

Visual Cognition III

Object Perception

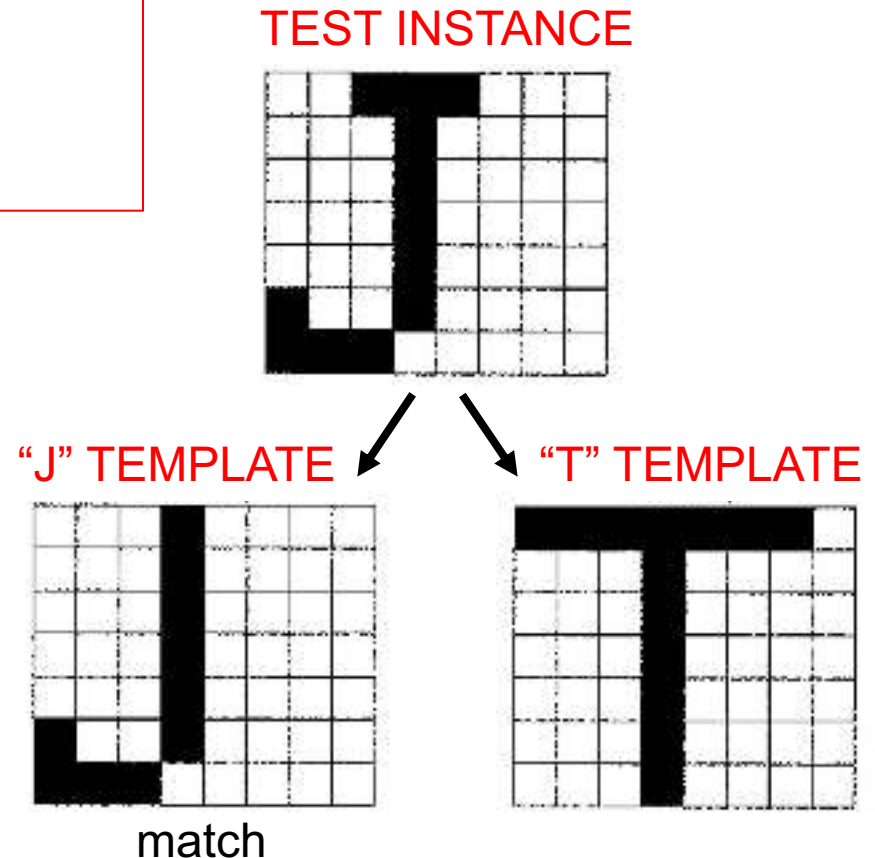
Theories of Object Recognition

- Template matching models (模板匹配)
- Feature matching Models (特征匹配)
- Recognition-by-components (成分识别)
- Configural models (构型模型)

Template Matching

A **template** is a pattern, like a cookie cutter or a stencil. It can be used to compare individual items to a standard.

The ***template-matching method*** as initially conceived is straightforward and useful as long as the item to be recognized and the template to which the system compares it are almost identical and different from others.



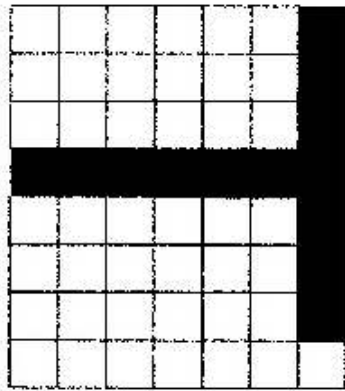
Detect patterns by matching visual input with a set of templates stored in memory – see if any template matches.

Template Matching

Problem:

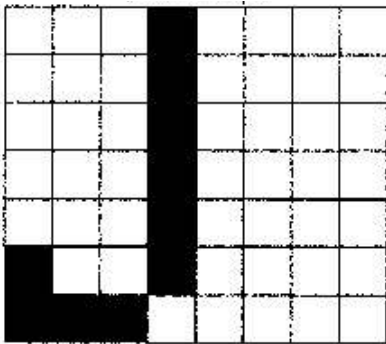
what if the object differs slightly from the template?
E.g., it is rotated or scaled differently?

TEST INSTANCE

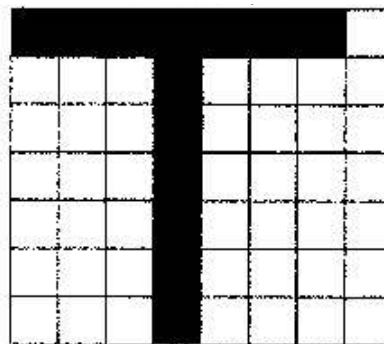


rotation
↓

“J” TEMPLATE



“T” TEMPLATE



match

Solution:

use a set of transformations to best align the object with a template (using translation, rotation, scaling)

Template-matching Works Well in Constrained Environments

STEPHEN R. DR CONSTANCE R. SCHMIDT 1805

2206 PINEHILL CT. PH. 896-9425
MURFREESBORO, TN 37130

_____ 19 _____ 87-7154
2641

PAY TO THE ORDER OF _____ \$ _____

_____ DOLLARS

Cavalry
Banking FA Murfreesboro, TN 37130

MEMO _____

⑆ 2641715421⑆ 1805

numbers scanned and recognized with template system

Fingerprints



弓形



环形



螺旋形

Feature Theories

- Detect objects by the presence of features
- Each object is broken down into features
- E.g. $A = / + \backslash + -$

Feature-matching models search for simple but characteristic features of an object; their presence signals a match. What constitutes a feature in these models? That varies with the type of object.

The feature-matching approach also lends itself well to the idea that processing of information in the brain is **parallel** (that is, happening at the same time) and distributed (that is, happening in different neural areas).

For a feature-matching model to be a plausible explanation of how we recognize objects, neurons or populations of neurons should show selectivity to parts of the input similar to the features in the model.

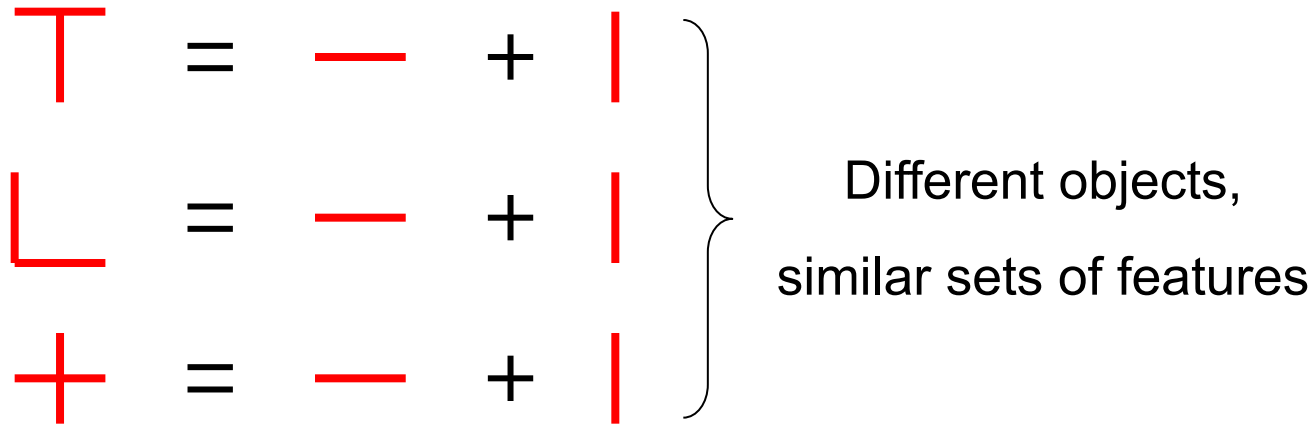
Sensory Binding Problem

- If spatially different areas are functionally specialized for processing visual attributes such as shape, color, orientation, and direction of motion....
- then how does the brain then “bind” together the sensory attributes of an object to construct a unified perception of the object?

