CHREST Tutorial

Key Results 1

verbal learning
categorisation
MOSAIC

Overview

We have now looked at the basic structure of CHREST, and understand its learning and recognition processes

In this section, we look at some standard results:

- Learning lists of new patterns
  - Time to learn the list
  - Forgetting
- Categorisation
- Language learning

Verbal Learning

- Behaviourism is a theory of human learning based on the idea of *conditioning* (e.g. Pavlov's dogs).
- The dominant experimental paradigm in the 1950s and 1960s for exploring this form of learning with humans was the *verbal learning* paradigm
- The early EPAM model (1959!) was developed as a computational model of these experiments
- No other computational model captures anything like as much of the data as EPAM (see EPAM VI for more information)

Verbal Learning Experiment

- Nonsense 'words' are used, QIL BAK ...
- A list of about a dozen words is formed, and presented in a series of *trials*
  - *Paired-associate task*, where subjects must learn that, e.g. BAK goes with QIL. Pairs are taken from consecutive items in list, but pairs are presented at random.
  - *Serial-anticipation task,* where subjects must learn the whole list in order
- The results from the experiment are recorded as a *subject protocol*

Subject Protocol

- A trial-by-trial list of the response of the subject for each item in the list
- Reveals many aspects of human memory:
  - *Stimulus generalisation*: where a stimulus is used to provide the response of a similar stimulus
  - *Perseveration*: where an incorrect response is repeatedly made
  - *Oscillation*: where the correct response is given first, then an incorrect response, etc.

Example of Subject Protocol

Stimulus	Response	Trial 10	Trial 11	Trial 12	Trial 13	Trial 14	Trial 15
D A G	B I F	B I F	B I F	B I F	B I F	B I F	B I F
B I F	D A X	D A X	G . .	D . .	D A .	D A X	D A X
D A X	Q I L	. . .	B I F	B I F	Q . .	Q I L	Q I L
Q I L	. . .						

- The stimulus and target response are on the left
- The actual response of the subject (human or model) is listed as a separate column
- Each pass through the list is a 'trial'

Example of Subject Protocol

Stimulus	Response	Trial 10	Trial 11	Trial 12	Trial 13	Trial 14	Trial 15
D A G	B I F	B I F	B I F	B I F	B I F	B I F	B I F
B I F	D A X	D A X	G . .	D . .	D A .	D A X	D A X
D A X	Q I L	. . .	B I F	B I F	Q . .	Q I L	Q I L
Q I L	. . .						

The highlighted area shows:
1. stimulus generalisation (because response for DAX is that for DAG)
2. perseveration (because error persists in two trials)

Example of Subject Protocol

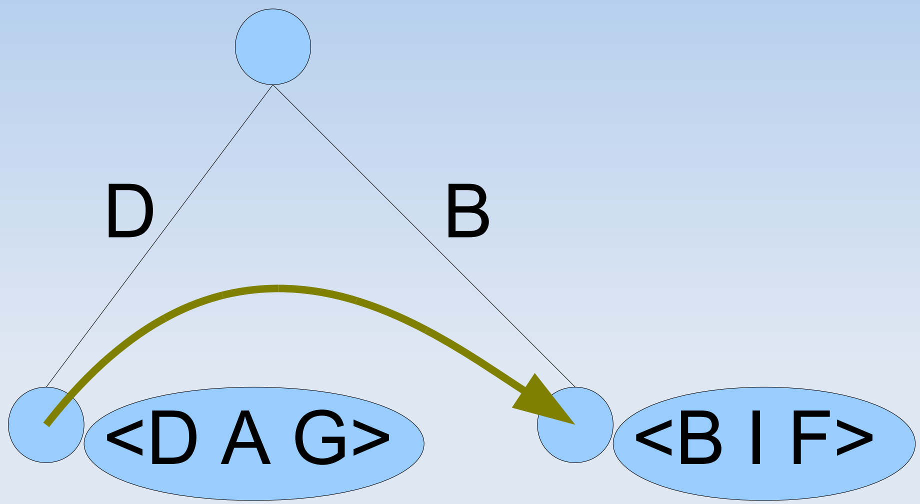
Stimulus	Response	Trial 10	Trial 11	Trial 12	Trial 13	Trial 14	Trial 15
D A G	B I F	B I F	B I F	B I F	B I F	B I F	B I F
B I F	D I X	D I X	G . .	D . .	D I .	D I X	D I X
D I X	Q I L	. . .	B I F	B I F	Q . .	Q I L	Q I L
Q I L	. . .						

