An Introduction to Learning

Lecture 9/13

Todd M. Gureckis
Department of Psychology
New York University
todd.gureckis@nyu.edu
Agenda for Today

- More complex forms of generalization: Categorization
  - Basic models: similarity and attention
  - Advanced models: Abstraction and multiple representations
  - Cognitive neuroscience approaches
1 Categorization
Basic Models
Categories and Concepts
Functions of Concepts

Categories have many functions:

- **Classification** - allows us to treat different things as the same
- **Communication** - we communicate using words that refer to more abstract ideas/concepts
- **Prediction and reasoning** - we can use categories to make predictions about unknown or unseen parts of the world
Example

What you see:
Red
Shiny
In a tree

What you can then infer:
Has seeds
Sweet
Edible
Healthy
A little history....
A little history....
A little history....

The Classical View

The Prototype View

Progress (hopefully)
A little history....
A little history....

- The Classical View
- The Prototype View
- The Exemplar View

Progress (hopefully)
A little history....

The Classical View

The Prototype View

The Exemplar View

The Theory View

Progress (hopefully)
A little history....
The Classical View

- According to the **classical view**, concepts are like definitions.

- The defining features of are both necessary and sufficient.
  
  - **Necessity**: If something is a category member, it has the defining features.
  
  - **Sufficiency**: If something has the defining features, it is a category member.
Example

- **Defining features:**
  - Closed figure, three sides, interior angles sum to 180 degrees

- **Sufficiency:**
  - If something is a closed figure, has three sides and angles sum to 180 degrees it is a triangle

- **Necessity:**
  - If something is a triangle, it is a closed figure, has three sides, and the angles sum to 180 degrees
Learning Classical Concepts

According to the classical view, category learning usually involves hypothesis testing or rule discovery:

- A search for the defining features

Hull, 1920 - phd thesis

Studied learning concepts defined by simple features
Learning Classical Concepts

According to the classical view, category learning usually involves hypothesis testing or rule discovery:

- A search for the defining features

Bruner, Goodnow, & Austin, 1956

Four dimensional concepts involving conjunctions of features, disjunctions, etc...
Rules are one basis for complex forms of generalization

... but, problems for the classical view

- Are all concepts represented in terms of rules? In the 70s there became strong philosophical and empirical arguments against this
One prediction

1: Do defining feature exist?

- Hampton (1979): Asked subjects for necessary and sufficient features of everyday categories (sofas, cars, dogs, chairs, birds, etc...)

- There was little agreement about what the defining features were

- However, people might not explicitly know the features and agreement between individuals doesn’t seem problematic per-se (although perhaps a little surprising)
One prediction

2: Category membership should be unambiguous

McClosky & Glucksberg (1978): Asked subjects to judge category membership of several everyday categories.

<table>
<thead>
<tr>
<th>Item</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A shelf</td>
<td>76%</td>
</tr>
<tr>
<td>A rug</td>
<td>52%</td>
</tr>
<tr>
<td>A lampshade</td>
<td>63%</td>
</tr>
<tr>
<td>Bookends</td>
<td>57%</td>
</tr>
<tr>
<td>Candlestick</td>
<td>28%</td>
</tr>
</tbody>
</table>

Many borderline cases
One prediction

2: Category membership should be unambiguous

McClosky & Glucksberg (1978): Asked subjects to judge category membership of several everyday categories.

<table>
<thead>
<tr>
<th>Item</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A shelf</td>
<td>76%</td>
</tr>
<tr>
<td>A rug</td>
<td>52%</td>
</tr>
<tr>
<td>A lampshade</td>
<td>63%</td>
</tr>
<tr>
<td>Bookends</td>
<td>57%</td>
</tr>
<tr>
<td>Candlestick</td>
<td>28%</td>
</tr>
</tbody>
</table>

Many borderline cases came back 1 week later... changed their minds about 22% of borderline cases!
One prediction

3: Ungraded category membership. All members are equally good

Rosch (1973): Asked people to rate “how good” different items are as an example of a category (1-7 scale)

<table>
<thead>
<tr>
<th>Item</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robin</td>
<td>1.4</td>
</tr>
<tr>
<td>Eagle</td>
<td>1.8</td>
</tr>
<tr>
<td>Wren</td>
<td>2.4</td>
</tr>
<tr>
<td>Chicken</td>
<td>2.8</td>
</tr>
<tr>
<td>Ostrich</td>
<td>3.2</td>
</tr>
</tbody>
</table>

Members vary in how good they are!
One prediction

3: Ungraded category membership. All members are equally good

What makes something typical? Rosch & Mervis (1975) investigated what makes an item typical by having subject list features of instance of many categories, other people rated typicality.

