# Knowledge Representation: Spaces, Trees, Features

# Announcements

- Optional section 1: Introduction to Matlab
  - Tonight, 8:00 pm
- Problem Set 1 is available

# The best statistical graphic ever?

Image removed due to copyright considerations. Please see:
Tufte, Edward. *The Visual Display of Quantitative Information.*
Cheshire CT: Graphics Press, 2001. ISBN: 0961392142.

# The worst statistical graphic ever ?

Image removed due to copyright considerations. Please see:
Tufte, Edward. *The Visual Display of Quantitative Information.*
Cheshire CT: Graphics Press, 2001. ISBN: 0961392142.

# Knowledge Representation

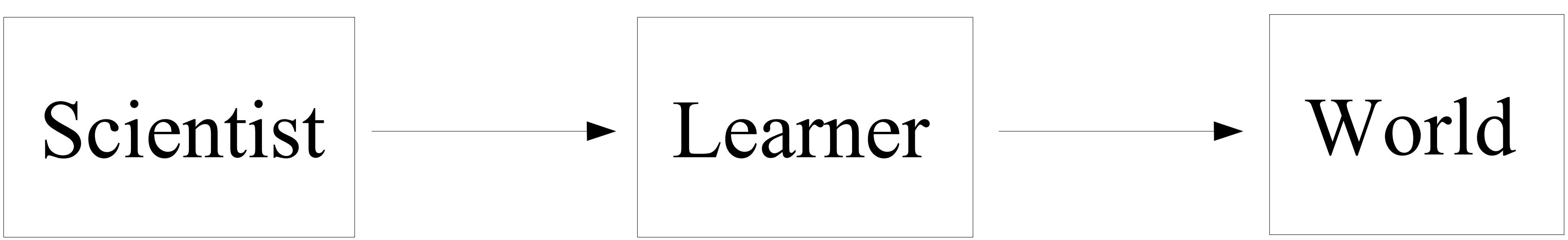
- A good representation should:
  - be parsimonious
  - pick out important features
  - make common operations easy
  - make less common operations possible

# Mental Representations

- Pick a domain: say animals
- Consider everything you know about that domain.
- How is all of that knowledge organized?
  - a list of facts?
  - a collection of facts and rules?
  - a database of statements in first-order logic?

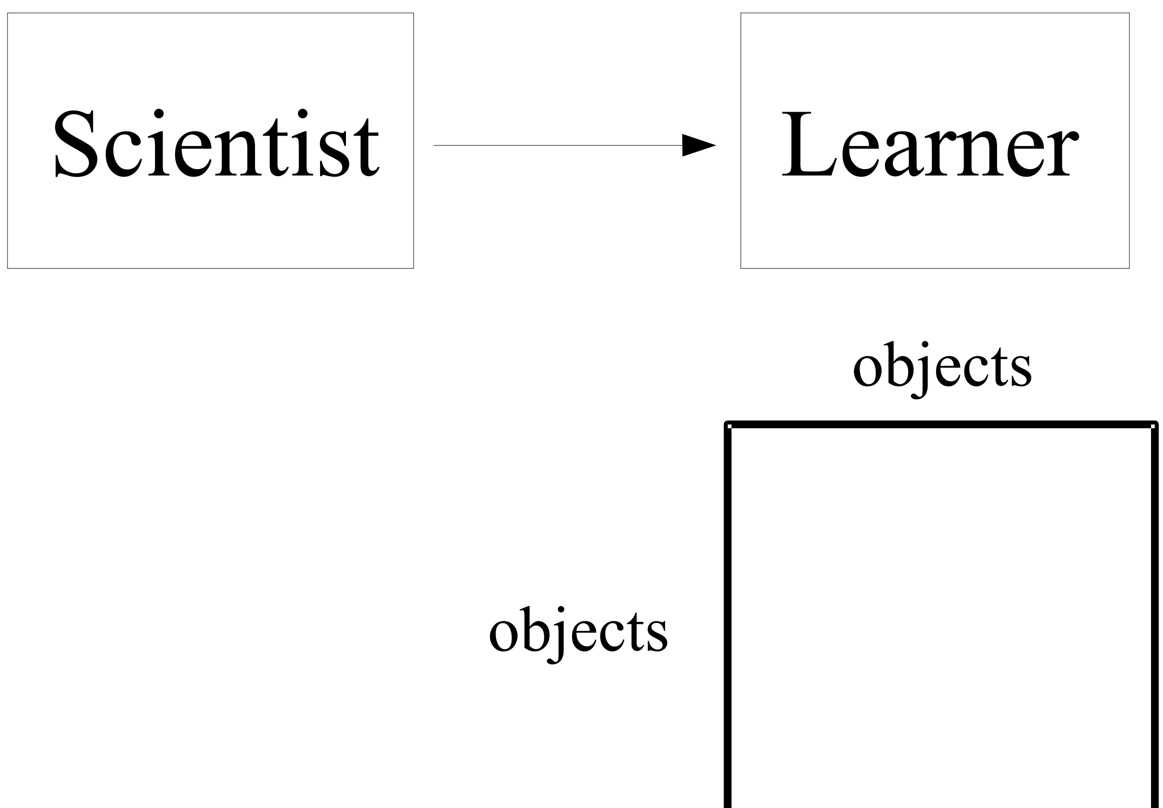
# Two Questions

1. How can a scientist figure out the structure of people's mental representations?
2. How do people acquire their representations?



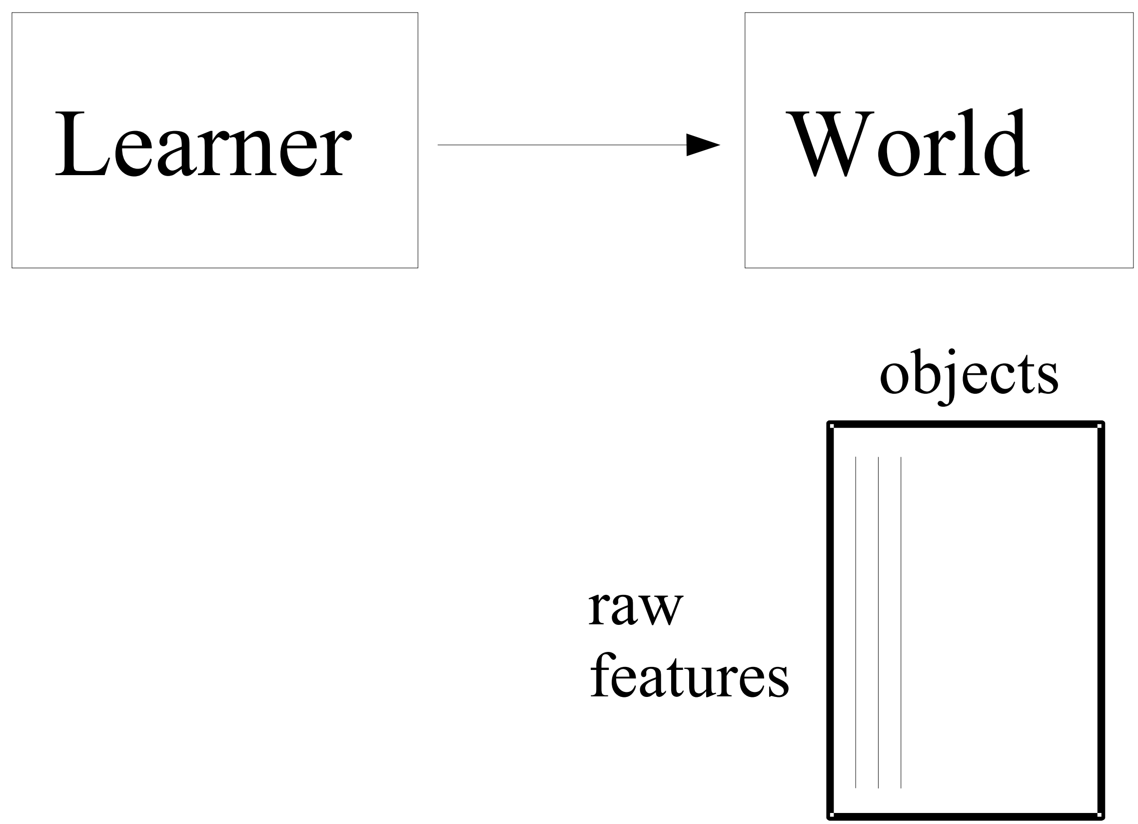
Q: How can a scientist figure out the structure of people's mental representations?

A: Ask them for similarity ratings



Q: How do people acquire their mental representations?

A: They build them from raw features – features that come for free



# Outline

- Spatial Representations
  - Multidimensional scaling
  - Principal component analysis
- Tree representations
  - Additive trees
  - Hierarchical agglomerative clustering
- Feature representations
  - Additive clustering

# Multidimensional scaling (MDS)

Image removed due to copyright considerations.

# Marr's three levels

- Level 1: Computational theory
  - What is the goal of the computation, and what is the logic by which it is carried out?
- Level 2: Representation and algorithm
  - How is information represented and processed to achieve the computational goal?
- Level 3: Hardware implementation
  - How is the computation realized in physical or biological hardware?

