An Introduction to Learning

Lecture 10/13

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

- More complex forms of generalization: **Categorization**
  - Abstraction and multiple representations
  - Rule-governed learning
  - Cognitive neuroscience approaches to category learning
- Learning by analogy
- Causal Learning
The Exemplar and Prototype Debate

• Exemplar and Prototype Models are titans in the field of cognitive psychology.

• These models are important beyond just the categorization literature because the issues of memory representation and stimulus generalization come up in many areas
  
  • “Prototypical” or “Average” faces are rated as more attractive (Langlois & Roggman, 1990)
  
  • The E and P models share deep similarities to Bayesian template-matching models in visual perception (Gold, Cohen, Shiffrin, 2006)
  
  • In Memory literature: MINERVA (Hintzman, 1988)
  
  • Speech Perception: Fuzzy Logic Model is a “prototype”-like model (Massaro, 1989); The prototype-magent effect (Kuhl, 1991), “Rich Phonology” (Port, 2007)

• However, is this really all there is?
3 Categorization
Rules and Associations in Learning
Multiple Strategies in Categorization

• Once again, the story is a shift to exemplar-like processes with extensive training.

• However, careful considering of the dynamics of learning, reveal a number of strategies and representations are in play.

• Abstract, effortful, (potentially explicit) strategies give way to effortless, automatic, memory-based processes later in training.
Sloman (1996)

Associative vs. Rule-Based Systems

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Associative system</th>
<th>Rule-based system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principles of operation</td>
<td>Similarity and contiguity</td>
<td>Symbol manipulation</td>
</tr>
<tr>
<td>Source of knowledge</td>
<td>Personal experience</td>
<td>Language, culture, and formal systems</td>
</tr>
<tr>
<td>Nature of representation</td>
<td>Concrete and generic concepts, images, stereotypes, and feature sets</td>
<td>Concrete, generic, and abstract concepts; abstracted features; compositional symbols</td>
</tr>
<tr>
<td>Basic units</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relations</td>
<td>(a) Associations</td>
<td>(a) Causal, logical, and hierarchical</td>
</tr>
<tr>
<td></td>
<td>(b) Soft constraints</td>
<td>(b) Hard constraints</td>
</tr>
<tr>
<td>Nature of processing</td>
<td>(a) Reproductive but capable of similarity-based generalization</td>
<td>(a) Productive and systematic</td>
</tr>
<tr>
<td></td>
<td>(b) Overall feature computation and constraint satisfaction</td>
<td>(b) Abstraction of relevant features</td>
</tr>
<tr>
<td></td>
<td>(c) Automatic</td>
<td>(c) Strategic</td>
</tr>
<tr>
<td>Illustrative cognitive functions</td>
<td>Intuition, Fantasy, Creativity, Imagination, Visual recognition, Associative memory</td>
<td>Deliberation, Explanation, Formal analysis, Verification, Ascription of purpose, Strategic memory</td>
</tr>
</tbody>
</table>
William James

- **Associationism** elaborate the Aristotelian view of associations to include networks of interrelated concepts (precursor to connectionist "Rummelhart" networks)

- Initiated study of habits and automatic associations

- Aimed to eventually detail the nature of his networks directly in the brain
Hypothesis Testing in the Rat: Tolman

(Fig. 14
(From I. Krechevsky (Now D. Krech), The genesis of "hypotheses" in rats. *U.S. Calif. Publ. Psychol.,* 1932, 6, No. 4, p. 46.)
Rips (1989)

- A central tenant of both the prototype and exemplar models is that similarity between representations stored in memory and to-be-categorized stimuli are the critical determinants of a response.

- Lance Rips found an simple way to dissociate similarity and categorization:

More Similar

More Likely
Rips (1989)

- The idea is that the rule *if diameter is not almost exactly .25 inches then it is not a quarter* overrides this “similarity” response
Sloman (1996)

- **Criterion S**: cases where one, presumably rule or logic based response, is held simultaneously with another associative response but are in conflict.