Hardly any examples of features that were in all category members!
One prediction

3: Ungraded category membership. All members are equally good

Typical features appear in many category members. # of typical features determines the typicality of a category member.

<table>
<thead>
<tr>
<th>Properties</th>
<th>Robin</th>
<th>Cardinal</th>
<th>Eagle</th>
<th>Penguin</th>
<th>Bat</th>
<th>Feature Score (a.k.a. “Weight”)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Has wings</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>5</td>
</tr>
<tr>
<td>Flies</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>4</td>
</tr>
<tr>
<td>Has feathers</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>4</td>
</tr>
<tr>
<td>Sings</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>2</td>
</tr>
<tr>
<td>Builds nests in trees</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>3</td>
</tr>
<tr>
<td>Eats worms/insects</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>3</td>
</tr>
<tr>
<td><strong>Family Resemblance Score</strong></td>
<td><strong>5+4+4+2+3+3=</strong></td>
<td><strong>5+4+4+2+3+3=</strong></td>
<td><strong>5+4+4+3=</strong></td>
<td><strong>5+4=</strong></td>
<td><strong>5+4+3=</strong></td>
<td><strong>21</strong></td>
</tr>
</tbody>
</table>
The Classical View

Probabilistic Approaches

- Category membership is a matter of degree. There can be better or worse members of a category!

- Two main theories - prototype models, exemplar models
According to **prototype theory**, the mental representation of a category consists of a prototype or central tendency of the examples.

Learning is about abstracting this schema or prototype across all the examples you have seen so far.
According to **prototype theory**, the mental representation of a category consists of a prototype or central tendency of the examples.

Learning is about abstracting this schema or prototype across all the examples you have seen so far.
How Most Models of Categorization Work
(Psychological Similarity Spaces)
How Most Models of Categorization Work (Prototype Theory)

Two key effects: prototype enhancement and borderline cases/graded structure
How Most Models of Categorization Work
(Prototype Theory)

- X is a bird.
  - Because it is closer to the bird prototype than to the insect prototype.

- Y is an insect.
  - Because it is closer to the insect prototype than to the bird prototype.
How Most Models of Categorization Work (Prototype Theory)

Penguins and ostriches are atypical because they are farther away from the bird prototype than robins and sparrows.
People remember many of the bird they have actually seen. People are influenced even in categorization contexts by the specific examples they’ve seen.

Brooks & Allen (1991)

- Asked subjects to discriminate between two animals (Diggers and Builders)
- Two kinds of animals could be distinguished in two different ways ("two-out-of-three" rule or based on animals environments context)
Classification rule:

2 of (long legs, angular body, spots) => Builder
2 of (short legs, curved body, not spots) => Digger

Example of Builder

Example of Digger
Environmental context

Forest scene => Builder
Arctic scene => Digger
Brooks and Allen (1991)

- Two conditions: memory group, or rule group

- **Memory group** - not given the rule (told would have to guess), learn incrementally based on feedback which is digger or builder

- **Rule group** - just told the rule from the outset

- At test, classified new items as Diggers or Builders

E.g., *Builder* in *Builder* context.

E.g., *Builder* in *Digger* context.
Brooks and Allen (1991)

Percent Errors (Non-Rule Answers)

- Memory
- Rule

- Rule/Environment Match
- Rule/Environment Mismatch
Brooks and Allen (1991)

Percent Errors (Non-Rule Answers)

Memory

Rule

Group

Rule/Environment Match

Rule/Environment Mismatch
Brooks and Allen (1991)

Response Time (ms)

- Rule/Environment Match
- Rule/Environment Mismatch
Even when an individual in the rule group responds correctly, they take longer if the environment is “telling them” to respond differently!

Even when subjects new the correct “featural rule” their classification decisions were affected by context.