# MDS: Computational Theory

$d_{ij}$ : distance in a low-dimensional space

$\delta_{ij}$ : human dissimilarity ratings

- Classical MDS: $d_{ij} \approx \delta_{ij}$
- Metric MDS: $d_{ij} \approx f(\delta_{ij})$
- Non-metric MDS: rank order of the $d_{ij}$ should match rank order of the $\delta_{ij}$

# MDS: Computational Theory

- Cost function
  - Classical MDS: cost = $\sum_{i,j}(d_{ij} - \delta_{ij})^2$

# MDS: Algorithm

- Minimize the cost function using standard methods (solve an eigenproblem if possible: if not use gradient-based methods)

# Choosing the dimensionality

- Elbow method

Image removed due to copyright considerations.

# Colours

Image removed due to copyright considerations.

# Phonemes

Image removed due to copyright considerations.

# What MDS achieves

- Sometimes discovers meaningful dimensions
- Are the dimensions qualitatively new ? Does MDS solve Fodor's problem?

# What MDS doesn't achieve

- Solution (usually) invariant under rotation of the axes
- The algorithm doesn't know what the axes mean. We look at the low-dimensional plots and find meaning in them.

# ideonomy.mit.edu

Image removed due to copyright considerations.

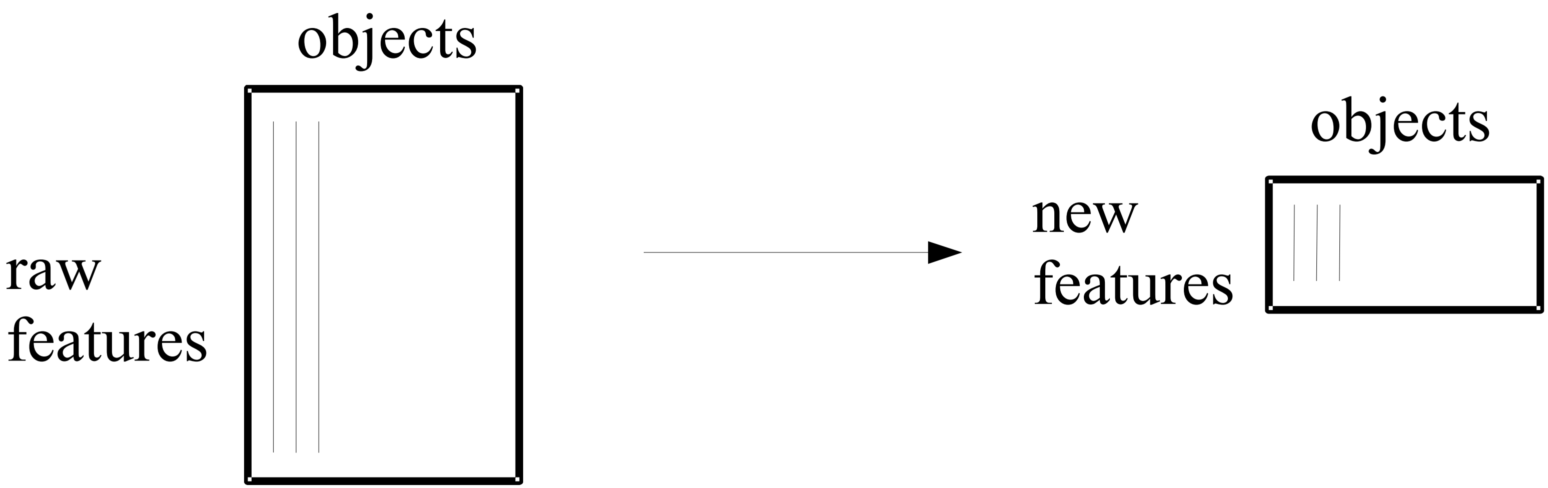
Please See: http://ideonomy.mit.edu/slides/16things.html

Patrick
Gunkel

# Two Questions

1. How can a scientist figure out the structure of people's mental representations?
2. How do people acquire their representations?

# Principal Components Analysis (PCA)



# PCA

Image removed due to copyright considerations.

# PCA

Image removed due to copyright considerations.

# PCA

Image removed due to copyright considerations.

# PCA

- Computational Theory
  - find a low-dimensional subspace that preserves as much of the variance as possible
- Algorithm
  - based on the Singular Value Decomposition (SVD)

# SVD

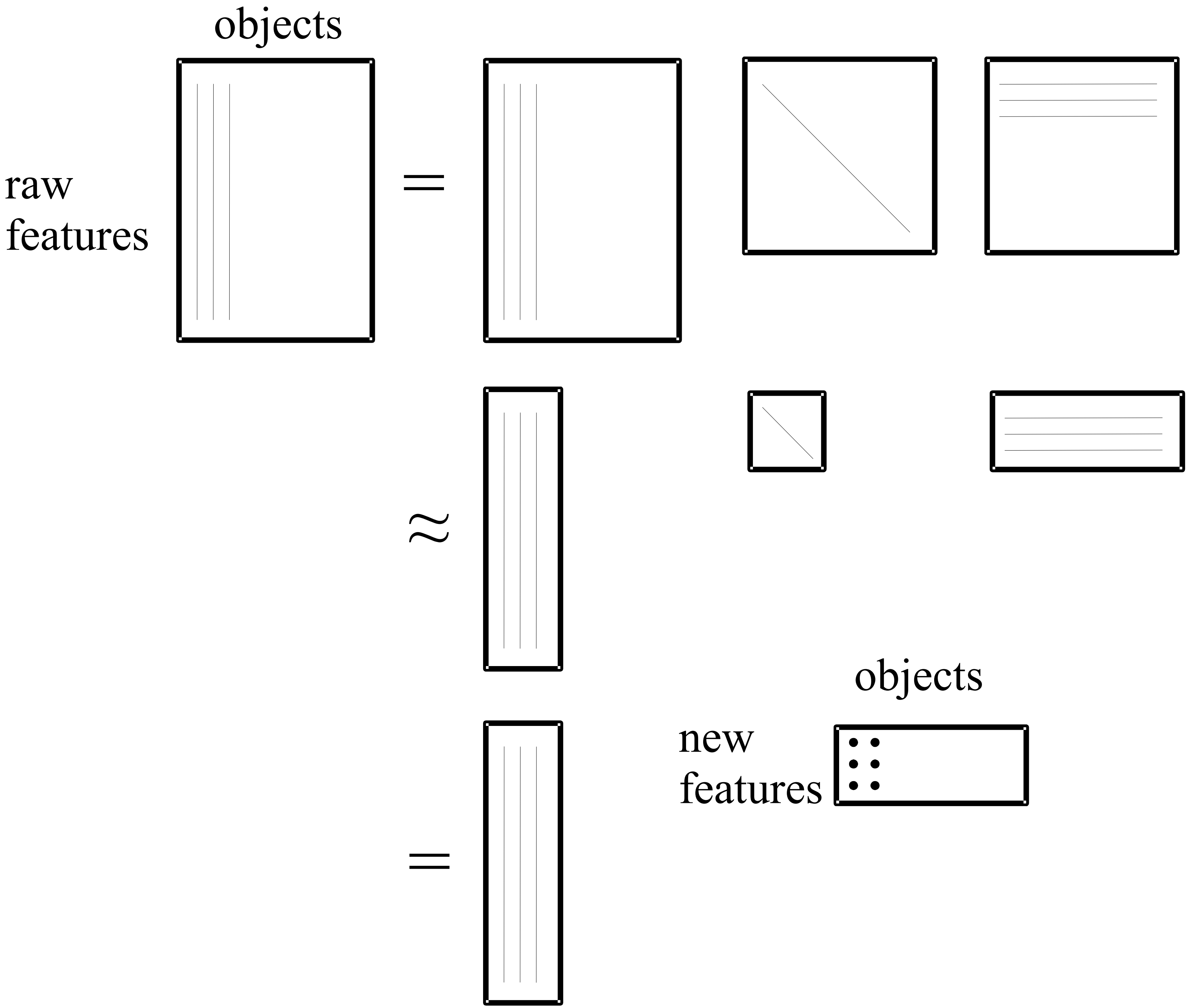
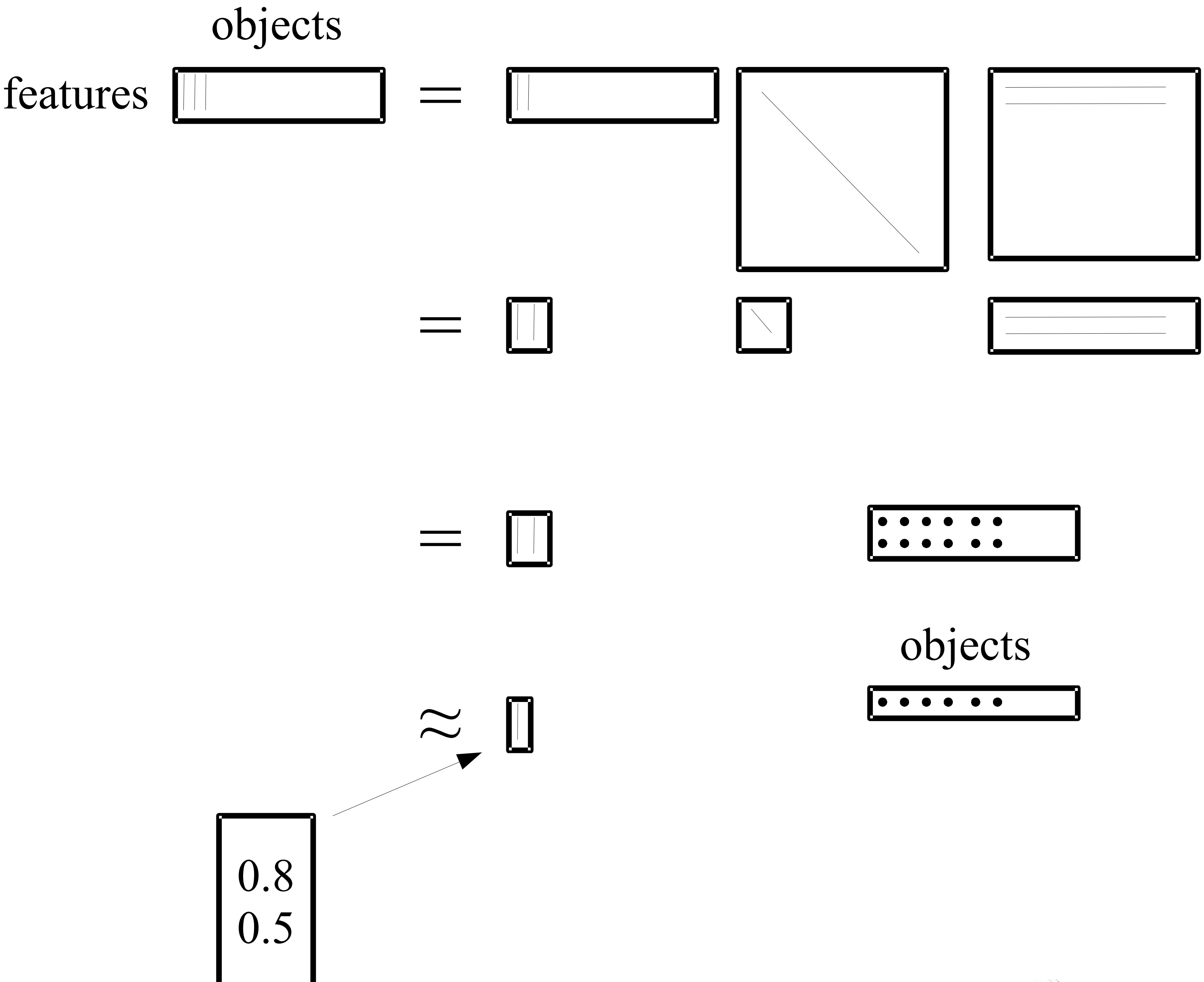


