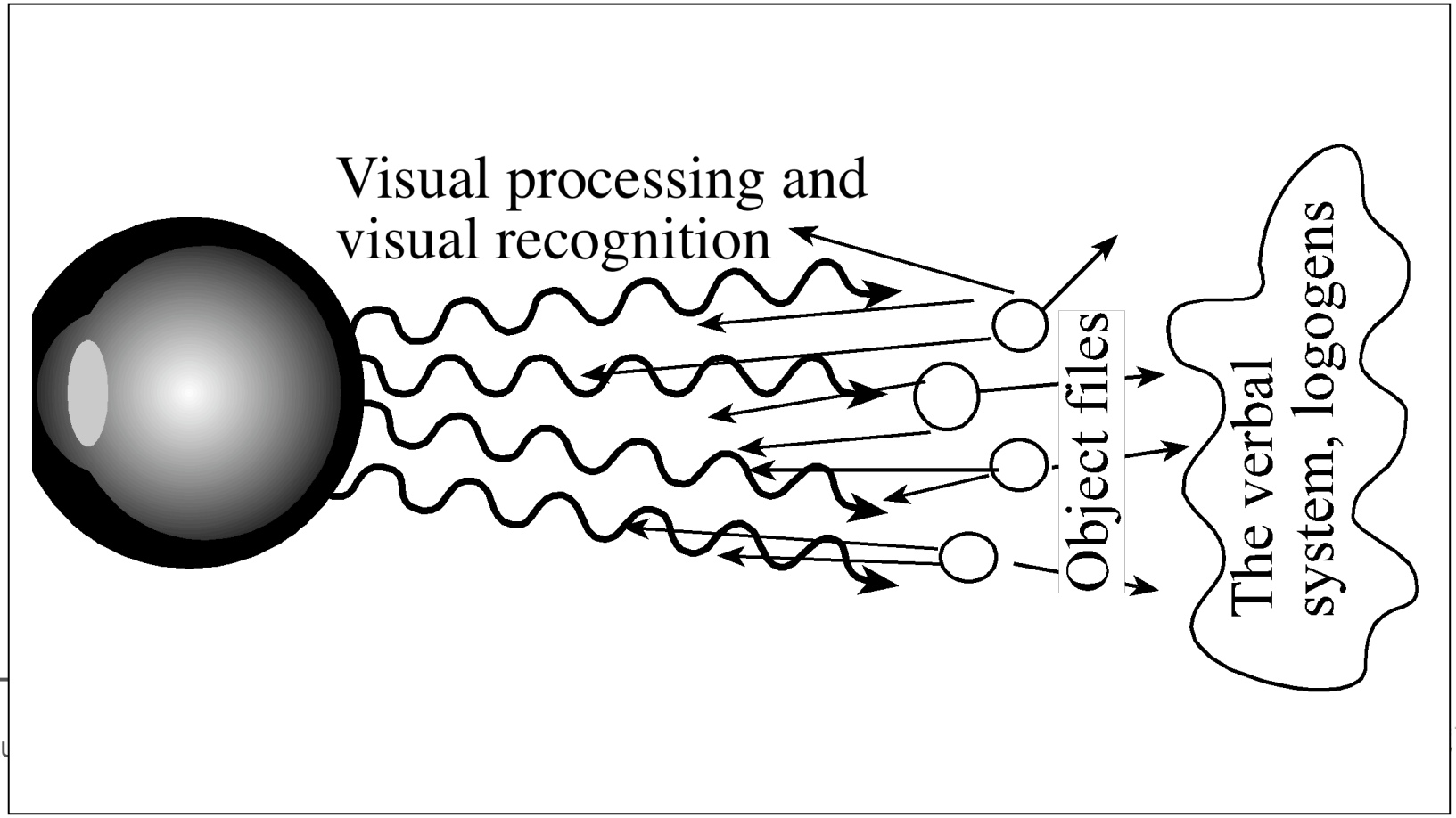


Capacity of verbal working memory

- Used to be though of a 7 +/- 2
- It is now thought of as more a duration of proto-verbal codes.

