Classic categorization models

CS786

4th November 2024

Assignment 4

- I have released the assignment already on HelloIITK
- Of the two models used in the assignment,
  - GCM is described in these slides
  - RMC is described in its own set of slides, and I am also sharing code for it with the assignment
- Assignment is due 13th November 11:59PM (one minute before midnight)

Functions of Concepts

- By dividing the world into classes of things to decrease the amount of information we need to learn, perceive, remember, and recognise: **cognitive economy**

- They permit us to make accurate **predictions**

- Categorization serves a **communication** purpose

Outline

- Hierarchical Structure
  - Is there a preferred level of conceptualization?
- Organization of Concepts
  - classical view: **defining-attribute approach**
  - **prototype theory**
  - **exemplar models**
- Semantic similarity
  - **Query likelihood model**

Is there a preferred level of conceptualization?

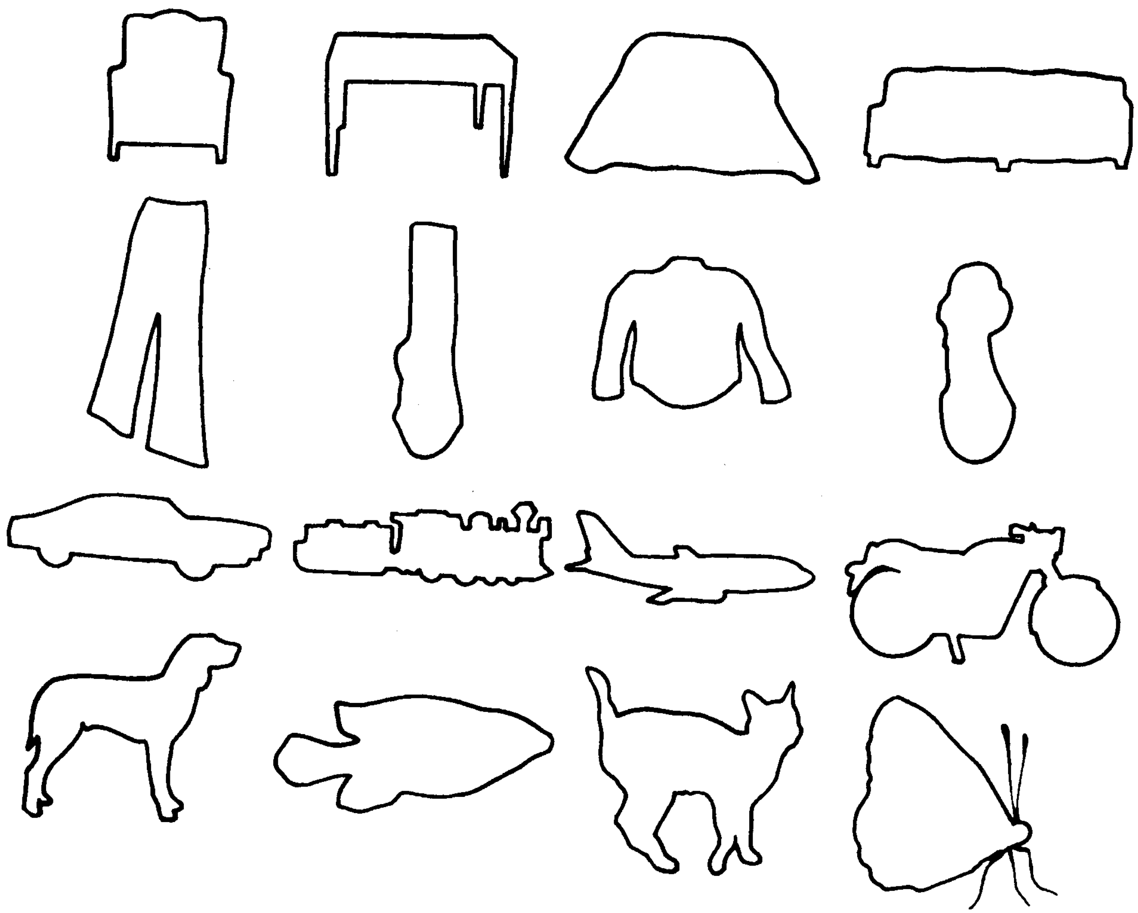
Superordinate level

Preferred level
BASIC LEVEL

Subordinate level

What’s special about the basic level

1) most abstract level at which objects have similar shapes



What’s special about the basic level

2) development

First words are learned at the basic level (e.g., doggy, car, ball)

