Cognition: Methods and Models

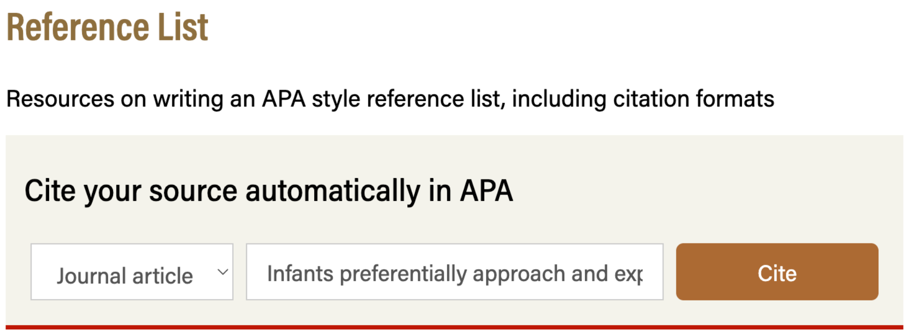
PSYC 2040

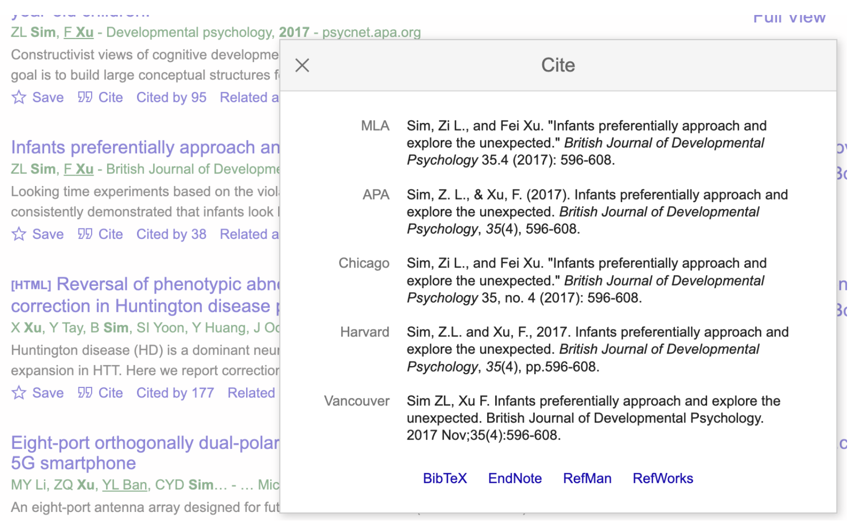
L9: Cognitive Models

Part 1

logistics: project milestone #4

- feedback returned to most groups
- use APA format to cite papers/references in first draft and final submission
  - Purdue OWL has a nice tool to obtain refs!
  - Google Scholar also has a straightforward tool
- focus on effective visualizations when making connections!
- milestone #5 (first draft) is due April 30!





recap: Apr 4/6, 2023

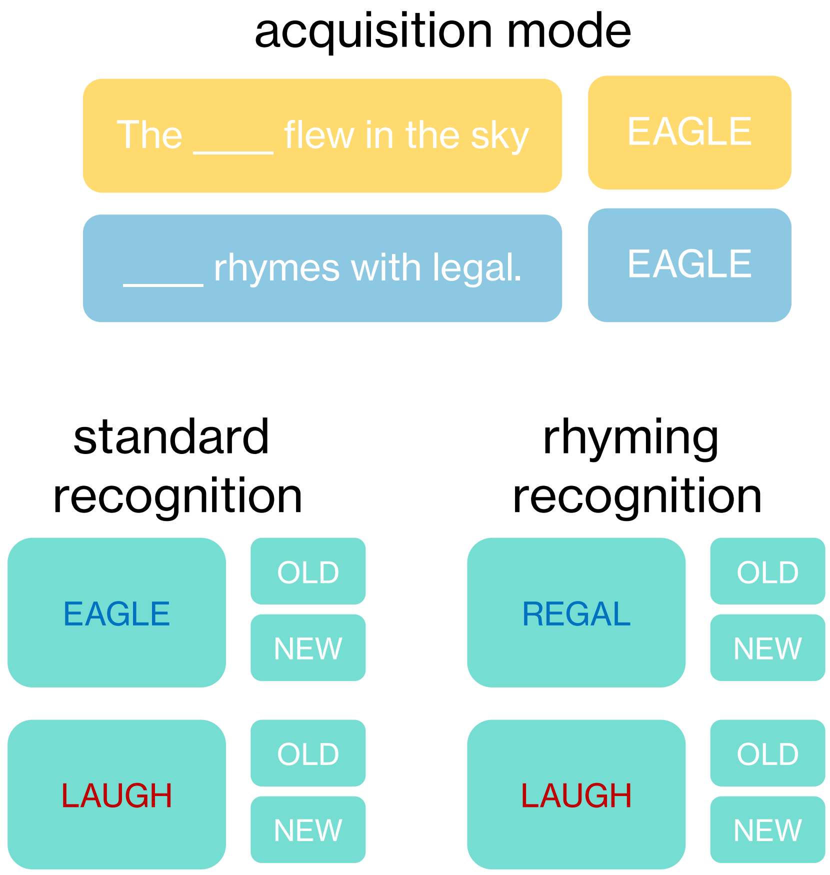
- what we covered:
  - memory tasks & phenomena
  - memory principles
- your to-dos were:
  - *finish:* L8 quiz/writing assignments
  - *submit:* project milestone #4 (outline)