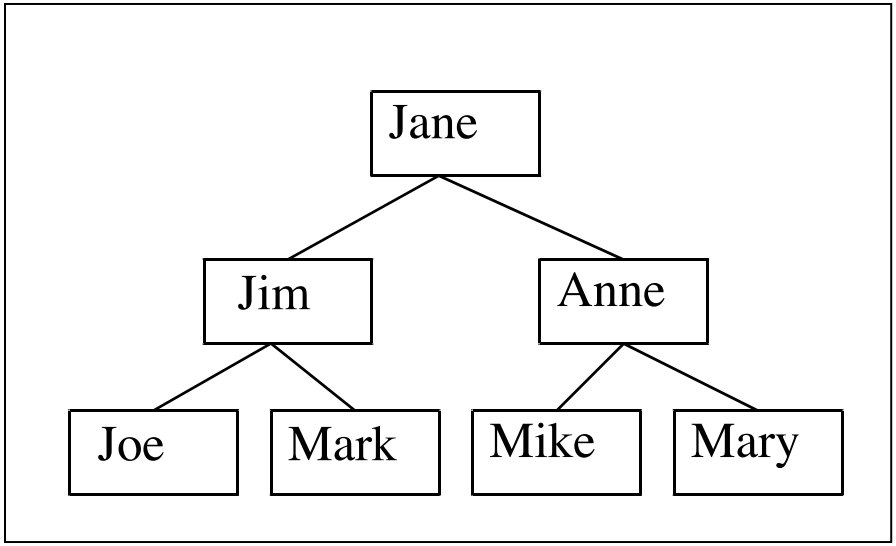
The visual system gives us

Rapid recognition and pattern finding



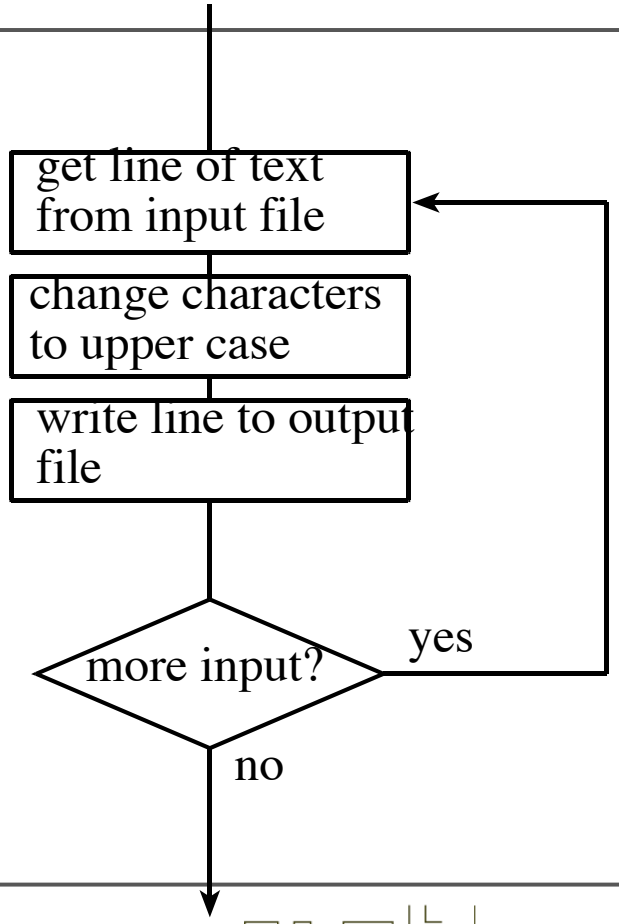
Abstraction Pattern

- Jane is Jim's boss
- Jim is Joe's boss
- Anne works for Jane
- Mark works for Jim
- Anne is Mary's boss
- Anne is Mike's boss



Visual and verbal pseudo-code

- While letters in stack
  - Take a letter
  - Put a stamp on it
  - Put it in the ‘out tray’
- We understand the simple diagram of flow
- More complex abstractions and flow are hard



Images vs. words

- Greatest advantage: words and language are ubiquitous
- Most elaborate, complete and share system of symbols available
- Visual techniques should be used only when there is a clear advantage
- One very simple example: the symbol for “Help” ?

Images vs. words

- Words are good for:
  - Framework and narrative of an extended communication
  - Detailed structure and elaboration of relationships in that structure
  - Procedural information, logic
  - abstract concepts
  - Conditions under which something should be done or not done (logical constraints)

Images vs. words

- Images are good for:
  - Spatial structures, location
  - Structural relationships
  - Detail and appearance (needs time)
  - Relative position and size
  - (physical constraints related to space)

Animated Images vs. words

- Animated images open up whole new range of possibility
  - Causality (single greatest enhancement)
  - Communication
  - Transformation over time
  - Sequence of data movements (Sorting out Sorting)
  - Complex spatial actions (perceptual-motor tasks)
    - Requires verbal instruction to accompany

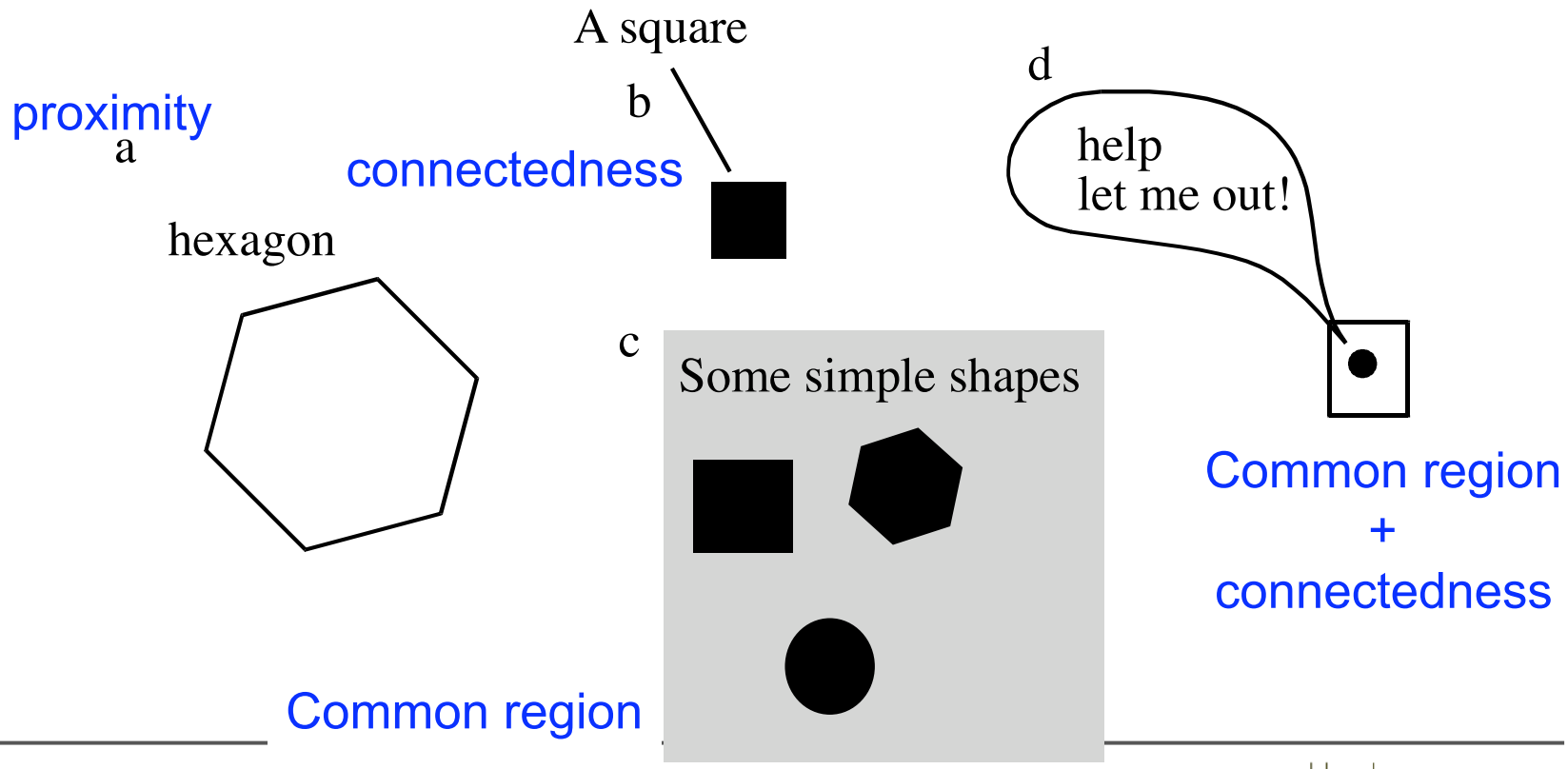
Linking images and words

- Central claim of multimedia is that more than one source (medium) of communication is better
- Theory: if active processing takes place in more than one subsystem learning will be better
- Dual coding is better than single modality coding

