BIROn - Birkbeck Institutional Research Online

Thomas, Michael S.C. and Forrester, N.A. and Ronald, Angelica (2016) Multi-scale modeling of gene-behavior associations in an artificial neural network model of cognitive development. Cognitive Science 40 (1), pp. 51-99. ISSN 1551-6709.

Downloaded from: https://eprints.bbk.ac.uk/id/eprint/11535/

Usage Guidelines:
Please refer to usage guidelines at https://eprints.bbk.ac.uk/policies.html or alternatively contact lib-eprints@bbk.ac.uk.

Multi-scale modeling of gene-behavior associations in an artificial neural network model of cognitive development

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

Developmental Neurocognition Lab
Birkbeck, University of London, UK

Abstract: 250 words
Main text: 13,417 words
References: 1837 words

Running head: Simulating gene-behavior associations in a multi-scale model

Abstract

In the multi-disciplinary field of developmental cognitive neuroscience, statistical associations between levels of description play an increasingly important role. One example of such associations is the observation of correlations between relatively common gene variants and individual differences in behavior. It is perhaps surprising that such associations can be detected despite the remoteness of these levels of description, and the fact that behavior is the outcome of an extended developmental process involving interaction with a variable environment. Given that they have been detected, how do such associations inform cognitive-level theories? To investigate this question, we employed a multi-scale computational model of development, using a sample domain drawn from the field of language acquisition. The model comprised an artificial neural network model of past-tense acquisition trained using the backpropagation learning algorithm, extended to incorporate population modeling and genetic algorithms. It included five levels of description, four internal: *genetic*, *network*, *neurocomputation*, *behavior*; and one external: *environment*. Since the mechanistic assumptions of the model were known and its operation was relatively transparent, we could evaluate whether cross-level associations gave an accurate picture of causal processes. We established that associations could be detected between artificial genes and behavioral variation, even under polygenic assumptions of a many-to-one relationship between genes and neurocomputational parameters, and when an experience-dependent developmental process interceded between the action of genes and the emergence of behavior. We evaluated these associations with respect to their specificity (to different behaviors, to

function versus structure), to their developmental stability, and to their replicability, as well as considering issues of missing heritability and gene-environment interactions. We argue that gene-behavior associations can inform cognitive theory with respect to *effect size*, *specificity*, and *timing*. The model demonstrates a means by which researchers can undertake modeling multi-scale modeling with respect to cognition, and develop highly specific and complex hypotheses across multiple levels of description.

Keywords: Multi-scale models, artificial neural networks, population modeling, gene-behavior associations, gene-environment interactions, missing heritability, socio-economic status, development, individual differences

1. Introduction

Developmental cognitive neuroscience is an intrinsically multi-disciplinary endeavor, where theoretical findings from multiple levels of description are integrated into an overall account of the origins of behavior. One source of empirical data that increasingly constrains theories is that of statistical associations between levels of description; for example, gene variants that correlate with individual differences in behavior, or structural and functional properties of the brain that correlate with behavior across individuals or within individuals over time. However, it is a significant challenge to construct causal accounts of development that span levels of description and thereby unify the correlations by appeal to explanatory mechanism (Johnston & Lickliter, 2009). This is particularly true for gene-behavior associations, because so many levels of description can be specified in between, and so many contributory factors interact to produce high-level behavior. Genetic effects are cellular but must be linked to behavior via neural circuits and global brain function. Moreover, the contribution of some genetic activity to individual differences in behavior occurs via an extended developmental process.

One recent response to this challenge is the use of *multi-scale computational modeling*. This approach originated in systems biology, where the availability of more powerful computers has enabled the coupling of complex models across multiple spatial and temporal scales and for multiple physical processes (Southern et al., 2008). The aim of multi-scale models is to integrate relevant information at multiple levels of organization to recreate dynamic interactions, where the complexity of the underlying interacting non-linear processes necessitates simulation via computational methods. Within biology,

Southern et al. (2008, p.67) define a multi-scale model as one 'which includes components from two or more levels of organization (multiple length scales) or if it includes some processes that occur much faster in time than others (multiple time scales)'. The levels they characterized ranged from the quantum to molecular, macro-molecular, sub-cellular, tissue, organ, organ system, organism, and environment. Southern et al. exemplified the approach via research on the dynamics of ion channels and on cardiac modeling. The work of Karr et al. (2012) represents a more recent example, where the authors constructed a multi-scale model of a whole cell, including all of its molecular components, in order to predict phenotype from genotype.

Dammann and Follett (2011) have argued that multi-scale computational models may be equally applicable to developmental cognitive neuroscience. In particular, they considered the use of computational models with respect to developmental disability. They identified *in silico* approaches as complementary to *in vivo* and *in vitro* studies in teasing apart the complicated inter-relationships between etiological exposures and pathological mechanisms on developmental outcomes. Dammann and Follett reviewed work at the systems level, where the target outcomes are located at the behavioral level, and the lower levels of description comprise phenomena such as activity-dependent plasticity and the response of neural networks to neuronal dysfunction.

In this paper, we employed multi-scale computational modeling to investigate gene-behavior associations, and in particular, the extent to which reliable associations from the low level of genes to the high level of behavior shed light on the causal processes that

take place at the intervening levels of description. Since the mechanistic assumptions of the model were known and its operation was relatively transparent, we could evaluate whether cross-level associations gave an accurate picture of causal processes. More specifically, where genes are taken to impinge on learning abilities, we could explore how the developmental process itself, involving interaction with a structured learning environment, impacted on the relationship between gene variants and eventual behavioral outcomes. As a sample domain, we used a well-known cognitive model drawn from research on language acquisition, which captured the development of past tense formation. The architecture we utilized combined artificial neural network models of development with genetic algorithms and population modeling techniques. In the following paragraphs, we characterize the way in which association analyses have been used as a source of constraining data in developmental cognitive neuroscience, before identifying the key phenomena that were the target of our multi-scale model.

1.1 Association studies in developmental cognitive neuroscience

Based on quantitative behavioral genetic methods such as twin studies, individual differences in behavior, including cognitive skills and personality dimensions, have been found to be highly heritable (Plomin et al., 2012). Frequently, between a half and three quarters of the phenotypic variability may be explained by genetic factors in the populations that have been studied. Separately, indices of brain structure have also been found to be highly heritable – though notably, these indices are not always tightly correlated with behavior. For example, in one study by Posthuma et al. (2003), the heritability of global grey matter volume was reported to be 82% and the heritability of

verbal comprehension was reported to be 84%, while the correlation between these two indices was only .06 (see also Wallace et al., 2010). Given evidence of high heritability in individual differences at brain and behavioral levels, we should in theory be able to find gene variants across individuals that predict these differences.

Two main approaches have been used to uncover gene variants associated with phenotypic variability (see Ronald, 2011, for discussion). In *candidate gene association studies*, researchers have identified variants in genes that are hypothesized to play a role in brain development and function. The genes are involved in processes such as neurotransmitter regulation, synaptic plasticity, or neural migration. Researchers have then investigated whether the variants show reliable associations with differences in high-level behavior, either in explaining normal variation or occurring more frequently in atypical populations. As examples of studies using this approach, genetic variations have been proposed to modulate attention skills via a pathway that alters the efficiency of dopamine receptors in the fronto-striatal systems delivering behavioral control (Posner, Rothbart & Sheese, 2007). Developmental language impairment and autism have both been linked to a gene variant (CNTNAP2) that alters production of a protein sitting in the membranes of neurons. The protein influences interactions between different cells during the development and wiring up of the nervous system (Vernes et al., 2008; see Peñagarikano & Geschwind, 2012). Developmental dyslexia has been linked to four gene variants (DYX1C1, KIAA0319, DCDC2 and ROBO1) associated with neuronal cell adhesion, perhaps pointing towards regional disruptions of neural migration and axonal guidance in early brain development (Galaburda et al., 2006).

On the whole, associated gene variants appear to relate to fairly 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 their effects in the brain are likely to be widespread in terms of structure and function (Kovas & Plomin, 2006; Plomin & Kovas, 2005). Where gene-behavior associations have been found, effect sizes are usually small, each explaining less than 1% of the behavioral variance. The implication is that multiple gene variants contribute jointly to variations at the level of behavior (Plomin et al., 2012). Even though effect sizes are small, they can nevertheless be observed for one behavior and not for another even in the same domain. For example, in individuals with specific language impairment, an association was observed between variants of two genes on chromosome 16 (CMIP and ATP2C2) and non-word repetition performance, but no association was observed for recalling sentences or for reading (Newbury et al., 2009). Since the contribution of individual gene variants to predicting behavior is usually so small in association analyses, even with large populations, there are many false alarms and failures to replicate across different samples in candidate gene association studies (Posthuma & de Geus, 2006).

The second main approach used to uncover gene variants associated with phenotypic variability is *genome-wide association studies* (GWAS). In GWAS, researchers seek associations with markers of genetic variation that span the whole genome. If an association is found between a particular marker and a high-level trait, researchers infer

that the location of the causal variant is close to the marker (based on the principle of *linkage disequilibrium*, whereby locations that are closer on a chromosome have a greater probability of being inherited together; see Visscher et al., 2012). A large number of markers are used, allowing some localization of causal variants on the genome, though the actual causal variants must then be identified. To date, GWAS have been more often used to study genetic variation associated with complex diseases, often conceptualized as a dichotomous outcome. Visscher et al. (2012) reported that well over 2000 locations have now been significantly and robustly associated with one or more disease traits, generating novel hypotheses about causal pathways generating disease. In most cases, multiple loci are associated with a given trait, implicating the joint contribution of multiple gene variants to variations in behavior (so called *polygenic* effects).

Visscher et al. (2012) interpreted genetic findings from the study of disease to support the *common disease-common variant hypothesis*. This hypothesis states that disease causing gene variants are common in the population, with a large number of variants each conferring a small amount of additional risk of disease. Thus a given variant may increase the odds of having a disease 1.1 to 1.5-fold (Altshuler et al., 2008). For an odds ratio of 1.1, the variant will be found in 11 individuals who have the disease for each 10 controls who do not. Gene variants also appear to be associated with more than one trait (known as *pleiotropy*) (Trzaskowski et al., 2013). However, the total phenotypic variation explained by observed associations tends not to exceed 10-20%, less than the heritability implied by twin studies. This has led to the proposal that there is 'missing heritability' (Manolio et al., 2009). New methods might reduce or eliminate the problem of missing

heritability: Yang et al. (2010) introduced the method of genome-wide complex trait analysis (GCTA). In GCTA, the genetic similarity between individuals is assessed not by family relatedness but by number of shared single nucleotide polymorphisms (SNPs; these are differences in a single 'letter' of the genetic code). This between-individual genetic similarity is then used to predict phenotypic variance. Using this approach, Benyamin et al. (2013) found that the similarity between SNPs could explain between 22 and 46% of phenotypic variation in childhood intelligence in three large cohorts totaling 18,000 individuals aged between 6 and 18 (see also Plomin et al., 2013). Despite this encouraging result, when it comes to cognitive and behavioral phenotypes rather than complex diseases, GWAS have generally struggled to find significant associations with markers of genetic variation, possibly suggesting a greater problem with missing heritability for these phenotypes than medical disease (Ronald, 2011). Rietveld et al. (2013) recently used a GWAS to identify SNPs predicting variation in educational achievement in a large sample of 120,000 individuals. Together, the identified markers of genetic variation predicted around 2% of variation in educational achievement, compared to around 10% in a similar study of height (Speliotes et al., 2010). This led the authors to propose that the genetic architecture of complex behavioral traits may be more diffuse than that of complex physical traits.

GWAS are not ideal for detecting the contribution of rare variants to disease, since by definition these will have low frequency in the population, thereby compromising the statistical power to detect associations. There is increasing evidence that rare copy number variations (CNVs) and de novo mutations may also play a role in producing

phenotypic variation. For example, the contribution of rare CNVs and de novo mutations has been identified in cases of autism (e.g., Levy et al., 2011) and schizophrenia (e.g., Kirov et al., 2012; The International Schizophrenia Consortium, 2008).

1.2 The puzzle of gene-behavior associations

From one perspective, it is surprising that it is possible to detect *any* associations between individual gene variants and high-level behavior.[1] This is for two reasons: the *remoteness* of these levels of description, and the fact that behavior is the outcome of an extended *developmental process* involving interaction with a variable environment. We expand on each of these points in turn.

With respect to *remoteness*, the genetic level of description here pertains to variation between individuals in the DNA code which codes for the production of proteins in cells, while behavior pertains to the whole organism as a single system embedded in a physical and social context. The heritability of individual differences in behavior tells us that there are genetic effects, but unpacking the causal pathways through which they operate on behavior is a daunting prospect. Genetic effects on cognition must, presumably, operate via their effect on neurocomputation and/or network topology. However, two examples suffice to illustrate the complexity of the problem at hand.

[1] Under the hypothesis that common variants contribute to normal variability. It is less surprising where a rare mutation causes a (large) pathological effect on the organism.

First, a gene codes for a protein; Plomin et al. (2008) pointed out that each synapse is affected by more than a thousand protein components. Understanding the factors that cause variations in the efficiency of the synapse is still a long way from understanding even a functional neural circuit, let alone brain networks generating behavior. There must be many points of convergence of genetic variation as one ascends levels of description. Moreover, recent research has pointed towards the complexity of the process by which genes contribute to cellular function, identifying their role as part of a dynamical system that includes multiple points of regulation of gene expression, such as modification of messenger RNA, DNA methylation and histone modification (Charney, 2012).

Second, Sapolsky (2005) outlined the multiplicity of low-level variations that one might conservatively expect to contribute to the functioning of neural circuits: at the level of individual neurons, 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 glutamate receptor types, the level of calcium ion release, and the level of

phosphorylation of the receptors. It is possible that a range of gene variants contribute to each of these neural parameters. It does not follow that all these variations would necessarily be meaningful, and development must in some sense be robust to variations in such low-level properties to be successful. Nevertheless, finding significant associations between *individual* gene variants and high-level behavior through this conflagration of causal processes is both impressive and somewhat unexpected; and perhaps even more so, given that genotyping data and behavioral data are both likely to contain measurement error.

With respect to *development*, cognitive abilities are the outcome of an extended and dynamic developmental process involving interaction with the physical and social environment, an environment that the individuals themselves play a role in specifying (Flynn et al., 2013). The environment also varies, contributing to individual differences in behavior. The nature of the developmental process itself is considered to be an important component of the explanation of cognitive variability (Karmiloff-Smith, 1998). This is illustrated by the fact that relationships between genotypes and phenotypes are not stable across development, even for neurogenetic developmental disorders. For example, Paterson et al. (1999) found that the relative pattern of cognitive strengths and weaknesses in Down syndrome and Williams syndrome altered between infancy and adulthood; that is, the effects of the respective genetic mutations depended on what stage of development the phenotype was measured. Association studies only give an askew picture of the developmental process because they rely on differences between individuals of similar ages or at similar developmental stages. Development can be

studied with association studies by examining whether the associations between gene variants and individual differences in behavior are stable across development, or whether associations reduce or increase (Ronald, 2011). Changes in gene expression are expected since they are a key component of development. However, the actual relationship between individual differences and development *as mechanistic processes* (Bechtel, 2011) has yet to be determined, and quite diverse hypotheses are still in play. For example, within the study of cognition, there are competing theoretical proposals that range from the idea that individual differences and development represent variations along orthogonal mechanistic dimensions, to the idea that they are variations over the same dimensions (see, Thomas & Karmiloff-Smith, 2003a, for discussion). For example, under one hypothetical scenario (borrowing proposals from the psychology literature), it might turn out that individual differences are generated by differences in *inhibitory control*, while development corresponds to changes in *processing capacity*; here the dimensions would be orthogonal. Under an alternative hypothetical scenario, both individual differences and development might represent variations in *processing speed*; here there would be a single common dimension. Now, if the dimensions are orthogonal, then the study of individual differences will tell us little about the developmental process; but if they are common, the study of individual differences will provide a direct window onto the developmental process.

From a computational modeling perspective, development and individual differences have rarely been considered within the same framework (see Garlick, 2002, for an exception), so these issues are not typically addressed. Developmental computational

models that specify mechanisms of experience-dependent learning usually attempt to capture the development of the 'average child', while models of individual differences usually focus on the intrinsic and extrinsic factors contributing to the variation at a single age, excluding the developmental origins of behavior. There is a pressing need to begin to consider development and individual differences within a common computational framework.

1.3 Using multi-scale models to understand the implications of associations between levels of description

In principle, multi-scale modeling can complement genetic association analyses by demonstrating how, in a system where multiple levels of description are implemented, associations from low to high levels of description reflect the causal mechanisms best characterized as operating at the intermediate levels. In practice, the contribution of a given multi-scale model depends on the constraints it embodies at different levels, the interfaces it specifies between levels, and the set of simplifying assumptions.

The notion of 'level' here is somewhat tricky, because it combines several distinctions. These include intra-personal versus extra-personal (e.g., brain processes versus the environment); levels of a mechanism that characterize the combination of smaller components into larger components; and levels of analysis in describing a phenomenon (e.g., one might describe a real neural network as performing a computational function) (see, e.g., Bechtel, 2008; Bechtel & Mundale, 1999; Craver, 2007; Eliasmith, 2002, 2013; Marr & Poggio, 1976; Potochnik & McGill, 2012). Our modeling framework indexes

each of these ideas, but our main theoretical reference point is the causal modeling approach proposed by Morton (2004) to understand the causes of developmental disorders. In this approach, the individual is distinguished from the environment; within the individual, Morton then distinguishes biological, cognitive, and behavioral levels. In our multi-scale model, the biological level is represented by a genetic level, the cognitive level is represented by neurocomputation, and the behavioral level is represented by the output of the model (see Figure 1).

To construct the current multi-scale model, we began by taking advantage of the fact that artificial neural networks have been used as models of cognitive development (see, e.g., Elman et al., 1996; Mareschal & Thomas, 2007). Behavioral change is captured as the outcome of an experience-dependent developmental process taking place in a structured learning environment. These models therefore allow us to separately characterize *behavior* and the structure of the learning *environment*. Artificial neural network models are based on abstractions of neurocomputation, and include parameters that are analogous to neurocomputational properties. Moreover, the networks encode knowledge by changing their structure, in terms of their connectivity. We can therefore discern the intra-personal properties of *neurocomputation* and *network structure*. Lastly, using methods from genetic algorithms within machine learning, the parameters of the artificial neural networks can be encoded in an artificial genome. Variations in the genome specify variations in network parameters, which then influence learning ability. We therefore posit a lowest level of *artificial genome*. The artificial genome is part of the biological level whereby many smaller components produced the operation of the larger component

that is the neural network. This sets the stage to investigate associations that span levels of description.

In order for us to simulate association studies, two further steps were necessary. First, such studies take place at a population level. Therefore we needed to simulate a population of artificial neural networks undergoing development (see Thomas, Baughman, et al., 2012; Thomas, Knowland & Karmiloff-Smith, 2011). Second, association studies rely on variability. We created both genetic and environmental sources of variation to produce variability in acquired behavior. These methods ensured that we could consider association analyses within a developmental framework: the associations between individual differences in the artificial genome and individual differences in behavior could be assessed at any point in development, whilst simultaneously capturing the developmental origins of behavior via an experience-dependent process. This was the principal innovation of our model.

The aim of our multi-scale model was to investigate associations between levels of description, such as genes to behavior, genes to network structure, and neurocomputational parameters to behavior. In particular, because the mechanistic assumptions of the model were known and its operation was relatively transparent, the model could inform the extent to which gene-behavior associations gave an accurate picture of neurocomputational causal processes operating at the intermediate level. For example, if we know that variation in two artificial genes contributes independent

influences on the operation of two neurocomputational parameters, do we observe additive statistical effects of these genes in their associations to behavior?

Given the assumptions of the model, the simulations addressed the following specific questions: (1) Can statistically significant associations be observed between artificial gene variants and individual differences in behavior, given many-to-one gene-to-neurocomputational parameter mappings and an intervening experience-dependent developmental process? (2) Do such associations show specificity to different behaviors generated by the system or are they general? (3) What is the stability of the associations over developmental time – are associations modulated by the developmental process? (4) Do associations replicate across populations? (5) Are associations observed from artificial genome to network structure and activation levels, and if so, are these the same as the associations observed from artificial genes to network output (behavior)? (6) Are associations modulated by the quality of the environment, producing gene x environment interactions? (7) Can interactions between genes be observed in the way that they influence behavior? (8) When all sources of variability are known, is all the population variance explained or is some 'missing'? We then discuss whether observed cross-level statistical associations accurately reflected the causal operation of the model.

2. Method

The model we utilized to simulate gene-behavior associations was taken from the domain of language development, and has been successfully used to simulate socio-economic status effects on language development (Thomas, Forrester & Ronald, 2013), as well as

sub-types of language delay (Thomas & Knowland, 2014). The model was addressed to the domain of English past tense formation. Here, we employed the model in an illustrative setting. The model is intended only as an example of a developmental system applied to the problem of extracting the latent structure of a cognitive domain through exposure to a variable training environment. Past tense has been used similarly to study phenomena such as critical periods in development (Marchman, 1993) and developmental regression in autism (Thomas, Knowland & Karmiloff-Smith, 2011). The English past tense provides a useful sample domain because it is *quasi-regular*. It is characterized by a majority of past tenses that follow a productive rule (add 'ed' to the verb stem) but a minority of exceptions to this rule, forming their past tenses in a variety of ways. Performance on regular verbs and irregular verbs form two different behaviors that the system must acquire. A range of empirical research indicates that children's and adults' performance on regular and irregular verbs differs in its characteristics, sufficiently so that some have argued that different processing mechanisms are needed to acquire the verb types (e.g., Pinker, 1994). The two types of behavior allow us to test the *specificity* of associations between artificial genes and behavior.

In the following sections, we first outline the base model. We then consider the implementation of constraints at each level: Environment, Behavior, Network structure and activation, Neurocomputation, and Artificial Genome. Finally, we outline the simulation design. Further implementation details can be found in Supplementary Materials.

2.1 Base model

A 3-layer, backpropagation network was used to learn to map between a phonological representation of verb stems and their past tense forms. The architecture is shown in Figure 1.

The results we report come from the simulation of 6000 artificial neural networks. Some simplifications of network scale were employed for computational tractability. First, an artificial language was used rather than a corpus of real English verbs, per the work of Plunkett and Marchman (1991, 1993). The training set comprised an artificial language constructed to reflect many of the important structural features of English past-tense formation. Artificial verbs were monosyllabic and encoded used articulatory feature-based codes drawn from English phonology. Second, the model employed a simplified architecture in restricting mappings to be between phonological codes. More recent, larger scale models have included additional information in the input, such as lexical semantic information (e.g., Joanisse & Seidenberg, 1999; Woollams et al., 2009), and acquire multiple inflectional paradigms rather than just the past tense of verbs (e.g., Karaminis & Thomas, 2010). These simplifications are not relevant given the abstract aims of the model.

