Modelling the mechanisms underlying population variability across development: Simulating genetic and environmental effects on cognition

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

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

Word count (main text): 25K

Running head: Modelling genetic and environmental effects on cognitive development

Address for correspondence:

Dr. Michael Thomas
Developmental Neurocognition Lab
Centre for Brain and Cognitive Development
Department of Psychological Science,
Birkbeck College,
Malet Street, Bloomsbury
London WC1E 7HX, UK
Email: m.thomas@bbk.ac.uk
Tel.: +44 (0)20 7631 6468
Fax: +44 (0)20 7631 6312

Abstract

We set out to implement a model of one aspect of language development that linked between multiple levels of description: genome, neurocomputation, overt behaviour, and environment. The goal was to render explicit assumptions linking the three levels of description and to evaluate their consequences via implementation in a complex learning system. By combining Artificial Neural Networks with Genetic Algorithms, we simulated population variability in development. We employed English past tense acquisition as an illustrative problem domain and compared model performance to empirical data from a population of 442 English 6-year-old children (Bishop, 2005). Assuming a polygenic framework and additive genetic effects, we assessed the contributions of individual 'genes' (which coded for neurocomputational parameters) and environmental factors (which coded for variations in information content, implementing the effects of *socio-economic status*) in generating behavioural variability across development in a large number of simulated individuals. Results explored: (1) changes in population variability over development; (2) effect sizes of genetic and neurocomputational parameters in predicting behavioural variability, including simulated *association analyses*; (3) the power of the environment to predict behavioural variability, including a novel prediction subsequently confirmed by the empirical data; (4) mechanisms of resilience in development, including gene-environment interactions and apparent gene-gene interactions. The model unifies an account of developmental change and individual differences within a single explanatory framework, and highlights possible disparities between causal and statistical relationships in mapping between genes, environment and behaviour.

Keywords: Genes, environment, socio-economic status, cognitive development, language development, gene-environment interactions, resilience, association analyses, computational modelling, artificial neural networks, genetic algorithms.

Introduction

The aim of this paper

The aim of this paper is to explore the mechanisms that contribute to population variability in cognitive development; and specifically, to consider the contribution of genetic variation and environmental variation to the range of behaviours exhibited by a population during the acquisition of a given ability. To do so, we employ connectionist models of development, and exemplified our approach with the case study of English past tense acquisition. The behavioural variability exhibited by a population of neural network models was compared to the performance of 442 6-year-old children on a past tense elicitation task (Bishop, 2005). Our goal was to simulate empirical data on the distribution of performance levels exhibited by these children and the role of the environment in predicting variation; and then to explore what associations might be expected between behaviour and, respectively, neurocomputational and genetic levels of description.

Beyond the average child – individual differences in development

Most studies of cognitive development seek to characterise the average child, and trace the improvement of cognitive abilities with age. Variation of cognitive abilities around the average profile of development is more often considered within the framework of individual differences or intelligence research, or in cases of more extreme variation, within the study of development disorders and giftedness at the respective ends of the distribution. Studying average development allows for more focused sampling, because the average of a sample can give a reasonable approximation of the average of a larger

population. Sampling and averaging naturally de-emphasise the consideration of variability within a developmental context.

Population studies of cognitive development are challenging to carry out for practical reasons. Large numbers of children must be assessed over an extended time period. Such studies are usually undertaken within a clinical framework, to identify the expected range of abilities at different ages so that children falling outside of this range can be targeted for intervention. Population studies of development are effectively undertaken when researchers seek to create standardised tests, for example of verbal versus non-verbal abilities, or at a finer granularity, receptive versus productive language. By their nature, standardised tests often have limited sensitivity (in the sense that they rely on accuracy data) and constitute fairly blunt measures of individual cognitive processes. By contrast, the kinds of sensitive experimental measures that target individual cognitive processes are impractical to give to population samples. Nevertheless, in principle, population studies would allow us to address a range of interesting questions. For example, one might ask whether the causes of cognitive variability are the same across the range of ability observed in the population, or whether they differ in the lower and upper tails. One might ask whether the causes of variability remain constant over developmental time or whether they alter. What mechanisms cause children to change their rank order in the population, for example in the case of late developers? Are the mechanisms that explain within-age variation the same as those that explain between-age variation, so that being more intelligent is in fact equivalent to having extra cognitive development? Or does intelligence operate over orthogonal dimensions to development?

We will argue here that shifting our focus to individual differences with respect to development is, in addition, a powerful way to relate genetic, neural, and behavioural levels of observation. Such a foundation is likely to prove important for targeting educational and intervention strategies to those children who will maximally benefit. By understanding the influence of internal constraints, the effects of environmental variation can be optimised.

Using population studies to partition genetic and environmental influences

Despite the practical difficulties, population studies are more common within the fields of epidemiology, behavioural genetics, and molecular genetics. While these fields have not focused on characterising the mechanisms of cognitive development, the findings emerging from behavioural genetics have had increasing implications for the study of cognition. These stem from evidence of high degrees of heritability for cognitive abilities (e.g., Plomin, Owen & McGuffin, 1994), to sophisticated designs to unpick the influence of genes versus parenting in the transmissions of behaviours across generations (e.g., Caspi et al., 2004; Jaffee et al., 2005), to an increasing number of candidate genes for explaining population variation in behaviours such as language, attention, and general cognitive ability (e.g., respectively, Smith, 2007; Posner, Rothbart & Sheese, 2007; Deary, Spinath & Bates, 2006), as well as genes implicated in disorders such as autism, dyslexia, schizophrenia, and attention deficit hyperactivity disorder (e.g., respectively, Abrahams & Geschwind, 2008; Fisher & Francks, 2006; Schwab & Wildenauer, 2009; Gizer, Ficks & Waldman, 2009). Moreover, as large cohorts of twins are followed longitudinally over development, there are increasing opportunities to gain a richer

understanding of causal relations over time. With regard to the investigation of mechanisms of cognitive development, however, these population studies still incur practical limitations. Specifically, they often use standardised tests and questionnaire data that serve as only coarse measures of cognitive processes. Nevertheless, a number of key findings are now becoming apparent.

What have behavioural and molecular genetic studies told us about the mechanisms underlying cognitive development?

Behavioural genetic researchers use population studies to evaluate the extent to which relatedness between individuals predicts similarity in behaviours, while making assumptions about the similarity of the environment in which the individuals are raised. The workhorses of this approach are twin studies and adoption studies. For example, under the assumption that twins raised together experience the same environment, then any greater similarity in behaviour in identical or monozygotic (MZ) twins over fraternal or dizygotic (DZ) twins must come from their greater genetic similarity. Large sample sizes are typically required to achieve the requisite statistical power in behavioural genetic analyses, hence the need for population studies, sometimes involving thousands of twin pairs. Behavioural genetics has now offered a plethora of findings to the researcher in cognitive development, of which there is space here to mention only a few (see Plomin, DeFries, McClearn & McGuffin, 2008).

Most well known is the finding that genes exert a large influence on individual differences in behaviour, often accounting for 50% of the variance. Moreover, heritability often increases with age. Environmental effects tend to be unique to individuals even if

objectively the environment seems similar (e.g., being raised in the same family). This has led to the view that genes operate to make individuals similar while environments operate to make them different (Plomin et al., 2008). Analytical techniques permit researchers to investigate whether the genes that account for very low or very high performance are the same as those accounting for variation in the normal range. For cognitive abilities, the answer appears to be yes, leading to the claim that 'the abnormal is normal'; that is, that in many cases disorders are the extremes of normal population variability (e.g., Kovas, Haworth, Dale, & Plomin, 2007).

Multivariate methods permit researchers to assess whether the variation in two different abilities is caused by the same or different genes. For the normal population, this approach has led to the claim that the general factor of intelligence (*g*) accounts for nearly all of the genetic variance of diverse psychometric cognitive tests (Plomin & Spinath, 2002). The finding that relatedness mainly contributes to the similarity in children's performance across different abilities, while the environment contributes to differences in performance, has led to claims that genes are generalists and environments are specialists (Kovas et al., 2007). Multivariate analyses can also be used on the same measure at different points in time: are the same genes responsible for variability in behaviour at time 1 as time 2? Longitudinal analyses of children's cognitive abilities suggest that age-to-age stability is primarily mediated genetically, while the environment contributes to change from age to age (e.g., Kovas et al., 2007). Lastly, multivariate methods can be used to explore the genetic unity of the behavioural characteristics of developmental disorders such as autism (e.g., Ronald, Happé, Price, Baron-Cohen & Plomin, 2006).

Evidence of heritability implies that variation in DNA *causes* behavioural variation. The next step is to find the DNA sequences in order to understand the mechanisms by which genes affect behaviour (Plomin et al., 2008). In many cases, the behaviours we are studying are not all-or-nothing traits but vary quantitatively, e.g., performance on a reasoning task or a language task or a visuospatial task. Research has focused on genes that contribute to this variation, referred to as Quantitative Trait Loci. It has indicated that the heritability observed in quantitative traits typically arises from the contribution of many genes, each contributing a small amount to the variability in behaviour. In the normal population, single gene variants of large effect size have proved hard to find for complex cognitive abilities. Association analyses, which test the possibility that a particular gene variant is responsible for some portion of the behavioural variation, have found candidate variants that increase risk for disorders such as schizophrenia, autism, dyslexia, speech and language disorders, and attention deficit hyperactivity disorder (see, e.g., Plomin et al., 2008; Smith, 2007; Posthuma & de Geus, 2006, for reviews). Gene variants have also been found that contribute to individual differences in the normal population, for example in general cognitive ability (e.g., Harlaar et al., 2005). Moreover, in some cases, it has been found that gene variants that contribute to disorder risk also predict variance in the normal population (e.g., attention processes in normal children vs. those diagnosed with attention deficit hyperactivity disorder; Schmidt et al., 2001). However, since the contribution of individual gene variants to predicting behaviour is usually so small in association analyses, even with large populations, there are many false alarms and failures to replicate across different samples (Posthuma & de Geus, 2006).

For current purposes, two points are worth noting regarding molecular genetic findings on cognitive ability. First, where the contribution of individual genes to cognitive variability has been uncovered in the normal population, the genes appear to relate to general neurocomputational properties. For example, two genes whose variants have been much studied (COMT: catechol-*O*-methyl-transferase, and BDNF: brain-derived neurotrophic factor) have basic neural functions and it is argued that their effects in the brain are likely to be widespread in terms of structure and function (Kovas & Plomin, 2006; Plomin & Kovas, 2005; though see Marcus & Rabagliati, 2006, for speculations on specificity). Second, although behavioural genetics has led to a focus on genetics as a cause of behavioural similarities across individuals, the methodology also stresses the importance of the environment. The environment frequently accounts for half of the variability in behaviour. However, because children are usually raised by relations, genetic and environmental effects are often conflated. Behavioural genetic methods allow environmental causes to be isolated by controlling for genetic influences, for example by assessing the extent to which MZ twins diverge in their behaviours (e.g., Caspi et al., 2004).

Three current controversies

The challenge facing researchers in cognitive development is to integrate these behavioural and molecular genetic findings into their understanding of the mechanisms driving developmental change. However, several controversies make this integration more complicated. First, the greater focus on genetic factors (driven in part by their amenability to study via the methods of modern biology) has masked our limited

understanding of the environmental causes of cognitive variation. Thus Plomin et al. (2008, p. 303) recently commented: 'we know much more about genes than we do about the environment . . . where in the brain are environmental influences expressed, how do they change in development, and how do they cause individual differences in behaviour?' Second, there has been disagreement regarding the extent to which partitioning behavioural variation into separate genetic and environmental contributions moves us closer to understanding the mechanisms by which development occurs. This is because developmental processes will always involve on-going interactions between genetic and environmental factors. Third, it is still a challenge to identify what the involvement of genes (either in the guise of relatedness or particular gene variants / mutations) tells us about the mechanisms of development at the much higher level of the whole organism, that is, in terms of the behaviours over which the patterns of heritability are observed (e.g., language, attention, reasoning, and so forth).

Let us briefly consider these points in turn. Beginning with the first, some researchers have argued that despite plentiful evidence for unique environmental effects on individual variability, it has proved difficult to identify the actual sources of this variation, even when researchers have deliberately set out to do so (Turkheimer & Waldron, 2000). This extends to animal studies under laboratory conditions, where all effects of the environment can in principle be controlled or measured (see Kan et al., in press, for discussion). Indeed, Kan et al. have proposed that the statistical phenomenon of unique environmental variation (e.g., the extent to which the behaviour of MZ twins differs) may be intrinsically hard to access at a mechanistic level because it results from non-linear properties of brain development. Under this view, small differences between

individuals in the starting conditions lead to ever greater differences across developmental time. Other researchers have resisted this conclusion as premature. For example, Plomin, Ashbury and Dunn (2001) argued that together, a range of measures of differential environments in children's development could explain 13% of the variance in adjustment, personality and cognitive outcomes (Plomin et al., 2008). These measures included family constellation, differential parenting behaviour, sibling interaction, and peer or teacher interaction. Nevertheless, even where such predictors pick up variance, these are correlation data; it is a yet a further step to identify the relevant causal mechanisms and pathways. For instance, socio-economic status is known to explain variance in cognitive and language development (e.g., Hart, Petrill, Deckard & Thompson, 2007; Petrill, Pike, Price & Plomin, 2004), and even to predict differences in functional hemispheric specialisation (Raizada, Richards, Meltzoff & Kuhl, 2008). Yet as Stevens, Lauinger and Neville (in press) argue, socio-economic status indexes a range of confounded factors, among them differences in prenatal care, nutrition, depression, neglect, cortisol levels, money, quality of schooling, perception of inequality, parental attitudes, parental education, social support, and availability of books in the home. It is not clear by which causal route socio-economic status effects operate, or the relative weighting across the routes if (as is likely) several routes contribute. Yet for the developmental psychologist, the causal mechanisms are the key issue.

The second controversy concerns the utility of the variance partitioning methods of behavioural genetics, where genes and environment are modelled as main effects predicting behavioural variation. Researchers from the *Developmentalist* school (e.g., Gottlieb, 1995) argue that biological characteristics emerge from a matrix of

developmental interactions between biological and environmental factors. Partitioning variance and assigning greater or lesser weight to genetic or environmental causes does not necessarily shed light on the mechanistic basis of these interactions. Thus Turkheimer (2004, p.165) commented: 'it is still an open question whether multivariate behaviour genetics and molecular behavioural genetics will . . . [succeed in] . . . providing a quasi-experimental bridge between population-based variance partitioning and causally specified developmental models'. In the context of cognitive development, it is worth noting that behavioural genetic methods focus on individual differences and are not sensitive to species-universal properties (e.g., behaviours that show no variation; genes or environments that are shared by all members of the species). The window that individual differences provide on processes of cognitive development ultimately depends on the mechanistic basis of the relationship between these two sources of variation (within age and across age, respectively). Empirical work has yet to resolve the relationship and there are relatively few theoretical frameworks that incorporate both types of variation (see, e.g., Baughman, 2009; Davis & Anderson, 1999; Thomas & Karmiloff-Smith, 2003a).

The third point of controversy is that, even if we accept the importance of specifying causal mechanisms, the levels of gene and behaviour are remote (see, e.g., Fisher, 2006; Johnston & Lickliter, 2009). One pertains to the expression of individual proteins within a single cell, the other pertains to the behaviour of the whole organism embedded in a physical and social context. Taking the findings of gene-behaviour association studies at face value, thus far we are gaining only hints about the nature of the causal pathways between these levels of description. Here are four examples: genetic variations appear to modulate attention skills via a pathway that alters the efficiency of

dopamine receptors in the fronto-striatal systems that deliver behavioural control (Posner, Rothbart & Sheese, 2007); the disorders Rett syndrome and Fragile X have been argued to represent different causal pathways by which the production of proteins necessary for synaptic plasticity is demodulated (Kelleher & Bear, 2008); developmental language impairment and autism have both been linked to a gene variant that alters production of a protein sitting in the membranes of neurons, which influences interactions between different cells during the development and wiring up of the nervous system (Vernes et al., 2008); developmental dyslexia has been linked to four gene variants associated with neuronal cell adhesion, perhaps pointing towards regional disruptions of neural migration and axonal guidance in early brain development (Galaburda et al., 2006).

These recent findings are exciting and highly suggestive, but of course, further understanding is still required. First, we do not yet have an understanding of how the brain circuits deliver the high-level behaviours measured by cognitive tasks, sufficient to be confident about the consequences of the low-level differences in synapse formation or neural migration. Second, the properties picked out by the gene variants often appear too general for the more specific behavioural outcomes that they are intended to explain. And third, the developmental context of the causal account is often missing. It is unclear how differences in, say, dopaminergic signalling or synaptic plasticity or neuronal migration would lead to particular high-level behavioural consequences, because these low-level differences represent alterations to an adaptive system that must follow a sometimes protracted pathway of development before the whole organism can exhibit the target behaviour which led to the identification of the gene-behaviour association in the first place.

Nevertheless, attempts to close the explanatory gap between genes and behaviour have stimulated the emergence of a range of new fields. Plomin and colleagues have outlined the bottom-up approach of *functional genomics* beginning at the level of cells and molecular biology (e.g., Plomin, 1999). *Computational neurogenetics* follows the bottom-up approach and explores the integrations of dynamic neuronal models with gene models (Kasabov & Benuskova, 2004). *Behavioural neurogenetics* is similarly interested in elucidating the role of genes in brain development, and in the emergence of species-specific behaviour (e.g., Baumgardner, Green & Reiss, 1994). *Behavioural genomics* takes a top-down approach, attempting to understand how genes work at the level of the behaviour of the whole organism, spanning behaviour, psychology and neuroscience (e.g., Plomin & Crabbe, 2000). *Cognitive genetics* aims to link three levels of description, cognitive, neural systems, and cellular levels (Scerif & Karmiloff-Smith, 2005). Both bottom-up and top-down approaches face challenges. The challenge for the bottom-up approach is the sheer complexity of the molecular processes involved. As Plomin et al. (2008) point out, each synapse is affected by more than a thousand protein components; and a synapse is a long way from understanding a functional neural circuit. The challenge for the top-down approach is that researchers in cognitive neuroscience still only have a sketchy view of how the brain implements cognition. The top-down approach must move below the cognitive level to construe developmental theories in terms relevant to the influence of neurocomputational parameters, whilst retaining contact with the overt behaviours expressed by individuals. This is the aim of the current paper.

A top-down approach to closing the explanatory gap: computational models of development

In this paper, we pursue a top-down strategy to closing the explanatory gap between genes and behaviour. The strategy involves the use of implemented computational models that explicitly simulate the interaction between genetic and environmental factors during the development of high-level behaviours; and it involves simulating variability at the population level, thereby addressing development and individual differences within a single explanatory framework. The next sections lay some groundwork, before introducing the specific model and the target empirical data to which it will be applied. We briefly outline how computational models have been used to explore cognitive development over the last twenty years; we show how more recently these models have been extended to consider cognitive variation through the manipulation of neurocomputational parameters; and we show how computational models of high-level behaviours can be linked to a genomic level of description via the use of Genetic Algorithms.

Connectionist models of development

Over the past two decades, computational models that exploit machine learning have proved useful in developmental psychology because they permit formal specification of transition mechanisms. These mechanisms demonstrate how a child can move between producing different patterns of behaviour at different ages through interaction with a structured training environment. One of the most influential approaches, *connectionist modelling*, has demonstrated how the structure of the learning environment can interact

with the system's internal constraints to shape the developmental trajectories in behaviour. In this context, connectionist or artificial neural network models embody constraints drawn from neurocomputation while targeting high-level cognitive behaviours. Connectionist models have been applied to a range of phenomena including infant category development, language acquisition, and reasoning in children (Elman et al., 1996; see Thomas & McClelland, 2008; Mareschal & Thomas, 2007, for reviews; Spencer, Thomas, & McClelland, 2009, for comparisons to related dynamic systems approaches). However, as in developmental theory, such models have tended to characterise the 'average' child. Where variability in development does occur, for example through initial randomisation of the starting weights in the artificial neural networks or chance presentation orders of the items in the training set, the results are averaged out over multiple runs.

Neurocomputational parameters as mechanisms of variability

Computational models of development are well placed to explore the mechanistic basis of individual differences. Models contain parameters, such as the learning rate or level of resources, which affect both the model's efficiency to learn and the complexity of the representations it can acquire. Moreover, changes to the way in which the problem domain is presented to the network (i.e., its input and output representations, or the constitution of the training set) can alter its performance compared to the metric of normality (Thomas et al., submitted). In this way, connectionist models have been extended to explore potential causes of atypical development, for example in disorders such as autism, dyslexia, Specific Language Impairment (SLI), Williams syndrome, and

attention deficit hyperactivity disorder (e.g., Harm & Seidenberg, 1999; Joanisse, 2004; Lewis & Elman, 2008; Richardson & Thomas, 2006; Thomas & Karmiloff-Smith, 2003b; Triesch et al., 2006). In these cases, a single neurocomputational parameter is usually altered in the startstate of the model, based on psychological or neuroscientific data on the disorder, with the intention of demonstrating that the normal developmental trajectory of behaviour can then be deflected to exhibit the deficits observed in the disorder. Far fewer proposals have been made regarding methods to simulate individual differences (Garlick, 2002; Thomas et al., submitted; see Thomas & Karmiloff-Smith, 2003a, for discussion). The limited amount of population modelling of cognitive development that exists has been restricted to mathematical models or dynamic systems modelling, where no actual acquisition of a cognitive ability is simulated (e.g., Anderson & Nelson, 2005; Turkheimer, 2004; van der Maas et al., 2006).

Genetic connectionism: genomes to encode variation

In order to bring together computational models of development with studies of population variability that consider genetic and environmental influences, it is necessary to take three steps: (i) encode some of the computational properties of a connectionist network in a genome; (ii) allow variation in the genome between 'individuals', which then produces variations in computational properties; and (iii) allow variation in the structured training environment to which each system is exposed.

The idea of encoding the properties of a computer program in the form of a genome is a familiar one. *Genetic Algorithms* are a method of optimising computer programs by breeding generations of programs and selecting the 'fittest' (according to

performance on the target problem) to populate the next generation. Genetic algorithms constitute a valuable tool for optimisation in machine learning because they permit an efficient search of complex parameter spaces, where the settings of a large number of variables serve to optimise the performance of a computer program on a given problem. In principle, genetic algorithms can be applied to any computer program. The minimal requirement is that the parameter settings for the program must be encodable in a genome, and every version of the genome created by mechanisms that induce genetic variability (such as breeding) must correspond to a legal computer program. Genetic variability operators, inspired by sexual reproduction, include 'cross-over', where the genomes for two computer programs are combined in some way to produce two new genomes, and 'mutation', where elements of a genome are changed at random. Genetic algorithms typically proceed by generating a large initial population of genomes with random variability. These are translated into the equivalent computer programs, which are then exposed to the problem domain. A fitness measure rank orders performance on the problem domain. Poor performers are discarded, and good performers are copied. Operators of genetic variability are then applied to create the individuals of the next generation. The process is iterated and usually results in improvement in the average performance of the population over generations (see Mitchell, 1997, for introduction).

From the beginning of the field's recent incarnation, researchers in connectionism have seen the value of integrating their network models with genetic algorithms, leading one author to refer to the discipline of 'genetic connectionism' (Chalmers, 1990). Applied to artificial neural networks, genetic algorithms could in principle be used in three ways. First, they could be used as a type of training algorithm, with the aim of find the optimal

connection strengths for a network on a given problem. In this case, connection weight values would be specified by the genome and no learning would take place. Second, the genome could merely determine the initial parameter values of a network, and each network could then undergo a phase of learning on the target problem before the fitness metric was applied. Third, these two methods could be combined, so that some parameters would be fixed by the genome whilst others would be allowed to vary.

All three approaches have been pursued. For example, research has demonstrated the value of having fixed weights (phenotypic rigidity) versus permitting learning (phenotypic plasticity). One study indicated that in the presence of a stable environment, plasticity is advantageous for early generations, after which phenotypic rigidity becomes increasingly important (Hinton & Nowlan, 1987). Research has also explored the synergy between evolution and learning, where evolution can produce systems that are better placed to learn, known as the Baldwin effect (e.g., French & Messinger, 1994; Nolfi, Elman & Parisi, 1994). Simulation work has also pointed to the computational intractability of Lamarkian evolution, the proposal that learned behaviours are passed on to next generation via modifications to an individual's DNA during its lifetime. In complex systems, it is immensely difficult to solve the *reverse problem* of isolating the genes that are responsible for a given learned behaviour (Turney, 1996). Of course, this is the same problem that is targeted by gene-behaviour association studies.

Most of these computational investigations used highly simplified problem domains. However, Nakissa and Plunkett (1998) demonstrated how genetic algorithms could be used to evolve a self-organising learning system that rapidly acquired categorical representations for speech perception. The genome encoded both the

architecture and learning rules of the artificial neural network, and each network was trained by exposure to speech spectra from different human languages. Other research has used the marriage of artificial neural network and genetic algorithm approaches to explore the advantages of modular architecture and of age-dependent plasticity for cognitive development (Bullinaria, 2005, 2007, 2009).

The computational studies we present in this article differ from this previous work in that, while we use genetic algorithms to encode some of the computational properties of the learning systems, *we do not use selection*.[1] Selection is a key feature of optimisation. However, when genetic algorithms are used for optimisation, the consequence is to reduce population variability over subsequent generations, because only the better performing individuals are retained (see Appendix A). Variability will only persist across generations if there is some exact balancing mechanism that serves to refresh it or if the metric for selection is itself non-stationary across generations. By contrast, the starting point for the current investigation is that *behavioural variability is the empirical given*. Therefore, although we will employ the familiar mechanisms of genetic algorithms, in the absence of selection we will not expect to observe substantial change in the population variability across generations. In other words, in what follows, we will not consider (other than tangentially) the evolutionary origins of the genetic variation; rather, our focus is the mechanisms that explain the observed behavioural variability in a population of learning systems embedded in a given environment.

[1] Within the wider modelling framework, our assumption is not that selection is not operating at all. It is that selection is not operating with respect to the behaviours generated by the system we are simulating.

Mechanisms of environmental variation

While Genetic Algorithms provide a means to encode neurocomputational dimensions of variation within a computational implementation of development, we must separately consider how to implement variation in the quality of the environment. Psychological evidence on the role of the environment on cognitive development relies on circumstances where the quality of the environment varies. Three major sources of evidence are the role of socio-economic status (SES), the role of deprivation, and the effectiveness of enrichment programmes. We focused on the first of these literatures to motivate the assumptions of our model.

A great deal of research has accumulated on the role of SES in influencing both cognitive development and neurocognitive functioning. Reviewing recent findings, Hackman and Farah (2009) noted how the effect of SES is uneven across cognitive domains, affecting language development and executive functioning more than memory, spatial cognition, and visual cognition. Moreover, effects on neural processes (e.g., as measured by event-related potentials) can be observed even when there are no overt differences in behaviour (e.g., D'Angiulli et al., 2008). In terms of language development, Farah et al. (2006) created a language composite score of performance on vocabulary and phonological processing tasks in first graders (5-6 year olds), and were able to predict 32% of the behavioural variance using SES. Other reported effect sizes are more modest. Petrill et al. (2004) tested expressive vocabulary and grammatical complexity via parental report in a sample of 6000-8000 twins. SES predicted 3.2% of the variance at age 3 and 3.6% at age 4. Hart et al. (2007) tested 287 pairs of twins of

elementary school age on a test of general cognitive ability at two time points two years apart. SES predicted 5.8% of the variance at time 1 and 2.9% of the variance at time 2.

There are two main difficulties with employing SES as a marker for environmental variation. First, the exact definition of SES is controversial, typically involving factors such as parental education, income, and occupation (Hackman & Farah, 2009). Second, the causal mechanisms by which it exerts its effects remain to be elucidated. In their review, Hackman and Farah (2009) listed the following causal factors confounded with SES that might act on brain development: lead exposure, cognitive stimulation, nutrition, parenting styles, transient or chronic social hierarchy effects, and chronic stress. In a review of the effects of SES on language development in particular, Ginsborg (2006) considered the SES-related factors of parental education, home environment, the relationship between principal caregiver and child, the language environment experienced by the child and the nature of the interaction between mother and child (including the quantity of the speech addressed to the child and the nature of child-directed speech). Moreover, several causal pathways may operate simultaneously, and their respective contributions may differ at different stages in development.

