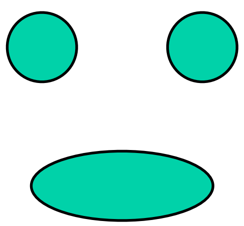
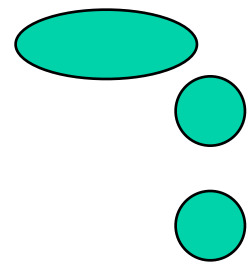
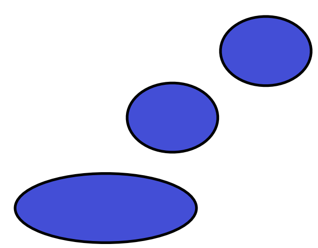
Face recognition: experience and intrinsic bias....

Infant Face Recognition







sub-cortical predisposition for eye-like things with mouth-like things?

details are learned (cortically?) from about 2 months

sheep raised with horned sheep develop "horn-cells", those without, don't

specificity comes from experience

Johnson and Morton (1991): neonates preferentially track a stimulus with 3 high contrast blobs. They do not prefer fully specified faces.

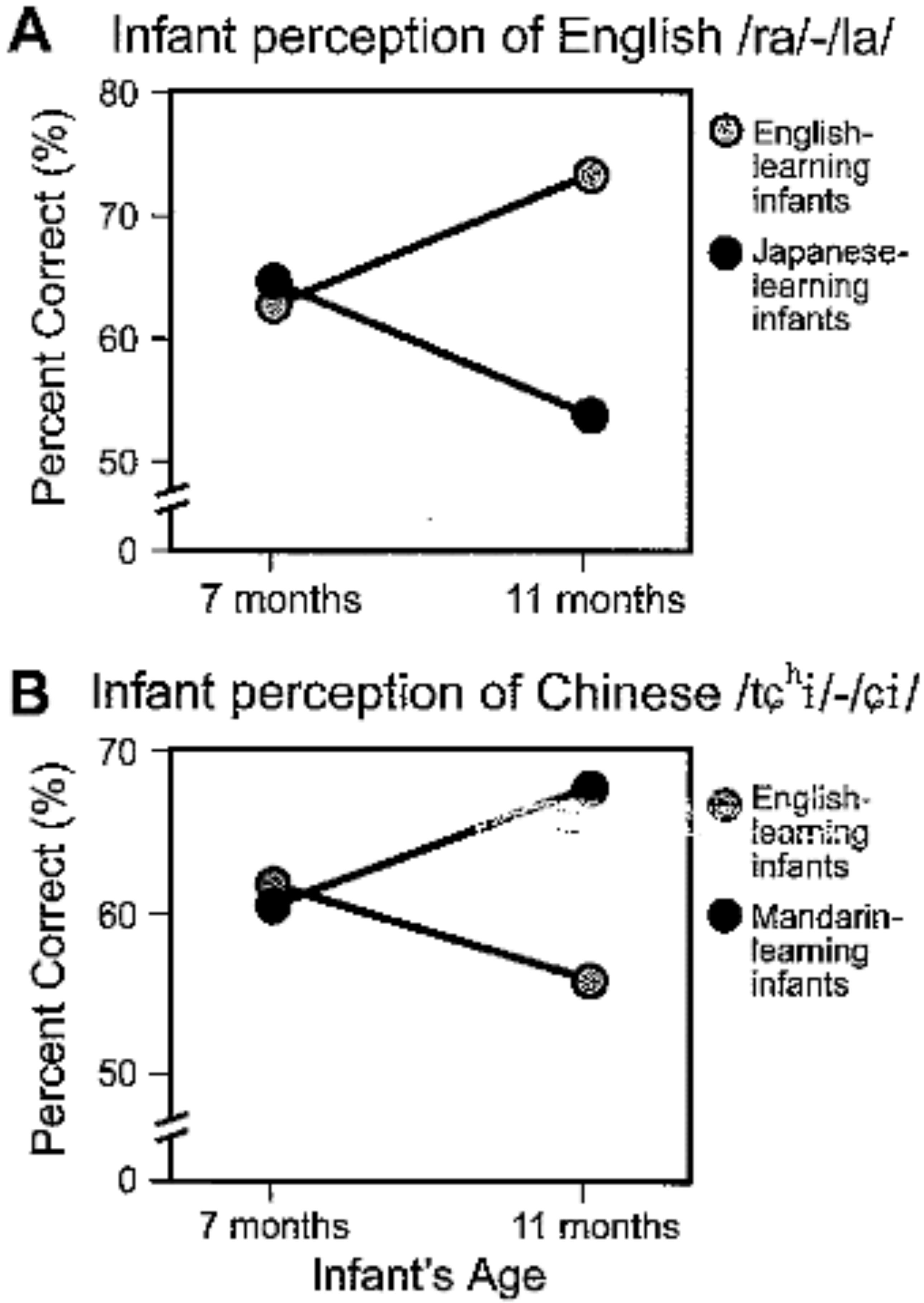
Sargent and Nelson (1992): 9 month old infants can discriminate monkey faces better than adults.

