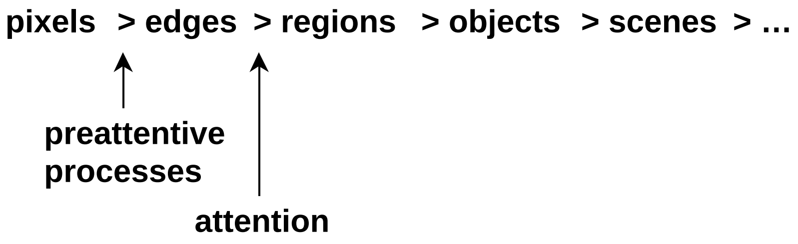
CPSC 532E — Week 10: Lecture

Scene Perception

- Virtual Representation
- Triadic Architecture
- Nonattentional Vision

How Do People See Scenes?

Older view: scene perception is carried out by
a sequence of operations:



pixels > edges > regions > objects > scenes > …

preattentive processes

attention

-> Build up a complete description of scene

However...

Change blindness shows that we can only integrate what we attend to

And we can’t attend to much (4-5 items)

Thus, observers have coherent representation of only a few objects at any moment (= objects that are attended)

If only a few objects are represented at a time,
why do we feel we see all objects at once?

Proposal: Virtual Representation (Rensink, 2000a)

Observation:

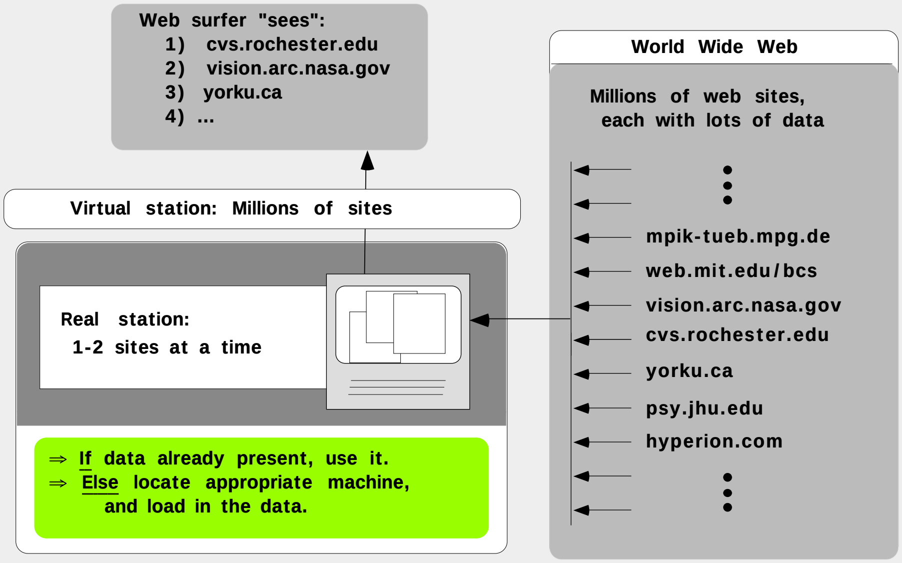
- Although objects appear to be present simultaneously, do not all need to be **represented** simultaneously
- All that is needed is that the properties of the objects can be **accessed when requested.**

This is virtual representation

If co-ordination is successful, it will appear to higher levels as if representation is “real”, i.e., as if all items present simultaneously.

In such a case, the sparse nature of the object representation is completely **transparent** to higher-level processes.

This strategy forms the basis of Web surfing…



Web surfer "sees":
1) cvs.rochester.edu
2) vision.arc.nasa.gov
3) yorku.ca
4) ...