3) Language

natural level at which objects are named

languages first acquire basic level terms most general

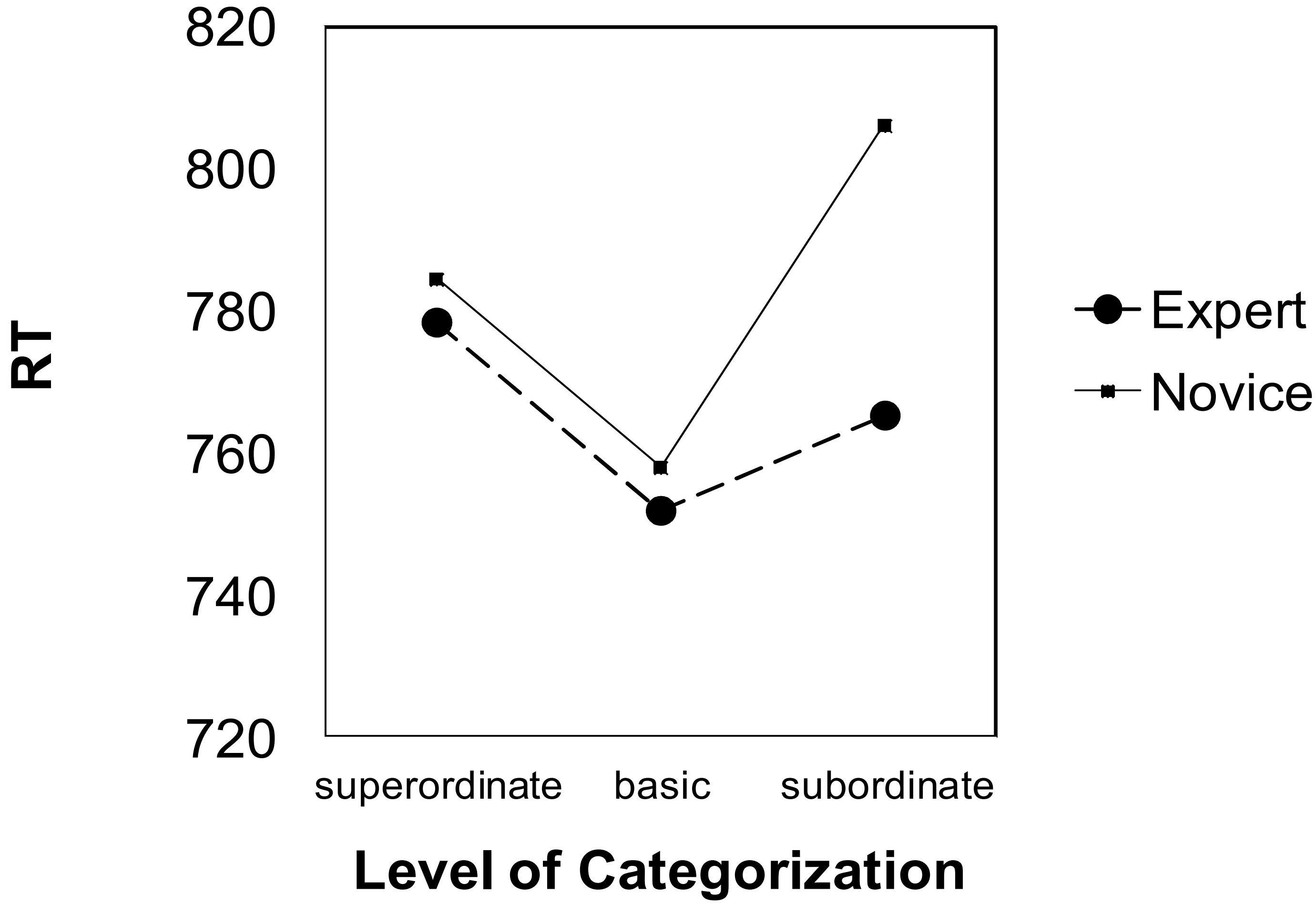
maximize accuracy
little predictive power

BASIC

maximize predictive power
little accuracy

most specific

Basic Level and Expertise



Organization of Concepts

Representation of Conceptual Knowledge

- How do we represent concepts? How do we classify items?