If you were to build a connectionist simulation of the development of infant face perception, what would you build in (bias)? What would you expect the data to provide?

Language: the bloodiest battlefield.

Learning sounds of a language

- Mature speech perception shows clear categorical perception
  - bad discrimination within category, good discrimination across category, variable response only at category boundary
  - categories are language specific (Eng: p/b; Thai: ph/p/b) (Eng: l/r; Jap: lr)
  - Categorical perception of consonants is well documented. vowels less so (!).
  - Some data suggesting animals may hear phonemic distinctions categorically too (non-linguistic?)
  - discrimination ability within category is lost as native language categories are learned to the exclusion of others
  - Do these argue for built-in linguistic discriminative abilities?


A Infant perception of English /ra/-/la/
Percent Correct (%)
80
70
60
50
0
7 months
11 months
English-learning infants
Japanese-learning infants
B Infant perception of Chinese /tɕʰi/-/ɕi/
Percent Correct (%)
70
60
50
0
7 months
11 months
Infant's Age
English-learning infants
Mandarin-learning infants

Categories and CP

- Categorization is fundamental to cognition
- We are discriminative
- Categorical Perception refers to a specific, limited, set of phenomena and associated theory.
- Don't confuse the two!

Categorical Perception

- To demonstrate CP, you must show
  - insensitivity to discriminations within-category
  - sensitivity to discriminations of similar magnitude across category boundaries
  - Abrupt shift in labelling at category boundary

Caution: Categorical Perception needs a health warning

- Classification of a portion of speech depends on a multiplicity of cues
- Evidence sources are combined and evaluated probabilistically
- Categorical Perception (strictly interpreted) may not be a good description of anything we do...

Schaefer and Mareschal: Modeling infant speech sound discrimination using simple associative networks, Infancy 2(1), 2001

- 8 month infants can make some discriminations that 14 month infants can not
- Qualitative shift?

Stager and Werker (1997) investigated the relationship between word learning and speech sound discrimination.

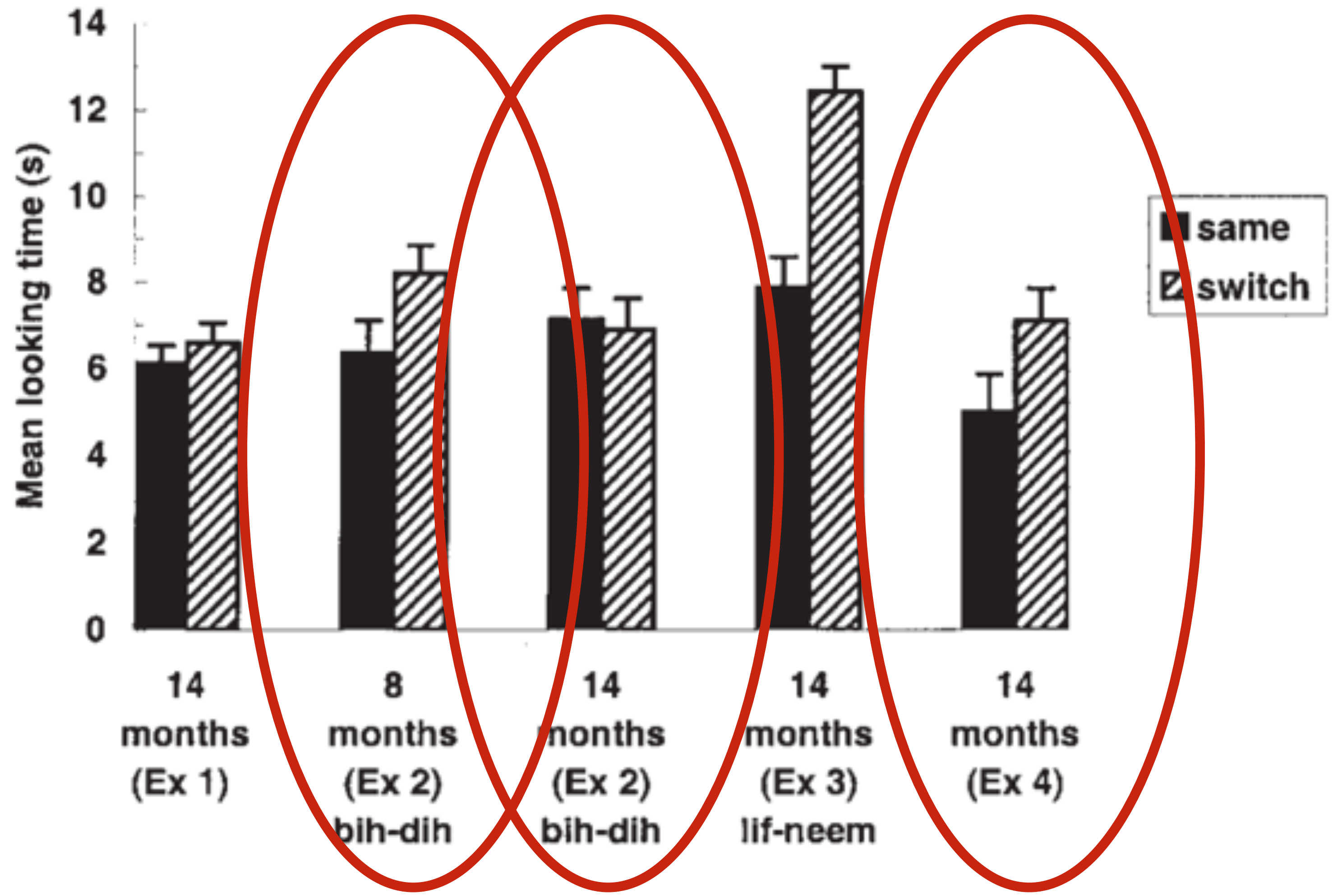
Infants learn sound-image combinations (habituation):