→ **Binding Problem**

Problem

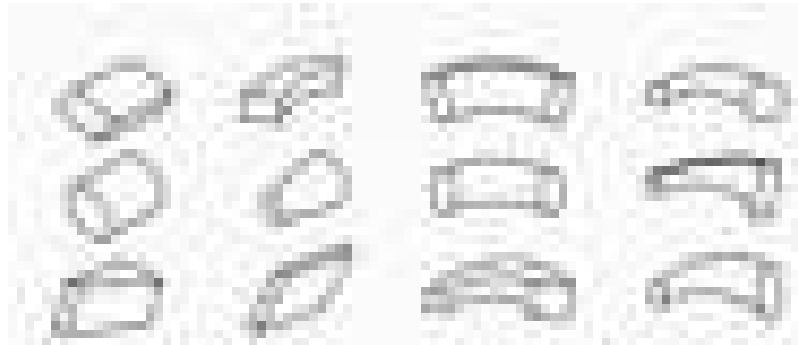
- Many objects consist of the same collection of features



- Need to also know how the features relate to each other
→ structural theories
- One theory is *recognition by components*

Recognition by Components (RBC)

- Biederman (1987): Complex objects are made up of arrangements of basic, component parts: *geons*
- “Alphabet” of 24 geons
- Recognition involves recognizing object elements (geons) and their configuration 组合，布置，结构



Recognition by Components (RBC)

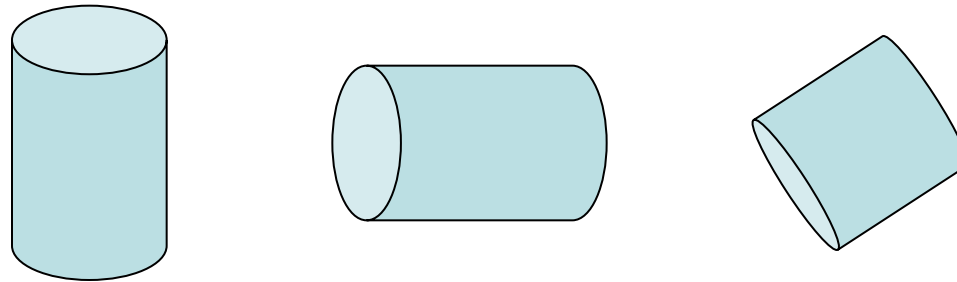
The *recognition-by-components (RBC)* model provides a possible method for recognizing three-dimensional objects across variations in viewpoint or exemplars (Biederman, 1987). (p. 78)

The model assumes that any three-dimensional object can be generally described according to its parts and the spatial relations among those parts.

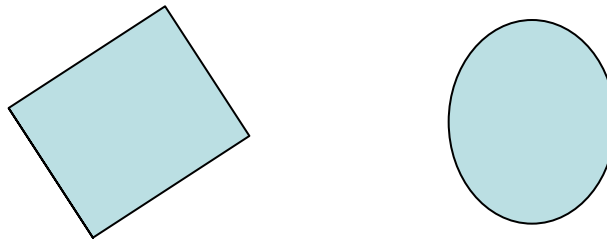
The current model proposes that a set of 24 geometrical three-dimensional shapes, such as cylinders and cones, can be used to represent just about any object; in the language of the model, these shapes are called **geons**.

Why these geons?

- Choice of shape vocabulary seems a bit arbitrary
- However, choice of geons was based on non-accidental properties. The same geon can be recognized across a variety of different perspectives: **Viewpoint Invariance**

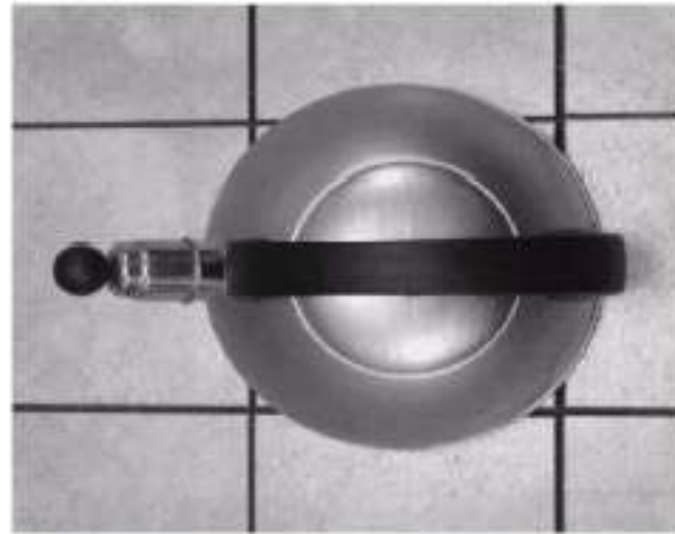


except for a few “accidental” views:



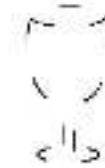
Viewpoint Invariance

- Viewpoint invariance is possible except for a few accidental viewpoints, where geons cannot be uniquely identified



Prediction

- Recognition is easier when geons can be recovered
- Disrupting vertices disrupts geon processing more than just deleting parts of lines