Image removed due to copyright considerations.

# SVD



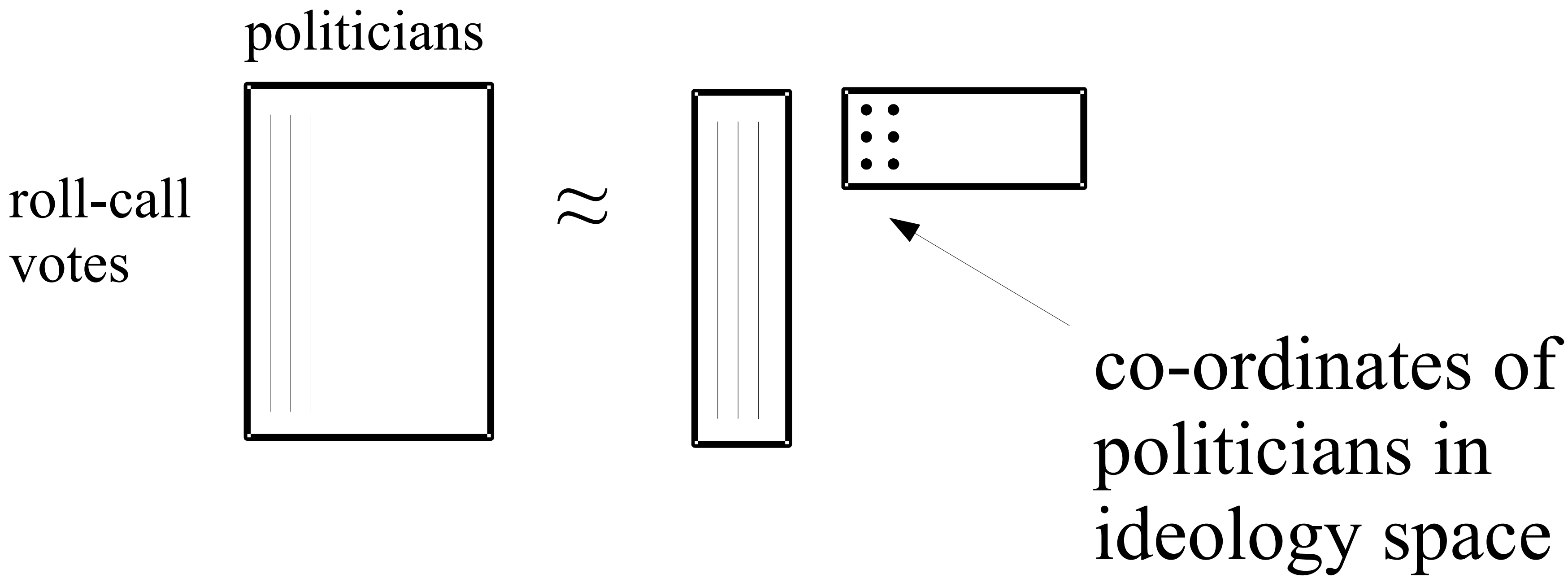
# PCA and MDS

PCA on a raw feature matrix

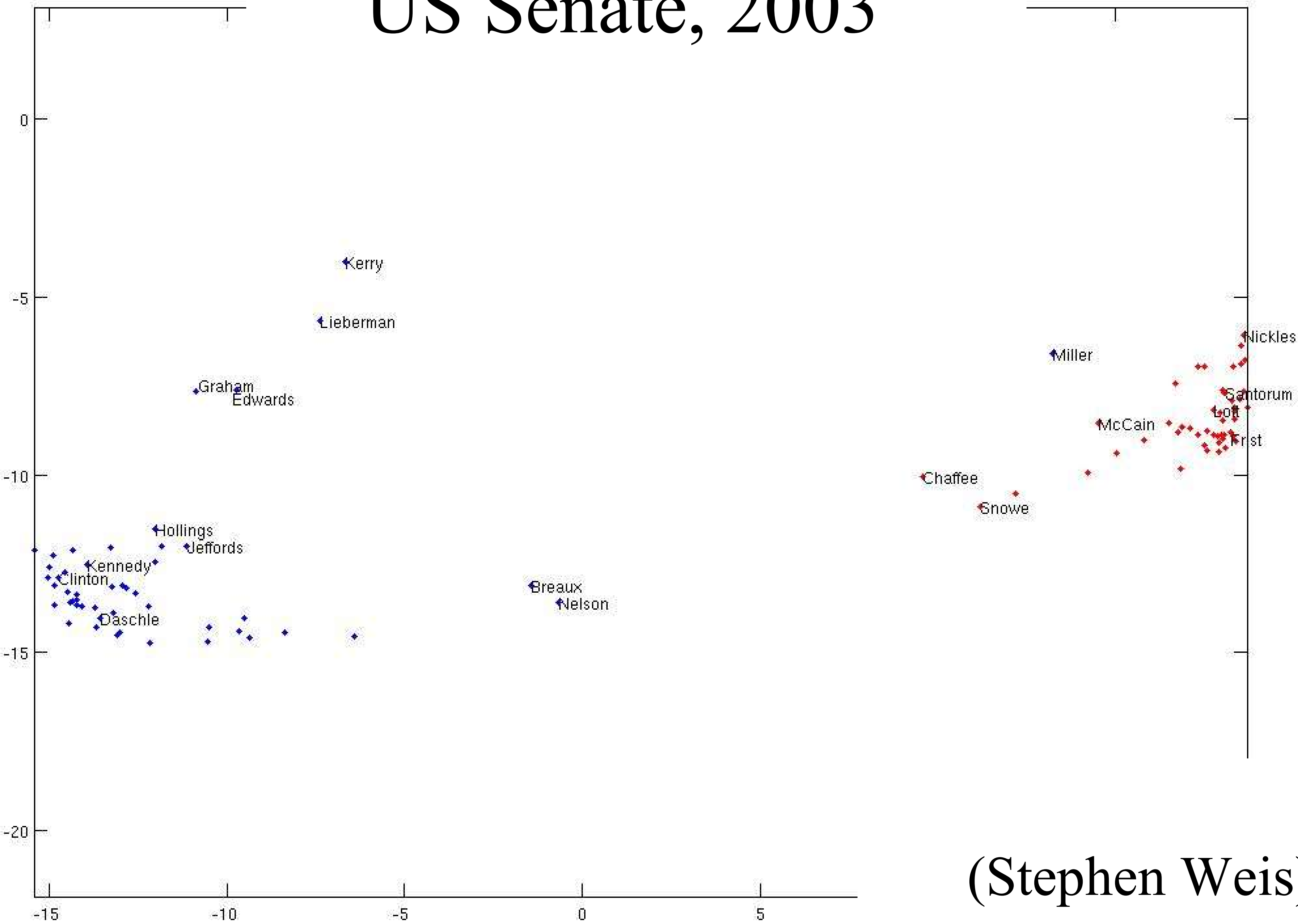
=