After habituation, on switch trials, the sound only is changed:

Operationalize measurement of “Difference in Habituation”

Looking_Time(switch) - Looking_Time(same)

Stager and Werker's results



8 month old infants can tell /bIh/ from /dIh/

14 month old infants can't: in the habituation/ dishabituation setting

14 month old infants *can* tell these apart in a simple discrimination task (checkerboard object)

Possible Account:

8 month old infants are learning (and hence sensitive) to fine phonetic detail which indexes language-specific contrasts

14 month old infants have finished that stage and are concentrating on learning word-referent relations. This biases their perceptual system differently, and accounts for the difference.

Is a change in behaviour indicative of a change in mechanism?

Schafer and Mareschal attempt to do 3 things:

[1] Demonstrate that change in behaviour does not necessarily require a change in processing mechanism or strategy

[2] present a method for modeling habituation phenomena in infants, and

[3] use their simulations to make predictions

Modelling assumptions:

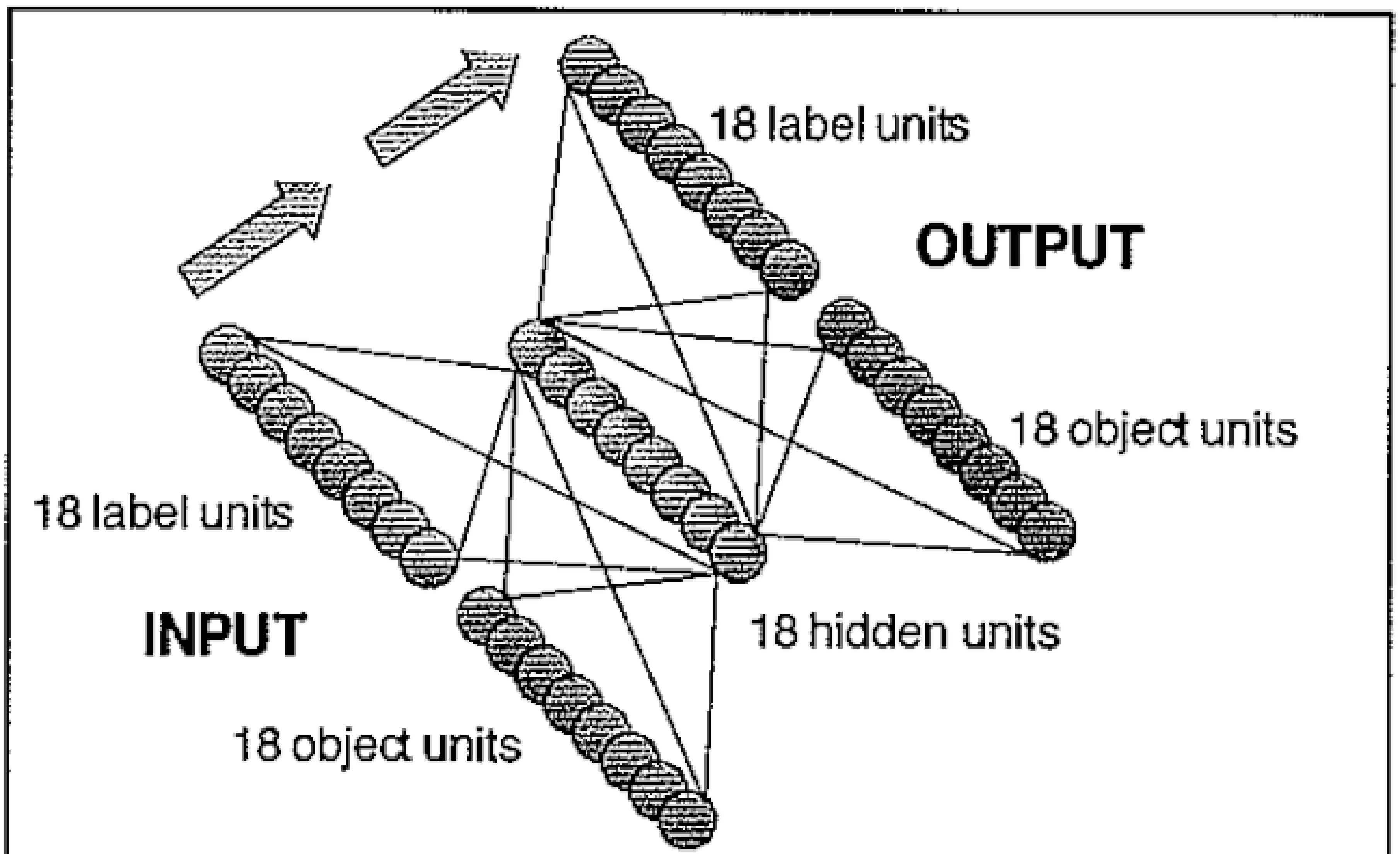
When presented with an object, infants compare the stimulus generated with an internal representation.