- An example: Tversky and Kahneman (1983)’s Linda the Bank Teller:

  Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

  Rank order the following according to the statement’s probability:

  -- **Linda is a bank teller.**
  -- **Linda is a bank teller and is active in the feminist movement.**

  The “logical”/”rule-based” response overwhelmed by the “representative heuristic” (a similarity based response).
Rules and the classical view of concepts

• The Allen & Brooks (1991) study described last week is consistent with this as well (even subjects in the rule condition were influence by exemplar similarity in both their judgements and response times)

• As the last lecture showed, the classical view of concepts was sorely damaged if not destroyed by findings of graded category structure.

• However, one way to view the general point in Sloman’s article is to appreciate the role of logical “reasoning”-based processes in categorization.

• Again, multiple ways of learning are engaged in any situation possibly in parallel
Bower and Trabasso (1963)

• Tested between rule and association based accounts of concept learning

• One of the key observations about rules is the idea of “one trial learning”. All learning takes place when the correct rule is under consideration and is tested.

• As a result, rule-based theory predicts that changing the rule during learning does not hurt when the correct rule hasn’t been discovered yet.

• In contrast, associationist accounts (like ALCOVE) gradually accumulate knowledge about the correct response and right dimensions to attend to. Thus, learning before the switch interferes with learning.
Bower and Trabasso (1963)

- Categories were defined by simple single dimensional rules
- Every time the participant made 2 mistakes in a row, the rule that defined the category switched.
- For example, first category A might be red things, category B, blue things. After two mistakes in a row, the whole category changes so that A is small thing, B is big things.
- RESULT: Little or no cost to learning in this switching environment
The Dangers in AVERAGING

- Group level data supports an ALCOVE-like account of gradual attention learning.
- But, is the appearance of gradual learning just an artifact of averaging over subjects?
- Let’s look at one Type I learner, subject 35 in Rehder & Hoffman (2005)
# of Dimensions
Fixated

Proportion Time
(Relevant Dimension)

Relative Priority
(Relevant Dimension)

Errors
Learning is Sudden, Not Gradual

- Subject 35’s results not consistent with Alcove’s gradual learning account!
  - Before trial 22:
    - All three dimensions (usually) fixated.
    - No preference for looking at relevant dimension...
      - …more often than other dimensions.
      - …earlier than other dimensions.
    - Performance at chance (many errors).
  - After trial 22:
    - Only relevant dimension fixated.
    - No errors.
Example Type I subject

Number of Dimensions Fixated
Example Type I subject

*Number of Dimensions Fixated*

![Graph showing number of dimensions fixated over trials](image-url)
Example Type I subject
Example Type I subject
Example Type I subject
Example Type I subject
Example Type I subject

\[\text{# of Dimensions} \]

Trial

-16
-8
0
8
16
Example Type I subject

# of Dimensions

Error Rate

Trial
Example Type I subject

![Graph showing the relationship between trial and dimensions, error rate, proportion time, and relative priority.](image)

- **Error Rate**
  - Red line
- **Proportion Time**
  - Green line
- **Relative Priority**
  - Blue line
- **Number of Dimensions**
  - Black line

<table>
<thead>
<tr>
<th>Trial</th>
<th># of Dimensions</th>
<th>Error Rate</th>
<th>Proportion Time</th>
<th>Relative Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>-16</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>-8</td>
<td>2</td>
<td>2</td>
<td>0.17</td>
<td>0.83</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0.33</td>
<td>0.67</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td>0.67</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** The graph illustrates the changes in the number of dimensions, error rate, proportion time, and relative priority across different trial stages.
Rules and Generalization

• A couple comments about rules
  • Fast (as in fast/efficient learning, usually slow response times)
  • Resistant to inference from previous learning
  • Discrete - all or none learning
  • *Abstract GENERALIZATION*

• Your experience on the New York Subway system has certain properties (get ticket, refill when empty, insert into turnstile and walk through) that apply to highly dissimilar situations but that follow the same pattern (the DC subway, London Underground).