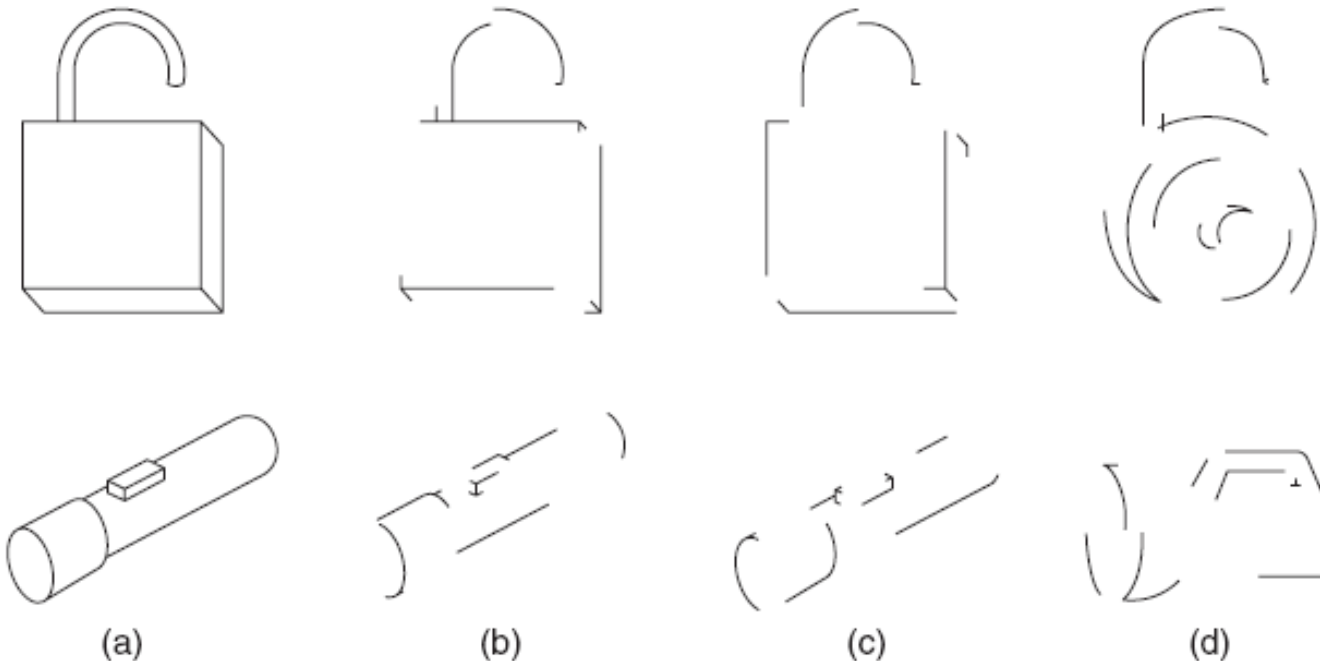
Object

Deleting line
segments

Deleting
vertices

Evidence from priming experiments

- Visual priming and recognition



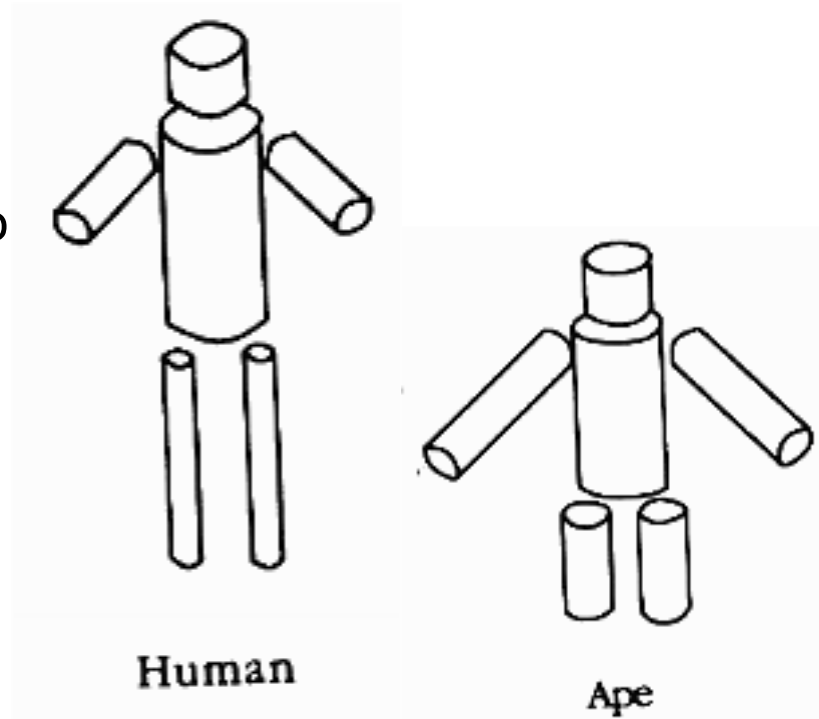
Two objects as (a) intact, (b) with half their contours removed, and (c) with the other half of the contours removed. (d) A different example of the same type of object but with different geons (again, half the contours are removed). It was easier for participants to recognize the objects when they saw (b) followed by (c) than when they saw (b) and (d).

Problem for RBC

- Theory does not say how color, texture and small details are processed. These are often important to tell apart specific exemplars or similar objects. E.g.:

Some problems for recognition by components theory:

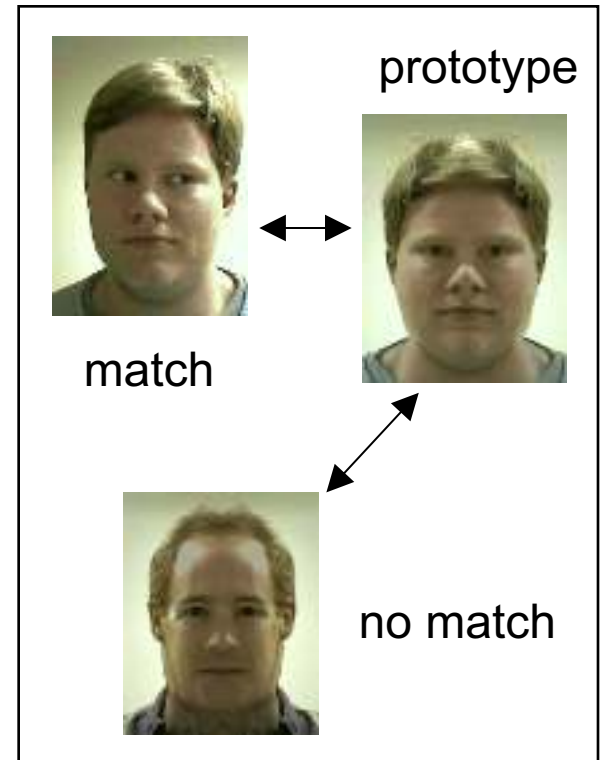
- 1) Structural description not enough, also need metric info to distinguish between specific exemplars or similar objects
- 2) Difficult to extract geons from real images
- 3) For some objects, deriving a structural representation can be difficult