Mismatch required updating internal representation(s).

Attending will be longer as mismatch is more severe.

Autoassociator.

Training cycles correspond to the development of an internal representation.
Network error taken as corresponding to looking times.



Language Exposure .... then Experimental Phase

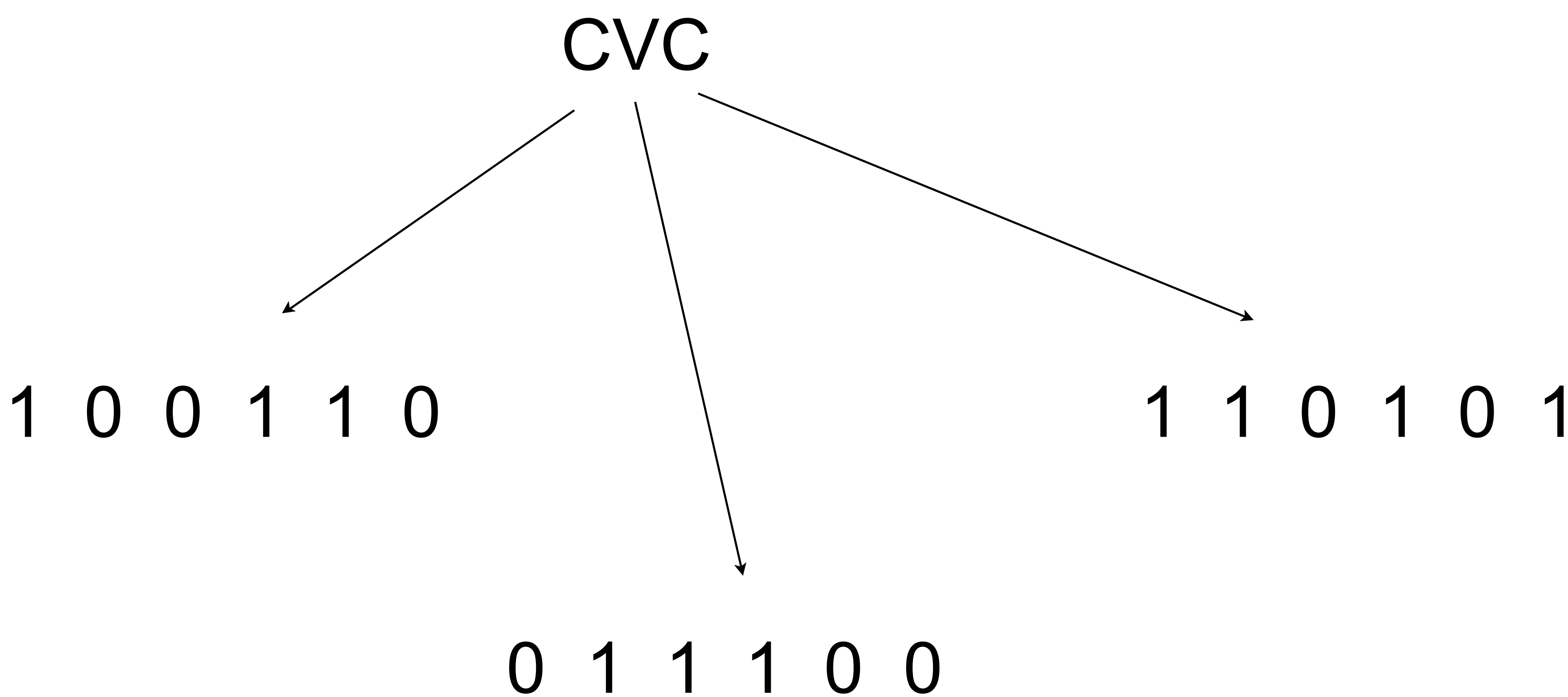
Habituation ..... then Testing

Language Exposure:

autoassociative learning of 240 label/object pairs

More training for ‘older’ networks (1000 vs 10000 trials)

Coding



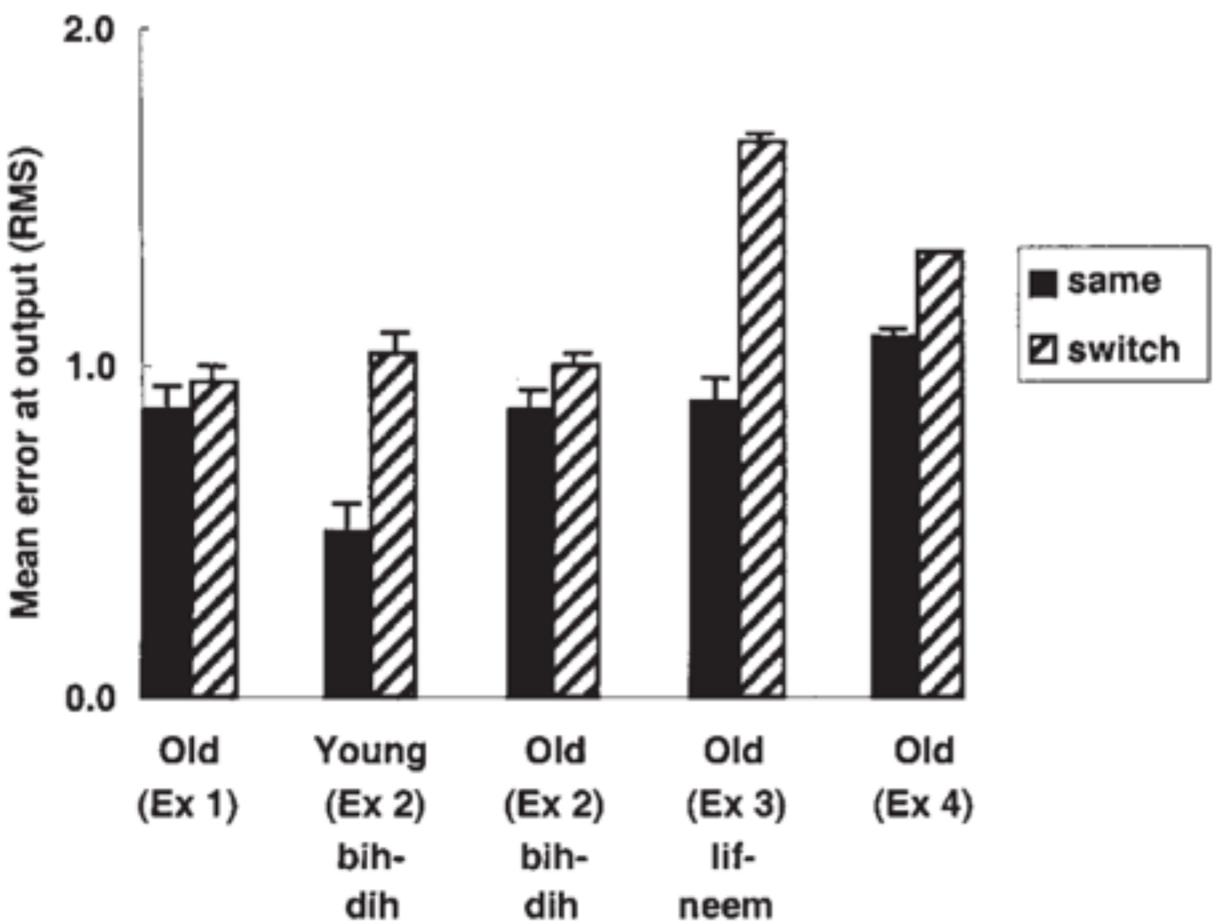
Consonantal? Voiced? Manner Place

Objects are essentially fixed, random 18-bit vectors.