In constructing a computational model, we must commit to a particular implementation of environmental variation. We considered two options. First, the environment may influence the computational properties of the learning system, perhaps via SES-related influences such as nutrition and stress. We did not pursue this option in the first instance (though see Thomas, Forrester & Ronald, in prep.). Second, SES may operate by influencing the nature of the information available in the environment. This might occur in three ways. It might alter the *quality* of the information available in the

environment. It might alter the *quantity* of the information in the environment. Or it might influence the motivation of the child to engage with the information available, perhaps through differences in reward and encouragement (effectively, the subjective information that the child actually exploits). We turned to the literature to evaluate which of these options was the case.

Although few in number, large-scale longitudinal observation studies of the development of spoken language in young children have addressed the role of SES. For example, in a landmark study, Hart and Risley (1995) followed children in 13 professional/managerial families, 23 working class families, and 6 families living on benefits, between the ages of 8 months and 3 years, recording all language produced by the child or available around the child for 1 hour per month. The most salient difference at different SES levels was not the quality of the language spoken to the child but the quantity. Professional families addressed 2100 words to their child in the average hour compared to 600 in the welfare families. Moreover, higher SES was associated with a greater incidence of affirmative feedback and lower incidence of prohibitions. Quantity and motivation therefore appeared to be most salient differences. This fits with the view that language acquisition is a 'data-crunching process and conversation is a delivery mechanism whose value lies . . . in the nature of the data that it delivers' (Hoff & Naigles, 2002, p.422). Notably, Huttenlocher, Vasilyeva, Cymerman and Levine (2002) found that SES did not predict the frequency of complex speech produced by children once parental input and teacher input were controlled for (see Ginsborg, 2006).

The study of children raised by their biological parents permits the possibility that correlations between parental and child language are mediated genetically.

Environmental variation could therefore play a weaker role than the correlations imply. However, research has explored variation in the school environment, specifically correlating gains in language development with the quality of teacher language over the school year. Data have tended to support the view the environmental variation in language indeed plays a causal role in development (Huttenlocher et al., 2002; Klibanoff et al., 2006; Vasilyeva et al., 2006). For example, the proportion of complex sentences produced by the teacher predicted 18% of the variance in the improvement in performance on a syntax comprehension task over the school year (Huttenlocher et al., 2002).

In sum, our review of the literature led us to the view that, as an initial assumption, it was reasonable to implement the effects of SES on development in terms of variations in the information content of the environment.

Target empirical domain: Acquisition of the English past tense

The principal aim of the current study is to explore the viability of linking three levels of description (genome, neurocomputational, overt behaviour) within an implemented developmental model. This is best achieved with reference to a concrete set of empirical data, and in the context of a body of research on what a plausible 'average' model of development might look like. Children's acquisition of the English past tense fulfils both of these criteria.

Children's acquisition of the English past tense, along with other aspects of inflectional morphology, has been the focus of a great deal of empirical research. This is because of the dual nature of the domain, with its split between regular and irregular

inflection. Verbs may form their past tenses following a rule, or they may be exceptions to that rule. Theorists have debated whether separate cognitive mechanisms may be required to acquire the two verb types, and whether this indeed constitutes an innate component of the language system (e.g., Marcus et al., 1992; Pinker, 1999; Rumelhart & McClelland, 1986). For our purposes, past tense is useful because research into the domain includes a significant body of work using computational models of development, particularly employing connectionist networks, starting with the pioneering work of Rumelhart and McClelland published in 1986 (see Thomas & McClelland, 2008, for recent review). Past tense abilities, and underlying cognitive processes, have been assessed in multiple human populations using multiple methods. These included the trajectory of acquisition in typically developing children (e.g., Marcus et al., 1992), differences in development due to gender (e.g., Hartshorne & Ullman, 2006), performance in children with behaviourally developmental disorders such as SLI and dyslexia (e.g., Joanisse et al., 2000; Rice, Wexler & Hershberger, 1998; van der Lely & Ullman, 2001), and in children with genetic disorders such as Williams syndrome and Down syndrome (e.g., Ring & Clahsen, 2005; Thomas et al., 2001). Methods have extended to the use of genetically sensitive twin study designs (e.g., Bishop, 2005) Research on past tense has also addressed variation in the quality of the environment to which children are exposed during acquisition (e.g., Hart & Risley, 1995). Finally, connectionist computational models have been applied to simulating the types of variability observed in developmental and acquired disorders (e.g., Joanisse & Seidenberg, 1999; Joanisse, 2000; Hoeffner & McClelland, 1993; Marchman, 1993; Thomas, 2008; Thomas & Karmiloff-Smith, 2003b; Westermann, 1998).

The English past tense possesses an interesting structure as a problem domain because it is characterised by a predominant rule (e.g., *talk-talked*, *drop-dropped*, etc.) that extends to novel stems (e.g., *wug-wugged*), but also contains irregular verbs of different types (*go-went*, *hit-hit*, *sing-sang*). Because of this structure, the past tense affords an examination of the effects of problem type on population variability in development, and in particular the effects of consistency, type frequency, and token frequency within the learning environment. These features of problem domains are argued to influence many aspects of cognitive development (e.g., Bates & MacWhinney, 1987, for wider arguments in language development). In particular, the regular past tense has the highest type frequency and forms a consistent set of mappings (reproduce the stem, add an inflection). The different classes of exception verb fall on a continuum of inconsistency, with no-change past tenses (*hit-hit*) least inconsistent (reproduce the stem but don't add an inflection), vowel change past tenses (*hide-hid*) at an intermediate level (partly reproduce the stem, no inflection), and arbitrary past tenses (*go-went*) most inconsistent (no relation between stem and past tense). Arbitrary past tenses have the lowest type frequency but require the highest token frequency in order to be acquired.

A limited amount of research has considered how environmental factors influence the acquisition of the English past tense. Rice, Wexler and Hershberger (1998) examined the associated between SES (as measured by maternal education) and the acquisition of regular past tense formation in 21 children with SLI and 20 typically developing children followed longitudinally over a 3-year period. In contrast to vocabulary and phonological awareness (Farah et al., 2006), Rice and colleagues reported that maternal education did *not* reliably predict variation in past tense development, explaining less than 1% of the

variance. Hart and Risley (1995) nevertheless reported large differences in the number of past tense verb forms addressed to the children, with professional families using 3 times as many as those on welfare. They noted there was little difference across SES in the richness of the past tense verbs used, where richness was defined as the number of past tenses per utterance. There is a question, then, of whether for the case of past tense, variations in the environment are associated with differences in rates of acquisition.

Together, the past tense problem presents a wealth of data to consider variability from at least three sources: differences in internal computational constraints, be they developmental or acquired; differences in behaviour due to the structure of the problem domain itself (e.g., the different influences of type frequency, token frequency, and similarity that distinguish regular verbs from irregular verbs); and differences in the quality of the language environment to which the child is exposed. Our target empirical data for the following simulations comprised performance on a past tense elicitation task of a population of 442 6-year-old children, for whom SES information was available (Bishop, 2005; Petrill et al., 2004). These data were originally collected as part of a twin study (Bishop, 2005).[2]

Research questions

In moving beyond using a computer model to simulate the average child, we initially simulated one thousand individual trajectories of development in past tense formation in each of four populations that differed in the relative influence of genetic and environmental factors. We then ask three broad questions of these simulated data. (1)

[2] We are very grateful to Dorothy Bishop for making these data available to us.

What is the nature of the population distributions of performance and how do distributions change across development? (2) Given that we have stipulated all sources of variability at a mechanistic level, how good are our various measures (genotypic, neurocomputational, environmental) in predicting population-level variability behaviour, either in the normal range or at the extremes? (3) In addition to these main effects, what interactions emerge between specified factors, for example, in the form of gene-environment interactions? Where possible, simulation results were compared to the empirical data. Table 1 includes a set of more specific research questions. (The reader is invited to predict the answers to each of these questions in advance, because one role of modelling, if successful, is to make the answers to some questions appear obvious in retrospect.)

==========

Insert Table 1 about here

==========

Method

Empirical methods

The English past tense data are taken from a study by Bishop (2005) on the performance of a sample of 442 6-year-old children, predominantly pairs of twins, on a sub-test of a language battery designed to assess risk of heritable language impairment (Rice & Wexler, 2001). The data represent accuracy levels in producing the inflected forms for regular and irregular English past tense verbs. In the Past Tense probe subtest, children are shown a picture of an action being carried out, while the experimenter describes the picture using a sentence containing the target verb in the present tense. The child is then shown a second picture with the action completed. The child is encouraged to describe the second picture. The task is designed to elicit the past tense in a full sentence with an overt subject. The test contains 11 regular verbs (*brush*, *clean*, *climb*, *colour*, *jump*, *kick*, *paint*, *pick*, *plant*, *play*, and *wash*) and 8 irregular verbs (*catch*, *dig*, *fall*, *make*, *ride*, *swim*, *throw*, and *write*). The test/retest reliability for the Past Tense probe subtest is r = .82 (Rice & Wexler, 2001).

Participant details: the population contained 442 children with mean age 6 years and 6 months (range: 6 years and 0 months to 7 years and 1 month). There were 250 boys and 192 girls. The children were originally recruited as twin pairs and in this sample, 224 were MZ twin pairs and 218 were DZ twin pairs. Based on a parental questionnaire carried out when the children were aged four, 215 of the children had been flagged as at risk for language impairment. The questionnaire combined responses indicating whether the children were talking in full sentences, whether they had low vocabulary, and whether the parents were worried that the child's language was developing slowly (see Bishop,

2005, for details); 10.6% of the original full sample of 5426 children had been so flagged. Children were excluded from the study if English was not the only language spoken at home and, for reasons of planned future molecular genetic studies, non-white families were also excluded to reduce the risk of ethnic stratification (Bishop, 2005). At six years of age, 304 of the 442 children were viewed as normal (i.e., had not been diagnosed with any disorder), implying that markers of risk had disappeared in about a third of the original at-risk children. Of the children subsequently diagnosed with disorders, 96 (21.7%) had been diagnosed with SLI, 7 (1.6%) had been diagnosed with autism, and 35 (7.9%) had been diagnosed with general delay.

A measure of socioeconomic status (SES) was also available for the families of each of these children (Petrill et al., 2004). Demographic information was obtained via questionnaire from the first contact with the family. Five pieces of information were collected: the father's highest educational level and occupational status, the mother's highest educational level and occupational status, and the age of mother at birth of the eldest child. From these data, an index of SES was created based on a factor analysis. Petrill et al. (2004, p.448) report the following method. Using principal components analysis, a single-factor solution yielded an eigenvalue of 2.51, accounting for 50% of the variance. Based on these results, a single composite was created. The five variables were standardized and then summed using unit weights. This method yielded a scale ranging from -1.57 (low SES) to +1.54 (high SES), with a mean of -0.16 and a standard deviation of 0.72. Of the 442 children, SES data were missing for 25 (11%). For these children, the missing data were replaced by the population mean.

Simulation methods

The base computational model of development

For our base model in the following simulations, we adopted a model proposed by Plunkett and Marchman (1991, 1993, 1996). These authors were part of a school of thought that argued that both the productive regular verbs and the exception verbs can be acquired by an otherwise undifferentiated connectionist network that learns to associate a vector-based representation of the phonological form of each verb stem to a similar representation of its past tense form. Our choice of base model is in part a practical one, because it is a relatively simple model and therefore tractable for large-scale population simulations.[3] More complex models of inflectional morphology have since been proposed, including models that embed past tense in a more general system responsible for inflecting all content words and involving the integration of lexical-semantic, phonological, and grammatical information (e.g., Karminis & Thomas, 2009). Some researchers maintain that symbolic approaches are more appropriate for explaining past tense acquisition, at least for regular verbs, where regularity is viewed as reflecting the operation of a rule-based mechanism (Pinker, 1999). However, in the main, these more linguistically oriented theories have not been realised in computational implementations of the developmental process. The Plunkett and Marchman model captures a number of

[3] In this paper, we report on results of 9 populations of 1000 individual networks. The performance of each network was measured 1000 times across development, on a training set of 500 items and a test set of a further 500 items. Simulation of a single network took a little over two hours on a standard PC running Linux, translating to around 95 days of processing time per population. Substantial gains in efficiency are likely to be available by the optimisation of code and use of parallel processing on machines with multiple chips.

basic empirical effects in child past tense acquisition, including differential performance depending on the type and token frequency of verbs, as well as error patterns across development (Plunkett & Marchman, 1991, 1993, 1996).

Simulation details

Training set

For our training set, we used the "phone" vocabulary from Plunkett and Marchman (1991, p. 70). This comprised an artificial language set constructed to reflect the most important structural features of English past tense formation. There were 508 monosyllabic verbs, constructed using consonant-vowel templates and the phoneme set of English. Phonemes were represented over 19 binary articulatory features, a distributed encoding based on standard linguistic categorisations (Fromkin & Rodman, 1988). Separate banks of units were used to represent the initial, middle, and final phonemes of each monosyllable. The output layer incorporated an additional 5 features to represent the affix for regular verbs. Networks thus had 3x19=57 input units and 3x19+5=62 output units. There were four types of verbs in the training set: (1) regular verbs that formed their past tense by adding one of the three allomorphs of the +ed rule, conditioned by the final phoneme of the verb stem (e.g., *tame-tamed*, *wrap-wrapped*, *chat-chatted*); (2) exception verbs whose past tense form was identical to the verb stem (e.g., *hit-hit*); (3) exception verbs that formed their past tenses by changing an internal vowel (e.g., *hide-hid*); (4) exception verbs whose past tense form bore no relation to its verb stem (e.g., *go-went*). The token frequency of this last type of exception verb had to be higher for the network to learn them successfully (see Plunkett & Marchman 1991), as is the case in

real languages. As a result, this verb type experienced three times as much training as the other verb types. The *type frequencies* were as follows. There were 410 regular verbs, and 20, 68, and 2, respectively, of each exception verb type. We added 8 additional arbitrary past tenses to allow for finer graduations of performance; although this slightly deviates from structure of English, it permits better assessment of role of high token frequency for items that must be rote learned. Following Plunkett and Marchman (1991), the verbs were given a token frequency structure. For computational convenience, *token frequency* was implemented by mediating the weight change generated by the difference between the actual output and the target output (Plaut et al., 1996). The weight change of high frequency arbitrary verbs was multiplied by 0.9 during a given training presentation and that of low frequency arbitrary verbs by 0.6. The weight change of all other high frequency verbs (regulars, no change, and vowel change) was multiplied by 0.3 and of all other low frequency verbs by 0.1. Finally, a separate set of novel verbs was constructed to evaluate the generalisation performance of the network. These verbs could differ depending on their similarity to items in the training set. For simplicity, 410 novel verbs were used, each of which shared two phonemes with one of the regular verbs in the training set. Generalisation was evaluated depending on the proportion of these novel verbs that were assigned the correct allomorph of the regular past tense rule.

Model architecture and parameters

The original connectionist model employed a three-layer artificial neural network, comprising an input layer, a layer of internal or 'hidden' units, and an output layer. It was trained using the backpropagation algorithm (Rumelhart, Hinton, & Williams, 1986), a

type of supervised learning. The free parameters in the model were the number of hidden units, the learning rate, and the momentum (see below). An expanded set of 14 parameters was employed in the current simulations, in many cases for additional analogues to known neurocomputational properties. However, backpropagation itself is not viewed as biologically plausible. We use it here in place of a more biologically plausible error correction algorithm for simplicity of implementation (see Thomas & McClelland, 2008, for discussion). The parameters and model architecture are depicted schematically in Figure 1. The parameters were as follows:

Building the network:

- *Architecture*: In addition to the 3-layer network, a 2-layer network without a layer of hidden units, and a fully connected network were used. A 2-layer network has less computational power. A fully connected network contains both direct connections from input to output and a hidden layer, and produces a computationally more powerful system.
- *Hidden units*: For networks with a hidden layer, the number of hidden units could vary.
- *Sparseness*: Of the potential connections between processing units in a given architecture, only a certain proportion were created. The sparseness parameter set the probability that any given connection would be created.
- *Weight variance*: Connection weights were assigned an initial random value within a range depending on this parameter. E.g., if set to 0.5, weights would be randomised between +/- 0.5.

Processing dynamics:

- *Processing noise*: The net activation a receiving unit receives from a given sending unit is a product of the sending unit's activation and the connection strength between them. Transmission noise was added to this net activation. Gaussian noise was used and the parameter specified the standard deviation of the noise distribution around zero.
- *Unit threshold function*: A receiving unit sums the net activation from all sending units and uses an activation function to determine its consequent output. We used a common activation function, the sigmoid or logistic function, equivalent to a smoothed threshold. This function has a parameter, the temperature, which makes the smoothed threshold either steeper or shallower. A shallow function denies a unit the opportunity to make large output changes in response to small changes in net input, whereas a steep function approximates a threshold thereby producing a unit with binary response characteristics. Either too shallow or too steep a function retards learning (see, e.g., Thomas, 2005).

Network maintenance:

- *Connection weight decay*: each connection's magnitude was reduced by a small proportion on each presentation of a training pattern, according to the weight decay parameter.
- We did not simulate the increase in synaptic density observed in human cortex during infancy and early childhood; we did, however, implemented the pruning of spare resources from mid-childhood (Huttenlocher, 2002). The pruning process

eliminated small connection weights and involved three parameters: onset, threshold, and probability

- *Connection pruning – onset*: Connections that were not being used were probabilistically pruned away after a certain point in training. The onset parameter determined the point in training when pruning began (see Thomas & Johnson, 2006).
- *Connection pruning – threshold*: Connections stood a chance of being pruned after onset only if their magnitude fell below a threshold determined by this parameter. The rationale is that small weights are assumed not to transmit strong activations and therefore not to be playing a key role in computations. They may therefore be removed to save on resources.
- *Connection pruning – probability*: If the magnitude of a connection fell below threshold after pruning has begun, it was eliminated probabilistically based on this parameter. High probability leads to faster loss of unused connections. Low probability leads to slower loss.

Network adaptation:

- *Learning algorithm error measure*: The backpropagation algorithm was used with two different metrics to determine the error signal marking the disparity between the network's current output and its intended target. These were Euclidean distance and cross-entropy (Hinton, 1989). Cross-entropy retains greater plasticity in the network even when it has committed to a solution that proves to be non-optimal.

- *Learning rate*: This parameter determined how much the connection weights were altered in response to a certain disparity between output and target during supervised learning. A large learning rate produces a system that learns more quickly but that also may be unstable, flipping between high performance on different parts of the problem domain.
- *Momentum*: This parameter allowed some proportion of the weight change on the previous learning trial to be carried over. It serves a smoothing function to prevent learning from getting stuck in local, sub-optimal solutions.

Network response:

- *Nearest neighbour threshold*: Network output comprised a vector of continuous activation values between 0 and 1, while legal responses of the network were binary vectors. An algorithm determined which legal phoneme was closest to the activation patterns at onset, nucleus, and coda. However, the phoneme was only recognised as a response if the activation was sufficiently close to the legal phoneme (using a root mean square or RMS measure). This was determined by the nearest neighbour threshold. (The legal phonemes could of course still be the incorrect ones for the target verb). The nearest neighbour computation may be viewed as equivalent to the settling of an unimplemented recurrent attractor network into a particular response state (see Plaut et al., 1996, for a model of reading development in which this attractor network was implemented). The nearest neighbour threshold parameter then indexes the efficiency of this attractor network to generate a response within some notional deadline. A high threshold allows an approximate output to be recognised as correct (i.e., larger error is

tolerated); a low threshold requires a more exact initial output. The use of a nearest neighbour algorithm allowed the network to generate accuracy levels. The disparity between the actual output activations and the target output (in terms of RMS error) has been used in the past as a proxy for reaction time, on the grounds that output activation states closer to the final response would enable faster settling of the notional attractor network (see Bullinaria, 1995).

==========

Insert Figure 1 about here

==========

Calibrating parametric variation

Two of these parameters were categorical: the architecture and learning algorithm metric. The others were continuously valued. In order to produce variability in the population according to these remaining parameters, we calibrated them as follows. An initial 'normal' set of parameters was defined based on previous research. Each of the continuously valued parameters was then varied in turn, holding the all other parameters at their initial values. For each parameter, the range was derived that produced failure of learning up to highly successful learning. In some cases, parameters were monotonic (e.g., hidden units, where more was better); in other cases, there was an optimal intermediate value (e.g., activation function). Our aim was to determine an average or adequate value for each parameter, which we defined heuristically as 'just enough to succeed and then a little bit more'. We then derived values that would cause increasingly poorer or increasingly better performance around this value. We attempted to make

poorer and better performance roughly symmetrical around average performance for each parameter. This caused some parameter ranges to be skewed. For example, 50 hidden units was determined as the average value in a 3-layer network. Values of 40 or 30 would cause poorer performance. However, to achieve equivalent differences above average level, 100 or 200 hidden units might be necessary. We chose to emphasise behavioural symmetry around the average parameter value rather than parametric symmetry, on the grounds that the symmetrical bell curve is a common pattern observed in human abilities. The ranges for each parameter are included in Figure 2.

We did not vary the encoding of the problem domain. This decision was made for practical reasons. Our previous work suggests that, within certain limits, varying the problem encoding has similar effects on the developmental trajectory to altering computational parameters (Thomas & Karmiloff-Smith, 2003b). However, recoding the problem domain can in principle have extreme effects on learnability, if key distinctions in the input or output are lost in the recoding. Some models of developmental language impairment and dyslexia propose that differences in encoding of phonology cause subsequent behavioural deficits in grammar and reading acquisition (e.g., Harm & Seidenberg, 1999; Hoeffner & McClelland, 1993; Joanisse, 2004).

Although only main effects of each parameter were considered as sources of variability during calibration, we fully expected interactions between these neurocomputational parameters in subsequent learning. For example, large numbers of hidden units can partially compensate for a shallow sigmoid function in those processing units. Having a more sparse initial connectivity is likely to reduce the amount of weights eliminated via pruning because their magnitudes will be larger. High weight decay can be

countered by a higher learning rate. An aberrant, over-keen pruning process (e.g., with a low threshold and high probability) can be alleviated if its onset occurs very late in training when weights have become large, but made worse if the onset is early. And so forth. Large numbers of parameter combinations were possible within our scheme: given the number of levels specified for each parameter, around two thousand billion unique parameter combinations were available.

Specifying a genome for the model and the mechanism for breeding

The use of genetic algorithms entails creation of a genome to encode the neural network's parameter values, such that all possible genomes correspond to legal parameter sets. In creating the genome, we made the following assumptions. There were two copies of each gene, with genes residing on pairs of chromosomes. For simplicity, each gene had only two variants or alleles, one which was more efficient for the gene's function, one which was less efficient. The influence of genes was purely *additive*: we did not include dominant or recessive effects, and genes had the same effect in combination as in isolation. This simplification is driven by the finding within behavioural genetics that the effect of genes variants is predominantly additive on phenotypic outcomes (Plomin et al., 2008). All neurocomputational parameters were *polygenic*. That is, their value was determined by the additive action of a collection of genes. However, for simplicity, we assumed that the action of genes was not *pleiotropic*; that is, with respect to neurocomputational parameters, we assumed that no gene affected the value of more than

one parameter at once.[4] The assumption of polygenicity was motivated by the fact we are using computational models to capture cognitive-level phenomena, and is a point worth emphasising. We expect many low-level neural variations to influence neurocomputational functions at the level of cognitive processes in neural circuits. We therefore view it as unlikely that a single gene would modulate a neurocomputational parameter responsible for normal cognitive variation.[5]

Next, in this first formulation of the framework, we assumed that the combination of alleles for each polygenic neurocomputational parameter had a deterministic relation to the value of that parameter in the instantiated network (see Discussion). We assumed

---

[4] Note that the current theoretical view is that the relationship between genes and *cognitive processes* is most likely pleiotropic (see, e.g., Kovas & Plomin, 2006).

[5] The following list gives some examples of the low-level variations one might expect (taken from Sapolsky, 2005). At the level of individual neurons, conservatively, one might expect variation between individuals in the number of dendritic spines, the number of axon terminals, the level of resting potentials, the size of the dendritic wavelet caused by pre-synaptic activity, the excitability of the axon hillock, and the speed of propagation of the axon potential. At the level of two neurons communicating, one might expect individual variations in the amounts of neurotransmitter released, the numbers of receptors, the efficiency of receptors in binding neurotransmitters, the efficiency of producing neurotransmitters, the efficiency of producing receptors, and the proportions of different types of receptors. At the level of long-term potentiation, one might expect variation between individuals in how much glutamate neurotransmitter is released, the number of glutamate receptors, the ratio of the two types of glutamate receptor, the level of calcium ion release, and the level of phosphoration of the receptors. It is likely that a range of gene variants contribute to each of these neural parameters. Our higher-level models encode much coarser neurocomputational parameters such as “level of processing noise” or “learning rate”, which would correspond to the combined effect of many of the more detailed neural properties.

(and did not instantiate) a much larger part of the genome that was species universal and coded for the basics of, for example, creating the processing units, the connections, the activation dynamics, the sensorium, the input-output connectivity, and the mechanics of experience-dependent systems.

Turning to mechanisms of breeding, we assumed that there was sexual reproduction, so that each gene had a 50% chance of being passed on to a gamete (egg or sperm), which combined with a gamete of another individual to create a new offspring. Although reproduction was sexual in this sense, we did not consider sex effects in these simulations (i.e., there were no genetic differences between males and females). During breeding, we assumed that there was uniform crossover and no linkage disequilibrium, the latter falling beyond the scope of the project. That is, the presence or absence of a given allele in a gamete was independent of the presence or absence of any others. This assumption is violated in humans because genes on the same chromosome have a greater than 50% chance of being transferred into a gamete together, and the closer they lie on a chromosome, the higher the chance.

When Genetic Algorithms are used for machine learning optimisation, the most successful individuals of the previous generation are often retained in the next generation. In our case, after breeding, the previous generation died. Breeding allowed individuals to be created with different degrees of relatedness, for instance as twins or siblings. MZ twins shared the same genome. Dizygotic twins and siblings were created by generating two offspring from the same set of parents, but from a different sperm and egg. Dizygotic twins and siblings shared 50% of their alleles on average. Also in contrast to the more common use of Genetic Algorithms, we did not include genetic mutation during

reproduction. In humans, the mutation rate is extremely low (e.g., Strachan & Read, 2003, cite a rate of between 1 and 4 mutations per 100,000 genes per generation). Mutations serve to reduce the average genetic similarity of siblings below 50%, and our preference was to maintain retain the 50% value, since it is the one deployed in standard behavioural genetic models. Several other aspects fell beyond the scope of this project. We did not model the effects of epistasis (interactions between genes) or epigenetic effects on gene expression; we did not model assortative mating – in our simulations, mates were selected at random from the population; and we did not model gene-environment correlations (Plomin et al., 2008) – variation in the quality of the environment had no correlation with the quality of an individual's genotype.

The total of number of genes used to encode the value of the 14 computational parameters was 126 (or two copies of 63). Figure 2 includes the range of values for each parameter and their frequency of occurrence in the population. The number of genes encoding each parameter is given in the legend. The translation of a genome into a parameter set was implemented by assigning efficient alleles the value 1 and inefficient alleles the value 0, and then deriving the total for all the genes influencing the parameter (thereby ensuring additivity). The parameter value was calculated from the total using a lookup table, created by hand for each parameter to reflect the range of values identified during the calibration stage. Table 2 shows an example table for the temperature parameter linking the number of more efficient alleles in the gene set to the parameter value.

==========

Insert Figure 2 and Table 2 about here

==========

A more constricted range of genetic variation was also considered for each computational parameter, shown in Figure 2 in grey. This required fewer genes to encode, leading to a genome with only 60 genes (2 copies of 30). Because the parameter of learning algorithm only had two values in the original formulation, we restricted the range of variation by fixing the parameter to use cross-entropy, thus removing variation in this gene. Two other conditions of genetic variation were considered when we came to assess the power of association analyses to predict behavioural variability from the values of the genes. In the first populations, efficient and inefficient alleles had an equal frequency. In these latter populations, the frequency of efficient and inefficient alleles was unbalanced. In one population, alleles were efficient with a probability of 70% and inefficient with a probability of 30%. In the second, alleles were efficient with a probability of 30% and inefficient with a probability of 70%.