- The row for BIF – DIX illustrates oscillation
- Note response is
  - correct in trial 10
  - incorrect for trials 11-13
  - correct from trial 14

Explanation of Oscillation

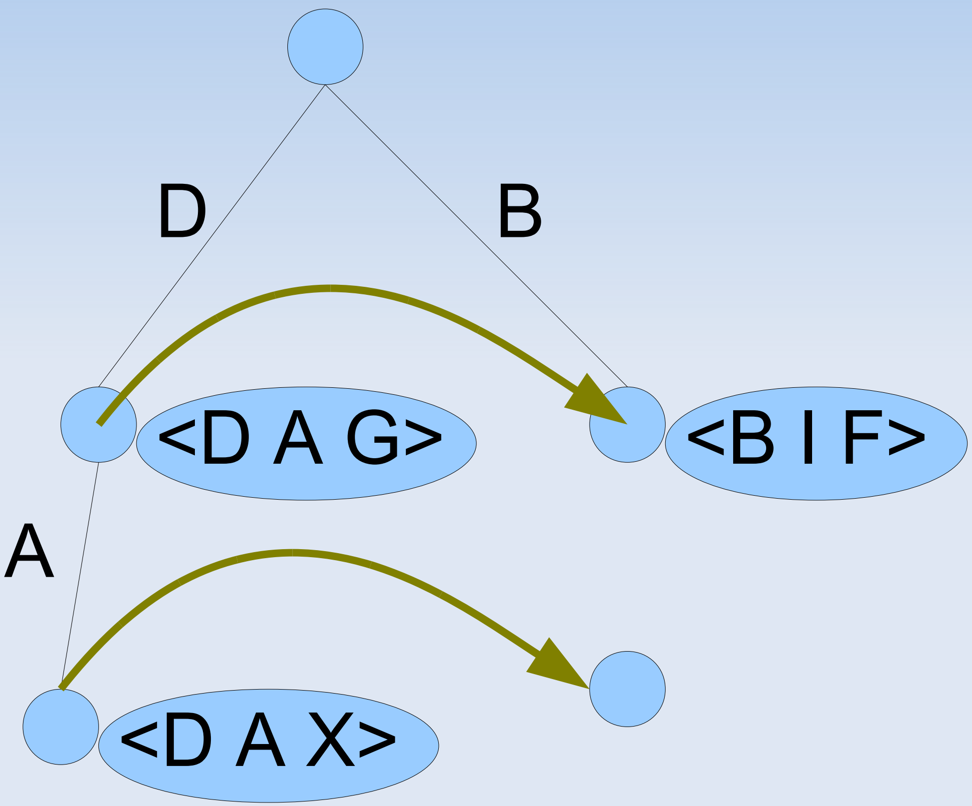
- The explanation of oscillation follows from the learning mechanisms within CHREST
  1. a pattern is learnt with its correct response
  2. a similar pattern is discriminated from the learnt pattern
  3. because of overlap, the old pattern now sorts down the new test link, so the wrong response is produced
  4. continued learning means the pattern is soon correctly associated with its response again

Explanation of Oscillation (1)