<Insert Figure 1 about here>

The training set was the "phone" vocabulary from Plunkett and Marchman's past tense model (1991, p. 70). There were 508 monosyllabic verbs, constructed using consonant-vowel templates and the phoneme set of English. Phonemes were represented over 19

binary articulatory features (Thomas & Karmiloff-Smith, 2003b), a distributed encoding based on standard linguistic categorizations (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 57 input units and 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 (examples from English: *tame-tamed*, *wrap-wrapped*, *chat-chatted*); (2) irregular verbs whose past-tense form was identical to the verb stem (e.g., *hit-hit*); (3) irregular verbs that formed their past tenses by changing an internal vowel (e.g., *write-wrote*; (4) irregular verbs whose past-tense form bore no relation to its verb stem (e.g., *go-went*). There were 410 regular verbs, and 20, 68, and 10, respectively, of each irregular verb type. A generalization set was also created with 410 novel verbs, each of which rhymed (shared two phonemes) with an existing regular verb. Generalization was assessed by the accuracy of outputting the regularized past tense form. Networks learned by repeated presentations of the training set, with verbs presented in random order, and operation of a gradient-descent supervised learning algorithm (backpropagation). One presentation of the training set is referred to as an 'epoch'. All networks were trained for 1000 epochs.

2.2 Environment

The environmental level was defined as an extrapersonal influence on development. Each network simulated a child raised in a given family, and families were assumed to vary in the richness of the language used. The language input was assumed to vary to some

extent according to socio-economic status (SES) (Hart & Risley, 1995). A training set was created for the past-tense information available in each family environment. SES was implemented through generating a *family quotient* for each simulated child. The family quotient was a proportion between 0 and 100%. This value was used as a probability to determine whether each verb in the full training set would be included in the family's vocabulary. The family training set was then fixed throughout development. Performance was always assessed against the full training set (analogous to a standardized test of past-tense formation applied to all children). The family quotient manipulation corresponded to a reduction in type frequency for both regular and irregular verbs. Based on the findings of Thomas, Forrester, and Ronald (2013) on the appropriate range of intrinsic versus extrinsic variation to capture data on past tense acquisition, family quotients were sampled from a uniform distribution from 60% to 100% of the perfect training set, corresponding to learning environments with reasonably high quality. This translates to the assumption that there is at least a minimum amount of linguistic information typically available to a child.

Note that, in principle, the extrapersonal environment may also play a role in influencing the value of neurocomputational parameters across child development, for example via prenatal maternal nutrition, post natal diet, stress, and other effects on brain development (see Hackman, Farah & Meaney, 2010; Thomas, Forrester & Ronald, 2013, for discussion). Whether environment primarily affects neurocomputational properties or the subjective information content of the environment may depend on the absolute level of

SES. For the purposes of the current model, we restricted extrapersonal environmental effects to information.

Environments were determined independently of artificial genomes. That is, we assumed no gene-environment correlations in our initial simulations.

2.3 Behavioral level

The past tense was an advantageous illustrative domain because the same processing system acquired both regular verbs and irregular verbs (Rumelhart & McClelland, 1986). The dimension of regularity permitted consideration of the specificity of simulated gene-behavior associations: were observed artificial gene-behavior associations always the same for regular verb performance as irregular verb performance or could they differ? Some degree of specificity might be predicted because it is known that in artificial neural networks, the two verb types are differentially sensitive to variations in the neurocomputational parameters (Kello, Sibley & Plaut, 2005; Mareschal et al., 2007; Thomas & Karmiloff-Smith, 2003b). Results will focus on the contrast between regular verb performance and performance on the most common irregular verb type, vowel-change irregulars.

2.4 Network structure and activation

When used as cognitive models, artificial neural networks are fairly rudimentary in terms of neural realism. Nevertheless, they can still offer some suggestive ideas on the relation of brain to behavior. For example, for the networks we used, two different network

properties showed similar developmental trajectories to those observed in, respectively, global gray matter volume and global white matter volume: total number of connections, and total magnitude of connection strength (both excitatory and inhibitory). This analogy is of course, tentative. There is more in both gray and white matter than connections. However, as well as cell bodies, gray matter does include facilities for local connectivity (dendrite arbors, synapses); and white matter includes myelin that enhances axonal conductance, reflecting activity-dependent strengthening of long-range connections. The analogy between properties of the model and these two types of brain matter is based on their respective developmental profiles (Gogtay et al., 2004; Shaw et al., 2008). After the onset of pruning, gray matter and number of connections in the model both show an exponential decline, while both white matter and total connection strength show a linear increase. In the model, number of connections offers plasticity, such that the network's ability to change reduces as pruning takes place, in line with sensitive periods observed in the cognitive system (Thomas & Johnson, 2006); while increasing connection magnitude reflects experience-dependent strengthening, in line with white matter changes that are observed during skills acquisition (Bengtsson et al., 2005; Scholz, Klein, Behrens & Johansen-Berg, 2009). These two metrics, total number of network connections and total connectivity magnitude, served as our indices of network structure, measured independently of behavior.

In addition, we took a measure of the activation states within the network. Individual networks varied in the number of hidden units they possessed. The average activity across the hidden units (that is, the sum of activation divided by the number of hidden

units in that network) was calculated, either over items in the training set or over items in the generalization set.

2.5 Neurocomputation

Artificial neural networks contain a range of parameters that increase or decrease their ability to learn a given training set. Parameters such as learning rate, momentum, and number of hidden (internal) processing units feature in most published simulations. In models of normal/average development, parameters are optimized to achieve best learning (usually in the presence of the perfect training set). In the current model, a number of parameters were simultaneously varied across individual networks, with learning ability determined by their cumulative affect. Multiple parameters were varied at the same time to reflect the expectation articulated by Sapolsky (2005) that many low-level neural properties are likely to vary between individuals. Variations occurred over fourteen computational parameters, in principle allowing for over 2 trillion unique individuals. Parameters determined four broad properties of the artificial neural networks: *network construction*, *network dynamics*, *network adaptation*, and *network maintenance*.

In line with the arguments of Plomin and Kovas (2005), the parameters had general computational functions, and no specific relation to the problem domain that the system was acquiring. The parameters were as follows. *Network construction*: Architecture, number of hidden units, range for initial connection weight randomization, and sparseness of initial connectivity between layers. *Network dynamics*: unit threshold function (or 'temperature'), processing noise, and response accuracy threshold. *Network*

adaptation: backpropagation error metric used in the learning algorithm, learning rate, and momentum. *Network maintenance*: weight decay, connectivity pruning onset, pruning probability, and pruning threshold. These parameters have derivations in neurocomputational theory, and differences in their settings have been used in models to simulate variations in cognition, including those found in general intelligence, specific language impairment, dyslexia, schizophrenia, autism, and ageing (see Supplementary Materials for description and citations). A range of variation in the population was established for each parameter (see Supplementary Materials for details of the calibration procedure, as well as plots of the sensitivity of network performance to variations in each parameter). Model performance was fairly robust to variations in each parameter: calibration was carried out to establish extremes.

2.6 Genetic level

An artificial genome was created, variation in which produced variation in the neurocomputational parameters. We assumed that a full genome would contain three portions, of which we only implemented one. The first portion would be genes not relevant to the functioning of our modeled system (though if measured in a GWAS, variations in these genes would be candidates to produce false positive associations). The second portion would be genes that were species universal and did not vary across individuals, and whose on-going dynamics of expression and regulation delivered the functionality of the network itself, in terms of the existence of processing units, connections, activation dynamics, the sensorium, the input-output connectivity, and the mechanics of experience-dependent learning systems. The third portion would be genes

that were influential in the initial growth of the network, and which influenced particularly the effective computational properties of the system once its experience-dependent properties came on-line. The neurocomputational properties were therefore conceived of as the outcome of a growth process, in the way that the number of neurons in different brain areas is the outcome of neural proliferation and migration. This portion of the genome was assumed to show variation across individuals, and was the only portion we implemented. For simplicity, we assumed that the relevant genes were the sole source of variance in the growth of neurocomputational parameters (i.e., contributing biochemical environmental factors were constant across individuals) and that the relationship was non-stochastic. We stipulated that multiple genes would contribute to the setting of each parameter (*polygenicity*), but did not implement *pleiotropy*, where a single gene could contribute to the setting of more than one parameter. The values of the neurocomputational properties for each individual were encoded in the artificial genome.

The idea of encoding the properties of a computer program in the form of an artificial genome is familiar from the machine learning technique of *genetic algorithms*. Genetic algorithms are a method of optimizing computer programs by breeding generations of programs and selecting the 'fittest' (according to performance on the target problem) to populate the next generation (see Mitchell, 1997, for introduction). In principle, genetic algorithms can be applied to any computer program. The minimal requirement is that the parameter settings for the program (here, artificial neural network) 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, that is, one

that obeys the syntax of the computer language. The combination of artificial neural network models, genetic algorithms, and population modeling has been used extensively to consider how evolution may serve to optimize properties for learning, for instance in the domain of language (e.g., Batali, 1994; Reali & Christianson, 2009).

For the current model, we encoded the values of the 14 neurocomputational parameters in an artificial genome and then produced a population of 1000 individuals with randomly created genomes. We did not produce further generations via breeding and selection, with one exception: in related work, we used breeding alone to create monozygotic and dizygotic twin pairs from the initial population. This allowed us to simulate twin study designs and thereby assess the heritability of various properties of the population, such as behavior and network structure (Thomas, Forrester & Ronald, in preparation; see Kohli, Magoulas & Thomas, 2012, for further discussion of the technique).

The artificial genome contained several simplifications. Our starting point was to create conditions that allowed a fair opportunity to observe gene-behavior associations. We therefore created a population where genetic variation rather than environmental variation was responsible for the majority of individual differences in behavior (i.e., behavior was highly heritable); and we allowed gene variants to be common, so that there was no reduction in statistical power associated with rare variants. Artificial genes were binary digits, holding the value 1 or 0. Thus there were only two variants of each gene. We consider populations where these variants were equally frequent (so the initial population of random genomes was generated by setting each bit to 1 or 0 with 50% probability of

each); or where one variant was more common than the other (either: 1-valued alleles had 70% probability and 0-valued alleles 30% probability, or the reverse). Several binary genes encoded the value of each parameter, with more binary genes employed where a parameter took up a wider range of values. For example, the unit threshold function was encoded over 10 binary genes. The binary gene set was converted into a parameter value using the following method. The number of 1-valued alleles was summed. A look-up table was then used to convert the sum to a parameter value. Tables were constructed such that increasing sums corresponded to monotonic changes in the parameter. Intermediate valued sums corresponded to the average value of the parameter, and lower or higher sums corresponded to more extreme settings of the parameter in either direction from the average. An example of the lookup table for the unit threshold function is included in Figure 3. The full set of lookup tables is included in the Supplementary Materials.

The polygenic, binary coding of parameters ensured that average values were most common in the population, and more extreme values less common. Lookup tables were constructed to ensure that parameter changes above or below the average value corresponded to symmetric improvements or decrements in behavior. This meant that parameter value changes were not always linear. For example, the 'average' number of hidden units, ensuring a mediocre rate and final level of development, was 50 (with all other parameters at average values). Reducing this value to 30 caused poor development, but an equivalent improvement above average required an increase to 200. Such a non-linear relation from artificial genome to parameter ensured strong genetic effects, and

thus the best chance of observing these effects in single gene-behavior associations (Thomas, Forrester & Ronald, in preparation).

The artificial genome comprised 126 bits (split into two strings or chromosomes of 63). The numbers of binary genes per parameter were: hidden units: 10; unit threshold function: 10; processing noise: 8; learning rate: 12; momentum: 8; weight variance: 8; architecture: 6; learning algorithm error metric: 4; response threshold: 10; pruning onset: 10; pruning probability: 8; pruning threshold: 10; weight decay: 10; sparseness: 12. These values were determined during a calibration phase in order to accommodate different ranges of variation for the respective parameters in how they influenced behavior (though in principle, the number of genes per parameter could be held constant).

2.7 Simulation Design

Six populations of 1000 networks were run. In each case, (i) artificial genomes were generated at random; (ii) each genome was converted into an instantiated network; (iii) a family training set was created for the individual; and, (iv) development was tracked for 1000 epochs (presentations of the training set). The majority of results are reported from the first population, where the gene variants at each location on the artificial chromosome were equally frequent. We then considered five further populations in order to evaluate the replicability of artificial gene-behavior associations. First, we took the same set of genomes and exposed the networks to different environments. Second, we re-sampled the genomes with random binary values, but used the same lookup tables and therefore probabilistic distribution of the parameter values in the population; and then exposed

these networks to different environments. This was carried out twice to create two resamplings. Fourth, we re-sampled the genomes but now changing allele frequencies, with the 1-valued allele given 70% probability and the 0-valued allele 30%. The same look-up tables were used to convert artificial genomes to neurocomputational parameter values. Last, we re-sampled the genomes, but with the 1-valued allele now given 30% probability and the 0-valued allele 70%.

3. Results

We first consider the variability present in the behavior of the population. Figure 2 shows the population distribution of performance on regular and irregular verbs at three points in training, which we will refer to as early (50 epochs), mid (100 epochs), and late (750 epochs) in development. These points were chosen to capture different developmental phases, but before performance had entrenched at its final performance level. Table 1 shows the mean performance level and standard deviation for each past tense verb type at each measure point. These are the data at the behavioral level. At the genetic level, the artificial genome constituted 126 binary values per individual, for 1000 individuals. For a given point in development and a given behavior, a correlation could be computed between the value of each artificial gene (1 or 0) and the target behavior. In what follows, we report the variance explained by the association (that is, the square of the correlation). Associations had to exceed a certain size to be rated greater than chance. This threshold was determined via bootstrap methods, by repeatedly generating a random gene (with two possible values, 0 and 1) and associating variations in this gene to the target measure. One thousand iterations generated a distribution of the association sizes one might expect

by chance. Ninety-five percent and 99% confidence intervals could then be generated for this distribution to identify the association sizes that would occur by chance less either than 1 in 20 times or 1 in 100 times. A similar approach was used to compute how large a difference between two associations had to be before it could be viewed as significant. For most target measures, the 0.05 criterion corresponded to an effect size of around 0.5% and the 0.01 criterion to an effect size of around 0.75%. At these levels, for each 100 other unrelated genes on the (unimplemented wider) genome that one associated with the behavioral or structural measure, 5 would be expected give false positive associations at the .05 level and 1 would be expected to give a false positive association at the .01 level. We could have used more sophisticated methods that corrected for multiple comparisons but chose not to, first for the sake of simplicity, and second because the sources of variation in the modeled system were well understood.

<Insert Table 1 and Figure 2 about here>

Figure 3 shows the possible associations between different levels of the model, for one neurocomputational parameter, the unit threshold function or 'temperature'. Figure 3(a) shows the relationship between the parameter value and behavior on irregular verbs established during calibration. Like many neurocomputational properties, the relationship is non-linear. Figure 3(b) shows this relationship when plotted from the full population, with unequal frequencies of parameter values and all other parameters varying, in this case at the early point of development. Extreme values of the parameter were relatively less frequent than the average value. Figure 3(c) shows the lookup table that was used to

convert the binary genes to the parameter value. Figure 3(d) shows the associations that were then observed between genes and behavior, when behavior was plotted according to genotype.

We now turn to our equivalent of a GWAS, examining effect sizes across the full 126-bit artificial genome. We report the results relevant to our 8 questions.

<Insert Figure 3 about here>

3.1 Can statistically significant associations be observed between artificial gene variants and individual differences in behavior, given many-to-one gene-to-neurocomputational parameter mappings and an intervening experience-dependent developmental process? Figure 4 depicts the association size between the *neurocomputational parameter* values and *behavior*, using individual linear regressions. It demonstrates there are large effect sizes, which are modulated both by behavior type (regular vs. irregular mappings) and over development. Were these associations observable at the level of artificial genes? Figure 5 shows associations between genome and behavior, again split by regular and irregular verb type, and for three points in development. Ninety-five per cent confidence intervals on effect sizes were produced by generating a random binary allele for each individual and using this to predict the individual's behavioral score; this procedure was repeated 1000 times to generate a distribution of effect sizes; the distribution was used to derive the effect size value that would be produced by chance less than 1 time in 20. The significance levels were therefore specific to the population size that was simulated.

A number of gene-behavior associations were indeed observable, despite the fact that the genes acted only on parameters in a many-to-one fashion and that behavior was the outcome of a variable developmental process. Early in development, for regular verbs there were 33 reliable associations from artificial genes to behavior at $p<.05$ and 24 at $p<.01$ out of a possible 126. For irregular verbs, there were 40 reliable associations at $p<.05$ and 26 at $p<.01$. By chance, 6 or 7 would be expected at .05 and 1 or 2 at .01. Across all three stages of development, effect sizes ranged from 0 to 4.4% of the variance (mean effect size: 0.4% standard deviation: 0.6%); 91 of the effect sizes fell between 0 and 0.5%, 19 between 0.5 and 1.0%, 8 between 1.0 and 1.5%, and 8 were greater than 1.5%. Larger effect sizes were seen on regions of the artificial chromosome influencing the neurocomputational parameters which themselves showed larger effect sizes on behavior in Figure 4. On the whole, a substantial number of small effect sizes were seen in the associations between artificial gene variants and behavior, despite the interceding developmental process.

Artificial gene variants were also assessed by their ability to predict whether an individual would fall in the top 10% or bottom 10% of the population by rank (simulating precocious or delayed development). Individual artificial gene variants altered the likelihood of falling in the tails of the population distribution by a maximum of 2.89 times (mean: 1.15, standard deviation: 0.16); 56 of the ratios were between 1 and 1.1, 35 between 1.1 and 1.2, 20 between 1.2 and 1.3, and 15 were greater than 1.3 (recall, an odds ratio of 1.1 means 11 individuals with the variant will show the phenotype, for

every 10 with the variant who will not). Thus artificial gene variants could predict performance in the tails, with relatively modest odds ratios.

<Insert Figures 4 & 5 about here>

3.2 Do associations show specificity to different behaviors generated by the system or are they general?

We compared associations to performance on regular verbs and irregular verbs, early in development. Once more, bootstrapping methods were used to derive 95% confidence intervals on the differences between effect sizes. Out of the 126 possible associations, there were 37 that differed significantly in effect size at $p<.05$ between the two types of behavior, 10 where effect sizes were larger for regular verbs, and 27 where they were larger for irregular verbs. Twenty-seven differences were significant at $p<.01$, 10 where effect sizes were larger for regulars and 17 where effect sizes were larger for irregulars. Thus, despite the general nature of the neurocomputational parameters, and the absence of processing structures specific to the types of behavior, associations from artificial genes to behavior could demonstrate specificity to behavior type. However, the majority of associations were not significantly different across the two behaviors, in line with the fact that these behaviors were generated by the same network structure.

3.3 What is the stability of the associations over developmental time?

Associations changed over development. Focusing on regular verbs, between early and mid development, there were 12 significant differences in effect size at $p<.05$ out of a possible 126. Seven of these 12 were cases where effect sizes were larger early in

development, 5 were cases where they were larger in mid development. Only 3 developmental changes in associations were reliable at $p<.01$, all larger early in development. A comparison between early and late development revealed 33 reliable differences in effect size at $p<.05$, 14 where effect sizes were larger early and 19 where they were larger late. There were 20 differences reliable at $p<.01$, 7 where effect sizes were larger early, 13 where they were larger late. Thus associations between artificial genes and behavior could both decrease and increase across development within the model.

Figure 4 indicates that developmental sensitivity was also apparent in the associations between neurocomputational parameters and behavior, with some associations strengthening across development and some weakening. Within a given neurocomputational parameter, the rank order of performance between individuals with different settings of the parameter value was generally stable across development. However, it was possible to find cases where individuals with one parameter value scored higher than individuals with another parameter value earlier in development, while later the order was reversed. For example, after 30 epochs of training, the 212 individuals with the temperature value of 1.25 scored higher on irregular verbs than the 254 individuals with a temperature value of 1.00 (23.9% versus 21.6% accuracy), while by epoch 200 the pattern of performance had reversed (60.6% versus 63.8%; interaction of epoch x parameter value: $F(1,464)=8.31$, $p=.004$, effect size $\eta_p^2=.018$). The behavioral advantage to an individual of possessing a given neurocomputational parameter value could, therefore, be specific to a particular developmental stage.

3.4 Do associations replicate across populations?

Figure 6 displays between-level associations when the same set of artificial genomes was instantiated as a new set of networks, and trained in new randomly sampled environments. The figure incorporates the effect sizes between neurocomputational parameters and behavior, and between artificial genes and behavior. We picked one of the behavior types, irregular verb performance, and one developmental stage, early, for our comparisons. There was a fairly close replication of associations at both levels. For artificial gene-behavior associations, there were only 8 significant differences at $p<.05$ and 3 at $p<.01$, close to chance levels. Figure 7 depicts the same plots when a new set of artificial genomes was sampled, with the same allele frequencies and parameter frequencies across the population; these new genomes were instantiated as networks and trained in new environments. Figure 7 includes two such resamplings. Here, the replication was fairly good at the neurocomputational-to-behavior level, but poorer at the artificial gene-to-behavior level. For the first resampling, 39 associations were significantly different at $p<.05$, and 17 were significant at $p<.01$, out of 126. For the second resampling, 36 associations were significantly different from the original at $p<.05$ and 20 different at $p<.01$. Figure 8 depicts the situation where allele frequencies were changed, either making the 1-valued allele more frequent than the 0-valued (70:30), or less frequent (30:70). Once more, a population of genomes was generated, instantiated as networks, and trained in new environments. Replication was now poor for both neurocomputation-to-behavior and gene-to-behavior associations. For the latter, there were 54 significant differences between the original and the 70:30 population at $p<.05$

and 33 at p<.01. There were 41 significant differences between the original and the 30:70 population at p<.05 and 30 at p<.01. In sum, replication was variable, depending on the details of the resampling, and the levels between which associations were observed.

<Insert Figures 6, 7 and 8 about here>

3.5 Are associations observed from artificial genome to network structure and activation, and if so, are these the same as the associations observed from genes to behavior (network function)?

Figure 9 shows associations between the artificial genome and two indices of network structure, the total magnitude of network connectivity, and the total number of connection weights, for early in development. Associations for irregular verb behavior (network function) are also included for comparison. Large effect sizes were apparent for both magnitude and number, with 28 and 15 associations significant at p<.01, respectively. When these two structural indices were compared with the effect sizes for irregular verb behavior at the same point of development (which had 26 reliable associations at p<.01), there were 41 and 35 significant differences at p<.01, for magnitude and number, respectively. In other words, for connection magnitude, 13 associations were shared with behavior and 41 differed, while for connection number, 6 were shared and 35 differed. Thus, the majority of the associations between artificial genes and network structure, and between artificial genes and behavior (network function), were separate – even though it was the structure of the artificial neural networks that generated their behavior.