- CLASSICAL VIEW
  - concepts can be defined in terms of *singly necessary* and *jointly sufficient* features

singly necessary:
every instance of the concept must have that property

jointly sufficient:
every entity having all those features must be an instance of the concept

Problems with Classical View

- Bachelor: unmarried, male, adult

- What is a game?
- Ludwig Wittgenstein (1953) proposed that games could not be defined or categorized by features.
- Rather, any game shares some **family resemblance** to some (but not all) other games.

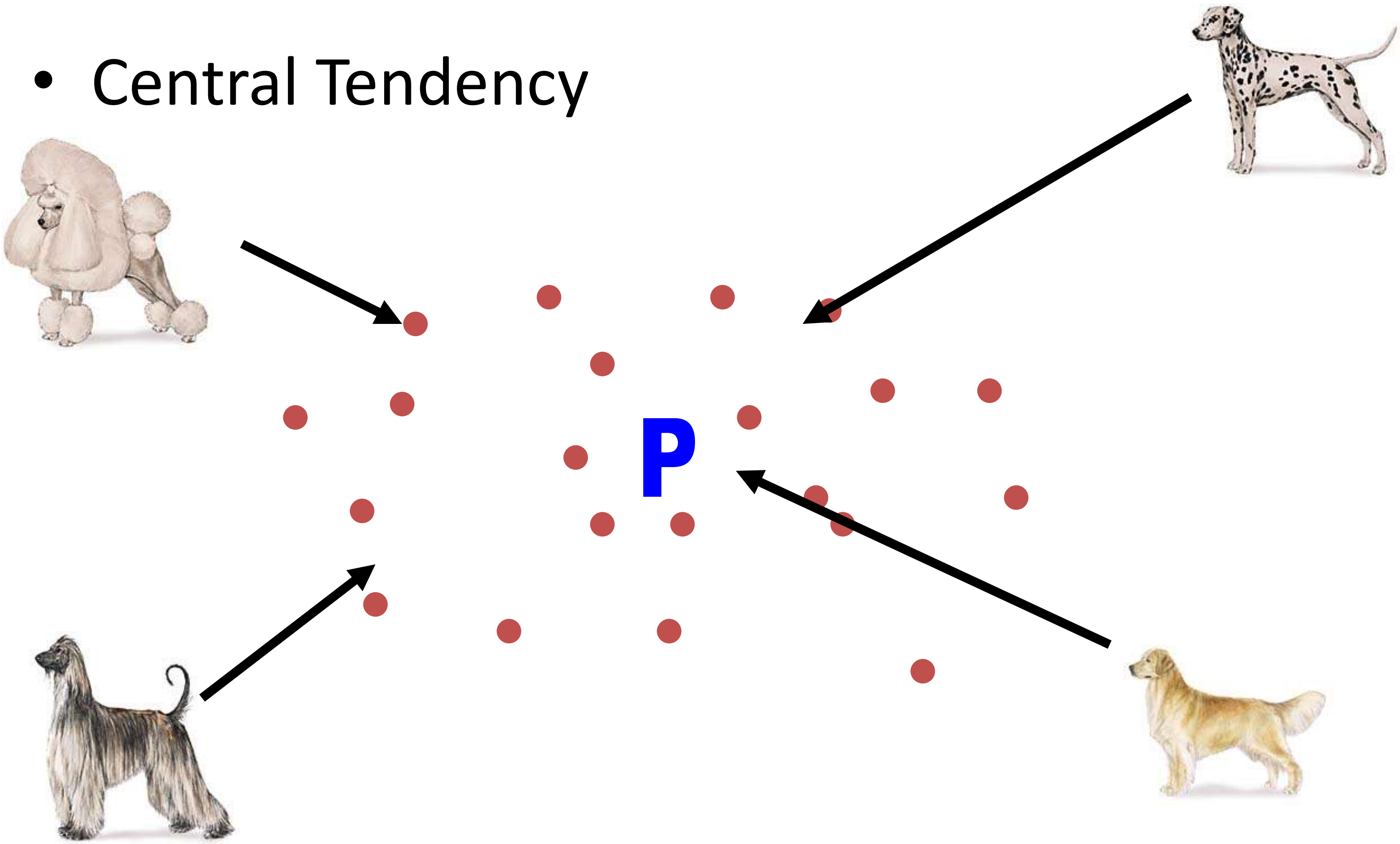
Prototype and Exemplar Models

- A new exemplar is classified based on its **similarity** to a stored category representation

