A simulated twin study exploring the heritability of past tense acquisition in a population of neural network models

Michael S.C. Thomas, Angelica Ronald & Neil A. Forrester

m.thomas@bbk.ac.uk, a.ronald@bbk.ac.uk, http://www.psyc.bbk.ac.uk/research/DNL/

Developmental Neurocognition Lab, Birkbeck College, University of London, UK

How do genes constrain language development?

Evidence of heritability indicates the influence of genes on behaviour

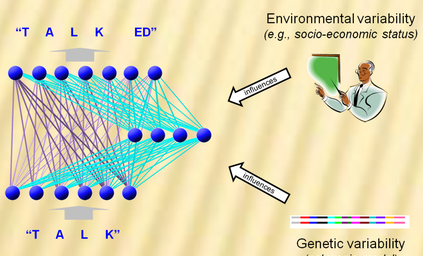
What are the processing mechanisms underlying genetic influences on cognition? How can this be fitted in a developmental framework?

Our approach: use neurocomputational models of cognitive development; simulate variability in a population from genetic & environmental causes

Aim: simulate a twin study for one aspect of language development: English past tense acquisition

Crucially, behaviour is the **outcome of an implemented developmental process in a psychologically plausible cognitive model**

Method



Connectionist models: Cognitive level simulations of behaviour implemented in artificial neural networks. Networks contain neurocomputational parameters. Parameter settings affect how efficiently a task domain is learned

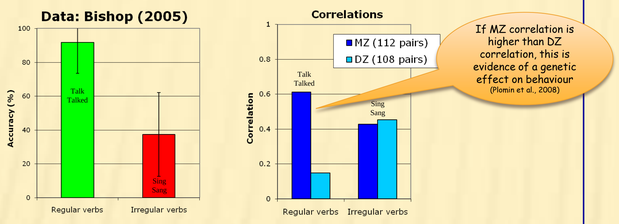
What's new: Network parameters are encoded as genomes. Using genetic algorithms, population of genomes are bred. Individuals are created with related genomes: MZ twins share the same genome, DZ twins share 50% of alleles on average

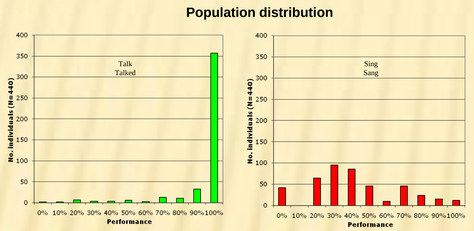
Genetically determined parameters	*Shared environmental factors*	*Unique environmental factors*
Network architecture Resources (internal units) Sparseness of connectivity Initial weight variance Rate of weight decay Unit discriminability Pruning onset Pruning threshold Pruning probability Learning algorithm Learning rate Learning momentum Level of processing noise Response threshold	'Perfect' training set Family quality (determines family training set) **Sources of variability**	Initial weight values Initial weight connectivity Initial unit thresholds Order of learning experiences On-line processing noise Probabilistic connection pruning Initial pre-training set (subjective experiences) **(Measurement error)**

Target empirical data

Performance of 440 6-year-old children on an English verb past tense elicitation task. Compare performance on regular verbs (*talk-talked*) with irregular verbs (*sing-sang*)

Twin study with 112 MZ twins and 108 DZ twins (Bishop, 2005)





Bishop, D. V. M. (2005). DeFries-Fulker analysis of twin data with skewed distributions: cautions and recommendations from a study of children's use of verb inflections. Behavior Genetics, 35, 479-490.

Plomin, R., DeFries, J. C., McClearn, G.E ., & McGuffin, P. (2008). *Behavioral Genetics (5th ed.)*. New York: Worth Publishers