Configural models of recognition

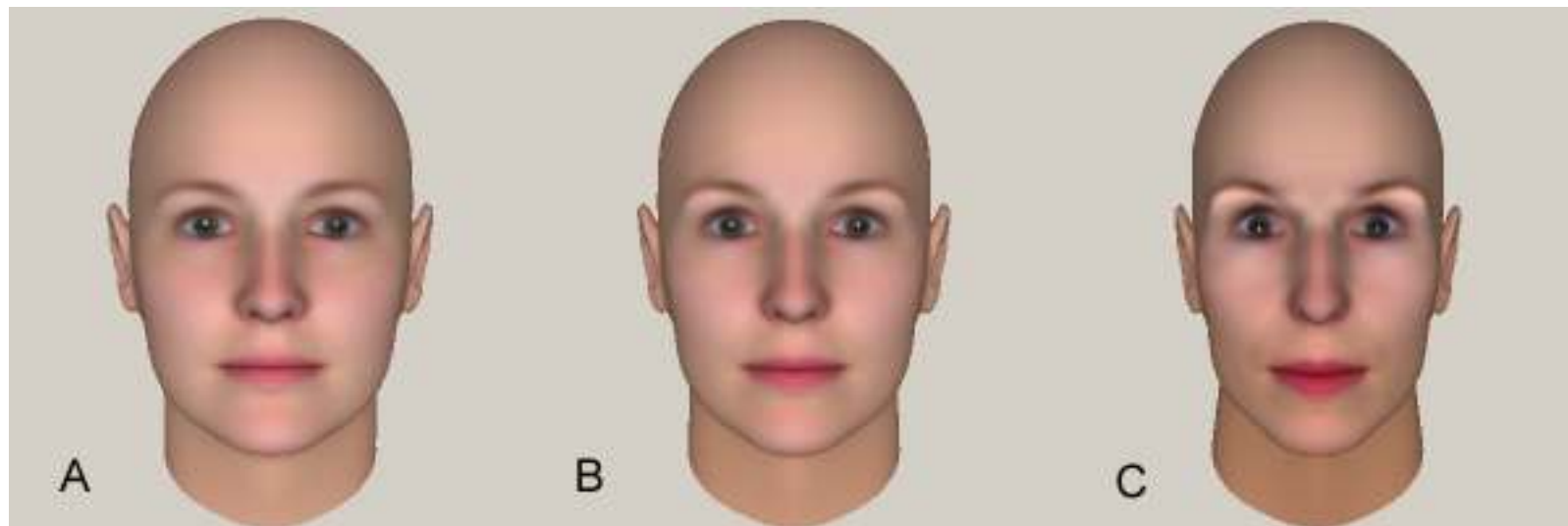
- Individual instances are not stored; what is stored is an “exemplar” or representative element of a category
- Recognition based on “distance” between perceived item and prototype

Configural models often can deal with the limitations of RBC models. They propose that objects that share the same parts and a common structure are recognized according to the spatial relations among those parts *and the extent to which those spatial relations deviate from the **prototype**, or “average,” object.*



“Face space”

Prediction: Caricatures might be better recognized than original face



A average female face

B “veridical” face

C caricature of B

The average female face for a particular face population is displayed (A), as well as a “veridical” image of an exemplar face (B). We create a caricatured version of the exemplar by moving away from the norm, thus exaggerating differences between the average face and the exemplar. The result is a face with “caricatured” shape and pigmentation (C). Such caricatures are recognized as well or better than veridical images.

Do these faces have anything in common?



- a) no, nothing
- b) same mouth
- c) same nose
- d) same eyes

How about these ones?



By disrupting holistic (configural) processing, it becomes easier to process the individual parts

Face Inversion

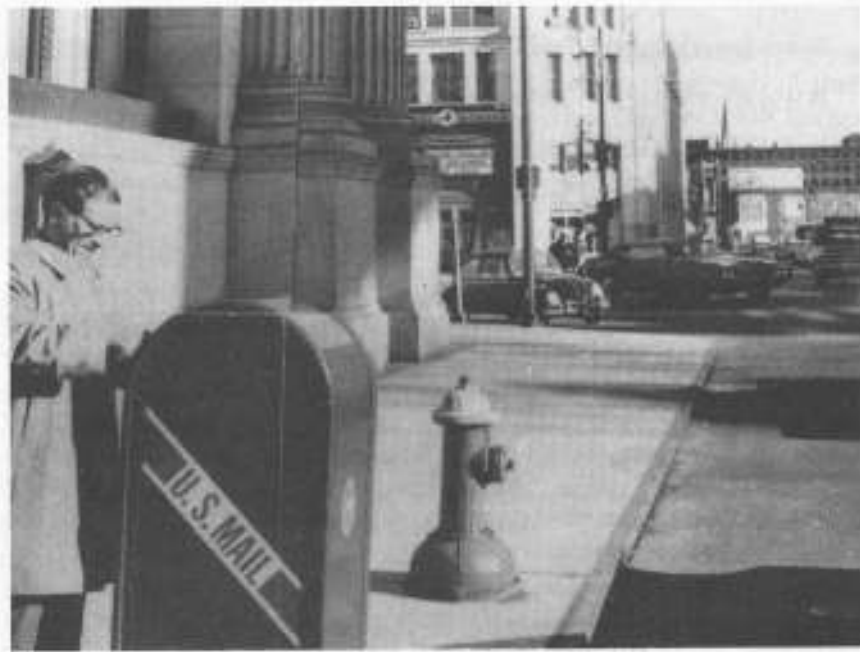
- Configural effects often disappear when stimulus is inverted



Context Effects in Object Recognition



Context can often help in identification of an object



(a)



(b)

Later identification of objects is more accurate when object is embedded in coherent context