Finally, although the genome of each network represented a deterministic specification of parameter values or ranges, the genome should not be viewed in terms of the old-fashioned 'blueprint' view of the relation of genotype to neural substrate. The parameters corresponded not only to how the network was built (e.g., number of layers, internal units, and connections) but also how it ran (processing noise, threshold functions), how it was maintained (weight decay and pruning), and how it adapted (learning rate, momentum). Some of these parameters could be seen to index genetic effects on brain development operating early on and not thereafter (e.g., neurogenesis, neural migration), while others could be seen to index on-going gene expression in the

maintenance of function (neural dynamics) and adaptive processes (plasticity, pruning, decay).

The stability of genetic variation across generations

The empirical starting point for our simulations was the presence of behavioural variability in the human population. We have proposed a framework in which sexual reproduction permits the transmission of the genetic contribution to this variability between generations, under the assumption that direct selection does not operate in respect of our target behaviour, past tense formation. It was important to demonstrate that the outlined mechanisms for genetically encoding parameter information and for sexual reproduction together served to maintain the level of variability across subsequent generations. Should these mechanisms actually lead to the elimination of genetic variation over a small number of generations, then our framework would have to be viewed as inadequate for explaining the origins of population variation, which persists across human generations.

The stability of allele frequencies across generations is in fact a well-known concept within population genetics (see Appendix A for further explanation and a demonstration using the breeding methods assumed in the model). It is known as the Hardy-Weinberg principle (see, e.g., Halliburton, 2004). The frequency of alleles, and therefore the genetically encoded variability, will remain constant at the population level so long as the following five conditions hold: there is no mutation, there is no immigration into or emigration out of the population, the population is large, individuals mate at random, and reproductive success is random. It is understood within population

genetics that the Hardy-Weinberg principle rarely holds in natural populations. However, it provides a foundation to consider how gene frequencies can change across generations in a given population (that is, the process of evolution) by observing which of the assumptions is violated. The fact that we stipulate that the five conditions hold in the current modelling framework marks another level of simplification in our investigation of cognitive variability.

Modelling variation in the environment

In machine learning terms, we operationalised environmental quality as a time-varying function with respect to the training set. We assumed that there was a 'perfect training set', in this case comprising all of the past tense verbs available in the language, along with their accepted past tense forms. We defined the function for variations in the training set as follows:

$$\text{Training set } P_nT_t = f\{perfecttrainingset, X, Y, Z\} \quad \text{(Eq.1)}$$

The training set for Person *n* at time *t* is a function *f* of the perfect training set and three parameters: *X* = proportion of valid training trials, *Y* = proportion of invalid training trials, and *Z* = proportion of noise trials. Invalid trials have the same input as a training pattern but a different output. Noise trials have different inputs and outputs or include partial information consistent with training patterns. In principle, the function could be influenced by Person *n*'s behaviour or experiences at *t-1*. This accommodates the

possibility of, for instance, reduced reward leading to reduced attention and therefore a subsequently smaller proportion of valid training trials.

We made the following assumptions in our implementation. First, once instantiated, we gave each network a *pre-conditioning* phase to produce divergent initial connection weights. This comprised a version of the training set with *X* set to 0, *Y* set to 0, and *Z* set to 1. The training set was made up of 30 random binary vectors at input and output, uniquely created for each individual, and trained for 50 epochs. This phase was intended to simulate the effect of early subjective experience prior to using the relevant learning system for, in this case, modulating phonological forms via tense information. Next, we created a training set for the past tense information available in each family environment. To do so, we generated a *family quotient* for each simulated child. This was a number between 0 and 1. This value was used as a probability whether each verb in the perfect training set would be included in the family's vocabulary. In terms of Equation 1, *X*=family quotient, *Y*=0, *Z*=0. The family training set was then fixed throughout development. However, performance was always assessed against the perfect training set (analogous to a standardised test of past tense formation applied to all children). The family quotient manipulation corresponds to a reduction in type frequency for both regular and irregular verbs, while the token frequency of each verb (3 times greater presentation for high than low frequency) was retained. Our final decision was how to sample the family quotient values. Inspection of SES distribution in Hart et al.'s (2007) maternal education data indicated a normal distribution with a large standard deviation but a positive skew. The distribution from Bishop (2005) for the composite SES measure had a similar normal distribution with large standard deviation but with a negative skew.

We selected a uniform distribution, which slightly exaggerated the incidence of high and low SES. We selected two ranges of environmental variation, a narrow range, sampling family quotient values between 0.6 and 1, and a wide range, sampling quotient values between 0 and 1.

Summary of implemented sources of variation

Table 3 summarises the causal factors producing variation in population development. These are split into the categories of genetic factors, environmental factors that are common to individuals raised in the same family, and environmental factors that are unique to individuals. In Table 3, we follow the standard behavioural genetic definition in treating instances of noise or stochastic variation as unique environmental causes. Measurement error also falls within this category. We consider the simulation of measurement error in the Results section. We return to the attribution of causes to the three categories in the Discussion section, where, for example, we discuss whether in reality neurocomputational parameters might themselves be partly environmentally determined.

==========

Insert Table 3 here

==========

Simulation design

Each population was created as follows. An initial 'breeding' population of 1000 individuals was produced with random binary genomes. From this, two individuals were

picked at random to mate and produce two offspring. This was repeated without replacement to produce the next generation of 1000 individuals, whose development was then assessed. Populations constituted pairs of MZ or DZ twins. Twins were used for two reasons. First, our target empirical data on past tense formation comprised MZ and DZ twin pairs. Second, the simulated twin pair data were amenable to later investigation of the heritability of acquired behaviour. We consider analyses of heritability in a separate work (Thomas, Forrester & Ronald, in preparation). For the rest of this paper, we will ignore the twin status of the individuals and include both twins in the population. However, every empirical result was also analysed using only the first of each twin pair in a reduced population of 500 individuals and, unless reported otherwise, the same qualitative pattern of results was found to hold.

Given a population of 1000 genomes, individuals were then trained as follows. The individual was trained initially on a unique *pre-conditioning training set* to produce unique and divergent starting weights. A family quotient value was then generated and used to generate the *family training set*. For separate populations, this was sampled to produce either narrow or wide environmental variation. Following pre-conditioning, each network was trained for 1000 epochs on its family training set. At each epoch, performance was measured on the perfect training set. Performance was assessed on regular verbs, on exception verbs of three types (no change, vowel change, and arbitrary), and on generalisation of the past tense rule to novel forms. Performance was measured in two ways, either in terms of nearest-neighbour accuracy (% correct) or in the RMS error between the network output and the target output (i.e., how close the activation values were to what they should have been for that verb).

==========

Insert Table 4 here

==========

Four populations were run in a 2x2 design, shown in Table 4. Genetic variation was either wide (full range of neurocomputational parameter values) or narrow (restricted range of values), and environmental variation was also either wide (full range of variation in family quotient) or narrow (restricted range of family quotient). This permitted evaluation of the respective influence of genetic and environmental factors for different levels of variation in each. The results for each population were compiled in a database that permitted recovery of the population distribution of performance for any measure at any point in training. Since the full genomes, the full set of neurocomputational parameters, and the family quotient values were available for all individuals, it was then possible to assess the extent to which these predictors explained population variability for any point in training.

Results

Mean performance and population variability on the past tense task

Empirical data

Bishop (2005) reported the performance of 442 6-year-old children on Rice and Wexler's (2001) Past Tense probe sub-test. While Bishop combined responses into a single score, here we separate performance into the accuracy of elicited regular past tenses (n=11 items) and irregular past tenses (n= 8 items). The irregular verbs in the sub-test were mainly vowel-change past tenses. Figure 3a shows the population mean and standard deviation on the two verb types. Regular past tenses were produced more accurately than irregulars ($t(441)=43.22$, $p<.001$), and with a smaller standard deviation due to a ceiling effect. Figure 4a and b depict the population distributions for the two verb types. Regular verbs exhibited a highly skewed distribution, with many children getting all 11 past tenses correct, while irregular verbs demonstrated a more normal distribution. There was some suggestion of bimodality in the irregular verb distribution, with a low performing sub-population centred on 35%, and a higher performing sub-population centred on 70%. The disparity between regular and irregular patterns demonstrates that the population distribution in performance depended on the measure used to assess knowledge of the past tense domain. In previous research, disparities between the quality of regular and irregular verb performance have been taken as evidence that separate cognitive mechanisms are responsible for acquiring the two verb types (see, Pinker, 1999).

==========

Insert Figure 3 and 4 about here

==========

Fitting the model to 6-year-old performance

Four simulated populations acquired the past tense. Our first requirement was to determine which point in development corresponded to that of 6-year-old children. We calibrated the models against performance on irregular verbs, determining for each population the epoch in training when the mean accuracy in vowel-change verbs was closest to the 37.4% exhibited by the children. Using the terminology that *G* stands for the range of genetic variation, *E* stands for the range of environmental variation, and one epoch corresponds to a single presentation of the family's training set, the calibration points of training were as follows: *G*-wide-*E*-narrow = 63 epochs, *G*-wide-*E*-wide = 177 epochs, *G*-narrow-*E*-narrow = 88 epochs, and *G*-narrow-*E*-wide = 190 epochs. The population means and standard deviations for the four populations are shown in Figure 3b. Mean performance levels on regular verbs were slightly lower than for the empirical data, reflecting the fact that we are using a simplified model of past tense acquisition[6], but all capture the qualitative difference in accuracy depending on verb type.

A brief comment is necessary regarding measurement noise. It is clear that the behavioural data include measurement noise, since we know that the test-retest reliability of the Past Tense Probe sub-test is 0.82 (Rice & Wexler, 2001). However, the performance of the models reported above represents a pure measure of the specific cognitive mechanism hypothesised to drive this behaviour (though some individual

[6] There are several ways to increase the relative efficiency of regular past tense learning, via changes to the input encoding or to the architecture. See Thomas and Karmiloff-Smith (2003), Karaminis and Thomas (2009) for discussion of more complex inflectional models.

networks possessed intrinsic variability in their behaviour due to processing noise within this mechanism). For the purposes of our analyses, we did not add measurement noise to our simulated behaviour. This is because we wished to later test the best-case scenario of predicting variations in behaviour from genetic, neurocomputational, and environmental measures. In the presence of noise, the predictive power of these variables would obviously be lower. In separate work, where we assessed the heritability of behaviour, it became more important to include measurement noise, since this is a component of unique environment effects (Thomas, Forrester & Ronald, in prep.).

In sum, the simulations gave a qualitative fit to the empirical data both for the mean performance level and population distributions for regular and irregular verbs.

Simulating population variation and its developmental context

For ease of exposition, we initially focus on the *G*-wide-*E*-narrow condition for considering population distributions for each past tense measure. We then consider how these distributions change across development, how they depend on measure sensitivity, and how they are influenced by the relative range of genetic and environmental variation. Figure 4c and 4d show the population distributions for regular and vowel-change irregular verbs when the population level of development was matched to 6-year-olds. These data qualitatively replicate the different distributions for each verb type: the high-skewed regular performance, with an extended lower tail, and the more normally distributed irregular performance. Note these differential distributions are not, in this case, an indication that separate cognitive mechanisms are responsible for acquiring the

two verb types; rather, they reflect the differential difficulty of a single mechanism in acquiring mappings with varying levels of input-output consistency.[7]

Interestingly, the distribution for irregular verbs was more uneven than that for regular verbs, suggesting that the distribution is made up of sub-populations defined by this measure. There are many simultaneously varying computational parameters that might have marked effects on learning, thereby generating apparent sub-populations, and we explore this in more detail later. Here, we picked two parameters that previous modelling work suggested might be important: did the network use a 3-layer network with hidden units (+) or a 2-layer architecture (–); and did the learning algorithm use cross entropy as an error signal for backpropagation (+) or Euclidean distance (–). These two contrasts defined four potential sub-populations with, respectively, 841, 57, 95, and 7 simulated individuals. Figure 5a and b plot the contribution of these individuals to the regular and irregular verb distributions. The sub-populations contribute equally to the skewed regular verb distribution. However, for the irregular verbs, the –/+ group contributed a low-performing normal distribution centred near 15% and the +/– group contributed to the number of individuals close to floor performance. This pattern is consistent with the view that in the model at least, computational parameters marked out sub-populations within irregular verb performance but not regular verb performance.

==========

Insert Figure 5 around here

==========

[7] Though see Thomas, Forrester & Ronald (in prep.). Performance on regular and irregular verbs in this population in fact showed different *heritability* as well.

One has to be cautious here, however. The regular verb distribution is skewed because regular verb performance is at ceiling in many individuals. If we traced performance back to an earlier developmental stage in the simulations, to a point where there was greater variability in regular verbs, perhaps we would find that the same sub-populations were now marked out in regulars too? Figure 5c and d address this question, now examining distributions for 10 epochs of training compared to the 63 epochs used above. Comparison of Figures 5c and 6b indicates that the story is more complicated. The +/– sub-population contribute low performance to both regular and irregular verbs (compare dashed lines in Fig.5b and 5c). However, the –/+ population (2-layer using cross entropy) only contributed low performance to irregular verbs (Fig.5b, grey line). In Figure 5c, the –/+ distribution follows that of the larger +/+ population and is similar to the population as a whole. In short, the ability to detect sub-populations depends on the variability present in the measure (which itself depends on the stage of development). But the simulations demonstrate that it is quite possible for sub-populations to exist in one behavioural measure of a domain (irregulars in past tense) but not in another (regulars). The differential pattern is caused by computational properties whose variation is more relevant for some behaviours than for others. While the empirical data also exhibited evidence of sub-populations in irregular verb performance, we would be cautious in viewing this as a real effect until it was replicated in a further large sample. Nevertheless, we do know that the children in our population contained a subset of 31% with a behavioural diagnosis of SLI.

One focus of the simulation work was to place the population distribution within its developmental context. Our empirical data only contained an age range of 13 months.

Age did not predict a reliable amount of the variance in regular verb performance within this range (R^2=.005, F(1,440)=2.37, p=.124) and only a borderline significant level of variance in irregular verb performance (R^2=.009, F(1,440)=3.91, p=.049). Our simulations therefore move beyond the empirical data at this point. Figure 6 includes five plots showing how population variation changed across development for each of our behavioural metrics: performance on regular verbs, generalisation of the past tense rule to novel verbs, and performance on the three types of irregular verb, no-change (*hit-hit*), vowel-change (*hide-hid*), and arbitrary (*go-went*). Distributions are plotted for nine points in learning, at epochs 1, 5, 10, 25, 50, 100, 250, 500, 750, and 1000 epochs. These data indicate that each verb type begins with a floor effect (the majority of the population failing to score), extends an upper tail through development, shows a roughly normal distribution midway through training and, depending on the verb type, finishes with a skewed distribution at ceiling. Verb type modulated this pattern by being relatively smoother for regular verbs and arbitrary irregulars, and by being relatively delayed and more uneven for no-change and vowel-change irregulars. One interpretation of this modulation is that high type frequency (regulars) and high token frequency (arbitrary irregulars) drive through both faster and more uniform development at the population level. Irregular verbs with lower frequency appeared more vulnerable to the vagaries of computational variations in the learning system and imperfections in the training environment.

==========

Insert Figure 6 around here

==========

One concern with these data is that the distributions were very much shaped by floor and ceiling effects, that is, limitations in the sensitivity of the behavioural measure. As Figure 5 demonstrates, this can cause difficulties in comparing across behaviours since, for a given point in development, the distributions were quite different due to the floor and ceiling effects. We therefore considered how population distributions changed across development using a more sensitive measure. Instead of measuring past tense performance via accuracy (which used a nearest neighbour calculation to 'clean-up' the output), we employed a distance metric, calculating how close the output activations were to the correct unit activations for each verb. The metric was the RMS error. To produce a more intuitive scale, we transformed the error using the negative of the log: for comparability with accuracy distributions, larger numbers now corresponded to better performance. We expected this measure to be more sensitive and therefore less likely to exhibit floor and ceiling effects. Figure 7 shows the resulting distributions. The more sensitive measure indeed retained greater variation in the population, with regular verbs in particular exhibiting a normal distribution through to the end of training. The measure also retained the respective ordering of the different verb types. However, our intention was later to exploit the greater range of variation in the *–log(error)* as a target for predicting population level variability. But the *–log(error)* measure proved hard to predict using parameter or allelic values, with summed variance dropping from around 50% when the dependent variable was accuracy to 3% when it was RMS error. The difficulty in predicting error compared to accuracy in part stemmed from a non-linear relationship between RMS error and accuracy. This implies that both measure sensitivity

and the linearity of a measure's relationship with predictors (such as age or ability) are important factors in studying the origins of population variability.

==========

Insert Figure 7 around here

==========

Lastly, we assessed the impact of the relative range of genetic and environmental variation for population distributions. Figure 8 depicts these distributions for our 2x2 design for regular verbs and vowel-change irregulars, for an early point in development (50 epochs) and a late point in development (750 epochs). The range of environmental variation proved to be the most important factor in determining the distribution, with wide variation producing a greater spread in performance, an effect that persisted across development. This disparity is mainly due to calibration: our manipulation of wider environmental variation encompassed poorer conditions, whereas a wider genetic range involved both better and worse neurocomputational conditions. The combination of wide genetic variation and wide environmental variation produced a population with the poorest performance. This is because for an individual, it maximised the opportunities for failure, either of a collection of poor gene variants or of a bad environment. Note that assortative mating, not implemented here, might attenuate this effect, since the poor gene variants would be associated with the poor environments.

In sum, these results showed the importance of the behaviour being measured (here, verb type) and the sensitivity of the measure, both in generating the population distribution and predicting how it would change across development.

==========

Insert Figure 8 around here

==========

The predictive power of the environment

Empirical data

For the children's data, the sole environmental measure available to pick up variation in past tense acquisition was the SES composite derived by Petrill et al. (2004), which combined information on parents' education and occupation along with the mother's age at the birth of her first child. The relationship between this measure and past tense production for regular and irregular verbs is shown in Figure 9a and b. SES predicted 0.7% of the variance in regular performance when all children were included, and 2.0% when only one twin was selected from each pair chosen at random (all: R^2=.007, F(1,440)=3.01, p=.084; one twin: R^2=.020, F(1,218)=4.37, p=.038). The predictive power of SES was compromised for regular verbs due to ceiling effects. Irregular verbs possessed a larger range of variability and here SES predicted 3.7% of the variance for all children, and 2.8% when only one twin was included (R^2=.037, F(1,440)=16.69, p<.001; R^2=.028, F(1,218)=6.15, p=.014). There was a reliable interaction between SES and verb-type when all children were considered (F(1,440)=6.68, p=.010, η_p^2 = .015). In the one-twin analysis, SES predicted more of the variance in regular and less in irregular verbs, rendering the SES by verb-type interaction non-significant (F(1,218)=.48, p=.489, η_p^2 = .002). While SES data were missing for 11% of the sample, analyses run with or without these children produced almost identical results.

SES therefore predicted reliable but small amounts of variance in past tense acquisition. In the full analysis, it predicted more variation in irregular verb performance than regular, although this may have arisen from ceiling effects in the latter. This empirical result contrasts with findings from a longitudinal study of Rice, Wexler and Hershberger (1998), which examined the association between SES as measured by maternal education and regular past tense formation in 20 typically developing children and 21 children with SLI. Rice et al. found that their measure of SES explained less than 1% of the variance, which was not statistically reliable. In the absence of other environmental predictor variables, these authors concluded that past tense development was under *maturational control* (Rice et al., 1998, p.1428). Differences in rates of acquisition across children were attributed to genetic differences in specification of the timing of linguistic properties (Rice & Wexler, 1996). By contrast, with the larger sample provided by Bishop (2005), we see small but reliable effects of environment on past tense acquisition for these 6-year-old children. Nonetheless, 96% of the variance must be explained by other sources, be they environmental or genetic.

Simulation data

Figure 9c and 9d plot the corresponding data for the *G*-wide-*E*-narrow population, at a point in development that matched accuracy on irregular verbs. The index of environmental variation was the *family quotient value*, which determined the size of the sub-sample of the 'perfect' past tense training set available in a given family. The lower the quotient, the more restricted the training set. The model successfully replicated the weaker predictive power of the environmental manipulation on regular verb performance.

Family quotient values explained 0.7% of the variance for regular verbs and 2.0% of the variance for irregular verbs (regular: R^2=.007, F(1,998)=6.76, p=.009; R^2=.020, F(1,998)=20.07, p<.001; interaction of family quotient x verb-type: F(1,998)=8.32, p=.004, η_p^2 = .008). While the variance proportions are numerically similar to the empirical data, this may be viewed as somewhat co-incidental, since genetic and environmental variation were not calibrated to achieve this result. Moreover, as indicated earlier, we included no measurement error in the simulation data. To achieve the requisite correlations in the presence of measurement noise, the contribution of environmental variation would have to be stronger than that used in the *G*-wide-*E*-narrow condition.

==========

Insert Figure 9 around here

==========

Once more, the simulation data allow us to consider this result within a developmental context, and more widely, what patterns would be expected under different relative influences of genes and environment. Figure 10a to d display these results, plotting the percentage of variance explained by the family quotient value for each population of 1000 networks for each verb type across training for the 2x2 design, using a linear regression in each case. Unsurprisingly, under conditions of wide environmental variation shown in Figure 10b and d, the predictive power of the family quotient value was much higher, explaining between 60-80% of population variation.

Three other results are of interest. First, in general, the predictive power of the environmental manipulation increased across training. Only generalisation to novel verbs showed a decrease. Second, differences in predictive power did not appear to be strictly

related to the degree of skew in the population distribution caused by floor or ceiling effects. Consider the degree to which environment explained variation in regular verb performance late in training. Figure 8c indicates that for *G*-narrow-*E*-narrow and *G*-wide-*E*-narrow conditions, late in training there was a skewed distribution in performance due to ceiling effects. Yet Figure 10a and c show that in the former case, environment predicted around 35% of the variance, while in the latter, it predicted only 2%. Differential amounts of variance can be predicted even when the population distribution is very skewed. Third, the range of the genetic and environmental influences interacted with verb type. There were clearly properties of the input-output mappings that were differentially sensitive to sources of variation. Arbitrary irregular past tenses tended to be predicted most strongly by environmental variation. This is because, if this idiosyncratic information was not available in the input, it could be inferred from other training instances. Perhaps most striking was the interaction between regular and irregular verbs under conditions of narrow environmental influence, when genetic influence can be narrow or wide (Figure 10a and c). When genetic variation was wide, genetic factors determined variation in regular verbs more than in irregular verbs (that is, environment predicted regular variation less, shown in Figure 10c). When genetic variation was narrow, it determined variation in irregular verbs more than regular verbs (that is, environment predicted regular variation more, shown in Figure 10a).

Putting these from environmental variation together, the model suggests the following interpretation of the empirical data. A pattern where the environment predicted only small amounts of variation in the acquisition of past tenses implies narrow environmental variation (Fig.10a or 10c). A pattern where environment predicted

irregular verb variation more strongly than regular verb variation fits with conditions of wide genetic variation combined with narrow environmental variation (Fig.10c). However, the simulation results do not directly equate to Rice et al.'s (1998) conclusion that regular verb acquisition is under the control of 'maturation', since the simulated acquisition process here is driven by experience-dependent changes in a general-purpose learning mechanism that provided with the appropriate information input-output information to acquire the skill. Identifying the weak predictive role of the environmental measure should not lead to the conclusion that environmental input is unnecessary for learning. Nevertheless, the simulation results concur with the view that genetic variation explains more of the behavioural variability than environmental factors, for this particular English-speaking population.

==========

Insert Figure 10 around here

==========

A novel prediction: environment as a predictor of giftedness versus delay

The current simulation framework is consistent with the view that the causal mechanisms operating to produce performance in the tails of the population distribution are the same as those causing variation in the normal range. There are no special mechanisms here that operate to produce delayed development or gifted development, simply an accumulation of less or more efficient alleles on the genetic side and poorer or richer family training sets on the environmental side. Nevertheless, we were interested in the role of the environment in predicting performance in the tails, and in particular whether it played a

symmetrical role in predicting delay as it did in predicting giftedness. For our 2x2 design, we considered performance on irregular vowel-change verbs at the stage of training matched to 6-year-old children. At this point, we classified networks by whether they fell in the bottom 10% or not, and whether they fell in the top 10% or not. Irregular verb performance was used because it provided a wide range of variation to perform this categorisation. The family quotient value was then used to predict the status in each case, either delayed or not, or gifted or not. Table 5 includes the resulting effect sizes of this analysis.

==========

Insert Table 5 around here

==========

When environmental variation was wide and genetic variation narrow, environment played a symmetrical role in predicting success and failure, because it explained the same amount of variance (~15%). In all other conditions, however, environment tended to be a more powerful predictor of success than failure, most markedly when both environmental and genetic variation were wide (top 10%: 20.4% of the variance; bottom 10%: 5.0% of the variance). For conditions of narrow environmental variation, environment didn't predict delay at all (neither was statistically reliable; all other values $p<.001$). What is most striking here is that although a poor environment did not predict delay, we know that *it did cause delay*. Since we constructed this model, we know that impoverishing the composition of the training set indeed causes poorer performance in acquisition. We return to this disparity in the discussion.

In the previous section, we saw that the *G*-wide-*E*-narrow condition was the closest fit to the empirical data. This condition indicated that we should not expect the environment to account for any of the variance in whether children fell in the bottom 10% of population for irregular past tense performance, but we should expect it to account for around 2% of the variance in whether children fell in the top 10%. We tested the prediction using the data of Bishop (2005) and the SES composite of Petrill et al. (2004). This constituted a fairly strong test of the modelling framework, since the empirical data were not collected with this comparison in mind, and the implementation of the real SES measures in terms of a training set manipulation was only provisional. Nevertheless, the results were in line with the model's prediction (Table 5, bottom row). SES did not reliably predict children falling in the bottom 10% (R^2=.001, F(1,440)=.55, p=.460; excluding cases with missing SES data: R^2=.001, F(1,394)=.56, p=.456); but SES did reliably predict children falling in the top 10%, accounting for 2.5% of the variance (R^2=.025, F(1,440)=11.51, p=.001; excluding cases with missing SES data: R^2=.029, F(1,394)=11.71, p=.001).

Predictive power of genes and neurocomputational parameters: Simulated association analyses

For the current simulations, there were several ways we could seek to predict population variance using factors intrinsic to each individual. Not all of these methods are currently available to behavioural and molecular geneticists working with human populations. For example, for the simulations, we know the actual value for each neurocomputational parameter for each individual; we know which genes specify the value of that parameter;

and we know the way that the gene variants are linked to the computational parameters. Molecular geneticists have identified gene variants showing correlations, and in some cases begun to unpack the biochemical processes in which protein products are involved. They do not, however, know the full set of variants affecting a parameter, nor the shape of the function linking the relevant variants with their consequences on behaviour via modulation of the neurocomputational parameter.

For an illustration, consider the temperature parameter, which alters the shape of the activation function determining how the activity of an artificial neuron is related to its net input. Figure 11a plots the function derived during calibration, where this parameter was altered while the other parameters were held constant (with the exception of stochastic factors during training; see Table 3, right column). A quadratic function linking the temperature value with behaviour accounted for 87% of the variance, suggesting that network performance worsened if the parameter was set either too high or too low. This type of relationship has been proposed to link, for example, cortical dopamine levels with working memory performance (Goldman-Rakic, Muly & Williams, 2000). Figure 11b plots the relationship between the temperature parameter value and population performance on irregular past tenses, for the *G*-wide-*E*-narrow population benchmarked to the performance of 6-year-old children. The parameter value now predicts much less of the variance, at only 12%. Additional variance is now produced by variation in the other parameters and by the environment. Moreover, there is no longer equal incidence of individuals with each value of the parameter.