- Types of representation
  - **prototype**
  - **exemplar**

Prototypes Representations

- Central Tendency



Learning involves abstracting a set of prototypes

Typicality Effects

- typical
  - robin-bird, dog-mammal, book-reading, diamond-precious stone
- atypical
  - ostrich-bird, whale-mammal, poem-reading, turquoise-precious stone

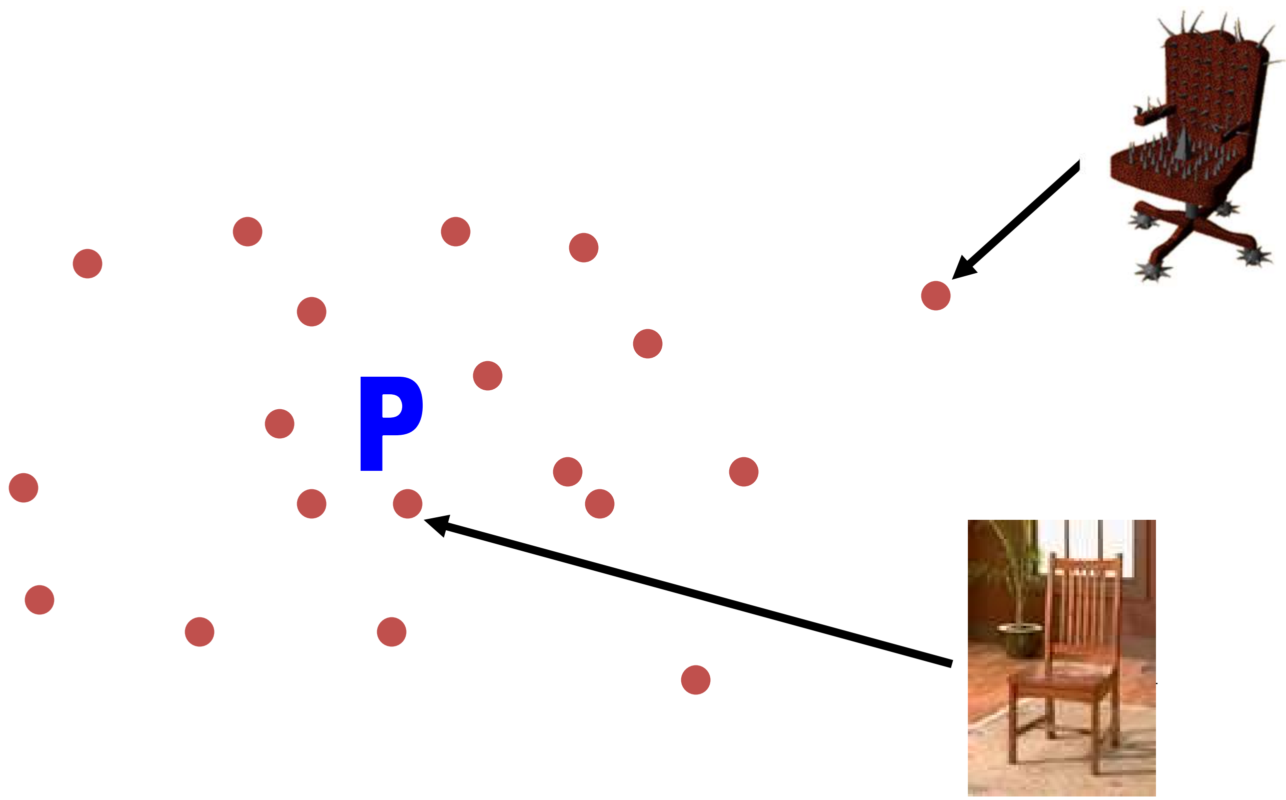
Is this a “chair”?