This is perhaps not surprising given the correlations between these structural indices and behavior. Table 2 shows the correlation matrix for structural (number and magnitude of connections) and functional (regular, irregular performance) indices for early, mid, and late in development. It reveals a pattern of strong correlations within structural indices and within functional indices, but weak correlations between structural and function indices. This pattern has also been observed in empirical studies (example data from Posthuma et al., 2003, are included in Table 3 for comparison). In the model, while, to some extent, more total connections necessarily entails greater total connection strength, the correlation is not guaranteed. Several factors can modulate the relationship. These include differential loss of connections through pruning, differential decay of connection strengths, differential strengthening of connections due to variations in learning environments, and the differential effect of other parameters that modulate how learning experiences strengthen the connections. Together, these factors can all serve to weaken the initial correlation between the two structural measures. This is confirmed in Table 2, which demonstrates how their correlation weakens over development.

We next assessed the correlation across individuals between measures of network structure and network activation, where the latter was calculated by the average hidden unit activation levels produced while generating behavior.[2] Hidden unit activation states were very similar when processing items in the training set and items in the generalization set (a correlation of 1.00), though greater activity was induced in networks by novel items than by items in the training set (training set: mean = .295, standard

[2] This calculation could only be performed for networks with hidden layers. The calculations therefore excluded the 102 networks with only a 2-layer architecture.

deviation = .159; novel: mean = .328, standard deviation = .179; t(897) = 50.19, p<.001, Cohen's d = .198). Novel items have also been observed to induce more neural activity than familiar items in some functional brain-imaging experiments, an effect which has been ascribed to greater neural efficiency in processing the latter (see, e.g., Poldrack, 2014). In the model, more activation represented less certainty about the identity of the input. However, the neural realism of the distributed codes acquired in backpropagation networks is too remote infer any much from the similarity to brain imaging results. Correlations between structural measures and activation levels were high, .89 for connection number and activity, and .63 for connection strength and activity (both p<.01). Figure 10 shows associations from the artificial genome to, respectively, number of connections, activation induced by processing novel verbs, and the generalization performance on novel verbs. Associations for activation states more closely tracked differences in the connectivity of the network rather than behavior. That is, variations in the representational codes across networks were tied to structural properties of those networks rather than how well the networks were performing in inflecting novel verbs.

<Insert Figures 9 and 10 about here>

<Insert Tables 2 and 3 about here>

3.6 Are associations modulated by the quality of the environment, producing gene x environment interactions?

Our illustrative model was drawn from the study of language development, where in another context, it has been used to simulate SES effects on past tense acquisition via

modulation of the information content of the environment (Thomas, Forrester & Ronald, 2013). Did variations in this information content affect the associations observed between levels? SES was modeled by the family quotient factor, which served as a filter on the full training set, and which varied in value from 0.6 to 1.0. Note, by design, genomes were randomly assigned to environments. We split the population into high and low SES groups at a quotient of 0.8, yielding sub-groups of N=502 and N=498, respectively. Figure 11 shows the neurocomputational parameter-to-behavior and artificial gene-to-behavior associations for irregular verbs early in development. There were modulations of effect size by SES in both cases. For artificial gene-behavior associations, there were 39 associations out of 126 that significantly differed between high and low SES groups at $p<.05$ and 24 at $p<.01$ (with confidence intervals recalculated to reflect the smaller sample size). An equivalent analysis of SES effects on regular verb associations yielded 38 at $p<.05$ and 15 at $p<.01$, respectively. This result demonstrates evidence of gene-environment interactions in our model system, at least in the way that SES modified gene-behavior associations. But did these effects translate into a modification of the relationship between SES and *behavior* according to genotype? We took the artificial gene with largest effect from Figure 11 (gene no. 68, predicting 6% of the variance in the high SES group but only 1% in the low SES group). In the group of individuals with the 1-valued allele, the effect of SES was to modulate behavioral performance by 8.9% (high SES, accuracy=40.4, N=230 versus low SES=31.5, N=248); for the 0-valued allele, the effect of SES was a negligible 0.4% (27.7, N=272 versus 27.3, N=250). This gene-environment interaction had a small effect size of 0.8% of the variance, but was statistically significant in our sample size ($F(1,996)=8.10$, $p=.005$, $\eta_p^2 = .008$).

<Insert Figure 11 about here>

3.7 Can interactions between genes be observed in the way that they influence behavior?

By design, at the level of artificial genome, there were no causal interactions between the genes in the way that they influenced different neurocomputational parameters. Thus, for example, the value of the hidden unit parameter depended only on the values of the relevant artificial genes encoding this parameter, and did not depend, say, on the values of the artificial genes determining the learning rate parameter.

However, based on machine-learning principles, we viewed it as likely that computational parameters in an artificial neural network would interact with each other in their effect on behavior. We explored whether this phenomenon might then generate statistical interactions between different gene-behavior associations, for the artificial genes encoding different computational parameters.

We took two parameters, number of hidden units and learning rate, which we expected on computational grounds to interact in their effect on behavior. Figure 12 plots the population performance for individuals split by whether they had 40 or 50 hidden units (where more hidden units implies greater computational power), and whether their learning rate was 0.075 or 0.125 (where a higher learning rate indicates a more plastic learning system). We compared regular and irregular verb performance and contrasted early and late phases in development. 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 numerical difference was present for both verb types and both stages of development but was significant only for irregular verbs (regular, early: p=.173; late: p=.323; irregular, early: F(1,111)=4.14, p=.044, η_p^2 = .036; late: F(1,111)=5.53, p=.020, η_p^2 = .047). As we expected, then, these two neurocomputational parameters interacted in their effect on behavior.

We explored whether this interaction was visible in artificial gene-behavior associations. We picked two alleles with significant associations, one from the hidden unit (HU) region and one from the learning rate (LR) region (effect sizes of 1.39% and 2.37%, respectively). We compared them with two alleles from these regions that showed non-significant associations (0.00% and 0.09%). The alleles with significant associations showed main effects but did not exhibit an interaction (main effect of HU: F(1,996)=13.78, p<.001, η_p^2 = .014; main effect of LR: F(1,996)=24.09, p<.001, η_p^2 = .024; HU x LR interaction: F(1,996)=1.27, p=.261, η_p^2 = .001). By contrast, the alleles *without* individually significant associations showed no main effects but a reliable interaction (main effect of HU: F(1,996)=.03, p=.871, η_p^2 = .000; main effect of LR: F(1,996)=1.21, p=.271, η_p^2 = .001; HU x LR interaction: F(1,996)=4.07, p=.044, η_p^2 = .004). The observed interaction was in the expected direction given Figure 12: for the 1-valued hidden unit allele (contributing to more hidden units), individuals with the 1-valued learning rate allele (contributing to more plasticity) scored higher. For the 0-valued hidden unit allele (contributing to fewer hidden units), individuals with the 0-valued learning rate allele (contributing to less plasticity) scored higher.

<Insert Figure 12 about here>

3.8 When all mechanistic sources of variability are known, can all the population variability in behavior be explained?

Due to the relative range of variation of genetic and environmental factors in the current simulation, the majority of the population variability was due to genetic factors. When the simulations were extended to a twin study design (see Thomas, Forrester & Ronald, in preparation), the monozygotic (MZ) twin correlation for regular verbs early in development was .98 and dizygotic (DZ) was .40 (note, behavior was computed without measurement error). An MZ correlation more than twice DZ implies dominant genetic effects. Could the population variability in behavior exhibited by the simulated population be explained by the associations at each lower level of description, respectively at the neurocomputational and genetic levels? Since we knew the contribution to individual differences in behavior due to the environment (stemming from a single parameter, the family quotient factor), the remaining variance should be accounted for by the genetically determined neurocomputational parameters. Together, the family quotient factor and the neurocomputational parameter values should predict all the population variance in behavior.

The neurocomputational parameters, along with the measure of environmental quality (family quotient), were used in independent linear regressions to predict regular verb performance early in development. The summed variance of behavior explained by the

parameter values was 48.1%, less than half the population variability, with the family quotient factor accounting for 0.7%. Separate regressions inflated the variability explained due to (in this case, chance) correlations between the neurocomputational parameters, thereby double-counting some of the variance that the parameters predicted. Simultaneously entering the parameters in a multiple linear regression reduced the explained variance to 43.4%. The inflation of independent fits was therefore around 5%.

Linear methods were not entirely appropriate, however. As exemplified in Figure 3(a), in most cases the relationship between a neurocomputational parameter and its effect on behavior was *non-linear*, with the appropriate function differing depending on the parameter. The best non-linear fit was computed for each neurocomputational parameter-behavior relationship from the set {linear, log, inverse, quadratic, cubic, power, logistic, growth, and exponential}. If only non-linear functions with two regression-parameters were used (the same number as a linear function), the total variance explained now rose to 70.1% (though this includes the inflation due to independent fitting). If non-linear functions with 2, 3, or 4 regression-parameters were permitted, the explained variance in population behavior rose to 77.1%, although the additional 7% explained variance was gained at the expense of 17 more regression-parameters (degrees of freedom). The maximum explained variance, combining knowledge of neurocomputational parameter values and environmental quality, was a little under 80%.

One possible source of the additional variance was higher order interactions between neurocomputational parameters. We saw once such interaction in the previous section.

Given that there were 14 parameters, there were a large number of possible interactions. To test the principle that interactions might account for missing variation, we re-ran the linear multiple regression, but now entering several interaction terms. These terms were educated guesses based on computational theory, and involved interactions between parameters such as hidden unit number (H), learning rate (LR), temperature (T), sparseness (S), architecture (A), response threshold (RT), noise (N), initial weight variation (W), and family quotient (FQ). (Two examples: a lower activation function temperature might mitigate the entrenchment caused large initial weights; a less representative view of the latent structure of the problem domain caused by a low family quotient might be mitigated by a more tolerant response threshold). Of the dozen interaction terms we guessed (entered into the regression as products of the parameter values), 5 explained statistically reliably amounts of the variance. These included three 2-way, one 3-way, and one 4-way interaction (H*LR, H*T, T*W, LR*RT*FQ, H*LR*T*S). The total variance explained in this regression model rose from 43.4% to 46.0%, a gain of 2.6%, thereby confirming that interactions between parameters could account for some of the missing population variance.

Finally, turning to the artificial-gene level, summing all associations plus variance explained by the environment yielded a total of 79.4%. Again, this method includes inflation due to independent fitting. Simultaneously entering all alleles plus environment into a multiple linear regression yielded a total of 61.3% variance explained.

In sum, in the absence of measurement error, subtracting the known contribution of the manipulated environmental factor to population variation in behavior, and the known contribution of stochastic factors computed from MZ correlations, we expected the other deterministic mechanistic factors producing variability to explain up to 97% of the variance. These mechanistic factors represented the genetic contribution to individual differences. However, only around 80% of the variance could be explained by these factors. In these simulations, one could say that around 20% of the variance expected to be explained by genetic factors was ‘missing’.

4. Discussion

We begin by considering the following specific question: For the preceding analyses, do observed cross-level statistical associations give an accurate picture of the causal processes which, with knowledge of the operation of the model, we know generated the behavior? We then turn to consider the broader theoretical issues raised by the multi-scale model, as well as the limitations of the simplified modeling framework.

4.1 Correlation and causality within the model

Small but statistically reliable associations were observed between the artificial genome and behavior from around a quarter of the alleles on the artificial genome. These were observable through the filter of the genes' many-to-one impact on neurocomputational parameters in a system that engaged in an extended, experience-dependent developmental process. On the one hand, this is impressive. On the other hand, no artificial genes were included in the genome that *did not* influence neurocomputational properties. Therefore, *every* artificial gene was causal. For three quarters of the artificial genes, there was causality without statistically significant correlation. There were two reasons why only a quarter showed reliable correlations to behavior: the polygenic relationship between the artificial genes and the neurocomputational parameters, and the differential predictive power of the neurocomputational parameters that they influenced. For the former, the reliable associations corresponded to the genes that happened, by chance, to contribute to setting the value of the computational parameter *for this population*. That is, causation was not fully manifested in correlations because of sampling. The divergence between correlation and causation was possible because of the many-to-one mapping between

artificial genes and neurocomputational parameters, and the many-to-one relationship between neurocomputational parameters and behavior.[3]

Some associations showed specificity to different behaviors. For this model, the specificity of artificial gene-behavior associations did not imply specificity of computational mechanisms responsible for processing each behavior type. Regular and irregular past tenses were generated by the same structure via parallel distributed processing. Specificity of associations occurred because the two behaviors had differential sensitivity to variations in different neurocomputational parameters. Therefore, the behaviors were able to show different associations to genes influencing the setting of those parameters. For example, irregulars are harder to learn and require more computational power. As a consequence, irregulars are more sensitive than regulars to

[3] Where several genes contribute to the setting of a neurocomputational parameter, in small samples, it may turn out that statistically, variation in some genes contributes disproportionately to predicting the value of the parameter (and by extension, its influence on behavior). As the population sample size gets larger, the combined contribution of the set of genes should become more apparent. The effect sizes of associations should become more even across the set. We verified this with a simple example where 10 binary artificial genes were used to determine the value of a notional parameter in an additive fashion. In a population of N=1000, the effect sizes of the associations between individual genes and the subsequent parameter value were somewhat uneven, with a mean of 11.1% and a standard deviation of 1.4% across the 10 artificial genes. When the sample was raised to 3000, the effect sizes become more even, with a mean of 10.3% and a reduced standard deviation of 0.8%. With a sample of 10,000, the mean effect size was 10.1% and the standard deviation was again reduced at 0.6%.

variations in the architecture of the network, one of the key determinants of processing power. In turn, variation in irregular verb performance can then show larger associations to variants of the artificial genes determining the architecture.

A particular relationship between variation in certain high-level behaviors and variation in certain low-level neurocomputational properties has been referred to as 'domain-relevance'. The concept has been used to explain why uneven cognitive profiles can occur in developmental disorders in the face of apparent brain-wide genetic effects (Karmiloff-Smith, 1998). A brain-wide parameter difference may differentially impact on behaviors for which the parameter is more developmentally relevant. In sum, while specificity of gene-behavior associations could indeed imply specificity of processing mechanisms, it need not and did not in our model. Instead, it could imply domain relevance of processing properties to problem domains.

We observed that some associations changed across development, either increasing or decreasing in size. In terminology sometimes used in association studies, the 'genetic architecture' of the system altered across development. However, by design, in the simulations there was no alteration in the genetic influence on variation in the neurocomputational parameters across development; the genes were taken to influence growth processes that led to a network with certain learning properties. In our simulations, associations changed, either rising or falling, because the computational properties that they influenced became more or less relevant to behavior at different phases of development. In the same way as computational parameters can be 'domain-

relevant', they can be 'phase-relevant'. For example, the response threshold parameter (indexing the notional settling of attractor networks at output; see Supplementary Materials) determined how 'clean' a response had to be before it could generate a behavioral output. Variations in the response threshold were more influential early in development when processing was less accurate; but when accuracy increased later in development, variations in the response threshold themselves became less relevant. By contrast, variation in the learning algorithm became increasingly relevant because it determined the final representational states that could be reached by the system by the end of development. In sum, developmental changes in gene-behavior associations could indeed (and presumably often do) imply changes in gene expression – after all, in many cases, biological development is defined by changes in gene expression; but they need not and did not in our model. Instead, they could imply phase-relevance at the computational and genetic levels.

Associations showed poor replication across populations. This was not due to an intrinsically noisy developmental process – replication of artificial gene-behavior associations was good if the population set of genomes was re-instantiated in a different set of randomly assigned environments. Lack of replication arose when the genomes were re-sampled, even with the same probabilistic distributions of parameter values. This is because, through polygenic coding, different alleles could be responsible for producing the same computational value in different populations. Neurocomputation-behavior associations were, however, more robust. If a move from a low level to a high level of description involves a sequence of many-to-one causal relations, associations become

more robust as more causal factors are resolved into fewer. This is related to the idea of *endophenotypes* (de Geus et al., 2001; Gottesman & Gould, 2003; Kendler & Neale, 2010). Proponents of endophenotypes argue that intermediate levels of description between the molecular level of genes and the whole system level of behavior are more likely to show links to the genetic level. For the model, we observed that measures at the intermediate level showed stronger and more replicable links to behavior. However, if allele frequencies differed between populations, while associations were still observed, these differed, even for neurocomputation-behavior mappings. This was because the computational balance of the systems had changed. For example, the population in which 1-alleles had 30% frequency and 0-alleles had 70% frequency, the corresponding computational parameters were less optimal and population performance was poorer. Networks, therefore, tended to rely on the response threshold far more to accept 'just good enough' output activations as correct answers, exaggerating the predictive power of variations in the response threshold parameter. Further simulations are required to consider scenarios where unequal allele frequencies are the norm. However, the implication of the differential allele frequency conditions was as follows. Given there will be a function linking the set of gene variants to their effect on neurocomputational properties, the cross-level associations that are observed will be influenced by the frequency of the different variants in a given population. Overall, then, the results point to the population-specific nature of between-level associations, and that many-to-one causal relations can lead more distant levels of description to have less replicable associations than more proximate ones (where distance refers to a hierarchy of larger components made from smaller components).

Associations were observed between artificial genome and both network structure and activation states. However, these were different from the associations observed between artificial genome and network function (behavior). This is consistent with the low correlation found between individual differences in network structure properties and individual differences in behavior, a divergence that has also been observed empirically (e.g., Posthuma et al., 2003). To some extent, one might expect weaker correlations between structure and function within the artificial neural network, at least at a global level, because the same structure has to produce different behaviors in a distributed processing system. Perhaps in our model, a more fine-grained analysis of network structure than total connectivity would have produced structural associations closer to those observed for behavior. This is far from guaranteed, because it was the same units and connections that processed regular and irregular verbs in this system, with specificity only arising via the different levels of activation propagating along different pathways (Thomas, Purser et al., 2012). One reason for the lack of overlap between genome-to-structure associations and genome-to-function associations was that some neurocomputational parameters contributed much more to structural variation. For example, the number of internal processing units greatly influenced structural measures based on total connectivity. However, behavior was more dependent on the *quality* of the processing occurring within that connectivity: function, therefore, was influenced by many other parameters with subtler effects not obviously detectable via the structural measures. One might expect a similar effect with current brain imaging techniques, since measures such as gray matter and white matter, or blood oxygenation, are unlikely to

capture all the properties that affect neurocomputation. Lastly, the degree of activation in the networks was more closely tied to structure than to function, indicating that alterations in computational capacity led to the adoption of different representational codes. In short, associations from gene-to-structure and gene-to-function can diverge, even when (as we know for the model) all the genes being measured influence aspects of neurocomputational processing.

Associations from artificial genes to behavior were reliably modulated by the quality of the environment (here, taking advantage of the fact that the model was drawn from work investigating the effects of socio-economic status on language development; the population could therefore be median-split into those developing in high SES and low SES families). It was also possible to identify artificial genes where the allele value altered the relationship between SES and behavior. In the simulated population, *gene-environment interactions* arose because those networks with better computational learning systems were more able to exploit the information available in better environments. Variation in performance due to the quality of the environment was therefore more apparent in those with higher ability than lower ability (Thomas, Forrester & Ronald, 2013). However, the proportion of gene-behavior associations showing modulation by SES was surprisingly high. Even though we expected gene-environment interactions for this model system, the overall behavioral effect sizes were relatively modest. For example, where intrinsic ability was taken to be the composite of all neurocomputational settings, and for performance early in development, the gene-environment interaction for regular verbs explained only 1.1% of behavioral variance

($F(1,996)=11.0$, $p=.001$), and that for irregular verbs explained only 0.4% of the behavioral variance ($F(1,996)=4.0$, $p=.047$; see Thomas, Forrester & Ronald, 2013, for the method of calculating these effects). The number of artificial gene-behavior associations modulated by the environment exceeded the size of the gene-environment interaction observed in behavior. The explanation is that many of these apparent modulations were a consequence of the between-subjects design – the low SES and high SES groups were different sub-populations; therefore one would predict the poor replicability of artificial gene-behavior associations discussed earlier. In short, the model suggests that although one might expect gene-environment interactions to be observed in gene-behavior associations, evidence of interactions may also be the artefactual / confounded consequence of measuring associations in populations with (stochastically) different genomes.

By design, artificial genes influencing variation in separate neurocomputational parameters did not interact with each other. Genes for a given parameter determined the value of that parameter independently of the genes for other parameters. Nevertheless, it was possible to detect statistical interactions between artificial genes for separate parameters in their associations with behavior. This is because the neurocomputational properties, which the artificial genes influenced, themselves interacted during the developmental process. In the example we gave, a system with more resources did better with higher plasticity than with lower plasticity, while a system with fewer resources did better with lower plasticity than 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. During training, this solution must be approached in the smaller iterative steps provided by a lower 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.

Finally, we used a system in which most of the population variability in behavior was caused by intrinsic factors, which we defined as genetic in origin. That is, the system generated highly heritable behavior. Given we knew all of the causal mechanistic settings that generated population variability in behavior (albeit via a developmental process), and given we had an estimate for the contribution of stochastic factors such as initial weight randomization of initial weights, pruning of weights, processing noise and randomization in exposure to the training set, could we then explain all of the observed behavioral variance, or was some of the variance 'missing'? It would be comforting if in a relatively simple system where the causal processes were transparent (even if some of the properties of the model were emergent), all the behavioral variability could be explained. However, around 20% of the behavioral variance remained unexplained. We identified two possible sources of this phenomenon in the simulations. First, there are limitations in the statistical techniques used to assess variance explained based on the predictor variables. In artificial neural networks, many of the relationships are non-linear. As we saw, use of linear methods under-estimates the variance that can be explained. Nevertheless, while use of non-linear statistical methods increased the amount of variance explained, it still left a fifth of the variance unexplained. Second, variance may

be left unexplained because there are complex interactions among the neurocomputational parameters, and between the parameters and the environment, during development. With many parameters, there are large numbers of possible interactions. We supported this source of unexplained variance by demonstrating that some sample interaction terms indeed accounted for reliable amounts of variance, although the few we chose only increased the explained variance by a small amount. To the extent that neurocomputational factors are genetically influenced, then, the interactions between them may constitute a source of missing heritability: variance that stems from genetic factors but that is not predicted by the factors in isolation.

4.2 Wider implications

Gene-behavior associations offer an exciting window onto the mechanisms by which the brain realizes cognition. Candidate gene association studies have suggested possible mechanistic pathways by which genetic variation produces individual variation, for instance via influences on neurotransmitter regulation, synaptic plasticity, or neural migration during development. Genome-wide association studies provide the opportunity for a systematic search for causal variants associated with variations in behavior. However, candidate gene studies have suffered from problems of replicability, while GWAS studies have had, as yet, more success in informing the biological pathways of common diseases than variations in high-level behavior.