As investigators, we might, however, have less information available than this. We might only have identified the gene-variants relevant to our target

neurocomputational property, without knowing how the variants altered its value. There are 10 binary alleles specifying the temperature value, organised into five pairs. The 10 alleles produced 11 possible sums of the number of more efficient alleles (from 0 to 10 1s; see Table 2). Under these circumstances, one logical way to organise the data would be to rank the genotypes from 1 to 11 in order of the performance levels they produced in behaviour. Figure 11c demonstrates this genotype-behaviour relation. The function is now monotonic (by definition). A linear fit to the categorical data explains a similar level of variance to the quadratic parameter-behaviour relation. This analysis shows that identifying the relevant genes and their variants sufficient to account for a significant proportion of behavioural variance does not in itself recover the non-linearity of the relationship between neurocomputational parameter and behaviour. Statistical relationships need to be complemented by investigation of the mechanisms delivering the computational function.

We may have still less information available than this. Through association analyses, we may have independently identified five genes that appear to predict some amount of variability in past tense development, while failing to link these genes to a common neurocomputational effect. Figure 11d displays the population performance split by the three genotypes for each gene, 00, 10/01, and 11. There are small differences in the means for each genotype, along with a great deal of variability. Three of these genes showed additive effects of the alleles (1, 2, and 4), while two showed interactions (3 and 5). Most of what we know about neurocomputation suggests that it must be non-linear, in order for subsequent layers of neurons to add computational power in a hierarchical system (see, e.g., Elman et al., 1996, p. 222). However, under the assumption of

polygenicity in the current model, variants of individual genes (defined by pairs of alleles) have limited opportunity to reveal the non-linearity of processing, because their contribution shifts performance only a small way up or down the function linking the neurocomputational parameter to behaviour. Thus the variation in behaviour produced by individual genes (as per Fig.11d) may provide too restricted a view of the relevant neurocomputational function to reveal its true level of non-linearity (as per Fig.11a) and make the contribution of small variations more additive than they actually are.

==========

Insert Figure 11 here

==========

Figure 11 identifies possible ways that we could relate behavioural variation to the intrinsic factors of the model.[8] To simulate association analyses, we proceeded as follows. We first considered the relationships between computational parameters and population variability. Second, we dropped to a lower level, considering what associations can be recovered by directly linking gene variants to behaviour. In both cases, we employed just a single condition from the 2x2 design, that of *G*-wide-*E*-narrow, to maximise the chance of identifying parameter-behaviour or gene-behavioural correlations. We then investigated two additional aspects. We sought to place these

[8] The data depicted in Figure 11 are for one behavioural measure, at one point in development, for one population. The framework employed 14 parameters, 5 behavioural measures, 1000 points in development, and initially 4 populations. Figure 11 therefore represents just 1 of 280,000 similar parameter-behaviour plots. The current section selects some of more salient findings from the simulated association analyses but a full description is beyond the scope of this report. All databases and parameter sets are available from the authors as Excel spreadsheets

correlations within a developmental context; and then assessed the extent to which parameter- / gene- behaviour correlations replicated across different populations. The latter issue is relevant for modern association analyses, because many gene-behaviour associations are found not to replicate across different samples. Under the artificial simulation conditions of zero measurement noise, could small gene-behaviour relations be consistently detected across different populations?

Parameter-behaviour associations

Figure 12 depicts the proportion of variance explained by each of the 14 computational parameters, for each verb type (regulars, novel verbs, no-change irregulars, vowel-change irregulars, and arbitrary irregulars). Effect sizes were calculated for population accuracy levels at three points in training, early (50 epochs), mid (100 epochs), and late (750 epochs). For practical reasons, we used linear methods. Since many parameter-behaviour functions were non-linear, this likely underestimated the amount of variance each parameter can explain. However, for the number of regressions involved, it was too cumbersome to individually model each non-linear function to maximise the explained variance. Most parameters explained less than 5% of the variance, but some parameters exerted stronger effects. The architecture, the learning algorithm, and the nearest-neighbour threshold explained from 10% to 25% of population variance.

It is notable that while the shape of the plots across the different verb types is broadly similar, there were differences. For example, learning algorithm was most important for regular verbs, while architecture played a stronger role for irregulars, and nearest-neighbour threshold played a marked role for novel verbs. These modulations

demonstrate that the mapping problems presented by each verb type rely differentially on the different computational properties of the general purpose learning system implemented by an artificial neural network. Karmiloff-Smith (1998) described this theoretical possibility in terms of *domain-relevant* computational properties. That is, learning algorithms and architectures do not encode information specific to the domain of regular verbs versus irregular verbs. Nevertheless, variation in these parameters can have differential influence in the acquisition of this domain of language morphology when a general purpose learning mechanism is exposed to the appropriate information. Karmiloff-Smith (1998) viewed domain-relevant properties within more general purpose learning systems[9] as candidate low-level differences to explain the origin of domain-specific deficits in high-level behaviours in developmental disorders.

Comparison of the effect sizes calculated at different points in training indicates that values are not constant across development. Given that these are statistical correlations, one might expect effect sizes to vary depending on the range of variation observed in the dependent variable. As we saw in Figure 7, the distributions change across development, exhibiting different levels of skew due to floor and ceiling effects. However, the fact that the predictive power of some parameters goes up across development (e.g., learning algorithm and architecture) while that of others goes down (e.g., nearest neighbour threshold and temperature) indicates that developmental differences cannot solely arise from changes in the performance distribution on each measure. Indeed, some of the patterns make computational sense. Nearest neighbour

[9] 'General purpose' here refers to computational principles. They may be instantiated in a domain-specific system, with respect to the information provided by input-output connectivity.

threshold becomes less important later in development. The threshold is a parameter that determines how close an output activation pattern has to be to a legal set of phonemes to be counted as correct. A lax threshold will give a network 'help' in getting an answer correct with only an approximate set of output activations. As the networks are better trained, their outputs will typically fall much closer to the target activations for the verbs they have acquired, so the setting of the threshold becomes less important: it is no longer required to 'clean up' the output. Temperature is a parameter that modulates the activation function of processing units. A low temperature means that processing units won't be sensitive to small differences in the input they receive. However, a low temperature can be compensated by larger weights on the connections arriving at a unit, since the larger weights can exaggerate the size of the differences in the input signal. Connection weights become larger with more training, so that a low temperature becomes less influential on performance. Conversely, both learning algorithm and architecture become more important in predicting population variability across development. This is because both parameters delimit the best representations that the network can develop. While the pathway to the end of development can be influenced by a range of factors (including the environment), its destination has fewer determinants. As performance asymptotes, learning algorithm and architecture become more important in predicting an individual network's ultimate performance level (at least, for the *G*-wide-*E*-narrow condition). In sum, it makes sense that the predictive power of computational parameters should alter across develop in an experience-dependent learning system.

==========

Insert Figure 12 here

==========

Gene-behaviour associations

Figure 13 drops to the lowest level of description, and plots the predictive power of all 126 alleles on behaviour on regular and vowel-change irregular performance. Each allele had the value of 0 or 1 and its value was used to predict population variability. (Here we have chosen to treat alleles independently rather than split them into pairs corresponding to genes. Due to the additive way in which the effect of alleles was exerted, identifying the summed effect for each gene / allele-pair would have constituted an arbitrary way of collapsing the effect of the alleles; it was the total of all relevant alleles that determined the parameter value.) Calculations were carried out at early, mid, and late points in training. Figure 13 also marks the computational parameter that each allele codes for, so that for instance, the first 10 alleles fall in the 'hidden unit' region of the genome. The analysis at a given point in training involved 126 independent regressions. Given the large number of comparisons, how large should an effect size be to be statistically reliable? To address this issue, a binary allele was created at random (1000 digits, corresponding to the 1000 individuals) and this value was used to predict population variance. The process was iterated one thousand times to create a distribution of the gene-behaviour effect sizes that could be produced by chance. Figure 13 includes the effect size at the 95^{th} percentile (solid horizontal line). An effect size larger than this value would have only a 1 in 20 probability of occurring at random. The criterion value altered depending on the distribution of the dependent variable but was typically around 0.5% of the variance.

Figure 13 demonstrates that a number of the alleles predicted reliable amounts of variance in behaviour, even though the effect sizes were typically small (mostly below 2%). These associations were visible despite the presence of an experience-dependent development process between the operation of the genes (on the parameters) and the measurement of the overt behaviour. Two further points are of note. First, the alleles with the strongest effects could differ depending on the measure of performance. That is, an allele in the nearest-neighbour threshold (NNT) region showed a stronger effect for regular verbs than irregular, while an allele in the learning rate (LR) region showed a stronger effect for irregular verbs than regular verbs. One could label these respectively as a 'gene for regular verbs' and a 'gene for irregular verbs' to mark the differential correlations. Second, the predictive power of alleles was sensitive to development. Some effect sizes increased across development, while others reduced. These changes were directly inherited from the computational level. If the predictive power of a parameter changed across training, for the reasons we outlined in the previous section, so too would that of any alleles that contributed to determining the parameter's value. Overall, two factors constrained the effect sizes that each allele could exhibit: (1) the effect size of the computational parameter for which it coded, and (2) the number of alleles that contributed to the variation in that parameter.

==========

Insert Figure 13 here

==========

The replicability of simulated gene-behaviour association analyses

To assess the stability of these effect sizes, we first re-ran the *G*-wide-*E*-narrow population, preserving the genomes but randomly exposing individuals to new environments. Without assortative mating, a network previously trained in a poor environment might now be trained in a rich environment. Figure 14a compares the effect sizes of the original and retrained populations at 50 epochs, while Figure 14b compares the effect sizes of the computational parameters. It is evident that the pattern of results replicated across the two runs at the levels of both parameter and allele.

We next created two new populations, re-sampling the genomes with the same method used to create the original *G*-wide-*E*-narrow population. Therefore, there was the same probability of parameter values in the population. These two new populations were exposed to new environments. Figure 15 compares the effect sizes of the original and two re-sampled populations for irregular vowel-change verbs at 50 epochs of training. In Figure 15a, it is evident that the *allele effect sizes did not replicate particularly well across genetic re-sampling*. Figure 15b, however, indicates that the effect sizes of the computational parameters *did* replicate fairly well. How can this disparity emerge? Closer inspection of Figure 15a reveals that re-samplings often have different spikes but all within the same region. For instance, pruning probability contains no reliable spikes, but learning rate contains many. The reason for the disparity is that in different populations, different alleles contributed to the variation in parameters. Different alleles threw their hats in the ring, as it were, on each re-sampling, whereas the parameter value was determined only by the number of hats in the ring. The implication of this simulation result is that under a polygenic model, individual association analyses from the allelic

level are predicted not to replicate, even under conditions of zero measurement error. The predictive power increases if one identifies all of the alleles that contribute to a particular parameter, equivalent to observing that spikes occur in the same region in Figure 15a.

==========

Insert Figures 15 and 16 here

==========

We considered one further condition for replication. In the original populations, there was an equal frequency of efficient and inefficient alleles (i.e., the bits for the alleles were equally likely to be set to 1 or 0). We created two new populations, one in which for the initial breeding population, there was a 70% incidence of more efficient alleles and 30% of less efficient, which we termed the *highskew* population; and a second in which there was a 30% incidence of more efficient alleles and a 70% incidence of less efficient, which we termed the *lowskew* population. Note that the functions linking allele values to computational parameter values were the same, but now the population frequencies of different computational values would be different. Figure 18 compares the effect sizes for the original population, *highskew*, and *lowskew* populations for irregular vowel-change verbs at 50 epochs of training. It shows that once more, effect sizes calculated from alleles did not replicate across populations. Now, however, effect sizes of parameters *did not replicate either*. This was because the computational balance of the systems had changed. For example, the *lowskew* population had generally worse parameters and poorer performance. Networks, therefore, tended to rely on the nearest-neighbour threshold far more to accept 'just good enough' output activations as correct answers, exaggerating the predictive power of this parameter in the *lowskew* population.

The simulation results imply that association analyses would not be expected to replicate across populations with different allele frequencies, but in this case, one would not expect parameter associations to replicate either (though note, in reality such neurocomputational associations are beyond our ability to measure at the present time).

==========

Insert Figure 17 here

==========

Correlation versus causation in association analyses

In some respects, the gene-behaviour plots shown in Figures 14-17 look a little like the results of real gene-behaviour association analyses, as if the peaks above the significance threshold served to identify the candidate genes involved in producing variance in past tense acquisition. But it is important to note, *all of the genes* shown in these figures were causal in producing parameter variation. We know this, because in the design of the model, each parameter value was determined by the summing the number of more efficient alleles. The differential pattern of correlations exhibited by the alleles arose from the way the parameter values happened to have been determined by the polygenic contributions in a particular population. In another population, different alleles could turn out to contribute more to the sum. In one sense, this is merely restating the point that, where there are additive polygenic effects, individual gene-behaviour associations may not replicate across populations. In another sense, it is highlighting the point that an absence of correlation is not definitive evidence for an absence of causality, a point to which we return in the discussion.

Mechanisms of resilience I: Gene x environment interactions

We have considered separately the power of environmental measures versus intrinsic measures in predicting population variability in behaviour. In this section, we consider how these two influences may interact. This does not refer to the causal process of experience-dependent change that necessarily involves the on-going interaction of environmental and intrinsic factors. Instead it refers to the possibility that the statistical, predictive power of a genetic factor may depend on the environment to which the individual is exposed. This is sometimes viewed in terms of resilience, where individuals exhibit unexpected levels of competence given the level of environmental adversity to which they have been exposed (Cicchetti & Blender, 2006). Knowledge of genetic variation may help identify individuals most at risk to adverse experiences, while protective functions of genes may also be discovered.

For the current simulations, we examined whether the variation in performance caused by changes in a given computational parameter was modulated by the family quotient value. This admittedly corresponds to a narrow notion of 'environment', following our implementation decision that for these simulations, environmental variation should translate only to manipulations in information content. For illustrative purposes, we took one parameter, the number of hidden units, and evaluated population performance in the *G*-narrow-*E*-wide population, split into quartiles of family quotient. We picked three hidden units levels, low (30), medium (50), and high (100). Our questions were: (1) did we observe gene-environment interactions; (2) was the pattern modulated by verb type; and (3) did the interaction change across development? Figure

17 depicts the population variation in regular verbs and vowel-change irregulars at an early and late point in development, depending on genetic and environmental conditions.

==========

Insert Figure 17 here

==========

Regular verb performance was influenced by the level of hidden unit resources, but this influence was independent of environment at both points in training ($p>.4$). Irregular verbs, by contrast, showed interactions between hidden unit resources and environmental variation at both points in training (early: $F(6,362)=3.05$, $p=.006$, $\eta_p^2 = .048$; late: $F(6,362)=3.95$, $p=.001$, $\eta_p^2 = .061$). Higher levels of hidden units allowed the environment to exert a greater influence on performance, while low levels compressed environmental variation. We compared these interactions against the results exhibited by the population trained with wide environmental variation. In that case, the interaction of hidden unit levels and environment in predicting irregular verb performance showed borderline significance early in training but was absent by late in training (early: $F(6,362)=2.03$, $p=.061$, $\eta_p^2 = .033$; late: $F(6,362)=.33$, $p=.921$, $\eta_p^2 = .005$). Overall, this illustrative set of data indicates that gene-environment interactions were sensitive to the mapping type demanded of the system, the relative variation of intrinsic and environmental influences, and the stage of development at which they were measured. Moreover, for hidden units at least, in these data we did *not* see a pattern where a parameter setting immunised a system to the disadvantages of an impoverished environment. Rather, parameter settings allowed a system to exploit a richer environment.

Lastly, we examined whether the gene-environment interactions observed when intrinsic variation was indexed by parameter values were also visible at the lowest level of alleles. We picked an allele in the hidden unit region with a large effect size (predicting 1.58% of population variance for irregular verbs at 50 epochs) and compared it to another allele in the hidden unit region predicting zero variance in behaviour. We contrasted performance in individuals experiencing low-average (family quotient 0.7-0.8) and medium-average (0.8-0.9) environmental conditions, depending on whether they possessed an allele value of 0 or 1. For the predictive allele, a gene x environment interaction was apparent, with a 2.7% performance difference between individuals possessing each allele value who were raised in a low-average environment, but a 12.5% difference for individuals who were raised in a medium-average environment (main effect of allele: $F(1, 516)=12.17$, $p=.001$, $\eta_p^2=.023$; main effect of environment: $F(1,516)=.55$, $p=.460$, $\eta_p^2=.001$; interaction of allele x environment: $F(1,516)=5.62$, $p=.018$, $\eta_p^2=.011$). For the non-predictive allele, there was no indication of an interaction (main effect of allele: $F(1, 516)=.06$, $p=.814$, $\eta_p^2=.000$; main effect of environment: $F(1,516)=.61$, $p=.435$, $\eta_p^2=.001$; interaction of allele x environment: $F(1,516)=.33$, $p=.565$, $\eta_p^2=.001$).

In sum, even when environment was viewed just in terms of as a manipulation on the information available, interactions between environmental variation and computational parameter variation were apparent. Moreover, traces of these interactions were observed at the lower level of the genome. Notably, verb type and developmental stage both influenced the pattern of the interaction, indicating that the effects arise from

causal processes involved in the formation of representational states in a non-linear learning system.

Mechanisms of resilience II: Pseudo gene x gene interactions

By definition, the contribution of genetic variation to computational parameters was additive. There could be, therefore, no true gene-gene interactions (e.g., of the form of dominance or epistasis). However, given that we know computational parameters are likely to interact with each other (see Method), we explored whether this phenomenon might generate the appearance of gene-gene interactions. We took two parameters, hidden units and learning rate, which we expected on computational grounds to interact (Thomas et al., submitted). Figure 18 plots the population performance for individuals split by whether they had 40 or 50 hidden units, and whether their learning rate was 0.075 or 0.125 (where a higher learning rate indicates a more plastic system). We looked at regular verb and irregular vowel-change verb performance and once more, we contrasted early and late phases in development.

==========

Insert Figure 18 here

==========

For the networks with 40 hidden units, the less plastic systems scored higher, while for 50 hidden units, the more plastic systems scored higher. The pattern appeared in both verb types but was only reliable for irregular verbs (regular, early: p=.173; late: p=.323; irregular, early: $F(1,111)=4.14$, p=.044, $\eta_p^2 = .036$; late: $F(1,111)=5.53$, p=.020, $\eta_p^2 = .047$). Once more, we explored whether this interaction was visible at the level of alleles.

We picked two alleles with predictive power on population variability, one from the hidden unit (HU) region and one from the learning rate (LR) region (effect sizes of 1.39% and 2.37%, respectively), and contrasted them with two alleles from these regions that showed no predictive power on behaviour (0.00% and 0.09%). Somewhat to our surprise, the alleles with predictive power showed main effects but no interaction (main effect of HU: $F(1,996)=13.78$, $p<.001$, $\eta_p^2 = .014$; main effect of LR: $F(1,996)=24.09$, $p<.001$, $\eta_p^2 = .024$; HU x LR interaction: $F(1,996)=1.27$, $p=.261$, $\eta_p^2 = .001$), while the alleles without individual predictive power showed no main effects but a reliable interaction (main effect of HU: $F(1,996)=.03$, $p=.871$, $\eta_p^2 = .000$; main effect of LR: $F(1,996)=1.21$, $p=.271$, $\eta_p^2 = .001$; HU x LR interaction: $F(1,996)=4.07$, $p=.044$, $\eta_p^2 = .004$). Moreover, the interaction was in the direction we were expecting: for the efficient hidden unit allele, individuals with the efficient learning rate allele scored higher; for the inefficient hidden unit allele, individuals with the inefficient learning rate allele scored higher.

The computational explanation of this interaction is that in networks with less representational capacity, a more precise combination of connection weight values must be reached to accommodate the set of mappings demanded by the training set. This solution must be approached during training in small iterative steps provided by a low learning rate. In a network with more resources, less exact weight values are necessary, and less care is therefore necessary in the adjustment of weight values; faster learning is merely developmentally advantageous. This example is illustrative. Many other such interactions between parameters are to be expected within an artificial neural network learning system. Perhaps most notably, the computational effect was observed as an interaction at the level of alleles. However, the statistical effect did not in this case index

any underlying causal interaction between the genes themselves; by stipulation, all genetic effects were additive. Moreover, the statistical interaction was observed between alleles that did not predict reliable amounts of variance individually. If such computational interactions exist in real neural systems, the statistical interactions of gene variants coding for the relevant parameters would be missed in an association analysis that only considered main effects. Of course, testing for interactions risks a combinatorial explosion of comparisons; in this case, we looked for the relevant interaction on *a priori* computational grounds.

In terms of resilience, the simulation framework provided at least one opportunity for a true effect, where variation in one dimension eliminates the influence of variation in another dimension: networks could have either direct input-output connections or an indirect pathway via a layer of hidden units. For those networks specified as two layer, genetically controlled variation in the number of units in the hidden layer was simply irrelevant for behaviour. (Since the hidden unit pathway was not activated during training in two-layer networks, it decayed quickly following the onset of pruning). Genes for 2-layer networks were therefore resilience genes that could protect against having very low hidden unit numbers. However, interactions between computational parameters provided other avenues for resilience, where variations in one parameter compensated for variations in another. Since these variations were encoded on the genome in the current framework, genes indexed individuals with greater or less resilience in this weaker sense.

Predicting the entire population variability

In our final analysis, we assessed how much of the population variability in performance could be predicted given that we know all the parameter values that contributed to this variation. For the 2x2 design, all computational parameter values along with the family quotient value for each individual were entered into a stepwise linear regression. The analyses were carried out on the four populations when their mean vowel-change irregular verb performance was matched to that of the six-year-old children. Total amounts of variance explained are depicted in Table 6. Variance explained ranged from 40% to 75%.

Why couldn't we predict all of the population variance? There were four reasons. First, the statistical model used linear relationships. We know that many of the parameter-behaviour relationships were in fact non-linear. A statistical model which used the appropriate non-linear function for each parameter would be required to explain the maximum amount of variance. Second, as we saw in the above sections, there were interactions among the computational parameters, and between the computational parameters and the training environment. When a single factor became dominant, as in the case of wide environmental variation, a greater sum of variance was explained. Third, there were stochastic factors in the networks, such as the initial weight values and pruning events (see Table 3, rightmost column). Fourth, there were factors that depended on the particular behaviour that was being measured. Table 6 indicates different amounts totals of explained variance for regular and irregular verbs. These differences likely stem from floor and ceiling effects in these measures, as well as subtler factors involving the interaction between learning events within the system.

In sum, even when we are in a position to know all the internal and external factors generating variation in population performance, there are several reasons why we can still fail to explain the observed variation in behaviour using our statistical models.

==========

Insert Table 6 here

==========

Discussion

The modelling framework set out to implement a model of one aspect of language development that linked between multiple levels of description: genome, neurocomputation, overt behaviour, and environment. The goal was to render explicit assumptions linking the three levels of description and to evaluate their consequences via implementation in a complex learning system.

One could debate whether, in using a connectionist model, we have provided the right level of analysis for linking variations in language development with lower level neurocomputational mechanisms. We might, for instance, have been guided more strongly by molecular genetics. Recent findings from the molecular genetics of language and literacy disorders have provided new insights about which aspects of brain development differ in these disorders. For instance, a considerable amount has been learned about the neural mechanisms disrupted by a mutation in the FOXP2 gene underlying a rare form of familiar dyspraxia of speech (Fisher, 2006). The expression of the CNTNAP2 gene is modulated by FOXP2; CNTNAP2 is involved in cortical development, possibly mediating intercellular interactions during neuroblast migration and laminar organisation (Vernes et al., 2008); it is expressed during cortical development in frontal grey matter regions including the inferior and middle frontal gyri (Abrahams et al., 2007); and Vernes et al. found that polymorphisms of the gene showed a significant association with behaviourally defined cases of SLI, explaining up to 4% of the variance in non-word repetition scores.

Similarly, recent candidate gene discoveries for dyslexia point towards neuronal migration and axon guidance as important neural mechanisms in that disorder, perhaps

producing subtle cortical malformations in peri-sylvian language areas that affect subsequent phonological development (Galaburda et al., 2006). By using discoveries such as these, low-level neurocomputational models of atypical language development could be constructed that are based on empirically constrained changes in brain development. The marriage of genetics and neurocomputational models should perhaps occur at the level of neural mechanisms implicated by actual gene discoveries.

The importance of such a research avenue is undeniable. The question is whether it is premature. We have presented an account at a much higher level of description, with neurocomputational properties that are abstracted away from actual neural circuits. The reason is threefold. First, our aim was to pursue a top-down approach retaining contact with the behavioural data that has motivated claims about the heritability of cognition. Second, in our view, models of language function (and development) are not yet sufficiently advanced to make it viable to construct detailed models in terms of low-level neural circuits. Structural and functional magnetic resonance imaging, connectivity analyses, dynamic causal modelling, and electrophysiological studies of brain function are making great strides, but their focus has been to identify the brain regions associated with various language tasks and the functional connectivity between them (Kuhl & Rivera-Gaxiola, 2008; Leff et al., 2008; Vannest et al., 2009; Vigneau et al., 2006). Moreover, this research has been hampered by an emerging mismatch in the ontologies of cognitive and anatomical models: classical cognitive models are being challenged because there is a high degree of overlap among the neural systems activated by tasks that share no cognitive components, suggesting a given neuronal structural can perform multiple functions that depend on the areas with which it interacts (Price & Friston,

2005). Overall, cognition-brain accounts cannot yet strongly constrain our view of how low-level differences in neural circuits would affect the high-level behaviour of these neural networks in the brain. Third, although association analyses have identified candidate genes involved in language disorders, and the function of these genes is fast being uncovered, it is another matter to identify how the *polymorphisms* in these genes modulate their function. As Galaburda et al. (2006) pointed out, none of the genetic variants consistently linked with dyslexia are mutations in the coding sequences of the genes, and it will be a challenge for future studies to identify how the polymorphisms lead to the specific changes in function in the developing and mature brain.

Given our top-down approach, one might ask how much can be read into our 'genetic' level of description. The relation of connectionist models to neurocomputation is already a fairly abstract one. In some senses, we view the genetic encoding of parameters in the model as notional. The appropriate analogy is to behavioural genetics, where genetic constraints are contained within parameters entered into structural equation models: the assumed genetic similarity between monozygotic twins (1), the genetic similarity between dizygotic twins under an additive model (0.5), and under a dominant model (0.25). In our framework, the genotypic level embodied several more detailed assumptions: genes exert their causal effect on behaviour via the modulation of general neurocomputational parameters; the relationship between genes and neurocomputational parameters is polygenic; and the effect of gene variants is mainly additive. Moreover, we created artificial chromosomes and simulated sexual reproduction to demonstrate that our scheme for encoding the genetic contribution to population level variability in cognitive

ability produced a stable distribution of ability across generations in the absence of selection (see Appendix A).

The aim of a top-down approach, however, is to provide hypotheses for the interpretation of data generated by bottom-up approaches. In providing explicit implementation of genomic, neurocomputational, and behavioural levels of description, our concern was to relate the mechanistic operation of the model to the statistical relationships observed between various predictor and variation in behaviour, given that the primary data used to identify gene-behaviour relations are statistical. This emphasis allowed us, for instance, to separate the *causal idea* that the process of development is an interaction between genes (contributing to the working of an experience-dependent learning system) and environment, from *the statistical idea* of a gene-environment interaction, where the contribution of genetic measures and environmental measures as predictors of population variance may be non-additive. Or more starkly, to distinguish the view that interaction with a structured learning environment may be causally essential for learning, from the view that variation in the quality of the environment may nevertheless statistically predict little of the variation in behaviour (as per Figure 9). As we saw in the Introduction, the link between causal models of development and the statistical models that employ genetic and environmental predictors of behavioural variation has been a source of concern for the *Developmentalists* (Gottlieb, 1995; Turkheimer, 2004).

