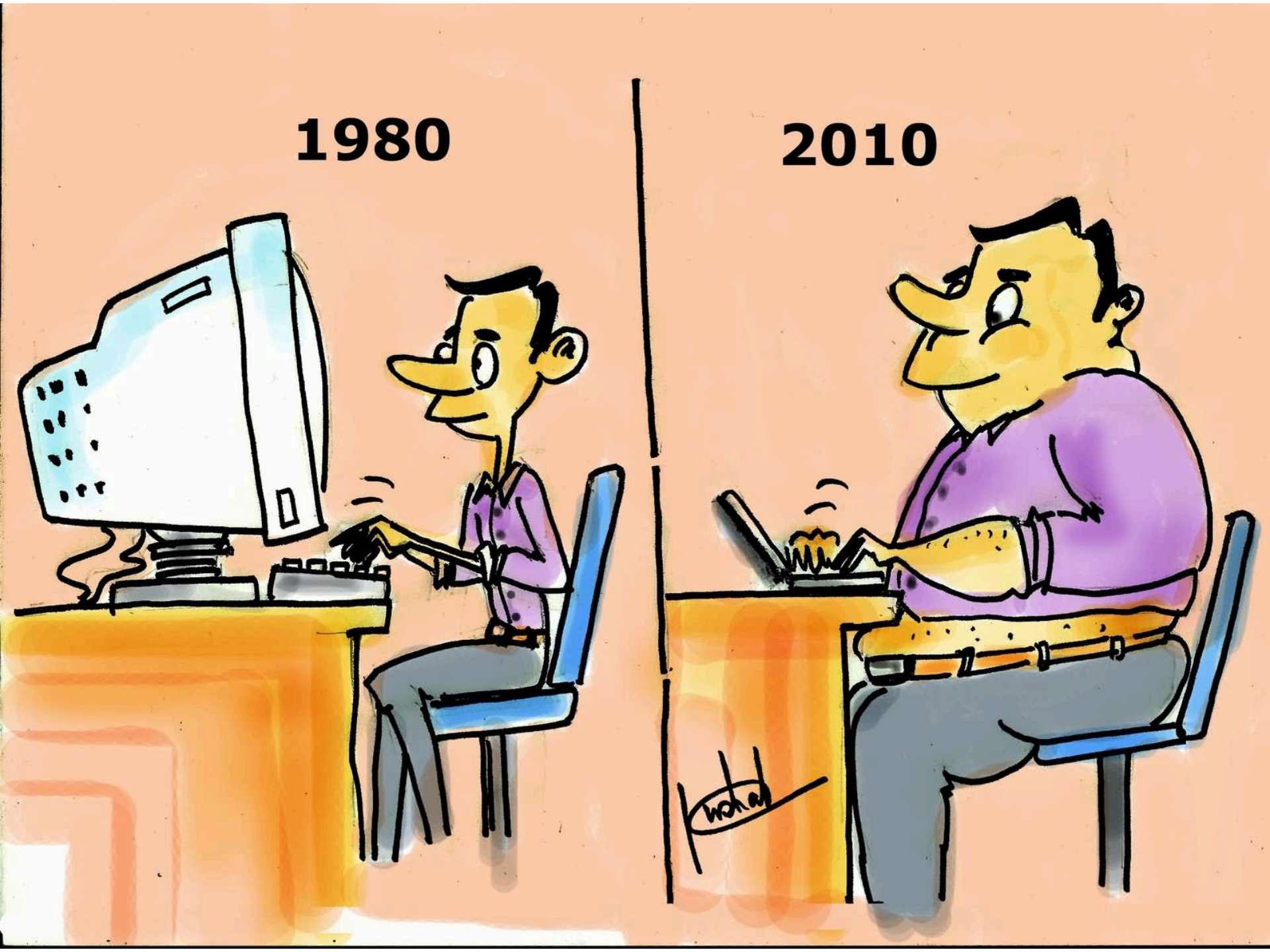
Object Recognition

"Now! – *That* should clear up a few things around here!"

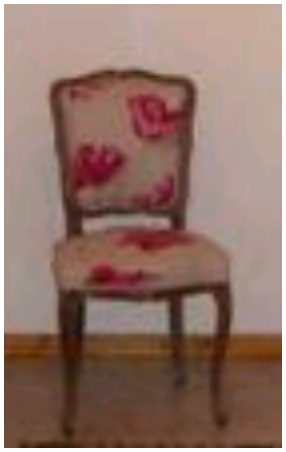
ELL788
Date: 03,05/10/16


1980
2010

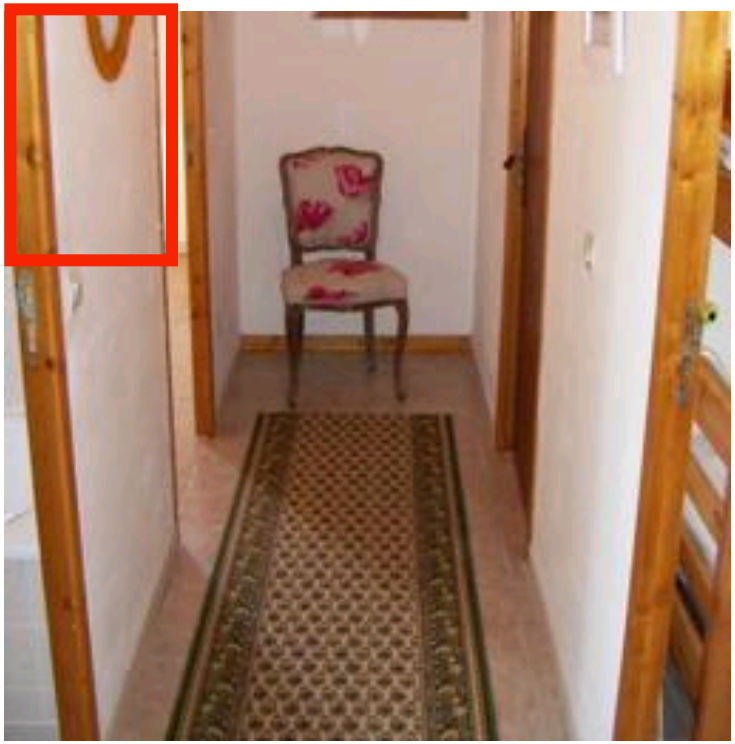

Fernbahnhof
Long distance trains
Gepäckausgabe
Baggage claim
Gleis
Track
4-7
Reisezentrum
Travel center
DB
DB

Object recognition
Is it really so hard?

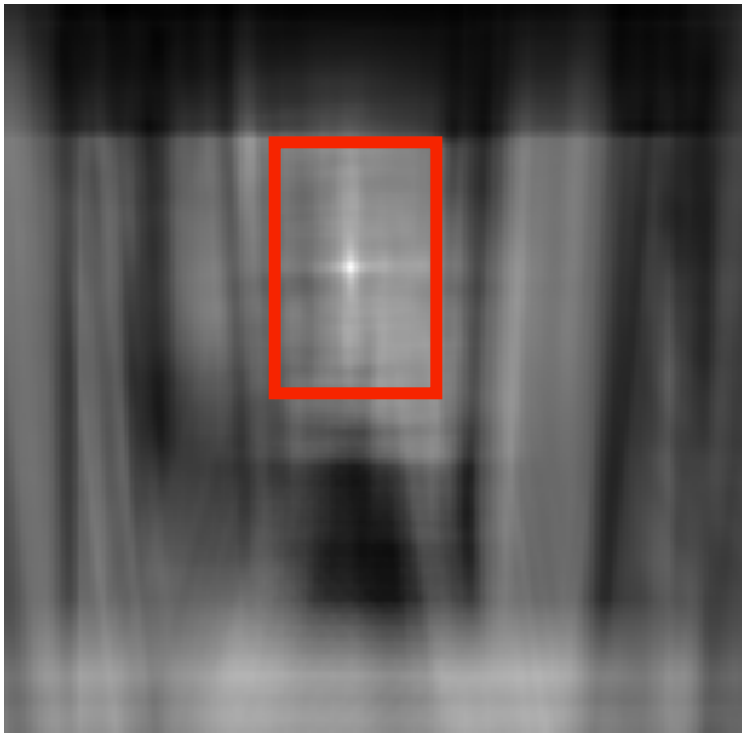
This is a chair

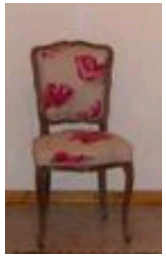


Find the chair in this image



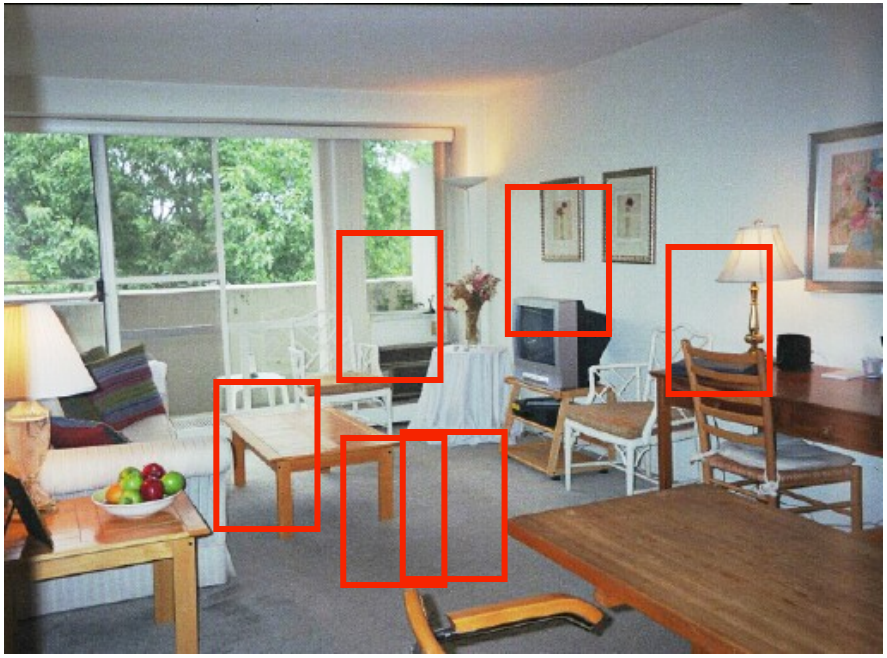
Output of normalized correlation

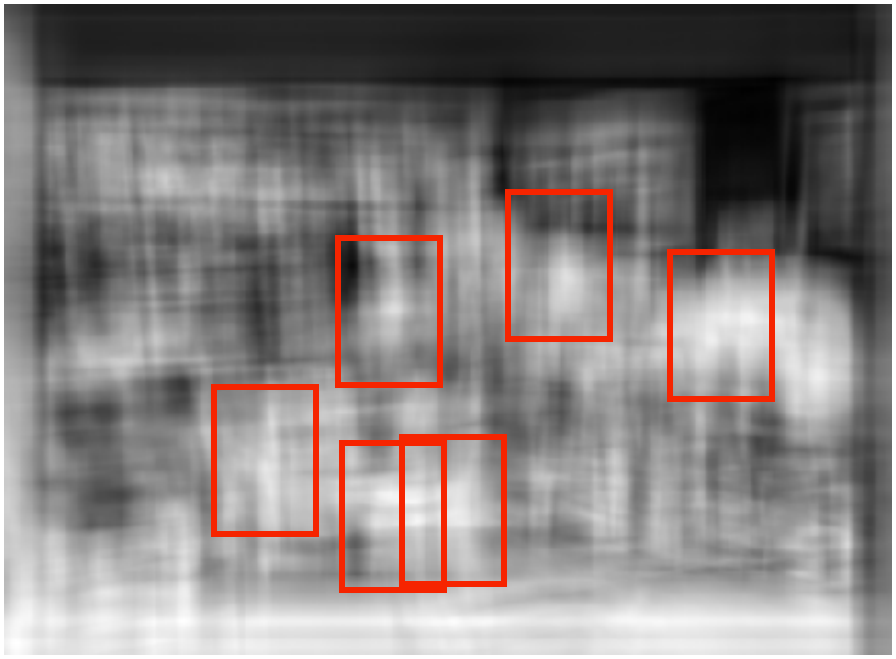




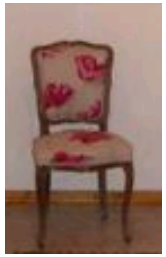
Object recognition
Is it really so hard?

Find the chair in this image





Pretty much garbage Simple template matching is not going to make it

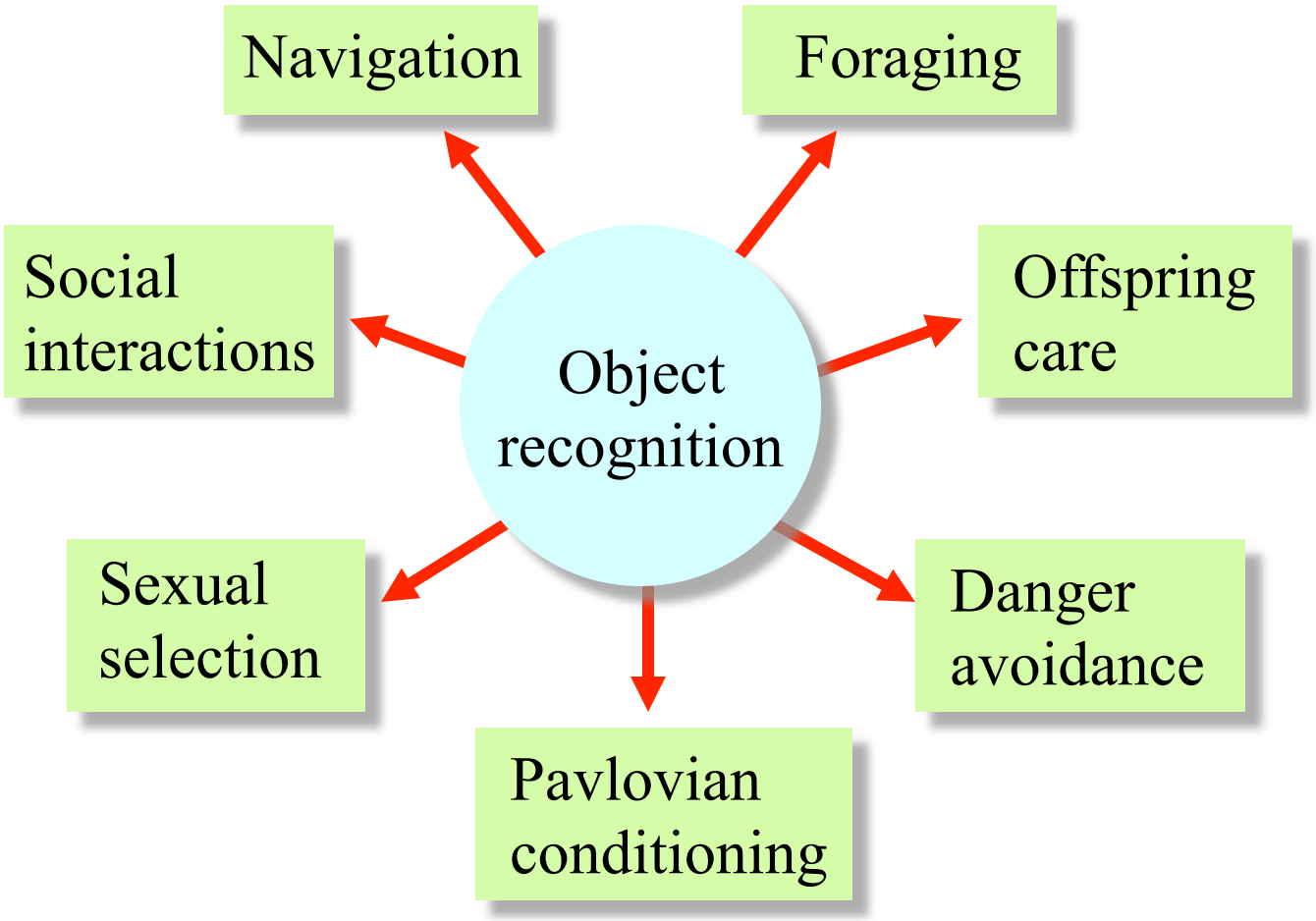


Object recognition
Is it really so hard?

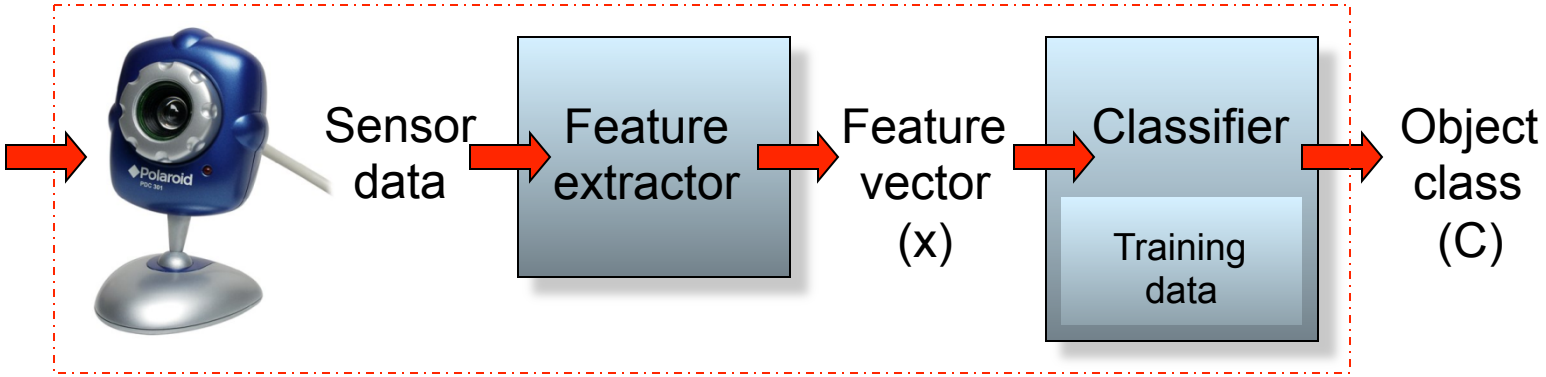
Find the chair in this image

A “popular method is that of template matching, by point to point correlation of a model pattern with the image pattern. These techniques are inadequate for three-dimensional scene analysis for many reasons, such as occlusion, changes in viewing angle, and articulation of parts.” Nivatia & Binford, 1977.

Why is recognition important?



What are the basic components of a recognition system?



Let's say the recognition task is to distinguish men from women

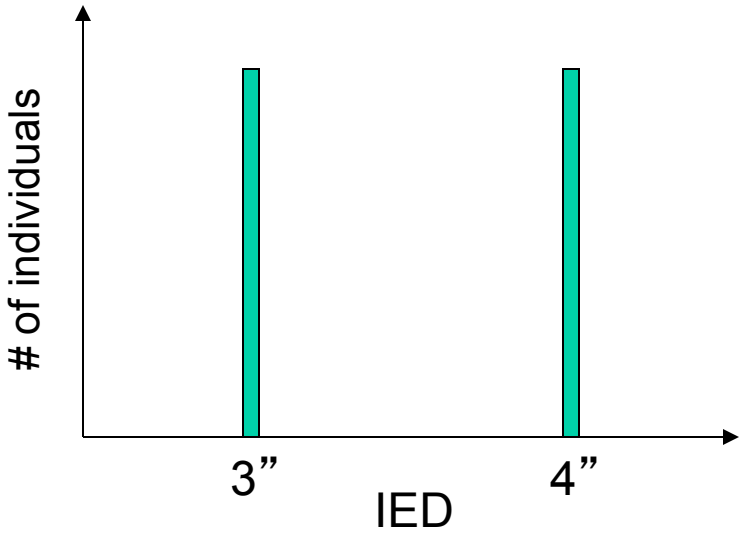
Start with a feature 'x', say the distance between the eyes.

How can we classify this feature?

Classifying feature vectors

Case I: Features are invariant and diagnostic

Imagine a tribe where all males have inter-eye distance of 4” and all women have an IED of 3”.



The classification task here is trivial.

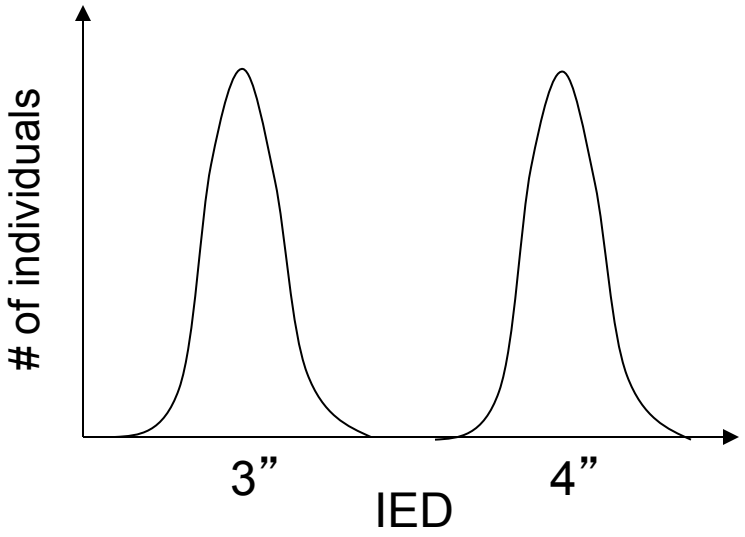
But, most real features are not invariant. They typically have scatter.

Classifying feature vectors

Case II: Features have scatter but are diagnostic

IED for men: $3.8 < x < 4.2$

IED for women: $2.8 < x < 3.2$



The classification task here is still simple, because x is an unambiguous feature.

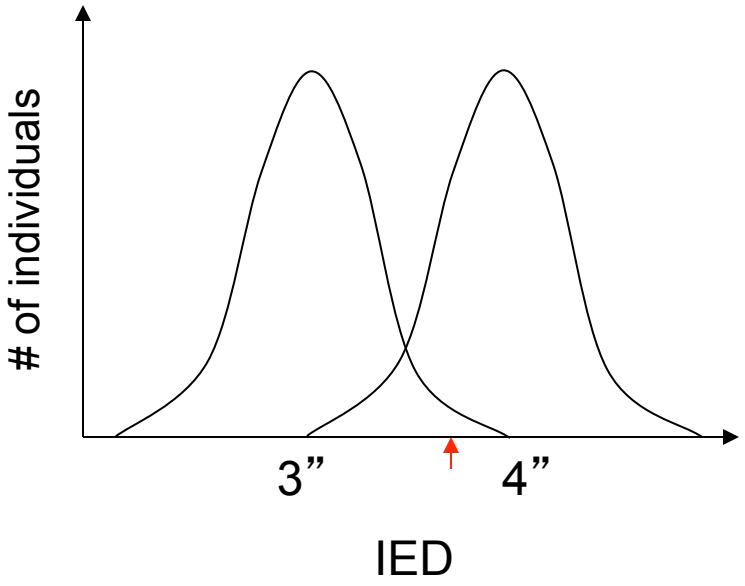
But, even this is highly unusual in the real world.

Classifying feature vectors

Case III: Features have scatter and are overlapping

IED for men: $3 < x < 5$

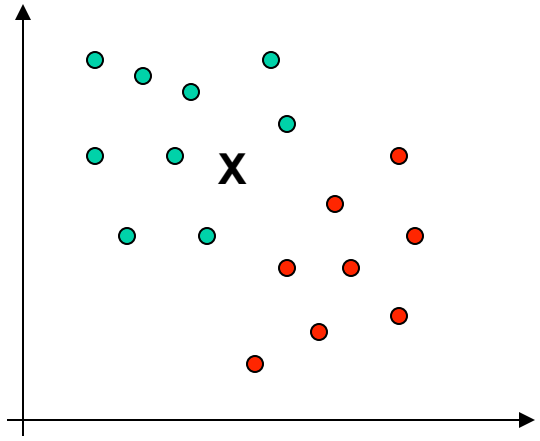
IED for women: $2 < x < 4$



The classification of at least some points is ambiguous.