today’s agenda

- first part:
  - TIP/TAP review
- second part:
  - cognitive models

TIP/TAP > levels of processing

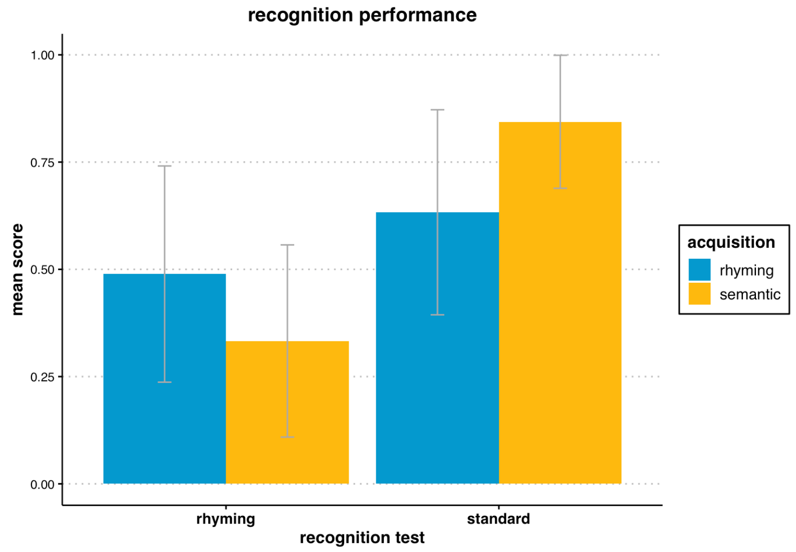
- claim: the tasks performed at encoding and retrieval take precedence over the nature of processing (shallow vs. deep)
- evidence: Morris, Bransford, and Franks (1977)
  - participants encoded words in a semantic or rhyming context
  - the test phase was either a standard recognition test or a rhyming-based recognition test



TIP/TAP > levels of processing

- claim: the tasks performed at encoding and retrieval take precedence over the nature of processing (shallow vs. deep)
- evidence: Morris, Bransford, and Franks (1977)
    - on standard test, recognition was higher for semantic vs. rhyme words
    - on rhyme test, recognition was higher for rhyme vs. semantic words

	Recognition test	
Acquisition mode	Standard	Rhyming
Semantic–Yes	.844 (.155)[a]	.333 (.224)
Rhyme–Yes	.633 (.239)	.489 (.252)



other principles

- task-appropriate processing/probe dependency: memory performance is influenced by the nature of the task and whether a particular cue is actively generated or provided
- resource demands: increasing the cognitive demands during encoding/retrieval can influence retention and performance

conceptual question #frequency

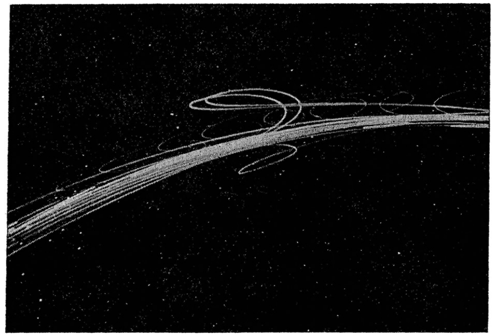
- something from our discussion of memory phenomena that interested me was the word frequency effect. It makes sense to me that low-frequency words are better recognized than high-frequency words, but I would expect this to be the same for a recall task. I know that studies have begun to explain this paradox, but why do manipulations of word frequency influence memory performance in different ways depending on the task?
  - likely has to do with task-appropriate processing, recall requires active generation and high-frequency words are easier to produce. recognition relies on familiarity and the smallest boost is useful for low-frequency words

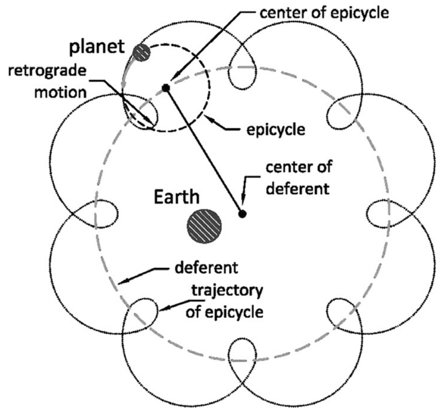
today’s agenda

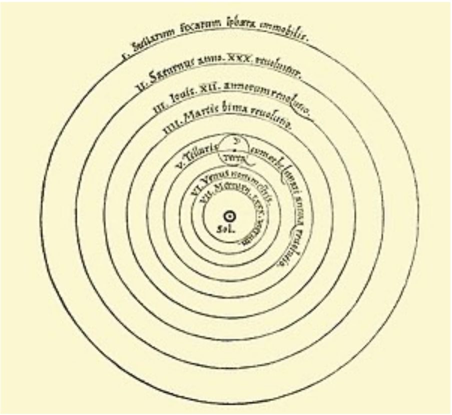
- first part:
  - TIP/TAP review
- second part:
  - cognitive models