We now turn to the findings of the model. We first consider how it fared in simulating our target empirical data. We then summarise the behaviour of the model when we moved beyond the target data. Our population data of the past tense formation in 442 6-year-old children exhibited three main features (Bishop, 2005). First, past tense

formation was more accurate for regular verbs than irregular verbs. Second, performance on regular verbs showed a highly skewed distribution with a long lower tail, while irregular verbs showed a more normal distribution, with some hint of bimodality. Third, variation in SES predicted small amounts of the variation in regular verb acquisition, and more in the acquisition of irregular verbs (around 1% and 3%, respectively). We should add some minor caveats about these empirical data. The data are not ideal for our purposes for three reasons. First, the data are from twins, which although fairly representative of a population, do exhibit some mean differences to singletons (Plomin et al., 2008). Second, the data only represent a snapshot in time and provide limited opportunity to understand how the distribution changes over time (the age range only spanned 13 months). Third, the data are over-sampled for children at risk of language impairment. Bishop (2005) used this population to investigate whether the heritability of variations in grammatical ability was different in normal versus language-impaired children. Therefore, the population includes 215 children who had been flagged as at risk for language impairment at age 4 on the basis of parental report. Notably, many of these children were not subsequently diagnosed with a language impairment at age 6, implying that some of the early detections were false positives triggered by individual differences in trajectories of early development (see Dale, Price, Bishop, & Plomin, 2003, for a similar effect when parental reports of language delay at 2 years of age were used predict language difficulties at 4 years). The sample contains 227 children who were not flagged as at risk for language impairment at age 4. Nevertheless, the data correspond to a much larger population of children than is typically assessed in psycholinguistic studies.

The model simulated genetic variation via individual differences in 14 computational parameters, and SES variation via manipulations to the composition of the past tense training set. The model design independently varied the range of genetic versus environmental variation in four simulated populations. There was no a priori basis to constrain the respective ranges of variation, so we employed a 2x2 design of wide versus narrow variation in genetic versus environment factors, thereby granting a view of the effects of the range of variation. The population with wide genetic variation and narrow environmental variation gave the best fit to the empirical data. The model succeeded in simulating all three of the main qualitative effects in the empirical data. In addition, it predicted that variation in particular computational parameters could produce sub-populations in irregular verb ability, some of which would not be observed in regular verb acquisition at any stage of development. Lastly, the model generated a novel prediction that SES would explain more of the variance in whether a child's performance fell within the top 10% of the population distribution than whether it fell in the bottom 10%. Numerically, it predicted that 2% of the variance would be explained in high performance and 0% in low performance. This novel prediction was tested on the empirical data and borne out. For the children, SES predicted 2.5% of the variance in high performance and 0% of the variance in low performance. The prediction formed a strong test of the model, since the empirical data were not collected with this analysis in mind, and the implementation of SES was (to our knowledge) the first of its kind and therefore somewhat provisional.

The model extended beyond the empirical data in two ways, first placing population distributions within a developmental context, and second simulating

associations between behaviour and, respectively, the computational parameter and the genome levels of description. The results indicated the importance of the verb type that the model was learning (e.g., the nature of the input-output mapping) for the evolution of population distributions across development, as well as the importance of measure sensitivity in producing floor and ceiling effects. The individual parameter values in each network predicted a reasonable amount of variance in behaviour (up to 20%), and the values of low-level alleles much smaller amounts (1-2%). Notably, the association sizes were to some extent dependent on the verb type, and also changed across development. Somewhat surprisingly, more sensitive measures of performance (RMS error compared to accuracy) proved harder to predict from model parameters. We considered the computational reasons for the way associations were modulated, respectively in terms of the *domain-relevance* of certain parameters for certain behaviours (Karmiloff-Smith, 1998), and in terms of the reliance of representations on different computational properties at different points in training.

More provocatively, it was possible to identify gene-behaviour associations in terms of 'genes for regular verbs' versus 'genes for irregular verbs', and in terms of 'genes for early development' versus 'genes for later development'. A computational analysis indicated that such labels would not index actual underlying causal processes in the model. Illustrative examples were then given of different types of *resilience*, where certain computational properties made the networks differently sensitive to variations in environmental quality (gene-environment interactions) or differently sensitive to variations in other computational properties (pseudo gene-gene interactions). In both cases, traces of these statistic effects were observable in direct links between gene

variants and high-level population behaviour. Lastly, we attempted to predict population variability using our complete knowledge of each individual's computational parameters and training environment, and found that at most 75% of the variance could be explained and sometimes as little 40%. Unexplained variance arose in part due to statistical limitations (use of linear methods), measurement methods (floor and ceiling effects), and causal processes (interactions between predictors).

Our goal of distinguishing statistical from causal relations in simulated gene-behaviour relations produced several interesting results. First, both empirical data and model indicated that there was no statistical relation between SES and whether a child fell in the bottom 10% of the population distribution on past tense formation. However, from the model, we know that a poor environment *did cause* poor acquisition. The reason for the absence of a statistical relation when a causal relation existed is that the relationship of cause to effect was many-to-one. In addition to an impoverished environment, there were many other possible causes of poor acquisition, resulting from the settings of computational parameters. The many-to-one causal relationship diluted the strength of the statistical association of any one cause. In a population with largely adequate computational parameters, environmental quality should be a more symmetrical predictor of success and failure, and this was confirmed by simulation of such a population (see Table 5 for the *G*-narrow-*E*-wide population). The asymmetry was therefore predicted to be sensitive to the sampling. Note our empirical data set, with its over-representation of children with language impairments may have corresponded to a sample that exaggerated genetic variance over environmental variance, thereby making the asymmetry more likely. More widely, if internal factors are more predictive of low

performance than mid-range performance, the implication is that manipulation of environment (in terms of early intervention) may be more effective for middle functioning children than low functioning children.

Second, our simulated association analyses found that only some of the alleles predicted reliable amounts of population variability in behaviour. However, we know from the model that *all alleles were causal* in determining the parameter value that led to behavioural variation. The gene variants, assigned binary values to represent efficient or inefficient alleles, were simply added together to derive the parameter value from a look-up table. The reason that some alleles produced reliable associations is that, in a given particular population, these alleles happened to contribute to a high total sum and a more beneficial parameter value, or to a low total sum and a less beneficial parameter value. In another population, other alleles for the same parameter might end up playing this role. Once more, there is absence of a statistical relation when a causal relation actually exists, and once more, it is because with polygenic encoding of parameter values at the genetic level, there is a many-to-one causal relationship. The simulations demonstrated that across re-sampling of populations with the same allele frequencies, one would *not* expect individual gene-behaviour associations to straightforwardly replicate, even in the absence of measurement error. Only the sum of associations within the alleles that coded for a given parameter was predicted to replicate. However, re-sampling of the populations did preserve the behavioural associations at the level of computational parameters. In reality, the replicability of association analyses in complex brain functioning has proved a problem (Posthuma & de Geus, 2006). This may be for quite other reasons – the methods are technically challenging, participant sampling is difficult, and real data are noisy.

Nevertheless, the simulations suggest that the strongest predictive power is via understanding the computational parameter values in a neural system, and that replication of associations will be improved by unifying gene variants around the neurocomputational parameter to whose variance those alleles contribute.[10]

Third, interactions between computational parameters produced apparent interactions between gene variants. However, we know these statistical interactions were *not causal*, because gene contributions were stipulated to be additive. The behavioural associations had been inherited by the alleles from the computational parameters whose variance they coded for. In this case, there was a statistical relationship without a causal link, via a common causal factor at a higher level of description. While we know in reality many genes do interact via the use of protein products to modulate expression (amongst other mechanisms), the current results indicate that statistical evidence for their interaction can also be produced without any causal interaction at the level of the genome.

The advantage of computational implementations is that they render assumptions explicit. The disadvantage is that by definition they involve simplifications. It is necessary, therefore, to be clear about which aspects of the modelling framework are

[10] We have stipulated the granularity of the genomic encoding by virtue of our assumption of a polygenic relationship between genes and computational parameters. However, one could take a different view: that the 1s and 0s of the genome correspond to 'base pairs' and the regions for each parameter correspond to the 'genes'. This view would predict much stronger associations between gene variants and behaviour, since each polymorphism would code for a computational parameter value. And it would predict greater replicability across association studies for whole genes but potentially lower replicability for associations between single nucleotide polymorphism (SNPs) and behaviour.

central to its functioning and which are merely simplifications necessary for implementation. We have already summarised the genome level assumptions. Here is a summary of the simplifications: no pleiotropy between genes and parameters, no epigenetic effects, no dominant or recessive genes, no epistasis, no gender effects, no linkage disequilibrium, no assortative mating, and no gene-environment correlations. The impact of each simplification must form the target for future work.

At the cognitive level, we view the following as key assumptions. Past tense represents a mapping problem between (at least) codes for phonological information corresponding to different forms of the verb. It is learned by an associative system using distributed computation among a network of simple processing units. These units are generic but gain domain-specificity via the architecture of which they are a part. A child's experience of past tense involves learning associations between uninflected and inflected forms of the verb, although these need not be simultaneously available in the environment. The English past tense is defined by particular patterns of similarity, type frequency, and token frequency. We believe that a variety of different implementations of past tense models that embodied these assumptions would produce the same simulation results. (Indeed our preferred model of this system is somewhat more complex, involving the convergence of more codes that simply phonological; Karaminis & Thomas, 2009).

However, we are more provisional about our assumptions regarding sources of environmental variation. We reviewed existing literature on the role of environmental variation in past tense acquisition, including an argument that it plays no role and acquisition simply reflects maturation (Rice et al., 1998). Research considered both the specific availability of past tense information in the environment in families of different

social status (e.g., Hart & Risley, 1995), and recent research on the widespread impact of SES on brain development. Implementation forces us to make a decision about how environment will affect acquisition. We chose to simulate the effects of SES on language acquisition only via a modulation of the information content available to the network, and via type rather than token manipulations to the training set with respect to verb frequency. The family training set was fixed and its composition was not dependent on the activity of the simulated child. Any of these assumptions might be revisited. Importantly, we did not allow environment to alter the computational parameters of the network. This simplification is certainly debatable, given what is now known about the effect of stress on brain function, and the possible contribution of dietary differences experienced in poverty (Hackman & Farah, 2009). We suspect that whether the environment impacts primarily on information content or on biological (non-cognitive) aspects of brain function will depend on the range of environmental variation involved. In the first world, even in the face of relative poverty, there is some minimal provision for the healthy upbringing of children. From the point of view of cognitive development, environmental variation may impact more on the information available, and on the particular schedules of reward and punishment experienced by a child. However, in the third world, the environmental range is much wider. Nutritional deficits during child development can be severe enough to cause stunting in growth and a statistically associated incidence of poor cognitive development (Grantham-McGregor et al., 2007). In this wider range, environmental variation may impact much more on biological aspects of neural function and therefore its computational properties. More work is needed to advance our understanding of how environmental variation impacts on cognitive

development at a neurocomputational level, and how specifically it affects the content and reward-based information available to the child.

Apart from the importance of implemented causal models, the modelling framework emphasised two ideas in particular. The first is the possibility of accounting for developmental processes and intelligence within a single framework. Variability across the normal range and at the tails was explained by continuous sources of variation, in implemented learning systems. Given the causally homogeneous source of variability across the range, the simulations still gave rise to different patterns in the tails, for example, the stronger predictive power of the environment in high performing individuals, consonant with emergenic theories of giftedness (e.g., Lykken, 1982, 2006; Simonton, 2005).[11] In the emergenic theory, everything has to be good for giftedness to result. If the environment is not good, then there will be no giftedness. Hence, environment has more unique predictive statistical power. This framework is also consistent with the view that, in the absence of overt environmental incidents or genetic mutations, the 'abnormal is normal' in terms of causal process (Kovas et al., 2007; Plomin et al., 2008). However, in principle, the modelling framework allows us to contrast the distributions generated by polygenic assumptions with those produced by single gene mutations producing large effects, and to explore how the frequency of such mutations might affect population distribution. Notably, with respect to intelligence research, the polygenic view of 'ability' in the current framework translates to the idea

[11] For reasons of space, we did not report on the power of intrinsic factors to predict membership of the tails. The patterns of asymmetry were more mixed and depended on the population and particular computational parameter involved.

that what a network can learn is the summation (and interaction) of the simultaneous variation of many computational parameters. Together, these alter the effective capacity and plasticity of each system with respect to learning. No single factor is responsible for variation in ability at a computational level. However, this conception of ability is entirely neutral to the specificity or generality of the effects. We have taken no view as to whether the genetic effects on parameters are specific to a particular neurocomputational system or are wide ranging; that is, whether the Hidden Unit genes, for example, would determine variation in computational resources just in the inflectional system or in other systems as well (Anderson & Nelson, 2005, for a consideration of general vs. specific intelligence in computational terms). We have made the assumption that genetic factors are at least general to all parts of the network we are studying, but that assumption too may be challenged. It is possible that genes code more specifically for neurons in different parts of the same network, for instance, in different layers.

The second issue highlighted by the current framework is the importance of taking a developmental perspective in considering the link between genes and behaviour. Simulation results suggested that each of the following effects was modulated by development: the population distribution, the predictive power of the environment, the relation between computational parameters and behaviour, associations between gene variants and behaviour, and gene-environment interactions. This should be informative for future empirical research into individual differences in the development of behaviour. Moreover, the effects of development were comprehensible in terms of the changing computational states of the learning system.

As we have seen, the modelling of causal systems urges caution in simple interpretations of statistical associations. In the current context, when gene-behaviour associations increased in strength across development, this did not reflect an increase in gene expression per se but a greater reliance on a given computation property whose variance the gene variant coded for. Nevertheless, the simulation framework did include genetic effects that *were* time sensitive, in the form of connection pruning parameters with late onset. These parameters, by contrast, showed relatively modest changes in their power to predict behavioural variance over developmental time.

It is a challenging goal to formulate accounts of variation in human behaviour that span multiple levels of description, from genome through neurocomputation, via cognition, to behaviour; yet more challenging to construe this variation in the context of the development of behaviour in children embedded in a rich environmental context. The current simulations take but a small step in this direction but, we think, indicate there is much to be gained in using implemented computational models to understand the interaction of genes and the environment in the development of behaviour.

Acknowledgements

We thank Dorothy Bishop for allowing access to her data set, and for her help and encouragement throughout this project. We would also like to thank Eddy Davelaar, Emma Hayiou-Thomas, Mark Johnson, Annette Karmiloff-Smith, Denis Mareschal, Robert Plomin, Mabel Rice, and Essi Viding, for helpful discussions and ideas. This research was supported by MRC Career Establishment Grant G0300188, British Academy grants LRG-35369 and SG-040400, and EC grant 0209088 (NEST).

Appendix A

The stability of genetic variation across generations

It may be self-evident to population geneticists that under the conditions we implemented, the genetic variability across generations must be stable in the absence of selection. However, it is perhaps less obvious to developmental psychologists. This is because psychologists are familiar with the idea of *regression to the mean* when measuring behavioural variability. When outlier behavioural scores are re-sampled, they are statistically more likely to fall closer to the mean. One might then reason as follows. Assume a normal distribution of cognitive ability encoded by genes and random mating between individuals at each generation. In terms of ability, it is statistically more likely that an individual with above average ability will breed with an individual who is less able, thereby producing more average children. Similarly, statistically, a below average individual will be more likely to breed with an individual who is more able, again producing more average children. Over generations, one might therefore expect the genetic contribution to population variability to narrow around the mean. We ran a brief set of simulations to demonstrate that such regression to the mean does not occur, and that our proposed mechanisms of encoding and breeding retained a stable range of genetic variation across generations.

First, let us demonstrate what happens to the population variability when there is a measure of fitness and when selection operates. Each individual had a genome of 126 bits. The genes came in two versions, more (1) or less (0) efficient. Let us assume that there is a direct mapping between the number of 1s in the genome and the fitness of the individual. We initially generated a population of 1000 individuals whose genomes were

set at random, so that for each gene, each allele had a 50% chance of being present (i.e., the gene was set to 1 or 0 with a probability of 0.5). Individuals then mated at random, using the mechanism outlined in the previous section. The first generation mated without selection. Thereafter, to generate each subsequent generation, the 100 fittest individuals in the current population (i.e., with the most 1s), were then designated as the breeding population. Reproduction occurred with uniform crossover and initially no mutation. The breeding population mated at random, each pair producing two offspring, in five iterations to produce 1000 individuals for the next generation. The whole process was iterated across 25 generations. Figure 3a depicts the distribution of the number of 1s in the genome in the original population and over 25 generations with selection in operation. It is evident that variability reduced, in particular as it approached the ceiling value of 126 1s. However, the reduction in variability under selective pressure occurred even before ceiling was reached.

Next, we repeated this exercise, now including a high mutation rate of 10%: in each gamete, each allele was stochastically flipped into the alternate state with a probability of 0.1. Figure 3b shows the same narrowing of population variability across generations. However, the population no longer tended towards the ceiling value of 126 1s. Instead, it asymptoted at a level where the fine-scale changes required to improve performance (by weeding out individuals with the remaining few 0s) were countered by mutations during reproduction (flipping 1s back to 0s). As it were, the 'signal' of selective fitness was transmitted across generations until a point at which further enhancements fell below the threshold of 'noise' caused by mutation during reproduction.

In this condition, then, some variability was retained in the population because a mechanism (mutation) replenished it even as selection reduced it.

Next, we re-ran the first mutation-free simulation but now removing selection. Breeding took place between all individuals in the population irrespective of fitness: 500 random pairs, two offspring per pair. Figure 3c demonstrates the resulting population distributions of allele frequencies over the 25 generations. Two results are evident from this figure. In the absence of selection, the distribution of variability remained constant across generations. Second, there was nevertheless some sampling noise between generations (technically referred to as 'genetic drift'). Regression to the mean did not occur because at the population level, subsequent generations merely represented reshuffling of the same set of gene variants between different individuals, with some sampling noise during breeding. Any regression to the mean was balanced by progression from the mean to retain the distribution.

==========

Insert Figure A1 here

==========

References

Abrahams, B. S., Tentler, D., Perederiy, J. V., Oldam, M. C. Coppola, G., & Geschwind, D. H. (2007). Genome-wide analyses of human perisylvian cerebral cortical patterning. *Proceedings of the National Academy of Sciences USA*, *104*, 17849-54.

Abrahams, B. S., & Geschwind, D. H. (2008). Advances in autism genetics: on the threshold of a new neurobiology. *Nature Reviews Genetics*, *9(5)*, 341-355.

Anderson, M. & Nelson, J. (2005). Individual differences and cognitive models of mind: Using the differentiation hypothesis to distinguish general and specific cognitive processes. In J. Duncan, L. Phillips, & P. McLeod (Eds.), *Measuring the mind: Speed, control, and age* (pp. 89-113). Oxford: Oxford University Press.

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157–193). Hillsdale, NJ: Erlbaum.

Baughman, F. D. (2009). Empirical and computational investigations of the relationship between intelligence and development: mental-age matching studies of cognitive variability in the normal range. *Unpublished Doctoral Thesis*. University of London, UK.

Baumgardner, T. L., Green, K. E., & Reiss, A. L. (1994). A behavioural neurogenetics approach to developmental disabilities: gene-brain-behavior associations. *Current Opinion in Neurology*, *7(2)*, 172-178.

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(4)*, 479-490.

Bullinaria, J.A. (1995). Modelling Reaction Times. In: L.S. Smith & P.J.B. Hancock (Eds.), *Neural Computation and Psychology*, (pp. 34-48). London: Springer.

Bullinaria, J.A. (2005). Evolved Age Dependent Plasticity Improves Neural Network Performance. In: *Proceedings of the Fifth International Conference on Hybrid Intelligent Systems* (HIS 2005), (pp. 79-84). Piscataway, NJ: IEEE.

Bullinaria, J.A. (2007). Using Evolution to Improve Neural Network Learning: Pitfalls and Solutions. *Neural Computing & Applications*, *16*, 209-226.

Bullinaria, J.A. (2009). The Importance of Neurophysiological Constraints for Modelling the Emergence of Modularity. In: D. Heinke & E. Mavritsaki (Eds), *Computational Modelling in Behavioural Neuroscience: Closing the Gap Between Neurophysiology and Behaviour*, (pp. 187-208). Hove, UK: Psychology Press.

Caspi, A., McClay, J., Moffitt, T., Mill, J., Martin, J., Craig, I. W., Taylor, A., & Poulton, R. (2002). Role of genotype in the cycle of violence in maltreated children. *Science*, *297*, 851-854.

Caspi, A., Moffitt, T. E., Morgan, J., Rutter, M., Taylor, A., Arseneault, L., Tully, L, Jacobs, C., Kim-Cohen, J., & Polo-Tomas, M. (2004). Maternal expressed emotion predicts children's antisocial behavior problems: Using monozygotic-twin differences to identify environmental effects on behavioral development. *Developmental Psychology*, *40(2)*, 149-161.

Caspi, A., Sugden, K., Moffitt, T., Taylor, A., Craig, I. W., Harrington, H., McClay, J., Mill, J., Martin, J., Braithwaite, A., & Poulton, R. (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science*, *301*, 386-389.

Chalmers, D. (1990). The evolution of learning: An experiment in genetic connectionism. In D. Touretzky, J. Elman, T. Sejnowski, & G. Hinton (Eds.), *Connectionist models: Proceedings of the 1990 summer school* (pp. 81-90). San Mateo, CA: Morgan Kaufmann Publishers.

Cicchetti, D., & Blender, J. A. (2006). A multiple-levels-of-analysis perspective on resilience: Implications for the developing brain, neural plasticity, and preventive interventions. *Ann. N. Y. Acad. Sci.*, *1094*, 248-258.

D'Angiulli, A., Herdman, A., Stapells, D., & Hertzman, C. (2008). Children's event-related potentials of auditory selective attention vary with their socioeconomic status. *Neuropsychology*, *22(3)*, 293-300.

Dale, P. S., Price, T. S., Bishop, D. V. M., & Plomin, R. (2003). Outcomes of early language delay: 1. Predicting persistent and transient language difficulties at 3 and 4 years. *Journal of Speech, Language, and Hearing Research*, *46*, 544-560.

Davis, H. & Anderson, M. (1999). Individual differences and development – one dimension or two? In M. Anderson (Ed.), *The development of intelligence*, (pp. 161-191). Hove: Psychology Press.

Deary, I. J., Spinath, F. M., & Bates T. C. (2006). Genetics of intelligence. *European Journal of Human Genetics*, *14(6)*, 690-700.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, Mass.: MIT Press.

Farah, M. J., Shera, D. M., Savage, J. H., Betancourt, L., Giannetta, J. M., Brodsky, N. L., Malmud, E. K., & Hurt, H. (2006). Childhood poverty: specific associations with neurocognitive development. *Brain Research*, *1110*, 166-174.

Fisher, S. E. (2006). Tangled webs: Tracing the connections between genes and cognition. *Cognition*, *101*, 270-297.

Fisher, S. E., & Francks, C. (2006). Genes, cognition and dyslexia: learning to read the genome. *Trends in Cognitive Sciences*, *10(6)*, 250-257.

French, R. M. & Messinger, A. (1994). Genes, Phenes and the Baldwin Effect: Learning and Evolution in a Simulated Population. In R. Brooks & P. Maes (eds.) *Artificial Life IV*, (pp. 277-282). MIT Press: Cambridge, MA.

Fromkin, V. & Rodman, R. (1988). *An introduction to language (4th Ed.)*. Holt, Rinehart and Winston, Inc.: London.

Galaburda, A. M., LoTurco, J., Ramus, F., Fitch, R. H. & Rosen, G. D. (2006). From genes to behavior in developmental dyslexia. *Nature Neuroscience*, *9*, 1213-1217.

Garlick, D. (2002). Understanding the nature of the general factor of intelligence: The role of individual differences in plasticity as an explanatory mechanism. *Psychological Review*, *109(1)*, 116-136.

Ginsborg, J. (2006). The effects of socio-economic status on children's language acquisition and use. In J. Glegg & J. Ginsborg (Eds.), *Language and social disadvantage: Theory into practice*. John Wiley & Sons Ltd.

Gizer, I. R., Ficks, C., Waldman, I. D. (2009). Candidate gene studies of ADHD: a meta-analytic review. *Human Genetics*, *126(1)*, 51-90.

Goldman-Rakic, P. S., Muly, E. C., III, & Williams, G. V. (2000). D1 receptors in prefrontal cells and circuits. *Brain Research Reviews*, *31*, 295-301.

Gottlieb, G. (1995). Some conceptual deficiencies in "developmental" behavior genetics. *Human Development*, *38*, 131-141.

Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L. et al. (2007). Child development in developing countries 1: Developmental potential in the first 5 years for children in developing countries. *The Lancet*, *369*, 60-70.

Hackman, D. A., & Farah, M. J. (2009). Socioeconomic status and the developing brain. *Trends in Cognitive Sciences*, *13(2)*, 65-73.

Halliburton, R. (2004). *Introduction to population genetics*. Pearson: Atlanta, GA.

Harlaar, N., Butcher, L. M., Meaburn, E., Sham, P., Craig, I. W., & Plomin, R. (2005). A behavioural genomic analysis of DNA markers associated with general cognitive ability in 7-year-olds. *Journal of Child Psychology and Psychiatry*, *46(10)*, 1097-1107.

Harm, M. & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, *106*, 491-528.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, Maryland: Paul H. Brookes Publishing Co.

Hart, S. A., Petrill, S. A., Deckard, K. D. & Thompson, L. A. (2007). SES and CHAOS as environmental mediators of cognitive ability: A longitudinal genetic analysis. *Intelligence*, *35*, 233-242.

Hartshorne, J. K., & Ullman, M. T. (2006). Why girls say 'holded' more than boys. *Developmental Science*, *9:1*, 21-32.

Hinton, G. (1989). Connectionist learning procedures. *Artificial Intelligence*, *40*, 185–234.

Hinton, G. E., & Nowlan, S. J. (1987). How learning can guide evolution. *Complex Systems*, *1*, 495-502.

Hoeffner, J. H. & McClelland, J. L. (1993). Can a perceptual processing deficit explain the impairment of inflectional morphology in developmental dysphasia? A computational investigation. In E.V. Clark (Ed), *Proceedings of the 25th Child language research forum*, (pp. 1-25). Stanford University Press.

Hoff, E. & Naigles, L. R. (2002). How children use input to acquire a lexicon. *Child Developmen*t, *73(2)*, 418-433.

Huttenlocher, J., Vasilyeva, J., Cymerman, E., & Levine, S. (2002). Language input and child syntax. *Cognitive Psychology*, *45*, 337-374.

Huttenlocher, P. R. (2002). *Neural plasticity: The effects of environment on the development of the cerebral cortex*. Cambridge, MA: Harvard University Press.

Jaffee, S. R., Caspi, A., Moffitt, T. E., Dodge, K. A., Rutter, M., Taylor, A., & Tully, L. A. (2005). Nature X nurture: Genetic vulnerabilities interact with physical maltreatment to promote conduct problems. *Development and Psychopathology*, *17*, 67-84.

Joanisse, M. F. (2000). *Connectionist Phonology*. Unpublished Ph.D. Dissertation, University of Southern California.

Joanisse, M.F. (2004). Specific language impairments in children: Phonology, semantics and the English past tense. *Current Directions in Psychological Science*, *13(4)*, 156-160.

Joanisse, M. F., Manis, F. R., Keating, P., & Seidenberg, M. S. (2000). Language Deficits in Dyslexic Children: Speech Perception, Phonology and Morphology. *Journal of Experimental Child Psychology*, *77*, 30-60.

Johnston, T. D. & Lickliter, R. (2009). A developmental systems theory perspective on psychological change. In J. P. Spencer, M. S. C. Thomas & J. L. McClelland (Eds.), *Toward a unified theory of development* (p.285-296). Oxford: Oxford University Press.

Kan, K. J., Ploeger, A. , Raijmakers M. E. J., Dolan, C. V., & van der Maas, H. L. J. (in press). Nonlinear epigenetic variance: Review and simulations. *Developmental Science*.

Karaminis, T. & Thomas, M. S. C. (2009). The Multiple Inflection Generator: A generalized developmental model of inflectional morphology. Poster presented at the *Biennial Meeting of the Society for Research in Child Development*, Denver, Co., 2-4 April 2009.

Kasabov, N. & Benuskova, L. (2004). Computational neurogenetics. *Journal of Computational and Theoretical Nanoscience*, 1, 47-61.

Kelleher, R. J., & Bear, M. F. (2008). The autistic neuron: Troubled translation? *Cell*, *135(3)*, 401-406.

Klibanoff, R. S., Levine, S. C., Huttenlocher, J., Vasilyeva, M., & Hedges, L. V. (2006). Preschool children's mathematical knowledge: The effect of teacher "Math Talk". *Developmental Psychology*, *42(1)*, 59-69.