What might be a principled way of saying which class a point belongs to?

Biological recognition systems often adopt a 'Nearest-Neighbor' classification strategy.



Determine the distance of new point from all points in the dataset. Pick the nearest point and assign the new point its label.

A variation: k-NN

What would a nearest-neighbors strategy mean for recognition?
Collecting a library of images and finding the best match of a new image within this library. This referred to as 'image-based' recognition.

Recognition by insects

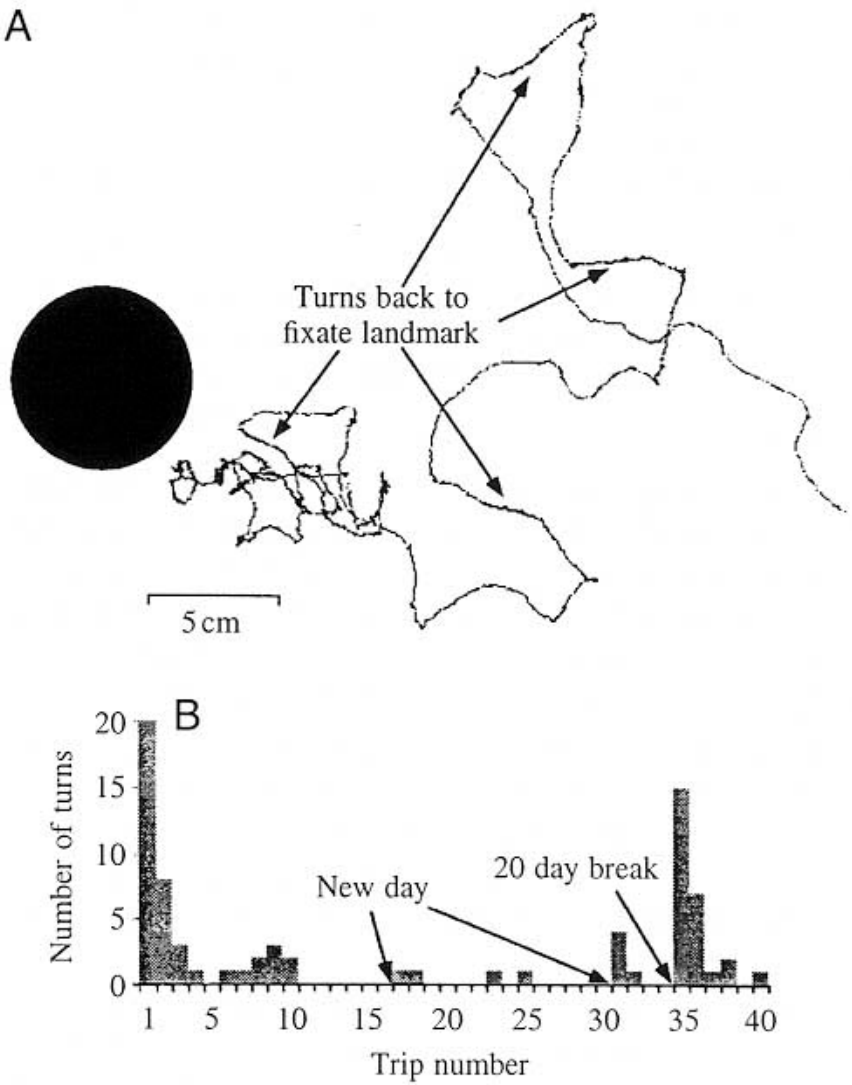
Recognition by primates

Evidence of image-based recognition in the desert ant

Sahara ant

http://www.youtube.com/watch?v=w9KDM4C1kVg

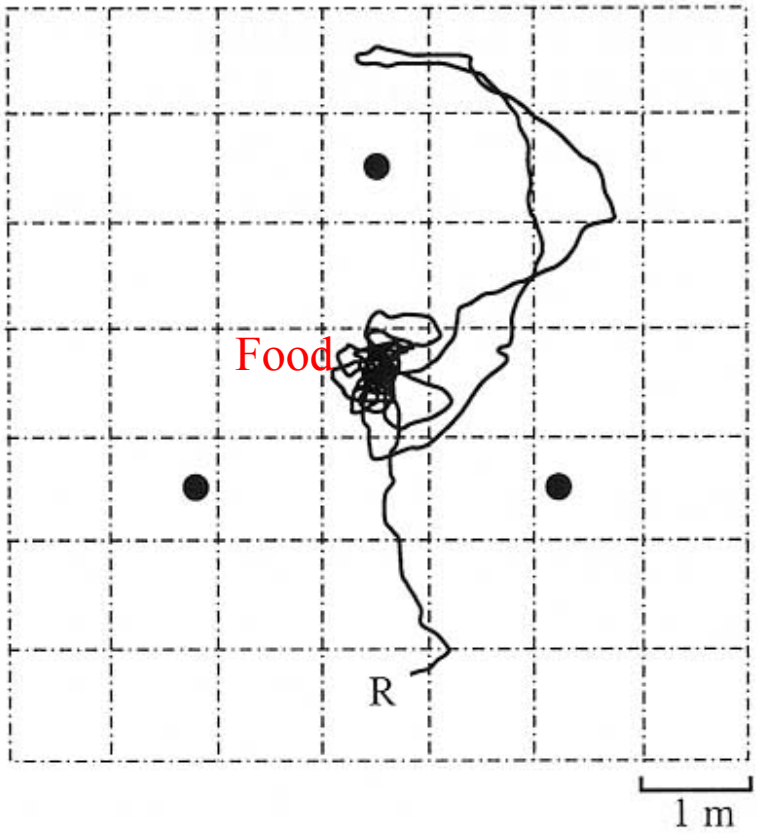
Step 1: Building a library of images

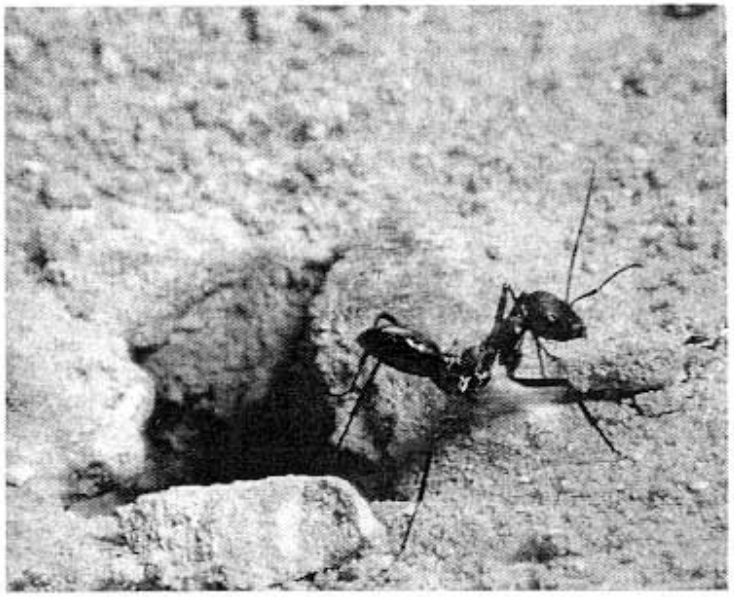


Multiple such objects can help to pinpoint locations precisely…

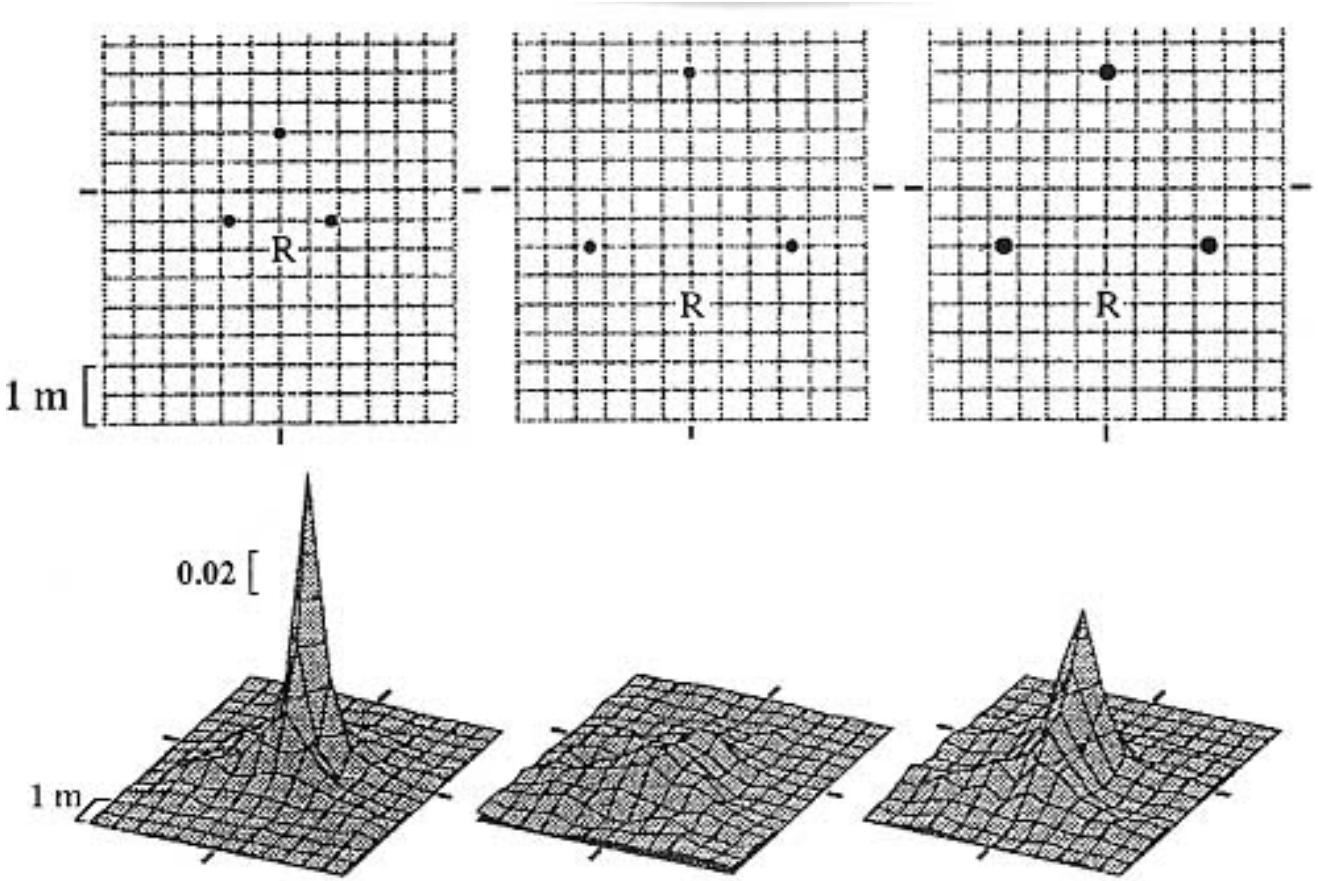
Step 2: Using images to pinpoint places by the desert ant

Wehner et al. (1996)





Step 2: Using images to pinpoint places by the desert ant (contd.)



The matching appears to be retinotopic

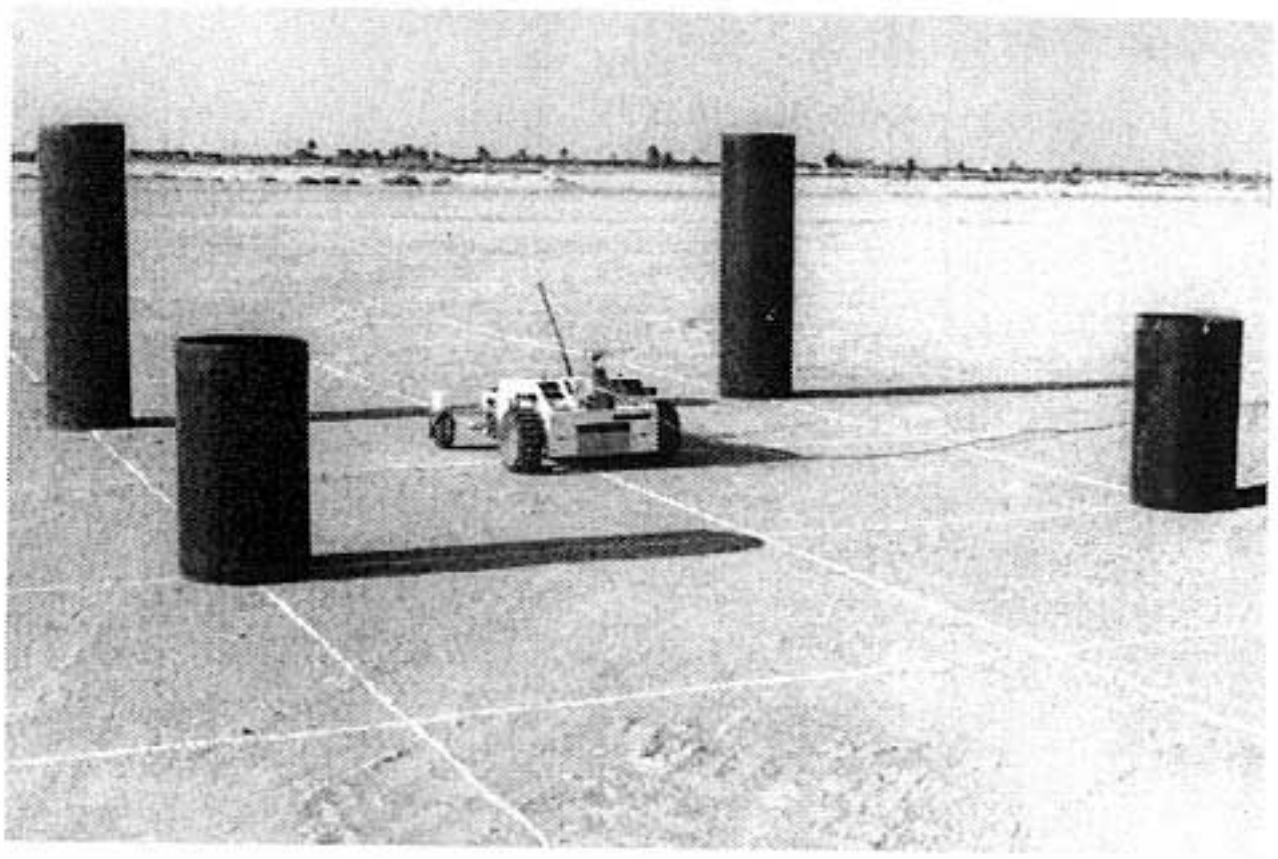
A mobile robot based on the desert ant's retinotopic matching strategy

Max-Planck Institute for Psychological Research, Munich

Sahara ant

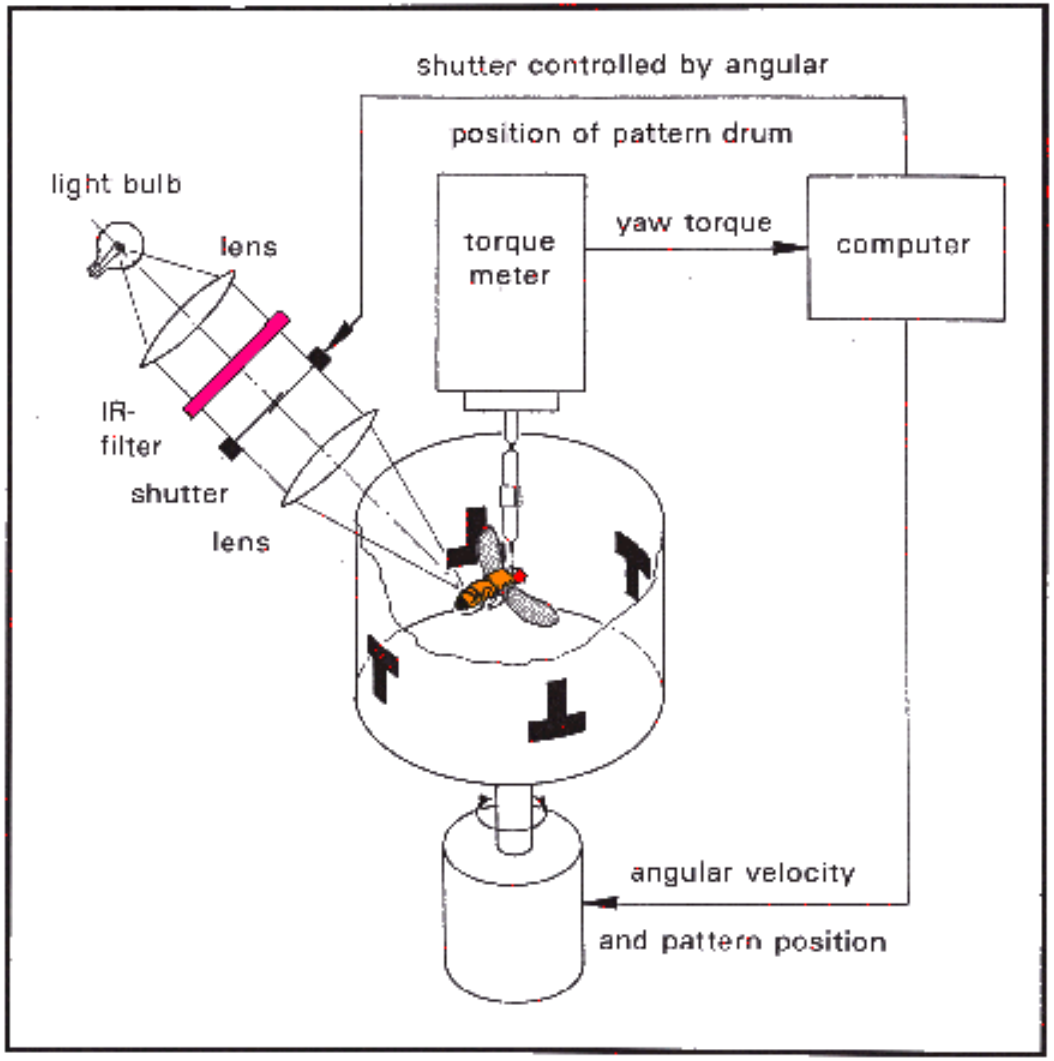
'Sahabot'

Showtime for Sahabot



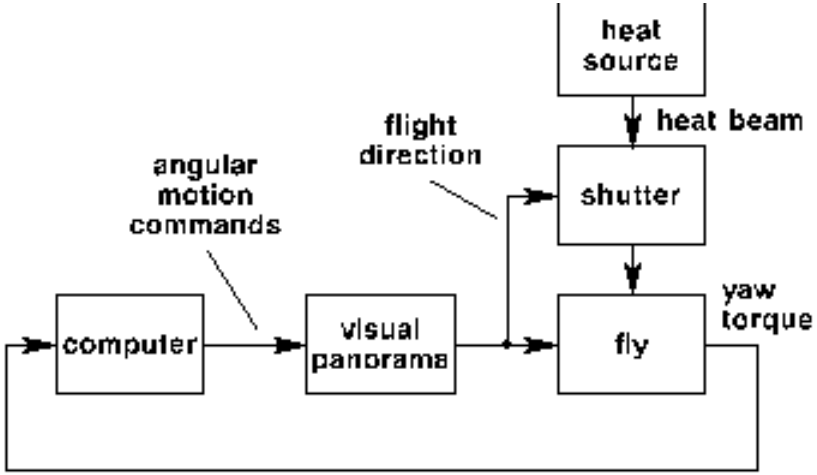
Evidence for retinotopic image matching by Drosophila (Dill et al, Nature, 1993)

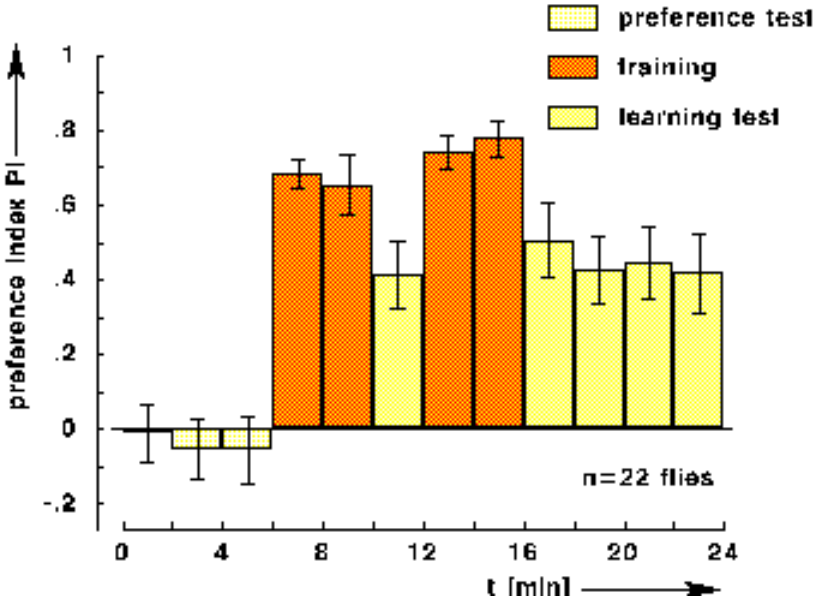
Experimental setup



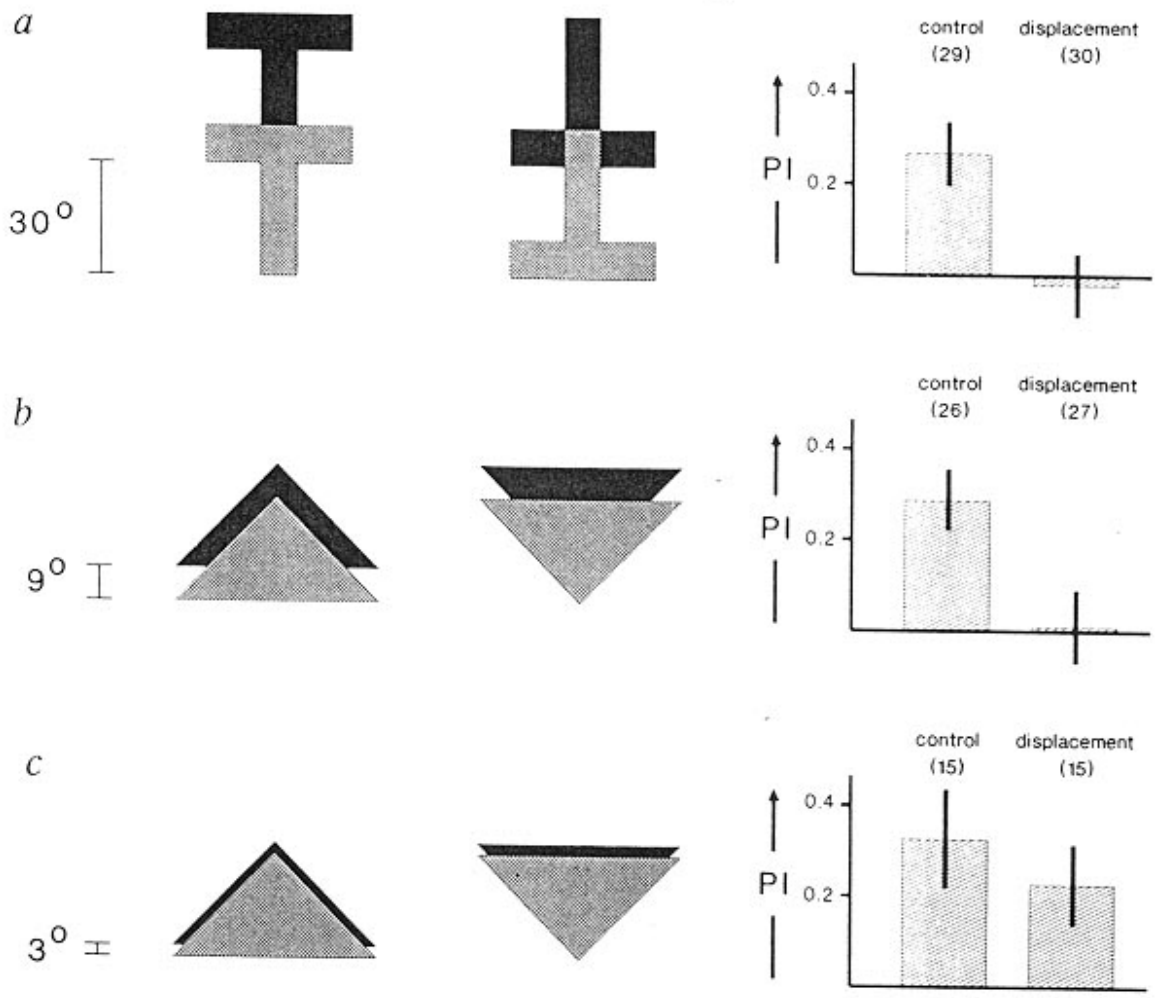
'Flight simulator'