Context can alter the interpretation
of an object

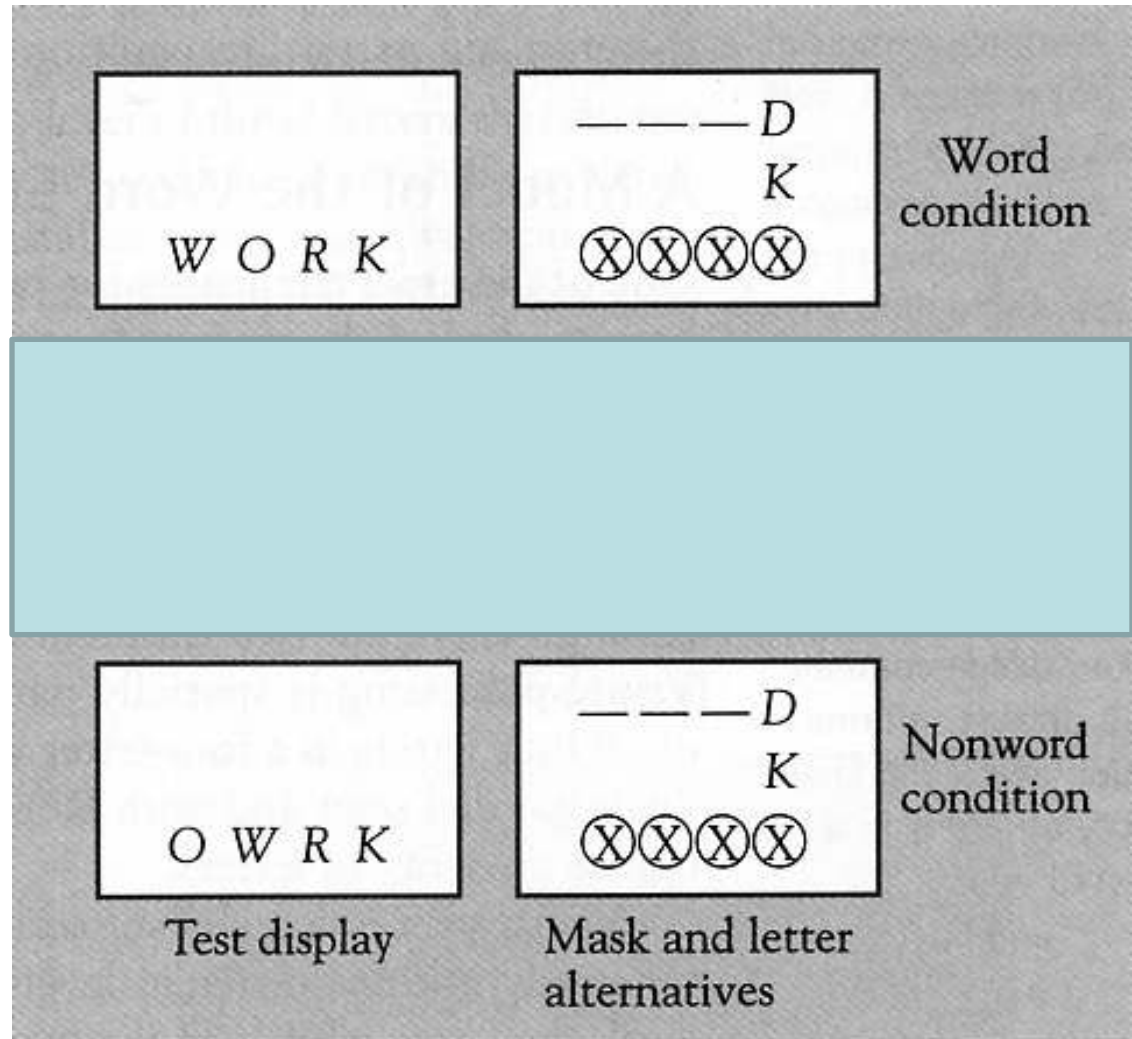
**THEY
BHKE**

THE CAT

Context Effects in Letter Perception

The word superiority effect: discriminating between letters is easier in the context of a word than as letters in the context of a nonword string.

Context improves people's sensitivity, not just bias!!

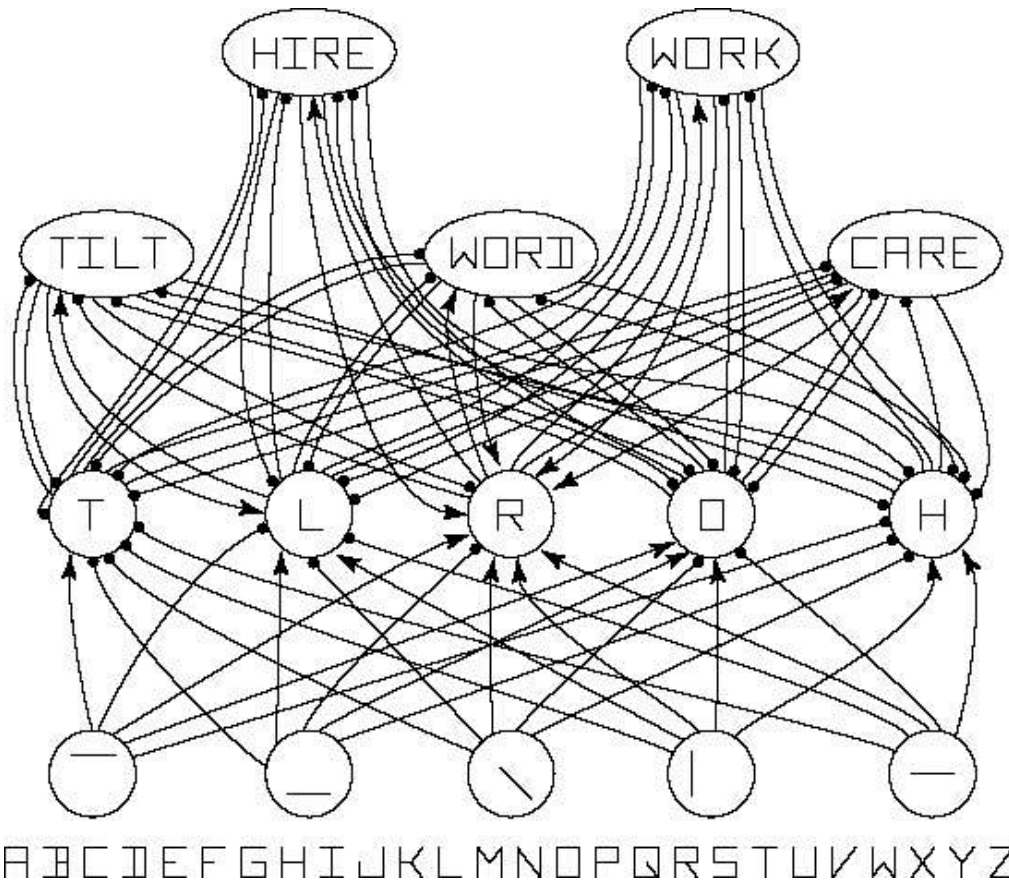


Interactive Activation Model

互动激发模式

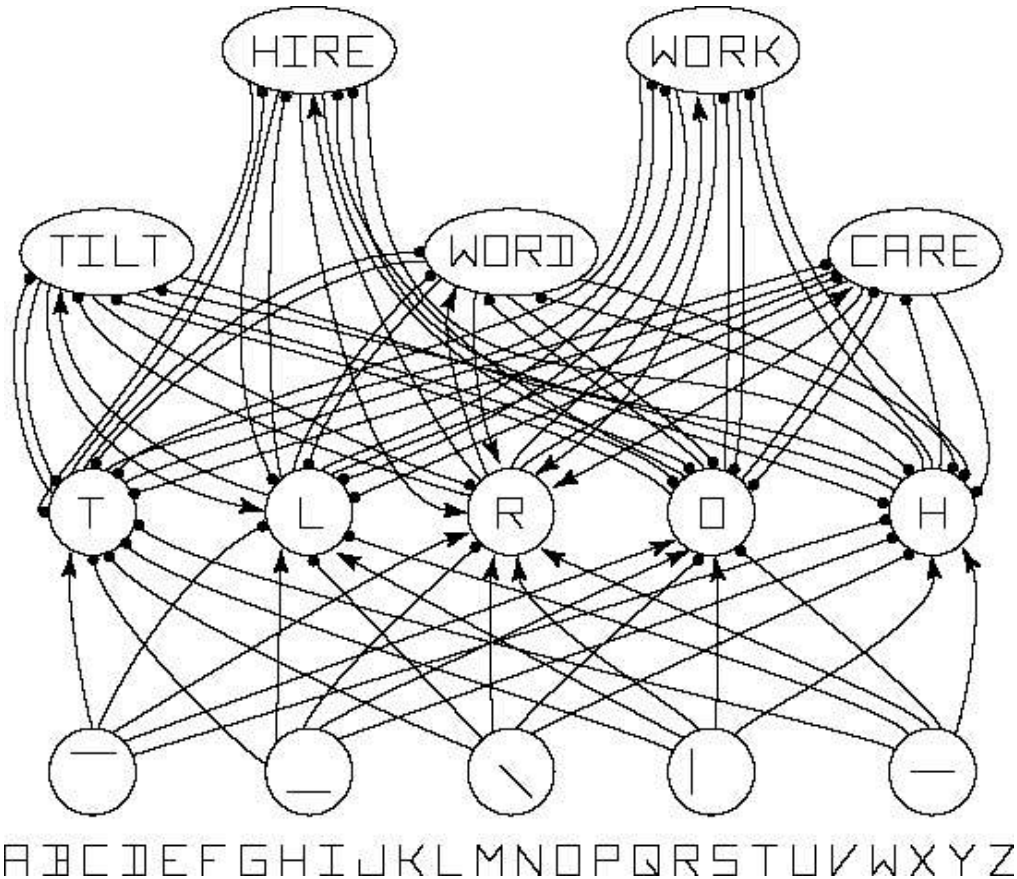
- Word superiority effect suggests that information at the word level might affect interpretation at the letter level
- **Interactive activation model**: neural network model for how different information processing levels interact
- Levels interact
 - bottom up: how letters combine to form words
 - top-down: how words affect detectability of letters