Kovas, Y., & Plomin, R. (2006). Generalist genes: implications for the cognitive sciences. *Trends in Cognitive Sciences*, *10(5)*, 198-203.

Kovas, Y., Haworth, C.M.A., Dale, P.S., & Plomin, R. (2007). The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monographs of the Society for Research in Child Development*, *Volume 72*, *whole number 3*, Serial No. 188, pp. 1-144.

Kuhl, P. K. & Rivera-Gaxiola, M. (2008). Neural substrates of early language acquisition. *Annual Review of Neuroscience, 31*, 511-534.

Leff, A. P., Schofield, T. M., Stephan, K. E., Crinion, J. T., Friston, K. J., & Price, C. J. (2008). The cortical dynamics of intelligible speech. The Journal of Neuroscience, 28(49): 13209-13215.

Lewis, J. D., & Elman, J. L. (2008). Growth-related neural reorganization and the autism phenotype: a test of the hypothesis that altered brain growth leads to altered connectivity. *Developmental Science*, *11(1)*, 135-155

Lykken, D. T. (1982). Research with twins: The concept of emergenesis. *Psychophysiology*, *19*, 361-373.

Lykken, D. T. (2006). The mechanism of emergenesis. *Genes, Brain, and Behavior*, *5*, 306-310.

Marchman, V. A. (1993). Constraints on plasticity in a connectionist model of English past tense. *Journal of Cognitive Neuroscience*, *5(2)*, 215-234.

Marcus, G. F., & Rabagliati, H. (2006). Genes and domain specificity. *Trends in Cognitive Sciences*, *10(9)*, 397-398.

Marcus, G., Pinker, S., Ullman, M., Hollander, J., Rosen, T. & Xu, F. (1992). Overregularisation in language acquisition. *Monographs of the Society for Research in Child Development*, *57* (Serial No. 228).

Mareschal, D. & Thomas M. S. C. (2007) Computational modeling in developmental psychology. *IEEE Transactions on Evolutionary Computation (Special Issue on Autonomous Mental Development)*, *11(2)*, 137-150.

Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill: New York.

Nakisa, R. & Plunkett, K. (1998). Evolution of a rapidly learned representations for speech. *Language and Cognitive Processes*, *13*, 105-127

Nolfi, S., Elman, J. L., & Parisi, D. (1994). Learning and evolution in neural networks. *Adaptive Behavior*, *3*, 5-28.

Petrill, S. A., Pike, A., Price, T. & Plomin, R. (2004). Chaos in the home and socioeconomic status are associated with cognitive development in early childhood: Environmental mediators identified in a genetic design. *Intelligence*, *32*, 445-460.

Pinker, S. (1999). *Words and rules: The ingredients of language*. Basic Books, New York.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. E. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*, 56-115.

Plomin, R., & Spinath, F. M. (2002). Genetics and general cognitive ability (*g*). *Trends in Cognitive Sciences*, *6*, 169-176.

Plomin, R. & Crabbe, J. C. (2000). DNA. *Psychological Bulletin*, *126*, 806-828.

Plomin, R. (1999). Genetics and general cognitive ability. *Nature*, *402 Supp*, C25-C29.

Plomin, R., & Kovas, Y. (2005). Generalist genes and learning disabilities. *Psychological Bulletin*, *131*, 592–617.

Plomin, R., Ashbury, K., & Dunn, J. (2001). Why are children in the same family so different? Nonshared environment a decade later. *Canadian Journal of Psychiatry*, *46*, 225-233.

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

Plomin, R., Owen, M. J. & McGuffin, P. (1994). The genetic basis of complex human behaviours. *Science*, *264*, 1733 -1739.

Plunkett, K. & Marchman, V. (1991). U-shaped learning and frequency effects in a multilayered perceptron: Implications for child language acquisition. *Cognition*, *38*, 1-60.

Plunkett, K. & Marchman, V. (1993). From rote learning to system building: acquiring verb morphology in children and connectionist nets. *Cognition*, *48*, 21-69.

Plunkett, K. & Marchman, V. (1996). Learning from a connectionist model of the English past tense. *Cognition*, *61*, 299-308.

Posner, M. I., Rothbart, M. K., & Sheese, B. E. (2007). Attention genes. *Developmental Science*, *10:1*, 24-29.

Posthuma, D., & de Geus, E. J. C. (2006). Progress in the molecular genetic study of intelligence. *Current Directions in Psychological Science*, *15(4)*, 151-155.

Price, C. J., & Friston, K. J. (2005). Functional ontologies for cognition: The systematic definition of structure and function. *Cognitive Neuropsychology*, *22(3/4)*, 262-275.

Raizada, R. D. S., Richards, T. L., Meltzoff, A., & Kuhl, P. K. (2008). Socioeconomic status predicts hemispheric specialisation of the left inferior frontal gyrus in young children. *NeuroImage*, *40*, 1392-1401.

Rice, M. L. & Wexler, K. (1996). A phenotype of specific language impairment: Extended optional infinitives. In M. L. Rice (Ed.), *Toward a genetics of language* (pp. 215-237). Mahwah, NJ: Lawrence Erlbaum.

Rice, M. L., & Wexler, J. (2001). *Rice/Wexler Test of Early Grammatical Impairment*. San Antonio: Psychological Corporation.

Rice, M. L., Wexler, K., & Hershberger, S. (1998). Tense over time: The longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *41*, 1412-1431.

Richardson, F., & Thomas, M. S. C. (2006). The benefits of computational modelling for the study of developmental disorders: Extending the Triesch et al. model to ADHD. *Developmental Science*, *9(2)*, 151-155.

Ring, M., & Clahsen, H. (2005). Morphosyntax in Down's syndrome: Is the Extended Optional Infinitive hypothesis an option? *Stem-, Spraak- en Taalpathologie*, *13(1)*, 3-13.

Ronald, A., Happé, F., Price, T. S., Baron-Cohen, S., & Plomin, R. (2006). Phenotypic and genetic overlap between autistic features at the extremes of the general population. *Journal of the American Academy of Child & Adolescent Psychiatry*, *45*, 1206-1214.

Rumelhart, D. E. and McClelland, J. L. (1986). On learning the past tense of English verbs. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 2: Psychological and Biological Models* (pp. 216-271). Cambridge, MA: MIT Press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland and The PDP Research Group, *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations* (pp. 318-362). Cambridge, MA: MIT Press.

Sapolsky, R. (2005). *Biology and human behavior: The neurological origins of individuality (2nd Ed.)*. Chantilly, VA: The Teaching Company.

Scerif, G. & Karmiloff-Smith, A. (2005). The dawn of cognitive genetics? Crucial developmental caveats. *Trends in Cognitive Sciences*, *9(3)*, 126-35.

Schmidt, L. A., Fox, N. A., Perez-Edgar, K., Hu, S., & Hamer, D. H. (2001). Association of DRD4 with attention problems in normal childhood development. *Psychiatric Genetics*, *11*, 25-29.

Schwab, S. G., & Wildenauer, D. B. (2009). Update on key previously proposed candidate genes for schizophrenia. *Current Opinions in Psychiatry*, *22(2)*, 147-153.

Simonton, D. K. (2005). Genetics of giftedness: The implications of an emergenic-epigenetic model. In R. J. Sternberg & J. Davidson (Eds.), *Conceptions of giftedness (2nd Ed.)*, (pp. 312-326). New York: Cambridge University Press.

Smith, S. D. (2007). Genes, language development, and language disorders. *Mental Retardation and Developmental Disabilities Research Reviews*, *13(1)*, 96-105.

Spencer, J., Thomas, M. S. C., & McClelland, J. L. (2009). *Toward a new unified theory of development: Connectionism and dynamical systems theory re-considered*. Oxford: Oxford University Press.

Stevens, C., Lauinger, B. and Neville, H. (in press). Differences in the neural mechanisms of selective attention in children from different socioeconomic backgrounds: An even-related brain potential study. *Developmental Science*.

Strachan, T., & Read, P. (2003). *Human molecular genetics 3*. Garland Publishing.

Thomas, M. S. C. (2005). Characterising compensation. *Cortex*, *41(3)*, 434-442.

Thomas, M. S. C. (2008). Ageing, plasticity, and cognitive reserve in connectionist networks. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 2089-2094). Austin, TX: Cognitive Science Society.

Thomas, M. S. C., Grant, J., Barham, Z., Gsodl, M., Laing, E., Lakusta, L., Tyler, L. K., Grice, S., Paterson, S. & Karmiloff-Smith, A. (2001). Past tense formation in Williams syndrome. *Language and Cognitive Processes*, *16 (2/3)*, 143-176.

Thomas, M. S. C. & Johnson, M. H. (2006). The computational modelling of sensitive periods. *Developmental Psychobiology*, *48(4)*, 337-344.

Thomas, M. S. C. & Karmiloff-Smith, A. (2003a). Connectionist models of development, developmental disorders and individual differences. In R. J. Sternberg, J. Lautrey, & T. Lubart (Eds.), *Models of Intelligence: International Perspectives*, (p. 133-150). American Psychological Association.

Thomas, M. S. C. & Karmiloff-Smith, A. (2003b). Modelling language acquisition in atypical phenotypes. *Psychological Review*, *Vol. 110*, *No.4*, 647-682.

Thomas, M. S. C., & McClelland, J. L. (2008). Connectionist models of cognition. In R. Sun (Ed.), *Cambridge handbook of computational cognitive modelling* (pp. 23-58). Cambridge: Cambridge University Press.

Thomas, M. S. C., Forrester, N. A., & Ronald, A. (2009). *Modelling heritability in language development: A simulated twin study of past tense acquisition in a population of neural networks*. Manuscript in preparation.

Thomas, M. S. C., Richardson, F. M., Forrester, N. A., & Baughman, F. D. (submitted). *Modelling individual variability in cognitive development*. Manuscript submitted for publication.

Triesch, J., Teuscher, C., Deak, G., & Carlson, E. (2006). Gaze following: Why (not) learn it? *Developmental Science*, *9(2)*, 125-147.

Turkheimer, E. & Waldron, M. (2000). Nonshared environment: A theoretical, methodological and quantitative review. *Psychological Bulletin*, *126*, 78-108.

Turkheimer, E. (2004). Spinach and Ice Cream: Why Social Science Is So Difficult. In. L. DiLalla (Ed). *Behavior genetics principles: Perspectives in development, personality, and psychopathology*. (pp. 161-189). Washington, DC, US: American Psychological Association.

Turney, P. (1996). Myths and legends of the Baldwin effect. In *Proceedings of the Workshop on Evolutionary Computing and Machine Learning at the 13th International Conference on Machine Learning*, p. 135-142.

Thomas, M. S. C., Grant, J., Barham, Z., Gsodl, M., Laing, E., Lakusta, L., Tyler, L. K., Grice, S., Paterson, S. & Karmiloff-Smith, A. (2001). Past tense formation in Williams syndrome. *Language and Cognitive Processes*, *16 (2/3)*, 143-176.

Van der Maas, H. L. J., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The

positive manifold of intelligence by mutualism. *Psychological Review*, *113(4)*, 842-861.

Vannest, J., Karunanayaka, P. R., Schmithorst, V. J., Szaflarski, J. P., & Holland, S. K. (2009). Language networks in children: evidence from functional MRI studies. *AJR Am J Roentgenol*. May;192(5):1190-6.

Vasilyeva, M., Huttenlocher, J., & Waterfall, H. (2006). Effects of language intervention on syntactic skill levels in preschoolers. *Developmental Psychology*, *42(1)*, 164-174.

Vernes, S. C., Newbury, D. F., Abrahams, B. S., Winchester, L., Nicod, J., Groszer, M., Alarcón, M., Oliver, P. L., Davies, K. E., Geschwind, D. H., Monaco, A. P., & Fisher, S. E. (2008). A functional genetic link between distinct developmental language disorders. *N. Engl. J. Med.* 359: 2337-2345

Vigneau, M., Beaucousin, V., Hervé, P. Y., Duffau, H., Crivello, F., Houdé, O., Mazoyer, B., & Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage*. 2006 May 1;30(4):1414-32.

Westermann, G. (1998). Emergent modularity and U-shaped learning in a constructivist neural network learning the English past tense. In *Proceedings of the 20th Annual Meeting of the Cognitive Science Society* (pp. 1130-1135). Erlbaum, Hillsdale, NJ.

Tables

Table 1. Research questions

Population distribution:

- For performance on a cognitive task, how does the distribution of a population change across development?
- How does the population distribution depend on the problem type / specific behaviour within the cognitive domain under study (e.g., regular verbs versus irregular verbs), and the type of measure being used to assess performance (e.g., accuracy versus error)?

Predicting variance from internal factors (genetic variants and neurocomputational parameters):

- How much behavioural variation is explained by particular neurocomputational parameters?
- Do parameters symmetrically predict high performance and low performance?
- How much behavioural variation is explained by particular genetic variants?
- Are some genetic variants more closely associated with some parts of the problem domain than others (e.g., regular verbs vs. irregular verbs) in an otherwise undifferentiated learning system?
- Do the above effect sizes change across development? That is, to what extent should be expect association analyses to be developmentally sensitive?

Predicting variance from the environment:

- What proportion of behavioural variability is predicted by the quality of the environment?

- Does the environment play a symmetrical role in predicting the tails of the distribution, for example in producing giftedness versus delay?

Interactions:

- What is the role of the environment in permitting the expression of genetic variability (gene x environment interactions)?
- Do gene-environment interactions depend on problem type (e.g., regular verbs versus irregular verbs)?
- Are gene-environment interactions stable across development?
- Since genes code for variability at the neurocomputational level, do interactions between neurocomputational parameters during development create the appearance of gene-gene interactions?

General:

- What is the relationship between average development, individual differences, developmental disorders, and giftedness at a mechanistic level?

Table 2. Example look-up table for deriving a computational parameter from the number of efficient alleles in the gene set determining variation. In this example, 5 genes (two alleles per gene) specify the temperature parameter, for the condition of wide genetic variation.

	Temperature parameter value										
Number of efficient alleles	0	1	2	3	4	5	6	7	8	9	10
Population probability	0.1%	1%	4%	12%	21%	25%	21%	12%	4%	1%	0.1%
Parameter value	0.0625	0.125	0.25	0.5	0.75	1	1.25	1.5	2	3	4

Table 3. Implemented sources of variation

Non-varying factors	*Varying Factors*		
	Genetically determined parameters	*Shared environmental factors* [a]	*Unique environmental factors* [a]
Processing units	Network architecture	Family quotient value	Initial weight values
Connections	Resources (internal units)	(determines family training	Initial weight connectivity
Activation dynamics	Sparseness of connectivity	set)	Initial unit thresholds
Adaptive processes	Range of initial weight		Order of learning experiences
	variance		On-line processing noise
Phonological input and	Rate of weight decay		Probabilistic connection
output	Unit discriminability		pruning
representations	(activation function)		
Exposure to relevant	Pruning onset		Initial pre-conditioning
training patterns	Pruning threshold		training set (subjective
(past tense)	Pruning probability		experiences)
Composition of			
'perfect' training set			
	Learning algorithm		*Measurement error*
	Learning rate		
	Learning momentum		
	Level of processing noise		
	Response threshold		

[a] According to the standard definitions of behavioural genetics, *shared environmental factors* are those (environmental) factors that serve to make individuals raised in the same environment (e.g., family) similar. *Unique environmental factors* are those that serve to make individuals raised in the same environment (family) different.

Table 4. Main simulation design. Percentages indicate range of family quotient values sampled (see text for details).

		Environmental variation	
		Narrow (60% - 100%)	Wide (0% - 100%)
Genetic variation	Narrow	1000 individuals	1000 individuals
	Wide	1000 individuals	1000 individuals

Table 5. Role of the environment in predicting giftedness or delay. Figures show proportion of variance explained by SES (empirical data) or family quotient value (simulation data) in whether an individual's performance falls in the top 10% or bottom 10% of the population (simulation: *G*-wide-*E*-narrow condition, irregular vowel-change verbs, matched for 6-year-old performance level).

		Environmental variation			
		Narrow		Wide	
		Delayed	Gifted	Delayed	Gifted
Simulation					
Genetic variation	Narrow	0.0%	5.2%	16.1%	15.0%
	Wide	0.0%	2.0%	5.0%	20.4%
Empirical data		0.0%+	2.5%*		

+ $F(1,440)=.546$, $p=.460$ * $F(1,440)=11.508$, $p=.001$

Table 6. The proportion of population variability explained by all neurocomputational parameter values and the family environment quotient, for the 2x2 design. Step-wise linear regressions were carried out for the four populations, at the point in training when irregular vowel-change verbs were most closely matched to the performance of the 6-year-old children. Analyses were run separately for regular verbs and vowel-change irregular verbs.

		Environmental variation			
		Narrow		Wide	
		Regular verbs	Irregular verbs	Regular verbs	Irregular verbs
Simulation					
Genetic variation	Narrow	39.7%	60.1%	74.0%	75.1%
	Wide	43.4%	49.9%	67.8%	65.1%

Figure captions

Figure 1. Architecture of the connectionist model of English past tense acquisition, showing the internal parameters that varied in the population.

Figure 2. Parameter values (x-axis) and their predicted frequency in the population (y-axis) for the wide-genetic (black) and narrow-genetic (grey) variation conditions, for each of the 14 computational parameters. Each gene had two alleles, coded as binary values. Several genes coded for each parameter value. Sets of binary values were summed and a look-up table used to derive each parameter value. The numbers of binary alleles for each parameter were as follows. *G-Wide* = hidden units: 10; temperature: 10; noise: 8; learning rate: 12; momentum: 8; weight variance: 8; architecture: 6; learning algorithm: 4; nearest neighbour threshold: 10; pruning onset epoch: 10; pruning probability: 8; pruning threshold: 10; weight decay: 10; sparseness: 12 (total 126 bits). *G-Narrow* = hidden units: 4; temperature: 6; noise: 6; learning rate: 4; momentum: 2; weight variance: 6; architecture: 2; learning algorithm: 4; nearest neighbour threshold: 4; pruning onset epoch: 4; pruning probability: 4; pruning threshold: 6; weight decay: 4; sparseness: 4 (total 60 bits).

Figure 3. Population means and standard deviations for accuracy of past tense production (a) Empirical data for 442 6-year-old children (from Bishop, 2005); (b) Population simulations for conditions with different ranges of variation: G = genetic variation, E = environmental variation, W = wide, N = narrow.

Figure 4. Population distributions for performance on regular and irregular past tense production. Empirical data are from Bishop (2005); simulation data are from a population with wide genetic variation and narrow environmental variation.

Figure 5. Contributions to the population distribution of sub-populations, distinguished by whether a network has 3-layer architecture or a 2-layer architecture (+/ or /-) and whether the network is trained using cross entropy or not (/+ or /-). (a) and (b) show performance at a point in training when irregulars exhibited a wide range, while (c) and (d) show performance at an earlier point in training when regulars show a wide distribution. +/- afflicts both regulars and irregular, while -/+ particularly afflicts irregulars. N per sub-group: +/+ 841, +/– 57, –/+ 95, –/– 7.

Figure 6. Population distributions of performance across development for five measures of past tense performance, for the condition using wide genetic variation and narrow environmental variation.

Figure 7. Population distributions of performance across development for five measures of past tense performance, now using RMS error as the performance metric (data are the negative of the log of the error, so that larger numbers correspond to better performance).

Figure 8. Population distributions for regular verbs and irregular vowel-change verbs, at two points in development (early = 50 epochs; late = 750 epochs). Plots compare

distributions for different conditions of genetic and environmental variation. *G* = genetic variation, *E* = environmental variation, *W* = wide, *N* = narrow.

Figure 9. The predictive power of the environment in explaining variability in behaviour in past tense formation, for 6-year-old children (Bishop, 2005) and the simulation (*G*-wide-*E*-narrow condition). For the empirical data, the predictor is an SES composite measure (Petrill et al., 2004). For the simulation data, the predictor is the family quotient parameter.

Figure 10. Simulated changes in the predictive power of the environment across development for the past tense verb types, for differing ranges of genetic and environmental variation. Note the different scales.

Figure 11. Different methods to carry out simulated association analyses depending on availability of system information. Accuracy data are for irregular vowel-change verbs from the *G*-wide-*E*-narrow condition at 63 epochs (matched to the children's empirical data), and depict population performance relating to variability in the temperature parameter: (a) calibration data showing the function relating temperature and performance, with other parameters held constant; (b) population performance split by values of the parameter; (c) performance split by genotypes for the parameter, in rank order; (d) performance for each genotype, per gene influencing temperature.

Figure 12. Variance in behaviour predicted by each internal computational parameter for the *G*-wide-*E*-narrow condition. Early = 50 epochs of training; Mid = 100 epochs of training; Late = 750 epochs of training. Effect sizes indicate R^2 values for independent linear regressions.

Figure 13. Simulate gene-behaviour association analyses. Effect sizes in predicting variation in population performance from individual binary allele values (0 or 1), for (a) regular verbs and (b) irregular vowel-change verbs. Early = 50 epochs of training; Mid = 100 epochs of training; Late = 750 epochs of training. There were 126 binary alleles, split into regions coding for each computational parameter: Hidden units (HU), temperature (TMP), noise (NS), learning rate (LR), momentum (MO), weight variance (WV), architecture (ARC), learning algorithm (LA), nearest-neighbour threshold (NNT), pruning onset (PO), pruning probability (PP), pruning threshold (PT), weight decay (WD), sparseness of connectivity (SP).

Figure 14. Replicability of simulation association analyses. (a) Comparison of effect sizes for original *G*-wide-*E*-narrow population and for a population trained with the same genomes but re-sampled environmental variation; (b) comparison of computational parameter effect sizes for those populations.

Figure 15. Replicability of simulation association analyses. (a) Comparison of effect sizes for original *G*-wide-*E*-narrow population and for two populations with re-sampled

genomes (same allele frequency) and re-sampled environments; (b) comparison of computational parameter effect sizes.

Figure 16. Replicability of simulation association analyses. (a) Comparison of effect sizes for original *G*-wide-*E*-narrow population and for *highskew* and *lowskew* populations with different allele frequencies; (b) comparison of computational parameter effect sizes.

Figure 17. Gene-environment interactions. Performance on (a) regular verbs and (b) vowel-change irregular verbs at early (50 epochs) and late (750 epochs) points in training, split by whether Hidden unit levels were low (30), Medium (50), or High (100), and by whether the family quotient was poor (0.6-0.7), low-average (0.7-0.8), medium-average (0.8-0.9), or high (0.9-1).

Figure 18. Pseudo gene-gene interactions. Performance on (a) regular and (b) vowel-change irregular verbs, for early (50 epochs) and late (750 epochs) of training, split by two Hidden Unit levels (40 or 50) and by two Learning Rate levels (0.125 or 0.075).

Figure A1. Changes in the population frequency of alleles across generations with and without selection. Dark grey line = first generation. Light grey line = 25^{th} generation. The ceiling number of more efficient alleles is 126, the total number of genes on the genome.

Figures

Figure 1

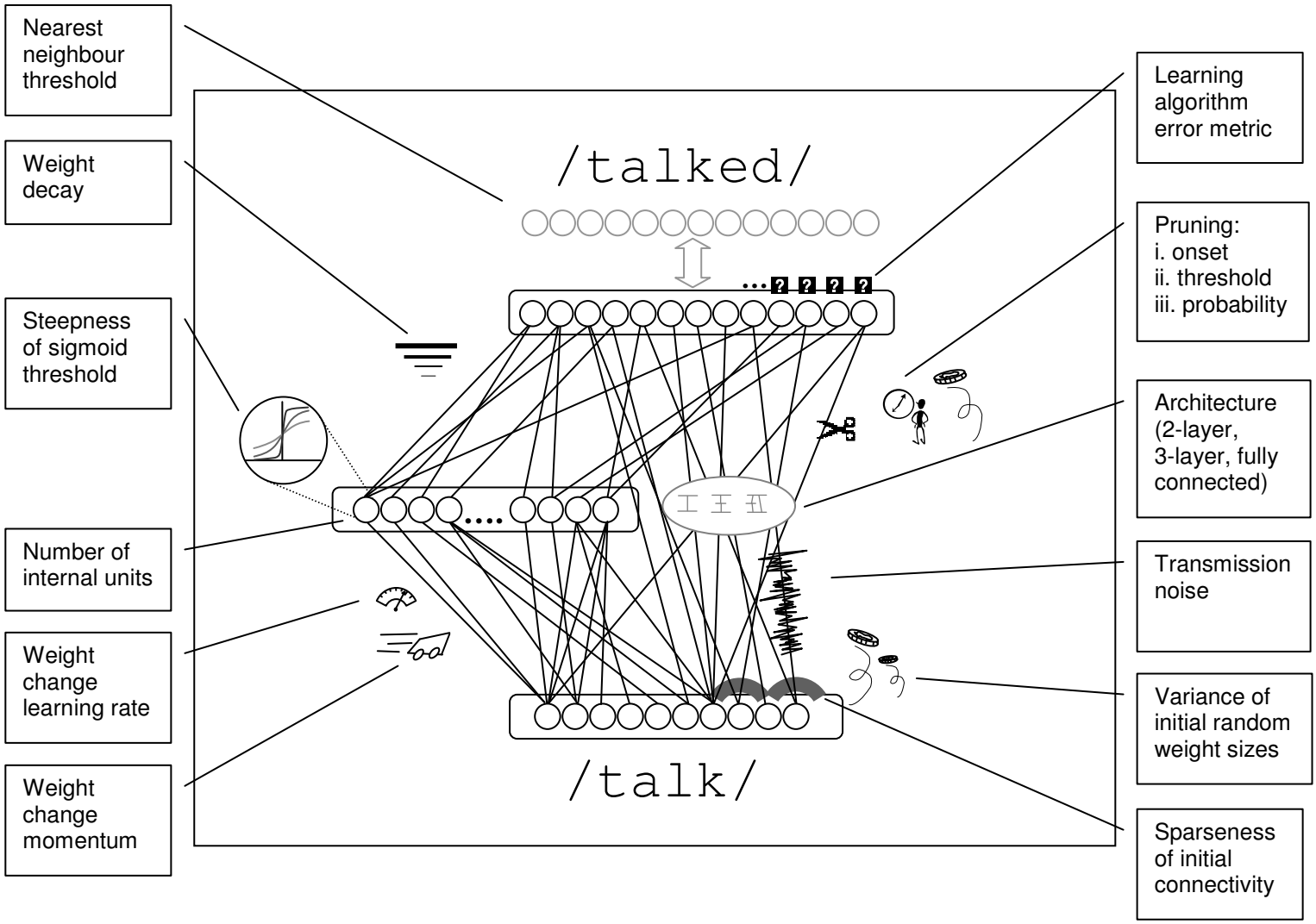

Nearest neighbour threshold
Weight decay
Steepness of sigmoid threshold
Number of internal units
Weight change learning rate
Weight change momentum
/talked/
/talk/
Learning algorithm error metric
Pruning:
i. onset
ii. threshold
iii. probability
Architecture (2-layer, 3-layer, fully connected)
Transmission noise
Variance of initial random weight sizes
Sparseness of initial connectivity