• Gary Marcus’ word learning studies with infants: ABA -> CDC
ATRIUM (Ericsson & Kruschke, 1999)
ATRIUM (Ericsson & Kruschke, 1999)
ATRIUM (Ericsson & Kruschke, 1999)
ATRIUM (Ericsson & Kruschke, 1999)
ATRIUM (Ericsson & Kruschke, 1999)
ATRIUM (Ericsson & Kruschke, 1999)
ATRIUM (Ericsson & Kruschke, 1999)
ATRIUM (Ericsson & Kruschke, 1999)
ATRIUM (Ericsson & Kruschke, 1999)

- ATRIUM combines an exemplar and rule based module into a single competitive network
- Each module learns independently and a gating mechanism decides how responses from these two systems should be integrated.
- Critically, the gating system can learn to apply stimulus-specific response strategies (for example, tagging a single exception item as it should be handled by the ALCOVE component while other stimuli should be handled by the rule).
General Conclusions

• People use a variety of representational strategies to learn categories

• Contemporary formal models have taken the approach of assuming multiple representation systems that learn independently and there is some competition between them for behavioral control

• **THIS IS AN OVERALL MORE COMPLEX STORY BUT GOOD:** This is a more accurate picture of what is going on (and is supported by some of the neuropsychological work as well that we will discuss next time)
Rational Arguments in Favor of this Distinction

- As in Marr’s paper, different representations make different kinds of processes more or less accessible: Roman Numerals vs. Base-10 Digits

- The distinction between rule-based and similarity-based/exemplar strategies have a similar flavor:
  - Rules are fast, efficient, and abstract but potentially OVER-regularize the true situation
  - Associations are slow, but in some sense make use of all the information experienced (i.e., statistically optimal)
3 Categorization
Cognitive Neuroscience of Category Learning
Multiple Systems in Categorization

- The bird’s eye view is that:
  - Distinct neural pathways may be tapped for different forms of learning
  - Different patient population have a specific (and somewhat predictable) pattern of deficits
  - Critically, the pattern of deficits varies as a function of the information structure of the category itself and the study conditions
Smith, Palatino, Jonides (1998)

- “Alternative Strategies of Categorization”: In contrast to unitary view, people use multiple strategies, often at the same time.
- Built off Allen & Brooks (1991) “Builders”/”Diggers” study
Smith, et al. (1998)
PET results

- Subtractions compared to a control phase where just made responses (to subtract non-specific motor components)
- Large number of areas unique to rule based categorization including: parietal lobes, dorsolateral PFC, supplementary motor cortex, cerebellum, right thalamus
- Common areas included visual cortex, cerebellum
- Only memory: visual cortex/cerebellum
Smith, et al. (1998) PET results

- **Parietal lobes** used in Rule condition are associated with controlled, effortful attention (i.e., selective attention)
- **Right Dorsolateral PFC** -> implicated on rule-based task switching like WCST
- **Supplementary Motor Cortex** = speech, motor preparation
Smith, et al. (1998)
PET results

- **Visual cortex** activated in all groups consistent with perceptual categorization, exemplar retrieval?
Executive Function and the Wisconsin Card Sorting Task (Milner, 1964)

- Sort a set of cards on a basis of a single dimension.

- Correct dimension switches after a certain number of correct responses (compare to the Bower and Trabasso study... here it is switch when correct, not when make mistakes)

- Generally, patients with impaired frontal lobe function show deficits in the task which presumably requires representation and maintenance of a rule in working memory.
General Conclusion

• The strategy used to learn the category not only influenced the pattern of behavioral results (consistent with Allen and Brooks original study) but also influenced the neural systems engaged.