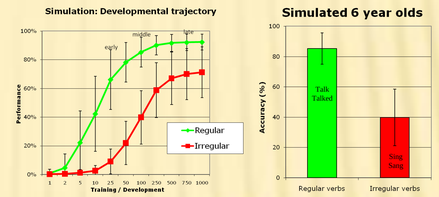
Network simulation details

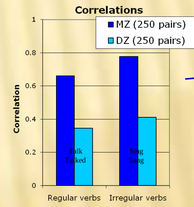
- We simulated a population of 1000 individuals, comprising MZ and DZ twins
- Each individual was a neural network for learning the English past tense. Networks were trained to associate phonological representations of present and past tense for regular (*talk-talked*) and irregular (*sing-sang*) verbs (Plunkett & Marchman, 1993)
- Each network had a genome
- 14 genes specified the computational parameters of the network that could vary in the population (encoded as binary strings)
- 55 alleles were available across the 14 genes = ~50 million possible genotypes
- Initially, 1000 random genotypes were generated and placed into 500 pairs (parents)
- Each family had a pair of twins
- 250 pairs of MZ twins, 250 pairs of DZ twins
- Each individual experienced some random pre-training on arbitrary input-output mappings to simulate unique experience
- The 'perfect' training set contained the full set of present-past tense pairs for regular and irregular verbs
- Each family was randomly assigned a 'quality' value between 0.6 and 1. Quality was used to randomly select a subset of the perfect training set. The family training set was used as shared environment for both twins in a pair. Manipulation can be seen as representing, e.g., SOCIO ECONOMIC STATUS
- Random noise was added to performance to simulate measurement error (based on known test-retest reliability of experimental measure of past tense = 0.8)
- Each individual generated a developmental trajectory (1000 presentations of training set)
- Simulation was initially matched to empirical data at appropriate accuracy level for 6 year olds
- For this population simulation, all causal factors could be assigned as (i) under genetic control, (ii) shared environment, or (iii) unique environment

Plunkett, K., & Marchman, V. (1993). From Rote Learning to System Building - Acquiring Verb Morphology in Children and Connectionist Nets. Cognition, 48(1), 21-69.

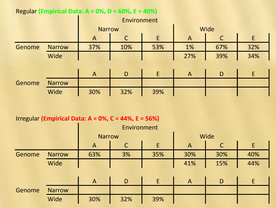
Simulation Results

(i) Phenotypic performance*

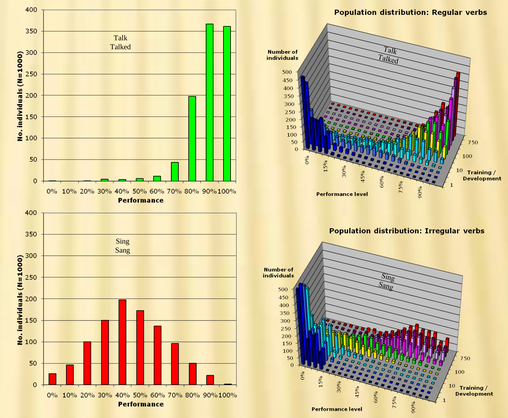




Genetic effects on regular verbs captured, but model does not capture strong shared environmental effects on irregular verb performance

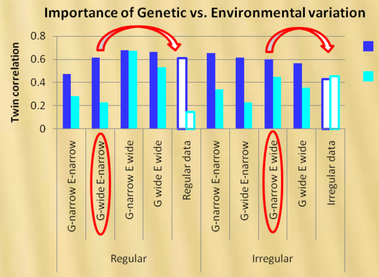


(ii) Population variability*



* For condition with wide genetic variation and narrow environmental variation

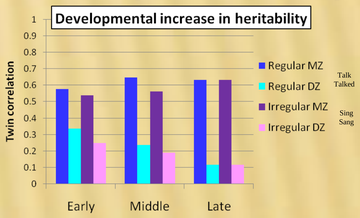
(iii) Heritability depends on relative genetic and environmental variation for the population



Genetic variation: wide vs. narrow

Environmental variation: wide vs. narrow

(iv) Heritability and development*



Discussion

- Demonstrates the viability of population studies of variability in cognitive development using implemented computational models
- Top-down approach to exploring the causal mechanisms that generate genetic effects on behaviour (complements bottom-up molecular approach)
- Allows genetic and environmental effects to be placed within a developmental context
- Supports polygenic view: many small genetic effects explain population variance
- Here, genetic effects act on lower-level neurocomputational parameters in experience-dependent processing structures
- **Heritability does not directly reflect genetic processes**, because it depends on the pre-existing range of genetic vs. environmental variation in the population
- Next challenge: calibrate the respective variability present in the genotype versus in the environment – key in determining heritability but not easily empirically constrained

This research was funded by UK MRC Career Establishment Grant G0300188 and EC grant 0209088 (NEST). We thank Prof. Dorothy Bishop for making the raw data available to us.