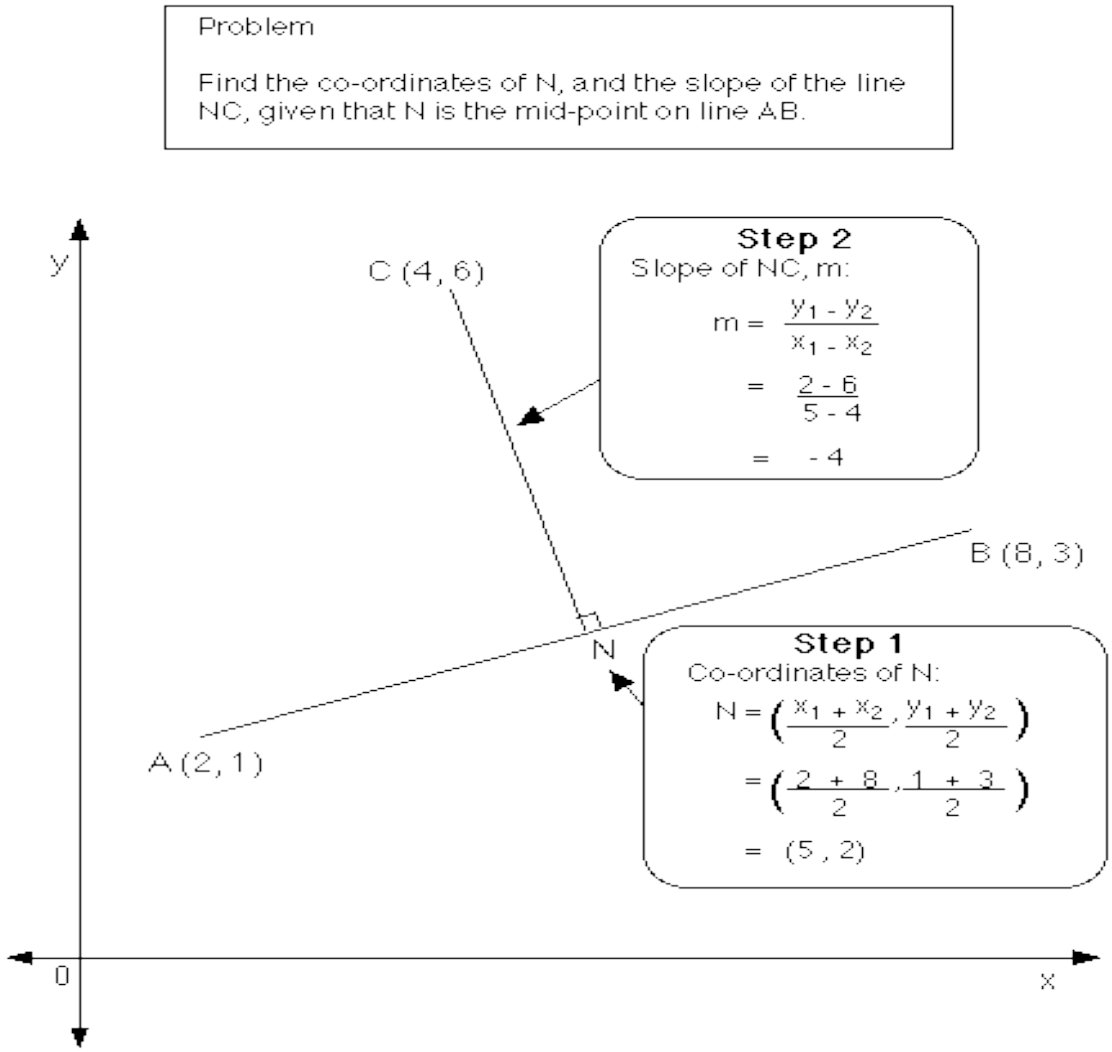
Linking images and words

- Links are critical:
  - Visual and verbal information have to be actively constructed together
  - association for cross-linking requires careful design