• Overall a strong effect of the Rule condition (network of regions involved in executive function, working memory, and attention)

• Pure effect of memory task was less pronounced as was the overlap

• Additional support for models like RULEX or ATRIUM that posit separate rule and similarity-based (i.e., exemplar) processing
Perhaps the most complete and well-tested multiple-systems theory of categorization that has taken the connections to the biological substrates of behavior seriously

(At least) two neural systems supporting category learning: An implicit/procedural system and an explicit/verbally mediated system
COVIS - The Verbal System


- Rule based tasks: exemplified by the categorization task shown
- Easy to verbalize category boundary
- (As in Sloman) system assumed to relied on verbal processes, logical reasoning, strategic processing, and semantic memory
COVIS - The Procedural System


- Some debate about if the “implicit” system is supported by an exemplar-like system (Erickson & Kruschke, ATRIUM) or by a procedural learning system
- (COVIS takes the latter view)
- Procedural category learning is associative with an inability to verbally express the “rule” for the category
- Based on a decision-bound system that gradually learns to associate particular regions of stimulus space with a particular response.
- Tied to the perceptual-motor system
Figure 2. A schematic depicting the neuropsychological underpinnings of COVIS (competition between verbal and implicit systems). The dotted lines denote dopamine projections. VTA = ventral tegmental area; SN = substantia nigra; NAC = nucleus accumbens; IT = inferotemporal cortex.
Selection
Switching
Working
Memory
Loops
Behavioral Dissociation in support of COVIS


- Procedural system relies on feedback, thus is impaired in unsupervised learning contexts
- Observational training and delayed feedback negatively impact the procedural system
- Changes in motor response will impact information integration more than rule-based learning
Feedback and Procedural Learning

Observational Learning

Observational Learning

Delayed Feedback

Delayed Feedback

Changing motor response

Changing motor response

Changing motor response

Changing motor response

Behavioral Dissociation in support of COVIS


- Dual task/Distraction impact rule learning more than information integration
- Reduced time to process feedback impacts rule learning more the information integration
- The procedural system is insensitive to the number of categories/ complexity of the category rule (relative to the rule system)
Dual Task


Fig. 15. Basic design for the single and dual task training procedures.
Reduced Time to Process Feedback

Cluster vs. Categories

Recent fMRI Evidence


- Compared II vs. RB categorization (i.e., diagonal versus vertical boundaries)
- Rule-based category learning associated with increased activity in MTL, Right Caudate, Anterior Cingulate, medial frontal gyrus
- II-based category learning associated with increased activities in caudate/striatum

Correct-Incorrect Trials
Criticisms of COVIS

• What are the limits of the verbal system? What says that one rule is harder than the other to learn? In COVIS this is left blank, and in the modeling the complexity of one rule compared to another is set by the experimenter (see Feldman, 2000 for a possible answer to which rules are easier or harder to learn).

• Can these dissociations be explained other ways? Nosofsky and colleagues have argued that the difference really is one of cognitive complexity. It is hard to control the relative difficulty of the tasks (Nosofsky, Stanton, & Zaki, 2005; Stanton & Nosofsky, 2007, etc...).

• Competition? The competition mechanisms in COVIS is relatively weak. The “model” is never really applied to the data. Instead, two different decision bound models are used: a rule-based one which is limited in the types of decision bounds it can employ, and a procedural one which includes the possibility of “diagonalized” boundaries. Winner is assumed to be “best” model.
Continuum of PFC-MTL Function

Various groups are ordered by their ability to individuate events.
Is Categorization Intact in Amnesia?

- Nosofsky & Zaki (1998) present an analysis with a single-system exemplar model the accounts of the same pattern of results.

- In exemplar model, similarity is based on exponential function of distance in metric space  \( s_{ij} = \exp(-c \cdot d_{ij}) \)

- As \( c \) is higher, the exemplar traces in memory are encoded more “sharply”

- If one assumes that amnesic individual acquire the exemplars normally, but have a lower setting of memory sensitivity (\( c \)) then they will be ok at categorization (which benefits from “blurred” representations) while impaired at recognition (which require fine-tuned representations)
Does Categorization Require Learning?

• Palmeri & Flannery (2001) presented an innovative study on this.

• Normal subject told to look at a random pictures of everyday objects.

• Then told they were subliminally presented dot patterns (like the K&S stimuli). However they weren’t really presented.

• At test make judgements of items that come from a single category and random patterns.

• Result: The do just as well at the task as those that studied the actual patterns!!