Is this a “cat”?

Is this a “dog”?

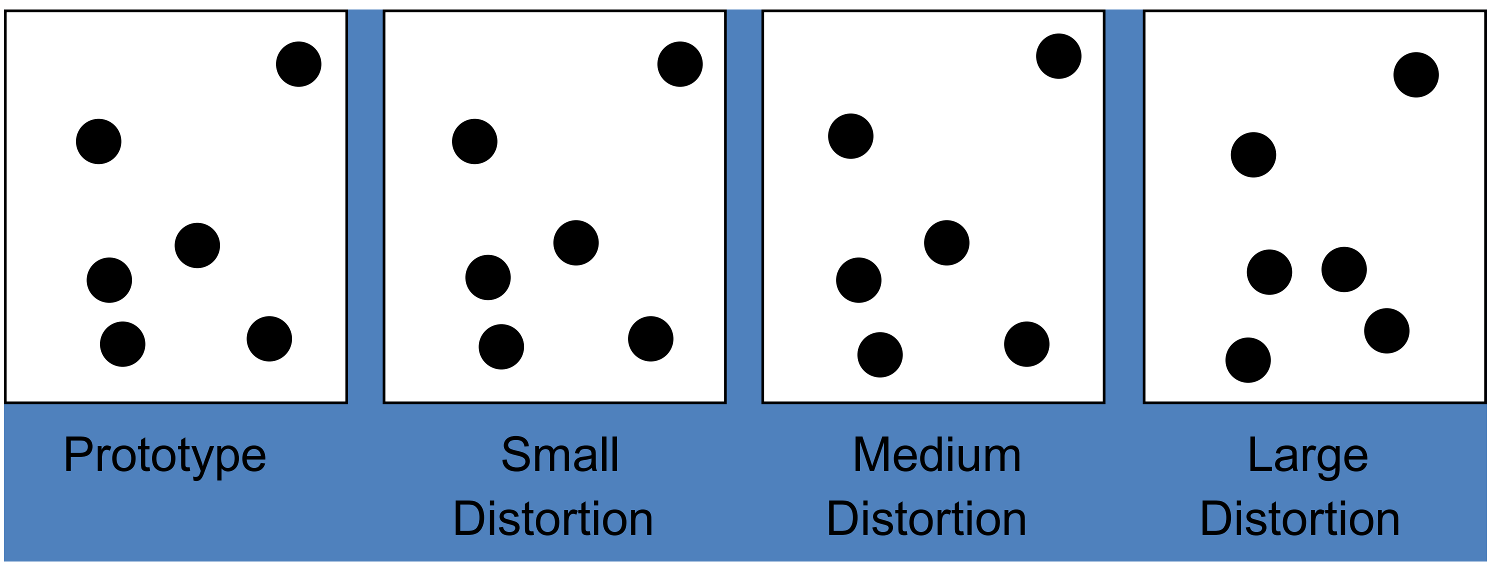
Graded Structure

- Typical items are similar to a prototype
- Typicality effects are naturally predicted



Classification of Prototype

- Prototype are often easy to classify and remember
- Even if the prototype is never seen during learning
- Posner & Keele DEMO:
  http://psiexp.ss.uci.edu/research/teaching/Posner_Keele_Demo.ppt