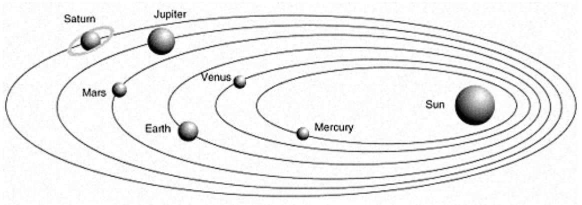
motivating models: planetary motion

- planets typically have curvilinear paths, but appear to have strange “loops”, referred to as retrograde motion
- explaining why this happens requires a model of an underlying process that generates this pattern
- models do not physically *exist,* they are “abstract explanatory devices” that people use to describe, predict, and explain *real data*
- several models may explain the data and scientists must select among different alternatives







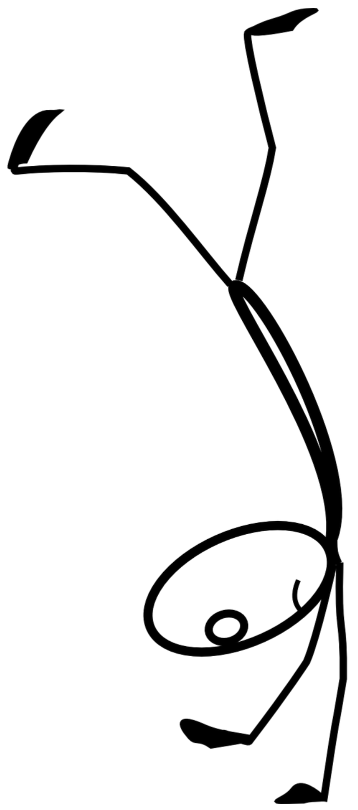


we use models all the time!

- any type of description of data can be considered a model
- averaging a set of numbers is a *model* of the data
  - means can be informative: examples?
  - means can be misleading: examples?
- the Rescorla-Wagner model of associative learning
- other examples?

theories of learning

- we know people get better over time at learning a new skill, but how exactly?
- the first time takes forever, the next few attempts lead to major improvements, and then improvements slow down
- two explanations/models:
  - power law: $\mathrm{RT} = N^{-\beta}$
  - exponential law: $\mathrm{RT} = e^{-\alpha N}$,

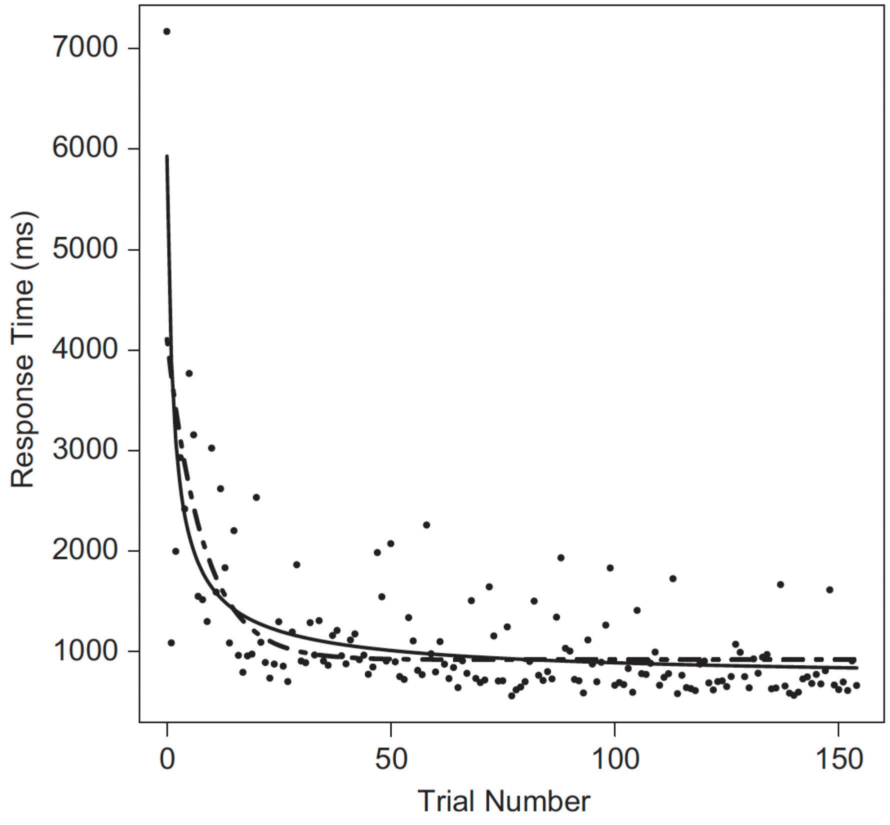


exercise: power vs. exponential

- go to power vs. exponential spreadsheet
- find your group's tab
- inspect the formulas in columns C and D and figure out which is the power function and which is the exponential
- select columns A, C, and D, and insert a chart that shows you the time taken to learn after N trials