Exemplar-based interference - even when a rule is known and easy to articulate, past examples can override application of the rule
Exemplar Theory

Birds
You’ve Seen

Bird?
Exemplar Theory

X is a bird because it is similar to many other birds.
Y is an insect because it is similar to many other insects.
<table>
<thead>
<tr>
<th>Empirical Effect</th>
<th>Classical View</th>
<th>Prototype Model</th>
<th>Exemplar Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>No defining features</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Borderline cases</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Graded typicality</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Prototype effect</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Exemplar effects</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>
Exemplar Theory

A is equally close to all birds and all insects
Exemplar Theory

Ostriches are not close to most other birds.
Exemplar Theory

Prototype is very similar to many birds.
What is the consequence?
Prototype is very similar to many birds.
What does it mean?
How does it work?
Similarity and Exemplar Models

- How is similarity to the stored examples computed?

- Medin & Schaffer (1978) proposed the context model of classification
  - A model of similarity for binary dimensions
  - A simple model of evidence accrual
  - A simple model of decision making
Similarity and Exemplar Models

- Each dimension has an associated importance or weight
  - An $s$ parameter (0-1) which controls importance
- When comparing two items, compute a match score, $m$, on each dimension
  - $m_i = 1$ if values on dimension $i$ match
  - $m_i = s_i$ if values on that dimension mismatch
- Overall similarity is the product of the $m$ values
Similarity and Exemplar Models
Evidence Accrual

- Similarity of item $S_i$ to a category $C_j$ is the sum of its similarities to the category’s exemplars

\[ \text{sim}(C_j, S_i) = \sum_k \text{sim}(S_k, S_i) \]

Decision Making

- The probability of classifying $S_i$ as a $C_j$ is the ratio of its evidence relative to other categories

\[ p(C_j | S_i) = \frac{\text{sum}(C_j, S_i)}{\sum_k \text{sim}(C_k, S_i)} \]
Application to Animal Learning Paradigms

- Pearce (1987): Many learning results arise from the similarity of cue-combinations at training versus test
  - Overshadowing
  - External inhibition
  - Overexpectation
  - Blocking
  - etc...

- Response drops off (quickly) if any cues or contextual elements are missing ("context model"). Compare to Gershman, Blei, & Niv latent cause model!!
The Generalized Context Model
Nosofsky (1984; 1986)

- The generalized context model (GCM)
  - Application of the context model to continuous dimensions.
  - Unification of Luce’s work on choice behavior and Shepard’s work on stimulus generalization
  - Similarity is a function of the distance between two objects in psychological space (Shepard!!).

\[
d_{ij} = c \left( \sum_{k=1}^{N} w_k |x_{ik} - x_{jk}|^r \right)^{1/r}
\]

\[
d_{ij} = \left( \sum_{k=1}^{K} |x_{ik} - x_{jk}|^r \right)^{1/r}
\]
The Generalized Context Model
Nosofsky (1984; 1986)

- Actual similarity of two objects is a function of their distance:

\[ \eta_{ij} = e^{-d_{ij}} \]

- Response rule

\[ p(R_j | S_i) = \frac{b_j \sum_{j \in C_j} n_{ij}}{\sum_{k=1}^{m} (b_k \sum_{j \in C_k} n_{ik})} \]
The Generalized Context Model

Nosofsky (1984; 1986)
The Generalized Context Model
Nosofsky (1984; 1986)

The $c$ parameter in the model matches the exponential generalization gradient in Shepard’s work.
The Generalized Context Model
Nosofsky (1984; 1986)
Selective Attention
Nosofsky (1986)

**Stimuli**
- Size and Angle
- Both size and angle varied along four levels

- Subject first made identification judgments (yielding a confusion matrix)
- MDS techniques from Shepard used to provide stimulus representation for each subject in appropriate “psychological space”
Selective Attention
Nosofsky (1986)

GCM model fit to each subject to estimate best fit values of w’s, c, etc...

```
DIMENSIONAL
2  2
2  2
1  1
1  1

CRISS-CROSS
1  2
1  2
2  1
2  1

INTERIOR-EXTERIOR
2  1
1  1
2  1

DIAGONAL
1  2
1  2
1  2
```