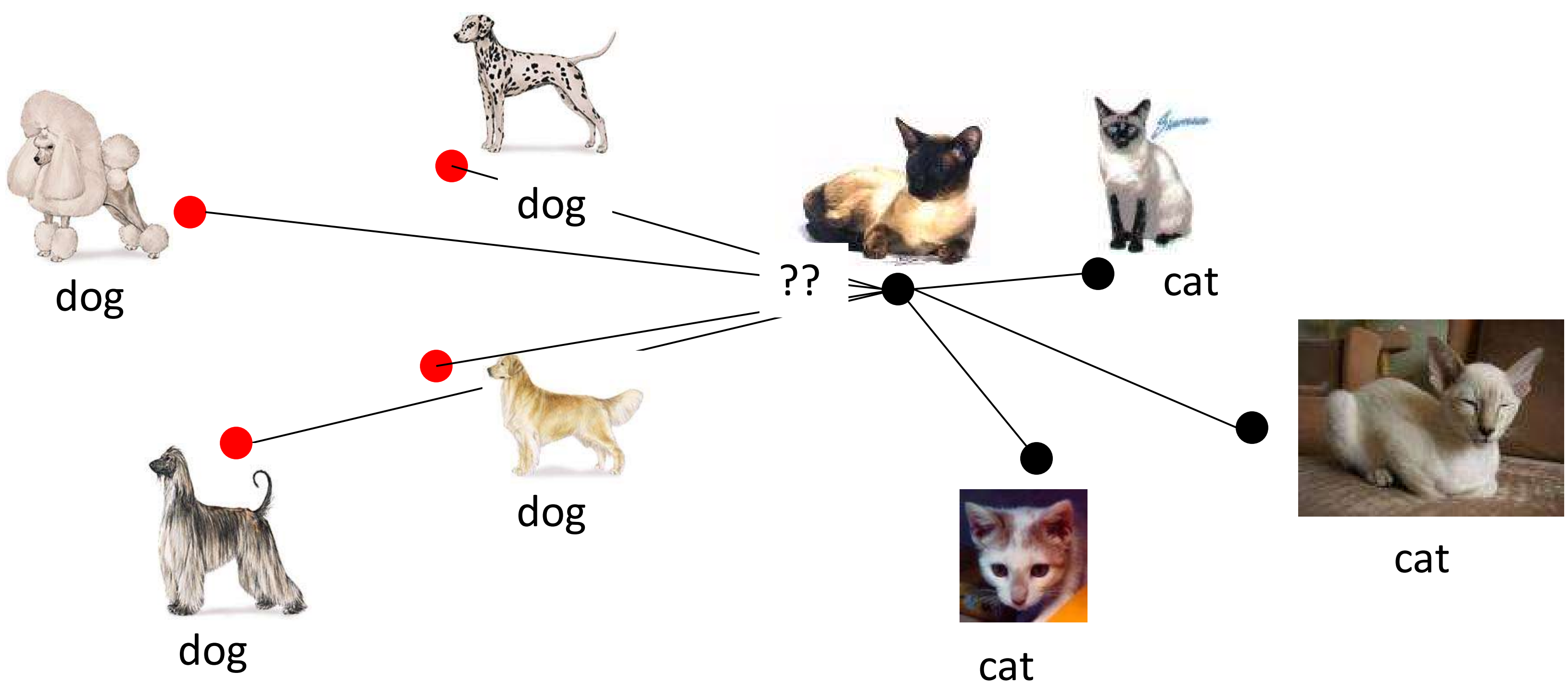


Problem with Prototype Models

- All information about individual exemplars is lost
  - category size
  - variability of the exemplars
  - correlations among attributes (e.g., only small birds sing)

Exemplar model

- category representation consists of storage of a number of category members
- New exemplars are compared to known exemplars – most similar item will influence classification the most



Exemplar Models

- Model can explain
  - Prototype classification effects
    - Prototype is similar to most exemplars from a category
  - Graded typicality
    - How many exemplars is new item similar to?
  - Effects of variability
    - pizzas and rulers
- Overall, compared to prototype models, exemplar models better explain data from categorization experiments (Storms et al., 2000)

Sample exemplar model

- Nosofsky’s 1986 Generalized Context Model (GCM) has been very influential
- Stimuli stored in memory as combinations of features
- Context for a feature are the other features with which it co-occurs
- Assumes that stimuli are points in interval-scaled multidimensional space

GCM similarity function

- Compute psychological distance between memory exemplar x and stimulus y as

$$d(x,y) = \sum_i \alpha_i (x_i - y_i)$$

- Alpha are attention weights
- Similarity is calculated as

$$s(x,y) = \exp(-\beta d(x,y))$$

- Note: Distance function is always greater than zero
  - Use abs outside the summation if necessary

Category response in GCM

- Exemplars vote for the category with which they are associated

$$p(R|y) = \frac{\gamma_R \sum_{x \in R} N(R,x) s(x,y)}{\sum_r \gamma_r \sum_{k \in r} N(r,k) s(k,y)}$$