Gene-behavior associations span many intermediate levels of description, including the cognitive level. What can gene-behavior associations tell us about cognition? Three

characteristics of associations are able to inform cognitive theories. These are *effect size*, *specificity*, and *timing*. A large *effect size* suggests how much of the causal pathway is being indexed by the genetic (or environmental) measure. *Specificity* suggests possible dissociations between mechanisms underlying different behaviors. Relatedly, modulation of environmental influences by genetic factors may point to mechanisms for resilience in development. With respect to *timing*, changes in associations over development may imply differential involvement of mechanisms at different ages (Ronald, 2011).

However, the use of genetic association findings to constrain cognitive theories is compromised by the complexity of the systems under consideration, and the fact that an extended developmental process is necessary before the emergence of high-level behaviors whose variation can be linked with genetic variation. We argued here that multi-scale models provide one method to investigate the relationship between associations that cross levels of description, and causal processes best characterized as operating at intermediate levels. In this case, we employed a modeling framework drawn from research on language development, which incorporated the levels of artificial genes, neurocomputation, network structure, behavior, and environment. Importantly, the model captured individual differences within a developmental framework. The results suggested the following.

Statistical associations spanning disparate levels of description will not always offer strong constraints on theories developed at intermediate levels of description, for a number of reasons. Specificity in associations may not be reflected in specificity of

mechanism. Timing effects in associations may arise for neurocomputational reasons without changes in genetic effects. Associations between structure and function may differ, even when genetic effects operate on the structure that realizes the function. Even without measurement error, non-linear relationships and complex interactions in learning systems may limit how much behavioral variance can be predicted from known parameters, leading to 'missing' variance. Many-to-one relationships between genes and neurocomputational parameters suggest inherent problems in replicability due to sampling differences across populations, and therefore difficulties with between-participants designs.[4] Some results from the model were more encouraging for the utility of cross-level associations. Measures that are intermediate to genes and behavior, where some of these many-to-one relationships have resolved, may improve replicability across populations, consistent with the idea of endophenotypes. Moreover, the presence of associations between artificial genes and behavior supports the principle that statistical associations can bear on intermediate-level mechanism, because in many cases these associations had clear computational explanations.

---

[4] To some extent, this result depends on the assumed scale of the model. We stipulated the granularity of the genomic encoding by virtue of our assumption of a polygenic relationship between genes and neurocomputational parameters. However, one could take a different view: that the 1s and 0s of the artificial 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 behavior, since each polymorphism would influence 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 behavior.

Multi-scale simulation framework that combines individual differences with development provide a foundation to consider wider issues, such as the causes of developmental deficits like autism, and mental health conditions like depression and schizophrenia. In particular, the specification of genetic and environmental causes of individual variation in high-level behavior firstly permits investigation of whether a disorder lies on a mechanistic continuum with normal variation; and secondly, where a distinct pathological effect is identified (of either genetic or environmental origin), how this effect interacts with protective and risk factors understood as population-wide causes of individual variation. For example, the current simulation framework has been applied to study of risk and protective factors for developmental regression in autism (Thomas, Knowland & Karmiloff-Smith, 2011), and the study of environmental factors contributing to the resolution of delay in language development (Thomas & Knowland, 2014).

4.3 How transferable are the model behaviors to real biological systems?

How severely to the simplifications of the model limit the generality of its findings to biological systems? In some senses, the modeling enterprise here is an unusual one. Mostly, models seek to capture a specific quantitative pattern of empirical data, or if they are more abstract (like the current model), seek to capture a wide set of phenomena using as few parameters as possible to provide a parsimonious causal account. In our model, we instead added a small degree of the complexity that we know exists in real biological systems. The aim was not parsimony but to evaluate the consequences of this complexity in drawing inferences from the kinds of cross-level association data emerging from

developmental cognitive neuroscience. We believe multi-scale modeling is an essential tool to address the complexity of the systems under consideration, but we recognize there is a clear tension in such models. This concerns simplification at the interface between levels of description. As Dammann and Follett (2011) put it, 'the trade-off between necessary simplification and necessary detail remains a major challenge in all computational modeling of complex processes. While the former is needed to achieve a reasonable level of modeling feasibility, the latter is needed to retain sufficient detail to render the model biologically meaningful. Moreover, assessment of reasons for model success or failure is difficult due to this tradeoff, especially in a multi-scale model, where important aspects of overall mechanistic complexity may have been sacrificed for the sake of modeling simplicity.'

One example of a simplification in the current model was the use of backpropagation networks to represent the neurocomputational level. The neural plausibility of the backpropagation algorithm has been questioned. At best, it represents a shorthand for a Hebbian-based algorithm that uses bidirectional connections to spread error signals throughout a neural network (Cowell, Bussey & Saksida, 2012; Thomas & McClelland, 2008; Xie & Seung, 2003). For a multi-scale model, contact with lower levels of description is important, and one might ask whether the use of the backpropagation learning algorithm restricts the generality of the findings. Certainly, it is possible that algorithms that are closer to those operating in neural systems might involve neurocomputational parameters with larger effects on behavior; if so, genes that influence their setting would produce larger associations in gene-behavior association studies. One

key distinction is between error-correction and self-organizing learning algorithms (O'Reilly, 1998). The former involves associations between codes, such as in the current model, while the latter involves the development of higher-order representations of input information without a training target. Kan et al. (2010) suggested that in self-organizing systems, initial (potentially stochastic) differences in start states could produce divergent developmental trajectories (see also Oliver, Johnson, Karmiloff-Smith & Pennington, 2000). Applied to the current framework, this would serve to reduce the size of gene-behavior associations. The choice of learning algorithm for a multi-scale model of development, and its implemented parameters, will clearly be important. The plausibility of the artificial neural network itself rests on a range of properties it shares with biological systems: its use of an associative network with distributed processing across a network of simple integrate-and-fire processing units, where behavior is acquired via an experience-dependent learning process involving interaction with a structured and variable learning environment, and the developmental trajectory and final representational states are constrained by parameters that have analogues in neurocomputation, such as the activation function of the neurons, the number of neurons, the connection density, the level of processing noise, and the onset and rate of pruning.

By design, the current modeling framework included significant simplification at the lower levels of description because it emphasized contact with the behavioral level, and the specification of a developmental process that was influenced by the information content of the environment. The gap between gene function – the production of proteins – and neurocomputational function remains large. Other models may emphasize inclusion

of more lower-level assumptions at the expense of making contact with high-level behavior. For example, the computational neurogenetics approach advocated by Kasabov and Benuskova (2004) restricts its focus to integrating the study of dynamic neuronal models and gene models. The ultimate challenge is to combine both.

Our model included assumptions that the relationship of genes to neurocomputational parameters is many-to-one, that gene variants relate to fairly general neurocomputational properties, and that gene variants are reasonably common in the population. These assumptions were sufficient to simulate a range of empirical effects, including the small effect sizes observed between gene variants and individual differences in behavior, the possibility that these associations can be behaviorally specific, the modest odds ratios when gene variants were used to predict performance in the tails of the population distribution, poor replicability of associations under certain conditions, and the divergence between structural measures and functional measures of the system despite tight correlations within these measures. We believe these results are likely to be transferrable to real biological systems.

The model's simplifications included a highly simplified and deterministic mapping from artificial genes to neurocomputational properties, a stationary environment, no gene-environment correlations, no alteration in the influence of genes on variation in neurocomputational processes during the model's acquisition of the domain (i.e., no consideration of earlier stages of biological development defined by changes in gene expression), two variants at each locus, an absence of rare gene variants with large

effects, no pleiotropy (i.e., genes only influenced variation in one parameter), no epistasis (interaction between genes), and no assortative mating. Moreover, since we only considered a single cognitive system, both the effects of developmental interactions with other systems, and issues surrounding the generality or specificity of genetic effects across multiple systems fell beyond the scope of the project. We should be frank, then, that this model only represents a small step, serving to demonstrate the importance of including multiple scale and combining development and individual differences in a single framework; serving to set out the implications for cross-level associations of the set of assumptions we initially incorporated; and serving to identify the way ahead for future models.

Gradual expansion of the complexity of the modeled system is necessary to evaluate how each of these simplifications would alter the main results with respect to *effect size*, *specificity*, and *timing* of associations. The results of expanding the complexity of the model are not necessarily anticipatable in advance. For example, pleiotropy might enhance gene-behavior associations if the multiple influences of a given gene variant on neurocomputation produced behavioral consequences in a similar direction; or pleiotropy might reduce associations if the influences mitigate each other. Gene-environment correlations might exaggerate associations, if the correlated environment contributes to the same behavioral characteristic that the gene is influencing (such as children with ADHD inheriting both genes influencing impulsivity and an unpredictable family environment); or gene-environment correlations might attenuate associations, if genes

and environment contribute opposite effects (such as a night-owls 'self-medicating' with coffee in order to be more alert in the mornings).

4.4 Conclusion

Associations between levels of description rely on the existence of individual differences at each level. In this paper, we have emphasized the importance of considering individual differences within a developmental framework. With respect to cognition, this implies an experience-dependent process involving interaction with a structured (physical and social) learning environment. What is the relationship between individual differences and development? We raised this question in the introduction and referred to theories that view them either as a single dimension or as different dimensions. The model's first important message is that this conceptualization may be incorrect. Individual differences and development are not two phenomena to be related. Instead, they are two views of the same thing. In a population, there are simply variations in developmental trajectories, with diverse genetic and environmental causes.

The model's second important message is that although one may be able to identify correlations between genes and behavior, this is only the beginning of the challenge – to understand these effects, one has to understand mechanisms at many different levels through which the effects are produced. Some of the model's findings could be deemed as skeptical about gene-behavior associations – for instance, as showing how hard it could be to learn anything from such associations in systems with many-to-one mappings and highly non-linear processes. The simulation was deliberately constructed in ways to enhance the possibility of finding gene-behavior associations. In

biological organisms, individual gene variants may have much smaller effects and so be harder to find. Perhaps one way to put the point is that if one cannot find correlations and interpret them the current model, the prospects with real cognitive systems would seem even more remote. It is therefore notable that even in the model, while all artificial genes contributed to variation, only some associations to behavior were detected for a given population.

One of the key motivations for constructing multi-scale models of complex systems is because the impact of individual assumptions cannot be anticipated in advance. The complexity of the underlying interacting non-linear processes necessitated simulation via computational methods. To finish, here are some of the main findings that we had not necessarily anticipated when we set out to build our model.

- Associations between artificial genes and behavior were observable despite an intermediate neurocomputational level of description where many-to-one causal relationships occurred, and despite extended developmental process involving interaction with a variable environment
- Larger effect sizes were seen on regions of the artificial chromosome influencing neurocomputational parameters which themselves showed larger effect sizes on behavior; but not all artificial genes in these regions showed significant associations.
- Despite the general nature of the neurocomputational processing properties and the absence of specific-specific processing structures, associations could be

specific to behaviors, due to the 'domain-relevance' of neurocomputational parameters.

- Associations between artificial genes and behavior could both increase and decrease across development without changes in gene regulation, due to the 'phase-relevance' of neurocomputational parameters.
- Replication of artificial gene-behavior associations was poor whenever the population of genomes was re-sampled (as in between-participant designs); but replication was better for associations between neurocomputational parameters and behavior.
- The majority of associations between artificial genes and network structure, and between genes and behavior (network function), were separate, even though it was the network structure that was generating the behavior.
- The environment could modulate the size of gene-behavior associations.
- The multi-scale model suggested some possible limitations on the inferences that can be drawn from cross-level associations in the absence of specification of intermediate level mechanisms.

Acknowledgements

We would like to thank Dorothy Bishop, Eddy Davelaar, Emma Hayiou-Thomas, Mark Johnson, Annette Karmiloff-Smith, Yulia Kovas, Denis Mareschal, Robert Plomin, Kostas Papageorgiou, and Essi Viding, for helpful discussions in preparation of this article. This research was supported by MRC Career Establishment Grant G0300188, ESRC grant RES-062-23-2721, and a Leverhulme Study Abroad Fellowship held at the University of Chicago.

References

Altshuler, D., Daley, M. J., & Lander, E. S. (2008). Genetic mapping in human disease. *Science*, *322*, 881-888.

Batali, J. (1994) Innate biases and critical periods: Combining evolution and learning in the acquisition of syntax. In R. Brooks and P. Maes, editors, *Artificial Life IV*, pages 160-171. Cambridge, MA: MIT Press.

Bechtel, W. (2001). Mechanisms and biological explanation. *Philosophy of Science*, *78*, 533–557.

Bechtel, W. (2008). *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. New York: Lawrence Erlbaum Associates.

Bechtel, W. and Mundale, J. (1999). Multiple realizability revisited: Linking cognitive and neural States. *Philosophy of Science*, 66(2), 175-207.

Bengtsson, S. L., Nagym Z., Skare, S., Forsman, L., Forssberg, H., & Ullen, F. (2005). Extensive piano practicing has regionally specific effects on white matter development. *Nature Neuroscience*, 8 (9), 1148-1150.

Benyamin, B., Pourcain, B., Davis, O. S., Davies, G., . . . & Visscher, P. M. (2013). Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. *Molecular Psychiatry*. Epub 29 Jan: 2013

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.

Charney, E. (2012). Behavior genetics and postgenomics. *Behavioral and Brain Sciences*, 35, 1-80.

Cowell, R. A., Bussey, T. J., & Saksida, L. M. (2012). Empiricists are from Venus, modelers are from Mars: Reconciling experimental and computational approaches in cognitive neuroscience. *Neurosci Biobehav Rev*, 36(10):2371-9

Craver, C. (2007). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Oxford University Press.

Dammann, O., & Follett, P. (2011). Toward multi-scale computational modelling in developmental disability research. *Neuropediatrics*, *42(3)*, 90-96.

De Geus, E. J. C., Wright, M. J., Martin, N. G., & Boomsma, D. I. (2001). Genetics of brain function and cognition. *Behavior Genetics*, *31(6)*, 489-495.

Eliasmith, C. (2002). The Myth of the Turing Machine: The Failing of Functionalism and Related Theses. *Journal of Experimental & Theoretical Artificial Intelligence*, 14(1), 1-8.

Eliasmith, C. (2013). *How to Build a Brain: A Neural Architecture for Biological Cognition*. Oxford: Oxford University Press.

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

Flynn E. G., Laland K. N., Kendal R. L. & Kendal J. R. (2013). Developmental niche construction. *Developmental Science*, 16(2), 296-313.

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 neural plasticity as an explanatory mechanism. *Psychological Review*, 109(1), 116-136.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., Nugent, T. F., Herman, D. H., Clasen, L. S., Toga, A. W. et al. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proc Natl Acad Sci USA*, 101, 8174-8179.

Gottesman, I. I. & Gould, T. D. (2003). The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry*, *160(4)*, 636-45.

Hackman, D. A., Farah, M. J., & Meaney, M. J. (2010). Socioeconomic status and the brain: mechanistic insights from human and animal research. *Nat Rev Neurosci*. 2010 Sep;11(9):651-9. doi: 10.1038/nrn2897.

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

Joanisse, M. F., & Seidenberg, M. S. (1999). Impairments in verb morphology after brain injury: A connectionist model. *Proceedings of the National Academy of Sciences of the United States of America*, *96*, 7592-7597.

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. (2010). Nonlinear epigenetic variance: review and simulations. Developmental Science, 13(1), 11-27.

Karaminis, T. N., & Thomas, M. S. C. (2010). A cross-linguistic model of the acquisition of inflectional morphology in English and Modern Greek. To appear in: S. Ohlsson & R. Catrambone (Eds.), *Proceedings of 32nd Annual Conference of the Cognitive Science Society*, August 11-14, 2010, Portland, Oregon, USA.

Karmiloff-Smith, A. (1998) Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences*, *2 (10)*, 389-398.

Karr, J. R., Sanghvi, J. C., Macklin, D. N., Gutschow, M. V., Jacobs, J. M., Bolival, B., Assad-Garcia, N., Glass, J. I., & Covert, M. W. (2012). A whole-cell computational model predicts phenotype from genotype. *Cell*, *150(2)*, 389–401.

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

Kello, C. T., Sibley, D. E., & Plaut, D. C. (2005). Dissociations in performance on novel versus irregular items: Single-route demonstrations with input gain in localist and distributed models. *Cognitive Science*, *29*, 627-654.

Kendler, K. S. & Neale, M. C. (2010). Endophenotype: a conceptual analysis. *Molecular Psychiatry*, *15(8)*, 789-797.

Kirov, G., Pocklington, A. J., Holmans, P., Ivanov, D., . . . & Owen, M. J. (2012). De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Molecular Psychiatry*, 17(2), 142-53. doi: 10.1038/mp.2011.154. Epub 2011 Nov 15.

Kohli, M., Magoulas, G. D., & Thomas, M. S. C. (2012). Hybrid computational model for producing English past tense verbs. In *Proceedings of 13th Engineering Applications of Neural Network Conference (EANN2012)*, London, UK, from September 20-23, 2012.

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.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., …, & Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, *461(8)*, 747-753.

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

Mareschal, D., Johnson, M., Sirios, S., Spratling, M., Thomas, M. S. C., & Westermann, G. (2007). *Neuroconstructivism: How the brain constructs cognition*. Oxford: Oxford University Press.

Marr, D. & Poggio, T. (1976). *From Understanding Computation to Understanding Neural Circuitry*. Artificial Intelligence Laboratory. A.I. Memo. Massachusetts Institute of Technology. AIM-357.

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

Morton, J. (2004). *Understand developmental disorders: A causal modelling approach*. Oxford: Blackwell Publishing.

Newbury, D., Winchester, L., Addis, L., Paracchini, S., …, & Monaco, A. P. (2009). CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *American Journal of Human Genetics*, *85*, 264-272.

O'Reilly, R. C. (1998). Six principles for biologically-based computational models of cortical cognition. *Trends in Cognitive Sciences*, *2*, 455–462.

Oliver, A., Johnson, M. H., Karmiloff-Smith, A., & Pennington, B. (2000). Deviations in the emergence of representations: A neuroconstructivist framework for analysing developmental disorders. *Developmental Science*, *3*, 1-40.

Paterson, S.J., Brown, J. H. , Gsodl, M. K. , Johnson, M. H. & Karmiloff-Smith, A. (1999) Cognitive modularity and genetic disorders. *Science*, *286*, 2355-2358.

Peñagarikano, O. & Geschwind, D. H. (2012). What does CNTNAP2 reveal about autism spectrum disorder? *Trends in Molecular Medicine*, *18(3)*, 156-63. doi: 10.1016/j.molmed.2012.01.003. Epub 2012 Feb 25.

Pinker, S. (1994). *The language instinct*. Penguin books

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

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2012). *Behavioral Genetics (6th Edition)*. New York: Worth Publishers.

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

Plomin, R., Haworth, C. M. A., Meaburn, E. L., Price, T. S., Wellcome Trust Case Control Consortium & Davis, O. S. P. (2013). Common DNA markers can account

for more than half of the genetic influence on cognitive abilities. *Psychological Science*. Epub 2013 March 15. DOI: 10.1177/0956797612457952

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.

Poldrack, R. A. (2014). Is "efficiency" a useful concept in cognitive neuroscience. *Developmental Cognitive Neuroscience*. Published on-line 13 June 2014. Doi: 10.1016/j.dcn.2014.06.001

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.

Posthuma, D., Baare, W. F. C., Hulshoff Pol, H. E., Kahn, R. S., Boomsma, D. I. & De Geus, E. J. C. (2003). Genetic correlations between brain volumes and the WAIS-III dimensions of verbal comprehension, working memory, perceptual organisation, and processing speed. *Twin Research*, *6(2)*, 131-139.

Potochnik, A. and McGill, B. (2012). The limitations of hierarchical organization. *Philosophy of Science*, *79(1)*, 120-140.

Reali, F. & Christianson, M.H. (2009). Sequential learning and the interaction between biological and linguistic adaptation in language evolution. *Interaction Studies*, *10*, 5-30.

Rietveld, G. A. et al. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. Science. 2013 Jun 21;340(6139):1467-71. doi: 10.1126/science.1235488. Epub 2013 May 30.

Ronald, A. (2011). Is the child 'father of the Man'? Evaluating the stability of genetic influences across development. *Developmental Science*, *14(6)*, 1471-1478.

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.

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

Scholz, J, Klein, M. C., Behrens, T. E. J., & Johansen-Berg, H. (2009). Training induces changes in white matter architecture. *Nature Neuroscience*, *12 (11)*, 1370-1371.

Shaw, P., Kabani, N. J., Lerch, J. P., Eckstrand, K., Lenroot, R., Gogtay., N, et al. (2008). Neurodevelopmental trajectories of the human cerebral cortex. *J Neurosci*, 28, 3586–94

Southern, J., Pitt-Francis, J., Whiteley, J., Stokeley, D., Kobashi, H., Nobes, R. Kadooka, Y., & Gavaghan, D. (2008). Multi-scale computational modelling in biology and physiology. *Progress in Biophysics and Molecular Biology*, *96*, 60-89.

Speliotes, E. K. et al. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet. 2010 Nov;42(11):937-48. doi: 10.1038/ng.686. Epub 2010 Oct 10.

The International Schizophrenia Consortium (2008). Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*, *455*, 237–241.

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. & Knowland, V. C. P. (2014). Modelling mechanisms of persisting and resolving delay in language development. *Journal of Speech, Language, and Hearing Research*, 57(2), 467-483.

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., Baughman, F. D., Karaminis, T., & Addyman, C. (2012). Modelling development disorders. In: C. Marshall (Ed.), *Current Issues in Developmental Disorders*. Psychology Press.

Thomas, M. S. C., Forrester, N. A., & Ronald, A. (2013). Modelling socio-economic status effects on language development. *Developmental Psychology*. 2013 Apr 1. Epub ahead of print.

Thomas, M. S. C., Forrester, N. A., & Ronald, A. (in preparation). Multi-level modelling of heritability: How transparent is the window on to cognitive process? *Manuscript in preparation*.

Thomas, M. S. C., Knowland, V. C. P., & Karmiloff-Smith, A. (2011). Mechanisms of developmental regression in autism and the broader phenotype: A neural network modeling approach. Psychological Review, 118(4), 637-654.

Thomas, M. S. C., Knowland, V. C. P., & Karmiloff-Smith, A. (2011). Mechanisms of developmental regression in autism and the broader phenotype: A neural network modeling approach. *Psychological Review*, *118(4)*, 637-654.

Thomas, M. S. C., Purser, H. R. M., Tomlinson, S., & Mareschal, D. (2012). Are imaging and lesioning convergent methods for assessing functional specialisation? Investigations using an artificial neural network. *Brain and Cognition*, *78*, 38-49.

Trzaskowski, M.,, Davis, O. S., DeFries, J. C., Yang, J., Visscher, P. M., & Plomin, R. (2013). DNA Evidence for Strong Genome-Wide Pleiotropy of Cognitive and Learning Abilities. *Behavior Genetics*, 43(4), 267-73.

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

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *The American Journal of Human Genetics*, *90*, 7-24.