Rapid learning of discrimination



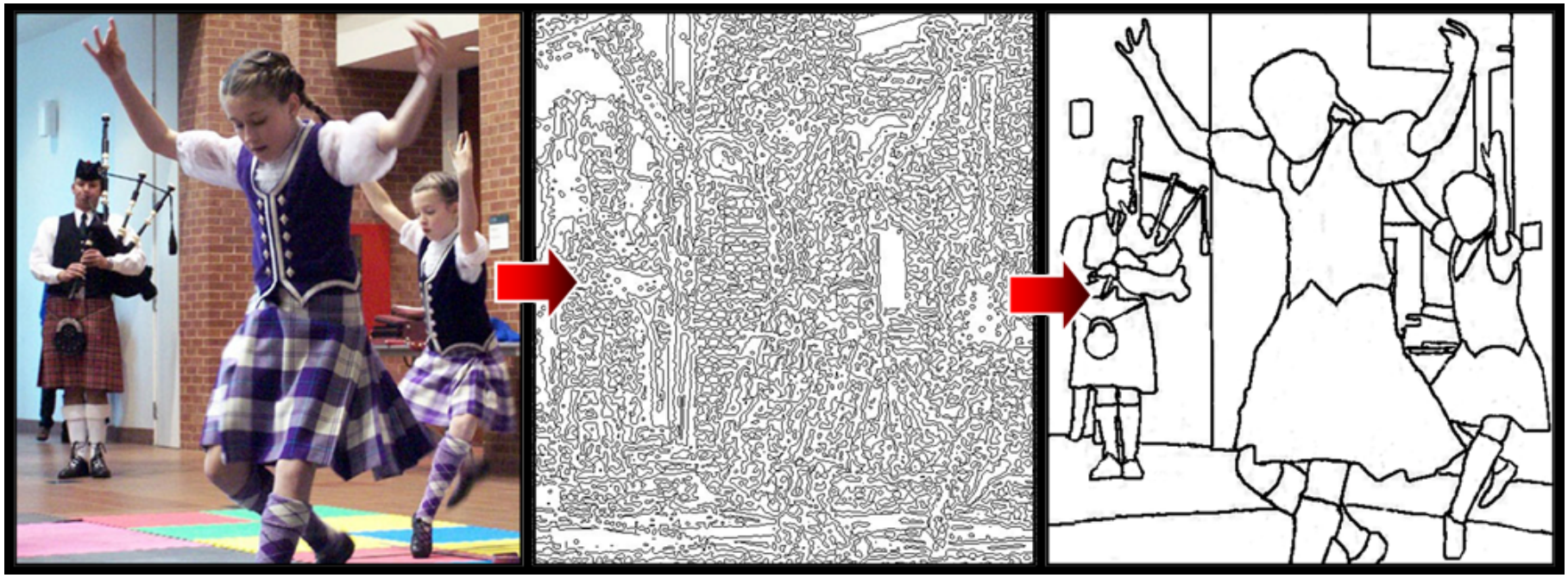


Effects of translations

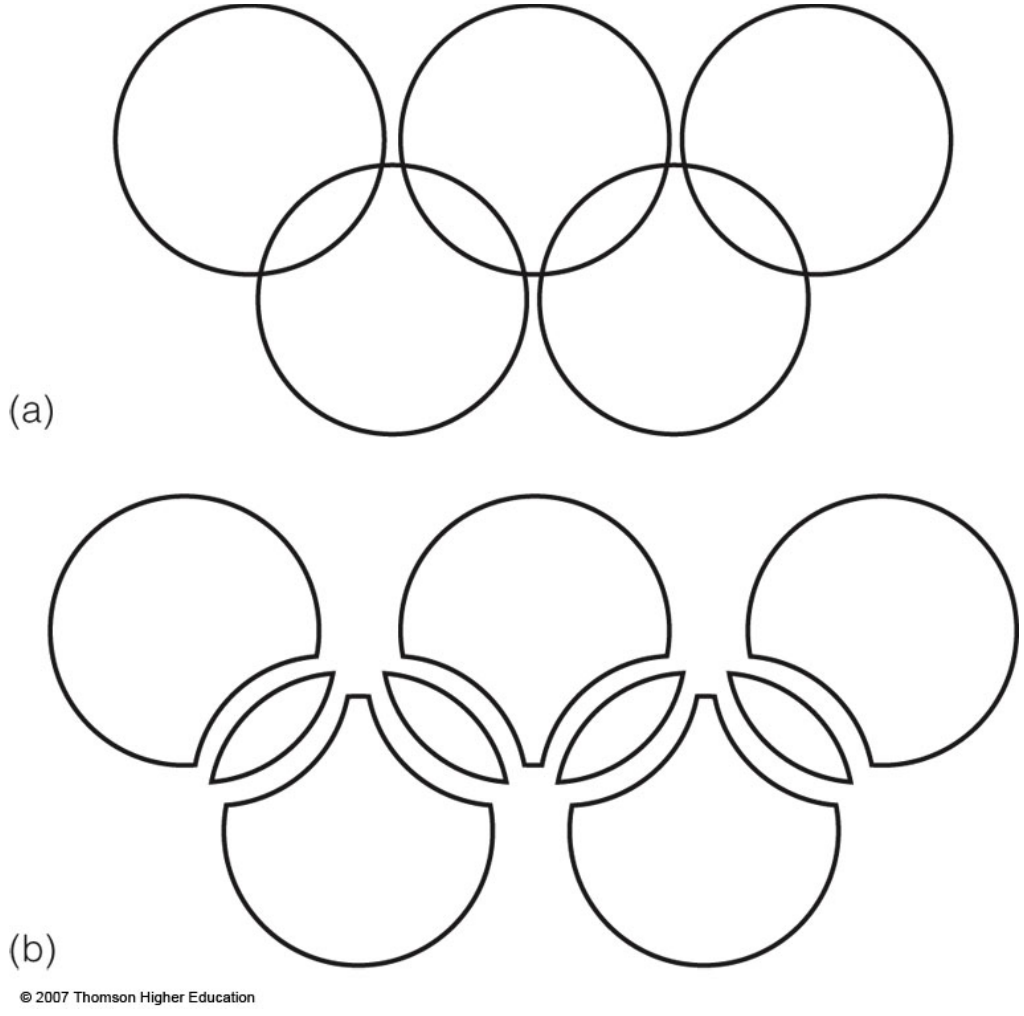


The fly is dramatically impaired by even small translations. This suggests retinotopic encoding of images.

Object Perception by Primates



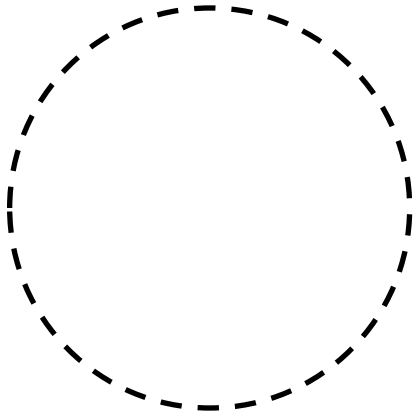
How does the visual system put together the fragments to form meaningful objects?

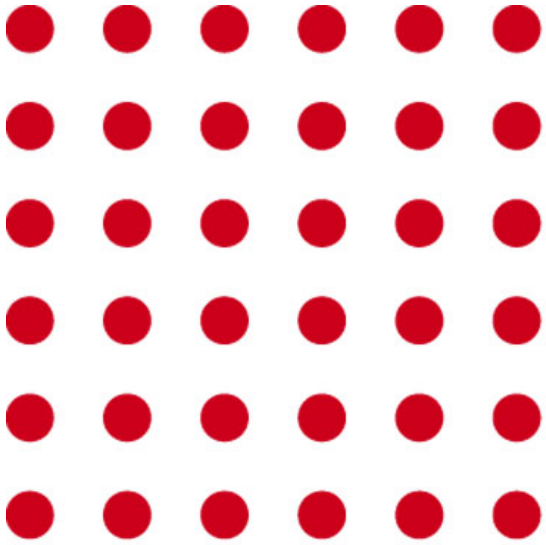


(a) This is usually perceived as five circles, not as the nine shapes in (b).

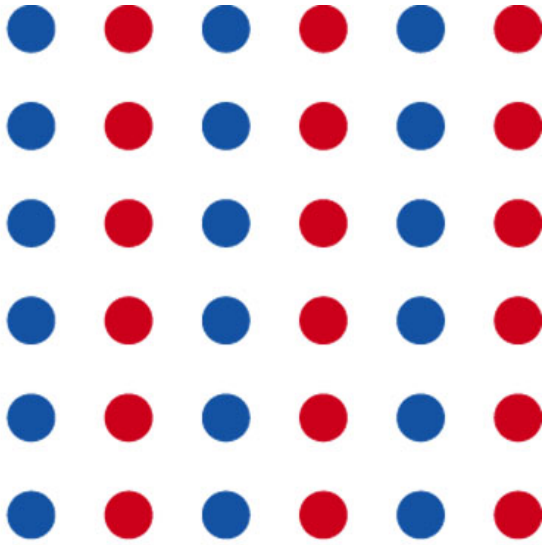
The Gestalt Approach

- The whole differs from the sum of its parts and is a result of perceptual organization





(a)



(b)

(a) Perceived as horizontal rows or vertical columns or both. (b) Perceived as vertical columns

Evidence for use of Gestalt grouping by digger wasp

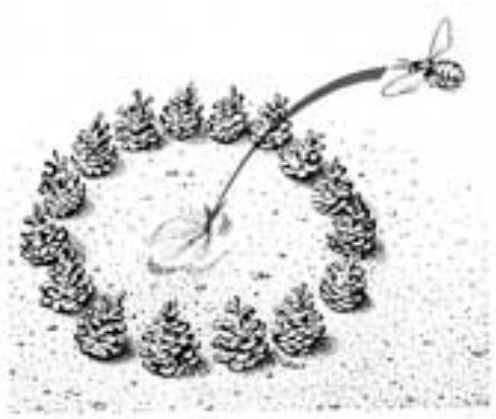
Nobel Prize, 1973

Tinbergen (1932)

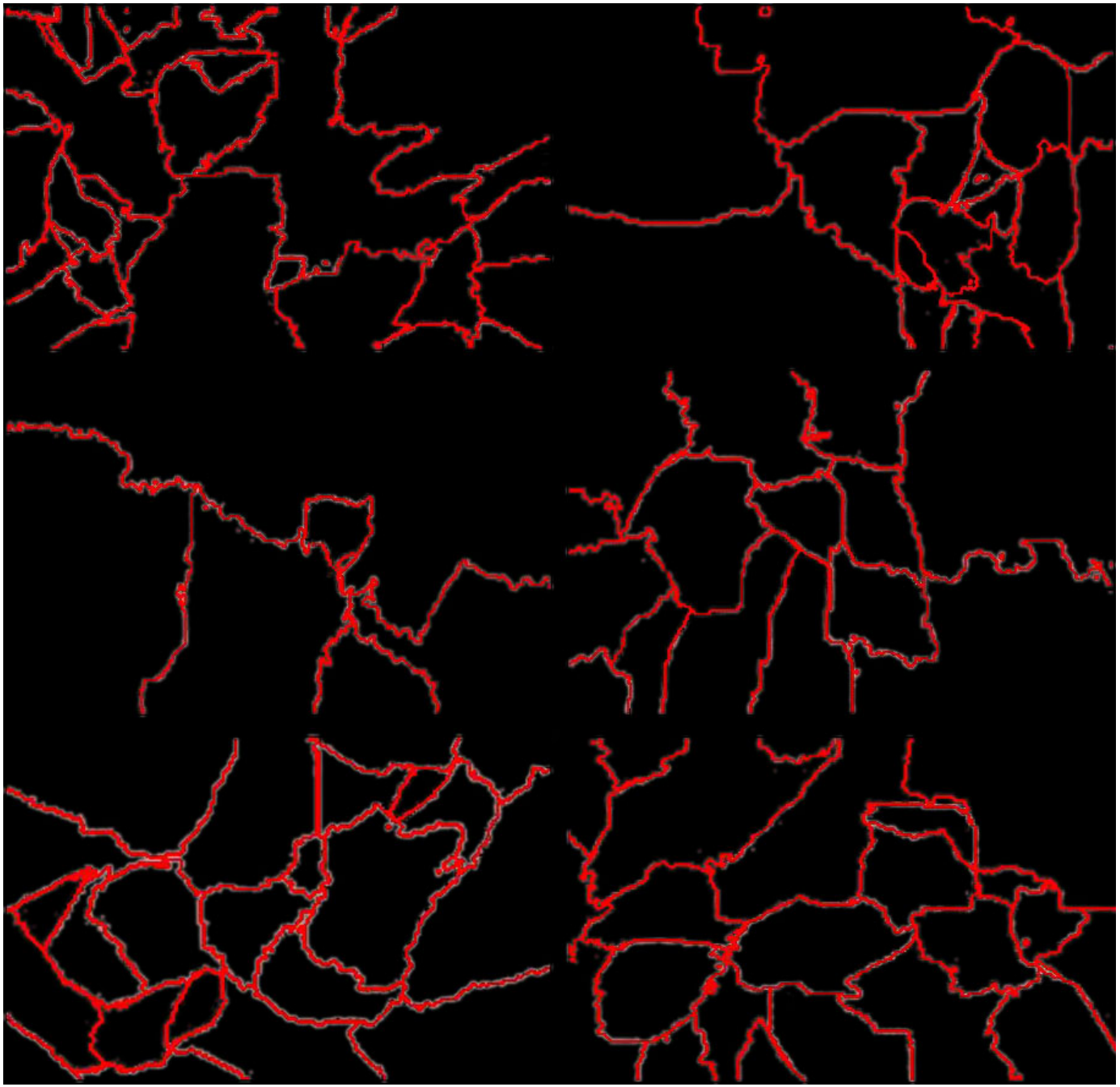
Evidence for use of global grouping by digger wasp

Nobel Prize, 1973

Tinbergen (1932)



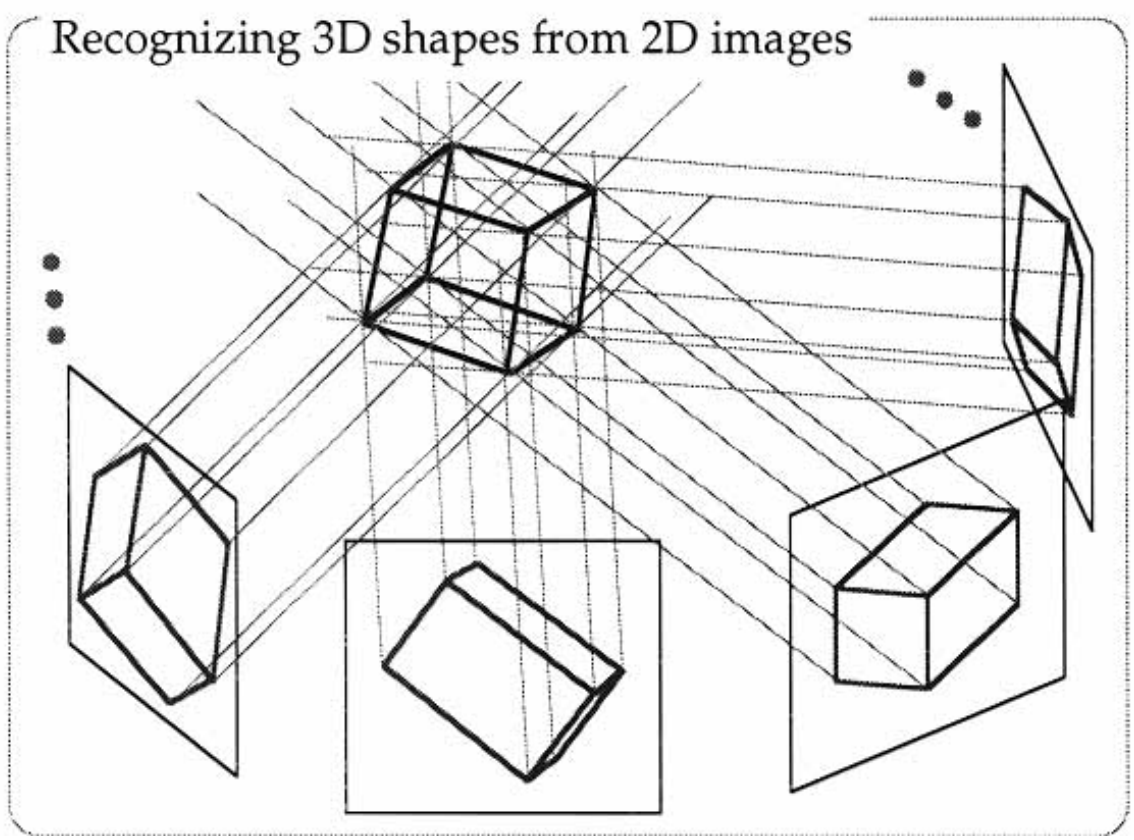
A few results



A few results

Courtesy: Malik et al, 2001

The challenge of shape-based recognition



A 3D object can look very different from different view-points

Two basic issues:

1. How are 3D objects represented?

2. What is the nature of processing underlying recognition?

Two basic issues:

1. How are 3D objects represented?

   -as 3D models
   -as collections of 2D views

2. What is the nature of processing underlying recognition?

   -feedforward
   -iterative via feedback

Some background material:

Terminology

Objects

Terminology: What exactly do we mean by the term ‘recognition’?

In normal usage, recognition refers to categorizing an object as an instance of a particular object class.
e.g. chair, dog, tree.

This is called ‘basic level’ or ‘entry level’ recognition.

Basic-level recognition is believed to be easier and faster than…

… subordinate level recognition

e.g. kitchen chair, dachshund, Oak

Which is easier than…

… superordinate level recognition

e.g. furniture, mammal, plant

Open issues:

What might be the reason for the primacy of entry-level recognition?

Are objects that we are intimately familiar with, still recognized first at the basic level?

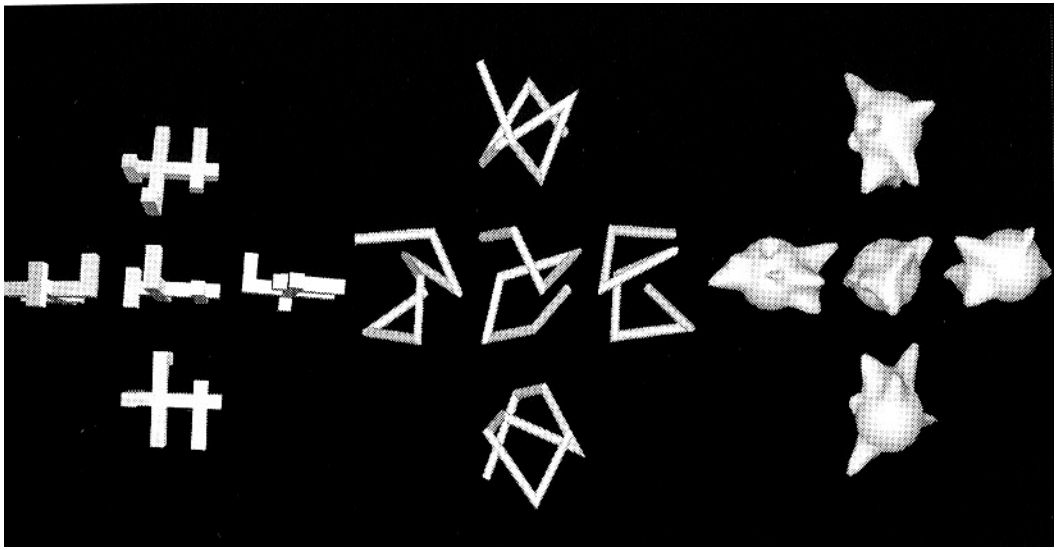
Having defined 'recognition' what kinds of object stimuli should we use to assess it?

Objects used as stimuli to investigate recognition

Familiar, everyday objects:

Pertinent to real world tasks but difficult to control subjects' prior exposure to the objects. Also difficult to systematically study the influence of specific attributes.

Unfamiliar objects:



Some background material:

Terminology

- -Basic level
- -Subordinate
- -Superordinate

Objects

- -Familiar
- -Unfamiliar

Two basic issues:

1. How are 3D objects represented?

   -as 3D models
   -as collections of 2D views

2. What is the nature of processing underlying recognition?

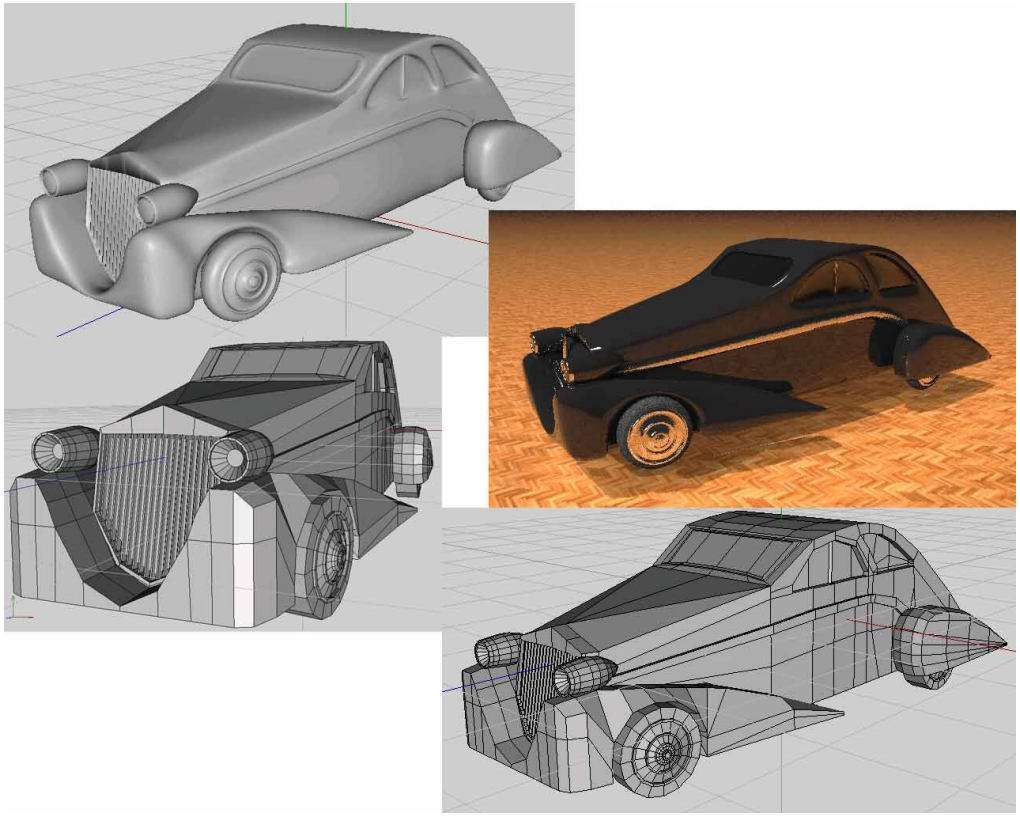
   -feedforward
   -iterative via feedback

Representing 3D objects via CAD like 3D models

Advantages:

Low memory requirements

Invariance over various transformations (view, lighting…)



A prominent proposal: Marr and Nishihara (1978)

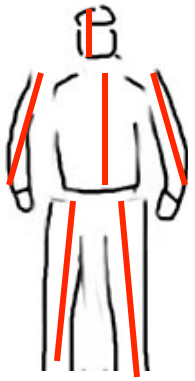
Marr and Nishihara's model [1978]

Early processing

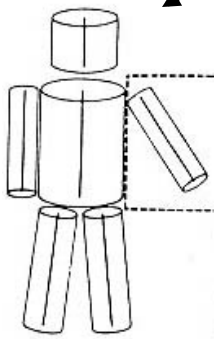
Figure-ground segregation; Edge extraction

Part segmentation

Axis estimation



Volumetric modeling



3D Object centered structure

3D models

Match?