... experimental manipulation involves only label change, not objects

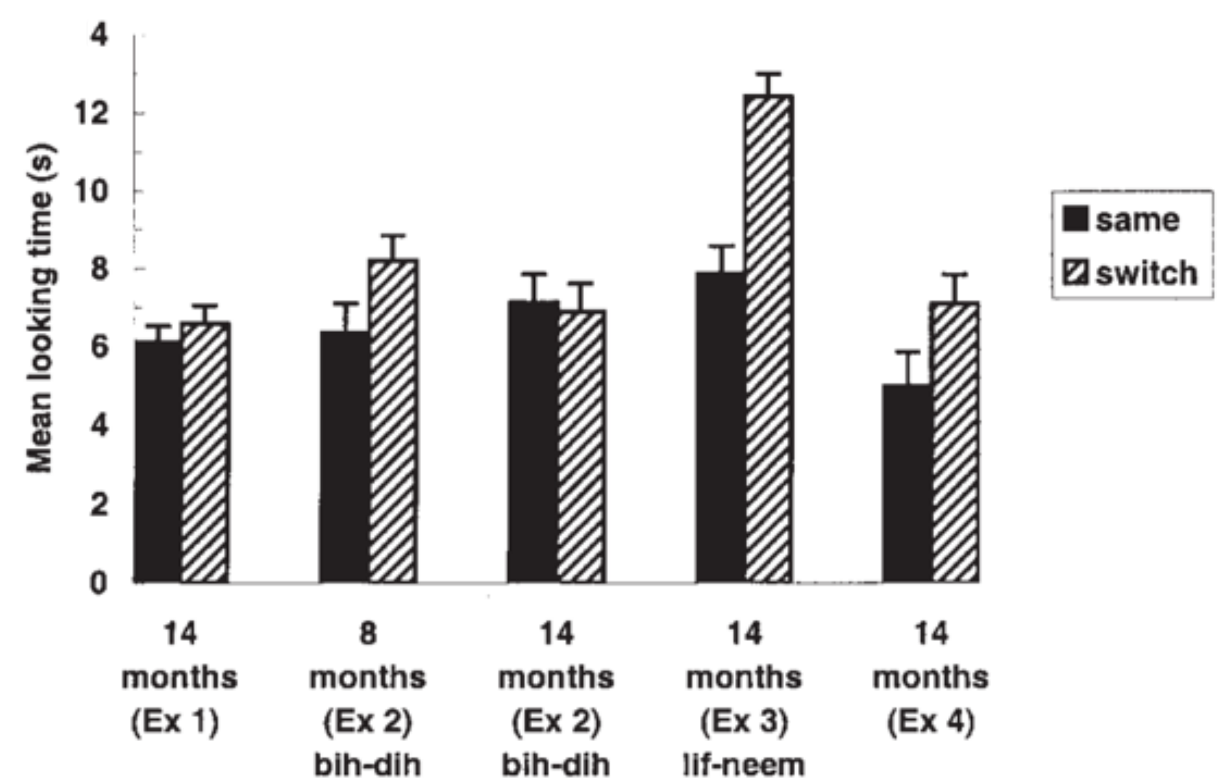
... Checkerboard pattern (non-object) has all object bits set to 0.5.

For each of the 4 experiments to be modelled, 20 networks with different initial conditions were used.