Virtual station: Millions of sites

Real station:
1-2 sites at a time

⇒ **If data already present, use it.**
⇒ **Else locate appropriate machine, and load in the data.**

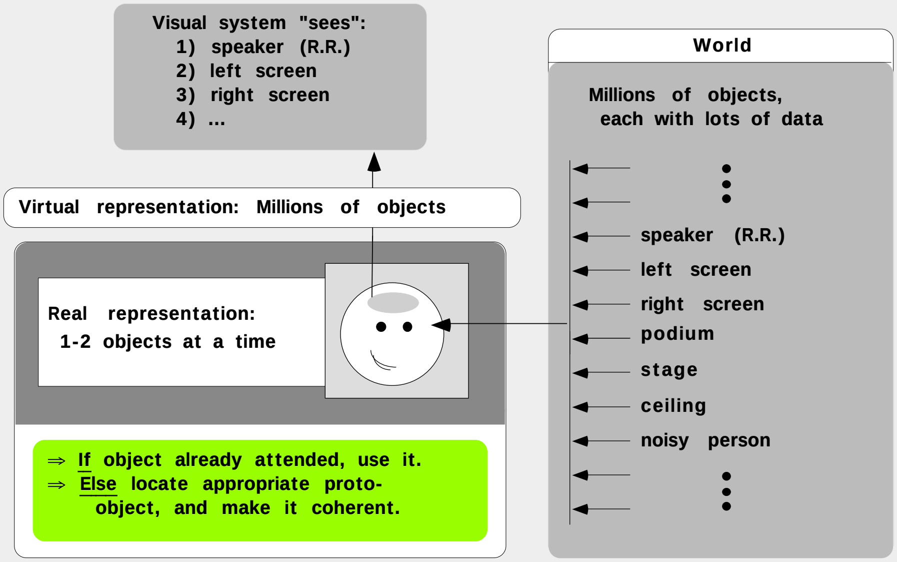
World Wide Web

Millions of web sites, each with lots of data

Can this work for the visual system? **Yes!!**

- can always obtain information from the world
  - **use the world itself as an external memory**
    (***Stroud, 1955; Brooks, 1991***)

- to build a coherent representation of an object, focus eyes and attention on appropriate location **whenever that object is needed**
  - this representation “dissolves” once it is no longer needed



Visual system "sees":
1) speaker (R.R.)
2) left screen
3) right screen
4) ...

Virtual representation: Millions of objects

Real representation:
1-2 objects at a time

⇒ <u>If</u> object already attended, use it.
⇒ <u>Else</u> locate appropriate proto-object, and make it coherent.

World

Millions of objects,
each with lots of data

Note 1:

- **attended representations never contain a complete description of an item at any instant in time**

(e.g., in airplane example, don't see engine change, even though it is part of attended object [i.e. airplane])

- attention traverses the object hierarchy, holding onto only the few details needed at the time.

-this makes it appear as if the object is always seen in full detail

Note 2:

- although world is an **external memory**,
  it is **<u>not</u> an external representation**
  (as proposed by e.g., Brooks, 1991)

- representations are still needed at early levels for various purposes, e.g.
  - compensating for object occlusion
  - linking together elements in the image that are related in the scene

Note 3:

- using the world as an external memory means that **perception is not carried out in isolation**
  - → perceiver and environment form a **partnership**

- environment is not only an **external memory;**
  it can also be an **external processor**
  - → example of situated cognition (see Clark, 1997)
  - → mechanisms already exist to couple perceptual system to environment.
    Can use these as the basis for effective human-machine interactions

Question:

How might a virtual representation be implemented in the human visual system?

Need a cognitive architecture that is compatible with what is known about human vision.

Proposal : Triadic Architecture (Rensink, 2000a)

Nonattentional extraction of aspects of scene (I):

Gist: abstract meaning of scene (farm, harbor, etc.)
- obtained within 150 ms (***Biederman, 1981***)
- obtained without attention (***Oliva & Schyns, 1997***)

Proposal : Triadic Architecture (Rensink, 2000a)

Nonattentional extraction of aspects of scene (I):