Attaching words to images
- use gestalt laws

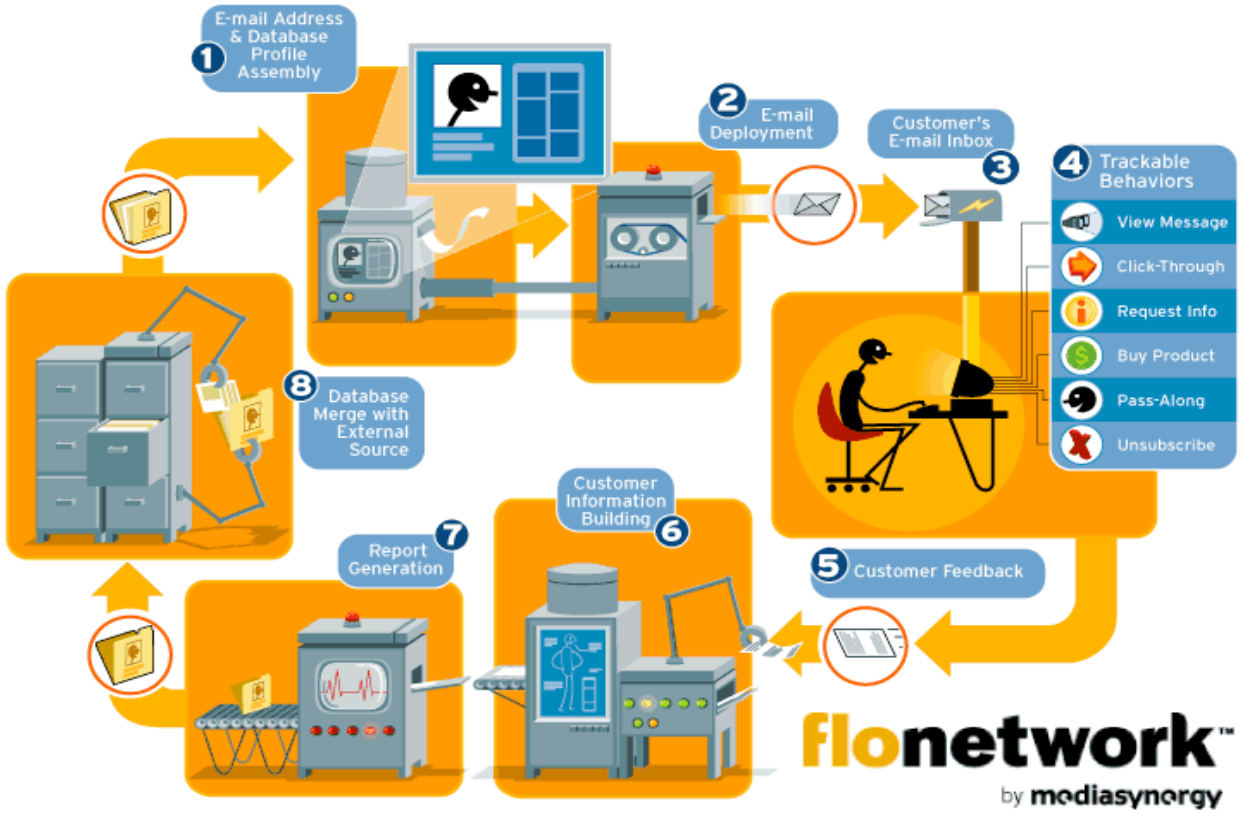


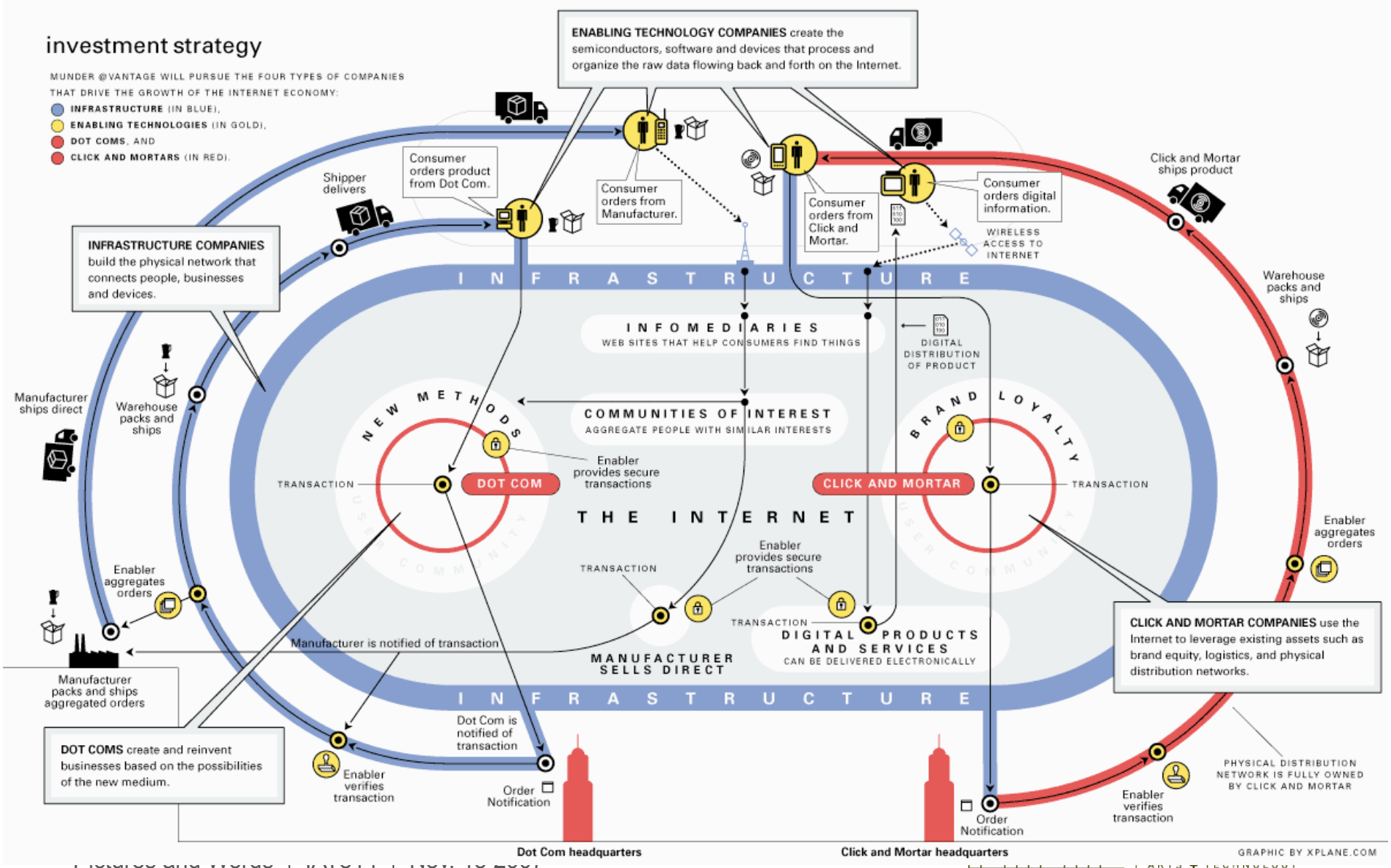
Integrated pictures and words more effective: Chandler and Sweller 1991