Classical MDS on Euclidean distances between feature vectors

# Applications: Politics

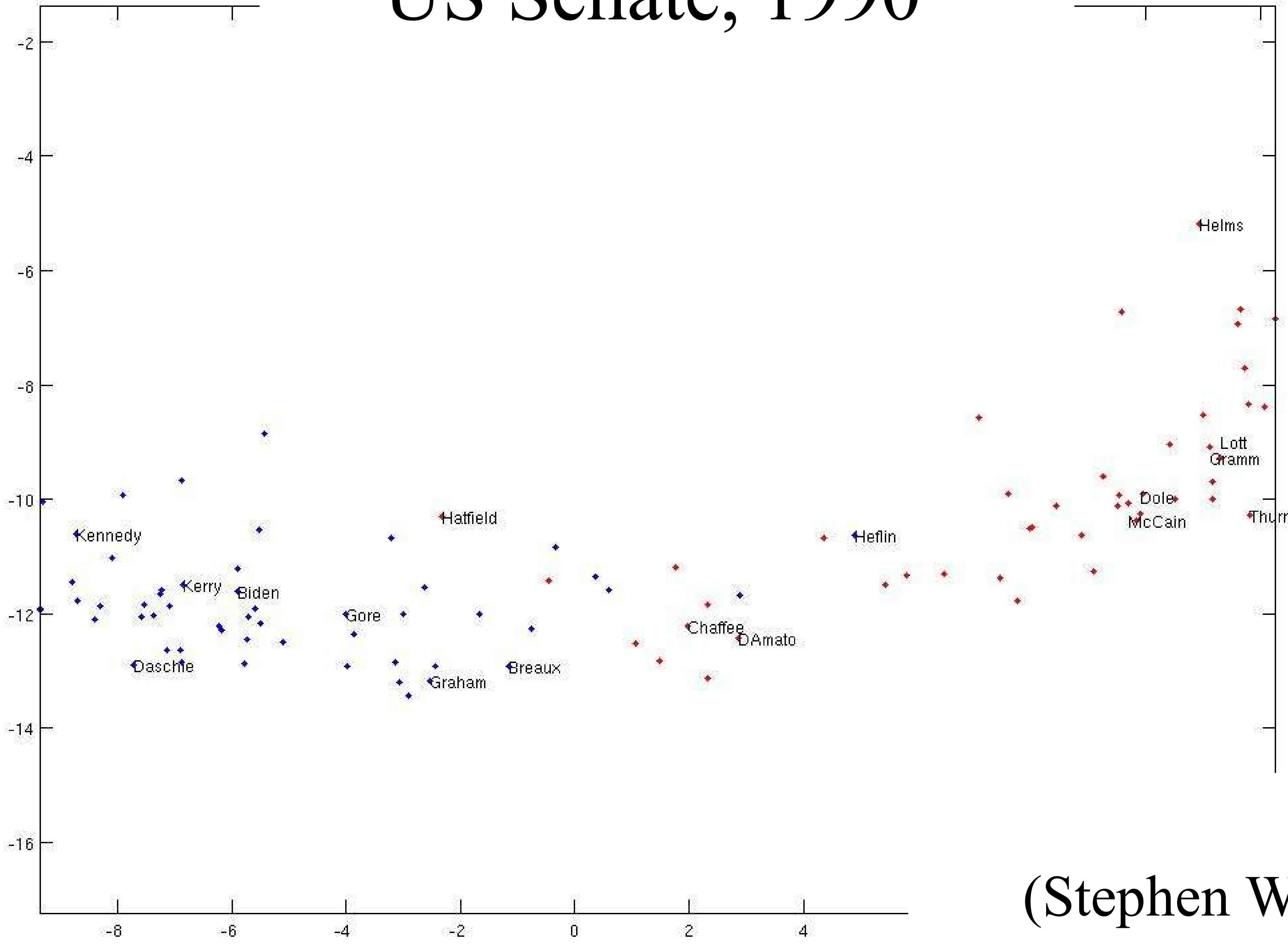


# US Senate, 2003




(Stephen Weis)

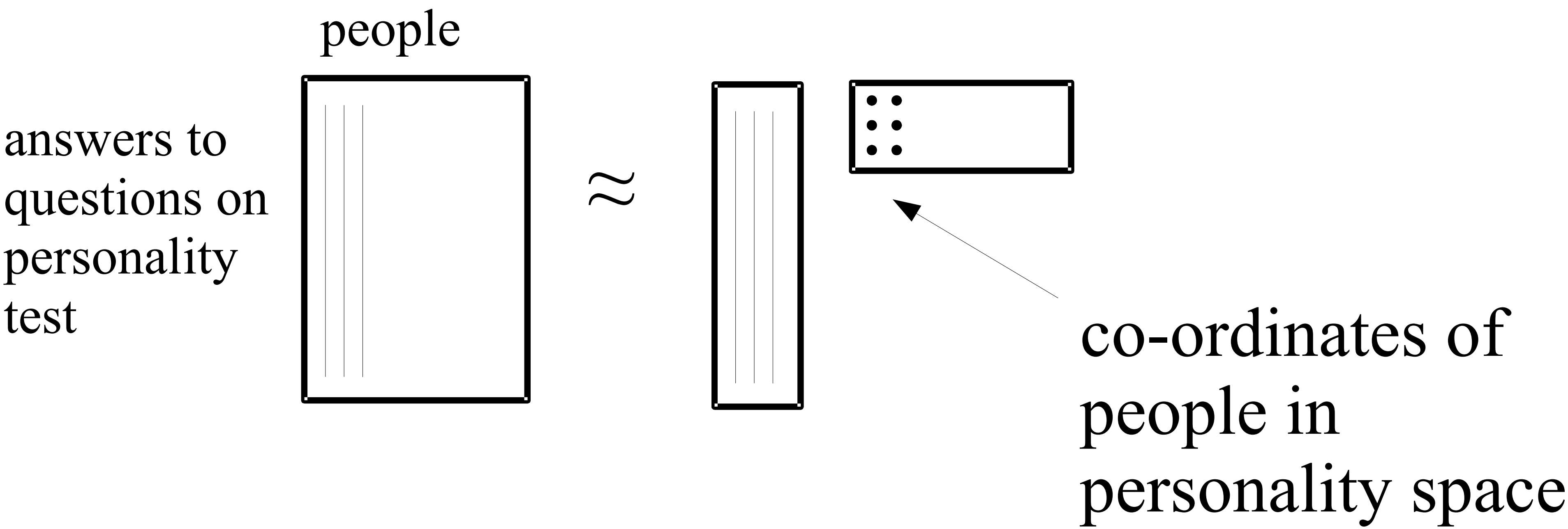
# US Senate, 1990




(Stephen Weis)