Yes/no

Representing 3D objects via 3D models

Advantages:

Low memory requirements

Invariance over various transformations (view, lighting…)

Shortcomings:

3D models hard to construct from 2D images

Complete invariance over various transformations not observed in actual experiments.

Irving Biederman proposed a recognition scheme to address these shortcomings…

Irving Biederman

Objects are represented by the HVS as collections of geometric primitives ('geons'). 2D cues in an image specify which geons constitute the object.

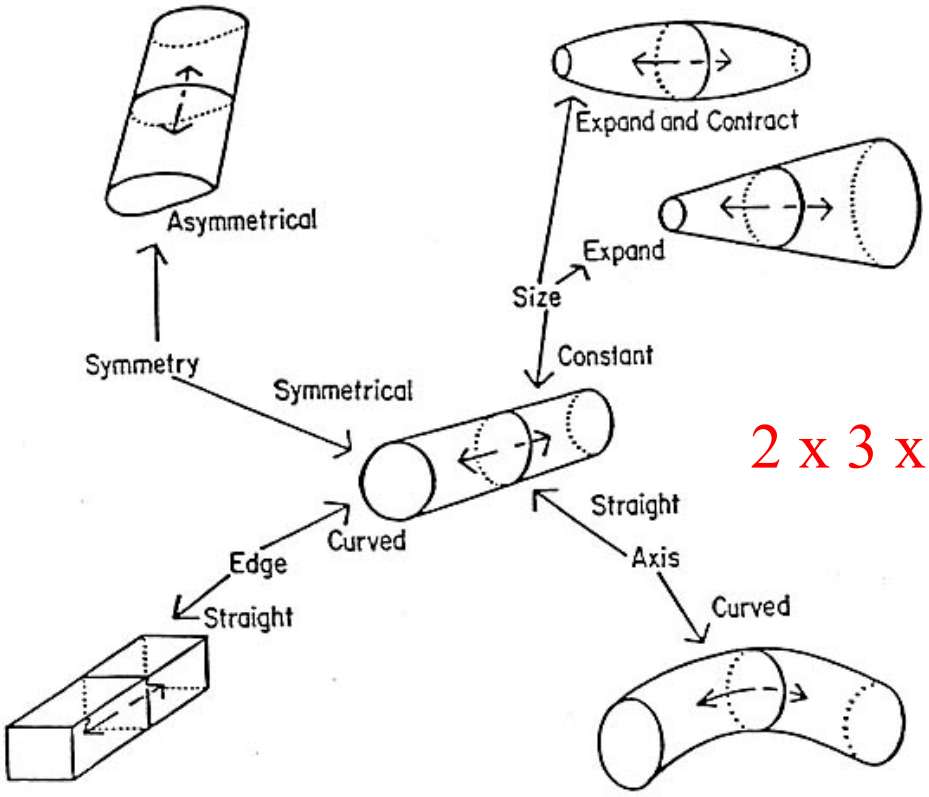
Object representation: Geons and their qualitative mutual relationships

Since the geons intuitively correspond to the 'parts' of an object...

…Biederman's theory is called Recognition by Components

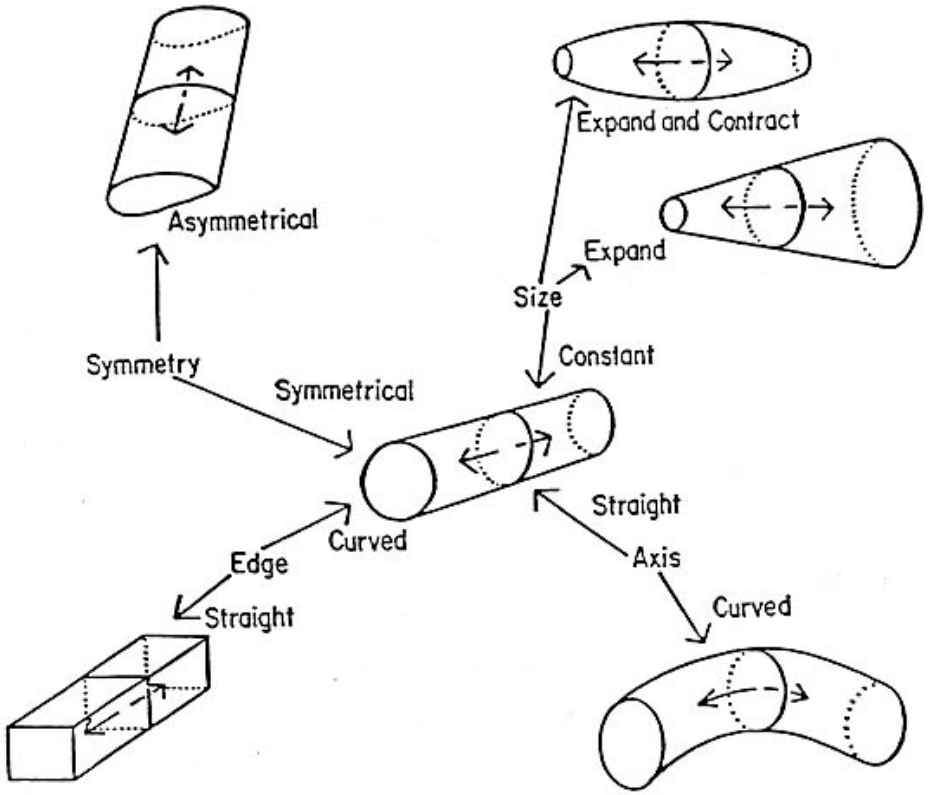
What primitives does Biederman propose as 'geons'?

Geons are generalized cylinders where the cross-section can vary over the length of the axis, which itself may not be straight.

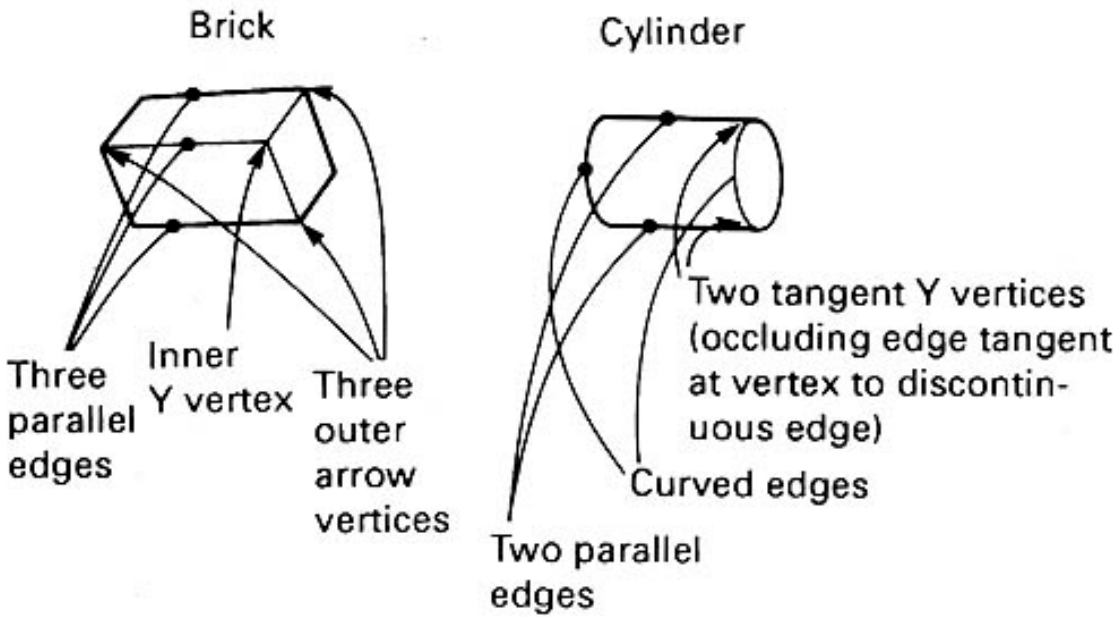


2 x 3 x 2 x 2 = 24 geons

But, what's special about these geons?

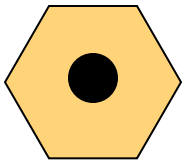


The existence of these geons can be hypothesized using 2D features in an image…



'Invariant' features for two geons that are evident in the 2D image

Views where these invariants are not satisfied are hard to recognize.



(a)

(a) It is difficult to identify the object behind the mask because its geons have been obscured.

(a)

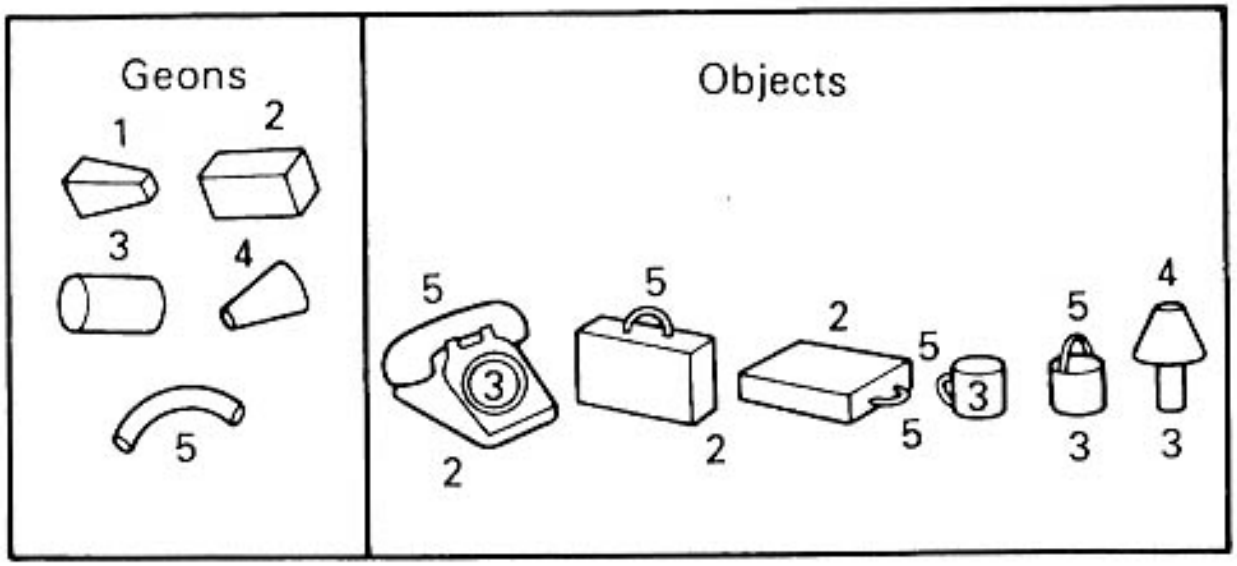
(a) It is difficult to identify the object behind the mask because its geons have been obscured.

(b)

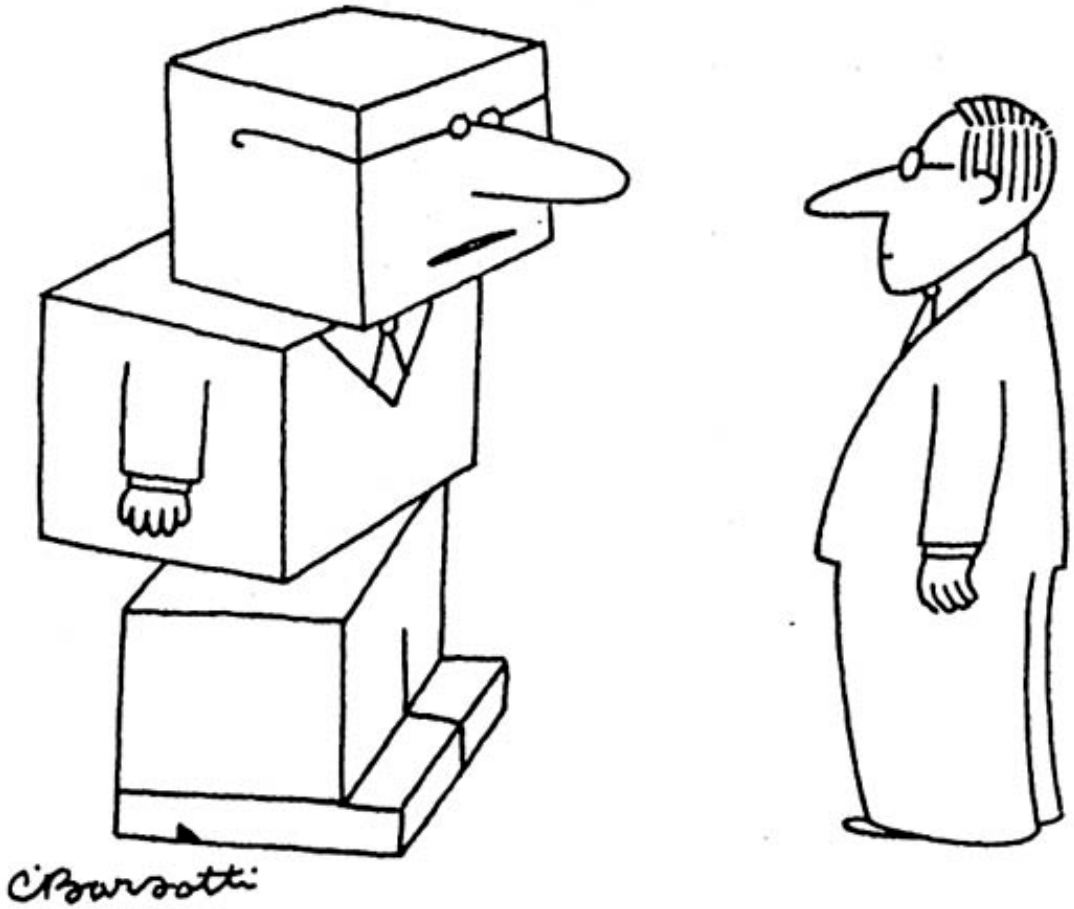
(b) Now that it is possible to identify geons, the object can be identified as a flashlight.

Does it make intuitive sense for the HVS to represent objects as collections of geons?

Biederman: "Yes, geons constitute a versatile vocabulary. Different compositions of the geons can be used to create various common objects"



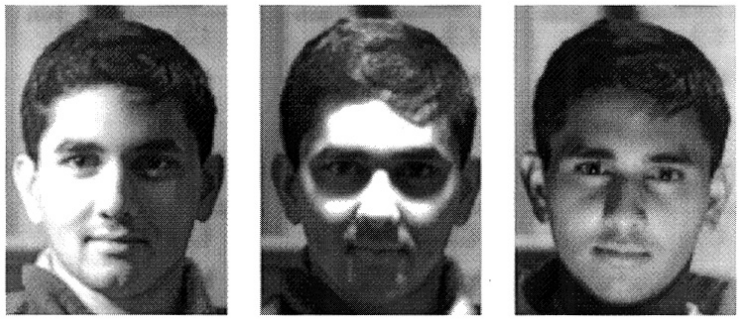
But, is the real-world truly amenable to a geon based representation?

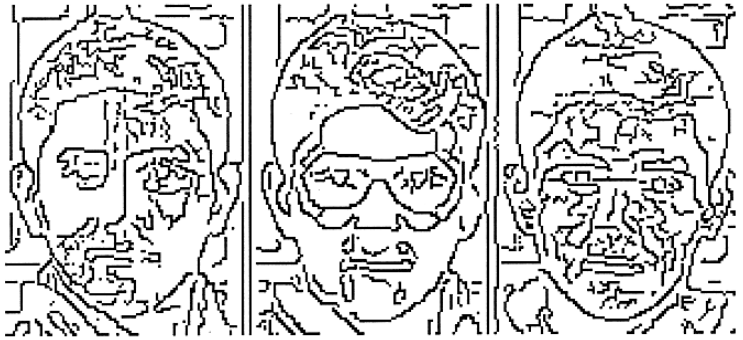


"We're offering you the job on probation, Whitlock. You have three months to become one of us."

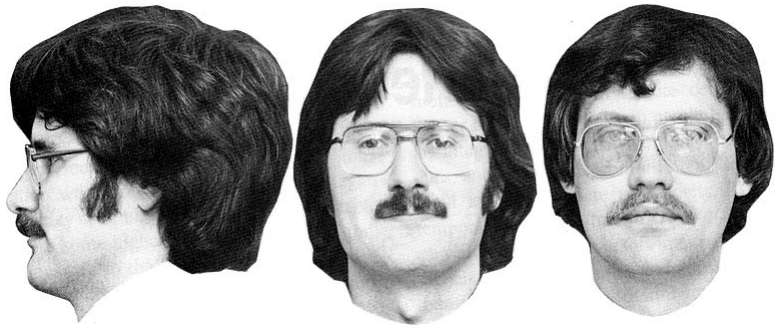
Biederman's Recognition By Components (RBC) theory:

Potential concerns

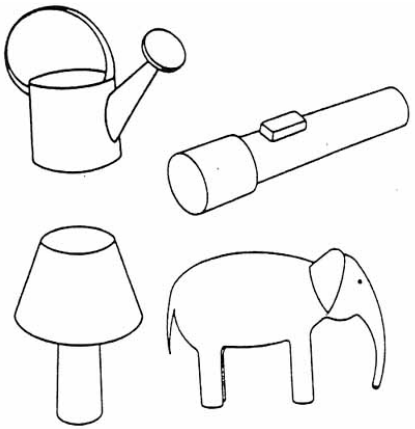




1. Invariant geon features difficult to extract in real images

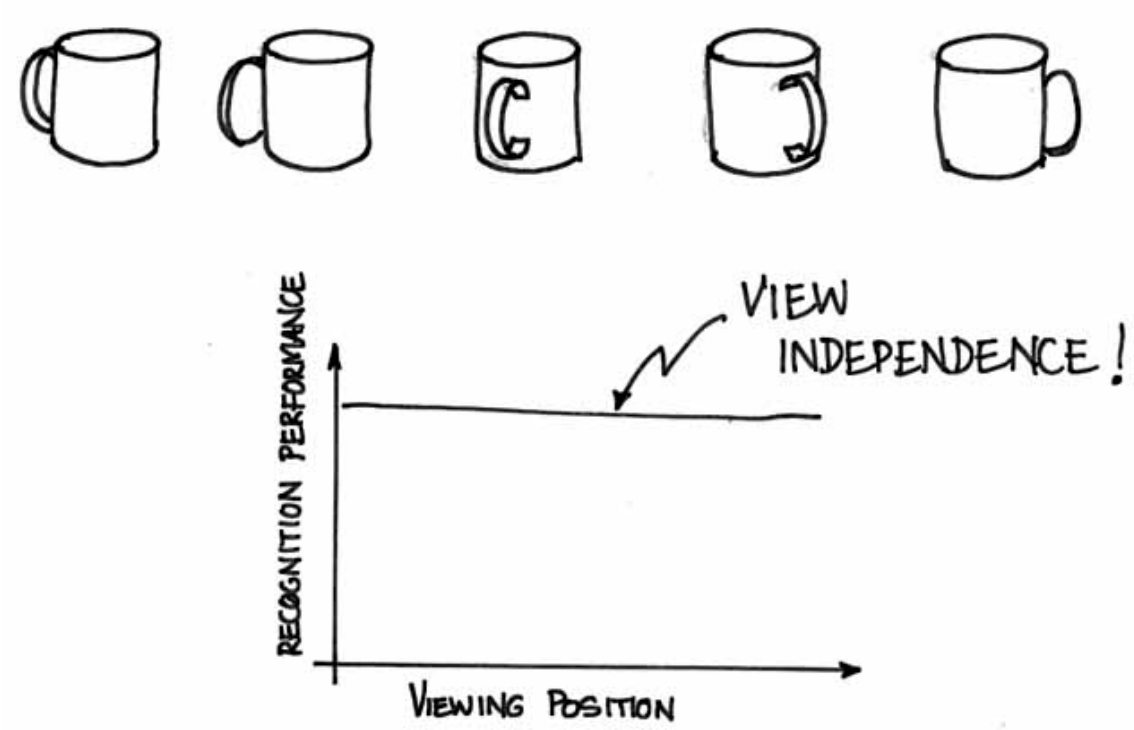


2. RBC does not address subordinate level recognition



3. Geons do not appear well-suited to representing many natural objects which may not have simple parts-based descriptions.

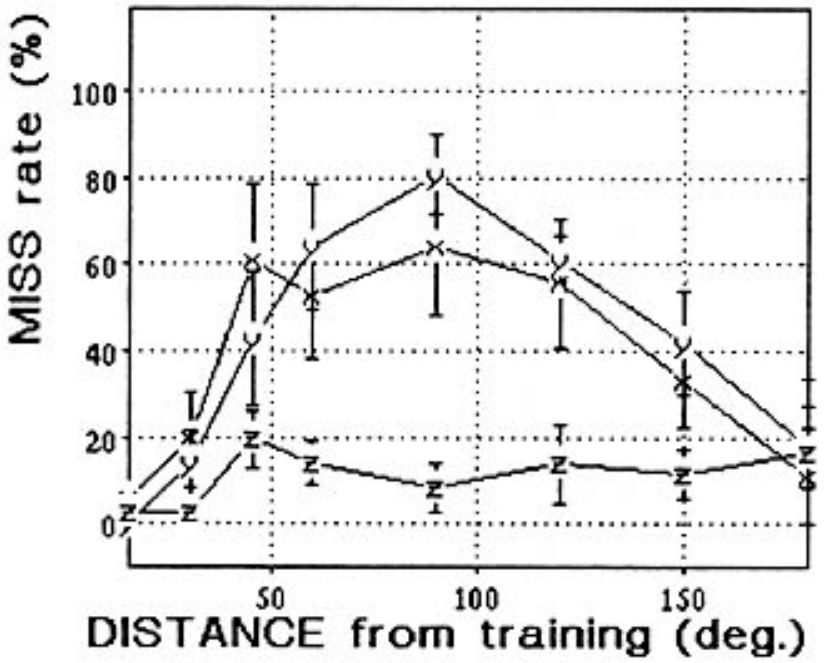
Piecewise invariance – a prediction of the RBC theory:



Many different views lead to the same set of geons and their relations. Therefore, the model predicts **<u>piecewise view invariance</u>**.

Is this prediction supported by actual experimental data?