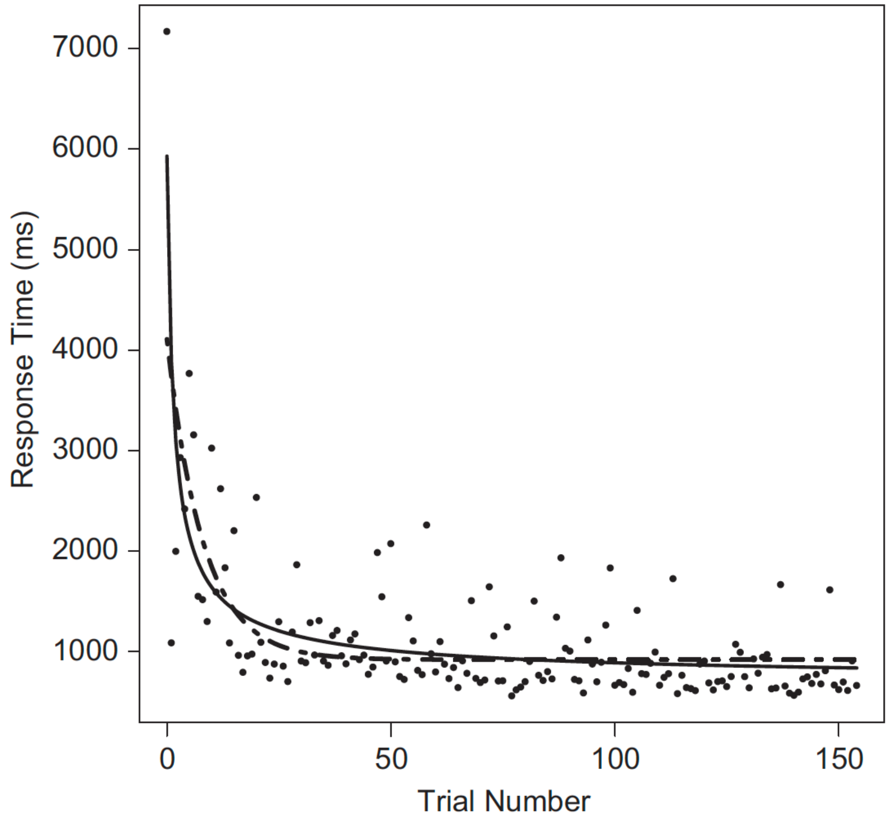
learning: why does it matter?

- the fit of both models is very similar so why does it matter which one is more accurate?
- the exponential form suggests that the relative learning rate remains constant, i.e., regardless of practice, your learning continues to be enhanced by a constant fraction
- the power law suggests that the relative learning rate is slowing down, i.e., as you practice more, you are actually learning less over time
- which model is correct has important practical implications: how much should you practice a new skill?



learning: why does it matter?

- Heathcote (2000) showed that the exponential function better fit the trial-level data
- learning curve is better explained by the exponential function
  - the more you learn, the more you retain
- implications for forgetting
  - learning is not the same as forgetting: forgetting follows a function closer to power law (Wixted, 2004), so you lose more initially and lose lesser over time



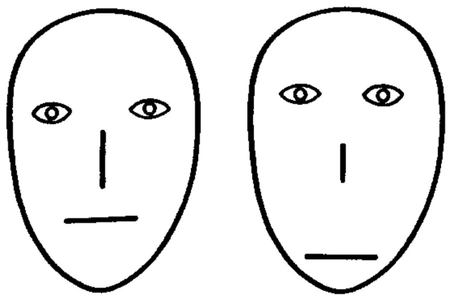
descriptive vs. process models

- descriptive models emphasize describing the data, typically through some type of mathematical formulation and/or statistic
  - examples include the exponential/power laws, means, proportions, etc.
- process models emphasize the underlying mechanism that directly produces the data, and can often generate predictions
  - examples include the Rescorla-Wagner model

