What does it mean to say a behavior (or knowledge) is innate?

- Brain is highly structured at birth and continues to develop/mature through adolescence
  - But not enough room in the genome to code complex knowledge (about 99% shared between humans and chimps/bonobos)
- Experience is rich and massive and begins before birth
- All complex behavior emerges from the interaction of structure and experience

“That which is inevitable need not be innate.”
— Jean Piaget
Object permanence

- When a toy is moved and hidden, young children will continue to search in the old location, even though much younger infants “know” that hidden objects continue to exist.

Munakata et al. (1997)

Puzzle: Why do infants show sensitivity to occluded objects and show an ability to retrieve objects but fail to retrieve occluded objects?

- Perhaps infants “know” the occluded object exists but lack the means-ends ability to retrieve it.

Munakata et al. trained 7 mo infants to retrieve distant objects lying on a towel, separated by a transparent or opaque barrier:

- Infants had no difficulty retrieving object viewed through transparent barrier.
- Lack of “object permanence” when reaching (vs. looking) not due to differences in means-ends demands.

Model (Munakata et al., 1997)

Internal Representation Units

Recurrent Weights

Encoding Weights

Prediction Weights

Input Units

time0  time1  time2  time3  time4  time5  time6  time7
Training

Analysis of hidden representations

Persistence of object information over occlusion steps
Memory for object “features”

Dissociation of looking vs. reaching

Memory for familiar vs. unfamiliar objects

A “critical period” in language learning? (Elman, 1991)

- Simple recurrent network trained to predict next word in English-like sentences
- Context-free grammar, number agreement, variable verb argument structure, multiple levels of embedding
- 75% of sentences had at least one relative clause; average length of 6 words.
- e.g., Girls who cat who lives chases walk dog who feeds girl who cats walk.
- After 20 sweeps through 4 sets of 10,000 sentences, mean absolute error for new set of 10,000 sentences was 0.177 (cf. initial: 12.45; uniform: 1.92)
The importance of “starting small” (Elman, 1991)

- Training was successful only when “starting small”
  - Trained on only simple sentences before gradually introducing embedded sentences
  - Trained on full language but with initially limited memory that gradually improved
- Consistent with Newport’s (1990, *Cog. Sci.*) “less is more” hypothesis
  - Child language acquisition is helped rather than hindered by maturational limits on cognitive resources

Alternative Hypothesis: Need to start small was exaggerated by lack of important soft constraints inherent in natural language

- SRN’s learn long-distant dependencies better when intervening material is partially correlated with distant information (Cleeremans et al., 1989, *Neural Comp.*)
- Soft semantic constraints—distributional biases on noun-verb co-occurrences across clauses—provide such correlations

Simulation 1: Semantic constraints (Rohde & Plaut, 1999)

- Replication of Elman (1993) simulation with addition of constraints on verb arguments
- Parametric variation of reliability of semantic constraints across clauses
  - \( A = \) none, ..., \( E = 100\% \) reliable
- Minor improvements in technical aspects of simulation (e.g., error function, initialization)

Comparison of two training regimens:

- **Complex**: Trained on full language throughout
  - 25 epochs through 10,000 sentences (75% complex)
- **Simple**: Trained incrementally
  - 5 epochs on simple sentences; 5 on 25% complex; 5 on 50% complex; 10 on 75% complex

Results: Prediction error

- Disadvantage for “starting small” that increases with reliability of semantic constraints

Relation to Elman’s (1993) results

- Exact replication, varying magnitudes of initial random weights
  - Simulation 1 used ±1.0; Elman used ±0.001
  - Very small initial weights prevent effective accumulation of error derivatives
Simulation 2: Native vs. late bilingual acquisition

Languages
- English: Analogous to language from Simulation 1
- German: German vocabulary (“hund” vs. “dog”), gender marking, case-marking in masculine, verb-final relative clauses
- Phoneme-based input and output representations

Training Conditions
- Monolingual: Trained on either English or German
  - 6 million sentence presentations sampled from corpus of 50,000 sentences
- Native Bilingual: Trained on both English and German (50/50)
  - 6 million sentence presentations sampled from two corpora of 50,000 sentences each
  - Language selected randomly every 50 sentences
- Late Bilingual: Monolingual training followed by bilingual training

Testing
- Late Bilingual tested on L2 (new sample of 5,000 sentences)
- All results counterbalanced for English vs. German

Results: Acquisition
- Initial monolingual training impedes subsequent bilingual acquisition
- Native bilingual acquisition is only slightly worse than monolingual acquisition

Results: Early-bilingual acquisition
- Even relatively brief exposure to monolingual L1 impacts subsequent L2 acquisition

Language learning: Conclusions
- Introducing soft semantic constraints aids learning of pseudo-natural languages by simple recurrent networks
  - No need to manipulate training environment or cognitive resources
  - Networks inherently learn local dependences before longer distance ones
- Critical-period effects may reflect entrenchment of representations that have learned to perform other tasks (including other languages)
  - No need to introduce additional maturational assumptions (e.g., “less is more”)