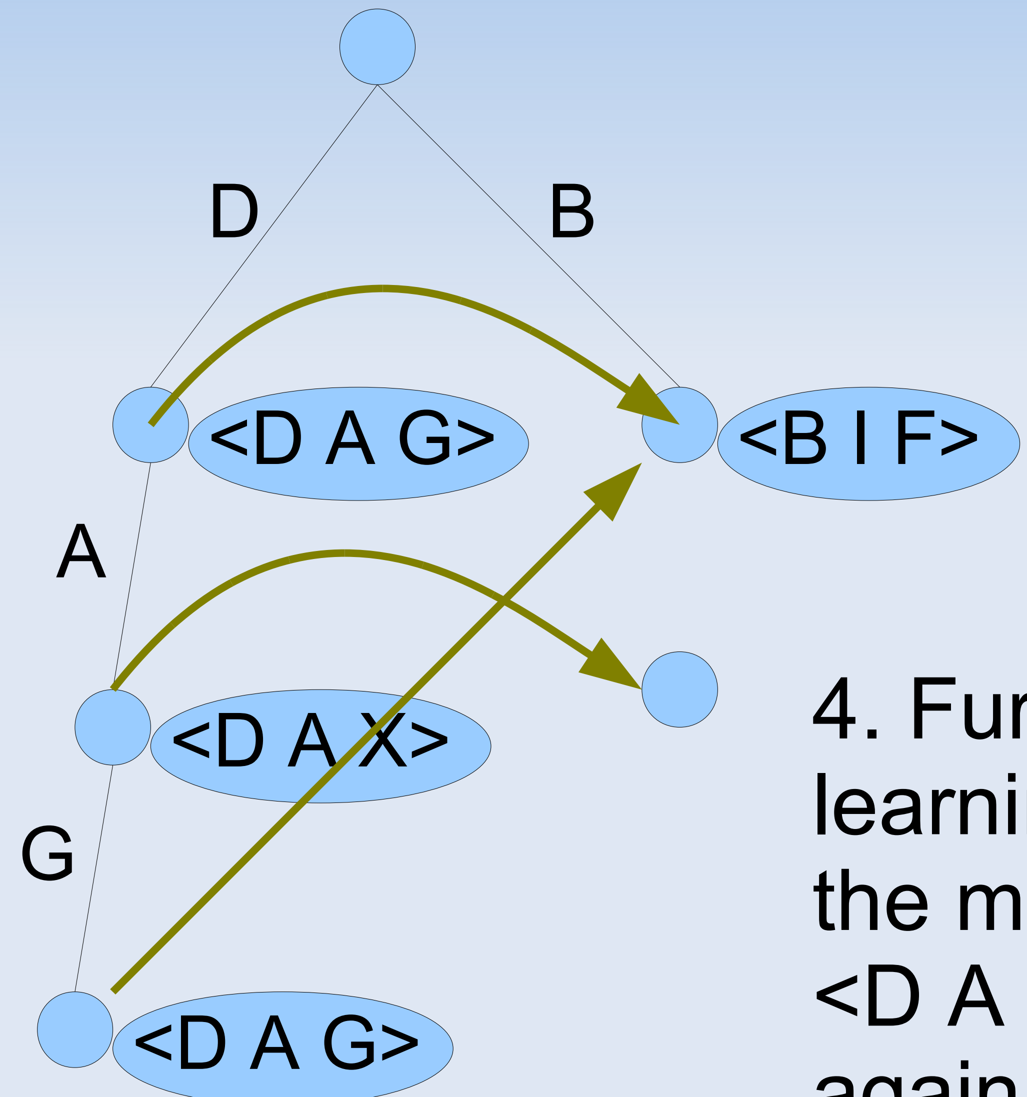
1. <D A G> is correctly associated with <B I F>

Explanation of Oscillation (2)



2. <D A X> is now learnt
3. <D A G> now sorts to the new node, and incorrect response will be given

Explanation of Oscillation (3)