Wallace, G. L., Lee, N. R., Prom-Wormley, E. C., Medland, S. E., Lenroot, R. K., Clasen, L. S., Schmitt, J. E., Neale, M. C., & Giedd, J. N. (2010). A bivariate twin

study of regional brain volumes and verbal and nonverbal intellectual skills during childhood and adolescence. *Behav Genet*. Mar;40(2):125-34. doi: 10.1007/s10519-009-9329-1. Epub 2010 Jan 30.

Woollams, A. M., Joanisse, M., & Patterson, K. (2009). Past-tense generation from form versus meaning: Behavioural data and simulation evidence. *Journal of Memory and Language*, *61*, 55-76.

Xie, X., & Seung, H. S. (2003). Equivalence of backpropagation and contrastive Hebbian learning in a layered network. *Neural Computation*, *15*, 441-454.

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E., & Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, *42(7)*, 565-9.

Tables

Table 1. Population mean and standard deviation for verb types at early (50 epochs), mid (100 epochs) and late (750 epochs) of training. Generalization was assessed by correct application of the past tense rule to novel verbs that rhymed with existing regulars in the training set.

		Early	Mid	Late
Regular		75.3 (23.8)	82.3 (19.7)	89.3 (13.9)
Irregular	Identity	45.3 (23.3)	57.5 (24.8)	74.6 (22.3)
	Vowel change	31.5 (24.0)	47.2 (28.0)	68.6 (26.4)
	Arbitrary	51.3 (31.6)	61.3 (29.7)	71.6 (23.8)
Generalization	Rule	59.3 (19.7)	63.0 (16.5)	65.9 (12.9)

Table 2. Correlations between structural indices (summed magnitude of connection weights, total number of connection weights) and functional indices (performance on regular verbs, performance on irregular verbs) for the simulated population, at early, mid, and late points of development.

Early				
		Magnitude	Number	Regular
Structure	Magnitude			
	Number	.623 **		
Function	Regular	.007	.086 **	
	Irregular	.106 **	.185 **	.640 **
Mid				
		Magnitude	Number	Regular
Structure	Magnitude			
	Number	.602 **		
Function	Regular	.036	.083 **	
	Irregular	.073 *	.120 **	.698 **
Late				
		Magnitude	Number	Regular
Structure	Magnitude			
	Number	.583 **		
Function	Regular	.149 **	.160 **	
	Irregular	.122 **	.199 **	.720 **

*. Correlation is significant at the 0.05 level (2-tailed)
**. Correlation is significant at the 0.01 level (2-tailed)

Table 3. Empirical data from Posthuma et al. (2003, Table 2) for structural indices of white matter volume and grey matter volume, and functional indices of performance on verbal comprehension and on working memory tests. Correlations within structural indices and within functional indices are shown in boxes.

		White matter volume	Grey matter volume	Verbal comprehension	Working memory
Structure	White matter volume				
	Grey matter volume	.59 **			
Function	Verbal comprehension	.01	.06		
	Working memory	.28 **	.27 **	.54 **	

*. Correlation is significant at the 0.05 level (2-tailed)
**. Correlation is significant at the 0.01 level (2-tailed)

Figure captions

Figure 1: The architecture of the target developmental system, identifying separate levels according to the causal modeling framework of Morton (2004).

Figure 2: The population distribution of performance on regular and irregular verbs at three points in training, early (50 epochs), mid (100 epochs), and late (750 epochs) in development.

Figure 3: Example of associations between levels of description for one neurocomputational parameter, the unit threshold function or 'temperature', for irregular verb behavior early in development. (a) The function linking behavior with the parameter value, with all other parameters held constant. (b) The association between behavior and parameter in the population, with uneven parameter frequencies and all parameters varying. (c) The look-up table used to derive the neurocomputational parameter from the artificial genome. (d) The association between behavior and the artificial genes, with the 10 alleles split into 5 genotypes.

Figure 4: Effect sizes of (linear) associations between neurocomputational parameter values and behavior, for regular verbs and irregular (vowel-change) verbs.

Figure 5: Effect sizes of artificial gene-behavior associations. Variation in population performance was predicted from individual binary allele values (0 or 1), for (a) regular

verbs and (b) irregular 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-neighbor threshold (NNT), pruning onset (PO), pruning probability (PP), pruning threshold (PT), weight decay (WD), sparseness of connectivity (SP).

Figure 6. Replicability of simulated association analyses. (a) Comparison of effect sizes for original population and for a population trained with the same artificial genomes but re-sampled environmental variation; (b) comparison of computational parameter effect sizes for those populations.

Figure 7: Replicability of simulated association analyses. (a) Comparison of effect sizes for original population and for two populations with re-sampled genomes (same allele frequency) and re-sampled environments; (b) comparison of computational parameter effect sizes.

Figure 8: Replicability of simulated association analyses. (a) Comparison of effect sizes for original population and for two populations with different allele frequencies. In the 70:30 population, the 1-valued allele had a frequency of 70% while the 0-valued allele had a frequency of 30%. In the 30:70 population, the 1-valued allele had a frequency of

30% while the 0-valued allele had a frequency of 70%. (b) Comparison of computational parameter effect sizes.

Figure 9: Associations between the artificial genome and the structural indices of total magnitude of network connection strengths, and total number of connection weights, for early in development. Associations for the functional index of irregular verb behavior are also included. (a) Artificial gene to structural / functional index; (b) computational parameter to structural / functional index.

Figure 10: A comparison of associations between the artificial genome and: (1) the structural index of total number of connection weights, (2) the mean network activation level in processing novel verbs, and (3) the behavioral performance on novel verbs (correct application of the past tense rule). Associations were computed for the early point of development. (a) Artificial gene to structural / activation / behavioral index; (b) computational parameter to structural / activation / behavioral index.

Figure 11: Associations when the population was split by (simulated) socio-economic status (SES). (a) Effect sizes for associations between artificial genome and behavior (irregular verb performance early in development); (b) effect sizes for associations between neurocomputational parameters and behavior.

Figure 12: Interactions between the effects of neurocomputational parameter values on behavior. Performance on (a) regular and (b) irregular verbs, for early (50 epochs) and late (750 epochs) in training, split by two Hidden Unit levels (40 or 50) and by two Learning Rate levels (0.125 or 0.075)

Figures

Figure 1

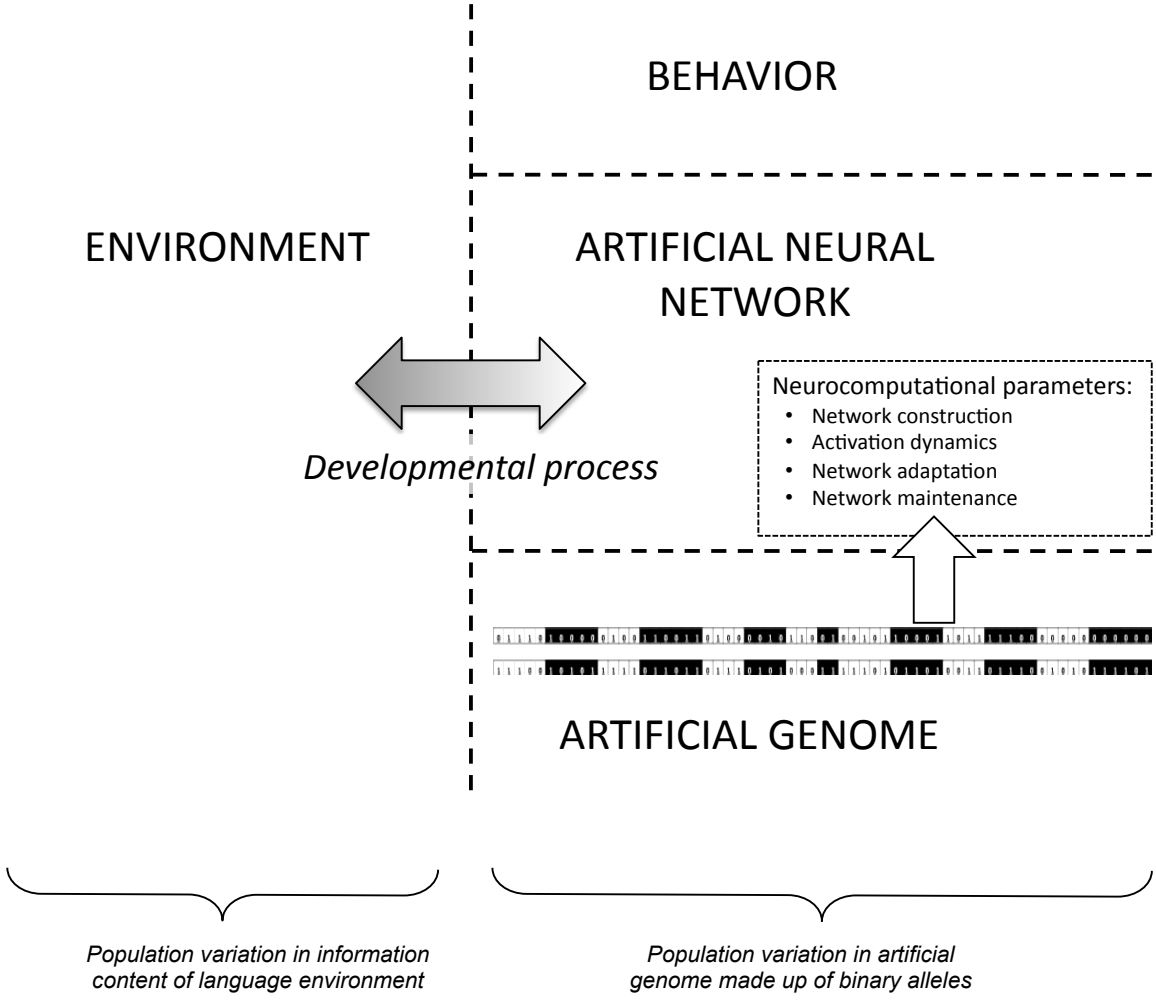
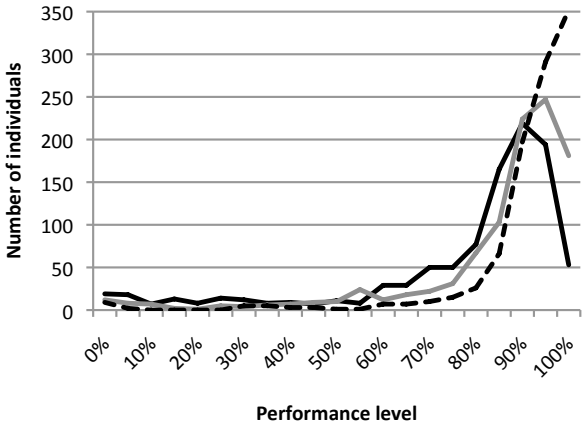


Figure 2

(a) Regular verbs



(b) Irregular verbs

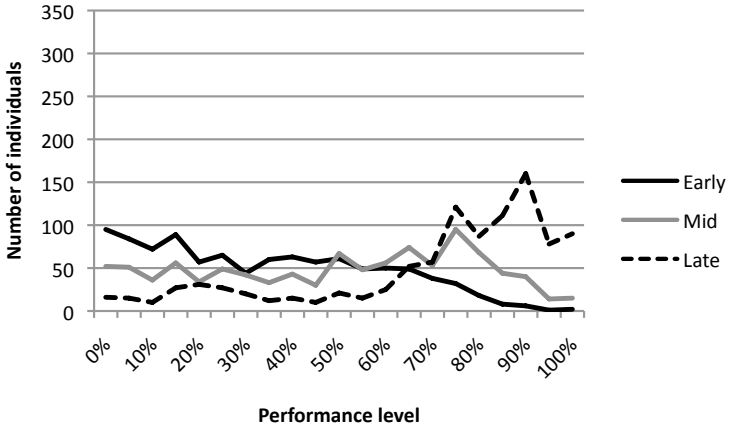
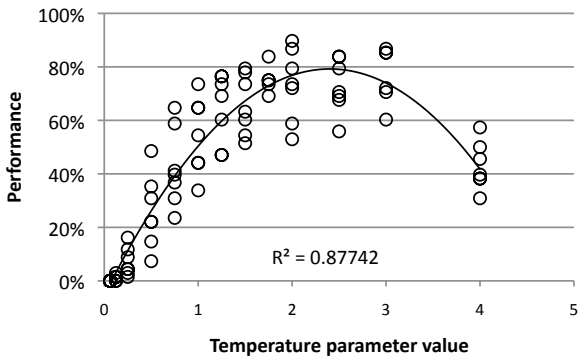


Figure 3

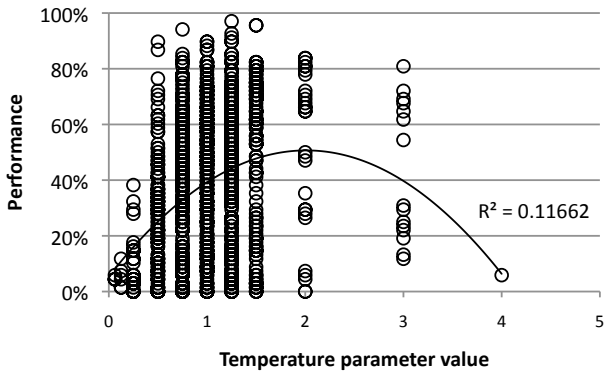
(a) Parameter-behavior function

(other parameters constant)



(b) Parameter-behavior association

(full population, all parameters varying)



(c) Look-up table for deriving the computational parameter from the artificial genes

	Temperature parameter value										
Number of 1-valued 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

(d) Simulated gene-behaviour associations

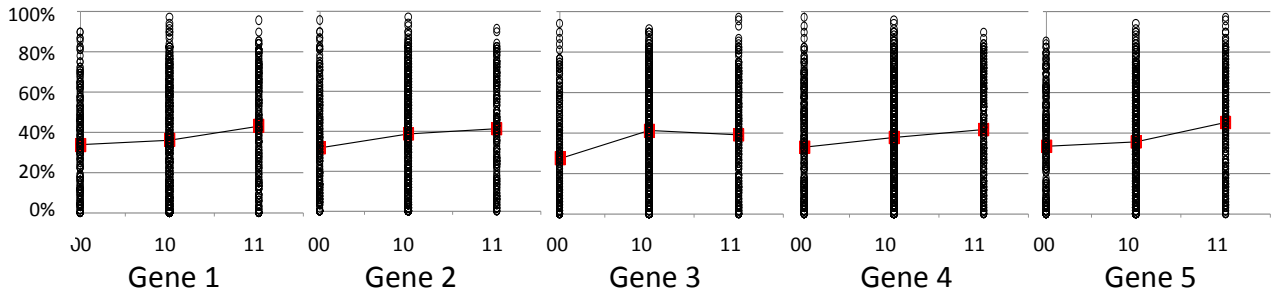
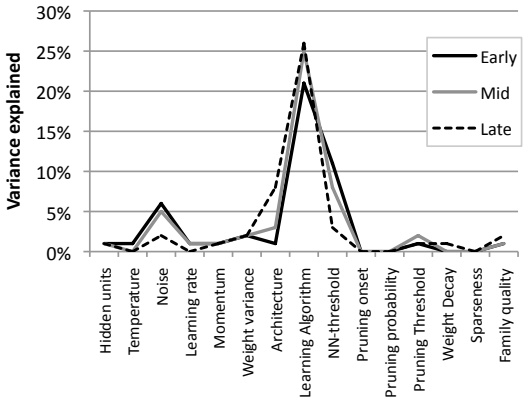


Figure 4

(a) Regular verbs



(b) Irregular verbs

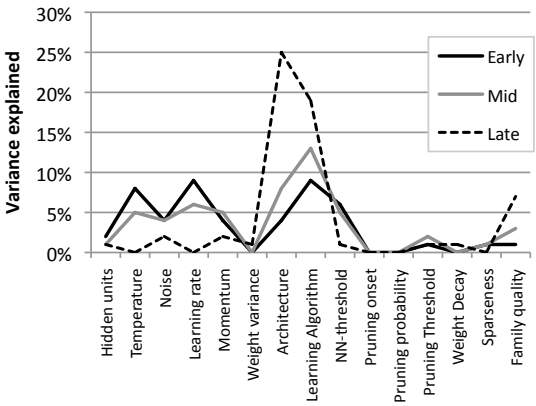
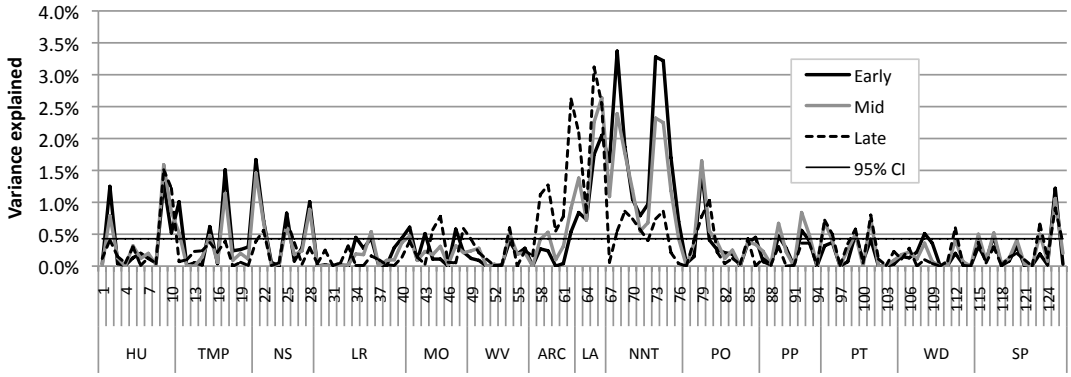


Figure 5

(a) Regular verbs



(b) Irregular vowel change verbs

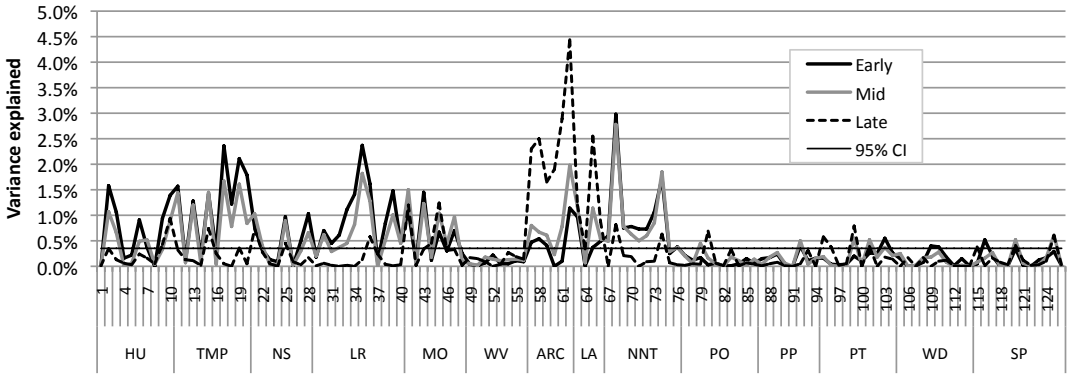
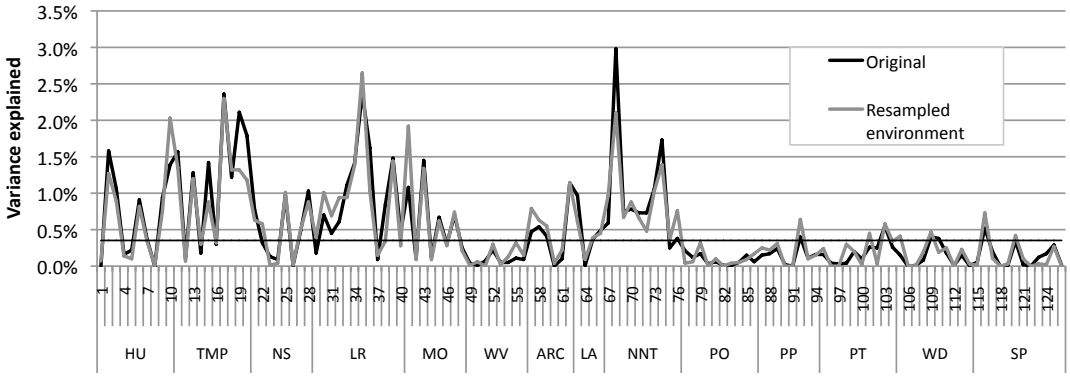


Figure 6

(a) Replication with re-sampled environment



(b) Equivalent parameter effect sizes

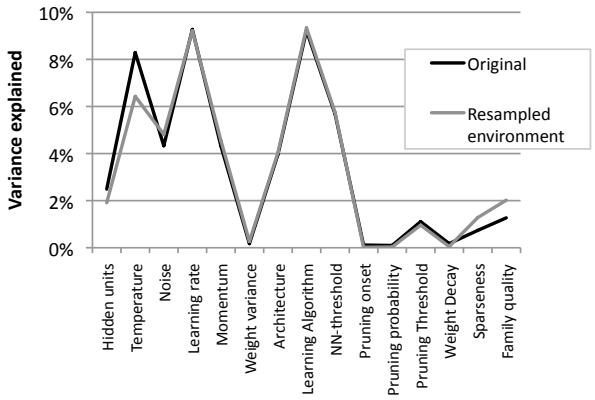
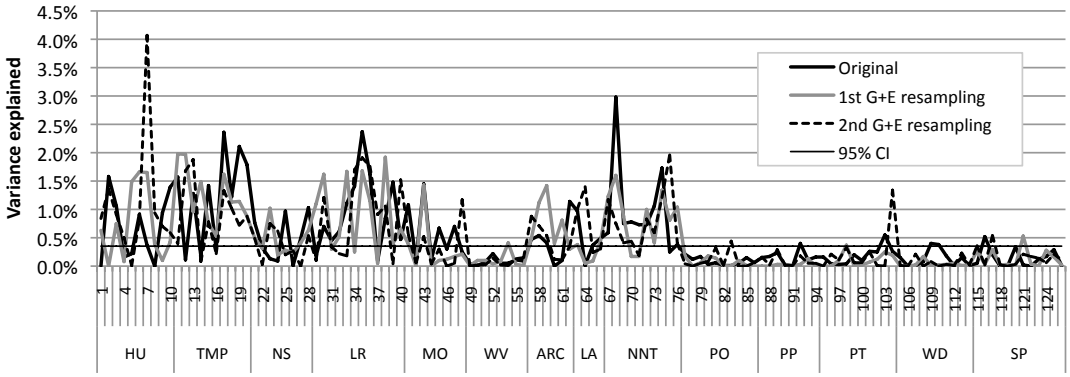