XPlane

BEFORE




investment strategy
MUNDER @VANTAGE WILL PURSUE THE FOUR TYPES OF COMPANIES THAT DRIVE THE GROWTH OF THE INTERNET ECONOMY:
INFRASTRUCTURE (IN BLUE),
ENABLING TECHNOLOGIES (IN GOLD),
DOT COMS, AND
CLICK AND MORTARS (IN RED).
ENABLING TECHNOLOGY COMPANIES create the semiconductors, software and devices that process and organize the raw data flowing back and forth on the Internet.
INFRASTRUCTURE COMPANIES build the physical network that connects people, businesses and devices.
Shipper delivers
Consumer orders product from Dot Com.
Consumer orders from Manufacturer.
Consumer orders from Click and Mortar.
Consumer orders digital information.
WIRELESS ACCESS TO INTERNET
Click and Mortar ships product
Warehouse packs and ships
INFRASTRUCTURE
INFOMEDIARIES
WEB SITES THAT HELP CONSUMERS FIND THINGS
DIGITAL DISTRIBUTION OF PRODUCT
Manufacturer ships direct
Warehouse packs and ships
NEW METHODS
COMMUNITIES OF INTEREST
AGGREGATE PEOPLE WITH SIMILAR INTERESTS
BRAND LOYALTY
TRANSACTION
DOT COM
Enabler provides secure transactions
CLICK AND MORTAR
TRANSACTION
THE INTERNET
Enabler aggregates orders
Enabler provides secure transactions
TRANSACTION
TRANSACTION
DIGITAL PRODUCTS AND SERVICES
CAN BE DELIVERED ELECTRONICALLY
Enabler aggregates orders
Manufacturer is notified of transaction
MANUFACTURER SELLS DIRECT
CLICK AND MORTAR COMPANIES use the Internet to leverage existing assets such as brand equity, logistics, and physical distribution networks.
Manufacturer packs and ships aggregated orders
INFRASTRUCTURE
Dot Com is notified of transaction
DOT COMS create and reinvent businesses based on the possibilities of the new medium.
Enabler verifies transaction
Order Notification
Enabler verifies transaction
Order Notification
PHYSICAL DISTRIBUTION NETWORK IS FULLY OWNED BY CLICK AND MORTAR
Dot Com headquarters
Click and Mortar headquarters
GRAPHIC BY XPLANE.COM

Recap

- Words for procedure, logic and abstract
- meanings/ images for pattern and structure

The nature of language

- Chomsky, innate deep structures
  - Syntactic structures generalise across cultures.
  - Cornerstone of computer languages
- Basis of modern linguistics
- But being verbal is not essential to language development
- Sign languages for the deaf are the most perfect examples of visual language

The nature of language

- Critical period for language development from birth to 10
- Most easily acquired from 0-4
- Need to obtain fluency in some language early on for fluency in any language to be possible
- Sign language needs to be acquired early in life

Language is dynamic

- We take in spoken, written, and sign language **serially**; it can take a few seconds to hear or read a short sentence.

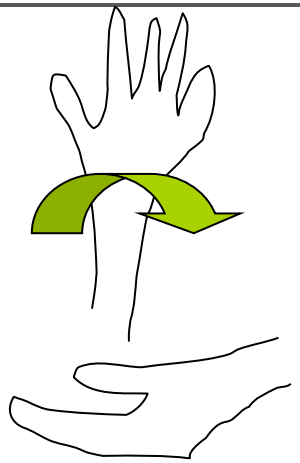
- • In contrast with the dynamic, temporally ordered nature of language, relatively large sections of static pictures and diagrams can be understood **in parallel.**

- We can comprehend a complex visual structure in a fraction of a second, based on a single glance.

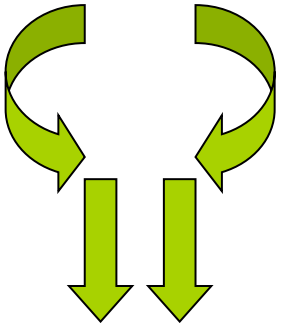
What is language

- Description
- Communication of intention
- The ability to communicate procedures and sequences of operations – including logic – if, but, causes, do ***a*** then ***b*** then ***c***
- ***Thus far we have only dealt with description***

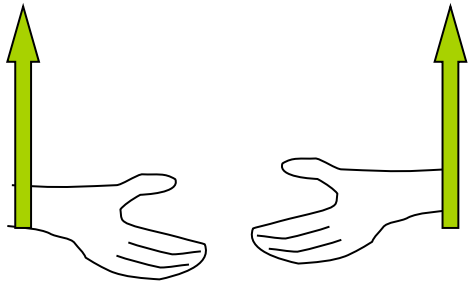
Sign languages



ASL



Danish SL



Chinese SL

Arose spontaneously

Are not related to verbal languages

Have syntax

Become more abstract

NSL

Goldin-Meadow

To be fluent in visual language we should be trained from early in life

Created visual languages

- The difficulty of writing and understanding computer programs has led to the development of a number of so-called visual languages in the hope that these can make the task easier.
  - Based on the notion that visual representation of abstraction is easier to process because of ease of processing structure

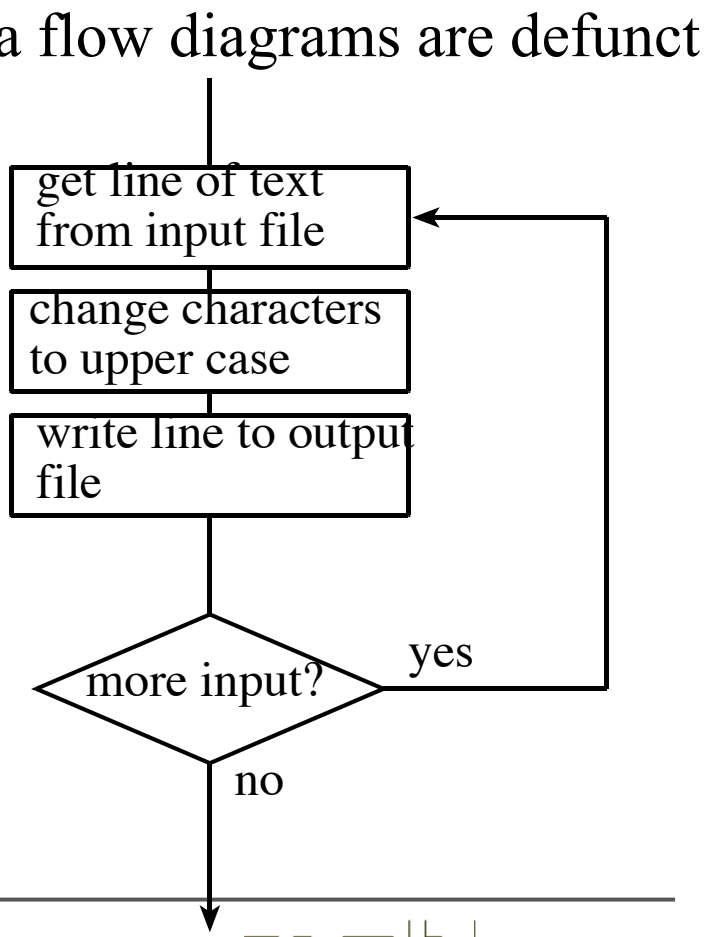
- • But we must be very careful in discussing these as languages.

- • Visual programming languages are mostly static diagramming systems, so different from spoken languages that using the word language for both can be more misleading than helpful.