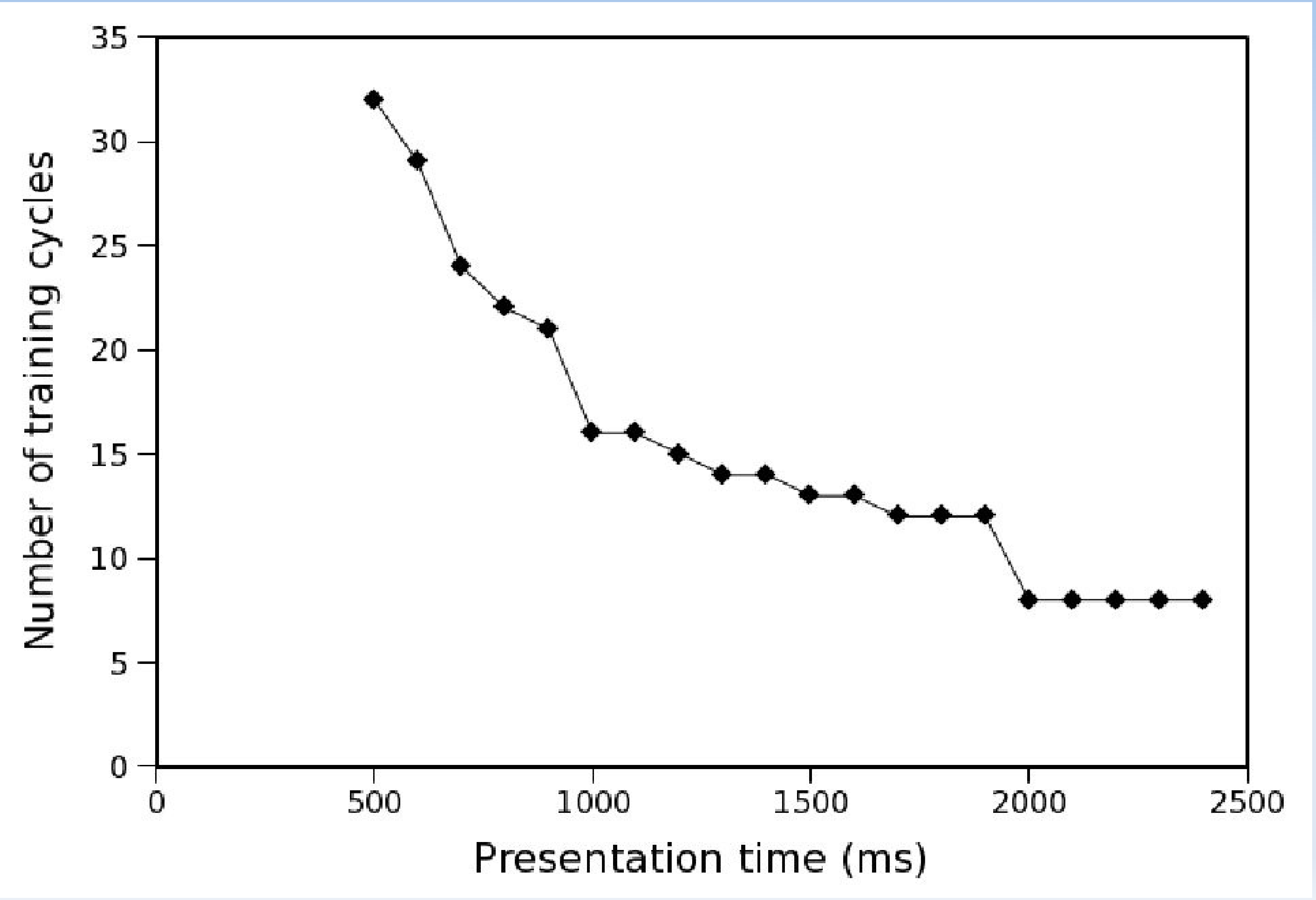


4. Further learning enables the model to get <D A G> correct again, in future

Experiment on Learning Time

- The presentation of a list takes place over time
- Each item is shown for a given amount of time, e.g. 2 seconds, or 10 seconds.
- Bugelski (1962) looked at the interaction between the number of times a list had to be presented and the amount of time that each item was shown
- An experiment with CHREST presents a list of words, varying the presentation time and recording the number of training cycles required in each case