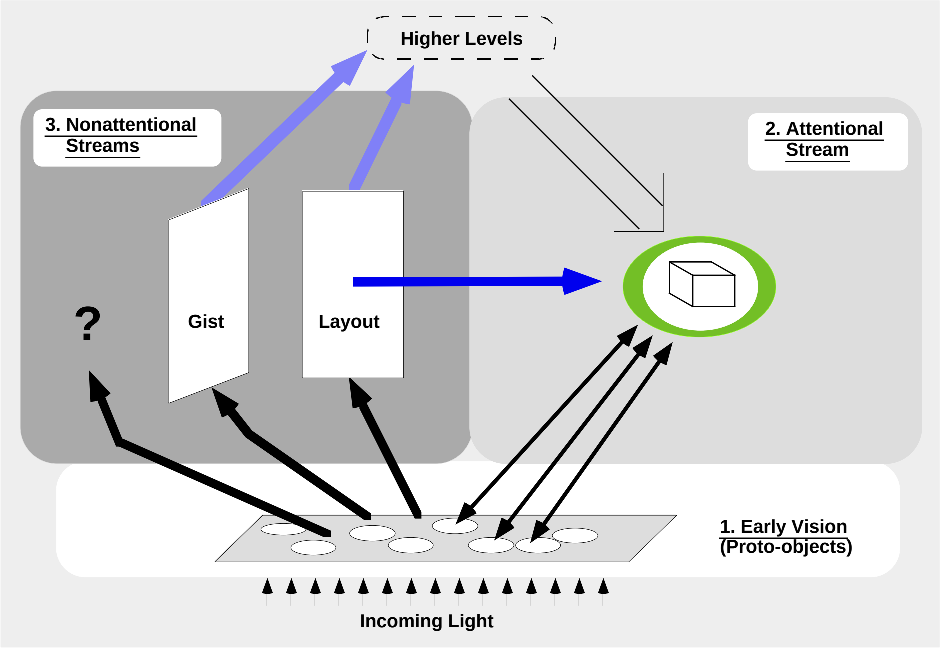
- Gist: abstract meaning of scene (farm, harbor, etc.)
- obtained within 150 ms (***Biederman, 1981***)
- obtained without attention (***Oliva & Schyns, 1997***)

Possibly derived via statistics of low-level structures
(***e.g. Swain & Ballard, 1991***)

Nonattentional extraction of aspects of scene (II):

Layout: arrangement of items in the scene.

- nonvolatile - held without attention (***Simons, 1996***)
- can be learned without attention (***Chun & Jiang, 1998***)



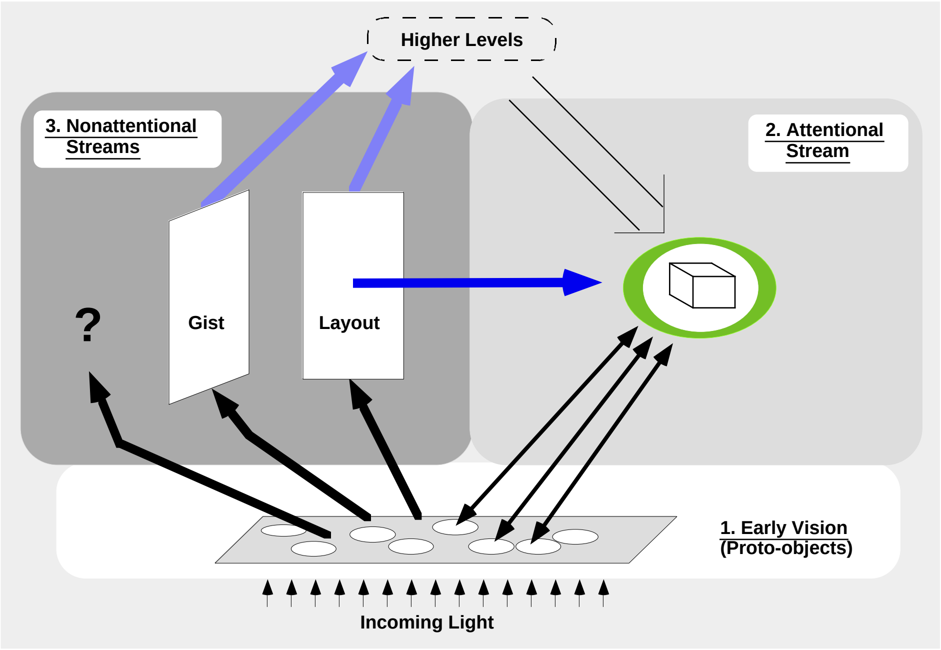
Layout supports much of the attentional guidance that makes operation “transparent”

An important issue:

What are the constraints on the representation of layout?