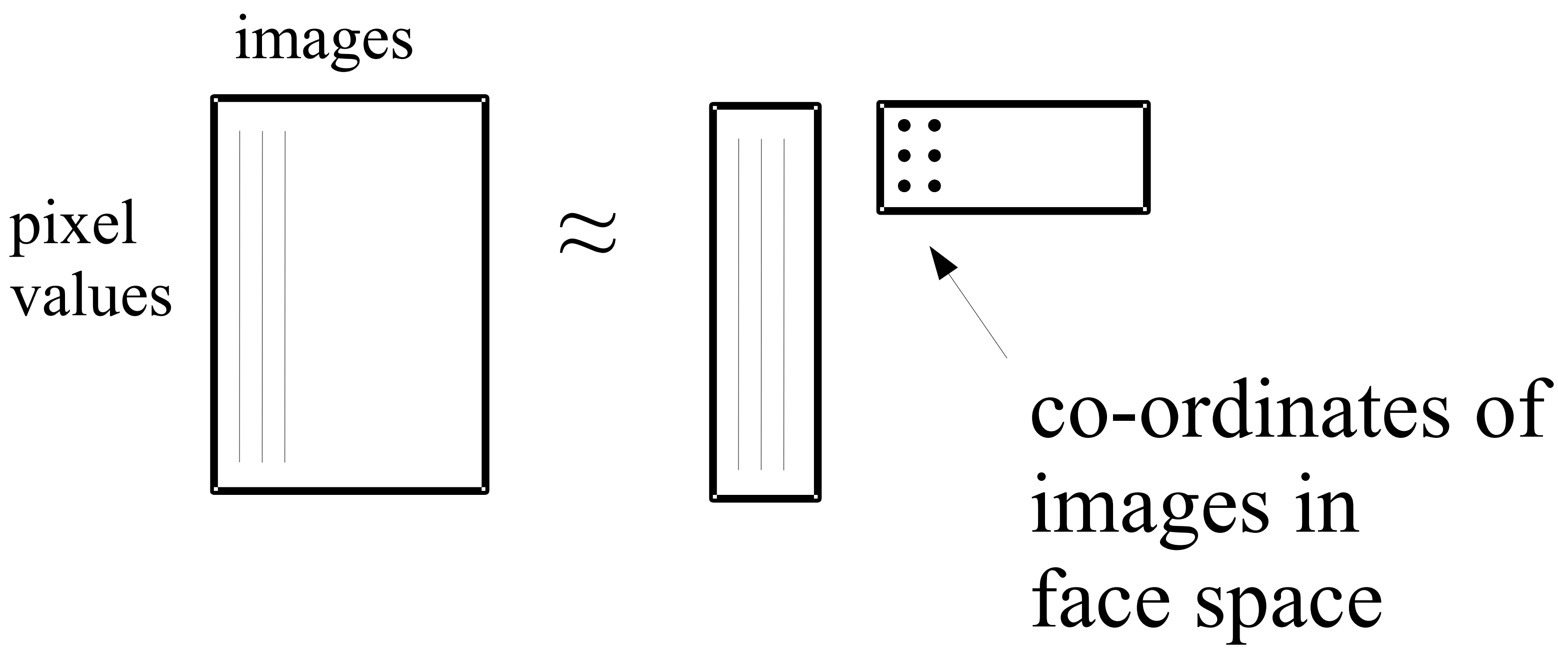
Courtesy of Stephen Weis. Used with permission.

# Applications: Personality




- The Big 5
  - Openness
  - Conscientiousness
  - Extraversion
  - Agreeableness
  - Neuroticism

# Applications: Face Recognition



# Original faces

Image removed due to copyright considerations.

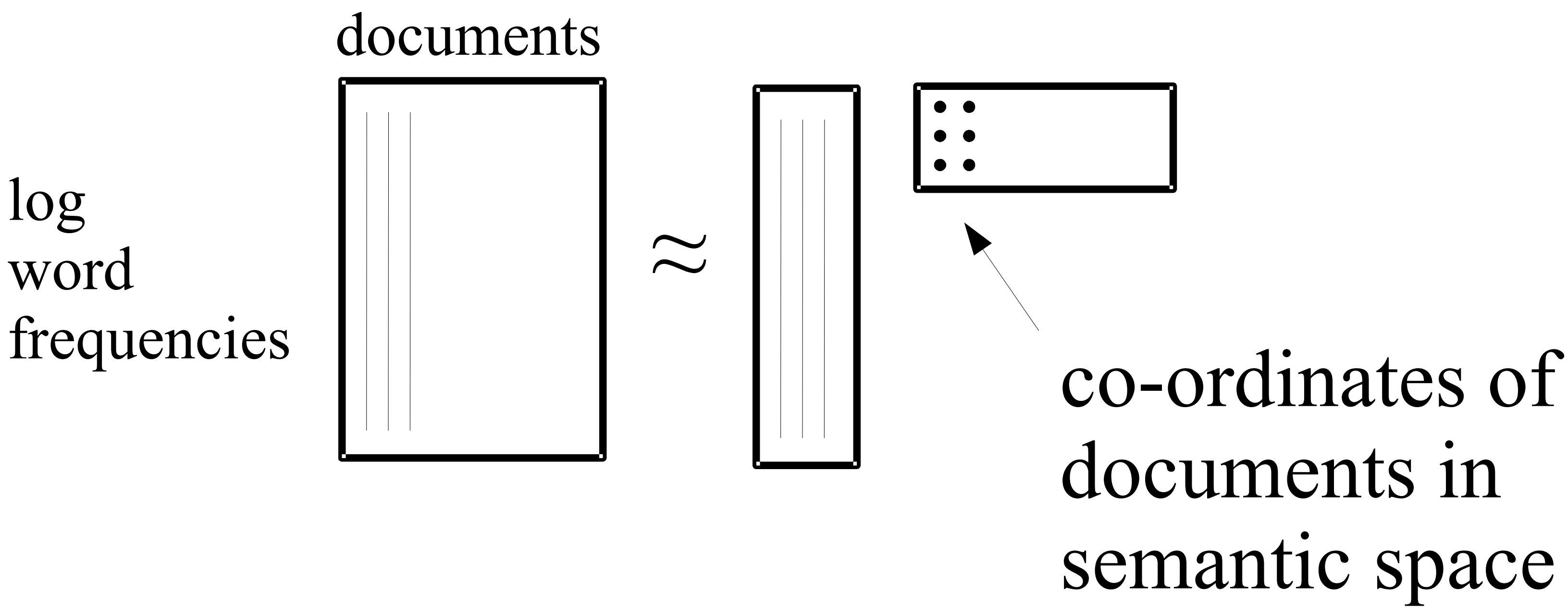
# Principal Components

Image removed due to copyright considerations.

# Face Recognition

- PCA has been discussed as a model of human perception – not just an engineering solution
  - Hancock, Burton and Bruce (1996). Face processing: human perception and principal components analysis

# Latent Semantic Analysis (LSA)




- New documents can be located in semantic space
- Similarity between documents is the angle between their vectors in semantic space

# LSA: Applications

- Essay grading
- Synonym test

# LSA as a cognitive theory

- Do brains really carry out SVD?
  - Irrelevant: the proposal is at the level of computational theory
- A solution to Fodor's problem?
  - Are the dimensions that LSA finds really new?

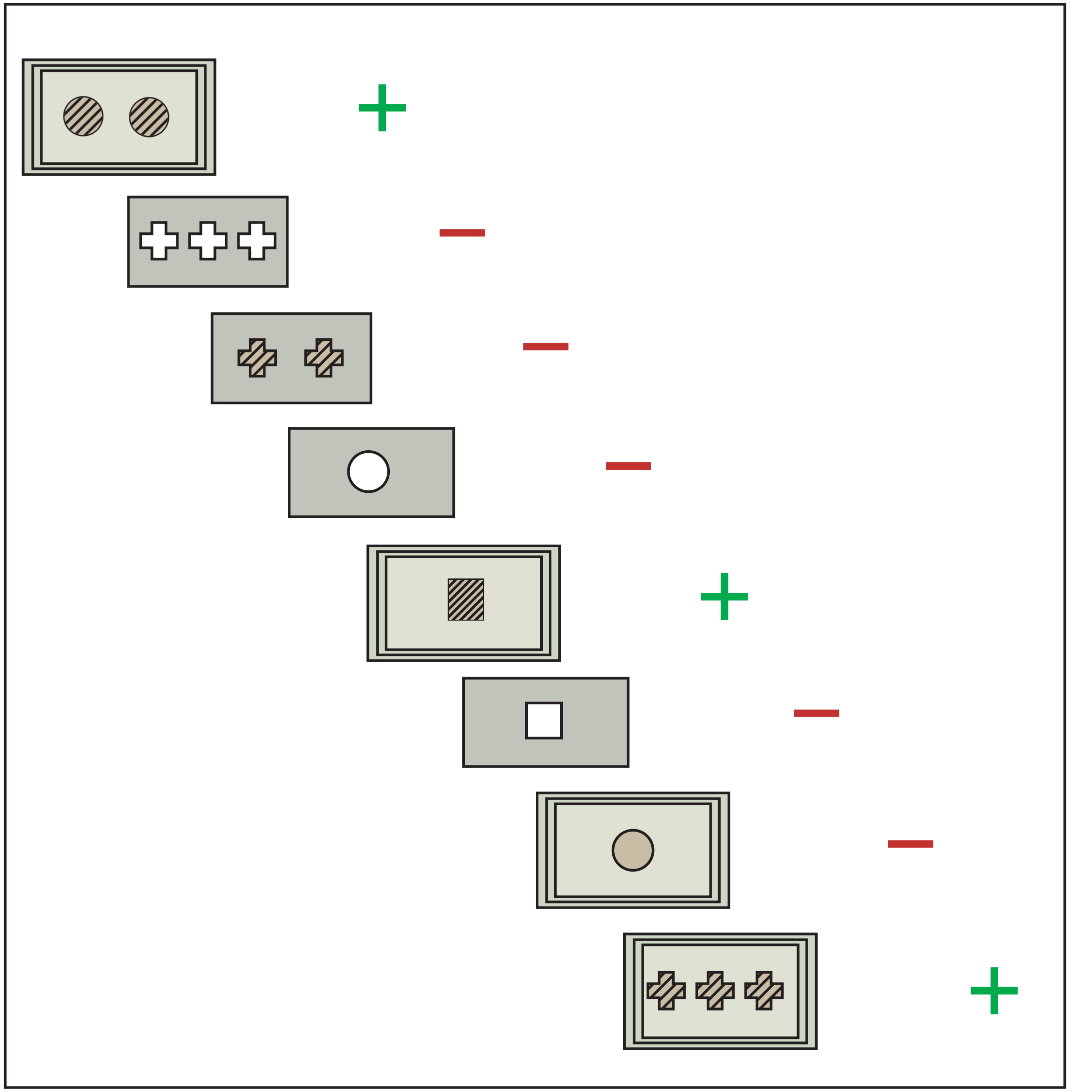


Figure by MIT OCW.