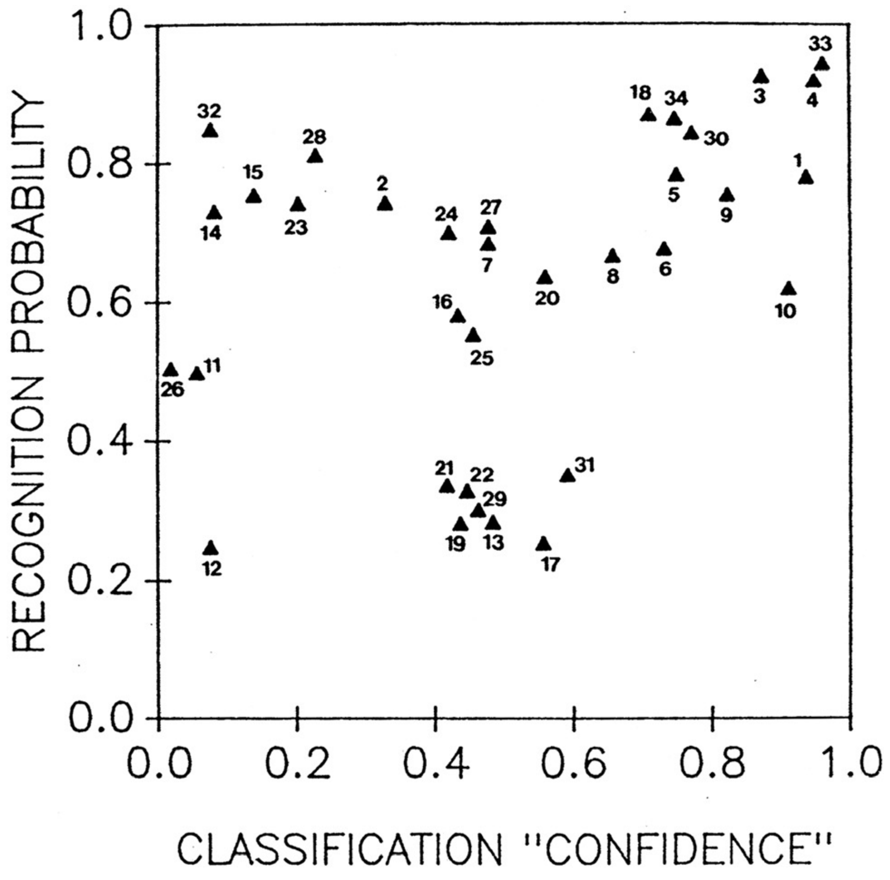
activity: cartoon face experiment

- go to https://lhclkdgwwg.cognition.run
- do the experiment (use laptop!)
- come back and discuss
    - how did you do the task?
    - was there anything special about MacDonalds or Campbells?

Nosofsky (1991) experiment



- training phase: classify cartoon faces
  - MacDonalds and Campbells
- test phase:
  - classification: classify faces and rate confidence
  - recognition: provide old/new judgments
- classification and recognition had a moderate correlation (r = .36) suggesting barely much of a relationship between the two tasks
- if we knew the classification confidence, then we may not be able to predict the recognition probability

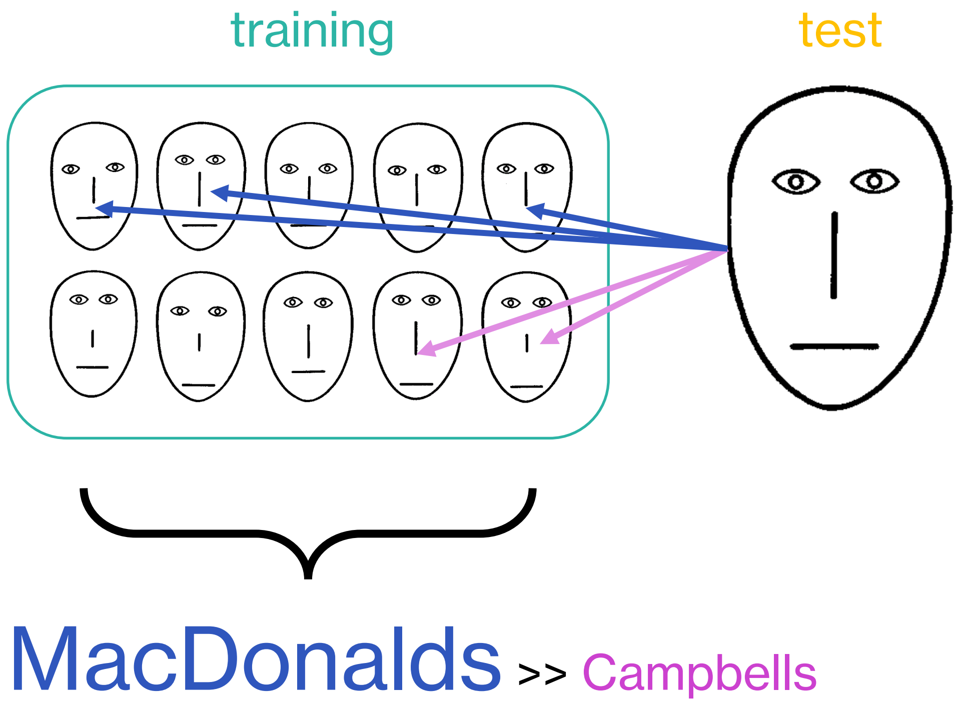


modeling classification

- Nosofsky (1991) set out to explain *how* people classify new faces after having seen examples from two different classes
- a prominent account of classification was the prototype model, which suggested that people create “general” representations of concepts to which new examples are compared
- Nososky (1991) proposed an alternative exemplar model