Figure 7

(a) Replication with re-sampled genomes and environment



(b) Equivalent parameter effect sizes

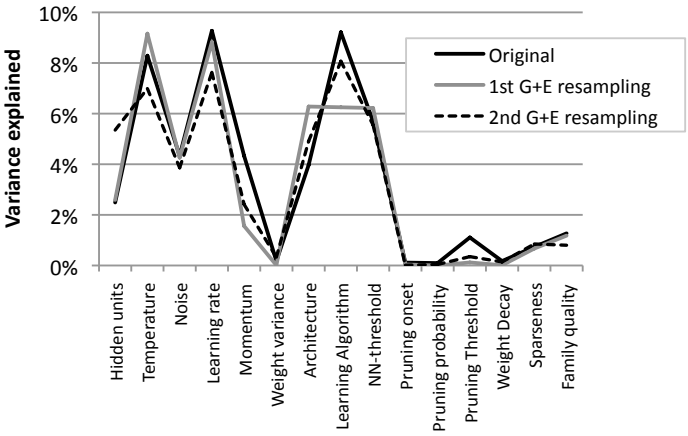
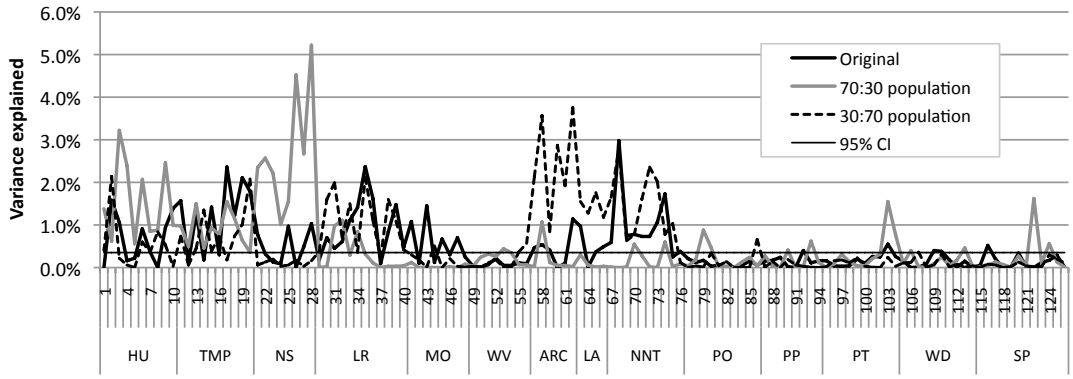


Figure 8

(a) Replication with populations with different allele frequencies



(b) Equivalent parameter effect sizes

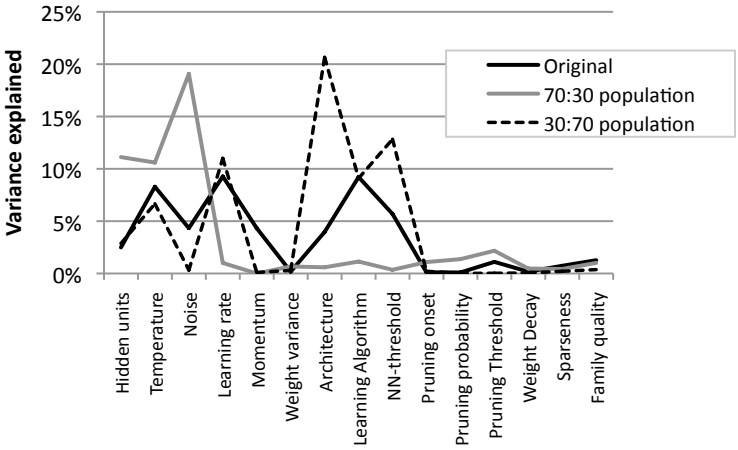
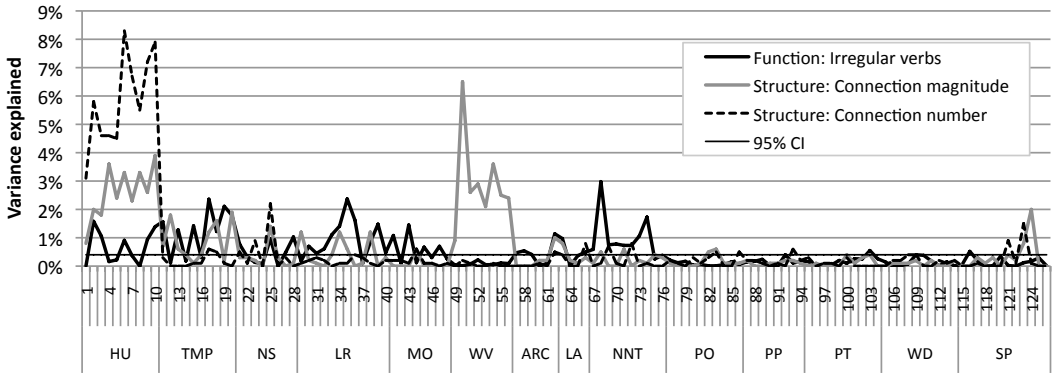


Figure 9

(a)


Variance explained
9%
8%
7%
6%
5%
4%
3%
2%
1%
0%
Function: Irregular verbs
Structure: Connection magnitude
Structure: Connection number
95% CI
HU
TMP
NS
LR
MO
WV
ARC
LA
NNT
PO
PP
PT
WD
SP

(b)

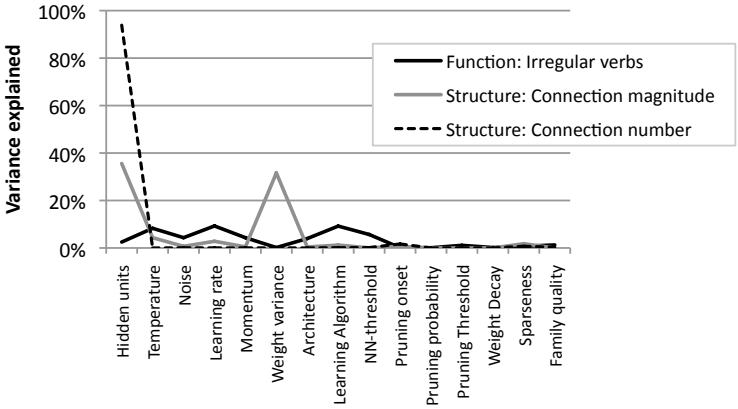
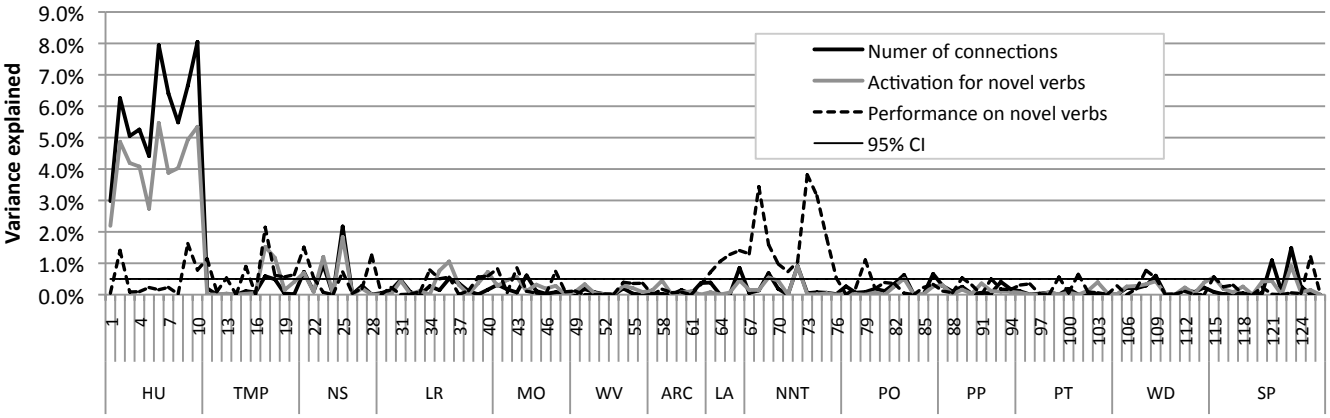

Variance explained
100%
80%
60%
40%
20%
0%
Function: Irregular verbs
Structure: Connection magnitude
Structure: Connection number
Hidden units
Temperature
Noise
Learning rate
Momentum
Weight variance
Architecture
Learning Algorithm
NN-threshold
Pruning onset
Pruning probability
Pruning Threshold
Weight Decay
Sparseness
Family quality

Figure 10

(a)


Variance explained
9.0%
8.0%
7.0%
6.0%
5.0%
4.0%
3.0%
2.0%
1.0%
0.0%
Numer of connections
Activation for novel verbs
Performance on novel verbs
95% CI
HU
TMP
NS
LR
MO
WV
ARC
LA
NNT
PO
PP
PT
WD
SP

(b)

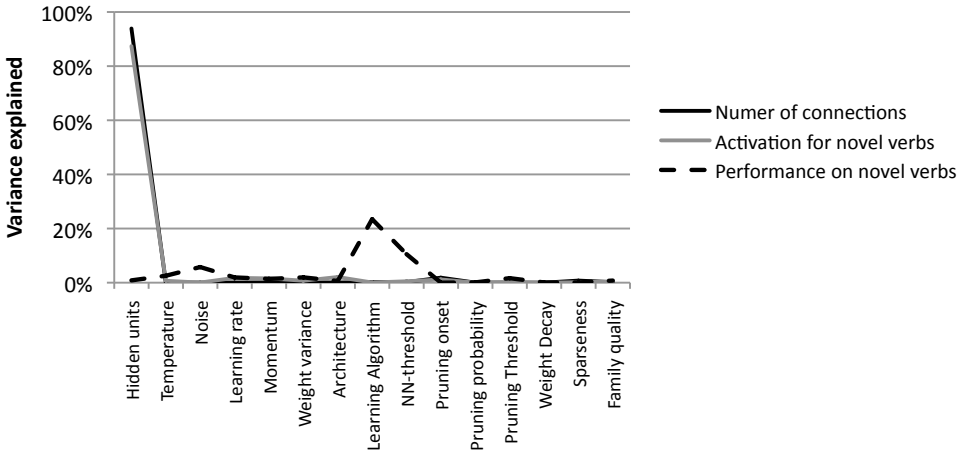
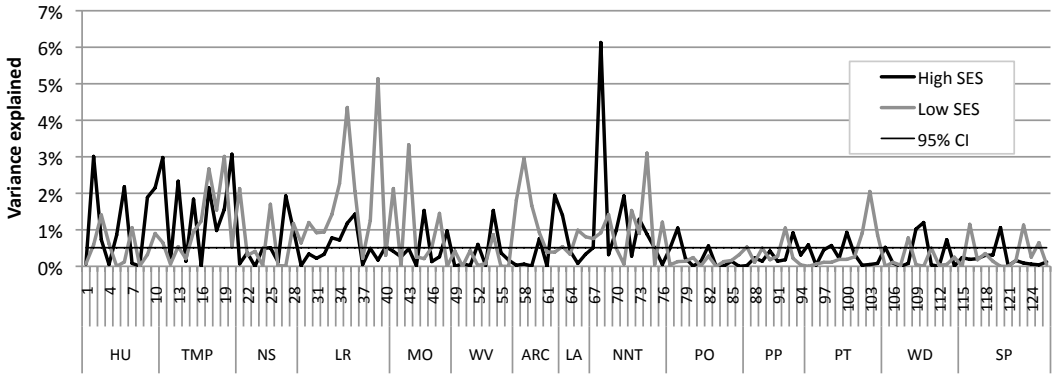

Variance explained
100%
80%
60%
40%
20%
0%
Numer of connections
Activation for novel verbs
Performance on novel verbs
Hidden units
Temperature
Noise
Learning rate
Momentum
Weight variance
Architecture
Learning Algorithm
NN-threshold
Pruning onset
Pruning probability
Pruning Threshold
Weight Decay
Sparseness
Family quality

Figure 11

(a)


Variance explained
7%
6%
5%
4%
3%
2%
1%
0%
High SES
Low SES
95% CI
HU
TMP
NS
LR
MO
WV
ARC
LA
NNT
PO
PP
PT
WD
SP

(b)

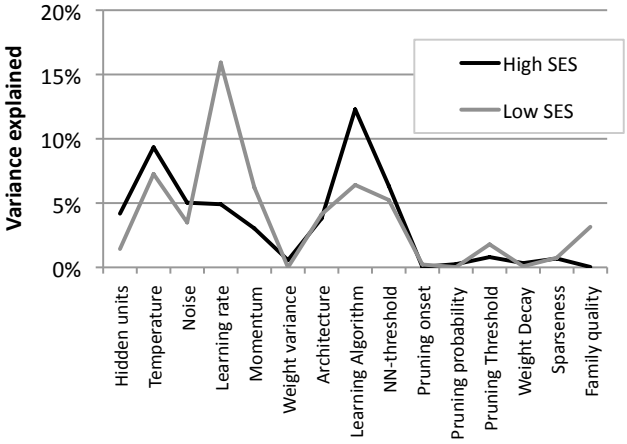
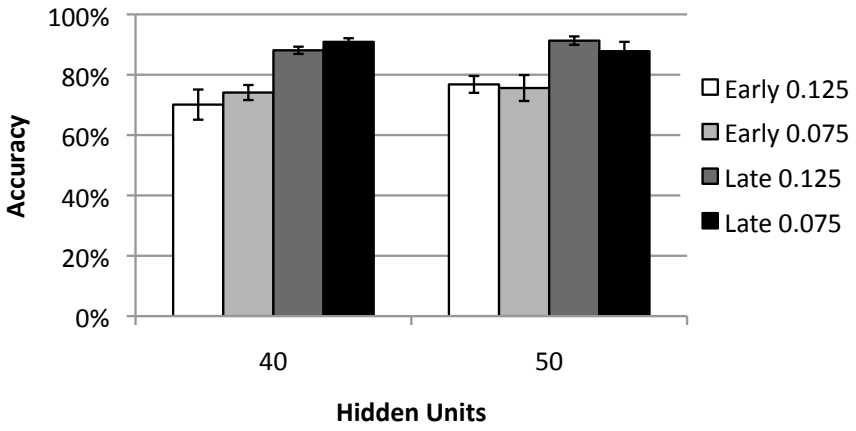
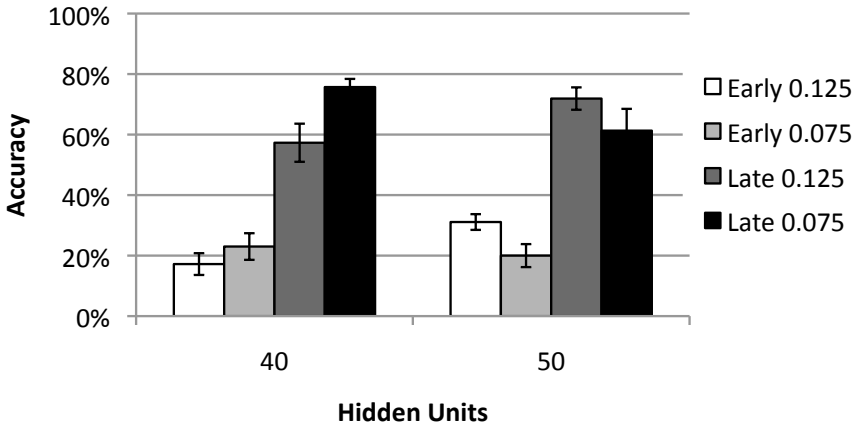

Variance explained
20%
15%
10%
5%
0%
High SES
Low SES
Hidden units
Temperature
Noise
Learning rate
Momentum
Weight variance
Architecture
Learning Algorithm
NN-threshold
Pruning onset
Pruning probability
Pruning Threshold
Weight Decay
Sparseness
Family quality

Figure 12

(a) Regular verbs



(b) Irregular verbs



Supplementary Material

To accompany the article 'Multi-scale modeling of gene-behavior associations in an artificial neural network model of cognitive development' by M. S. C. Thomas, N. A. Forrester, and A. Ronald

Introduction

This document contains technical details to accompany computer simulations that explore a population of networks acquiring the past tense domain. This domain is here used as a representative abstract learning problem within cognition. Individual variability is included both in the parameters of the artificial neural networks which model the children's learning systems, and the learning environment to which they are exposed. The parameters of the artificial neural networks are encoded in an artificial genome. Population variability in parameters is created by generating populations of artificial genomes. Each genome is realized as a parameterized network. The network is exposed to an individualized learning environment, generating a trajectory of behavioral development. The inclusion of an artificial genome level in the simulations allows us to study the associations that can arise between values on the artificial genome and behavioral variability that is the product of an implemented developmental process.

In the following, we describe the computational parameters that varied in the artificial neural networks. We outline how the range of variation for each parameter in the population was established. We then describe the method for designing the artificial genome, and the assumptions that this method embodies. Finally, a set of lookup tables is

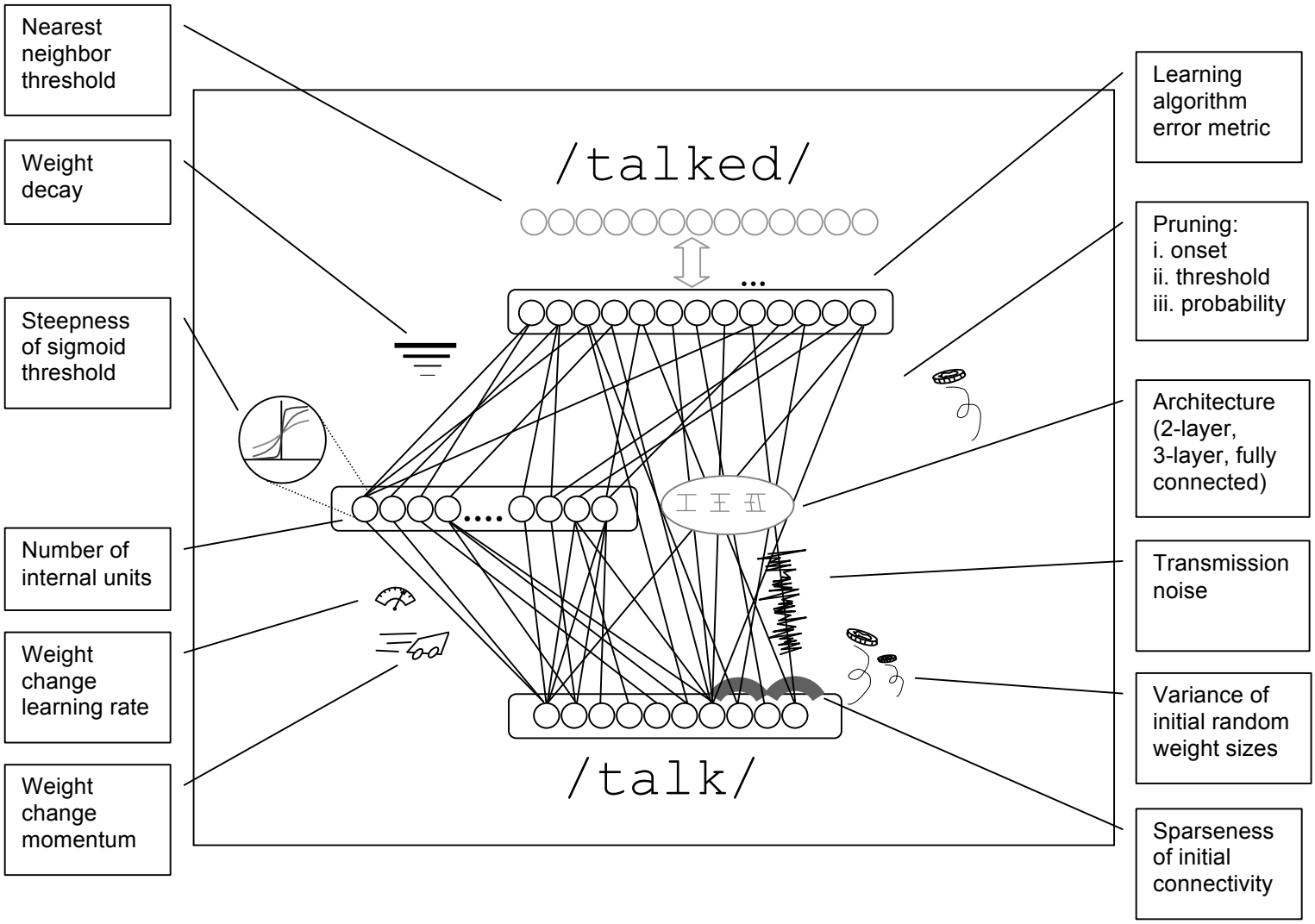
included detailing how values on the artificial genome were mapped to computational parameter values in the artificial neural networks.

The base past-tense model

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 to allow for additional analogues to known neurocomputational properties. However, backpropagation itself is not viewed as fully biologically plausible. We use it here in place of a more biologically plausible error-correction algorithm (see Thomas & McClelland, 2008, for discussion). An introduction to the idea that parameters in connectionist models can explain types of cognitive variability can be found in Thomas and Karmiloff-Smith, 2002a). The parameters and model architecture are depicted schematically in Figure 1.

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



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 than a 3-layer network but learns more quickly. A fully connected network contains both direct connections from input to output and a hidden layer, and produces a computationally more powerful system. Networks could therefore have 1, 2, or 3 layers of connection weights. Previous connectionist models have proposed single or multiple pathways may be available to connect input and output (e.g., Westermann, 1998; Zorzi, Houghton & Butterworth, 1998), and that differential use of routes may explain individual differences in behavior (Harm & Seidenberg, 2004; Plaut, 1997; Thomas & Karmiloff-Smith, 2002b). Recent functional brain imaging of reading lend support to this proposal (e.g., Richardson et al., 2011; Seghier et al., 2008).
- *Hidden units*: For networks with a hidden unit layer, the number of hidden units could vary. Variations of the number of hidden units have been proposed to account for developmental deficits such as dyslexia (e.g., Harm & Seidenberg, 1999) and autism (e.g., Cohen, 1998), as well as individual differences (Richardson et al., 2006a, b). We did not vary the number of hidden layers. More hidden units within a layer increases computational power and the rate of learning, while more layers of hidden units increases computational power but slows down learning, since error must be

propagated from the output more deeply into the network to improve learning (see Richardson et al., 2006a,b, for a comparison of these conditions).

- *Sparseness*: The architecture determined how many layers of connection weights existed. Of the potential connections in a layer, only a certain proportion was created. The sparseness parameter set the probability that any given connection would be created. Greater connectivity increases computational power, but can lead to slower learning. Under some conditions, it can also lead to poorer generalization, since greater integration of information causes more item-specific and context-specific learning (see McClelland, 2000, for a proposal that conjunctive coding may cause autistic symptoms; and conversely, Beversdorf, Narayanan & Hughes, 2007, for a proposal that the symptoms arise from sparse connectivity).
- *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 randomized between +/- 0.5. Large initial weights take time to unlearn, which slows learning (an effect known as entrenchment; see Munakata & McClelland, 2003, for discussion).

<u>Processing dynamics:</u>

- *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. Noise has been used to simulate under-specified representations in development (e.g., to simulate Specific Language Impairment: Joanisse & Seidenberg, 2003; or as a

candidate explanation of autistic symptoms: Simmons et al., 2007), and has also been proposed as an essential primitive in neural processing (McClelland, 1993).