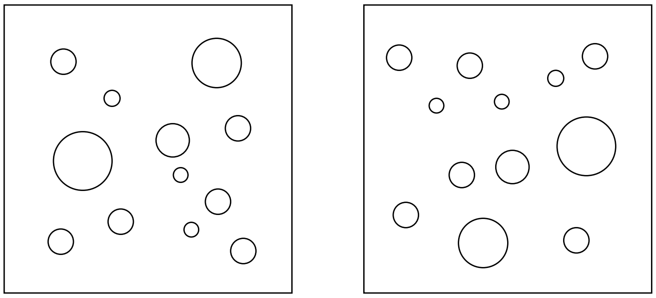
- could e.g., determine possibility of Web browsers capable of handling high-dimensional spaces in an effective (transparent) manner

---



Other aspects of scene may also be picked up rapidly:

Statistical properties (Ariely, 2001)



Observers sensitive to mean + range of sizes

Final note

Only a few objects are seen at any moment
(ie, those that are attended)

But attention is limited in capacity
-> only a few properties represented at any time

Question: Why does each object appear to be
represented completely, and in detail?

Possible Answer: Objects also have virtual representation
- attention travels up & down structural hierarchy
- latches onto whatever part or property is needed at that moment

Summary

Our impression that many coherent objects are represented simultaneously is an illusion.

- scenes represented via a **virtual representation**
- objects might be represented this way, too

Attention is not a central processing “gateway”

- it’s only the stream specialized for **coherent objects**
- other (nonattentional) streams help guide it
- different priorities of attending (based e.g., on cultural background or task requirements) can literally result in seeing the world differently

---

Implications for Display Design

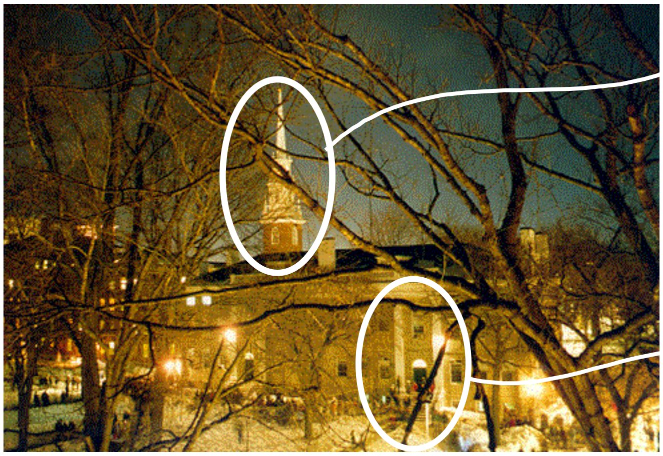
1. Pickup of Information

- optimally effective displays based on:
  - static aspects of visual perception (e.g. color)
  - dynamic aspects (e.g. attention management)

- what is attended depends on the & the
  - different people can literally see the same world very differently

→ can use flicker paradigm to measure which parts and properties of objects are attended first.
(= most easily seen to change)

Rensink et al (1997)



Central Interest

-mentioned by most observers

Marginal Interest

-mentioned by no observers

Central interests are objects/regions perceived to be important (or at least interesting)

Marginal interests are objects/regions that are not

Average time for detecting change (Rensink et al., 1997):

- **Central interests: 4.7 seconds**
- **Marginal interests: 10.9 seconds**

(Marginal changes are on average > 20% larger in area)