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0.58</th>
<th>0.64</th>
<th>0.60</th>
</tr>
</thead>
<tbody>
<tr>
<td>w1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>w2</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>0.09</th>
<th>0.66</th>
<th>0.66</th>
<th>0.56</th>
</tr>
</thead>
<tbody>
<tr>
<td>w1</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>w2</td>
<td>0.91</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>0.34</th>
<th>0.34</th>
<th>0.44</th>
</tr>
</thead>
<tbody>
<tr>
<td>w1</td>
<td>0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w2</td>
<td>0.91</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Selective Attention
Shepard, Hovland, Jenkins (1961)
## Selective Attention
Shepard, Hovland, Jenkins (1961)

<table>
<thead>
<tr>
<th>Problem Type</th>
<th>Within-Category Similarity</th>
<th>Predicted Learning Difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Very high</td>
<td>Very easy</td>
</tr>
<tr>
<td>III, IV, V</td>
<td>Moderately high</td>
<td>Moderately easy</td>
</tr>
<tr>
<td>II</td>
<td>Moderately low</td>
<td>Moderately hard</td>
</tr>
<tr>
<td>VI</td>
<td>Very low</td>
<td>Very hard</td>
</tr>
</tbody>
</table>

but actually
Type I < II < (III, IV, V) < IV
Selective Attention
Shepard, Hovland, Jenkins (1961)

<table>
<thead>
<tr>
<th>Problem Type</th>
<th>Within-Category Similarity</th>
<th>Predicted Learning Difficulty</th>
<th># of Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Very high</td>
<td>Very easy</td>
<td>1</td>
</tr>
<tr>
<td>III, IV, V</td>
<td>Moderately high</td>
<td>Moderately easy</td>
<td>3</td>
</tr>
<tr>
<td>II</td>
<td>Moderately low</td>
<td>Moderately hard</td>
<td>2</td>
</tr>
<tr>
<td>VI</td>
<td>Very low</td>
<td>Very hard.</td>
<td>3</td>
</tr>
</tbody>
</table>

For Type II, attention resources can be concentrated on two dimensions, making it easier.

Types III, IV, and V require attention to be diluted over three dimensions.
Eye Movement Data
Rehder & Hoffman (2005)

# of Dimensions Fixated

Block

Type VI
Type IV
Type II
Type I
Category Learning and Monkeys

SHJ’s Predictions for “associative” learner (based on confusion matrices)

Smith, et al.’s predictions using exemplar model’s multiplicative similarity function
Figure 3. Performance by humans in the Shepard et al. (1961) category tasks in three existing studies and the present study. A: Results from the Shepard et al. study. B: Results from the Nosofsky et al. (1994) study. C: Results from the Love (2002) study. D: Results from the present study.
**Category Learning and Monkeys**

*Figure 4.* A: Humans’ (H) forward learning curve in Type I tasks. Their percentage correct is shown at each eight-trial block. B: Humans’ backward learning curve in Type I tasks. Their percentage correct is shown for eight-trial blocks forward and backward from the start of their first run of three consecutive perfect blocks. C: Humans’ forward learning curve in Type II tasks. D: Humans’ backward learning curve in Type II tasks.
Category Learning and Monkeys

Figure 8. A: Monkeys’ (M) forward learning curve in Type I tasks. Their percentage of correct responses is shown at each of eighty-three 24-trial blocks. B: Monkeys’ backward learning curve in Type I tasks. Their percentage of correct responses is shown for 24-trial blocks forward and backward from their first perfect 24-trial block. C: Monkeys’ forward learning curve in Type II tasks. D: Monkeys’ backward learning curve in Type II tasks.
Category Learning and Monkeys

“It is noteworthy that all the monkeys, even across the six rotations of the six task types, instantiated so well Shepard, et al.’s (1961) idea of a cognitive system that learns categories associatively through cue-conditioning processes or similarity-based generalization processes.” - Smith, et al.
Category Learning and Kids
Minda, J.P. and Church (2008)

- Compared children (at various ages) and adults on SHJ type task
- Tested Type I, II, III, IV
- III is nonlinearly separable (rule+ex) category
- Hypothesized deficits for Type II and possibly III at early ages due to lack of verbal competence (and continued development of the PFC into adolescence)
Category Learning and Kids
Minda, J.P. and Church (2008)