- N(R,x) is the count of the number of times x has been recorded as being in category R before
- Gamma is a response bias parameter
- Equation is basically counting total votes cast for category R by exemplars divided by total votes cast

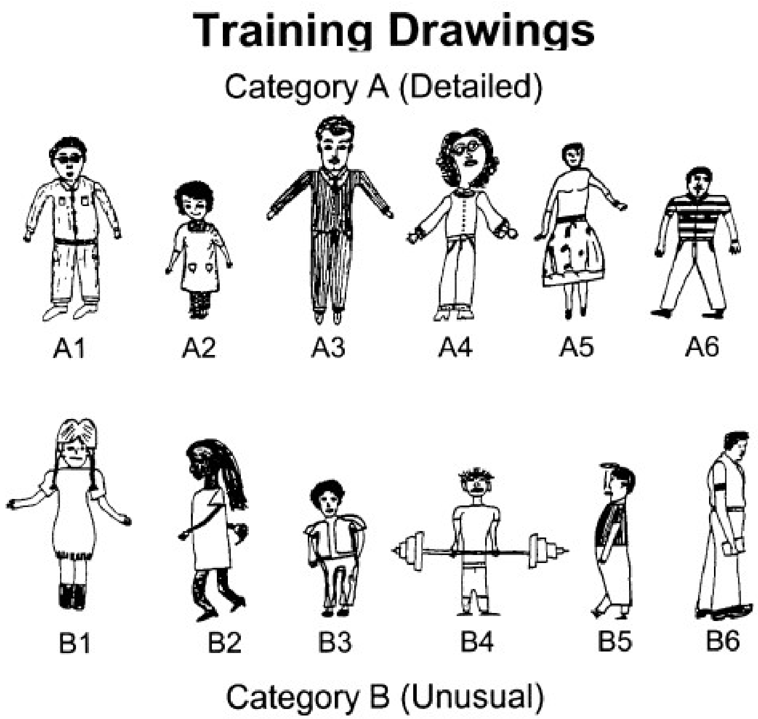
What do the parameters do?

- Gamma reflects environmental priors on categorization
- Beta reflects the bias-variance tradeoff in similarity judgments
- What does alpha do?
  - Reflects the role of semantic knowledge in categorization