The Interactive Activation (IA) Model



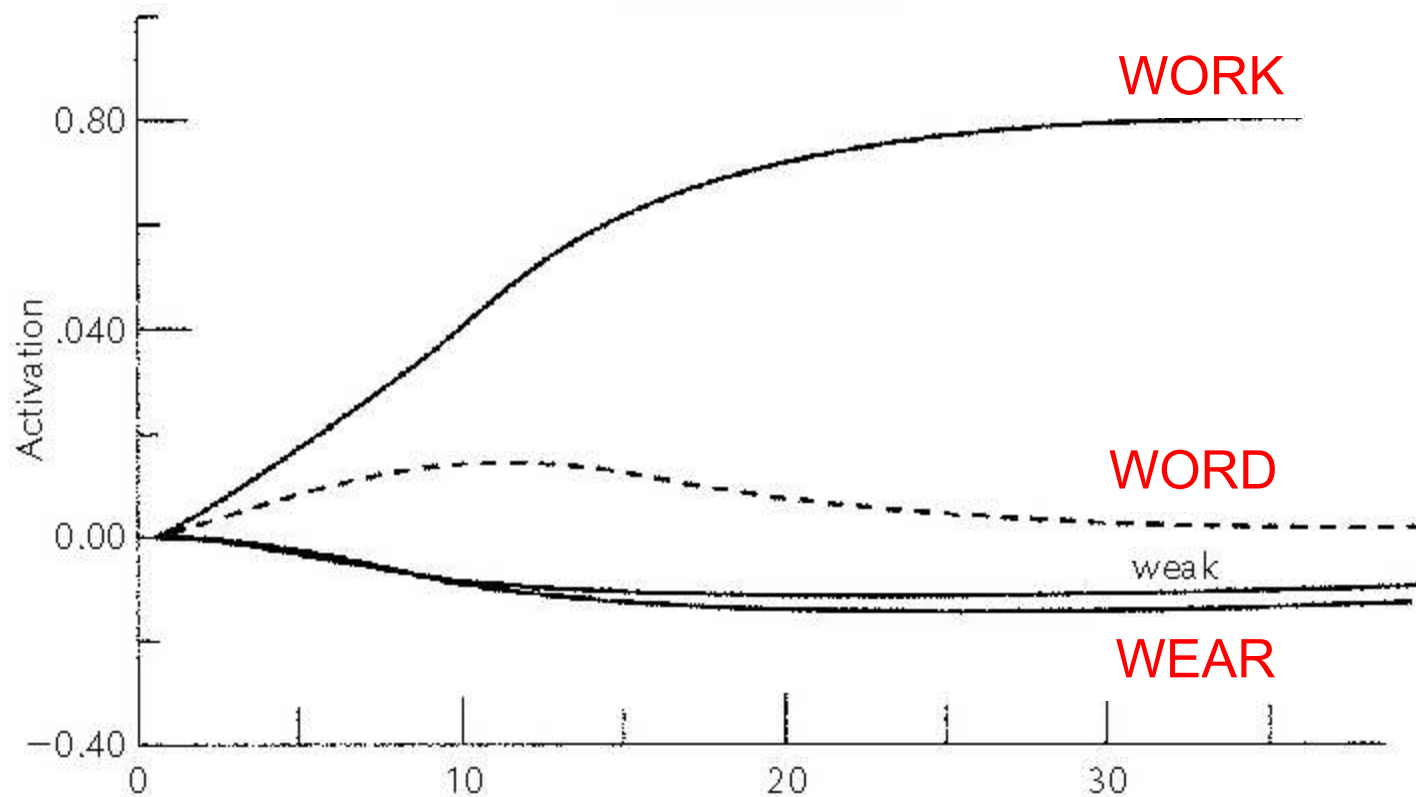
- Three levels: **feature, letter, and word level**
- **Nodes** represent features, letters and words; each has an **activation** level
- Connections between nodes are **excitatory** or **inhibitory**
- Activation flows from feature to letter to word level and back to letter level

The Interactive Activation (IA) Model



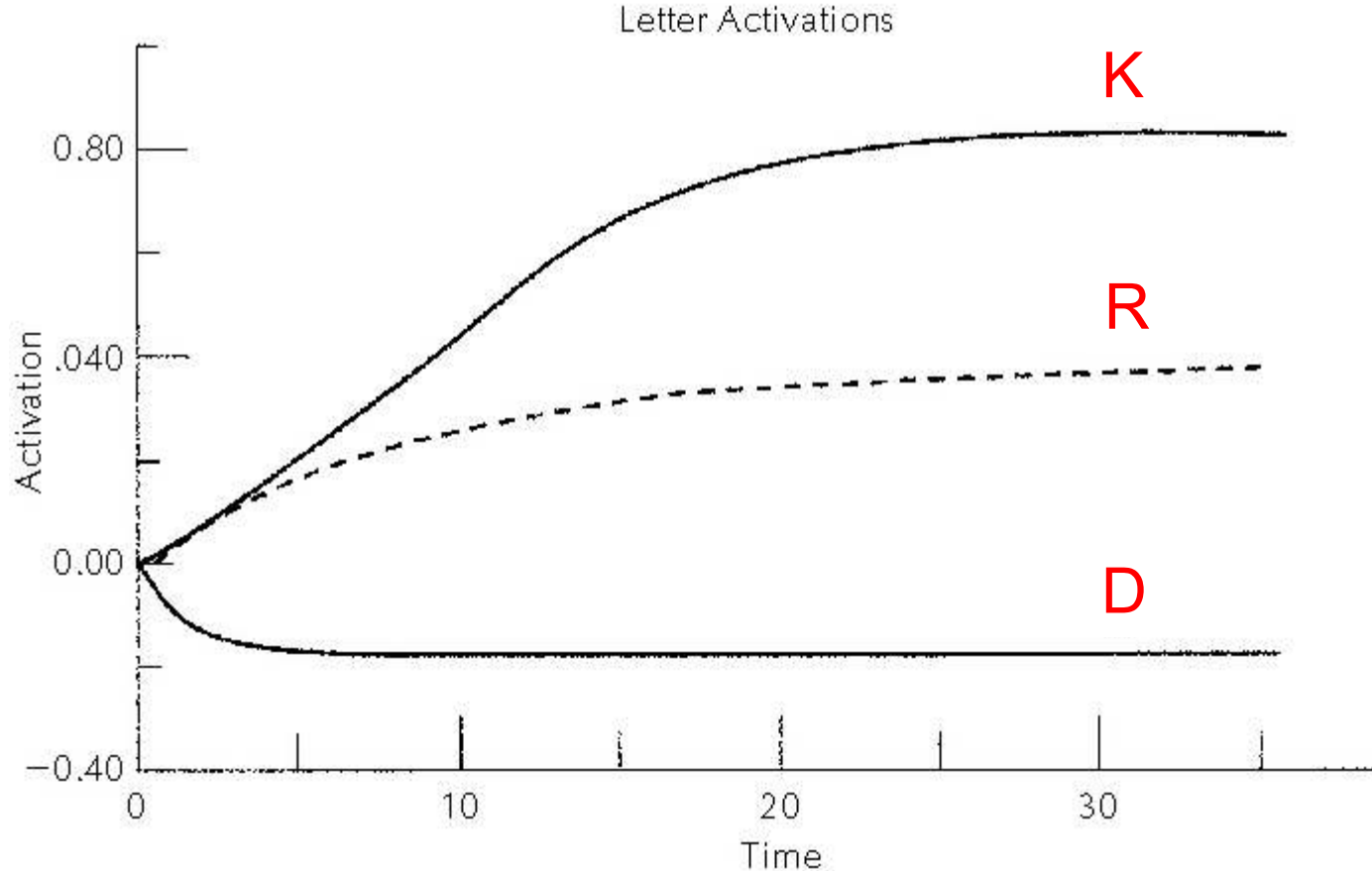
- Bottom-up:
 - feature to word level
- Top-down:
 - word to letter level
- Model predicts word superiority effect because of top-down processing