“striped and three borders”: *conjunctive* concept

- Bruner Reading:
  - Raw features: texture (striped, black)
    shape (cross, circle)
    number
  - Disjunctive and conjunctive combinations allowed

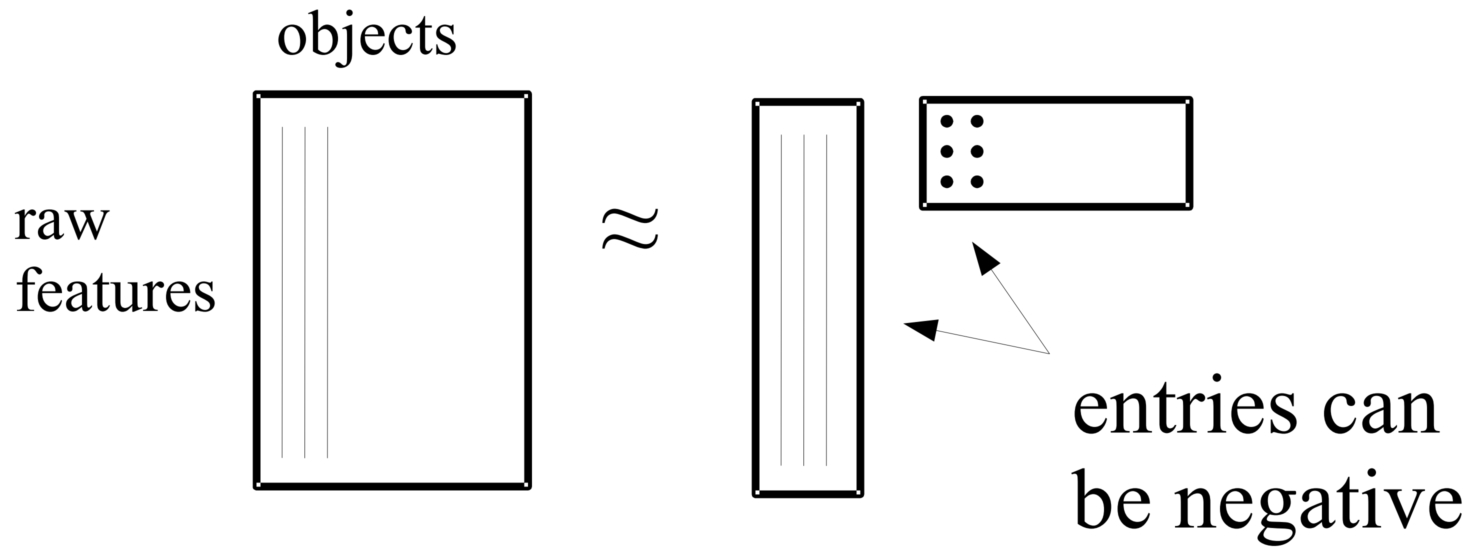
- LSA:
  - Raw features: words
  - Linear combinations of raw features allowed
    (new dimensions are linear combinations of the raw features)

# LSA as a cognitive theory

- Do brains really carry out SVD?
  - Irrelevant: the proposal is at the level of computational theory
- A solution to Fodor's problem?
  - Are the dimensions that LSA finds really new?
- What the heck do the dimensions even mean?

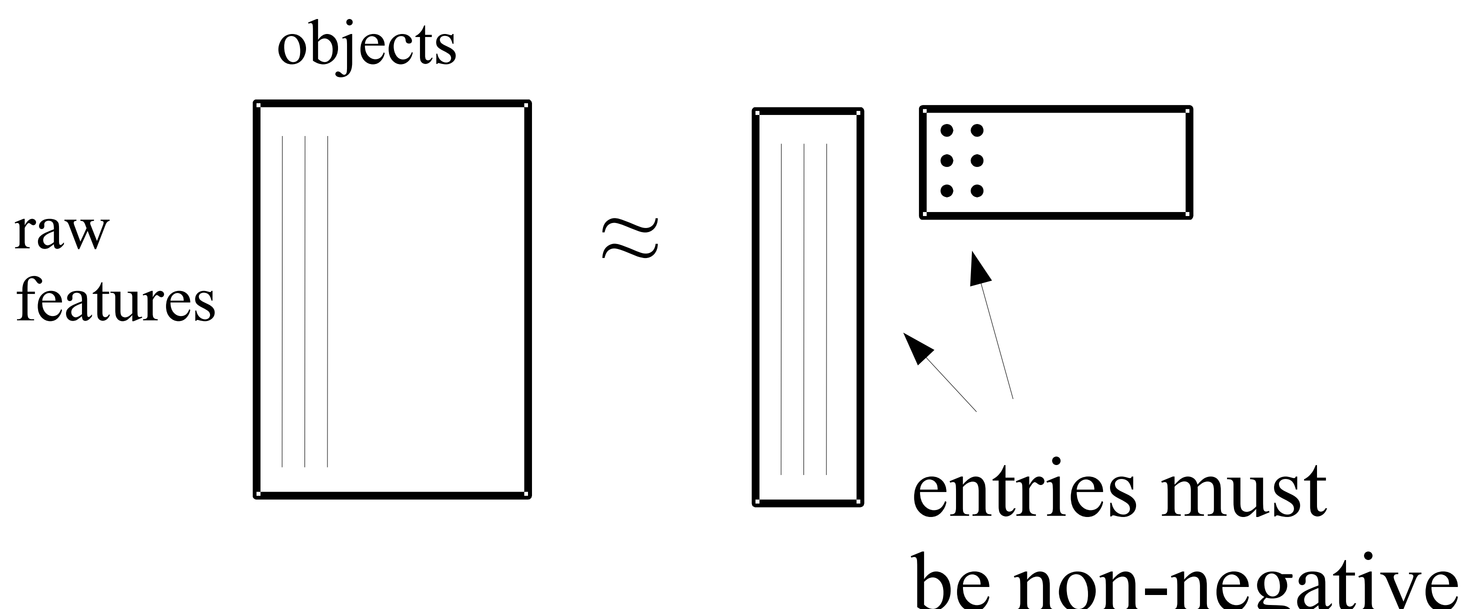
# Non-Negative Matrix Factorization

PCA:




NMF:

(Lee and Seung)



# Dimensions found by NMF

Image removed due to copyright considerations. Please see:

Lee, D. D., and H. S. Seung. "Algorithms for non-negative matrix factorization." *Advances in Neural Information Processing 13*. Proc. NIPS*2000, MIT Press, 2001.

See also Tom Griffiths' work on finding topics in text

# Outline

- Spatial Representations
  - Multidimensional scaling
  - Principal component analysis
- Tree representations
  - Additive trees
  - Hierarchical agglomerative clustering
- Feature representations
  - Additive clustering

# Tree Representations

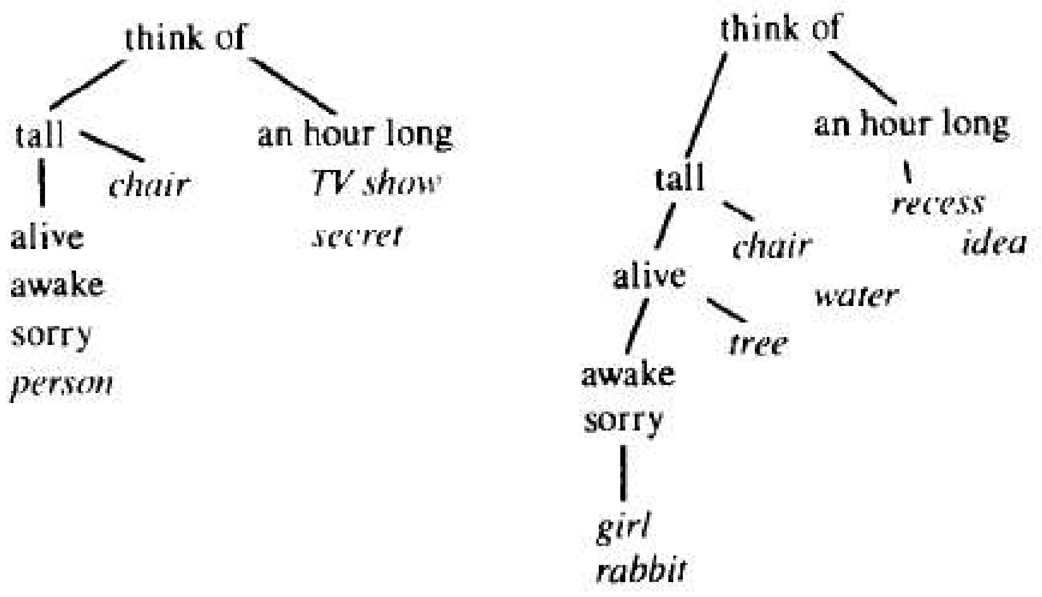
Image removed due to copyright considerations.