Graph of Cycles vs Time



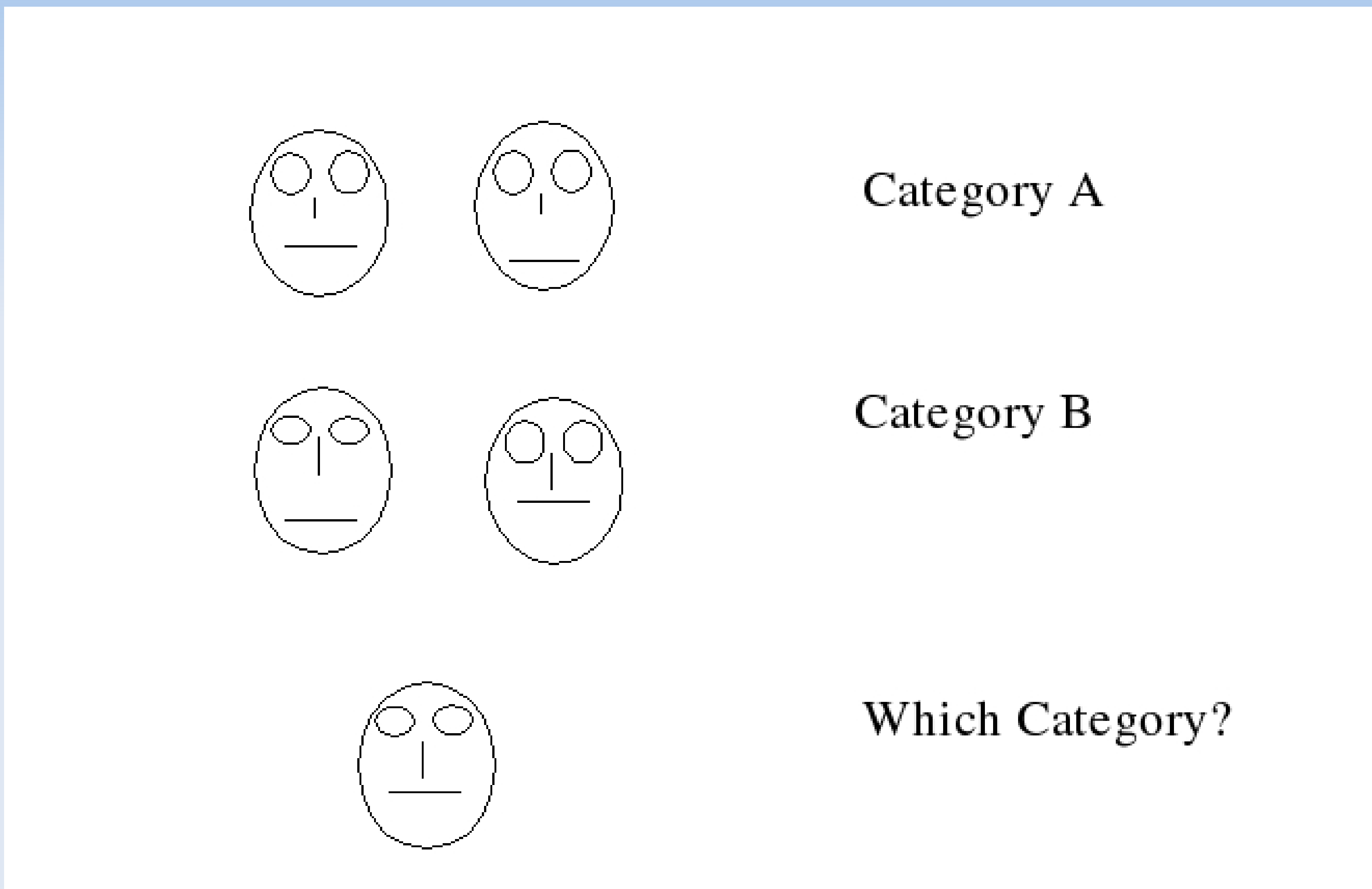
Results

- The graph shows the classic curve: the less time for presentation, the more training cycles are required
- The details of a graph such as this helps calibrate the model's learning parameters, such as the time for familiarisation and discrimination