Predictions of the IA model – stimulus is “WORK”



- At word level, evidence for “WORK” accumulates over time
- Small initial increase for “WORD”

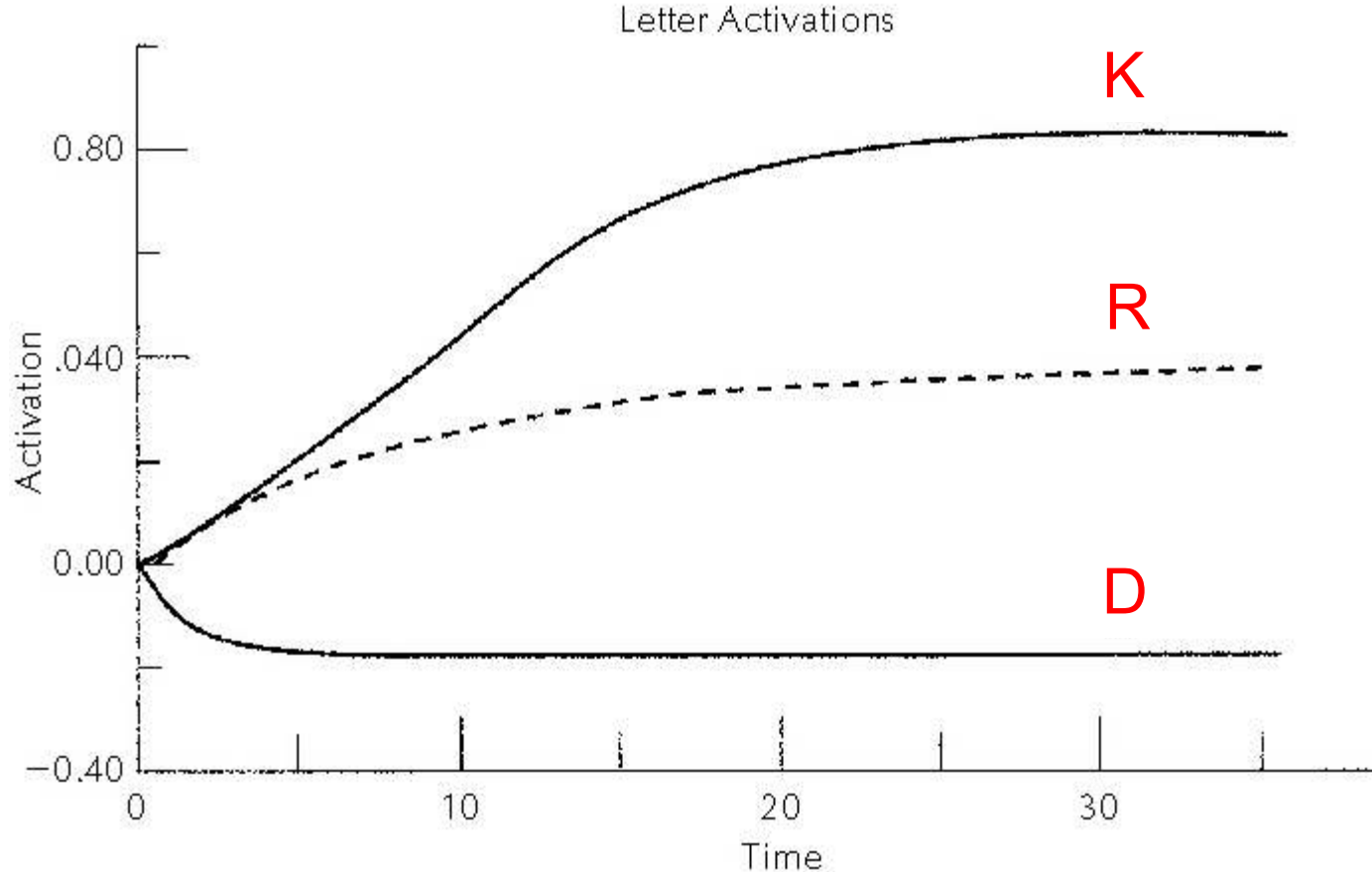
Predictions of the IA model – stimulus is “WORK”



Why does the letter “K” get activated?

- a) because of (partial) activation from feature level
- b) because of activation from word level back to feature level
- c) both a) and b)

Predictions of the IA model – stimulus is “WORK”



Why does the letter “R” get partially activated?

- a) because of (partial) activation from feature level
- b) because of activation from word level back to feature level
- c) both a) and b)

Perception

The Interactive Activation Model

Dr David Hardman
School of Psychology
London Metropolitan University

“Mind Reading”



Predicting What Somebody is Seeing

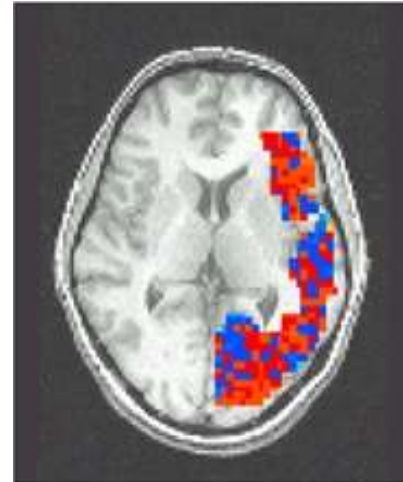
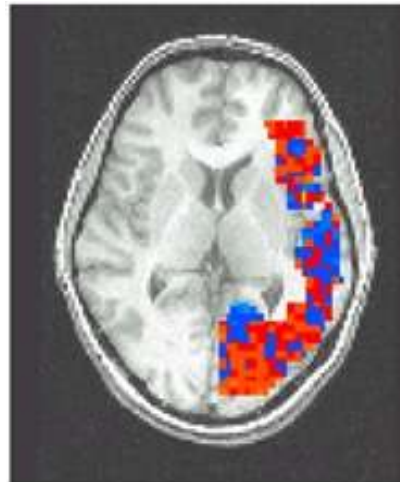
(“mind reading”)