# Tree Representations

- Library of Congress system
- Q335.R86

Q: Science
  Q1-Q385: General Science
    Q300-336: Cybernetics
      Q331-Q335: Artificial Intelligence
        Q335.R86: Russell & Norvig, AIMA

# Tree Representations




5-year-old's ontology

7-year-old's ontology

Keil, Frank C. *Concepts, Kinds, and Cognitive Development*. Cambridge, MA: MIT Press, 1989.

# Tree Representations

- We find hierarchical representations very natural. Why ?

  BUT

- Hierarchical representations are not always obvious. The work of Linnaeus was a real breakthrough.

# Today:

- Trees with objects located only at leaf nodes

# ADDTREE (Sattath and Tversky)

- Input: a dissimilarity matrix
- Output: an unrooted binary tree
- Computational Theory

  $d_{ij}$ : distance in a tree

  $\delta_{ij}$ : human dissimilarity ratings

  Want $d_{ij} \approx \delta_{ij}$

- Algorithm:
  - search the space of trees using heuristics

# ADDTREE: example

Image removed due to copyright considerations.

# ADDTREE

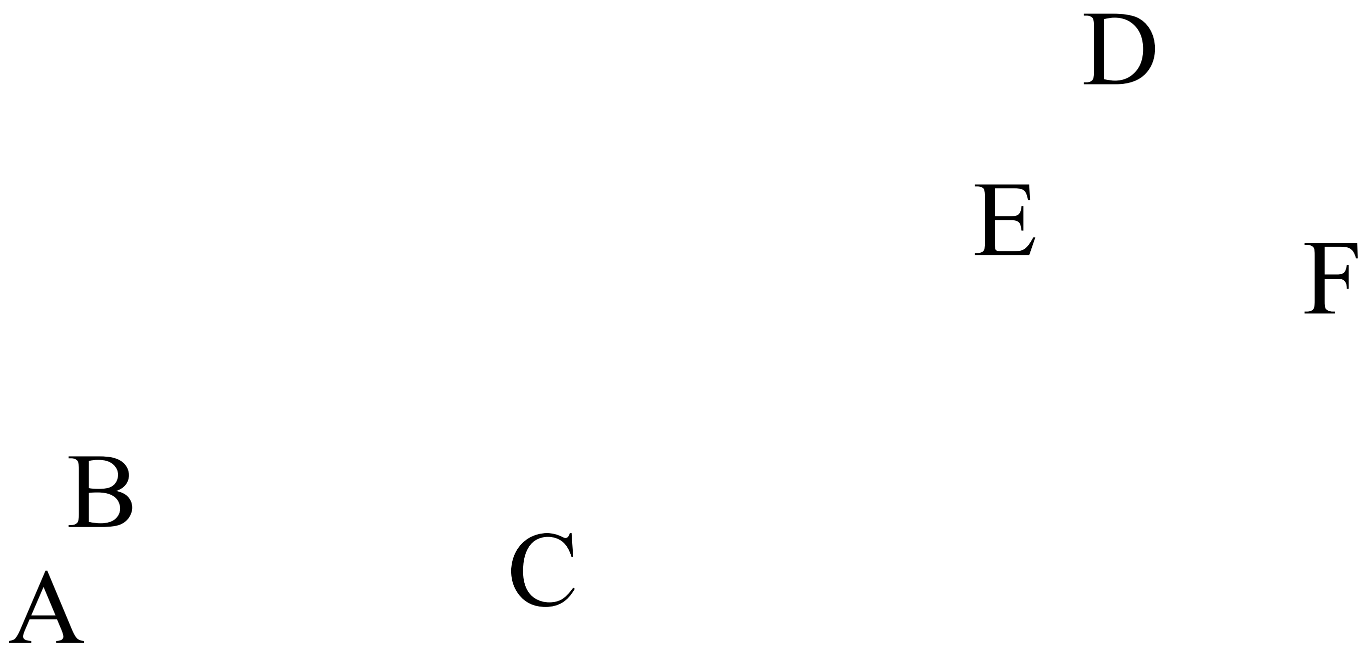
- Tree-distance is a metric

- Can think of a tree as a space with an unusual kind of geometry

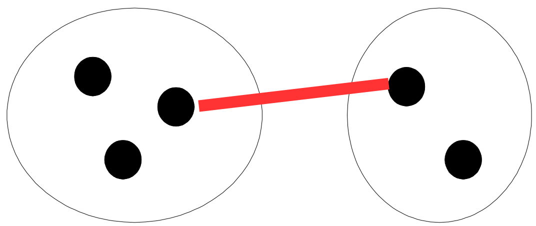
# Hierarchical Clustering

- Input: a dissimilarity matrix
- Output: a rooted binary tree
- Computational Theory
  - ? (but see Kamvar, Klein and Manning, 2002)
- Algorithm:
  - Begin with one group per object
  - Merge the two closest groups
  - Continue until only one group remains

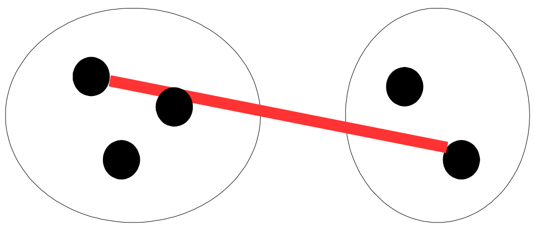
# Hierarchical Clustering



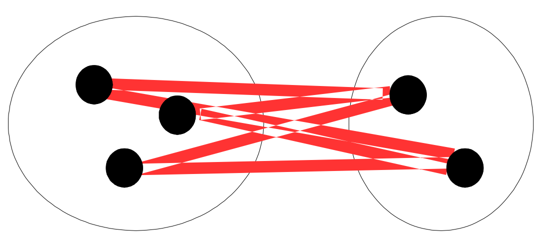
# How close are two groups?