Visual and verbal pseudo-code

- While letters in stack
  - Take a letter
  - Put a stamp on it
  - Put it in the ‘out tray’

Visual programming languages have a history of failure



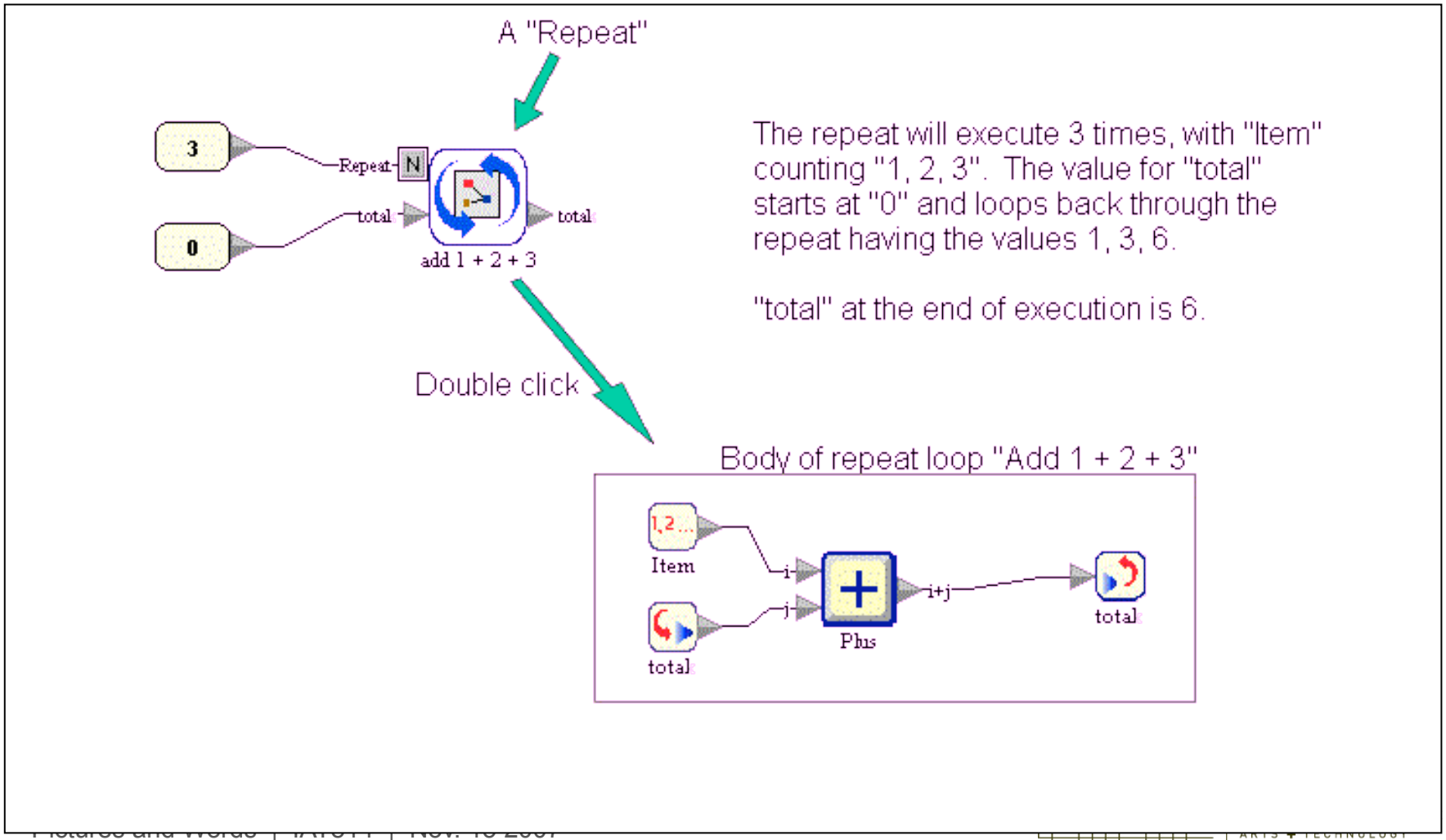
Examples of visual languages