Categorisation

- Categorisation is the process of putting patterns into named categories
- For example, you know the names of your friends, you know what a car looks like, etc.
- Categorisation is also a central problem in machine learning / data mining

Brunswik Faces



Categorisation: Experiment

- There are 16 possible faces (or other stimuli)
- 5 are defined for category A
- 4 for category B
- Participant is trained on these 9
- Participant then categorises all 16 faces
- Probability that each face is placed into category A is then computed by average across many participants

Categorisation: Results

- Example human data
  - E1: 0.78
  - E2: 0.88
  - E3: 0.81
  - E4: 0.88
  - E5: 0.81
  - E10: 0.59
  - E11: 0.31
  - etc
- Average difference of CHREST results to these data
  - 0.23
- Excellent correlation
- CHREST does not 'guess', therefore less likely to produce intermediate probabilities

Syntax Acquisition: MOSAIC

- Input is made of utterances of mothers talking to their child
- The model has no built-in syntactic knowledge
- Two main mechanisms
  - Probabilistic creation of nodes
  - Creation of lateral links based on similarity
- Similarity is defined by the context shared between two nodes

- There is a large debate in linguistics about the role of innate knowledge vs learnt knowledge

The Problem

- Children learn the syntax of their language rapidly
- Input is noisy, and feedback is rare
- Difficult inductive task
- Standard explanation in linguistics (Chomsky):
  - There is a universal grammar
  - Knowledge of universal grammar is innate

Distributional Analysis

- Statistics of the input provide significant amount of information about syntax
  - E.g., connectionism
- No approach provides detailed quantitative predictions of linguistic phenomena
- CHREST carries out a fairly simple distributional analysis

Performance Phase

- The model produces utterances in two ways:
  - By *recognition*
  - By *generation* (a lateral link is used)
- Mean Length of Utterances (MLU) increases with learning
- Identical and automatic analyses of the human data and the output of model

Optional-Infinitives in Three Languages

- English
  - Children produce utterances such as *He go*
  - Lack of an infinitival morpheme makes analysis difficult
- Dutch
  - Initially, virtually all children's utterances with verbs are root infinitives
  - This decreases to around 20% by MLU 3.5
- Spanish
  - Optional Infinitive errors are rare

Model for English

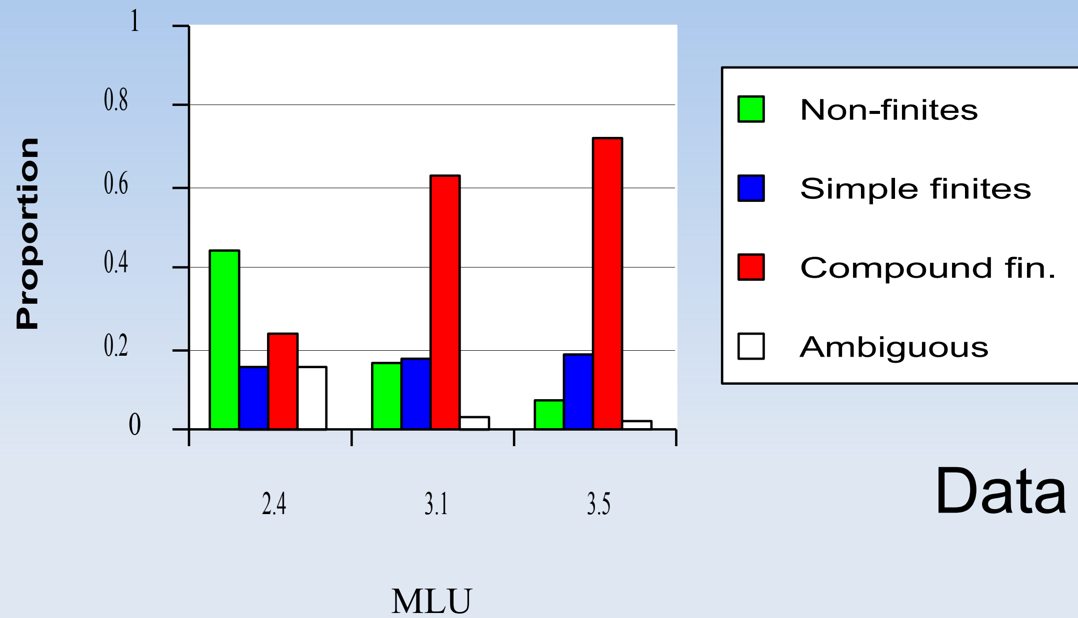
- Samples of the speech of two children were recorded
  - Becky: 2 years ⇒2 years 11 months
- Utterances with *he*, *she*, *it*, *this (one)*, or *that (one)* as subject
  - non-finites *she go*
  - simple finites *she goes*
  - compound finites *she has gone*
  - ambiguous utterances *she bought*
- Three developmental stages, defined by MLU
- Input for Becky: 27,000 maternal utterances