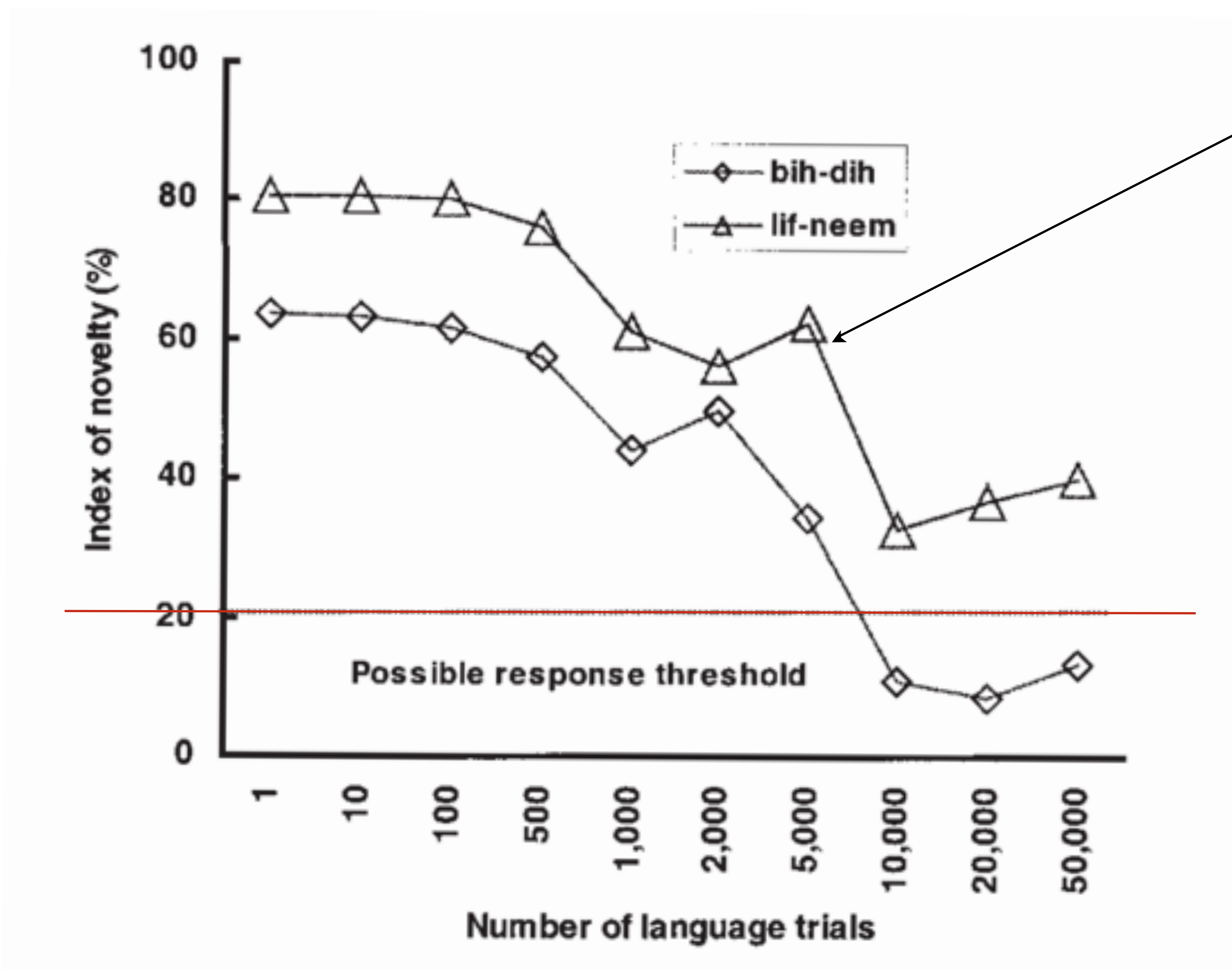


Model

Experiment



$$Novelty = (Error_T - Error_H)/(Error_T + Error_H)$$



Non-monotonicity

With threshold: different group behaviour to similar and dissimilar stimuli

What has been achieved?

Simple auto-associator learning pairs of patterns exhibits non-monotonic development,

...and apparent behavioural change despite a single underlying processing mechanism.

Some parameter fitting is ad hoc (threshold = 20%?, ‘young’ = 1,000 trials, `old’ = 10,000....), but the essential features do not depend critically on these...

`patterns’ (words/objects) are cartoonishly simplified

The Human Speechome Project

Deb Roy, MIT Media Lab

http://www.media.mit.edu/cogmac/projects/hsp.html

Also as a TED talk . . . (first 11 mins)

Learning the Past Tense

- Most verbs have regular past tenses: walk:walked, trounce:trounced...
- Many (including most frequent verbs) are irregular: go:went, spend:spent...
- During learning, many children over-regularize: go-ed, hitted, spended.....
- Overregularization typically follows learning of the irregular forms, suggesting "unlearning"
- Regular verbs: learn rules; irregulars: learn exceptions individually. 2 mechanisms.
- Early and frequent battleground for connectionist and anti-connectionists (WHY?)