- *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 non-linear activation function, the sigmoid or logistic function, equivalent to a smoothed threshold. This function has a free parameter, the 'temperature', which makes the smoothed threshold either steeper or shallower. The activation function was:

$$Output = \frac{1}{1 + e^{-temperature \times (netinput + bias)}}$$

where *netinput* is the summed activation to a unit, *bias* is the negative of the unit's threshold, and *Output* is the unit's activation state in response to this input. A shallow function (low temperature) denies a unit the opportunity to make large output changes in response to small changes in net input, whereas a steep function (high temperature) approximates a non-smoothed threshold, thereby producing a unit with binary response characteristics. Variations in the slope of the sigmoid function have been proposed as candidate explanations of disorders such as specific language impairment (Thomas, 2005) and schizophrenia (Cohen & Servan-Schreiber, 1992), as well as ageing (Li & Lindenberger, 1999). Changes to the slope of the sigmoid have a number of effects on learning. A shallow slope means that processing units are less sensitive to small differences in their input. This poor discriminability means they will be slow to learn categorizations that rely on small distinctions in the input. Secondly, in the backpropagation algorithm, weight update for a given error signal is proportional to the slope on the sigmoid (the differential of the function). If the function resembles a

gentle S-shape, then the slope across the range of unit activations will be small. A shallow sigmoid will lead directly to slower learning. Conversely, if the temperature is very high, producing a sigmoid similar to a step function, for most inputs to a unit, it will be jammed on or off ('saturated') rather than in its dynamic range. When a unit is saturated, the slope on the sigmoid function is flatter (the regions below or above the step). When it is in its dynamic range it is steep (the step). If a unit is predominantly saturated due to a high temperature, the flat slope will again lead to small weight changes for a given error signal and therefore slow learning. Finally, units with high temperatures flip between being saturated on or off. They are therefore ill suited to learning mappings requiring graduations of activation states. In sum, temperatures that are either too high or too low can delay learning.

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. The approximate range of weight decay values was derived by estimating a percentage of weight value that could plausibly be lost overall all of training (e.g., 50%), and then dividing this proportion by the number of training epochs (e.g., 1000) and the number of training patterns presented on each epoch (e.g., 508), to give a proportional reduction in the connection weights to be applied on each pattern presentation (e.g., $0.5/1000/508=9.84 \times 10^{-7}$). To our knowledge, weight decay has not been used as a candidate mechanism to explain individual variability.

- We did not simulate the increase in synaptic density observed in human cortex during infancy and early childhood, instead simulating the outcome of this process through variations in the sparseness of connectivity; we did, however, implemented the pruning of spare resources from mid-childhood (Huttenlocher, 2002). The *pruning process* eliminated small connection weights. Variations in pruning have been proposed as an explanation of autistic symptoms, and specifically developmental regression (Thomas, Knowland & Karmiloff-Smith, 2011). The pruning process 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, for simulations of pruning applied to sensitive periods in plasticity).
- *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 had 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). The Euclidean distance metric produces less weight change for a unit when it is committed to an erroneous response than the cross-entropy measure. That is, when a unit is stuck on in a saturated state but the learning algorithm requires it to be off, or vice versa, cross-entropy will lead to faster changes to its weights to change its activation state than Euclidean distance. Under some conditions, cross-entropy can therefore be a more plastic learning algorithm, leading to faster learning and higher ceiling performance.
- *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 good performance on different parts of the problem domain. Differences in learning rate have been proposed as explanations of individual differences in cognitive ability (Richardson et al., 2006a,b) and general intelligence (Garlick, 2002), as well as developmental deficits (e.g., dyslexia; Harm & Seidenberg, 1999).
- *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. While a parameter often

varied in connectionist models of development, it has not to my knowledge been used as a candidate explanation for individual differences in learning.

Network response:

- *Nearest neighbor 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 recognized 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 neighbor threshold. (The legal phonemes could of course still be the incorrect ones for the target verb). The nearest neighbor 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 neighbor 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 recognized as correct (i.e., larger error is tolerated); a low threshold requires a more exact initial output. The use of a nearest neighbor algorithm allowed the network to generate accuracy levels. Differences in the functioning of the attractor network (sometimes called 'clean-up' units) have been proposed as a candidate explanation of developmental deficits (e.g., dyslexia; Harm & & Seidenberg, 1999).

Calibrating parametric variation

Calibration was carried out to establish the full range of variation for each parameter over which the artificial neural network exhibited some degree of learning. In general, the network was fairly robust to variation in its parameters, as illustrated in Appendix A.

Two of the network 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, they were calibrated as follows. An initial 'normal' set of parameters was defined. These were estimated 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 had a monotonic relationship to performance (e.g., hidden units, where more was better); in other cases, there was an optimal intermediate value (e.g., activation function). The functions linking a given parameter and behavioral outcomes, with all other parameters held constant, are included in Appendix A. The aim was to determine an average or adequate value for each parameter, which was defined heuristically as 'just enough to succeed and then a little bit more'. Values were then derived 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 emphasize

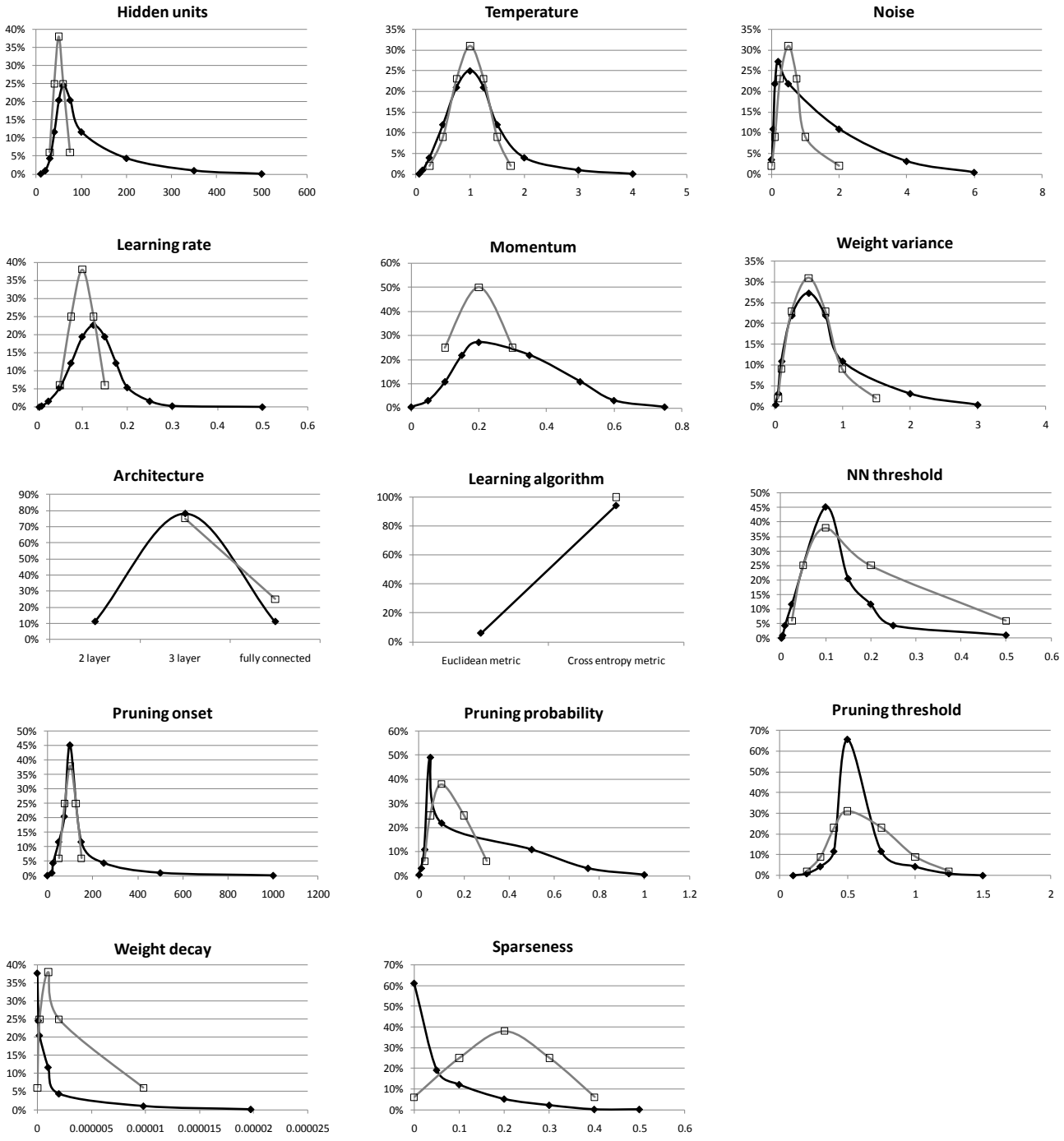
behavioral symmetry around the average parameter value rather than parametric symmetry, on the grounds that the symmetrical bell curve is a common pattern observed in human abilities. The ranges for each parameter for the phonology-to-phonology network are included in Figure 2.

We chose not to vary the input and output coding scheme. 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, 2003). 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 the representation of phonology cause subsequent behavioral 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 expected interactions between these neurocomputational parameters in subsequent learning. To pick four examples: (i) large numbers of hidden units can partially compensate for a shallow sigmoid function in those processing units; (ii) having a more sparse initial connectivity is likely to reduce the amount of weights eliminated via pruning because their magnitudes will be larger; (iii) high weight decay can be countered by a higher learning rate; (iv) an over-aggressive pruning process (e.g., with a high threshold and high probability) can be alleviated if its onset occurs very late in training when weights have become large, but exacerbated if the onset is early. Large numbers of parameter combinations were possible within our

scheme: given the number of levels specified for each parameter, approximately two trillion unique parameter combinations were available.

Figure 2: Parameter values and target population frequencies (dark lines).



In Figure 2, dark lines show parameter values (x-axis) and their target frequencies in the population (y-axis) 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: hidden units: 10; temperature: 10; noise: 8; learning rate: 12; momentum: 8; weight variance: 8; architecture: 6; learning algorithm: 4; nearest neighbor threshold: 10; pruning onset epoch: 10; pruning probability: 8; pruning threshold: 10; weight decay: 10; sparseness: 12 (total 126 bits). The grey lines show the functions for a condition in which parameter variation was narrower. We do not consider this condition further here.

Specifying an artificial genome for the model

The use of genetic algorithms entails creation of an artificial 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.
- The two alleles produced different outcomes in the functionality of the neurocomputational parameter which they encoded.
- The influence of genes was intended to be *additive*: we did not include dominant or recessive effects, and genes had the same effect in combination as in isolation.

This constraint was motivated by the finding within behavioral genetics that the effect of gene variants is predominantly additive on phenotypic outcomes (Plomin et al., 2008). Nevertheless, our method of implementing the mapping between gene variants and neurocomputational parameters did inadvertently produce some non-additive effects.

- All neurocomputational parameters were *polygenic*. That is, their value was determined by the additive action of a collection of genes.
- In the first instance, 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. This simplification likely will not hold in many cases, and certainly the current theoretical view is that the relationship between genes and *cognitive processes* is pleiotropic (see, e.g., Kovas & Plomin, 2006).

The assumption of polygenicity was motivated by the fact that we are using computational models to capture cognitive-level phenomena, and is a point worth emphasizing. 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.

We assumed, for reasons of simplification only, that the combination of alleles for each polygenic neurocomputational parameter had a deterministic relation to the value of that parameter in the instantiated network: that is, the allele set alone determined the

parameter value. Alternatively, this may be viewed as the assumption that the relation between alleles and parameter setting relied on an environment that did not vary across the individuals in the simulated population. We assumed (and did not instantiate) a much larger part of the genome that was species universal and was responsible for the basics of, for example, creating the processing units, the connections, the activation dynamics, the sensorium, the input-output connectivity pathways, and the mechanics of experience-dependent systems.

Parameter values and their link to the artificial genome for the past tense network

For the basic past tense network, the total of number of genes used to encode the value of the 14 computational parameters was 126 (or two copies of 63) as follows – hidden units: 10; temperature: 10; noise: 8; learning rate: 12; momentum: 8; weight variance: 8; architecture: 6; learning algorithm: 4; nearest neighbor threshold: 10; pruning onset epoch: 10; pruning probability: 8; pruning threshold: 10; weight decay: 10; sparseness: 12 (total 126 bits).

Figure 2 plots the range of values for each parameter against their target frequency of occurrence in the population. The translation of a genome into a parameter set was implemented by assigning alleles the value of 1 or 0, and then deriving the total for all the genes influencing the parameter. 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. The lookup tables for the 14 parameters (in the Wide Genetic used) condition used in the association simulations are shown below.

Table 1. Lookup table linking the artificial genome to the Hidden Unit parameter, for the Wide Genetic Variation condition

	Hidden Unit Parameter Value										
Number of 1-valued alleles	0	1	2	3	4	5	6	7	8	9	10
Population probability	0.001	0.01	0.04	0.12	0.21	0.25	0.21	0.12	0.04	0.01	0.001
Parameter value	10	20	30	40	50	60	75	100	200	350	500

Table 2. Lookup table linking the artificial genome to the Temperature parameter, for the Wide Genetic Variation condition

	Temperature Parameter Value										
Number of 1-valued alleles	0	1	2	3	4	5	6	7	8	9	10
Population probability	0.001	0.01	0.04	0.12	0.21	0.25	0.21	0.12	0.04	0.01	0.001
Parameter value	0.0625	0.125	0.25	0.5	0.75	1	1.25	1.5	2	3	4

Table 3. Lookup table linking the artificial genome to the Noise parameter, for the Wide Genetic Variation condition

	Noise Parameter Value								
Number of 1-valued alleles	0	1	2	3	4	5	6	7	8
Population probability	-	0.04	0.11	0.22	0.27	0.22	0.11	0.03	0.00
Parameter value	0	0	0.05	0.1	0.2	0.5	2	4	6

Table 4. Lookup table linking the artificial genome to the Learning Rate parameter, for the Wide Genetic Variation condition

	Learning Rate Parameter Value												
Number of 1-valued	0	1	2	3	4	5	6	7	8	9	10	11	12

alleles													
Population probability	0.0002	0.0029	0.02	0.05	0.12	0.19	0.23	0.19	0.12	0.05	0.02	0.0029	0.0002
Parameter value	0.005	0.01	0.025	0.05	0.075	0.1	0.125	0.15	0.175	0.2	0.25	0.3	0.5

Table 5. Lookup table linking the artificial genome to the Momentum parameter, for the Wide Genetic Variation condition

	Momentum Parameter Value								
Number of 1-valued alleles	0	1	2	3	4	5	6	7	8
Population probability	0.004	0.03	0.11	0.22	0.27	0.22	0.11	0.03	0.004
Parameter value	0	0.05	0.1	0.15	0.2	0.35	0.5	0.6	0.75

Table 6. Lookup table linking the artificial genome to the Weight Variation parameter, for the Wide Genetic Variation condition

	Weight Variation Parameter Value								
Number of 1-valued alleles	0	1	2	3	4	5	6	7	8
Population probability	0.004	0.03	0.11	0.22	0.27	0.22	0.11	0.03	0.004
Parameter value	0.01	0.05	0.1	0.25	0.5	0.75	1	2	3

Table 7. Lookup table linking the artificial genome to the Architecture parameter, for the Wide Genetic Variation condition. (0 = 2-layer, 1 = 3-layer, 2 = fully-connected)

	Architecture Parameter Value						
Number of 1-valued alleles	0	1	2	3	4	5	6
Population probability	-	0.109	-	0.781	-	0.109	-
Parameter value	0	0	1	1	1	2	2

Table 8. Lookup table linking the artificial genome to the Learning Algorithm parameter, for the Wide Genetic Variation condition. (0 = Euclidean distance error metric, 1 = cross-entropy error metric)

	Learning Algorithm Parameter Value				
Number of 1-valued alleles	0	1	2	3	4
Population probability	0.063	0.938	-	-	-
Parameter value	0	1	1	1	1

Table 9. Lookup table linking the artificial genome to the Nearest Neighbor Threshold parameter, for the Wide Genetic Variation condition

	Nearest Neighbor Threshold Parameter Value										
Number of 1-valued alleles	0	1	2	3	4	5	6	7	8	9	10
Population probability	0.001	0.010	0.044	0.117	0.451	-	0.205	0.117	0.044	0.011	-
Parameter value	0.0025	0.005	0.01	0.025	0.1	0.1	0.15	0.2	0.25	0.5	0.5

Table 10. Lookup table linking the artificial genome to the Pruning Onset parameter, for the Wide Genetic Variation condition

	Pruning Onset Parameter Value										
Number of 1-valued alleles	0	1	2	3	4	5	6	7	8	9	10
Population probability	0.001	0.01	0.04	0.12	-	0.45	0.21	0.12	0.04	0.01	0.001
Parameter value	1000	500	250	150	100	100	75	50	25	20	0

Table 11. Lookup table linking the artificial genome to the Pruning Probability parameter, for the Wide Genetic Variation condition

	Pruning Probability Parameter Value								
Number of 1-	0	1	2	3	4	5	6	7	8

valued alleles									
Population probability	0.004	0.03	0.11	-	0.49	0.22	0.11	0.03	0.004
Parameter value	0	0.01	0.025	0.05	0.05	0.1	0.5	0.75	1

Table 12. Lookup table linking the artificial genome to the Pruning Threshold parameter, for the Wide Genetic Variation condition

	Pruning Threshold Parameter Value										
Number of 1-valued alleles	0	1	2	3	4	5	6	7	8	9	10
Population probability (%)	0.001	0.01	0.04	0.12	-	0.66	-	0.12	0.04	0.01	0.001
Parameter value	0.1	0.2	0.3	0.4	0.5	0.5	0.5	0.75	1	1.25	1.5

Table 13. Lookup table linking the artificial genome to the Weight Decay parameter, for the Wide Genetic Variation condition

	Weight Decay Parameter Value										
Number of 1-valued alleles	0	1	2	3	4	5	6	7	8	9	10
Population probability	-	-	-	-	0.38	0.25	0.21	0.12	0.04	0.01	0.001
Parameter value	0	0	0	0	0	$1x10^{-7}$	$2x10^{-7}$	$9.8x10^{-7}$	$19.7x10^{-7}$	$98.4x10^{-7}$	$196.9x10^{-7}$

Table 14. Lookup table linking the artificial genome to the Sparseness parameter, for the Wide Genetic Variation condition

	Sparseness Parameter Value												
Number of 1-valued alleles	0	1	2	3	4	5	6	7	8	9	10	11	12
Population probability	-	-	-	-	-	-	0.61	0.19	0.12	0.05	0.02	0.003	0.0002
Parameter value	0	0	0	0	0	0	0	0.05	0.1	0.2	0.3	0.4	0.5

References

Beversdorf, D. Q., Narayanan, A., Hillier, A., & Hughes, J. D. (2007). Network model of decreased context utilization in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *37*, 1040– 1048.

Cohen, I. L. (1998). Neural network analysis of learning in autism. In D. J. Stein & J. Ludik (Eds.), *Neural networks and psychopathology* (pp. 274–315). New York, NY: Cambridge University Press.

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

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*, 662-720.

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

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.

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

Joanisse, M. F., & Seidenberg, M. S. (1999). Impairments in verb morphology after brain injury: A connectionist model. *Proceedings of the National Academy of Sciences of the United States of America*, *96*, 7592-7597.

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., Seidenberg, M. S. (2003). Phonology and syntax in Specific Language Impairments: Evidence from a connectionist model. *Brain and Language*, *86*, 40-56.

Karaminis, T. N., & Thomas, M. S. C. (2010). A cross-linguistic model of the acquisition of inflectional morphology in English and Modern Greek. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of 32nd Annual Conference of the Cognitive Science Society*, August 11-14, 2010, Portland, Oregon, USA.

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

Li, S.-C. & Lindenberger, U. (1999). Cross-level unification: A computational exploration of the link between deterioration of neurotransmitter systems and the dedifferentiation of cognitive abilities in old age, (pp. 103-146) in L.-G. Nilsson & H. Markowitsch (Eds.). *Cognitive neuroscience of memory*. Toronto: Hogrefe & Huber

McClelland, J. L. (1993) Toward a theory of information processing in graded, random, interactive networks. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence and cognitive neuroscience* (pp. 655-688). Cambridge, MA: MIT Press.

McClelland, J. L. (2000). The basis of hyperspecificity in autism: A preliminary suggestion based on properties of neural nets. *Journal of Autism and Developmental Disorders*, *30*, 497–502

Munakata, Y., & McClelland, J. L. (2003). Connectionist models of development. *Developmental Science*, *6(4)*, 413-429.

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes*, *12*, 767-808.

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

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.

Richardson, F. M., Forrester, N. A., Baughman, F. D., & Thomas, M. S. C. (2006b). Computational modeling of variability in the conservation Task. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (p. 2010-2015), July 26-29, Vancouver, BC, Canada.

Richardson, F. M., Seghier, M. L., Leff, A. P., Thomas, M. S. C., & Price, C. J. (2011). Multiple routes from occipital to temporal cortices during reading. *Journal of Neuroscience*, *31(22)*, 8239-8247.

Richardson, F.M., Baughman, F.D., Forrester, N. A., & Thomas, M.S.C. (2006a). Computational modeling of variability in the balance scale task. *Proceedings of the 7th International Conference of Cognitive Modeling*.

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.

Seghier, M. L., Lee, H. L., Schofield, T., Ellis, C. L., & Price, C. (2008). Inter-subject variability in the use of two different neuronal networks for reading aloud familiar words. *Neuroimage*, *42(3-3)*, 1226-1236.

Simmons, D. R., McKay, L., McAleer, P., Toal, E., Robertson, A., & Pollick, F. E. (2007). Neural noise and autism spectrum disorders. *Perception*, *36(Suppl.)*, 119–120.

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., Forrester, N. A., & Ronald, A. (2013, April 1). Modeling socio-economic status effects on language development. *Developmental Psychology*. Advance online publication. doi: 10.1037/a0032301.

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

Thomas, M. S. C. & Knowland, V. C. P. (2014). Modeling mechanisms of persisting and resolving delay in language development. *Journal of Speech, Language, and Hearing Research*.

Thomas, M. S. C., Knowland, V. C. P., & Karmiloff-Smith, A. (2011). Mechanisms of developmental regression in autism and the broader phenotype: A neural network modeling approach. *Psychological Review*, *118(4)*, 637-654.

Thomas, M. S. C. & Karmiloff-Smith, A. (2002b). Are developmental disorders like cases of adult brain damage? Implications from connectionist modeling. *Behavioral and Brain Sciences*, *25(6)*, 727-788.

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

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

Westermann, G (1998) Emergent modularity and U-shaped learning in a constructivist neural network learning the English past tense. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 1130-1135). Hillsdale, NJ: Erlbaum.