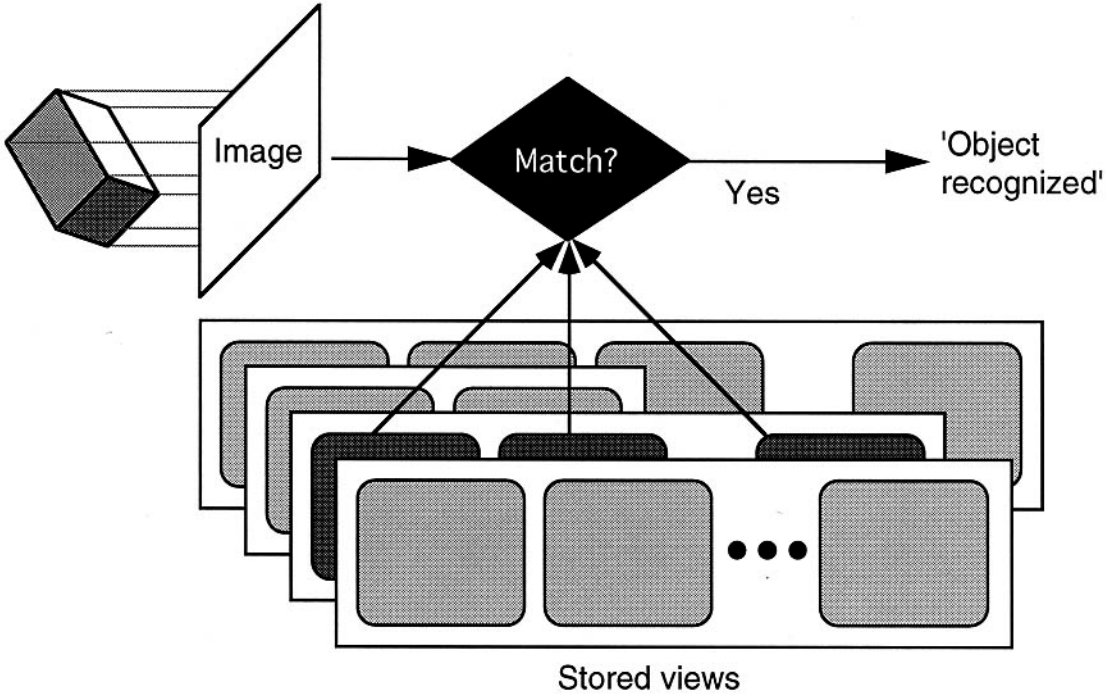
A more fundamental problem with Biederman's model:



3D object recognition is often not view-point independent…

A different proposal: View-based recognition scheme

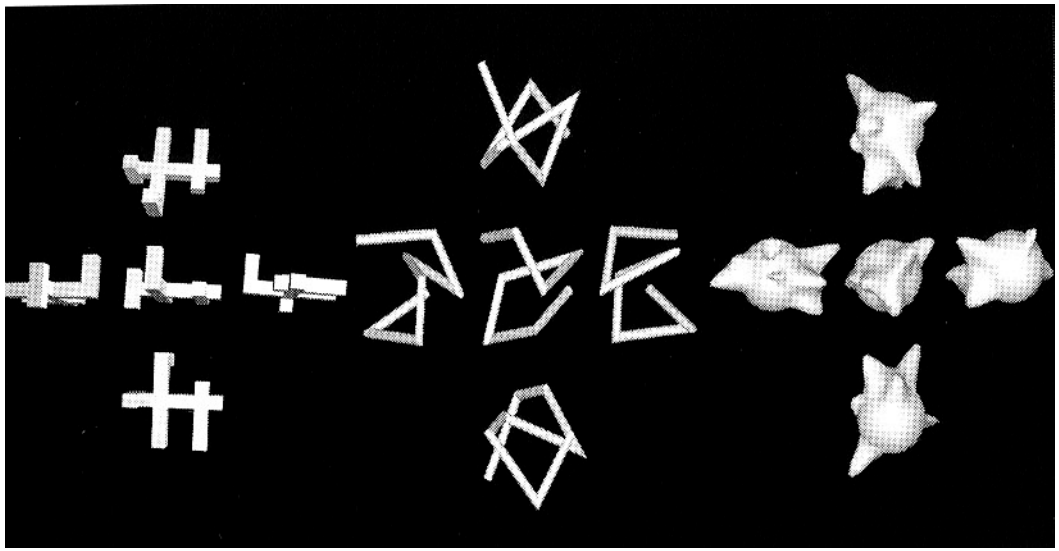
(a nearest-neighbors strategy using object views)



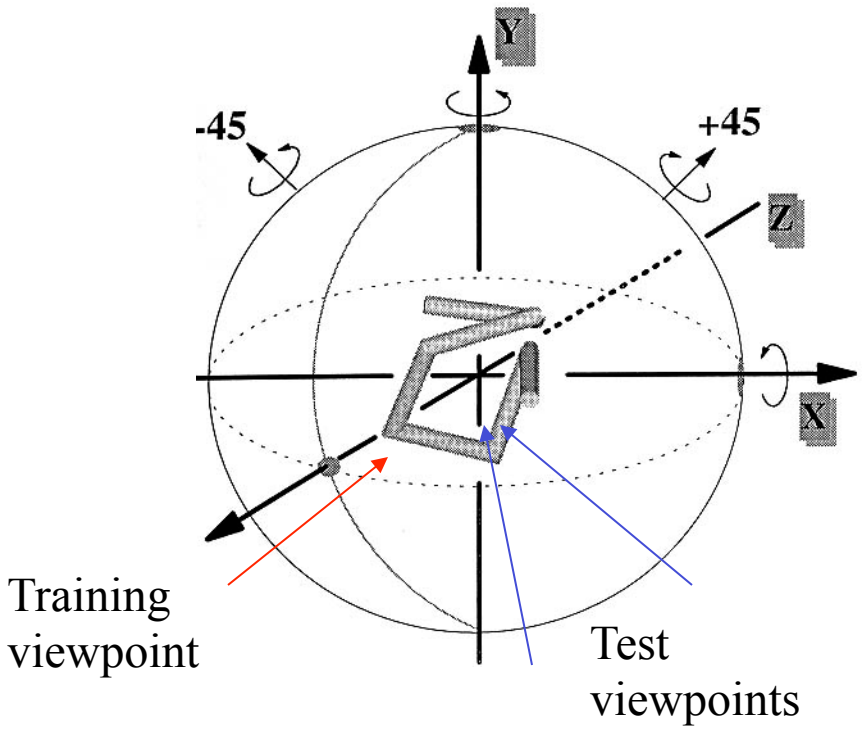
Each view has a limited generalization field.
Prediction: Very little invariance to transformations

Is this prediction supported by actual experimental data?

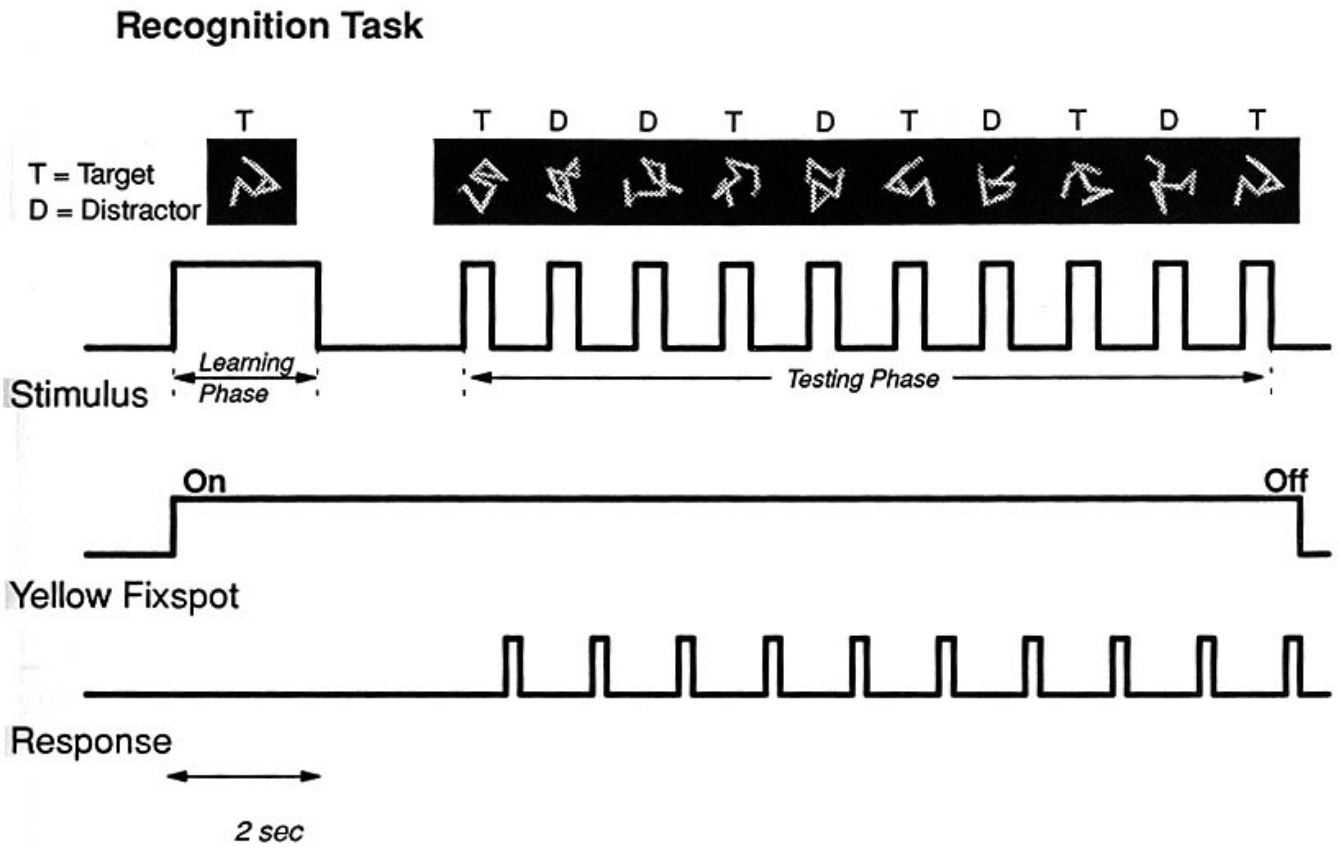
Assessing recognition performance as a function of viewpoint



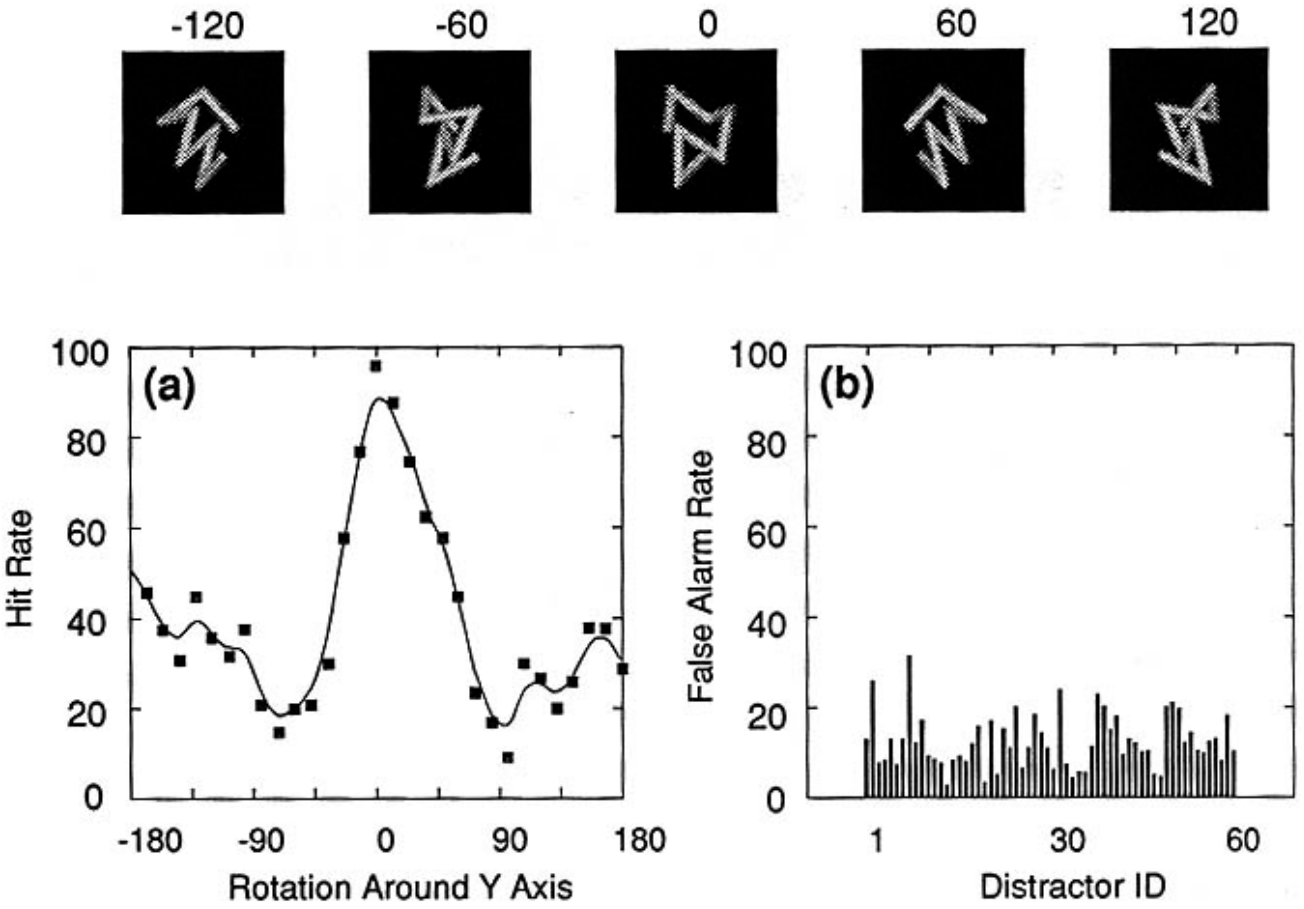
Experimental stimuli


Y
-45
+45
Z
X
Training
viewpoint
Test
viewpoints

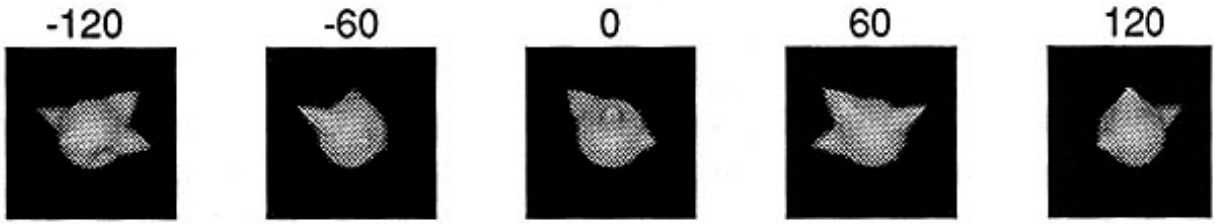
Experimental paradigm

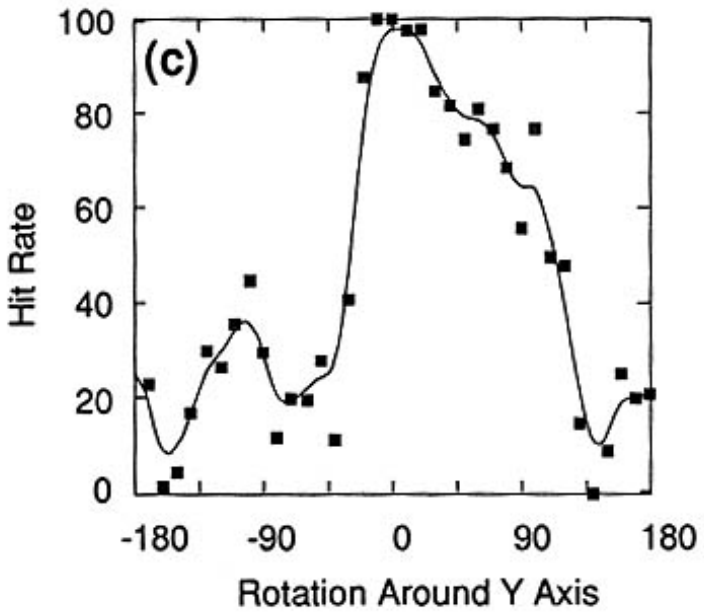


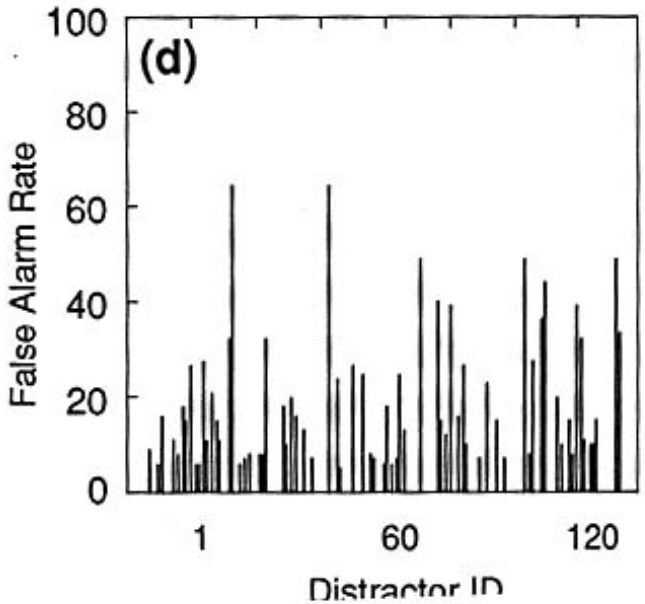
Psychophysical results



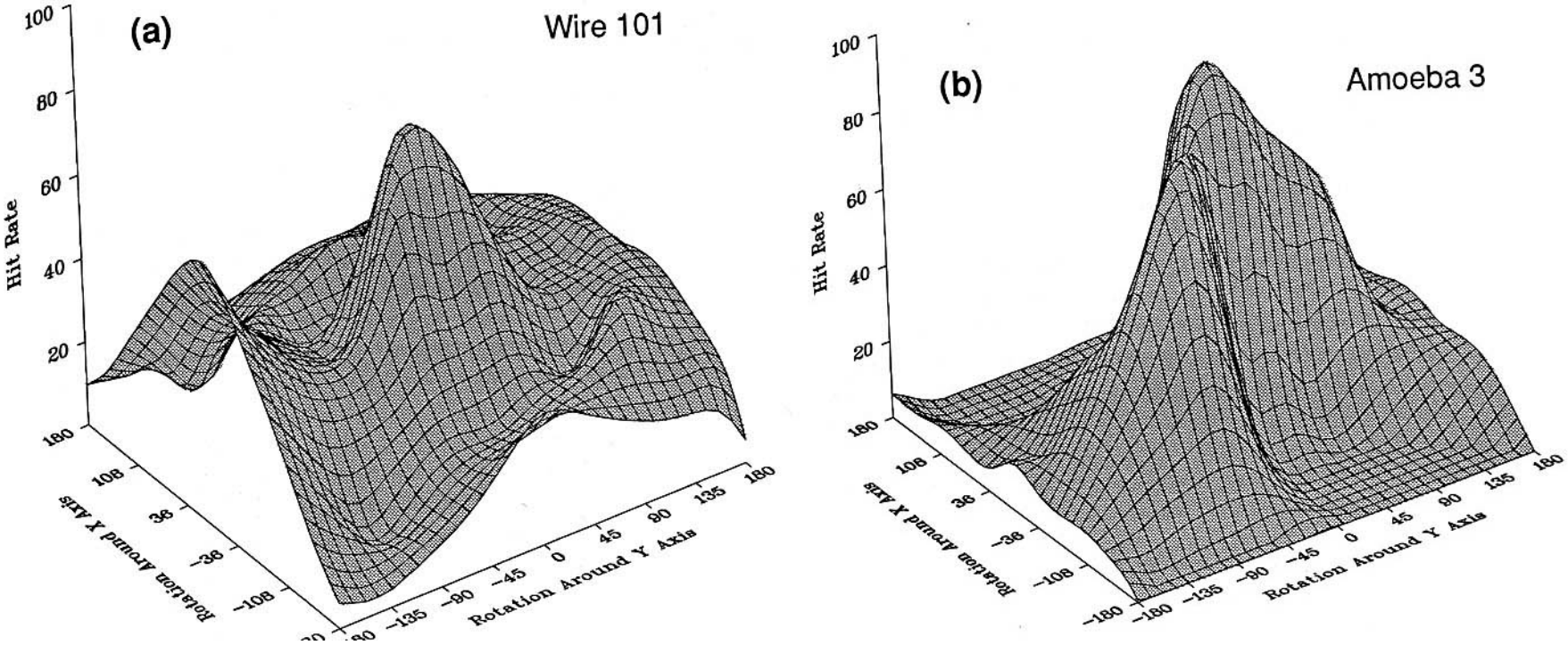
Psychophysical results







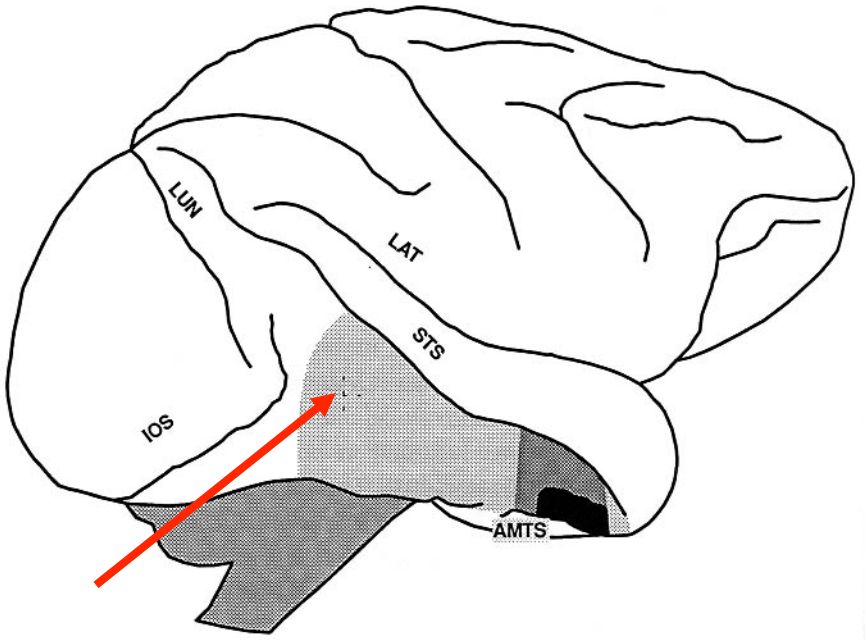
Psychophysical results



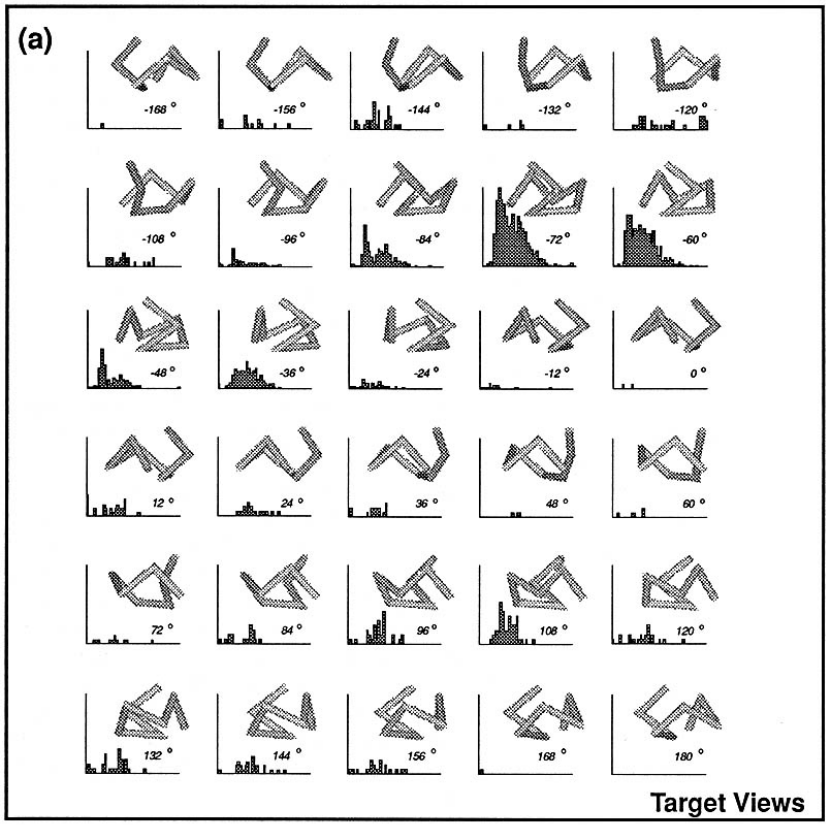
What's happening at the individual neuronal level?