Rummelhart and McClelland (1986)

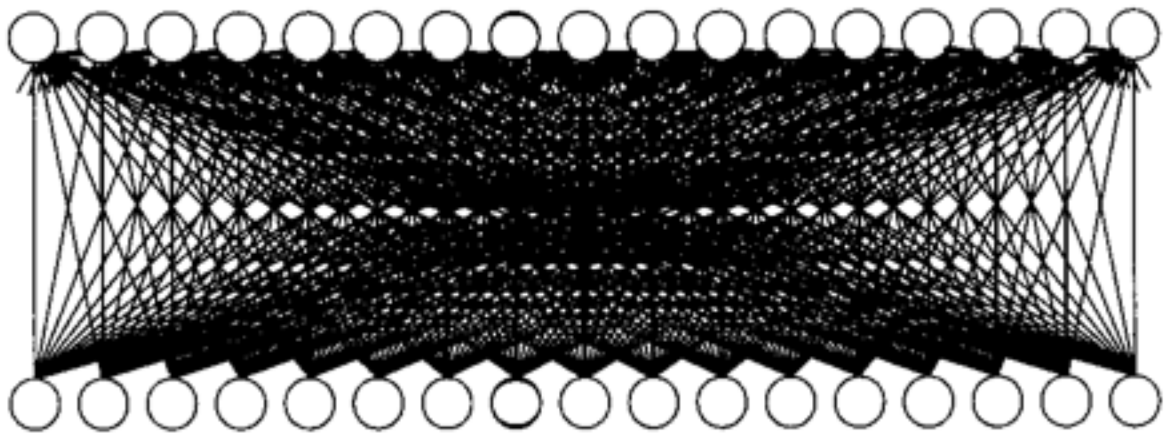


FIGURE 3.7 The learning network in the Rumelhart & McClelland (1986) model of the acquisition of the English past tense. The input is a distributed representation of the stem of the verb and output a distributed representation of the equivalent past tense.

- single layered perceptron
- mapping: stem → past tense form
- representations in terms of phonological features
- gradual training, using Perceptron Convergence Procedure

R&M's Perceptron results

- Trained on 420 stem/past tense forms
- During training, the network overregularizes, as regular verbs are more common
- Irregular verbs and regular verbs interfere during training
- After training, both are produced correctly
- Only one mechanism is used

Errors during training

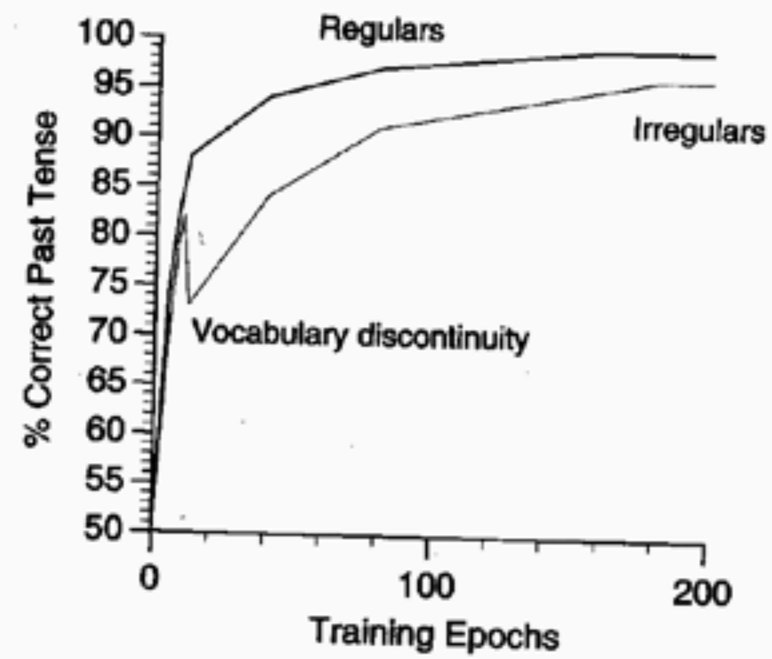


FIGURE 3.8 Performance on regular and irregular verbs in the Rumelhart & McClelland (1986) model of the acquisition of the English past tense. The vocabulary discontinuity at the tenth training epoch indicates the onset of overregularization errors in the network.

BUT.....the discontinuity in error coincides with a discontinuity in the training set used

Initial training is on 8 irregulars and 2 regulars. Then, after 10 epochs, all 420 verbs are introduced.....

Improvements to R&M's Perceptron

- Plunkett and Marchman (1991) used a network with hidden units, and a consistent training set, with relative frequencies approximating that available to children (irregulars very frequent).
- Error curves (p. 138) are quite similar to those of children
- Simultaneous learning of irregulars and regulars causes each to interfere with the other

- ....The debate continues, as models more closely approach child-like stages of development
- Children differ greatly, one from the other

Ontogenetic Development

- Modelling mature cognitive abilities differs greatly from modelling their development
- Elman attempts to show how a minimum of built-in structure (computational principles, architectural constraints) can give rise to highly differentiated mature structure
- Supervised connectionist learning is error-driven
- To what extent it this a plausible account of human development?