- Sanscrit
- Petri-nets
- Khoros

Sanscrit

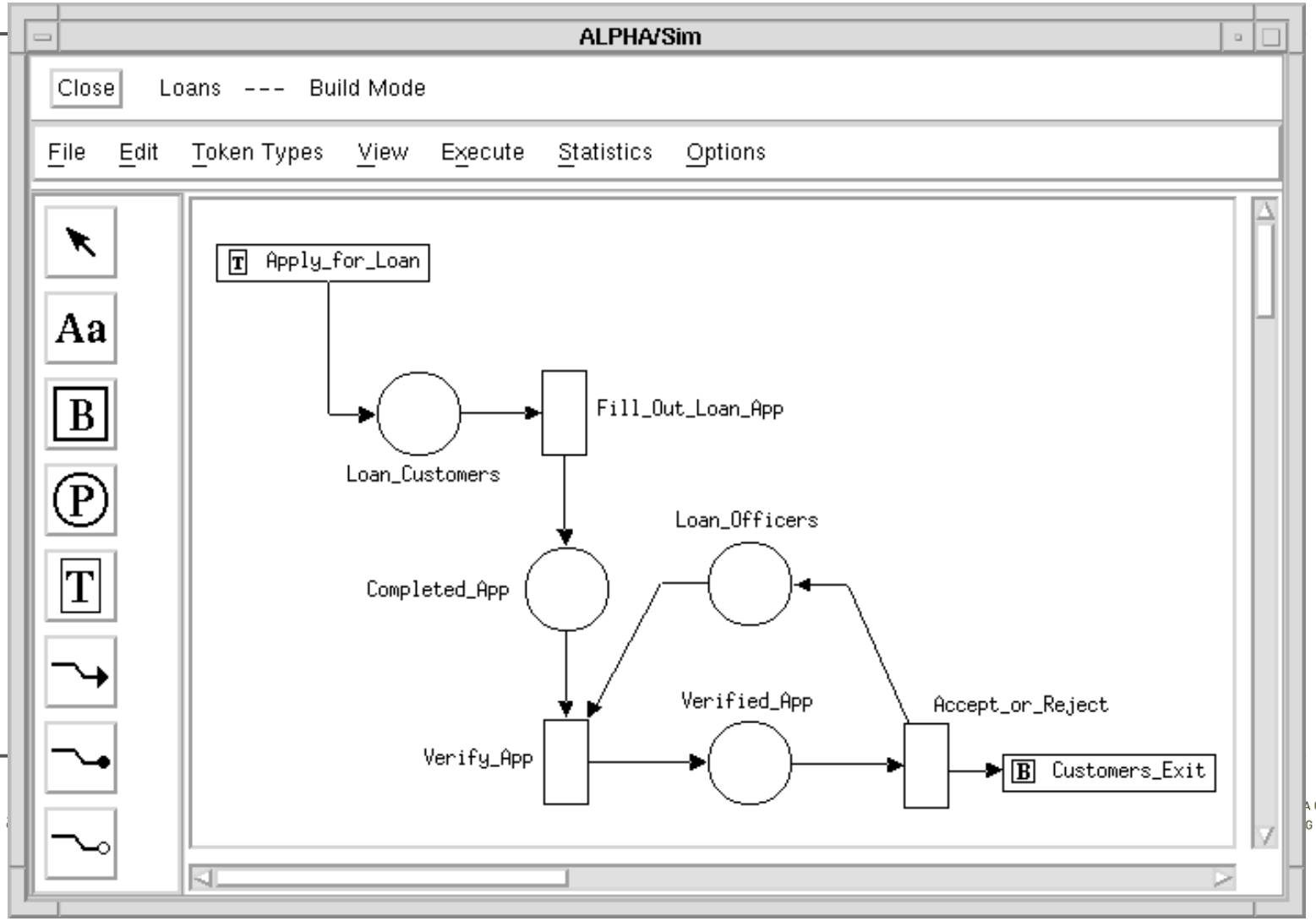
Count from 1 to 3

for i = 1 to 3 do



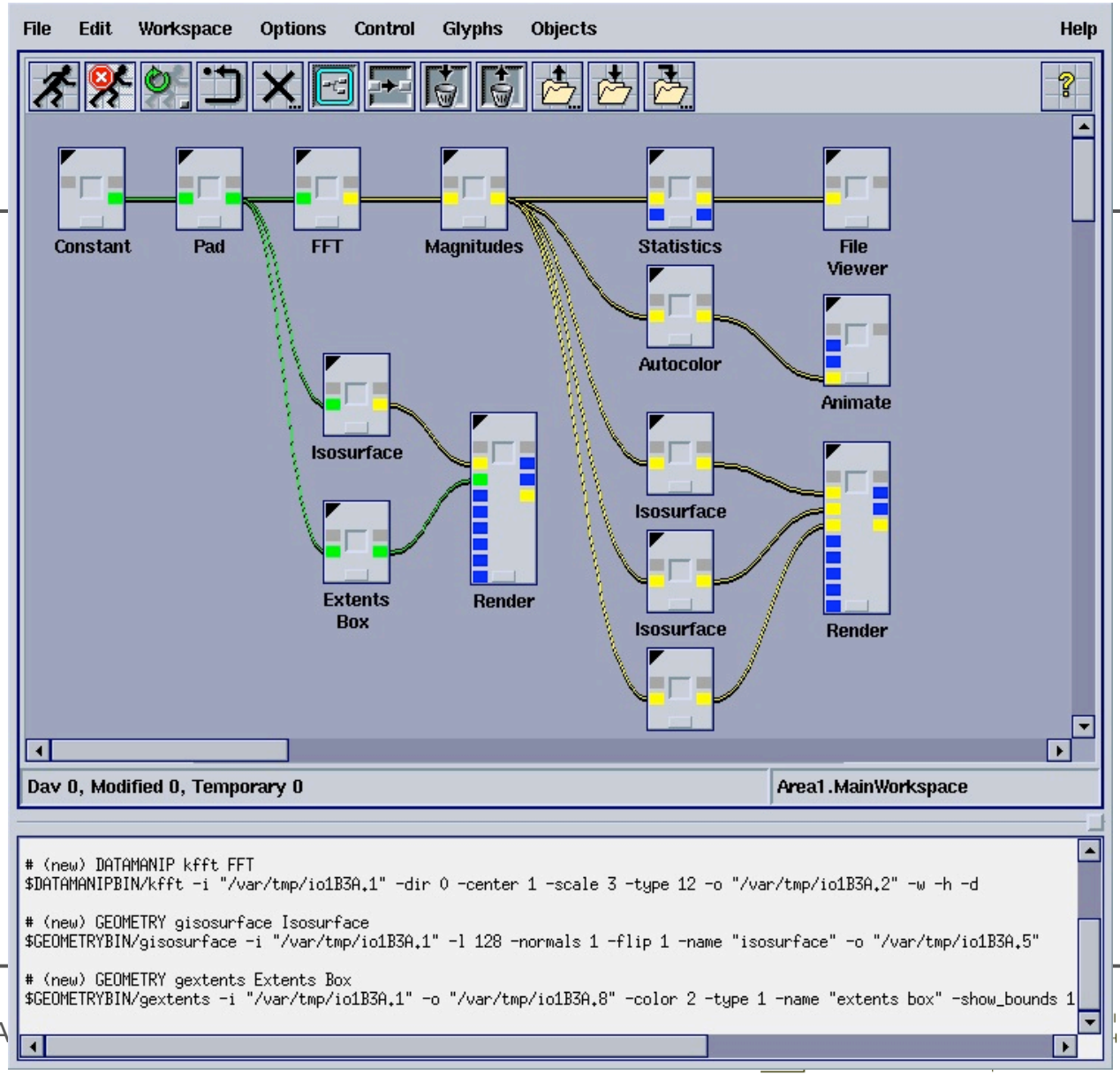
Petri Net language

Petri nets are stochastic – timed attributed (tokens on nodes, transitions)