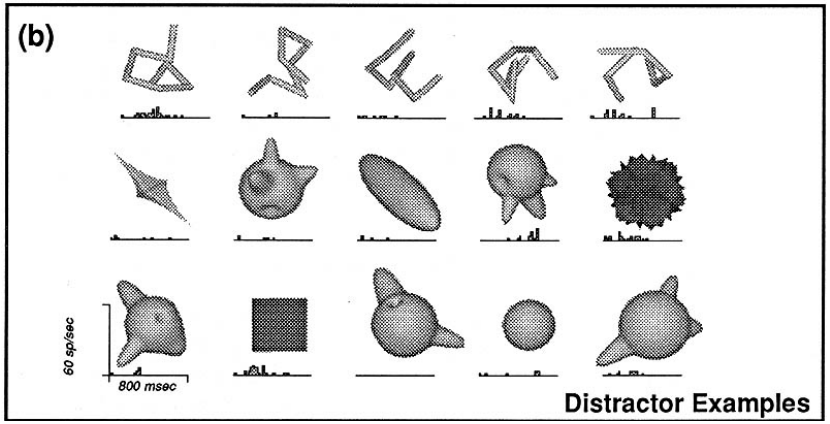
Physiological results

(Logothetis et al, 1995)

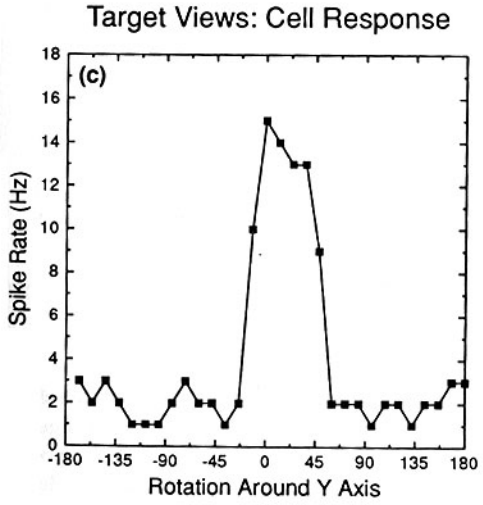


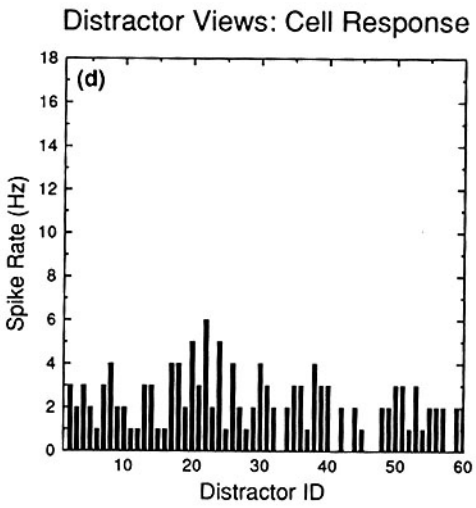
Physiological results



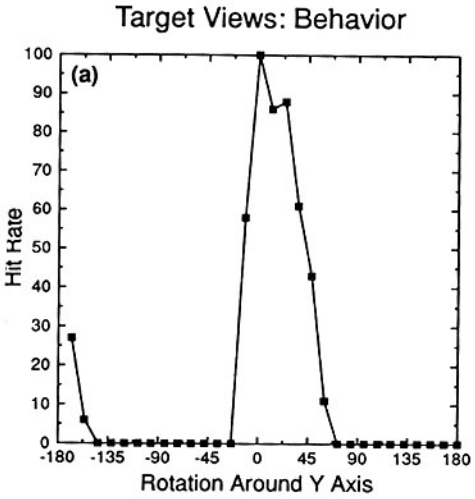


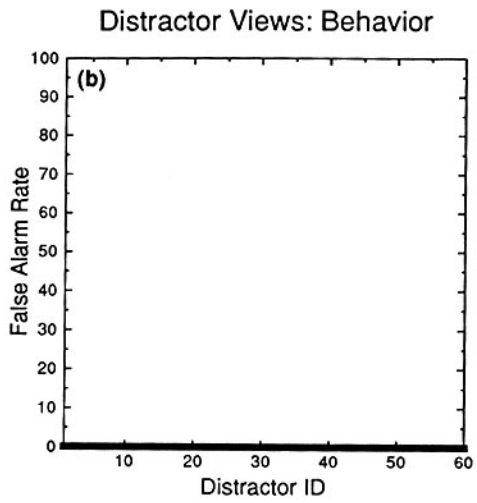
Physiological results

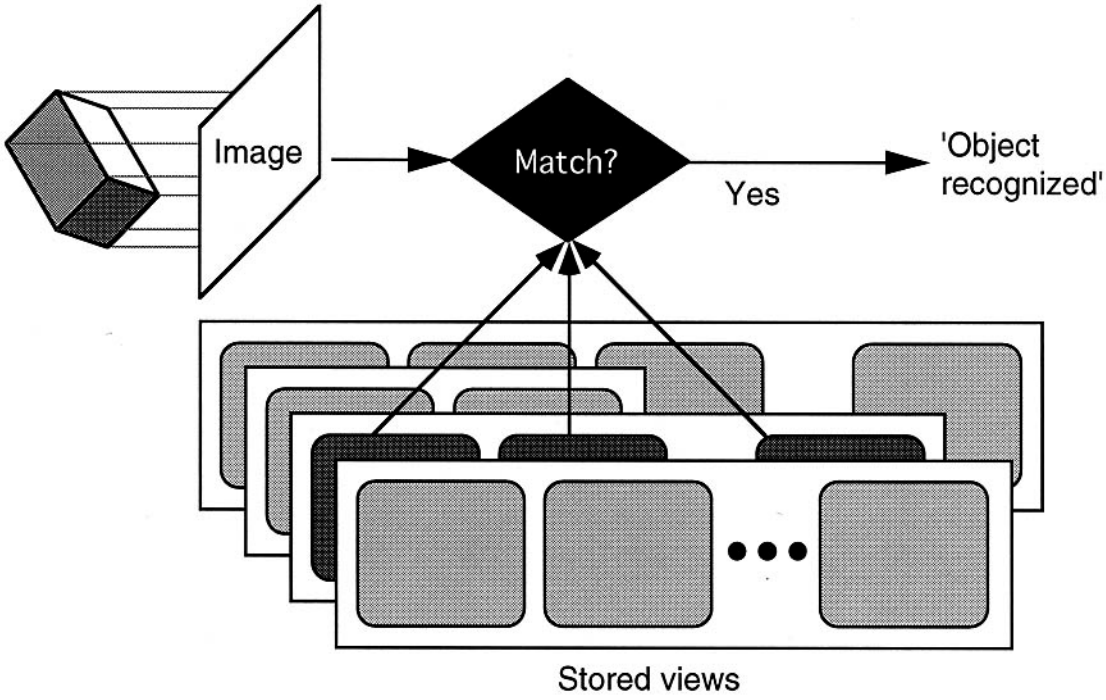




Behavioral results

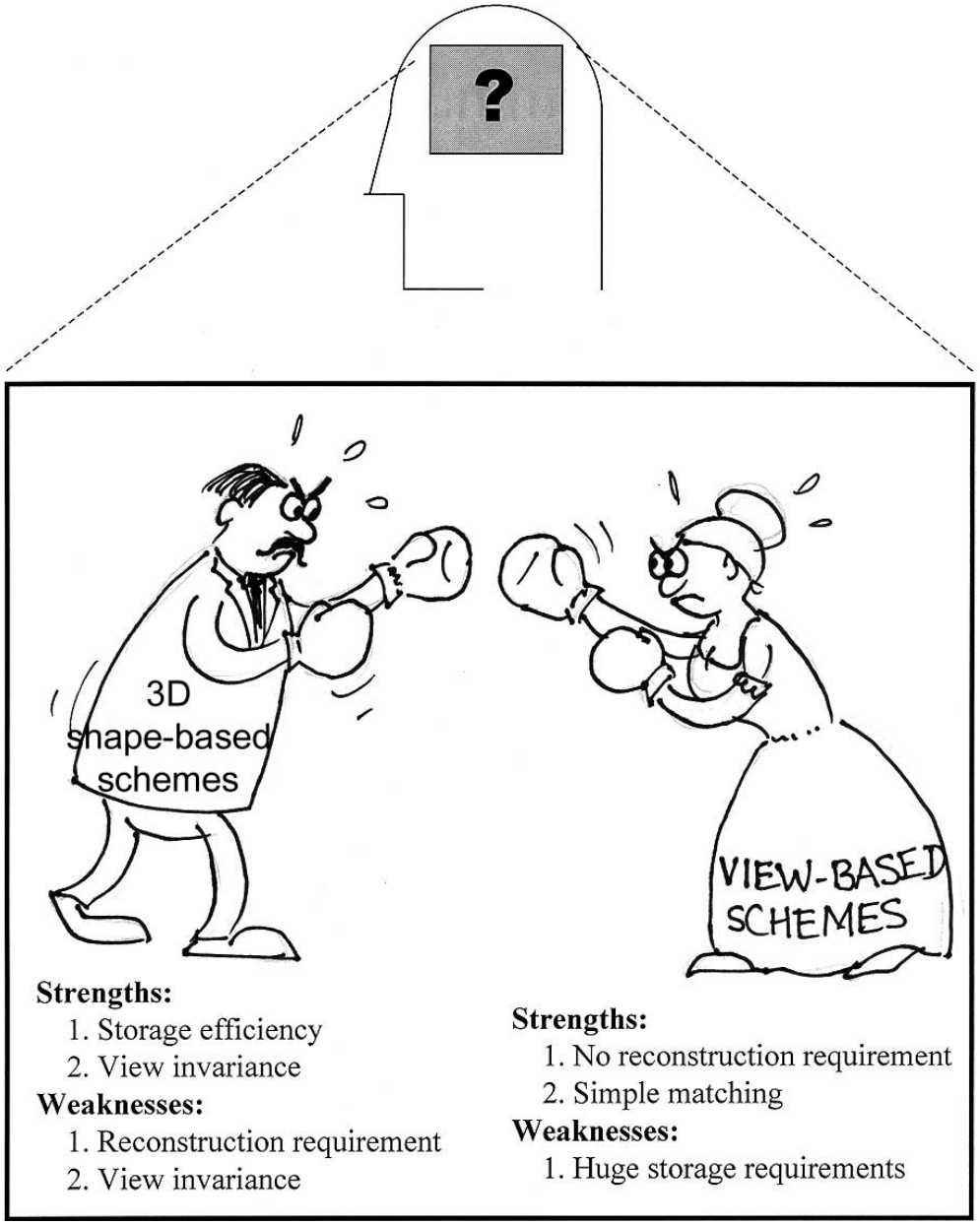






View-based recognition scheme
appears to be supported by experimental results
(somewhat like desert ants and drosophila)

Summary

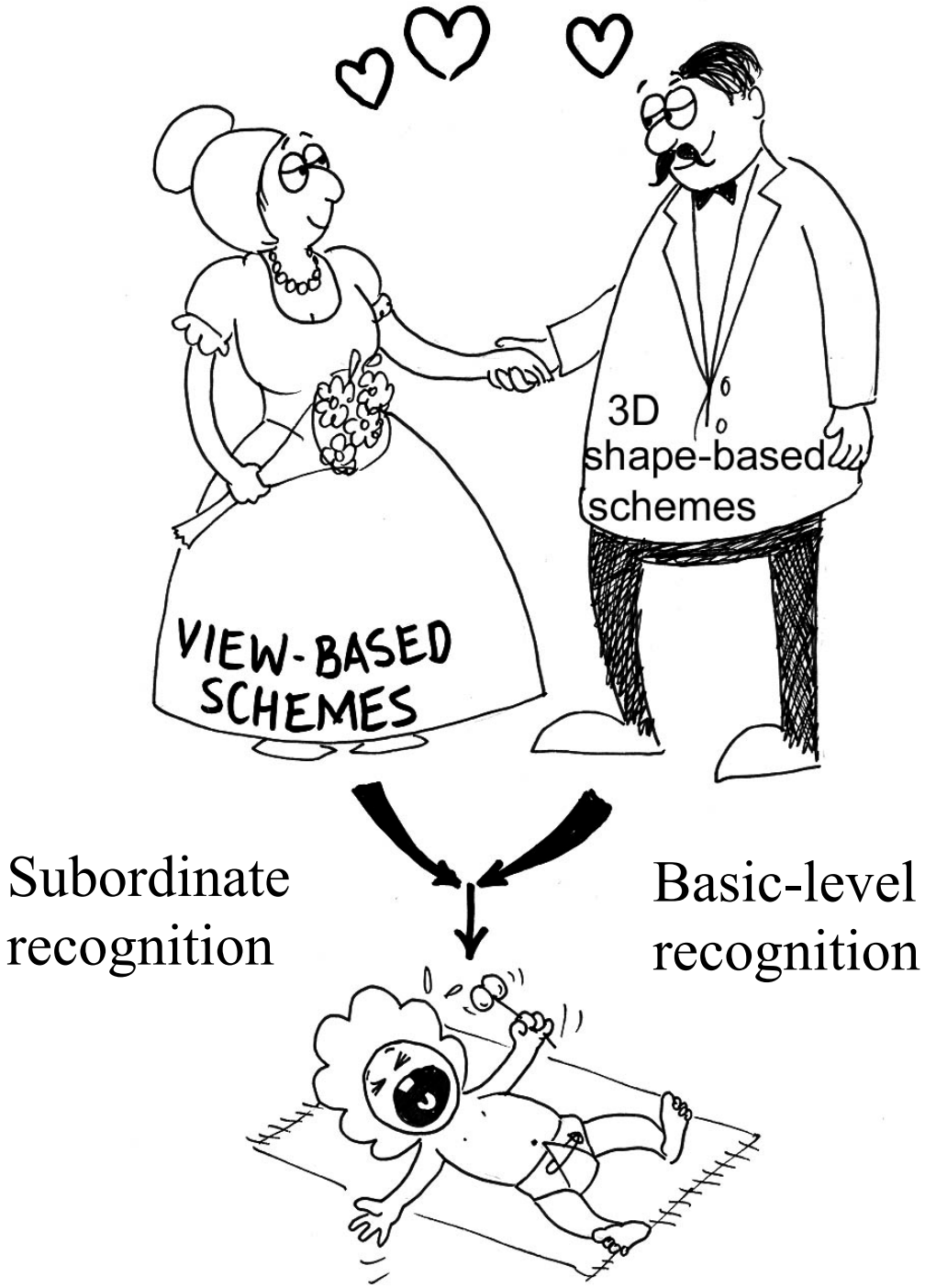


Strengths:
1. Storage efficiency
2. View invariance

Weaknesses:
1. Reconstruction requirement
2. View invariance

Strengths:
1. No reconstruction requirement
2. Simple matching

Weaknesses:
1. Huge storage requirements


3D
shape-based
schemes
VIEW-BASED
SCHEMES
Subordinate
recognition
Basic-level
recognition

Two basic issues:

1. How are 3D objects represented?

   -as 3D models

   -as collections of 2D views

2. What is the nature of processing underlying recognition?

   -feedforward

   -iterative via feedback

How can we make inferences about the flow of information within a 'black-box'?

How can we make inferences about the flow of information within a 'black-box'?

Speed of processing may yield some indirect clues.

Here is what we know:

1. The maximum firing rate of a cortical neuron is approx. 100 Hz. i.e. an action potential every 10 ms. Thus, we need at least 10ms to determine the rate of firing of a neuron.

2. Latency of V1 neurons is about 70 ms.

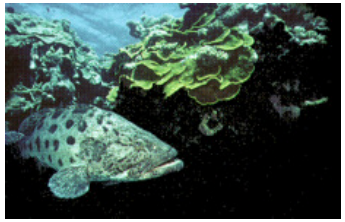
If we observe evidence of object recognition in neural responses after x ms, then the number of intervening stages I is $(x - 70)/10$.

If x (and therefore I) is small, then the processing is likely to be largely feedforward. Otherwise, there may be scope for iterative computations.

So, what is x?

The speed of processing of the human visual system

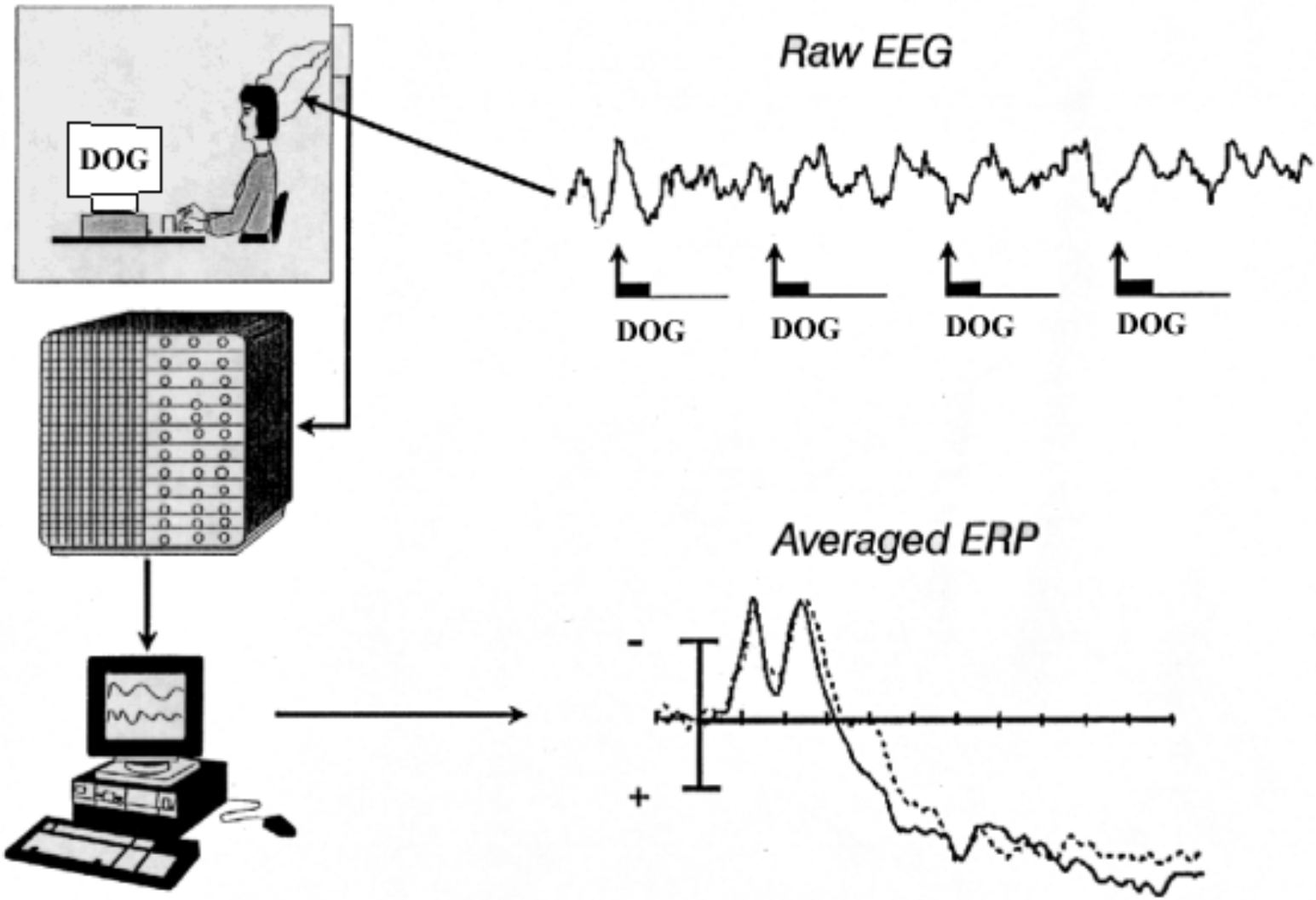
Thorpe et al. [1996] Nature

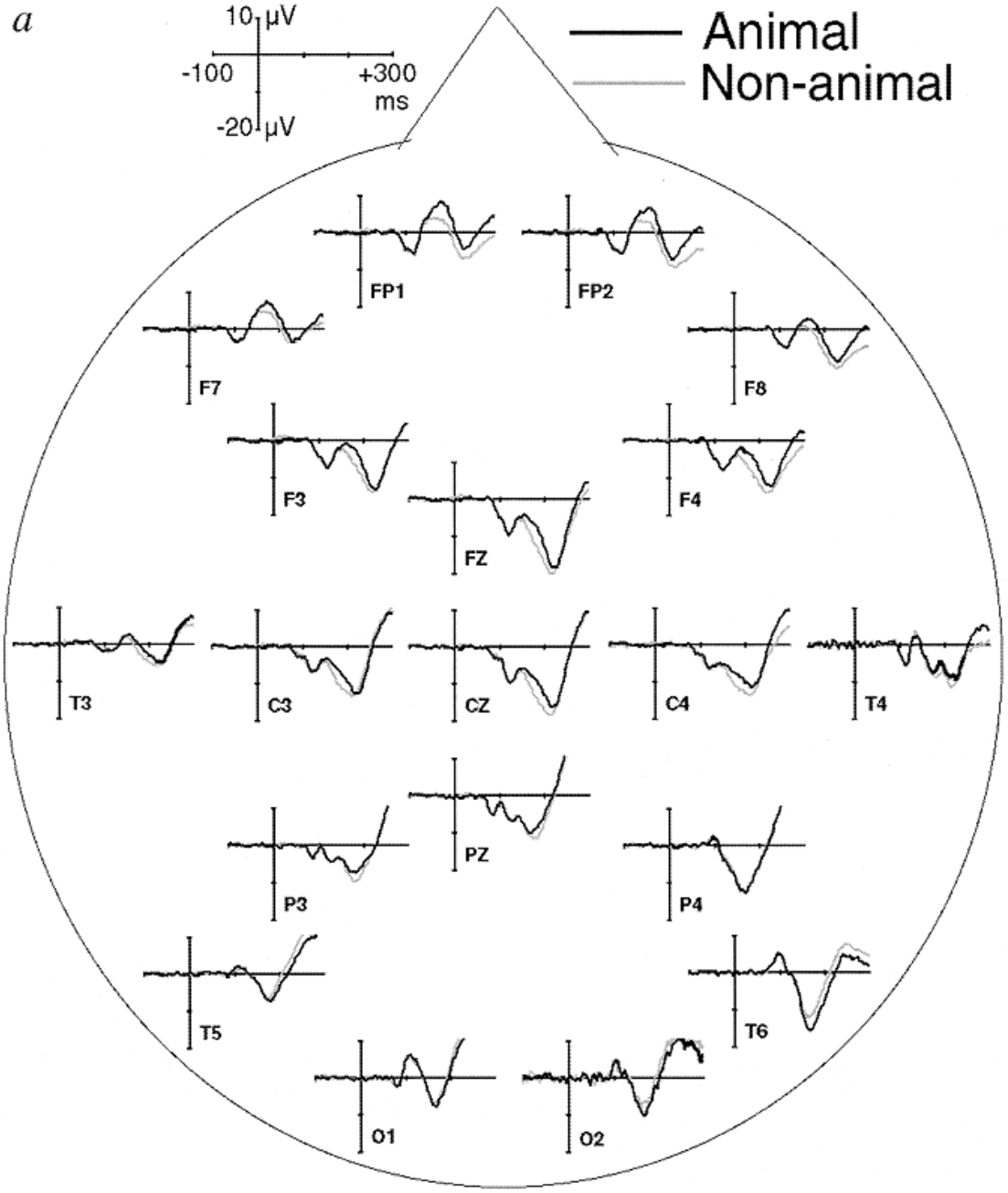



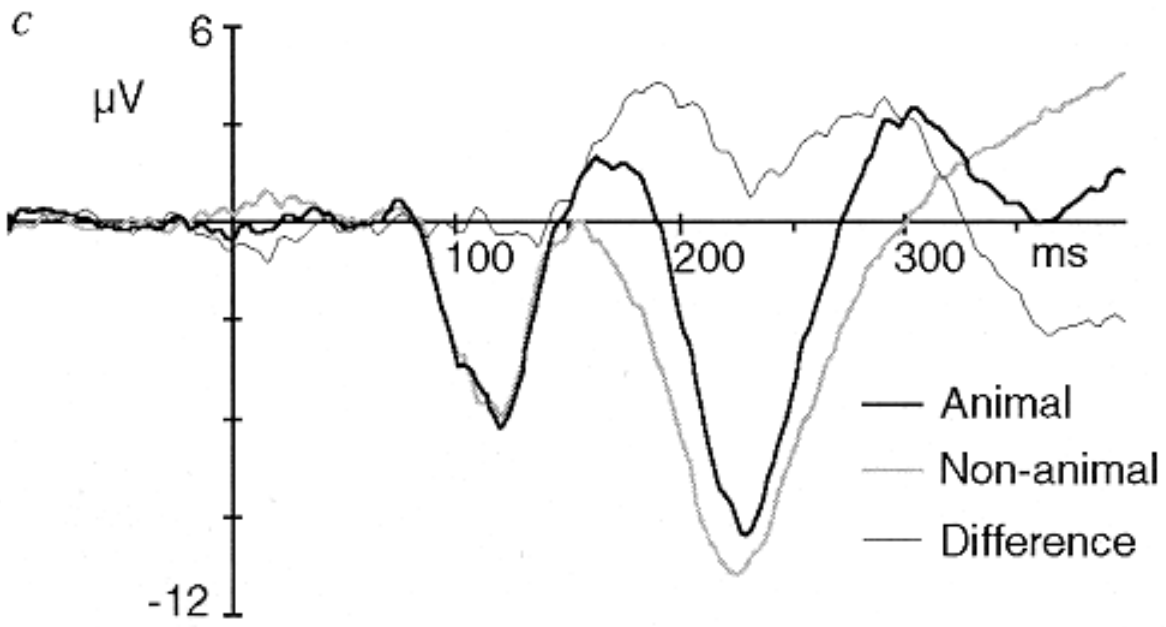

•••

Animal / non-animal images shown to subjects

Event-Related Potential Technique




a
10 µV
-100
+300
ms
-20 µV
Animal
Non-animal
FP1
FP2
F7
F8
F3
F4
FZ
T3
C3
CZ
C4
T4
P3
PZ
P4
T5
T6
O1
O2



There is evidence of image discrimination by 150 ms.