exemplar model: description

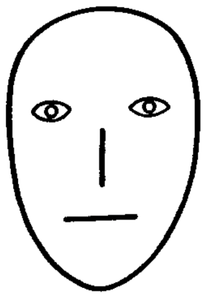
- during training, people store individual examples into memory
- during test, the training items are activated in proportion to their similarity to the test item
- the probability of responding with one label (MacDonald) vs. another (Campbell) depends on the sum of these activations



exemplar model: training

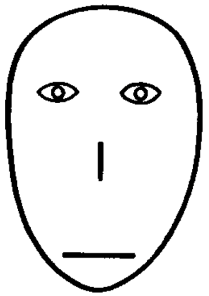
- x_i denotes the ith exemplar presented during training
- each exemplar can be defined along *m* dimensions

x_i



[-1.025, 0.493, 0.048, -0.666]

x_j



[-0.172, -0.557, 0.337, 0.163]

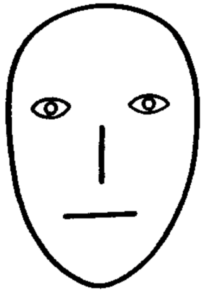
activity: computing similarities

- in groups, go to the face dimensions spreadsheet
- navigate to your group's tab
- select the columns containing face dimensions
- insert a chart and choose a "bubble" chart
- can you differentiate between MacDonalds and Campbells?

exemplar model: training

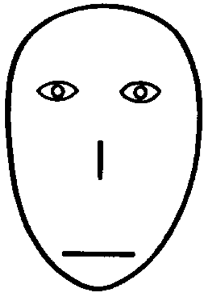
- Nosofsky (1991) varied the faces along 4 features (nose length, eye separation, etc.) such that there was a clear separation between the two classes (MacDonalds and Campbells)
- these features are often referred to as dimensions and can be placed in a multi-dimensional space

x_i



[-1.025, 0.493, 0.048, -0.666]

x_j



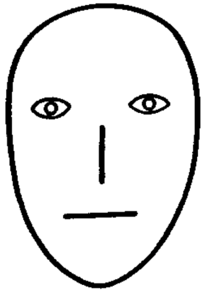
[-0.172, -0.557, 0.337, 0.163]

feature	face 1	face 2
eye height	23.5	19.5
eye separation	21.5	11.5
nose length	13.5	18
mouth height	16.5	12

exemplar model: training

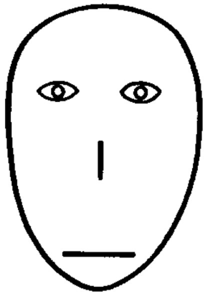
- Nosofsky (1991) varied the faces along 4 features (nose length, eye separation, etc.) such that there was a clear separation between the two classes (MacDonalds and Campbells)
- these features are often referred to as dimensions and can be placed in a multi-dimensional space

x_i

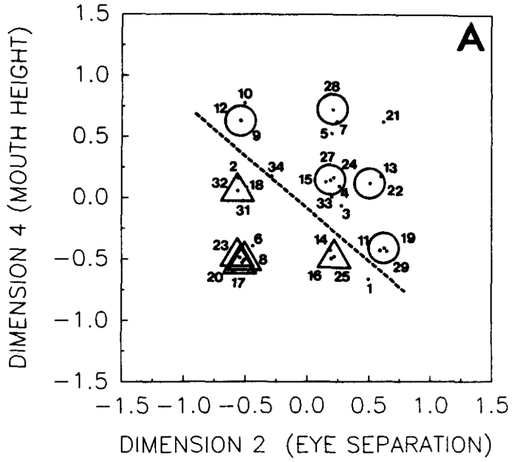


[-1.025, 0.493, 0.048, -0.666]

x_j

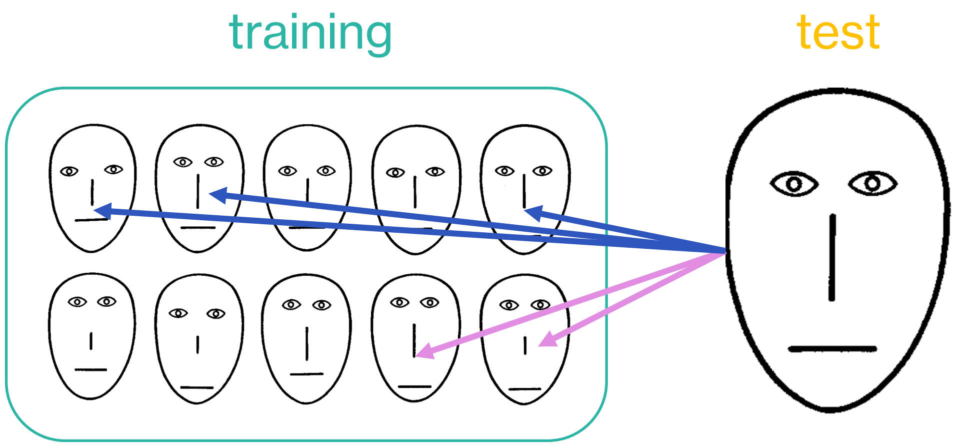


[-0.172, -0.557, 0.337, 0.163]