Model for English (2)

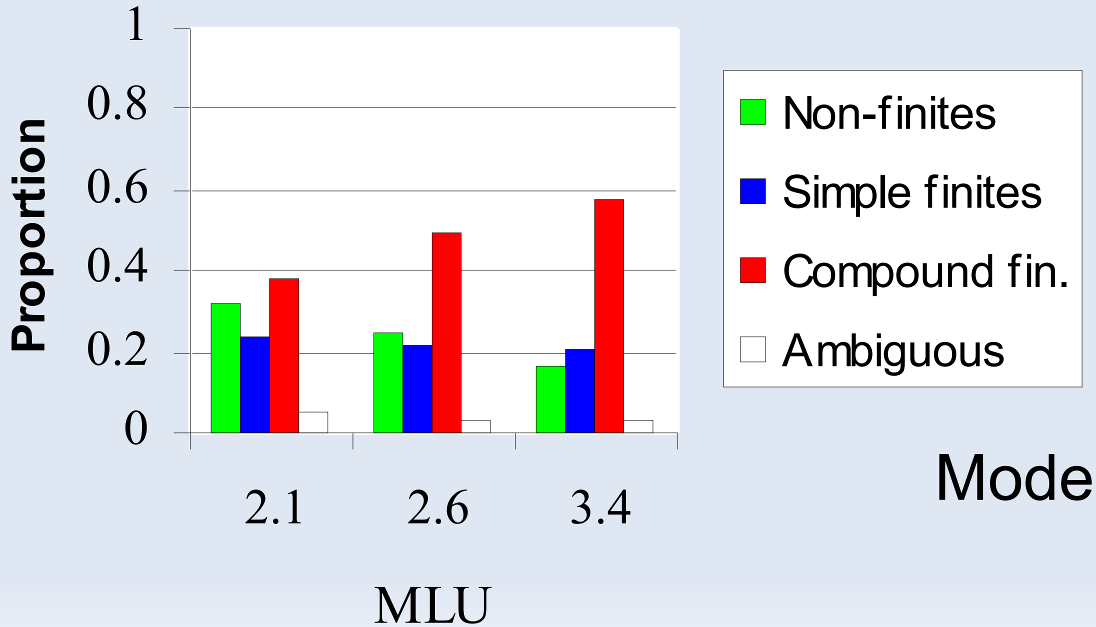
- Bias towards sentence final positions important, as non-finite utterances can be learned from compound finite questions:
  - *He walk home*

  can be learned from
  - *Did he walk home?*

Model vs Data for Becky



Data for Becky



Model for Becky

Summary

- The key mechanism for explaining OI errors is an interaction between
  - a sentence final processing bias
  - increasing MLU
- These mechanisms are fairly simple
- Identical model for simulating the same class of phenomena in four languages
  - English, Dutch, Spanish (and German)
  - In spite of obvious syntactic differences and very different proportions of optional-infinitive errors
- No need to appeal to innate linguistic knowledge to explain these phenomena