Viewing a Bottle

Bold
Response

Bold
Response

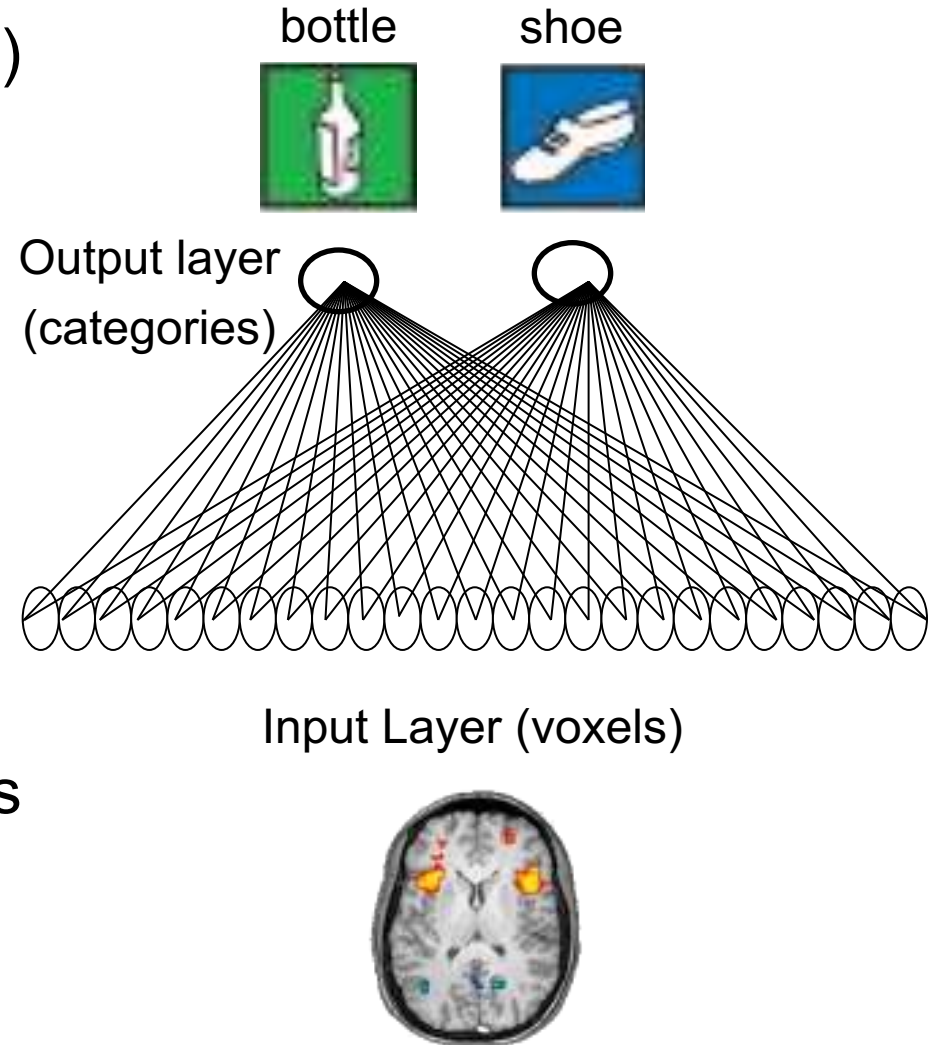
Viewing a Shoe



If the brain response is different for different kinds of stimuli, can we predict what somebody is thinking of solely based on the brain's response?

Pattern Classification Method

- Acquire brain data for different stimuli (e.g. bottles and shoes)
- Train a classifier (such as the neural network on right) to discriminate between bottle voxel patterns and shoe voxel patterns
- Test classifier on novel images



Haxby et al. (2001) can predict with 96% accuracy stimuli from 8 categories



Faces



Houses



Cats



Bottles



Scissors



Shoes



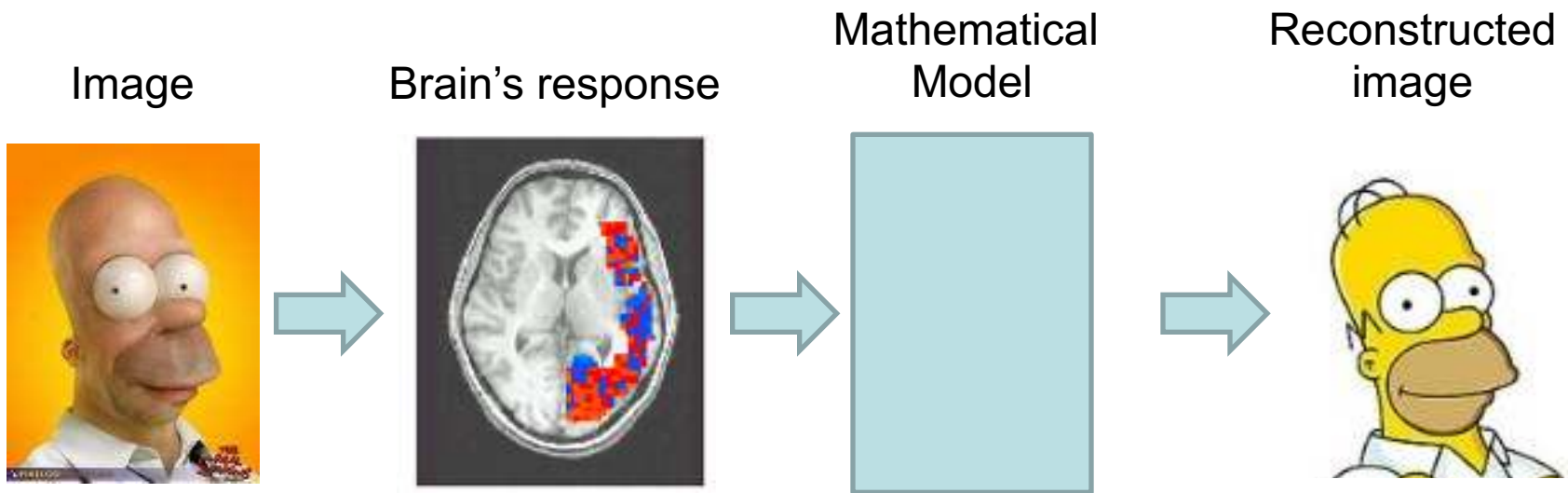
Chairs



Scrambled Pictures

Reconstructing the Mental Image

- If we can predict what somebody is looking at, can we also **reconstruct** what somebody might be looking at from just the brain's response?



Reconstructing simple patterns from fMRI

Miyawaki et al. (2008)

Presented
contrast pattern



Reconstructed
contrast pattern