Figure 2

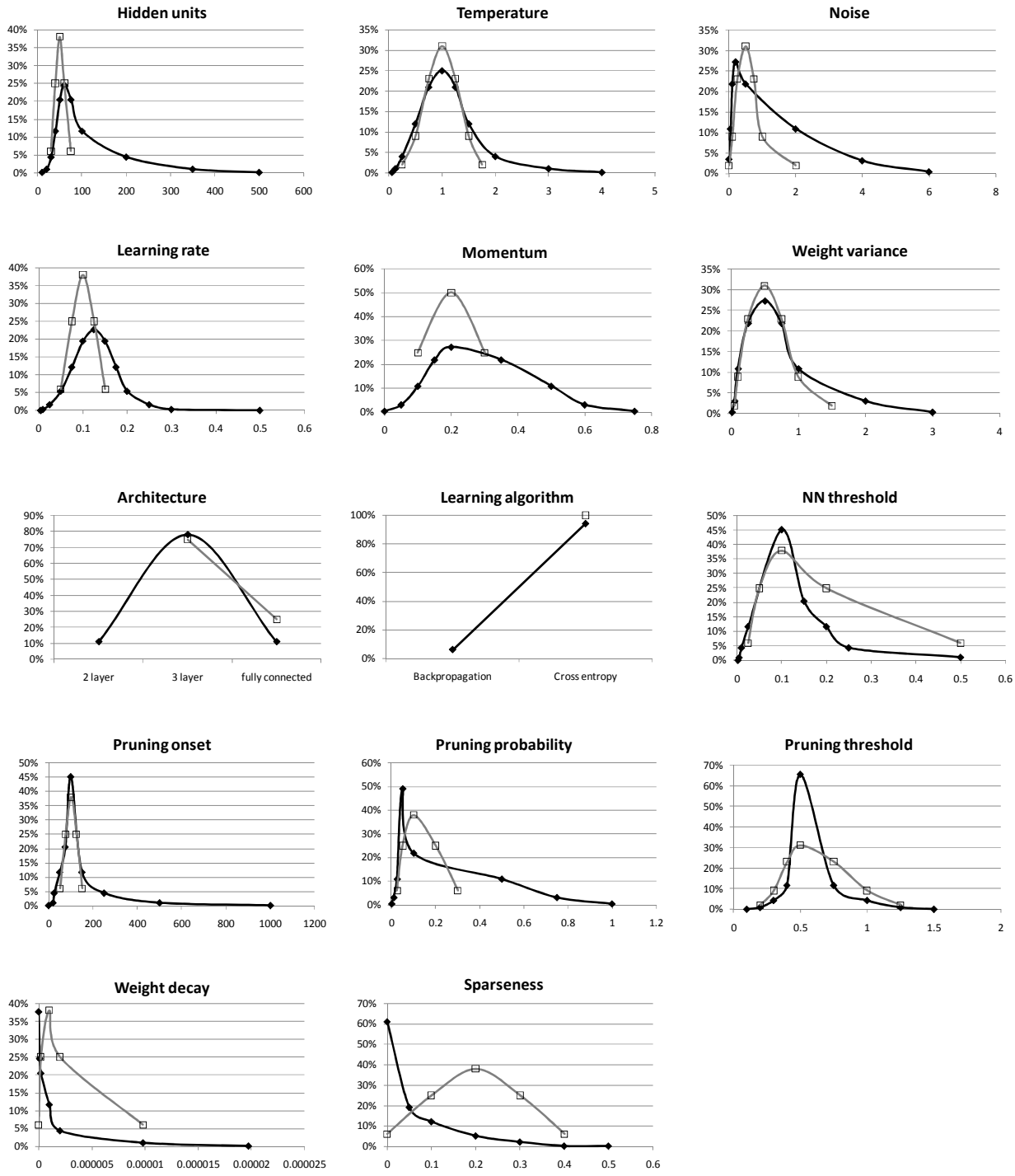
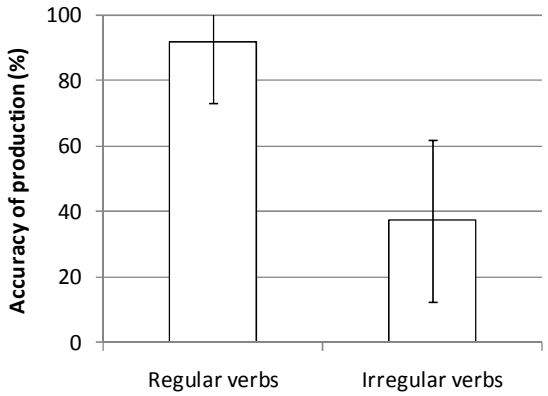

Hidden units
Temperature
Noise
Learning rate
Momentum
Weight variance
Architecture
2 layer
3 layer
fully connected
Learning algorithm
Backpropagation
Cross entropy
NN threshold
Pruning onset
Pruning probability
Pruning threshold
Weight decay
Sparseness

Figure 3

(a) Empirical data (Bishop, 2005)



(b) Simulation data

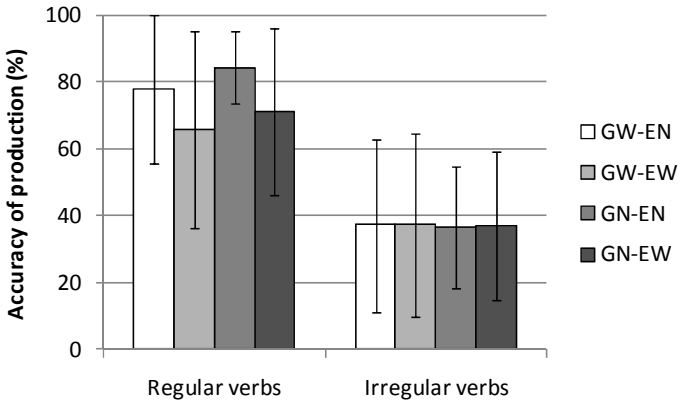
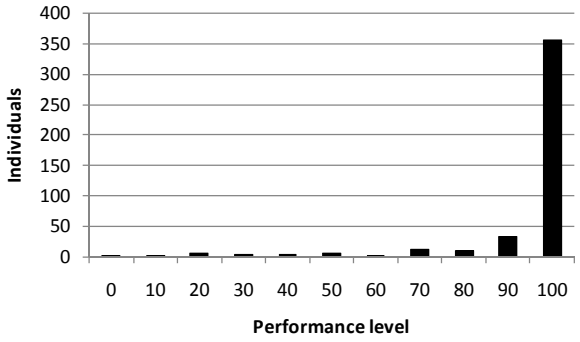
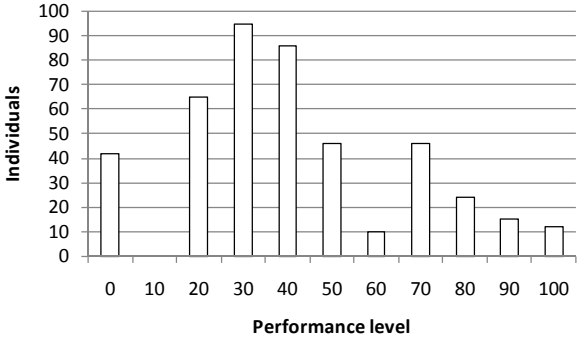


Figure 4

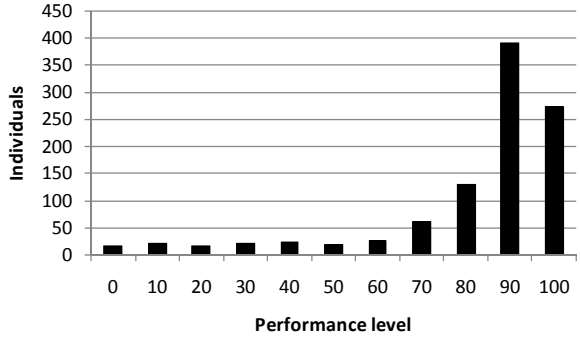
(a) Empirical data: Regular verbs



(b) Empirical data: Irregular verbs



(c) Simulation data: Regular verbs



(d) Simulation data: Irregular verbs

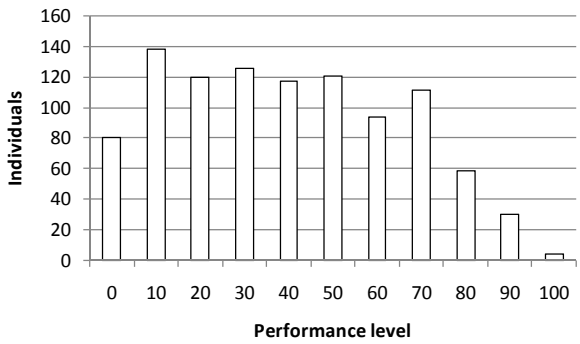
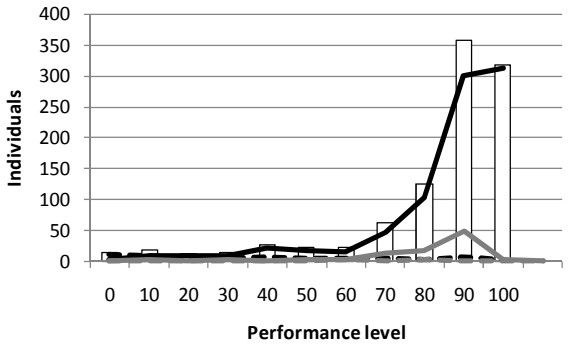
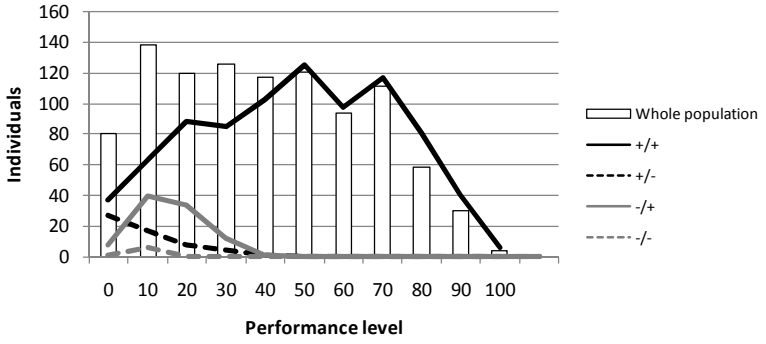


Figure 5

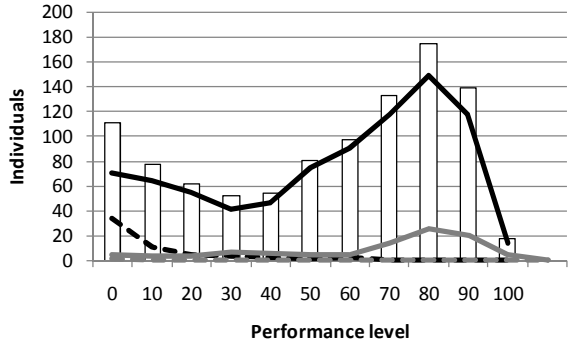
(a) Simulations: Regular verbs



(b) Simulations: Irregular verbs



(c) Simulations: Regular verbs (early)



(d) Simulations: Irregular verbs (early)

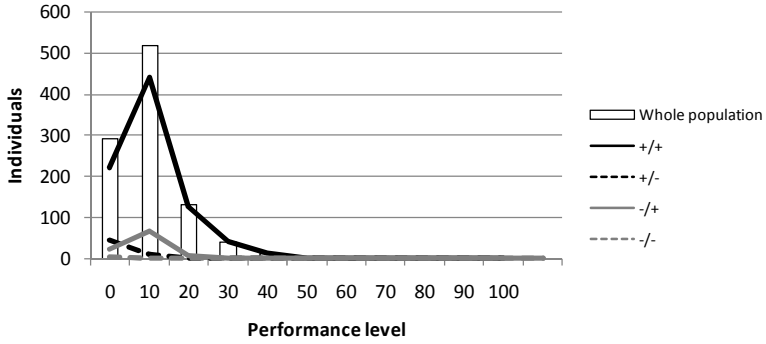
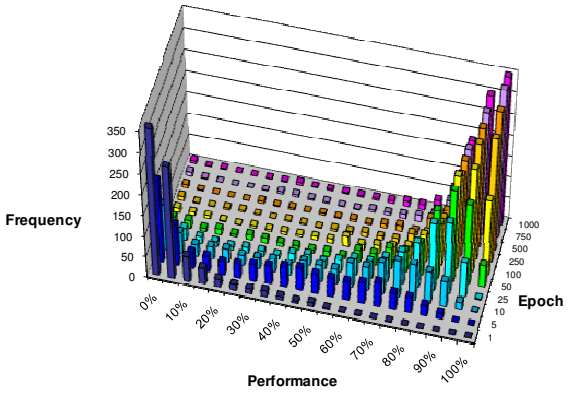
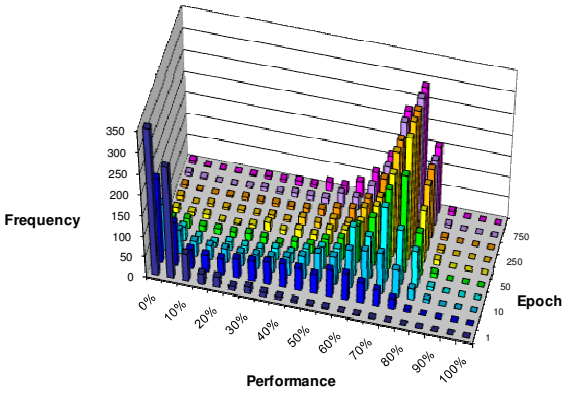


Figure 6

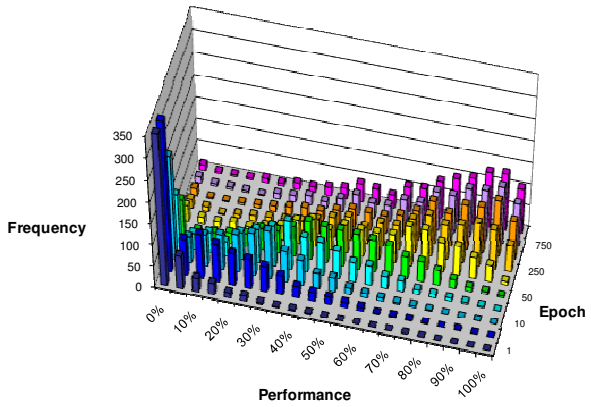
(a) Regular verbs



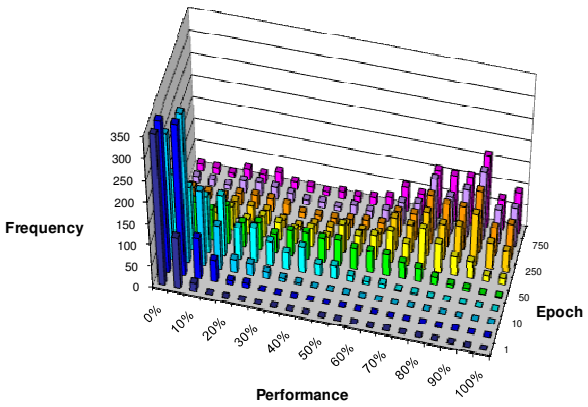
(b) Rule generalisation for novel verbs



(c) No-change irregular verbs (*hit-hit*)



(d) Vowel-change irregular verbs (*hide-hid*)



(c) Arbitrary irregular verbs (*go-went*)

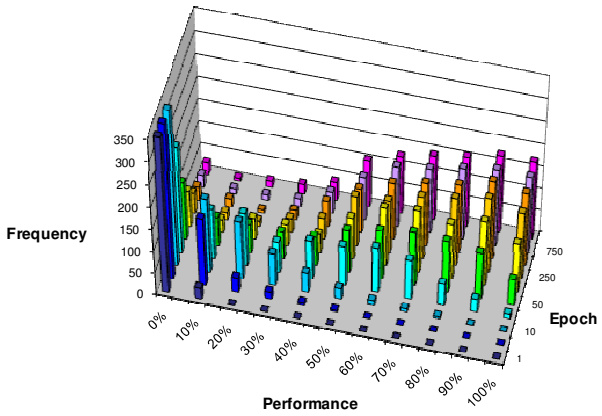
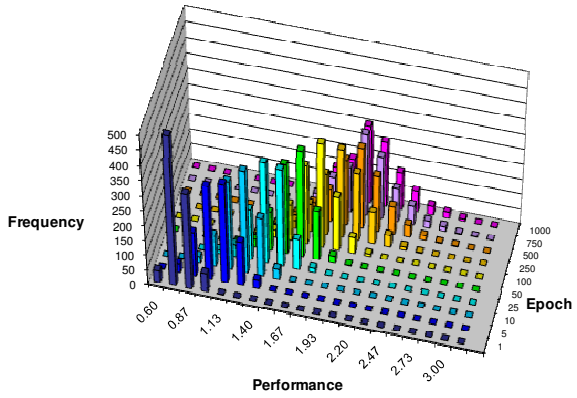
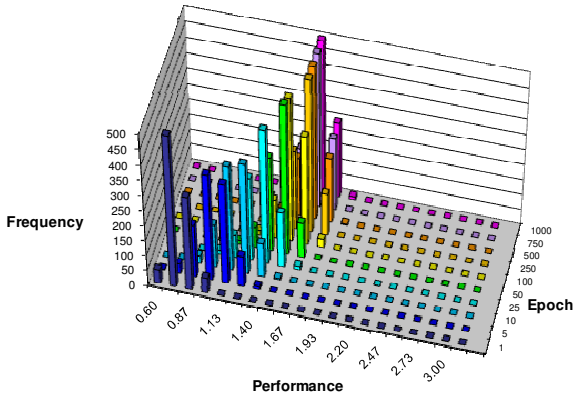


Figure 7

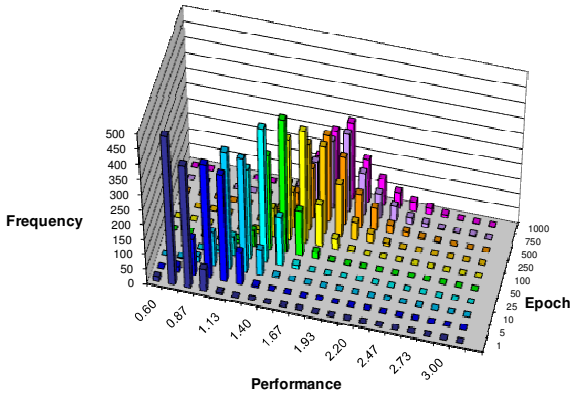
(a) Regular verbs



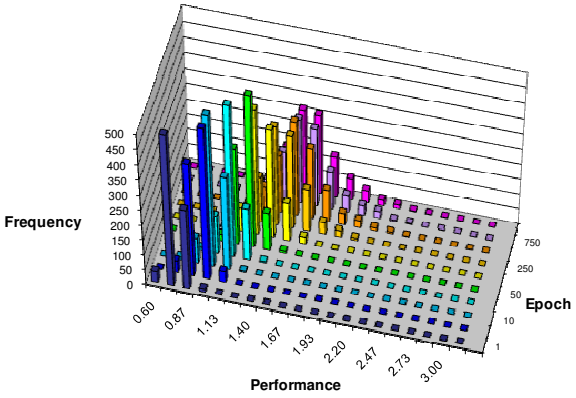
(b) Rule generalisation for novel verbs



(c) No-change irregular verbs (*hit-hit*)



(d) Vowel-change irregular verbs (*hide-hid*)



(c) Arbitrary irregular verbs (*go-went*)

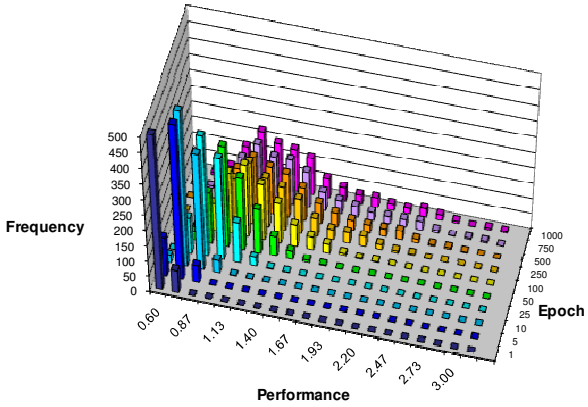
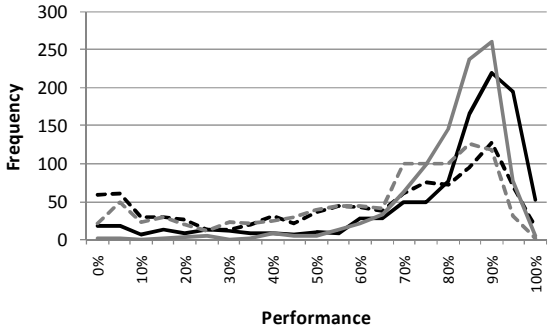
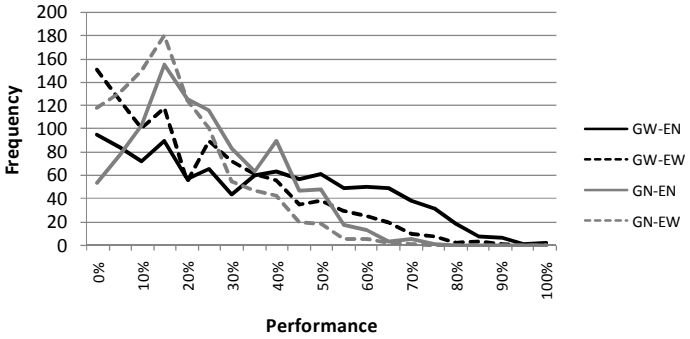


Figure 8

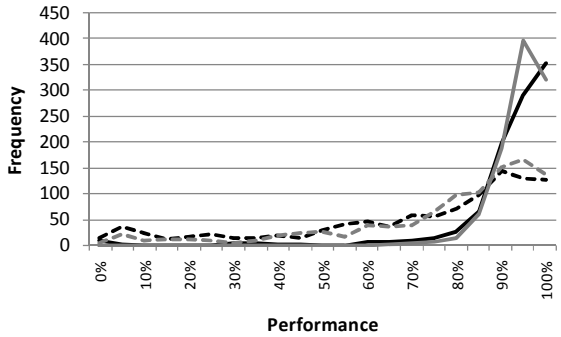
(a) Regular verbs at 50 epochs



(b) Irregular VC verbs at 50 epochs



(c) Regular verbs at 750 epochs



(b) Irregular VC verbs at 750 epochs

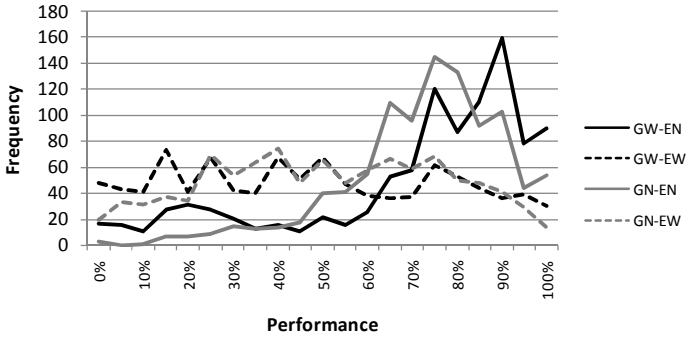
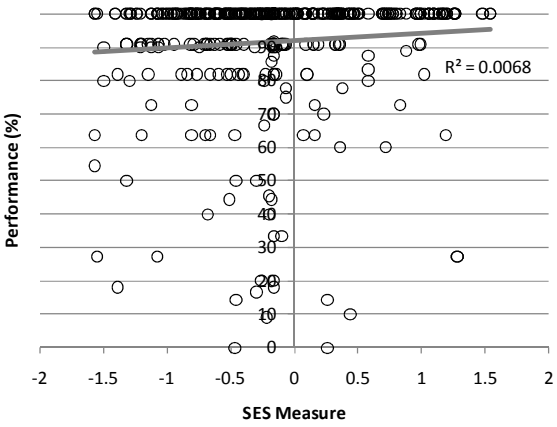
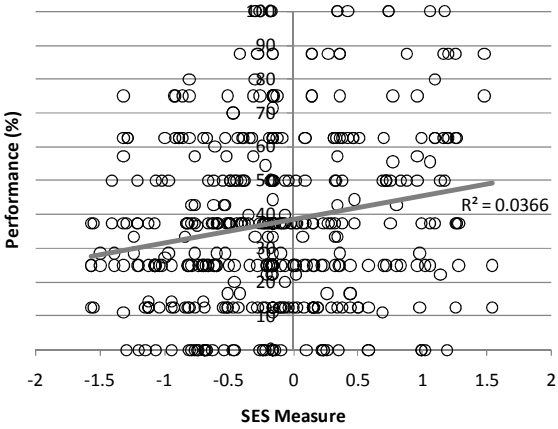


Figure 9

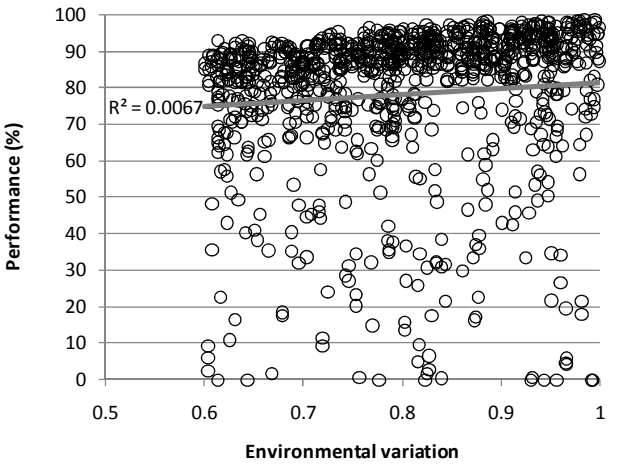
(a) Empirical data: Regular verbs



(b) Empirical data: Irregular verbs



(c) Simulation: Regular verbs



(d) Simulation: Irregular verbs

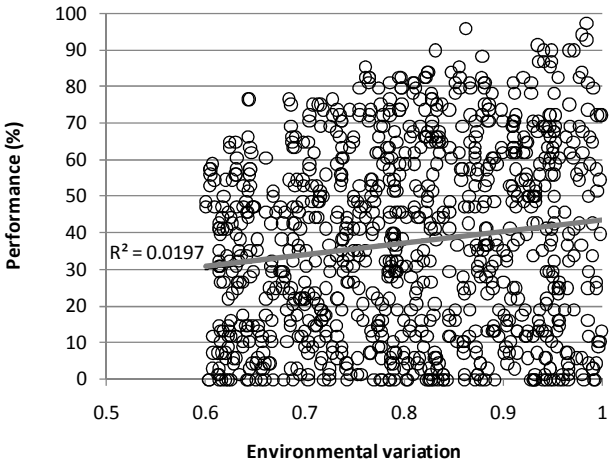
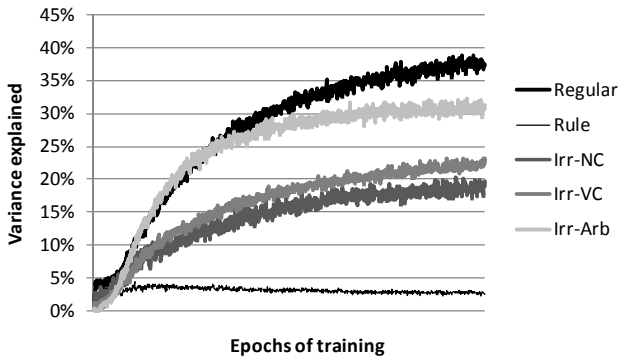
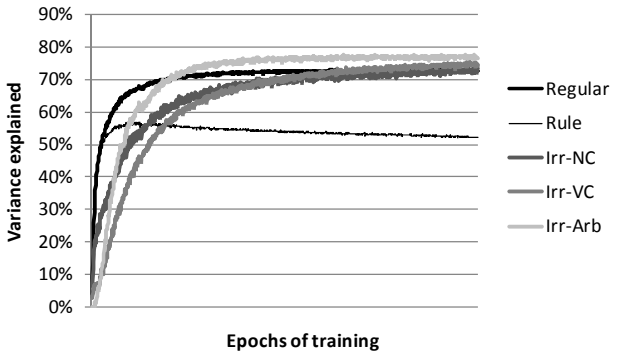