exemplar model: test

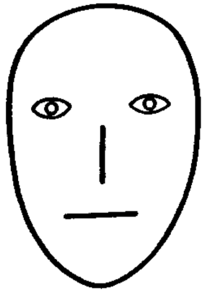
- when a new item (x_k) is presented, each training item is activated in proportion to its similarity to the test item
- but how do we calculate similarity??



exemplar model: similarity

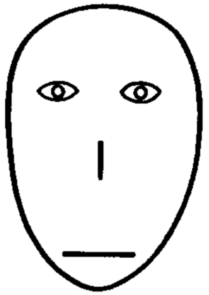
- the similarity between any two items (x_i and x_k) can be calculated using their coordinates in the multidimensional space
- this requires two steps:
  - calculating the Euclidean distance d_{ik} between the items i and k
  - translating distance to similarity through an exponential function

x_i



[-1.025, 0.493, 0.048, -0.666]

TEST (x_k)



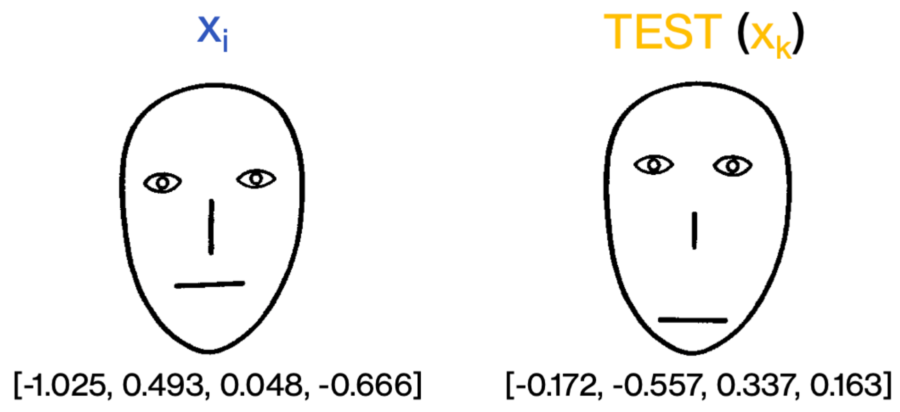
[-0.172, -0.557, 0.337, 0.163]

$$d_{ik} = \sqrt[2]{\sum_m |x_{im} - x_{km}|^2}$$

$$s_{ik} = e^{-cd_{ik}}$$

activity: computing similarities

- in groups, go to the similarity spreadsheet
- navigate to your group's tab
- use the formulas in columns F and G to compute distance and similarity of each face to the test item
- report back which face has the highest and lowest similarity to the test item

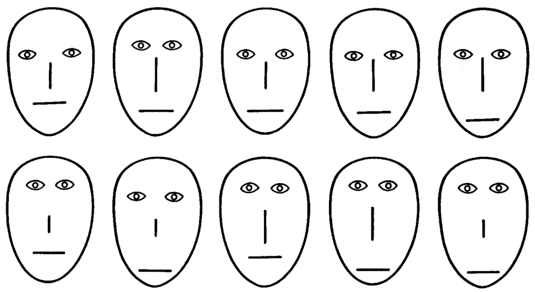


$$d_{ik} = \sqrt[2]{\sum_m |x_{im} - x_{km}|^2}$$

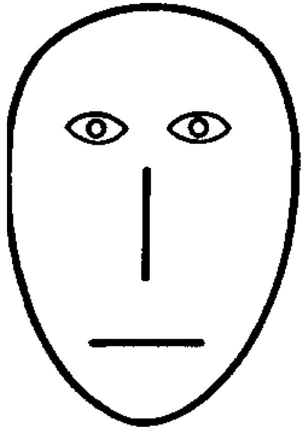
$$s_{ik} = e^{-cd_{ik}}$$

exemplar model: test

- when a new item (x_k) is presented, each training item is activated in proportion to its similarity to the test item
  - exemplar x_i is activated in proportion to its similarity to test item x_k