Thus, the likely number of intervening processing stages is (150 – 70)/10 = 8.

Given that there are about 5 neuronal links between V1 and the frontal cortex, this suggests feedforward processing.

Two basic issues:

1. How are 3D objects represented?

primarily as collections of 2D views

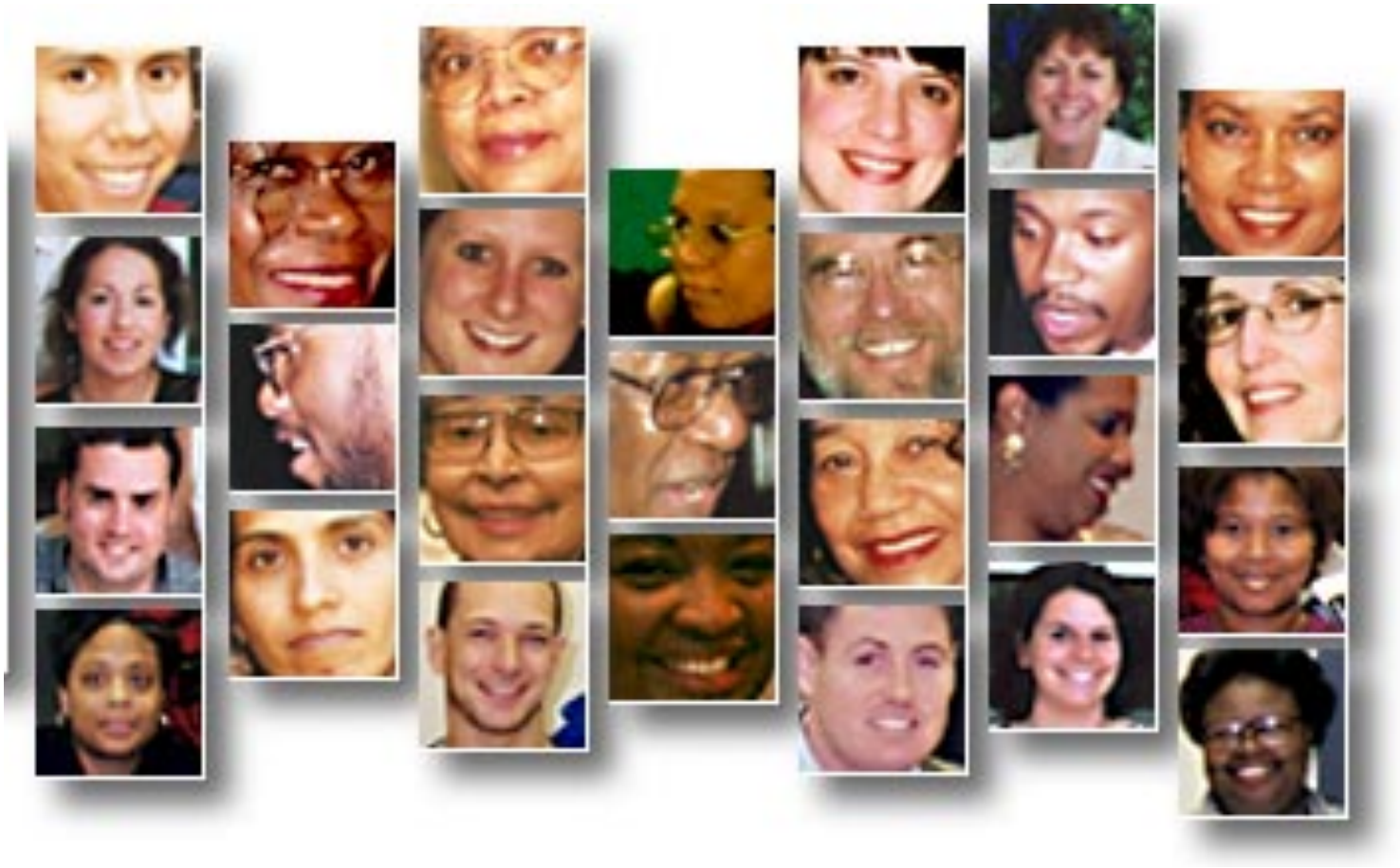
2. What is the nature of processing underlying recognition?

primarily feedforward

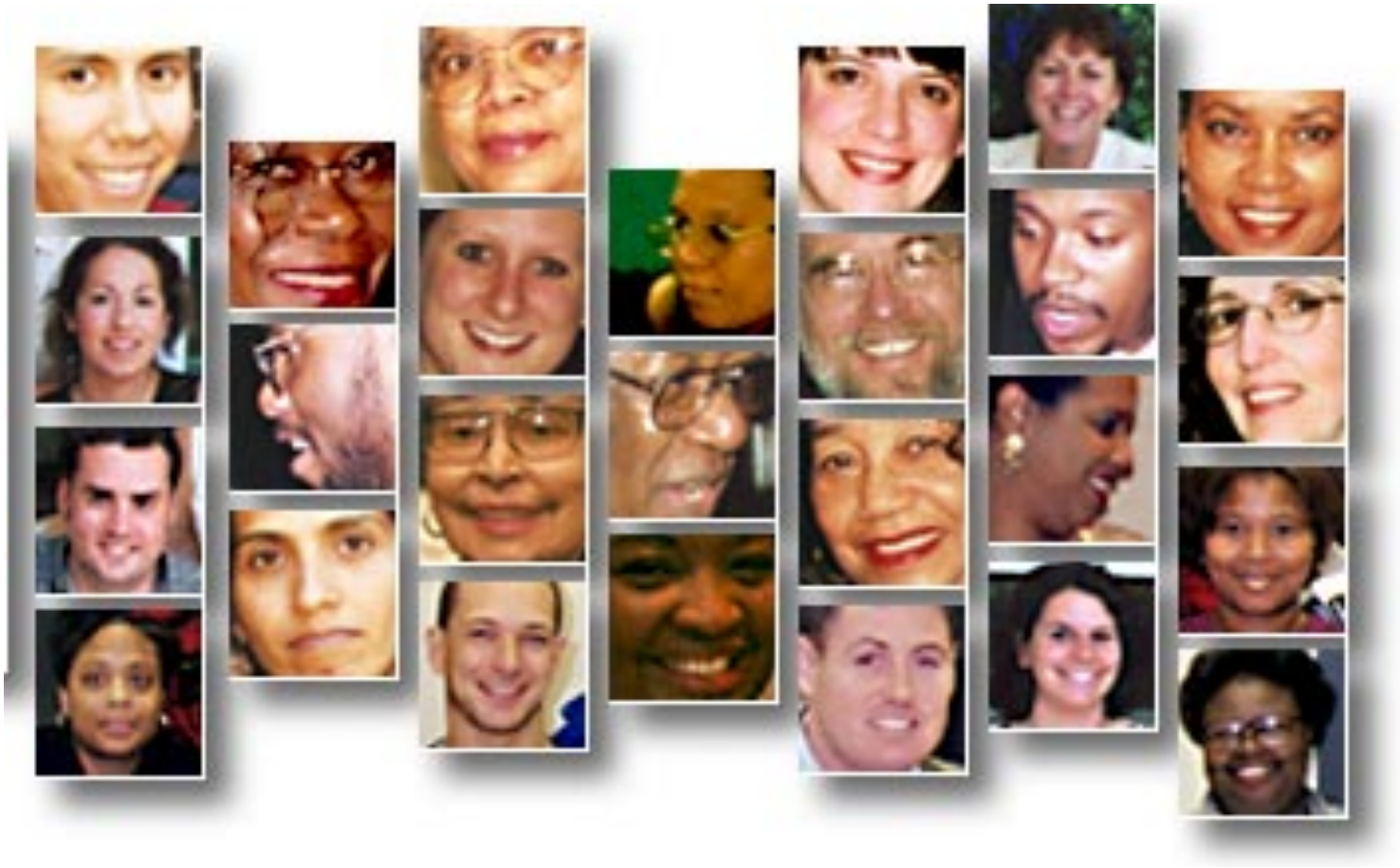
These are tentative answers. The questions are still largely open.

Faces recognition

an important sub-domain of object recognition



Faces: A special class of objects?



‘Special’ how?

Does the class of faces differ intrinsically from other objects?

- Ubiquitous and ecologically important

- Require many unique kinds of assessments
(age, identity, beauty, gender, emotion, personality)

- Require sensitivity to minor variations.

Herman Chernoff, 1973, “The Use of Faces to Represent Points in k-Dimensional Space Graphically"

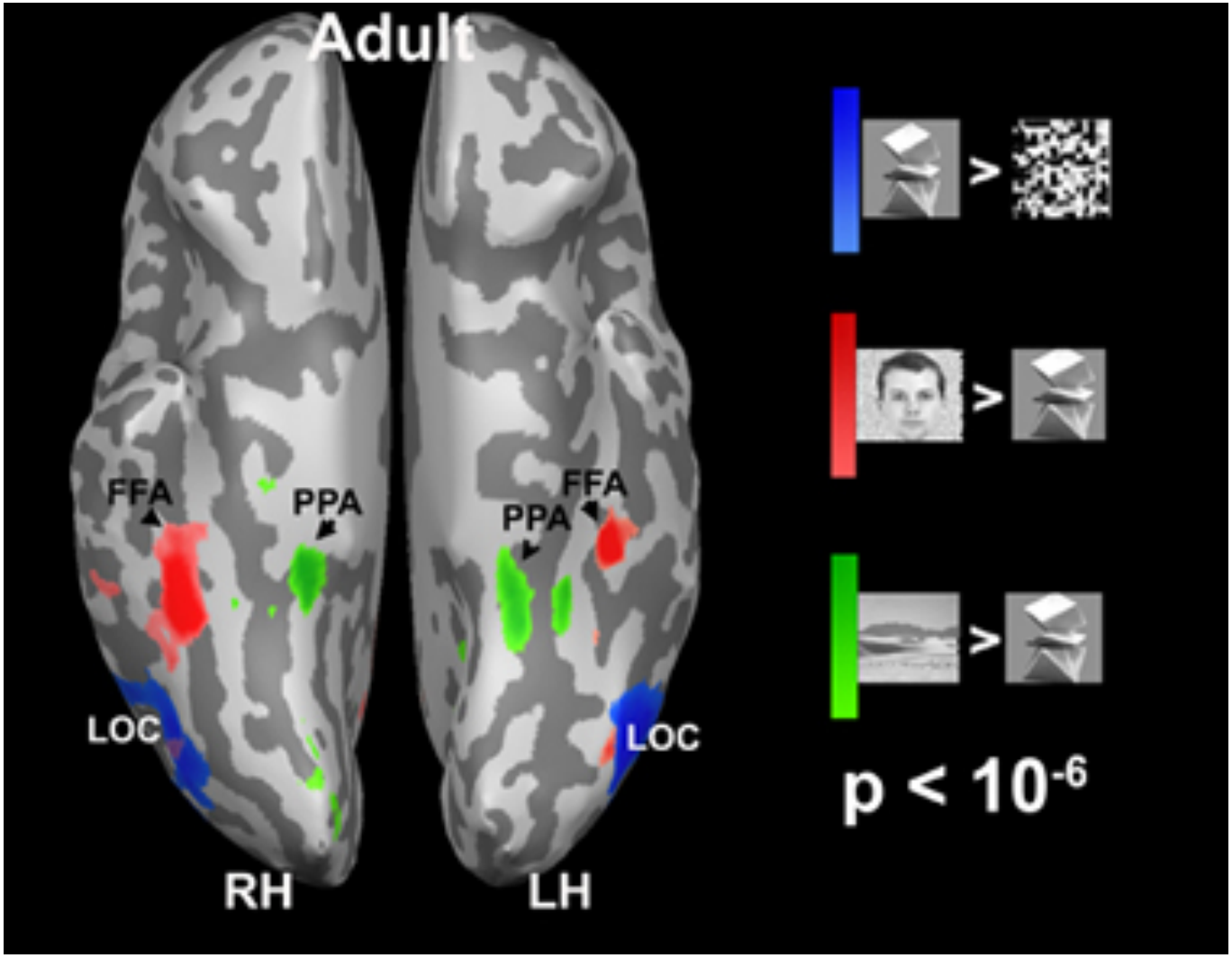
‘Special’ how?

Does the brain use special strategies for analyzing faces?

Suggestive evidence:

1. Distinct anatomical loci for face processing
2. Innate sensitivity to faces

fMRI evidence for distinct anatomical loci for faces



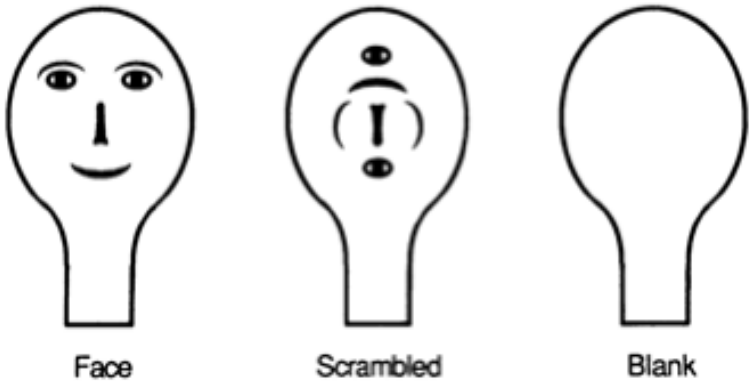
Evidence of innate face sensitivity from babies

Premise: Neonates have essentially no visual experience. If they exhibit face preferences, then the brain must come prewired for face processing




Babies seem to like looking at faces

Babies only days old have preferences (as judged by looking time or stimulus following) for faces and face-like stimuli over many other kinds of stimuli (including inverted faces; Walton et al. 1997, Pascalis et al. 1995).



Johnson et al., 1991

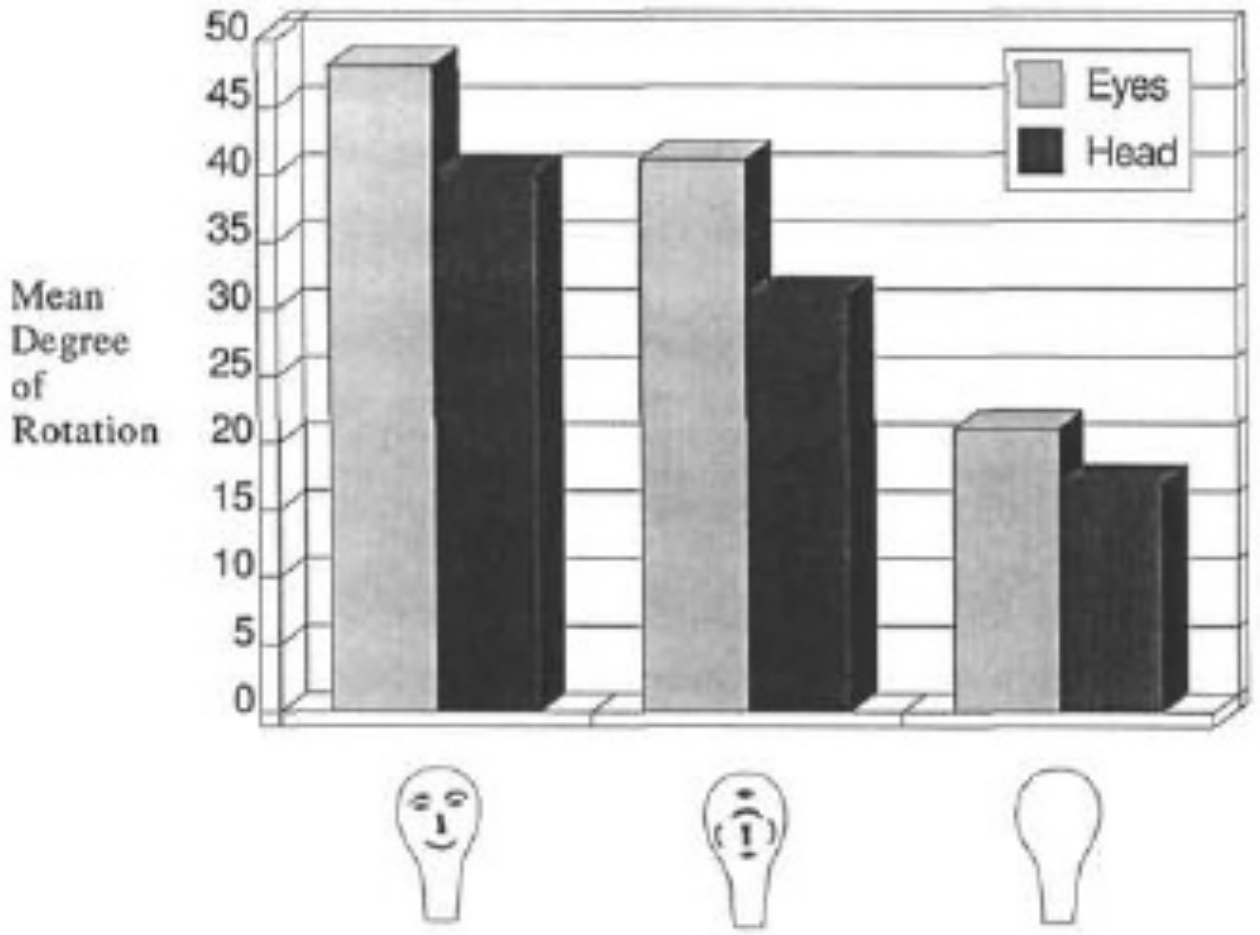


Figure 2. Data from Experiment 1 showing the extent of newborn eye and head turns in following the stimuli in Figure 1 (from Johnson, Dziurawiec, Ellis, & Morton, 1991). (Newborn infants follow the face farther than the other stimuli.)

Babies can imitate facial expressions (Meltzoff & Moore 1977)

Observers blind to purpose of experiment and experimental condition rated film of babies' facial movement (open mouth, tongue out, close mouth)

Babies were much more likely to adopt a facial pose when the experimenter was adopting the pose.

Babies can imitate facial expressions (Meltzoff & Moore 1977)

Babies can imitate facial expressions

Babies can imitate facial expressions

Babies can imitate facial expressions

Babies can imitate facial expressions

Babies can imitate facial expressions

Video

News Front Page

Africa
Americas
Asia-Pacific
Europe
Middle East
South Asia
UK
Business
Health
Science & Environment
Technology
Entertainment

Last Updated: Monday, 6 September, 2004, 11:02 GMT 12:02 UK

E-mail this to a friend Printable version

Newborns prefer beautiful faces

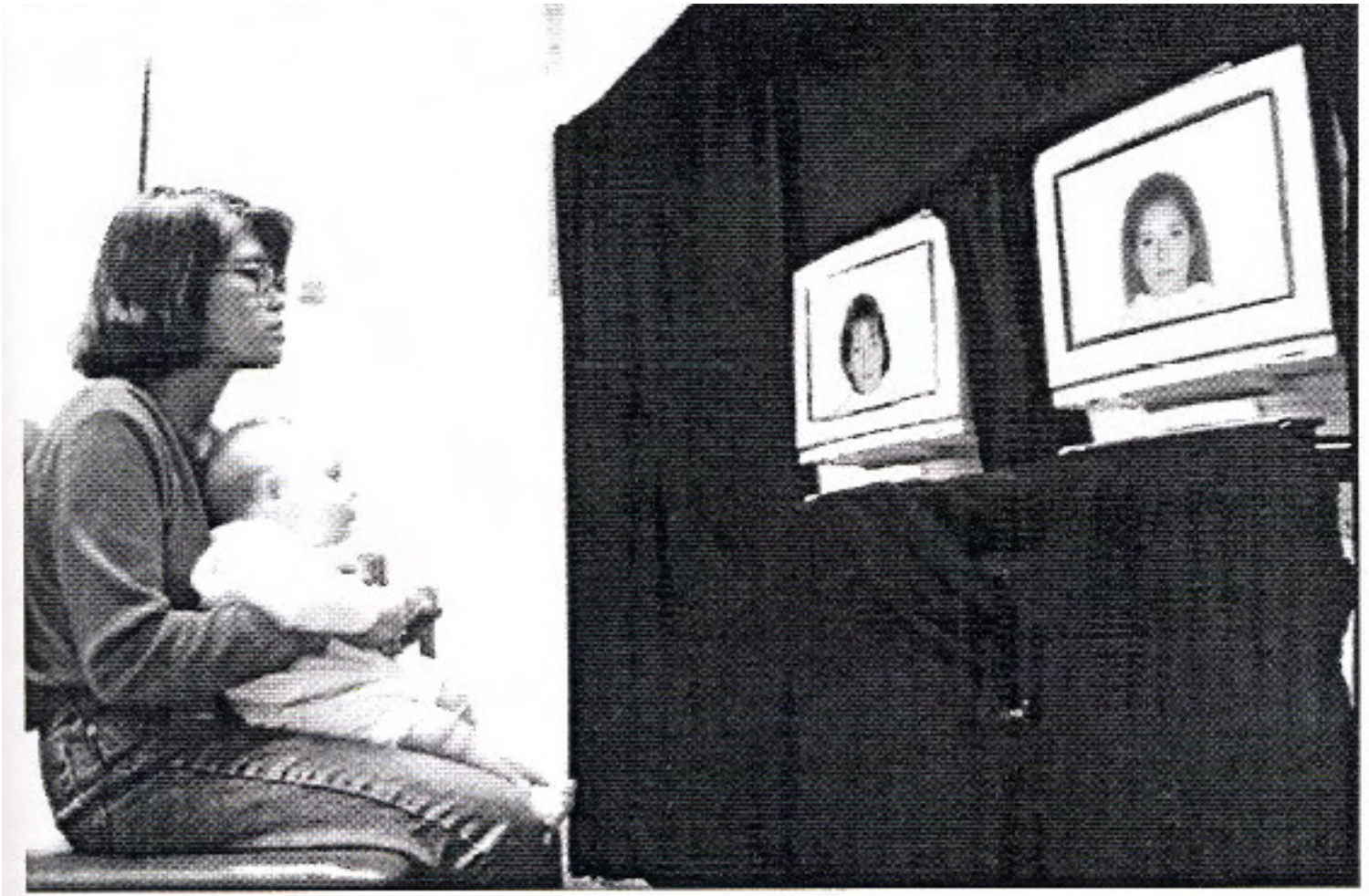
By Paul Rincon
BBC News Online science staff, at the BA festival

Newborn babies - just like adults - prefer to look at an attractive face, new research in the UK has shown.