Figure 10

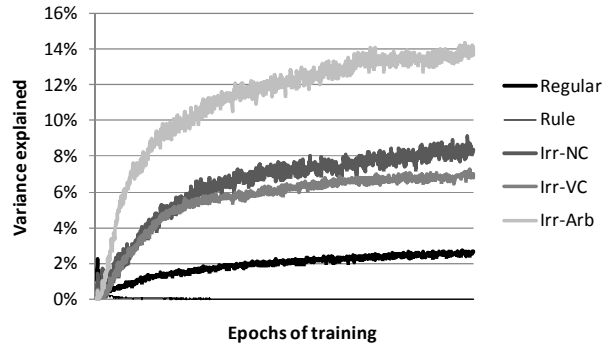
(a) Narrow genetic / Narrow environment



(b) Narrow genetic / Wide environment



(c) Wide genetic / Narrow environment



(d) Wide genetic / Wide environment

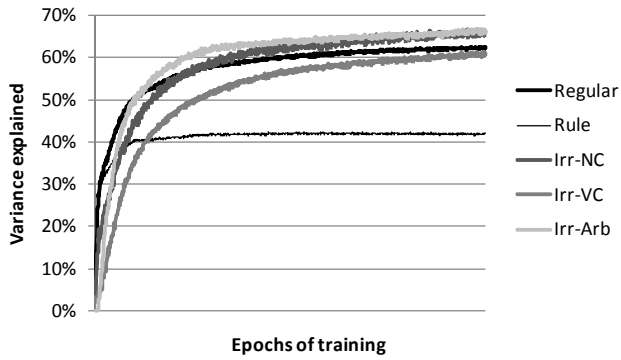
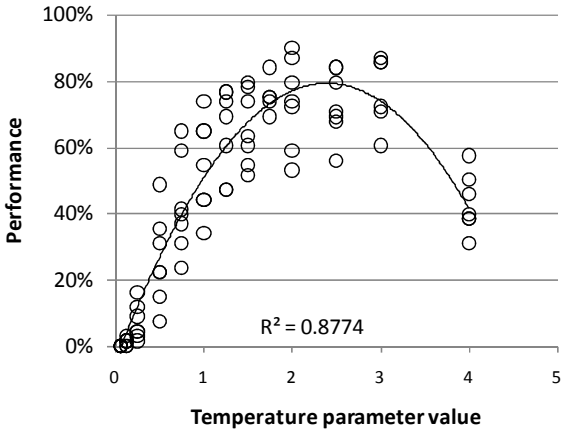
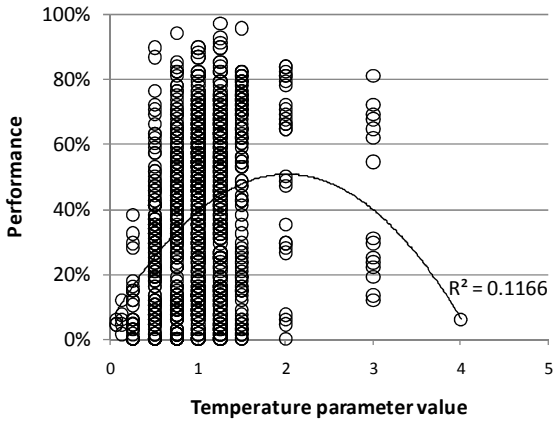


Figure 11

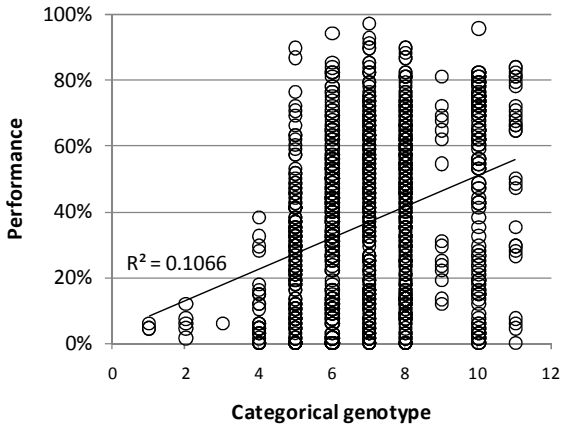
(a) Parameter-behaviour function (other parameters constant)



(b) Parameter-behaviour association



(c) Genotype-behaviour association



(d) Simulated gene associations

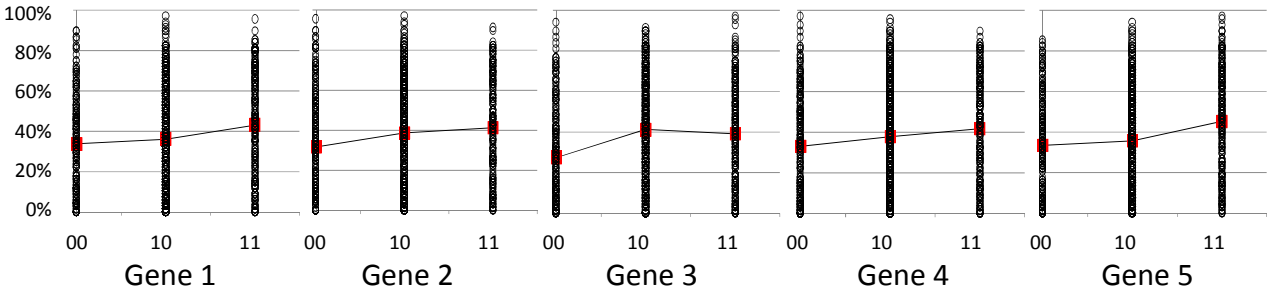
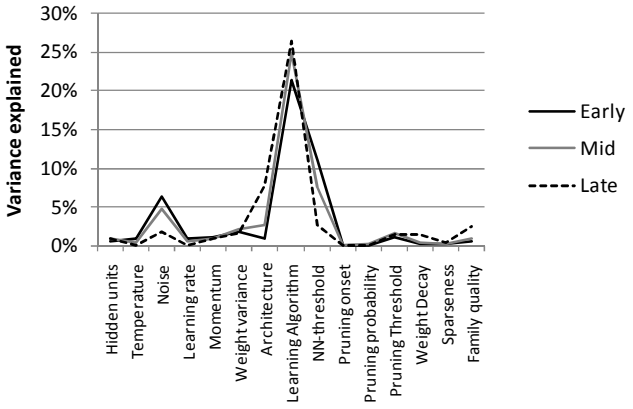
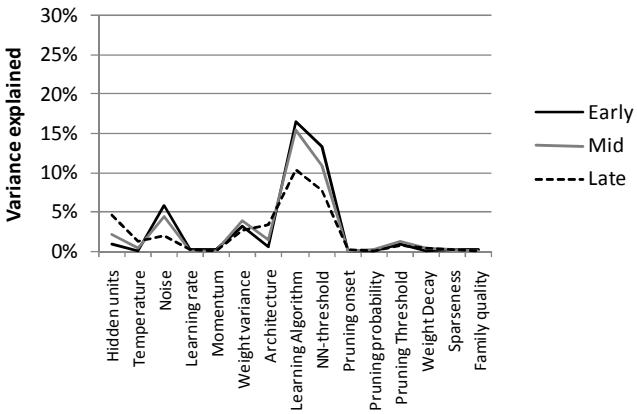


Figure 12

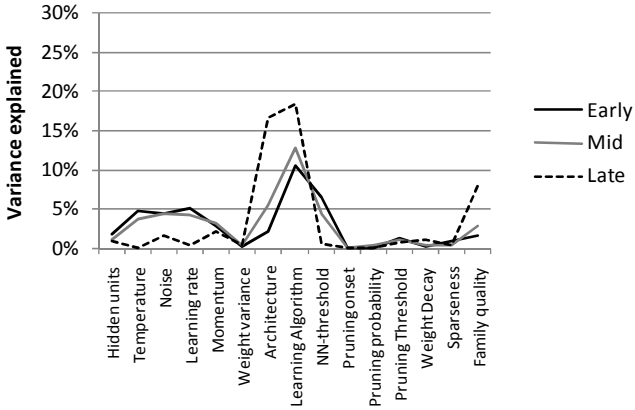
(a) Regular verbs



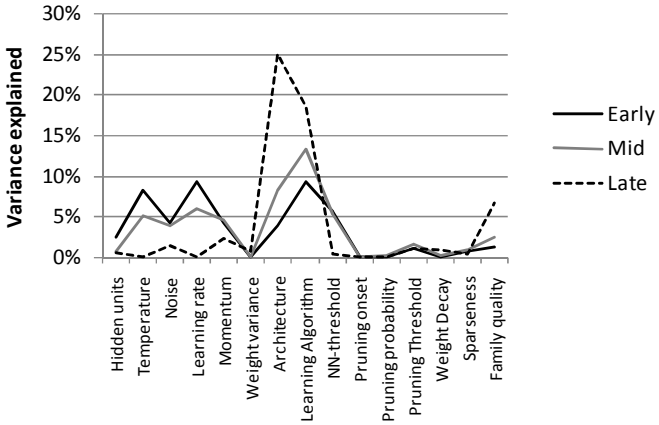
(b) Rule generalisation for novel verbs



(c) No-change irregular verbs (*hit-hit*)



(d) Vowel-change irregular verbs (*hide-hid*)



(c) Arbitrary irregular verbs (*go-went*)

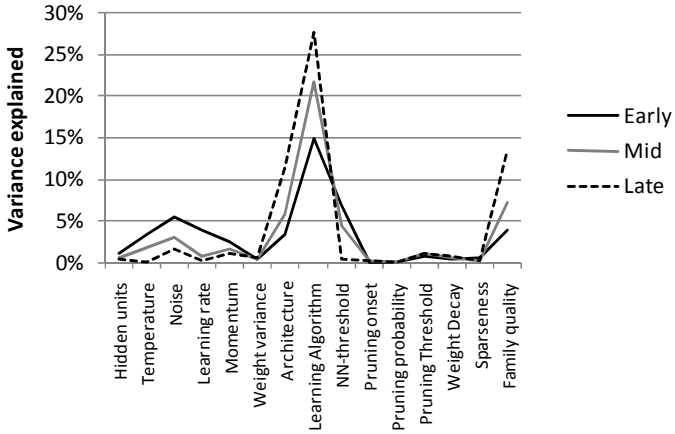
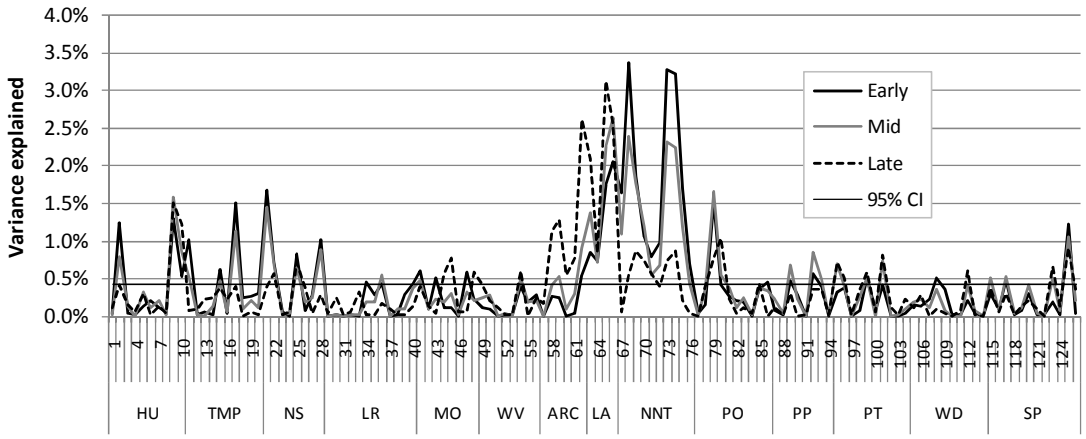


Figure 13

(a) Regular verbs



(b) Irregular vowel change verbs

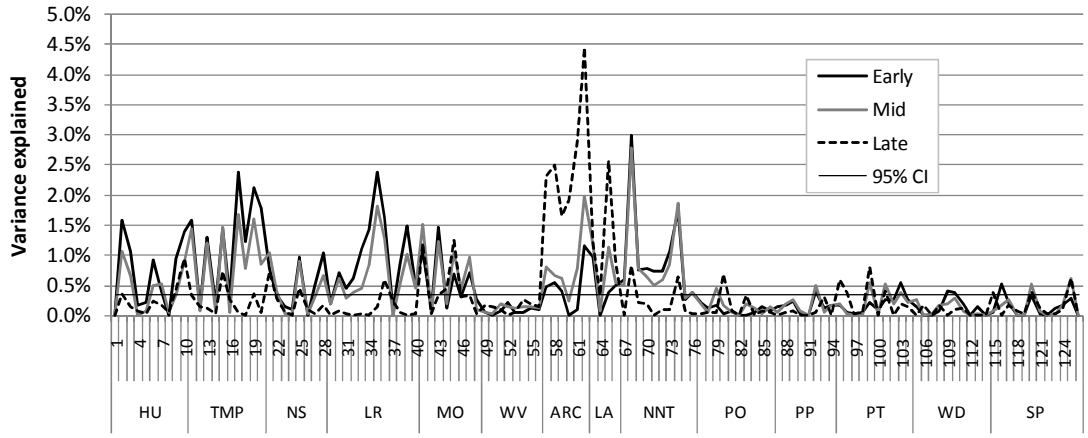
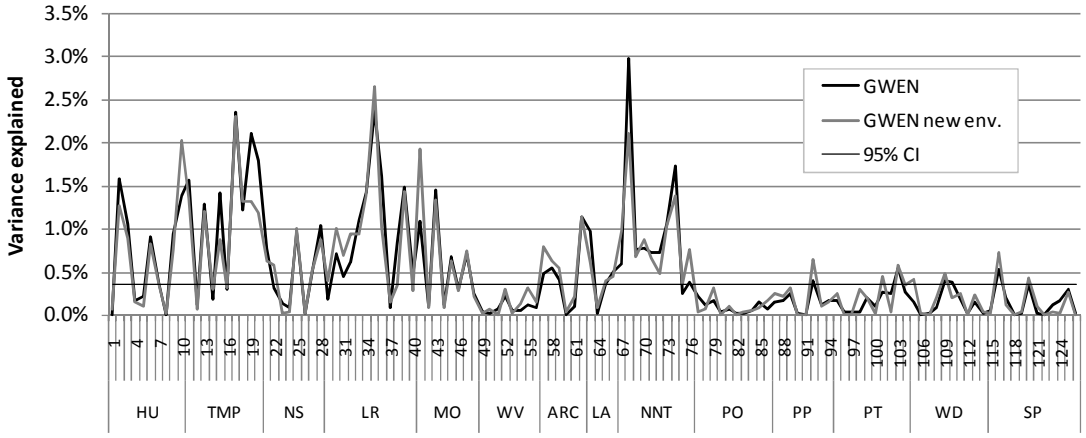


Figure 14

(a) Replication with re-sampled environment



(b) Equivalent parameter effect sizes

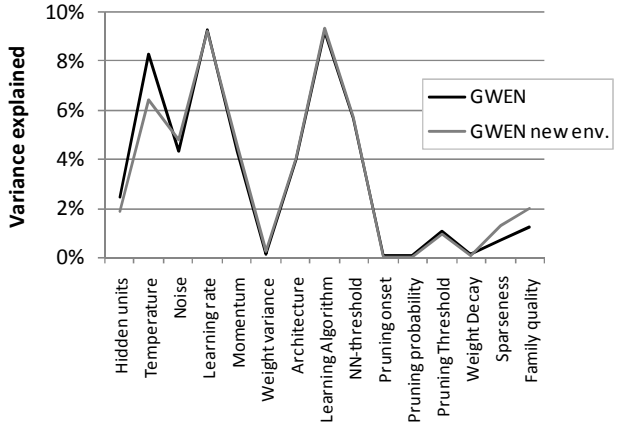
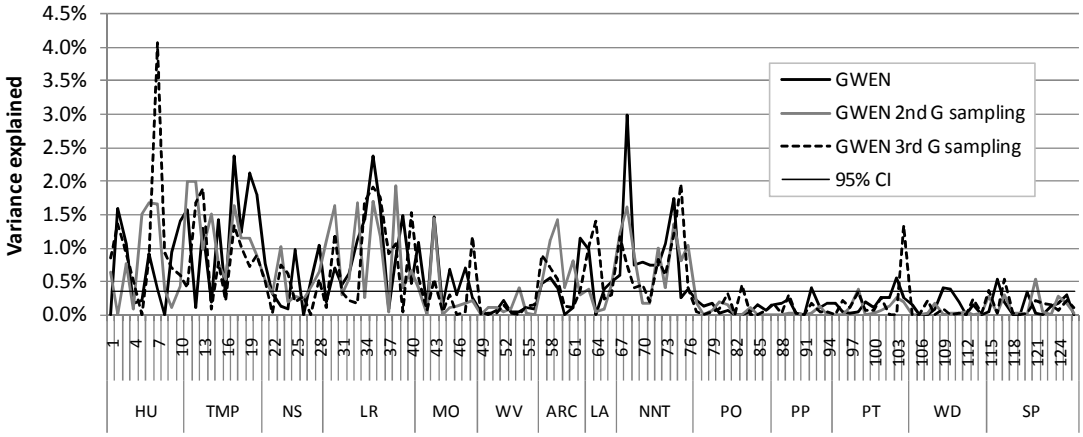


Figure 15

(a) Replication with re-sampled genomes and environment



(b) Equivalent parameter effect sizes

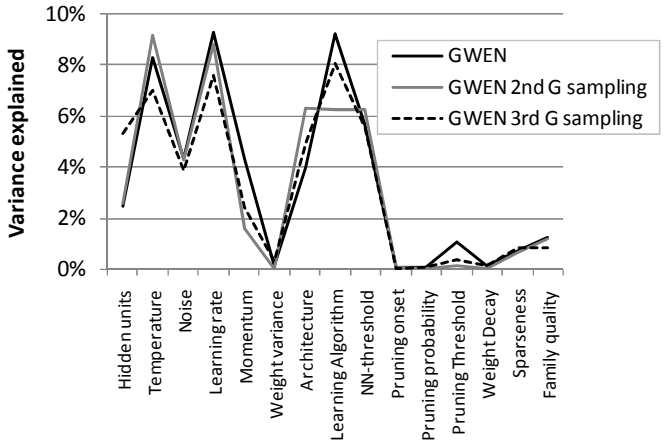
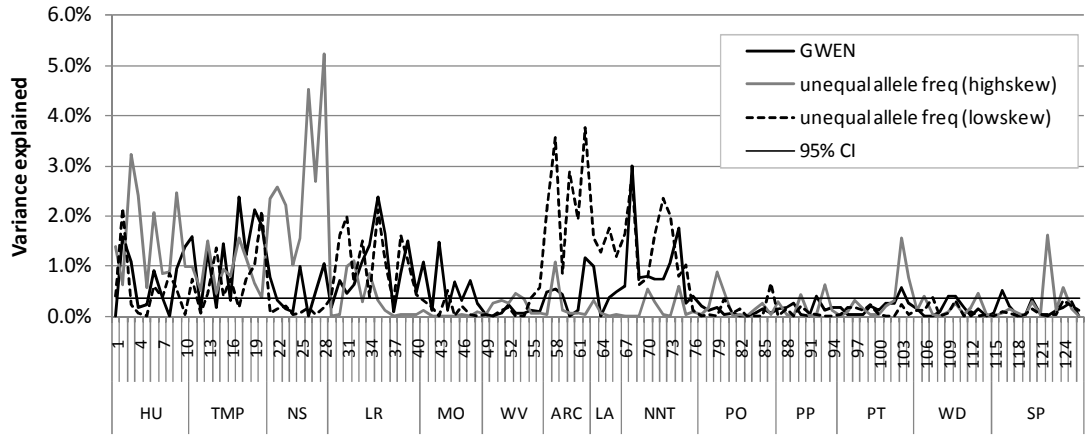


Figure 16

(a) Replication with populations with different allele frequencies



(b) Equivalent parameter effect sizes

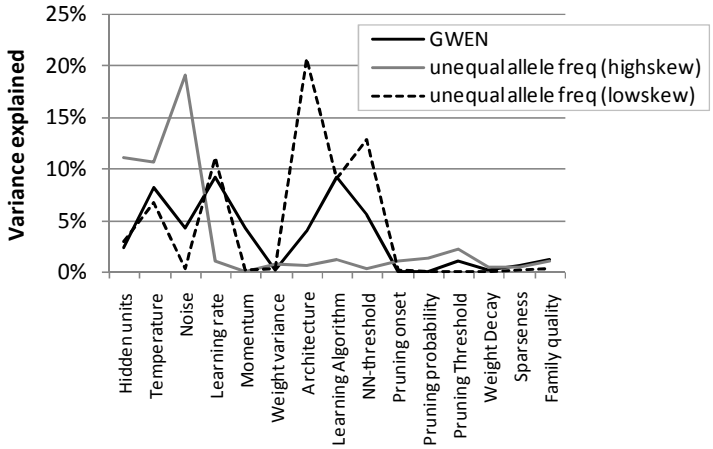
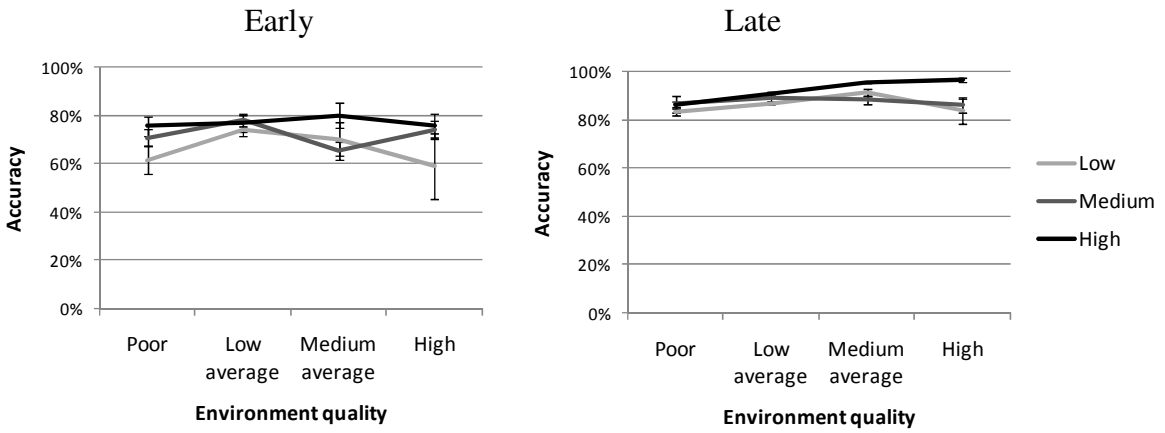


Figure 17

(a) Regular verbs



(b) Irregular vowel-change verbs

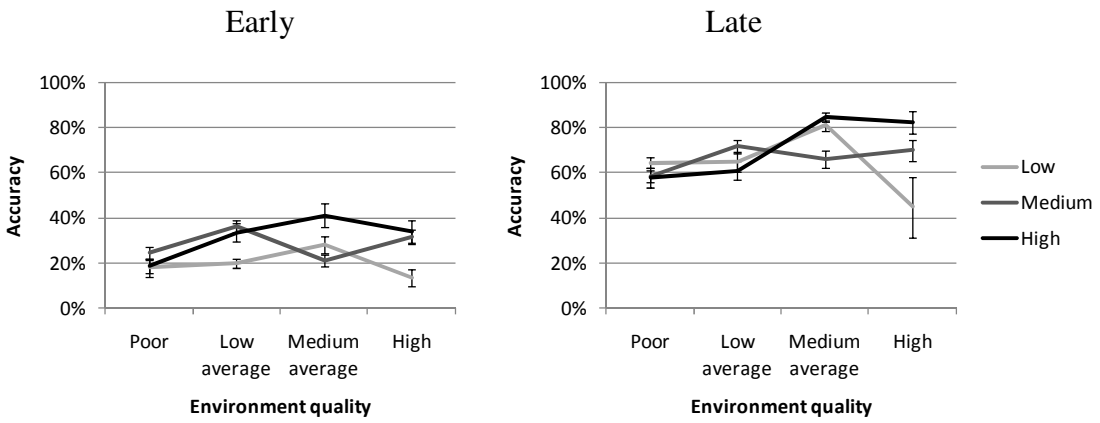
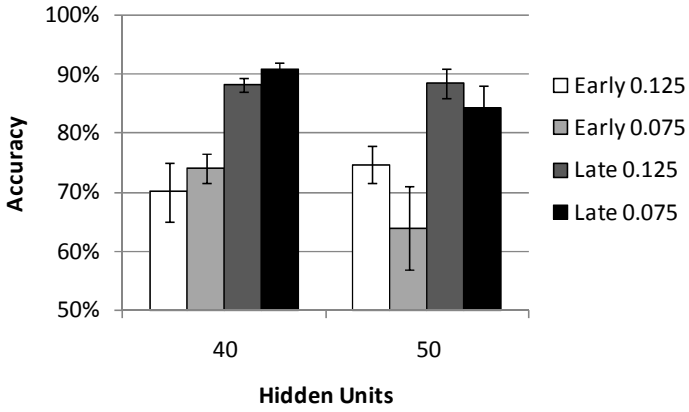


Figure 18

(a) Regular verbs



(b) Irregular vowel-change verbs

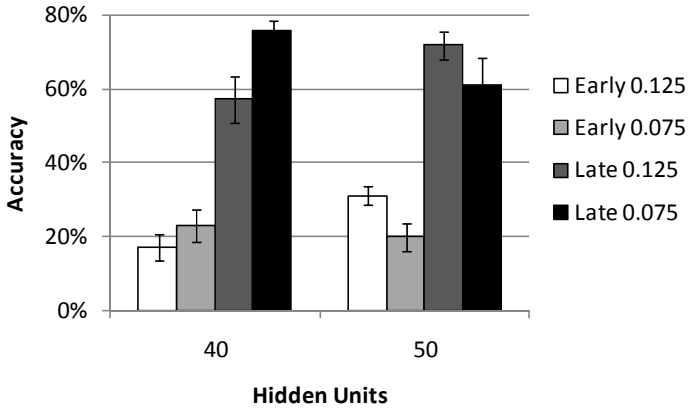
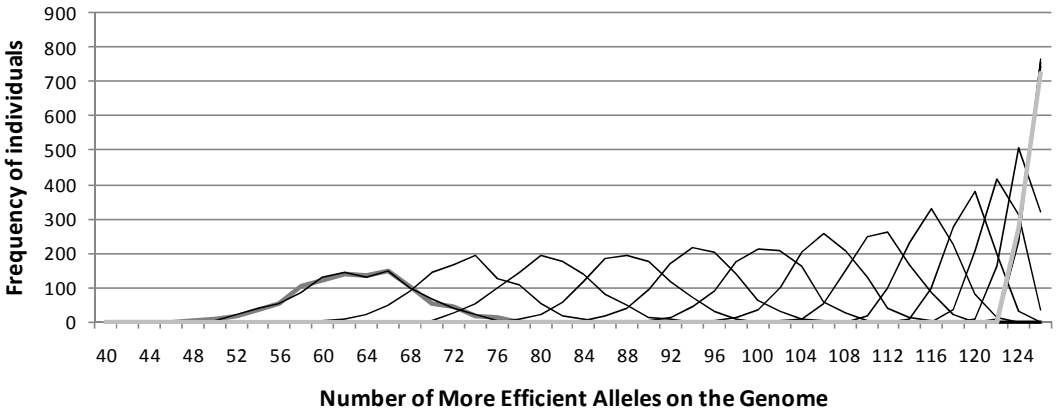
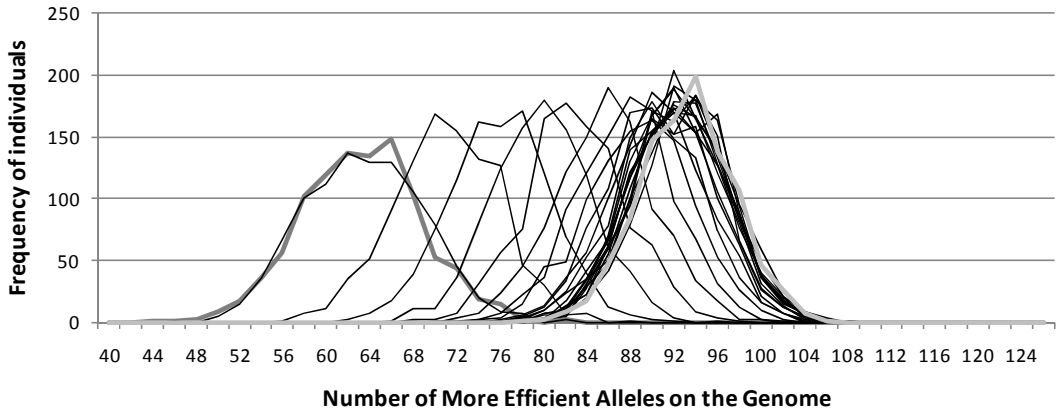


Figure A1

(a) Selection without mutation



(b) Selection with 10% mutation per generation



(c) Breeding without selection