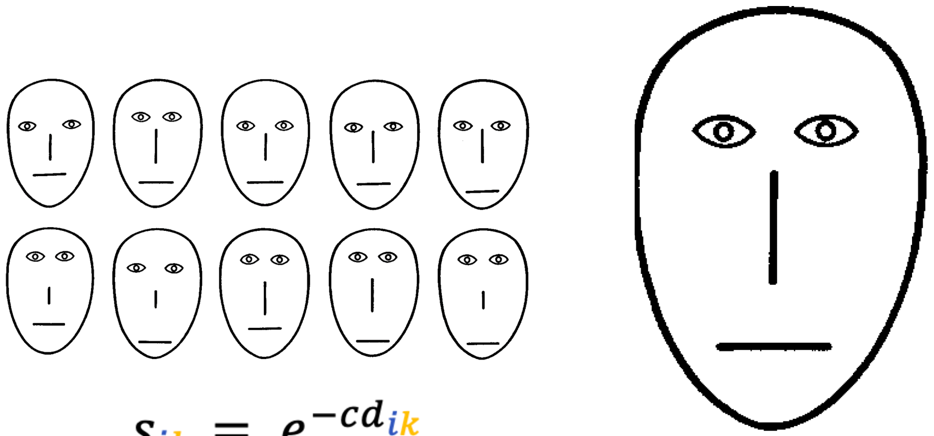


$$s_{ik} = e^{-cd_{ik}}$$



exemplar model: test

- when a new item (x_k) is presented, each training item is activated in proportion to its similarity to the test item
  - exemplar x_i is activated in proportion to its similarity to test item x_k
- activations of each exemplar in a class are added up to produce total activation for the class
- the probability of classifying the new test item is determined by whichever class has higher total activation



$$s_{ik} = e^{-cd_{ik}}$$

$$activation\ (MacDonald) = \sum_k s_{ik}$$

$$activation\ (Campbell) = \sum_k s_{ik}$$

$$P\ (MacDonald) = \frac{activation\ (MacDonald)}{activation\ (MacDonald) + activation\ (Campbell)}$$

activity: computing probabilities

- in groups, go to the probability spreadsheet
- navigate to your group's tab
- calculate the activation using the similarity values
- then calculate the probabilities
- report back the classification decision

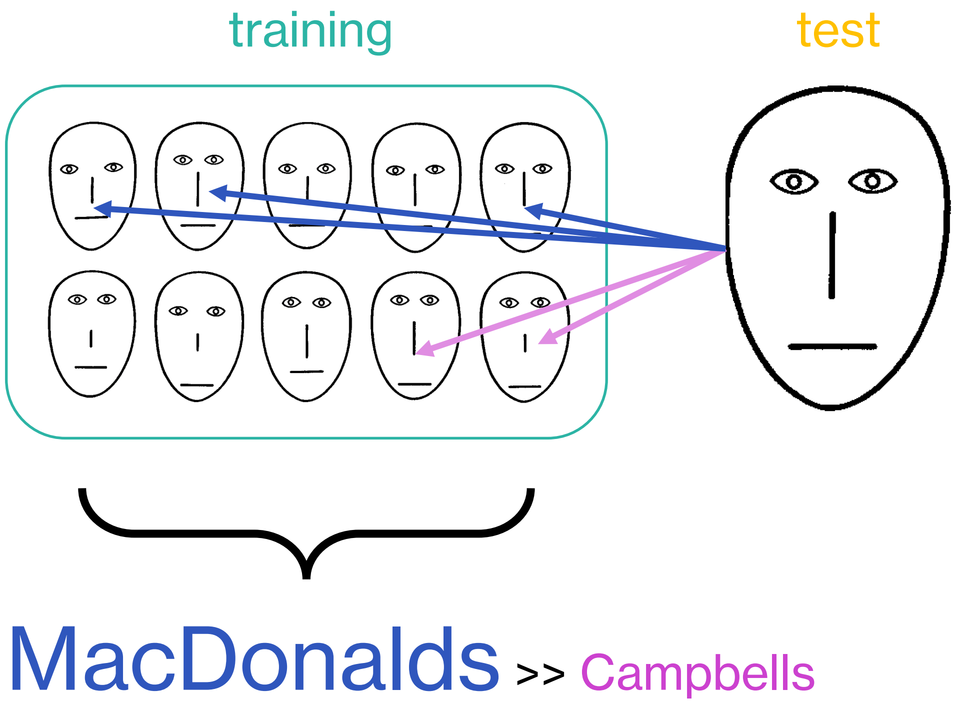
$$activation\ (MacDonald) = \sum_k s_{ik}$$

$$activation\ (Campbell) = \sum_k s_{ik}$$

$$P\ (MacDonald) = \frac{activation\ (MacDonald)}{activation\ (MacDonald) + activation\ (Campbell)}$$

exemplar model: description review

- during training, people store individual examples into memory
- during test, the training items are activated in proportion to their similarity to the test item
- the probability of responding with one label (MacDonald) vs. another (Campbell) depends on the sum of these activations



big takeaways

- get in groups of 3 and report key takeaways from today
- [takeaways document]

next class

- **before** class:
  - *finish:* L9 chapter
  - *post:* conceptual reflection
- **during** class:
  - prototype model
  - from exemplars to memory (MINERVA)