Could adapt this to:

different viewers x different tasks x different conditions

2. Interactive Displays / Perceptual Externalizations

- perception is an **inherently interactive** process.
  - representation is not a **copy** of the external world,
  - rather, it **co-ordinates** the actions of subsystems

- interaction with external displays has the potential of being as "natural" as visual perception. E.g.,
  - ***effortless navigation through dataspaces***
  - ***effective incorporation of external processors to aid thinking (visualization)***

→ key to effective interaction may be the use of nonattentional streams (e.g. layout) to provide sufficient **context** to guide attention

3. Coercive Graphics

- display can coerce user to attend to a given location
  - effectively "highjacks" virtual representation

→ can make observer see (or not see) given items

- attentional control via
  - high-level **interest** (cf. movies)
  - low-level **salience** (e.g. unique preattentive feature)
  - mid-level **directives** (e.g., pointing finger)

- such factors are what magicians use to control what audience "sees"

→ could create "magical" effects
- transitions at other locations made "invisible"?
- perception of events that could not occur in the real world. E.g.,
  - "unreal" (dis)appearances
  - "unreal" change to objects, regions

→ **soft warnings** (Rensink, 2002b):
- user automatically sees what they should see (e.g. "magically" notices incoming email)
- no need for hard warning (e.g. beep); attention is controlled in more natural way

Nonattentional Vision

Triadic architecture implies an important role for nonattentional streams in vision

These streams are not primarily concerned with **explicit** (= conscious) perception
- this is done via attentional mechanisms

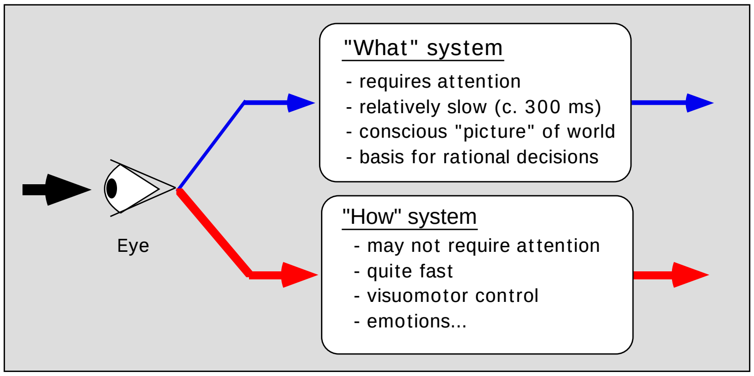
-> Mapped out via **implicit** (= unconscious) detection of change?

Implicit Detection of Change: Visuomotor

- Bridgeman et al. (1975) — **oculomotor response**
  - target moves while observer saccades to it
  - eye makes corrective saccade, even though observers have no explicit perception of change
- Goodale et al. (1986) — **manual pointing**
  - target moves while observer saccades to it
  - hand corrects its trajectory while reaching to target, even though observers have no explicit perception of change

---

Proposal: **Two visual subsystems** Milner & Goodale (1995)



"What" system
- requires attention
- relatively slow (c. 300 ms)
- conscious "picture" of world
- basis for rational decisions

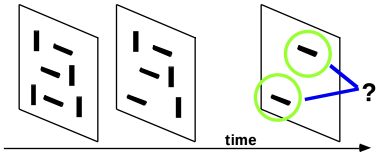
"How" system
- may not require attention
- quite fast
- visuomotor control
- emotions...

Two subsystems (submodalities) are largely separate
- supported by two separate neural pathways

The “how” system is essentially an “inner zombie”

Implicit Detection of change: Perceptual

- Fernandez-Duque & Thornton (2000)
  - observers view 2-display sequence; each display is a simple array of rectangles
  - observers tested on two items: the item changed, and the item diagonally across from it



  - If observer did not notice change, asked to **guess which item changed**.