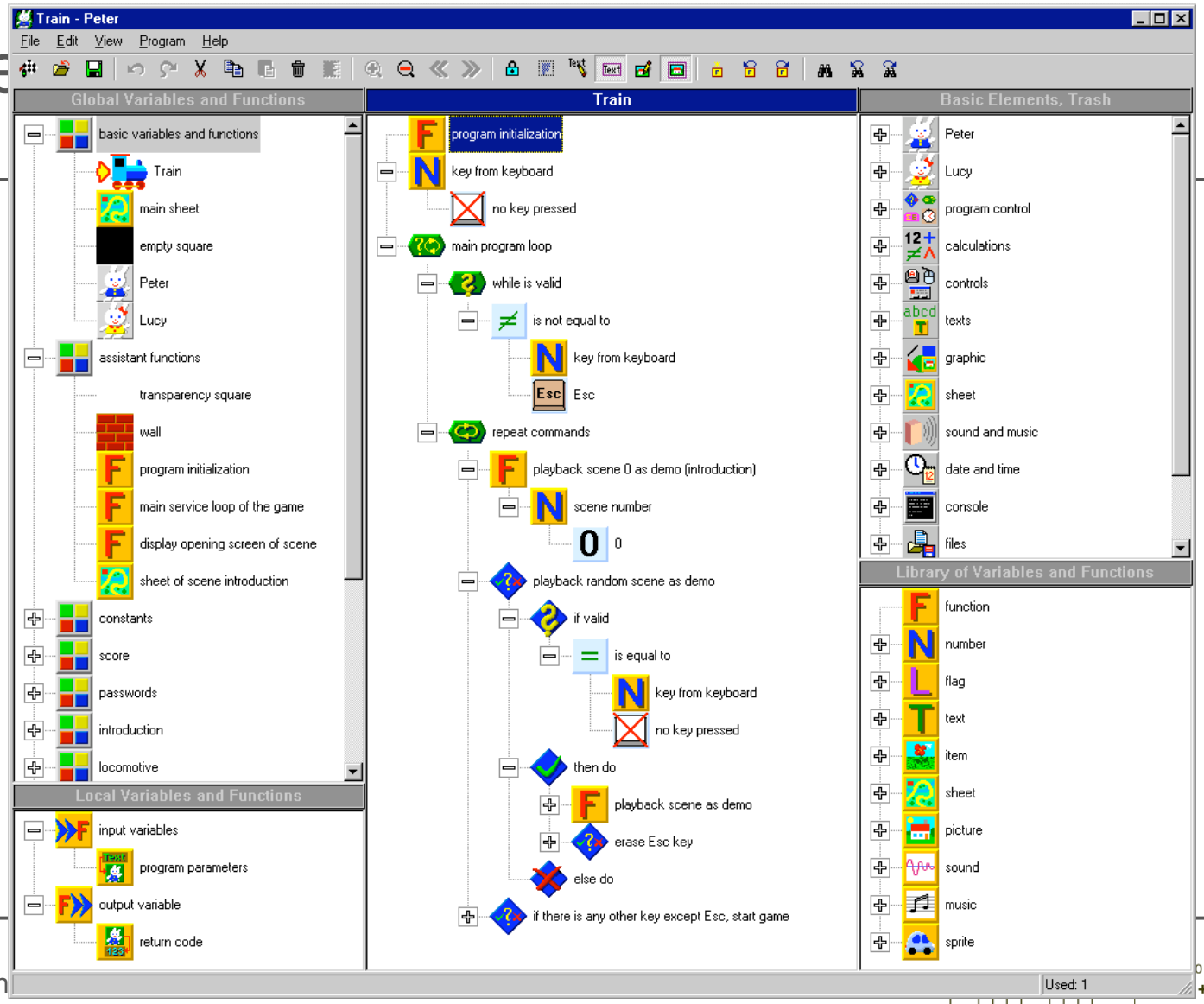


Pictures

Khoros


File
Edit
Workspace
Options
Control
Glyphs
Objects
Help
Constant
Pad
FFT
Magnitudes
Statistics
File Viewer
Autocolor
Animate
Isosurface
Extents Box
Render
Isosurface
Isosurface
Render
Dav 0, Modified 0, Temporary 0
Area1.MainWorkspace
(new) DATAMANIP kfft FFT
$DATAMANIPBIN/kfft -i "/var/tmp/io1B3A.1" -dir 0 -center 1 -scale 3 -type 12 -o "/var/tmp/io1B3A.2" -w -h -d
(new) GEOMETRY gisosurface Isosurface
$GEOMETRYBIN/gisosurface -i "/var/tmp/io1B3A.1" -l 128 -normals 1 -flip 1 -name "isosurface" -o "/var/tmp/io1B3A.5"
(new) GEOMETRY gextents Extents Box
$GEOMETRYBIN/gextents -i "/var/tmp/io1B3A.1" -o "/var/tmp/io1B3A.8" -color 2 -type 1 -name "extents box" -show_bounds 1

Pete



Pictures an

Linking images and words through gesture

- When possible spoken information should be used to accompany images
  - Text removes/reduces visual attention away from imagery

- Gesture is most natural way of linking spoken material with visual imagery

Linking images and words through gesture

- Deixis and the deictic gesture
- Can be a glance or a nod
- Pre-speech
- Shown to disambiguate verbal communications
- Why the mouse is so powerful
- The basis of shared environments

Linking images and words
Deixis

- Pointing is an elementary speech act.
- Pointing links images and words
- People like to point and talk at same time
- Put that (points) there (points)
- Subject verb predicate

Other kinds of gestures

- Beat gestures for emphasis
- Verb gestures showing how to do something
  - Conducting
  - Directing traffic
  - Visual languages of gesture
- McNeill Hand and mind
  - kinetographics

What is the syntax and semantics of gesture?

- Coproduction
- Symbols and “mogens” “
- Expressive gestures
  - Meaning from motion (Amaya, McNeill)

Visual momentum

- Woods (1984): continuity or flow maintained from one visual scene to another
- 4 principles (Wickens)
  1. Use consistent representations
  2. Use graceful transitions to retain context
  3. Highlight anchors
  4. Provide overview for context
     - “establishing shot”
     - Set context first

Issues in shared environments

- Speech + Pointer + Visuals – most important components
- Shared actions need to be synchronised with shared cursors
- Subtle ways of directing attention also important in meeting dynamics.
- New design space of “reference cues”

Animated visual languages

- Tendency to anthropomorphise program descriptions
  - “A tells B to” - function invocation
  - “C decides to “ - behaviour
- Attempts to use this in visual coding
  - KidSim
  - ToonTalk
- Some evidence that animation helps initial learning (synthesis)
  - Effect does not sustain, words have longer “shelf lfe”

Conclusion

- Complex decision about which is better
  - Intimately related to task
  - While words are mostly better for describing procedural information, some graphics are clearly effective
- Words are most common - so only use images for clear advantage
  - Recognition Is different from programming!
- Most visualisations are hybrid
- Linkages and cross-references are crucial