The University of Exeter study reveals that infants are born with in-built preferences which help them to make sense of their new environment.

The newly born show a clear preference for attractive faces

Newborn infants prefer attractive faces



To our own amazement, we found that 6-month-olds looked longer at the attractive faces than the unattractive faces. Even more amazing to us, we added a second group of even younger infants (2-3-month-olds) and found similar results. These very young babies also

Developmental Science

Developmental Science 11:1 (2008), pp 76–83 DOI: 10.1111/j.1467-7687.2007.00647.x

REPORT

Preference for attractive faces in human infants extends beyond conspecifics

Paul C. Quinn,[1] David J. Kelly,[2] Kang Lee,[3] Olivier Pascalis[2] and Alan M. Slater[4]

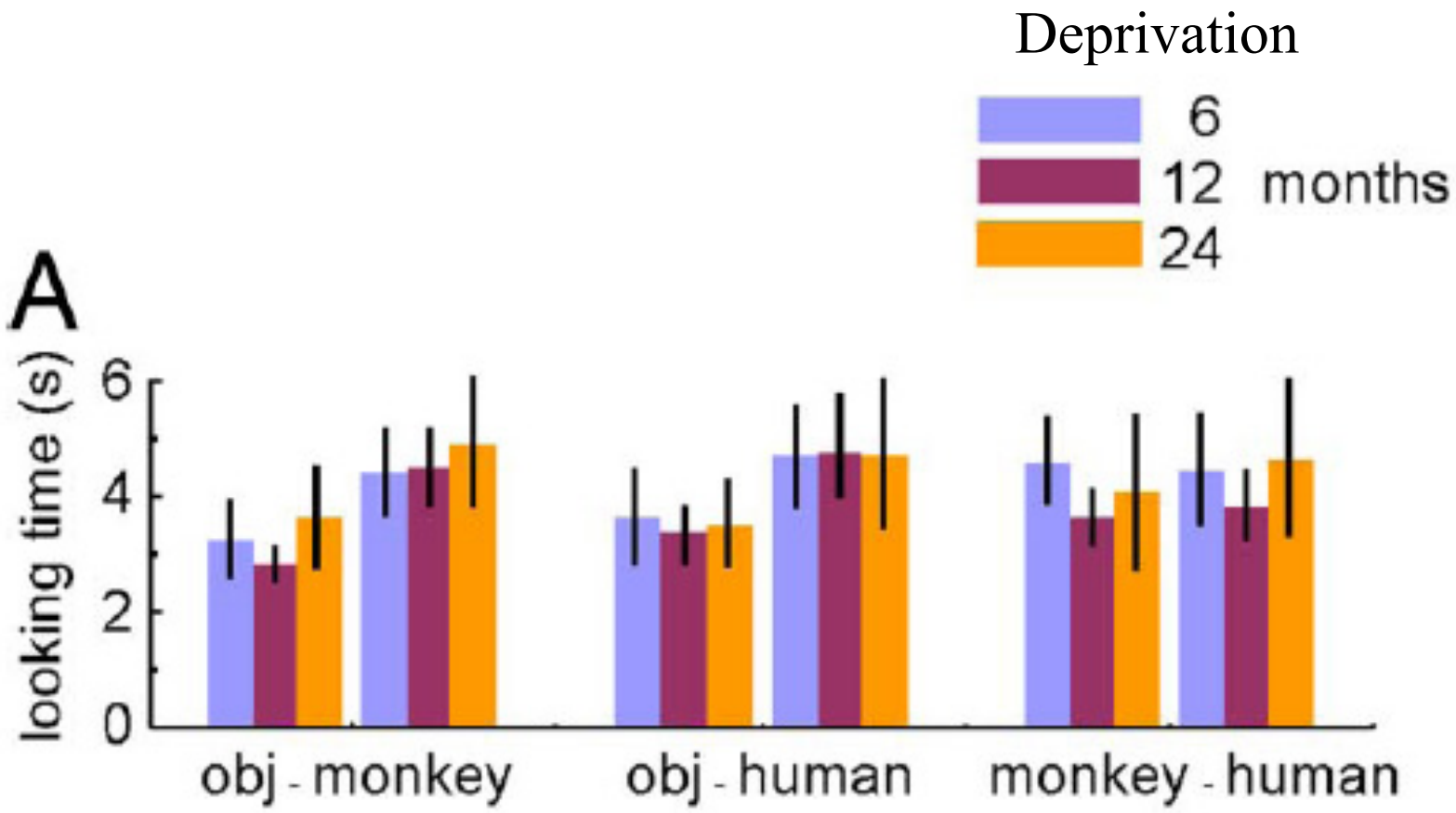
Figure 1 *Grayscale examples of the cat (top panel) and tiger (bottom panel) face stimuli used in the experiments. Attractive*

Face perception in monkeys reared with no exposure to faces

Yoichi Sugita

PNAS, 2008

Fig. 1. An infant monkey and her living circumstance. An infant monkey and a caregiver with (*A*) and without (*B*) a facemask. Both photos were taken after the face-deprivation period. (*C*) Toys placed in the monkey's home cage. (*D*) Decorations provided around the home cage.


Deprivation
6
12 months
24
A
looking time (s)
0
2
4
6
obj - monkey
obj - human
monkey - human

… infants are born with some information about the structure of faces.

Psychological Review
1991, Vol. 98, No. 2, 164–181

0033-295X/91/$3.00

CONSPEC and CONLERN: A Two-Process Theory of Infant Face Recognition

John Morton and Mark H. Johnson
Medical Research Council, Cognitive Development Unit
London, England

Evidence from newborns leads to the conclusion that infants are born with some information about the structure of faces. This structural information, termed *CONSPEC*, guides the preference for facelike patterns found in newborn infants. CONSPEC is contrasted with a device termed *CON-*
rning about the visual characteristics of conspecifics. In the
influence looking behavior until 2 months of age. The distinc-
echanisms allows a reconciliation of the conflicting data on the
human infants. Finally, evidence from another species, the
2-process theory has already been put forward, is discussed.
the chick and used as a basis for comparison with the infant.

LETTER Communicated by Marion Stewart-Bartlett

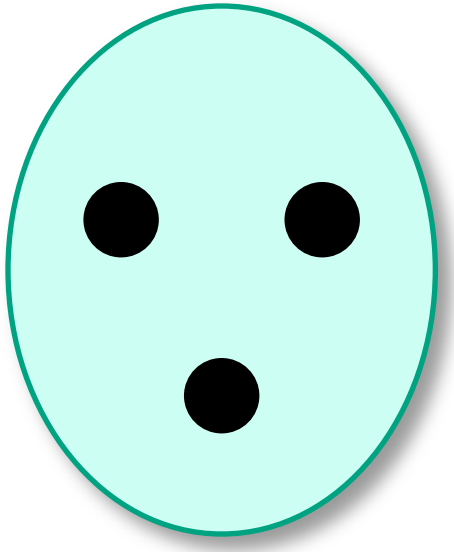
Learning Innate Face Preferences

James A. Bednar
jbednar@cs.utexas.edu
Risto Miikkulainen
risto@cs.utexas.edu
Department of Computer Sciences, University of Texas at Austin, Austin, TX 78712, U.S.A.

Newborn humans preferentially orient to facelike patterns at birth, bu
months of experience with faces are required for full face processing abilities to develop. Several models have been proposed for how the interaction of genetic and evironmental influences can explain these data. These models generally assume that the brain areas responsible for newborn orienting responses are not capable of learning and are physically separate from those that later learn from real faces. However, it has been difficult to reconcile these models with recent discoveries of face learning in newborns and young infants. We propose a general mechanism by which genetically specified and environment-driven preferences can coexist in the same visual areas. In particular, newborn face orienting may be the result of prenatal exposure of a learning system to internally generated input patterns, such as those found in PGO waves during REM sleep. Simulating this process with the HLISSOM biological model of the vi-

… prenatal exposure of a learning system to internally generated input patterns.

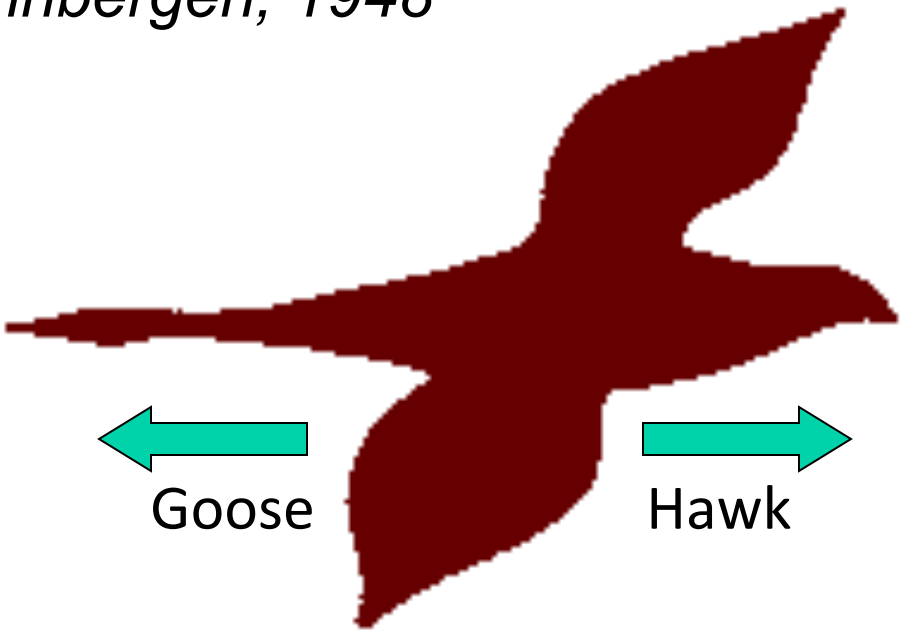
Proposed innately specified face-schema



Plausibility argument

Evidence for innate schemas in other species

Lorenz, 1939; Tinbergen, 1948



The case for ‘specialness’ of face perception seems very strong

fMRI studies:

A special section of cortical tissue appears to be devoted to face processing

Infant studies:

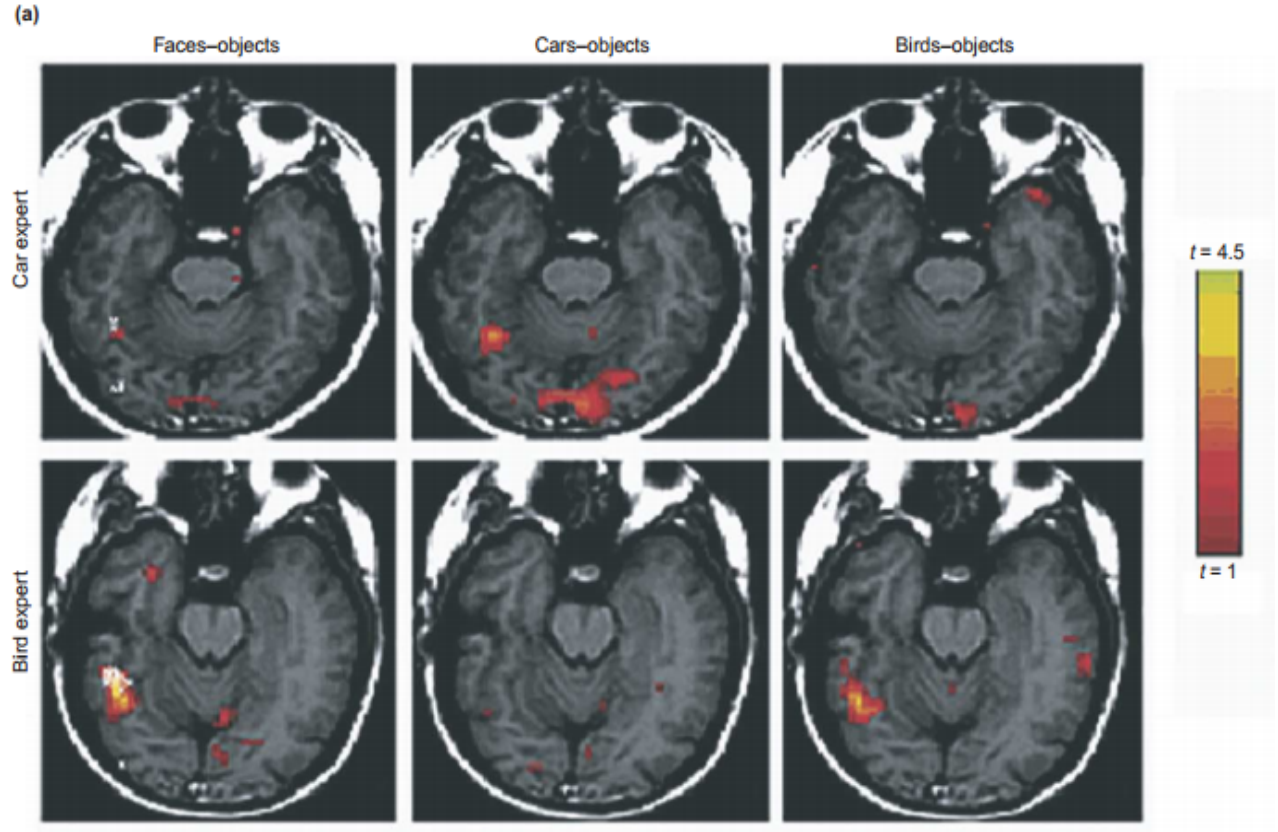
Experimental data suggest impressive face abilities in newborns

Even if a learning-based mechanism were available, it would not have enough time to become useful.

Counterpoint

Faces may not be intrinsically special for the brain.

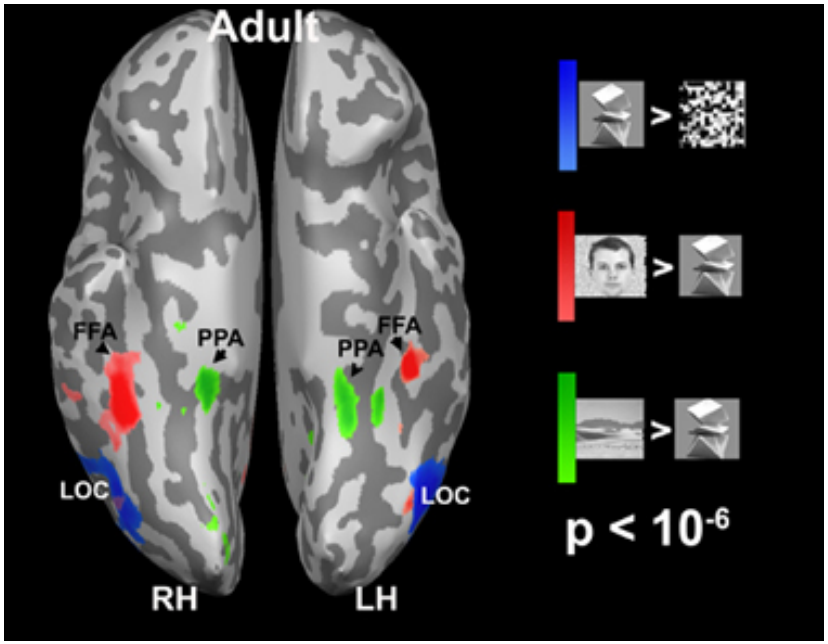
Is the FFA really devoted to faces or any class of objects with which we have lots of expertise?



What are the consequences of lesioning the FFA?

Expectation:

If the putative ‘face areas’ of the cortex are truly critical for face processing, we should be able to induce prosopagnosia through selective lesions of these areas.



Consequences of lesioning the face-cell area:
Heywood and Cowey, 1992

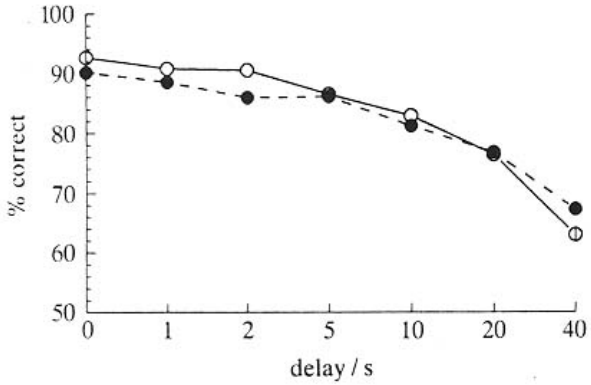
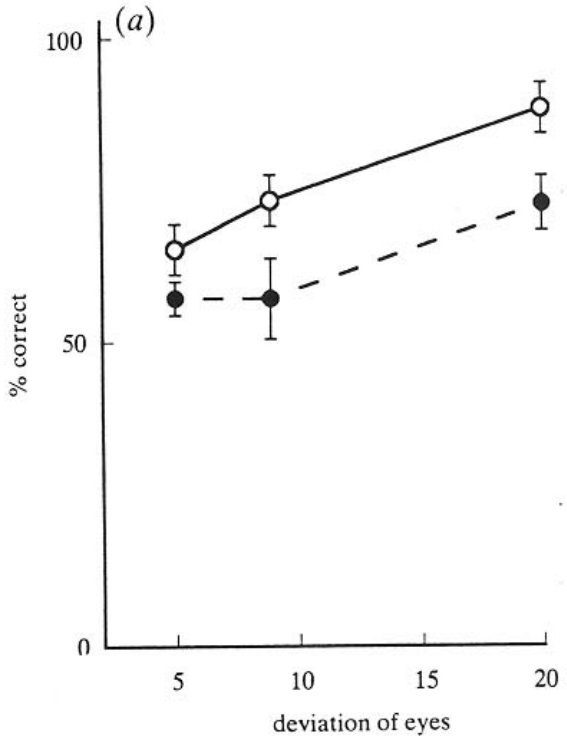
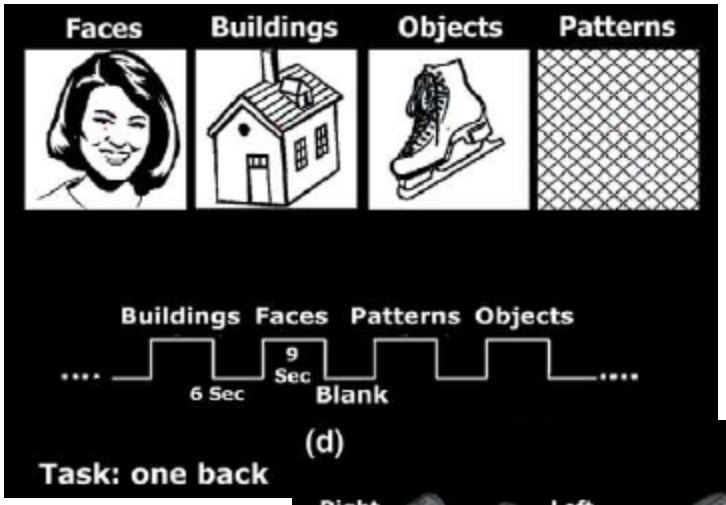


Figure 2. Mean percentage correct for unoperated (open circles) and STS (closed circles) group for delayed non-matching to sample where each point is the group mean of scores on novel and familiar faces and objects. Groups did not differ significantly in their performance on each of the tasks.

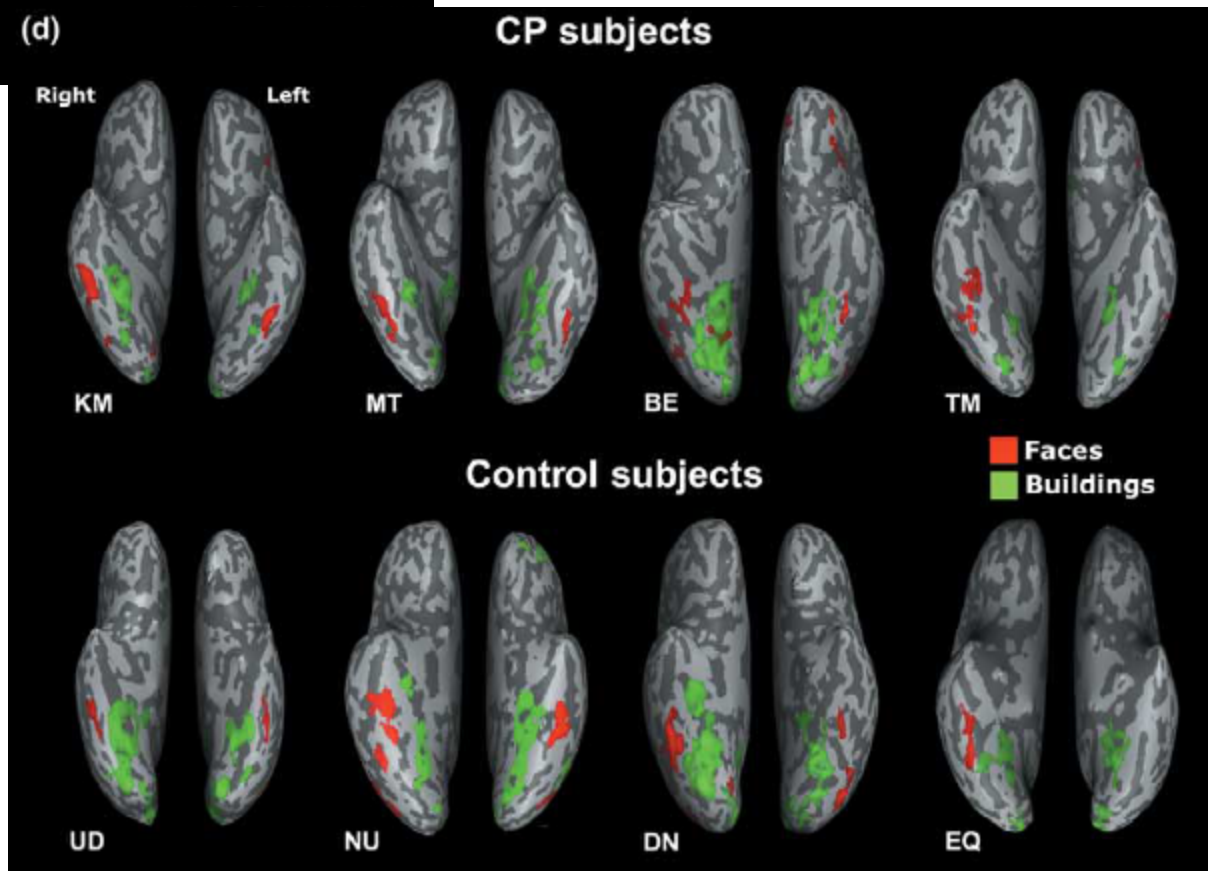


Inferences: No specific face recognition impairments
Mild impairments in gaze estimation.



Conversely, similarity of activations in ‘face areas’ does not predict behaviorally observed face-perception deficits

(Avidan, 2005)



Note the substantial similarity in the activation pattern exhibited by all CP subjects and controls.

"The search for an area of the brain entirely devoted to facial perception and memory and for recognition deficits specific to faces may be no more successful than the hunt for the Holy Grail"