Knowledge-based Views

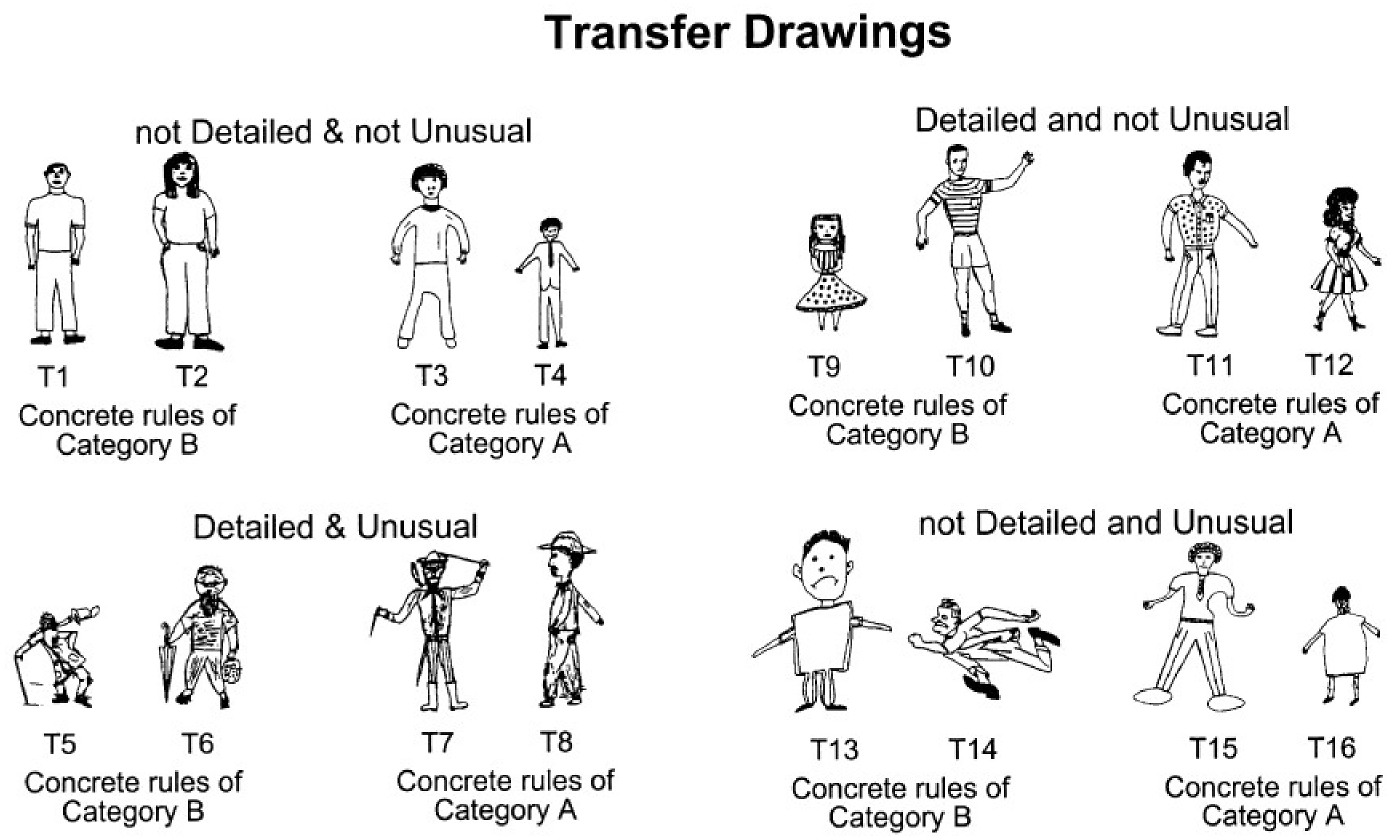
- Murphy (2002, p. 183):
  - "Neither prototype nor exemplar models have attempted to account for knowledge effects . . . The problem is that these models start from a kind of *tabula rasa* [blank slate] representation, and concept representations are built up solely by experience with exemplars."

Effect of Knowledge on Concept Learning



- Concept learning experiment involving two categories of children's drawings
- Two learning conditions:
  - neutral labels for categories (Group 1 vs. Group 2 children)
  - Category labels induced use of background knowledge: "Creative and non-creative children created category A and B drawings respectively"
- Note: same stimuli are used in both conditions

Palmeri & Blalock (2000)



- By manipulating the meaningfulness of the labels applied to those categories of drawings, subjects classified new drawings in markedly different ways. E.g., neutral labels led to an emphasis of concrete features. The “creative vs. non-creative” labels led to an emphasis of abstract features

- Background knowledge and empirical information about instances closely interact during category learning

Palmeri & Blalock (2000)

Learning an exemplar model from labels

- Original GCM model had no learning
  - Parameters fit to data
  - Basically just a clustering model (unsupervised)
- Later models offer learning mechanisms
- Kruschke's ALCOVE model (1992)
  - Assumes a supervised learning setting
  - Learner predicts categories
  - Teacher teaches true category

Supervised learning in ALCOVE

- Activation of category k given stimulus y

$$a_k = \sum_x w_{xk} s(x, y)$$

- Training loss function

$$L = \sum_k (t_k - a_k)^2$$

- Where t is a training label that is 1 if the predicted response is correct and 0 otherwise

Optimization using gradient descent

- All weights and parameters are learned using gradient descent
- Weight update

$$\Delta w_{kx} = \lambda_w (t_k - a_k) s(x, y)$$

- Exemplar-wise error

$$\epsilon_x = \sum_k w_{kx} s(x, y)(t_k - a_k)$$

- Attention update

$$\Delta \alpha_i = -\lambda_\alpha \sum_x \epsilon_x |x_i - y_i|$$

Variations

- GCM-class models assume the presence of interval-scaled psychological distances
- Can make different assumptions about similarity function, e.g. categorical instead of continuous scale
  - # of matches
  - # of mismatches
  - # matches - # mismatches
- Can make different assumptions about the learning mechanism
  - Anderson's Rational Model of Categorization
  - We will see this next