• The effect is learning at test (a large number of low distortions and prototypes in the testing phase support learning in the absence of training!)
Incidental Condition

- Incidental observers were shown 5 items (repeated twice) and were asked to identify the center dot in each pattern.
- The were not told that the patterns conformed to an underlying category.

Explicit Condition

- Shown the same items.
- Told all patterns came from same category and they should try to learn it while watching the patterns.
Test Phase

- Both groups see identical test condition while scanning
- 36 patterns came from same underlying category
  - 4 Prototypes
  - 16 Low Distortions
  - 16 High Distortions
- 36 patterns non-categorical patterns
Test Phase

- Test items presented in blocks of 9 items which had mostly categorical or non-categorical item
- 7:2 mix
- Analysis was based on blocks of trials which were mostly categorical or most non-categorical
Why this is interesting

- Unlike a number of other studies (such as K&S or the manipulations) this study is a demonstration of a dissociation at test based only on the study instructions.

- Falls in line with k&s work by showing that category learning of dot-pattern stimuli invokes brain regions unaffected by MTL damage...

- Dovetails with theories concerning the role of MTL in mediating explicit memory processes.
The role of cognitive neuroscience

• In many areas of psychology, cognitive neuroscience approaches have been criticized for being “just so stories”

• However, in the area of categorization the search has been focused on one theoretical issue: are there multiple memory or procedure systems support categorization (and memory)

• In some ways this kind of data has the potential to go beyond some of the empirical studies and modeling debates.
Weather Prediction Task

- Probabilistic classification task: discourages strategy based on memorization/explicit memory

- Poldrack, et al. found “interactive systems” in learning

- MTL early, Basal Ganglia later
Summary of Patient Data

- Parkinson’s Patients
- Huntington’s Patients
- MTL Amnesics/Early AD
- Korsakoff’s Amnesics
Parkinson’s/Huntington’s

- A broad spectrum of neurological change involving dopinergic processes in basal ganglia

- However the locus in basal ganglia suggest that PD patients may be impaired a both RB and II task.


- But also associated with procedural learning problems:
  - PD impaired at serial reaction time (SRT) tasks that PD are universally impaired a procedural learning
  - Knowlton, Mangels, & Squire (1996) report PD patients impaired at probabilistic categorization task (non-verbal) (Weather Prediction Task)

- HD generally impaired at information-integration tasks (Filoteo, et al., 2001; Maddox & Filoteo, 2001) while performance in rule-based tasks may be normal

- However, Filoteo, Maddox, and Davis (2001) found a early deficit in single-dimensional rule-based tasks for HD patients
MTL Amnesics/Early AD

- Impaired memory population, but appear not impaired at categories involving prototype learning
- May be impaired at certain rule based tasks
- Show late-training deficits in probabilistic tasks like the weather prediction task relative to healthy normals (suggesting a pattern of deficit that
- Impaired generally at conjunctive task (for example MTL amnesics show impairment in sequence learning tasks that depend on higher order conditional relationships) (Curran, 1997; Schendan, et al., 2003)
- Early AD appears to impact MTL first, thus these patients demonstrate classic MTL amnesic symptoms. In fact, often used as stand-in for “pure” MTL amnesics since the later are harder to come across and test.
Korsakoff’s/PFC Patients

- Korsakoff’s is a different amnesic syndrome affecting frontal lobe (caused by alcoholism commonly)

- Patient’s show similar deficits on memory tasks, however the locus of the memory effect may be different.

- In particular PFC patients fail to show anticipatory SCR to aversive stimuli (Bechara, et al., IGT)

- In SUSTAIN, these patients would be modeled as having an inability to detect surprising stimuli in the environment.
Key Principals for the Semester

- Learning and memory are closely related and intertwined states of information processing
- Major insights about learning and memory have come from studies of the brain
- The concept of multiple memory systems unifies the study of learning and memory
- The underlying bases of learning and memory are the same in humans and animals
- Our theoretical approaches to studying learning are always closely tied to technological advances that are unfolding in general society (e.g., today - machine learning)
Next time

The ivory towers of learning: Analogical learning, causal learning, and how learning a language might be possible (part I)