Results

- Observers could guess **better than chance** (55-63%) even though change was not consciously noticed
  - (a form of blindsight in normal observers)

    ⇐involvement of limited-capacity system

- No attentional priming at location of unnoticed change

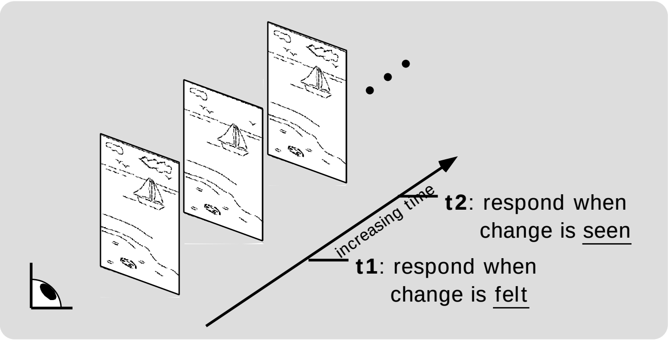
    ⇐involvement of a nonattentional system

Visual Awareness without Visual Experience

Origin - reports by some observers that they "**sensed**" the change long before they saw (= **visually experienced)** it.

Rensink (2000b)

- observers view flicker sequence (natural images)
- asked to hit button (t1) when change was **felt**
- then hit button (t2) when change was **seen**



Results

- 1/2 of observers had no feeling of change without visual experience of it
- 1/3 of observers could feel a change before seeing it
  - (t2-t1) > 1 second on 20% of trials
  - average duration = 3.7 seconds
- not a result of guessing:
  - accuracy on catch trials is good (82%)

Mindsight: Conscious (mental) awareness without
an accompanying visual experience

Different than seeing with visual experience
- different sensitivities to types of change

Mindsight due to a nonattentional system (alert?)

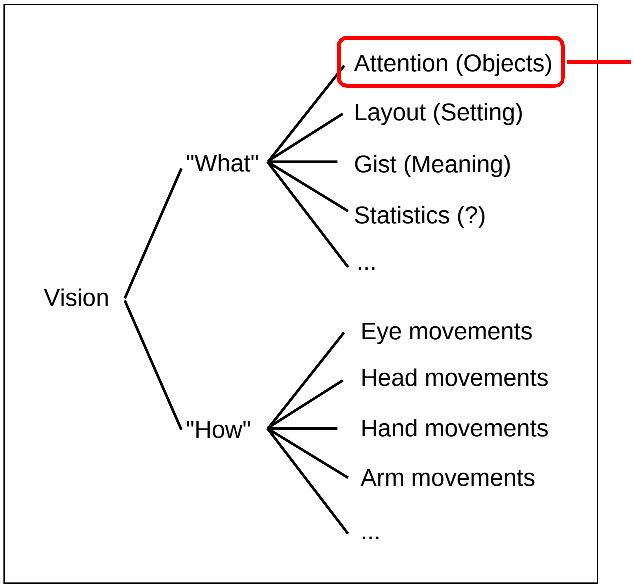
→ Basis of the belief in a “sixth sense”???

Summary

Vision appears to be carried out by a **set of subsystems**, each of which operates **concurrently**, and is largely **independent of others**

Primary division into
- “What” system: (conscious) visual perception
- “How” system: nonconscious visuomotor actions

Division of “What” system into
- Attentional system: object perception - conscious(?)
- Nonattentional system: visual context - nonconscious(?)



Only one of these systems is involved with conscious visual perception

Implications for Display Design

Display might influence other aspects of user’s experience besides conscious “image” of its contents

1. Visuomotor Actions

- displays designed for “zombie system”:

→ could guide user actions (e.g. control of mouse)

→ avoid problems with lag for visual feedback
  - pointing can be precise with open-loop control (Po, 2002)

→ might induce user to automatically “do the right thing”
  (no conscious noticing of this)

---

2. Displays for “Sixth Sense” Experience

- feeling that something is occurring, without an accompanying visual experience

→ use as a second form of “soft warning”
  - increase user vigilance without disrupting normal attentional allocation during a task (e.g. when driving)


STAR WARS
YODA