Single-link clustering

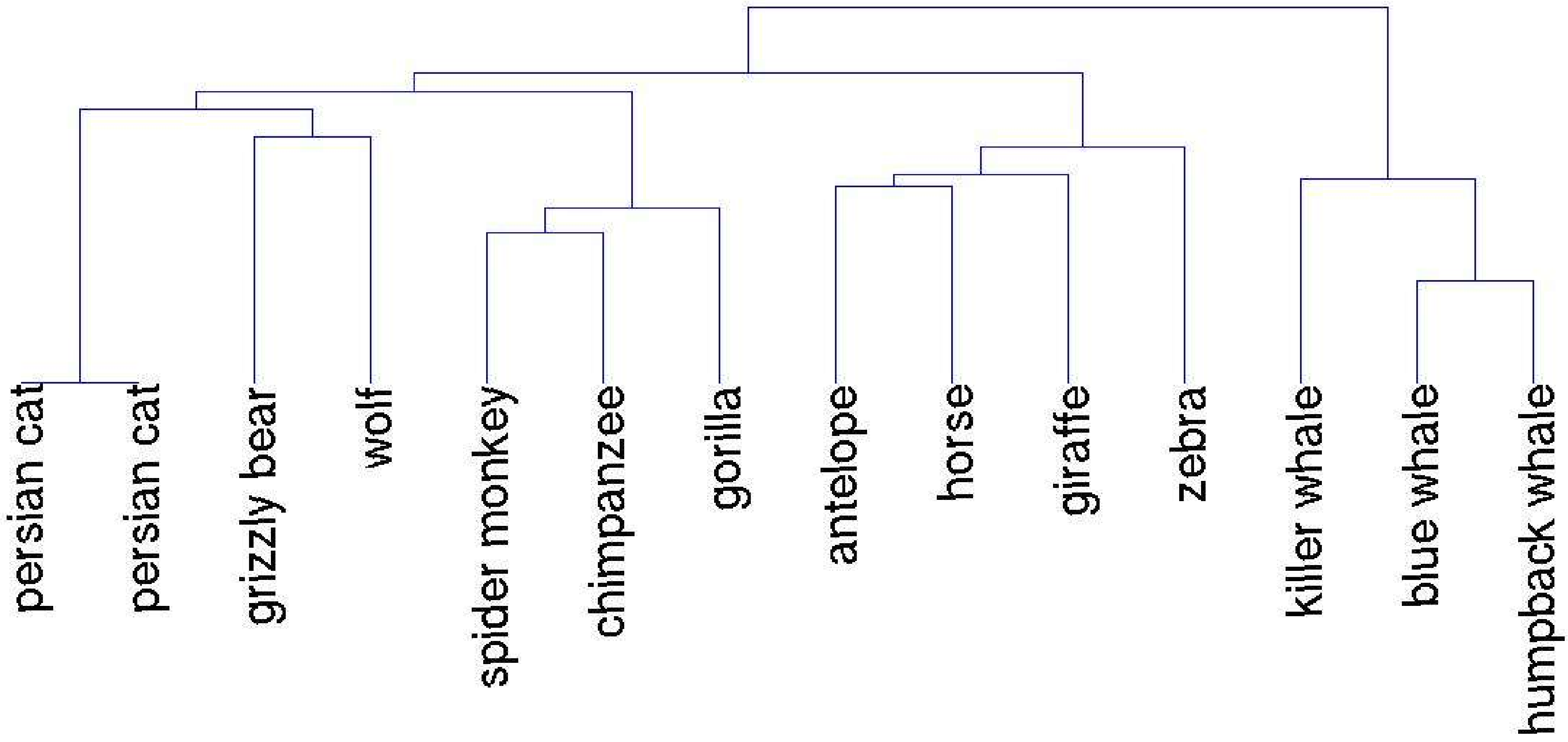


Complete-link clustering

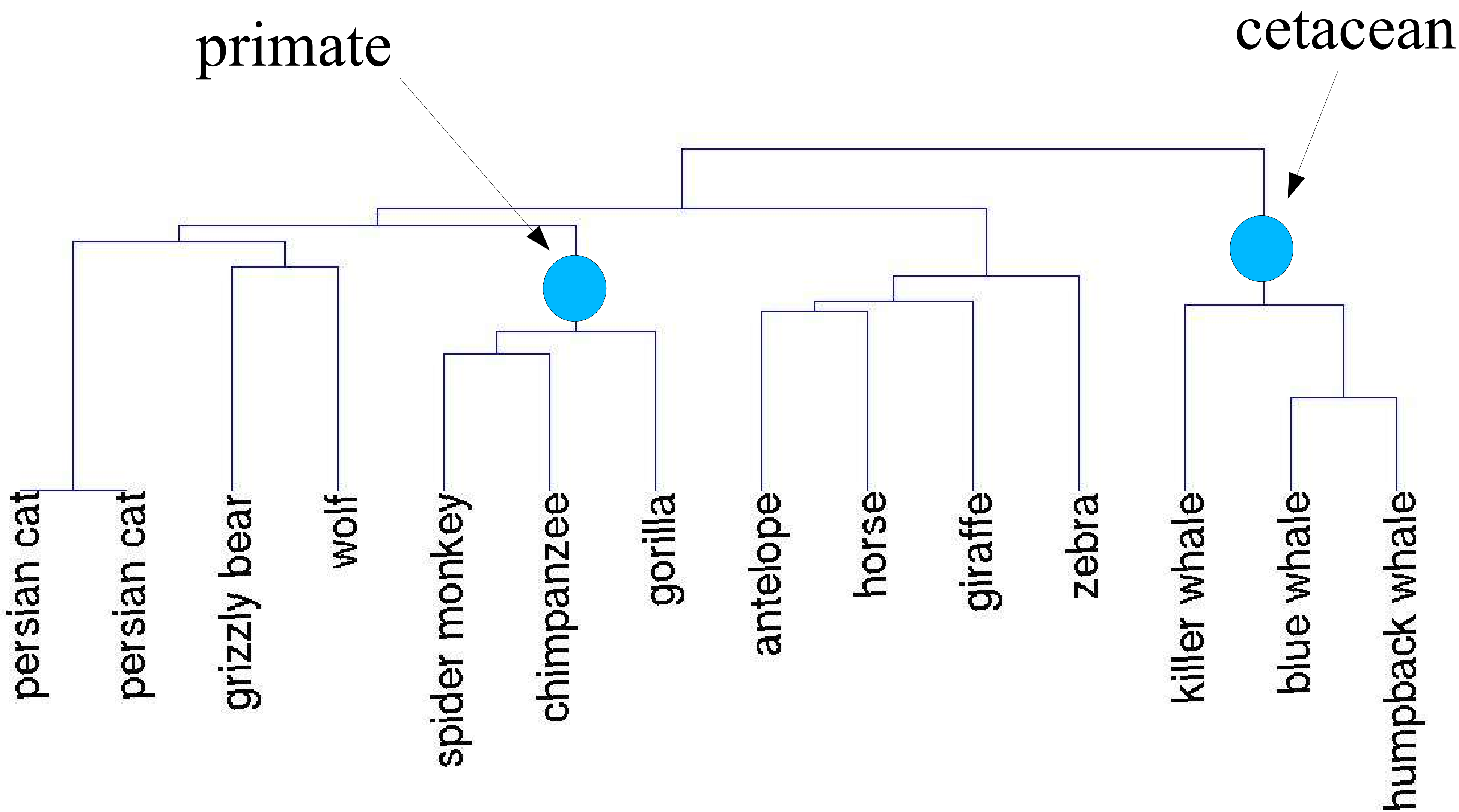


Average-link clustering

# Hierarchical Clustering: Example



# Tree-building as feature discovery



# Outline

- Spatial Representations
  - Multidimensional scaling
  - Principal component analysis
- Tree representations
  - Additive trees
  - Hierarchical agglomerative clustering
- Feature representations
  - Additive clustering

**WARNING: additive clustering is not about trees**

# Additive Clustering

- Representation: an object is a collection of discrete features
  - eg Elephant = {grey, wrinkly, has_trunk, is_animal ...}
- Additive clustering is about discovering features from similarity data

# Additive clustering

$$s_{ij} = \sum_k w_k f_{ik} f_{jk}$$

$s_{ij}$ : similarity of stimuli $i, j$

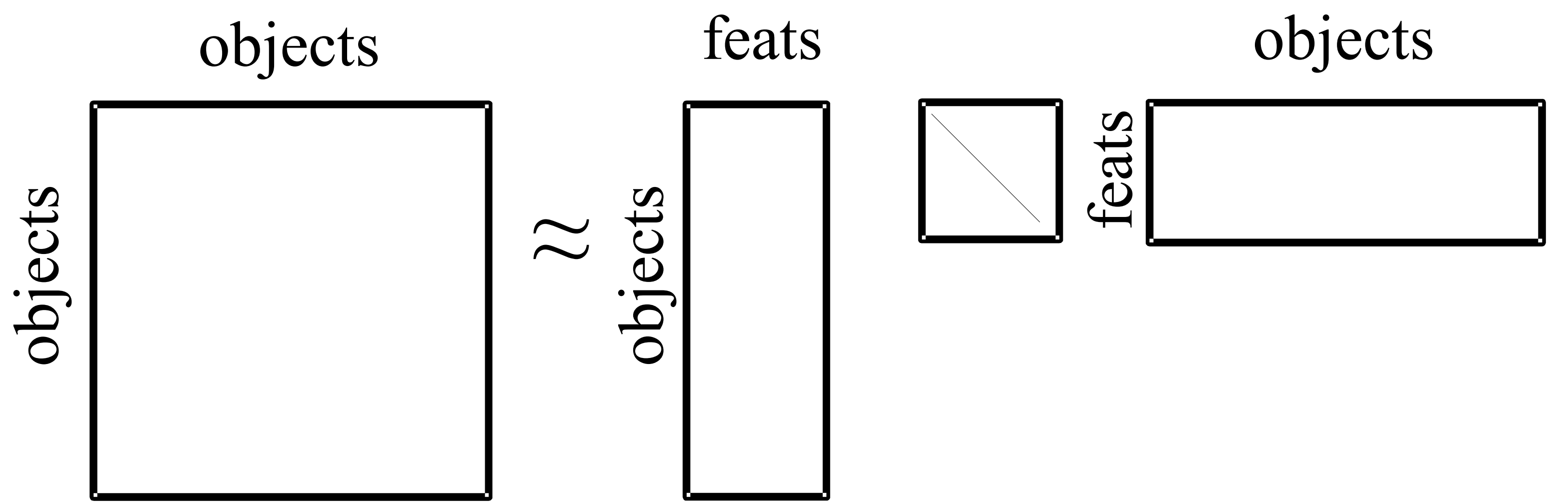
$w_k$ : weight of cluster $k$

$f_{ik}$ : membership of stimulus $i$ in cluster $k$

(1 if stimulus $i$ in cluster $k$, 0 otherwise)

Equivalent to similarity as a weighted sum of common features (Tversky, 1977).

# Additive clustering




$$S \approx FWF^T$$

$$s_{ij} \approx \sum_k w_k f_{ik} f_{jk}$$

# Additive clustering for the integers 0-9:

$$s_{ij} = \sum_k w_k f_{ik} f_{jk}$$

| **Rank** | **Weight** | **Stimuli in cluster** | | | | | | | | | | **Interpretation** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | |
| 1 | .444 | | | * | | * | | | | * | | powers of two |
| 2 | .345 | * | * | * | | | | | | | | small numbers |
| 3 | .331 | | | | * | | | * | | | * | multiples of three |
| 4 | .291 | | | | | | | * | * | * | * | large numbers |
| 5 | .255 | | | * | * | * | * | * | | | | middle numbers |
| 6 | .216 | | * | | * | | * | | * | | * | odd numbers |
| 7 | .214 | | * | * | * | * | | | | | | smallish numbers |
| 8 | .172 | | | | | * | * | * | * | * | | largish numbers |

# General Questions

- We've seen several types of representations. How do you pick the right representation for a domain?
  - related to the statistical problem of model selection
  - to be discussed later

# Next Week

- More complex representations