- Adults under verbal interference look like kids
Take home points

- Categorization is the study of how people learn and generalize from examples

- Early theories emphasized definitional rules, while this view was revised in favor of “probabilistic” approaches in the 70s and 80s

- Two major models are prototype and exemplar models. Exemplar models are powerful contenders that subsume many aspects of the prototype account. In addition, exemplar models were developed to bridge Shepard’s work on stimulus generalization into a theory of how people generalize from multiple examples

- A key concept is selective (and adaptive) weighting of stimulus dimensions during learning, and the ability to do so may vary across species and individuals.
2 Categorization
Multiple systems framework
Overview

• **Central Idea:** There are multiple representation schemes that people apply when learning and these are flexibly adjusted
  
  • The (new) case for abstraction (Smith & Minda, Knowlton & Squire)
  
  • General evidence in favor of “multiple” representations strategies*
  
  • “Hybrid” theories that combine exemplar, prototype, and rule-like processes/representations
  
  • Theories where these distinctions may be an “emergent” property of a more general (and flexible) representational strategy

* Note: purposefully avoiding use of “system” at this point. In next class we’ll talk about multiple-**systems** views of categorization from a neuro-biological perspective
The dominance of the exemplar view

- Examples included:

  - No cost to learning non-linearly separable categories (Medin & Schwanenflugel, 1982; Shepard, Hovland, Jenkins, 1961)

  - Patterns of generalization appear to favor the exemplar model in a variety of tasks (Medin & Schaffer, etc...)

  - Allen and Brooks (1991)-like effects were specific (exemplar) information intrudes even when you try to use a rule

  - The simple prototype model’s inability to account for category variability.
All may not be lost for the prototype model

- In recent years, there has been a resurgence of interest in the prototype account.

- In my view, there are two main reasons for this:
  
  - Knowlton & Squire’s work with category learning in amensics. The patients have impaired episodic memory, yet appear to learn prototype-centered categories just fine. i.e., no exemplars can “get into” memory, yet you can still categorize.

  - The second is a general dissatisfaction with the exemplar account for many of the practical reasons suggested in Monday’s class (what constitutes an exemplar? there must be some abstraction).
One argument is that the success of the exemplar model has been that it has been tested on highly trained participants who have a lot of experience with a small number of category members.

However, at various stages of learning the advantage of the prototype model might give way to what appears to be a prototype advantage.

Smith & Minda tested this in a set of experiments that tested both linearly separable (LS) and non-linearly separable (NLS) at various point during training.
Smith & Minda (1998) - Prototypes in the mist
Overall, Smith and Minda found that there was a small, but reliable, advantage for the prototype model early in learning, with a strong advantage for the exemplar model later.
Proposed a mixture model, with three components:

- A guessing parameter
- A standard additive prototype model
- A “restricted” memory store where examples are store in memory but there is no exemplar gradient/generalization around them (i.e., the exemplars only contributed a direct match, and there was no generalization around them)

The relative contribution controlled by two additional free parameters such that \( p(\text{guess}) + p(\text{prototype}) + p(\text{exemplar}) = 1 \)
Smith & Minda (1998) - Prototypes in the mist

The mixture model (while having a similar number of parameters as the standard exemplar model) has an overall advantage at accounting for the data. In addition, the parameters of the mixture model confirm the impression of a shift between the exemplar and prototype.
Smith & Minda (1998)

- There is clearly some value for considering strategies and representations that change in time rather than static performance at the end of training (i.e., “watching movies” of learning)

- Overall, the results appear somewhat consistent with Logan’s instance theory of automaticity (that over time performance begins to depend more on automatic processes drawing from memory)

- Combined prototype and exemplar model highlights importance of abstraction and combined strategies (Busemeyer, Dewer, and Medin, 1984)
What have been the conclusions of the “prototype revival?”

- Some of the successes of the exemplar approach may be due to the specific tasks (which are generally very difficult and non-intuitive... see Murphy’s *Big Book of Concepts* for a similar discussion)

- The field needs to take seriously issues of model complexity: Smith & Minda’s work highlighted how flexible the exemplar model is (it stores everything!!)

- The support for the exemplar approach is not absolute. While it clearly has its merits and have received considerable support, there is still an important role for abstraction when learning.