Woollams, A. M., Joanisse, M., & Patterson, K. (2009). Past-tense generation from form versus meaning: Behavioral data and simulation evidence. *Journal of Memory and Language*, *61*, 55-76.

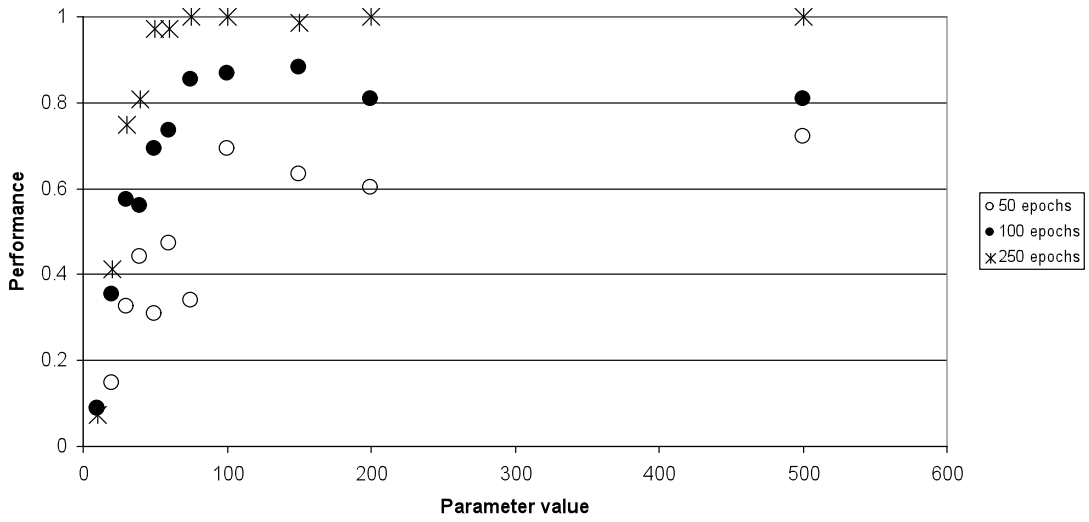
Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1131-1161.

Appendix A

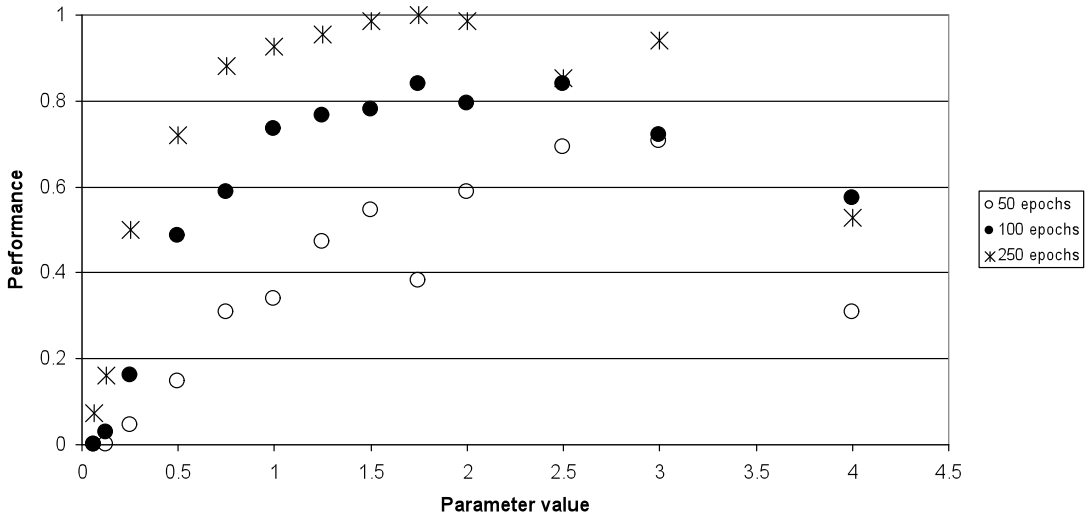
Parameter-behavior functions for the basic past tense network

All other parameters were held at default values, while a single parameter was varied. Performance is reported for regular verbs on the past tense task. The shape of the function is displayed for three points in training, 50, 100, and 250 epochs. The default parameter values were: hidden units: 50; temperature: 1; noise: 0; learning rate: 0.01; momentum: 0.2; weight variance: 0.5; architecture: 3-layer; learning algorithm: back propagation error measure; nearest neighbor threshold: 0.1; pruning onset epoch: 50; pruning probability: 0.1; pruning threshold: 0.5; weight decay: 0.000019; sparseness: 90% connectivity.

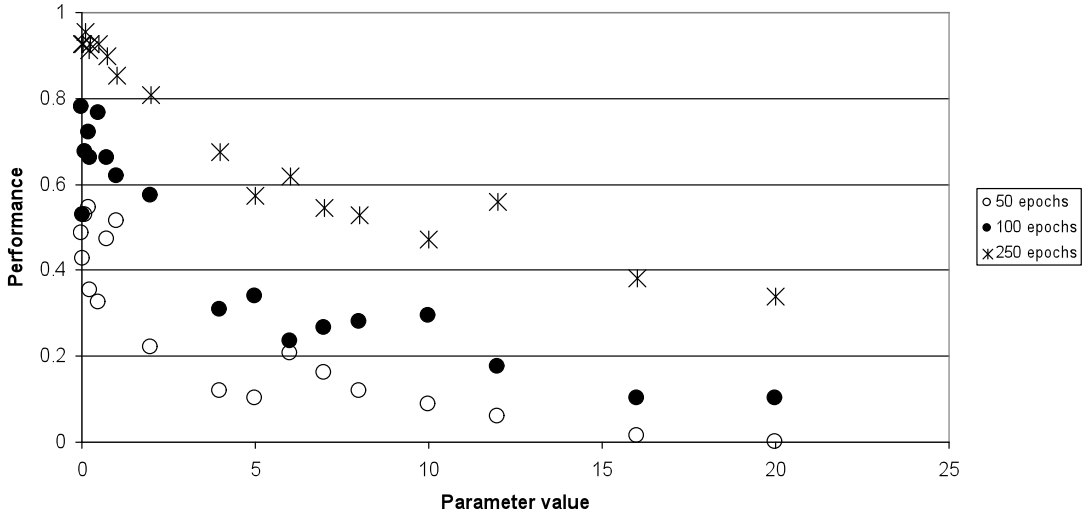
Hidden units:



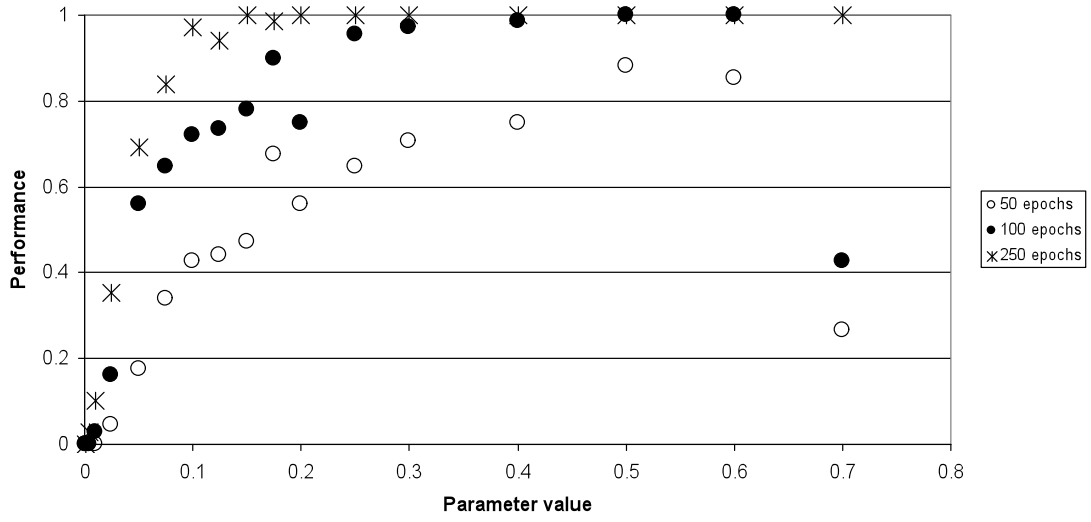
Unit threshold function (temperature)



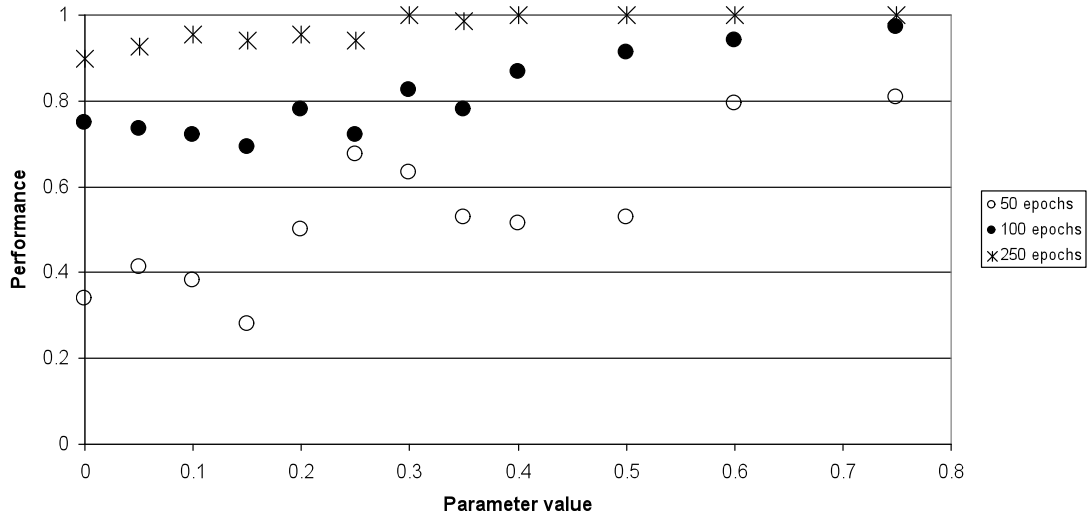
Processing noise:



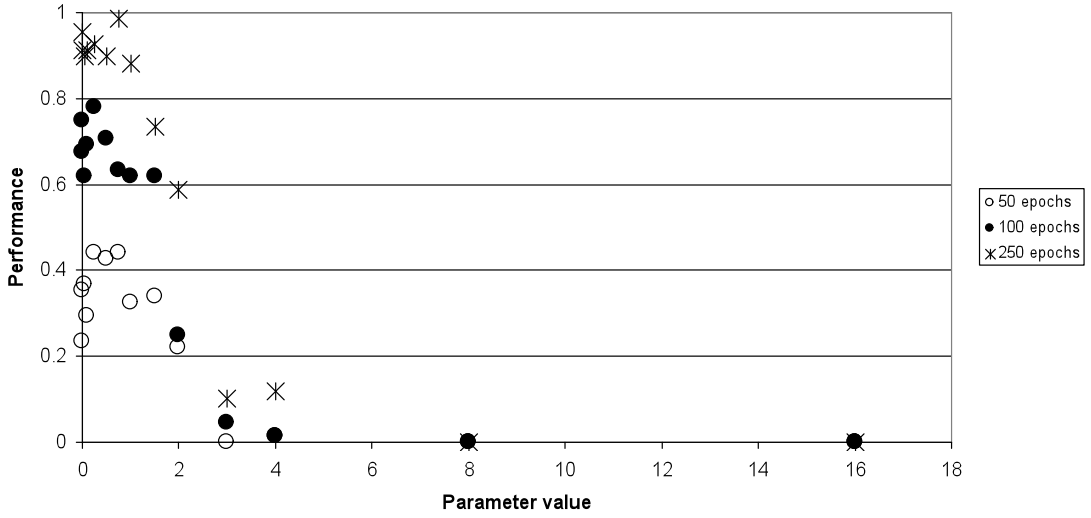
Learning rate:



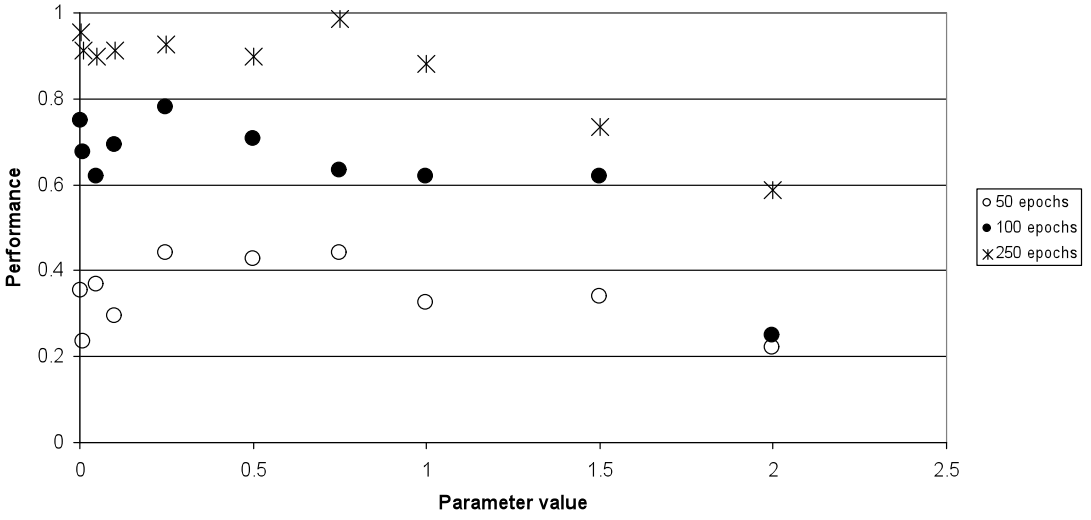
Momentum:



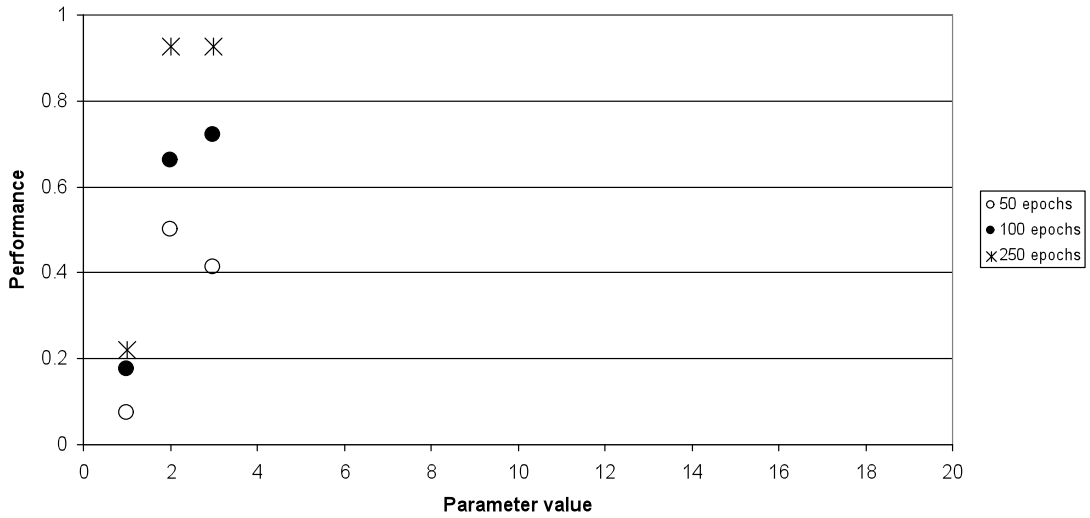
Initial weight variance???



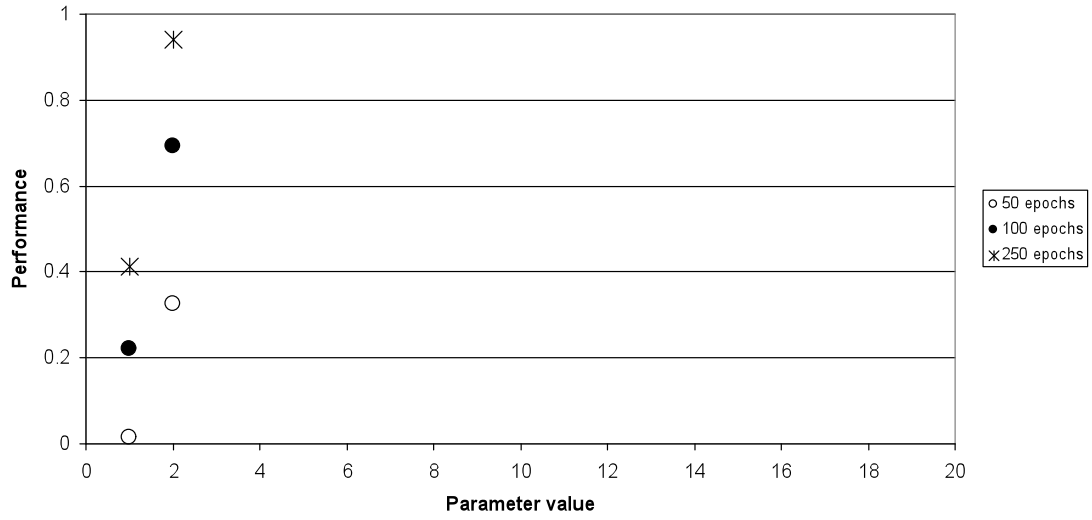
Initial weight variance????



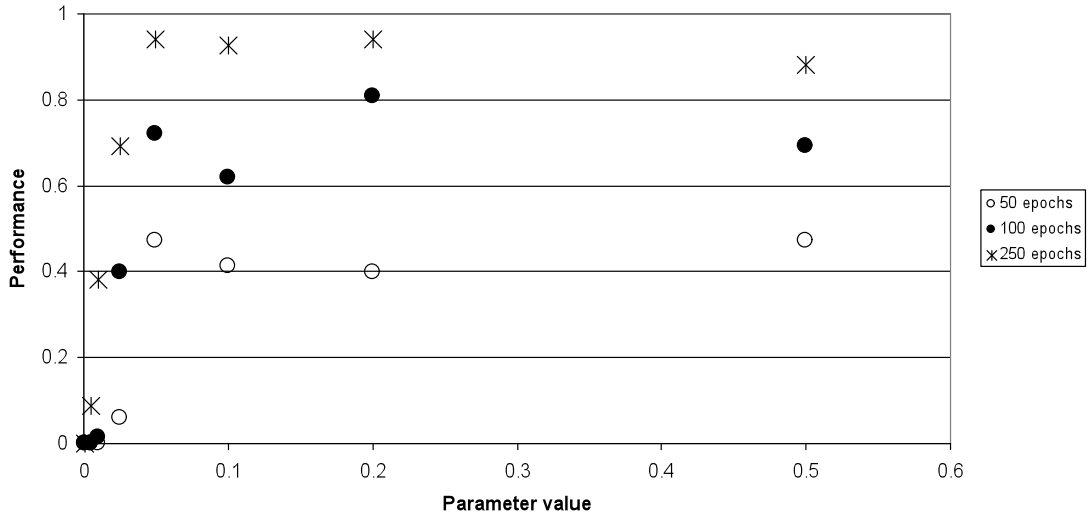
Architecture (0=2-layer; 1=3-layer; 2=fully connected):



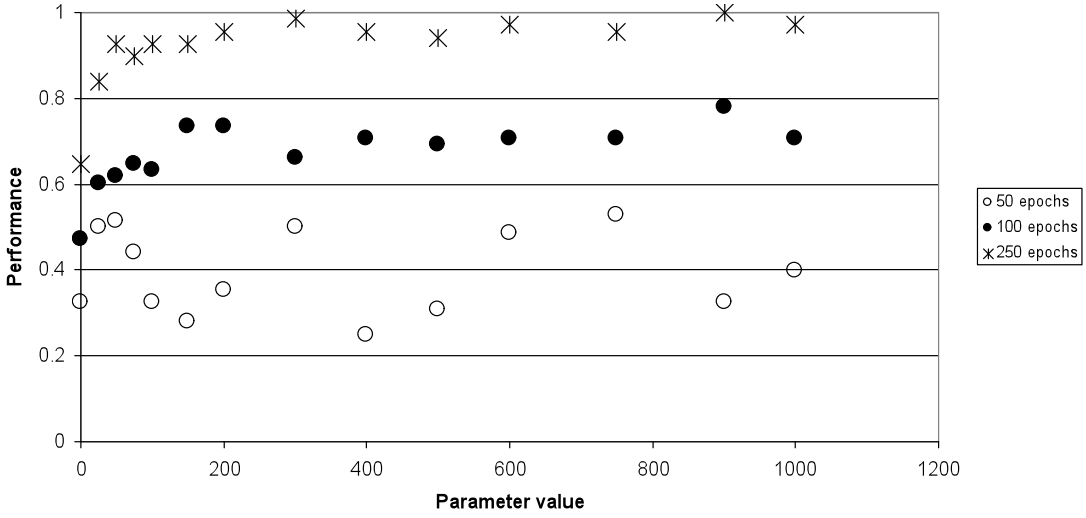
Learning algorithm error metric (0=Euclidean distance; 1=Cross-entropy):



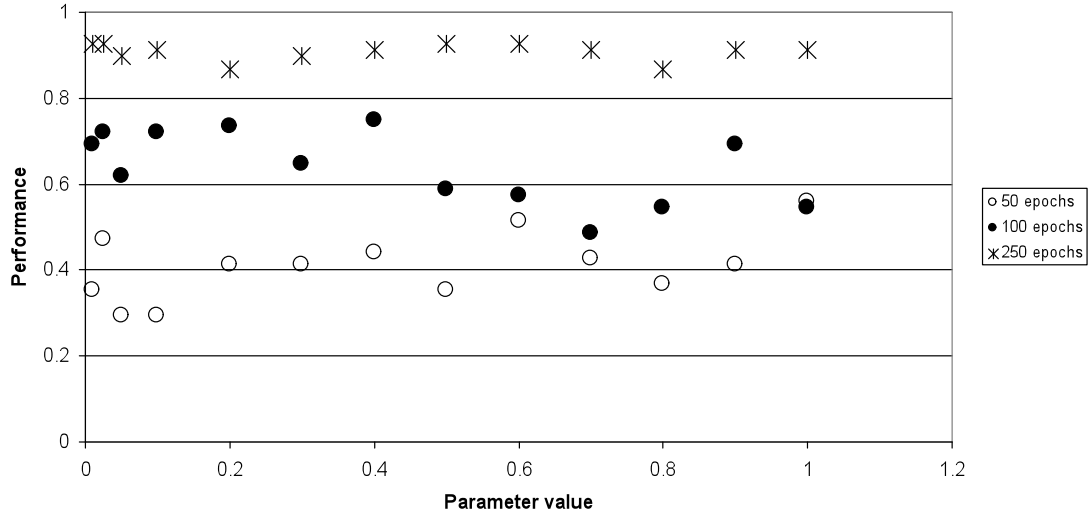
Nearest neighbor response threshold:



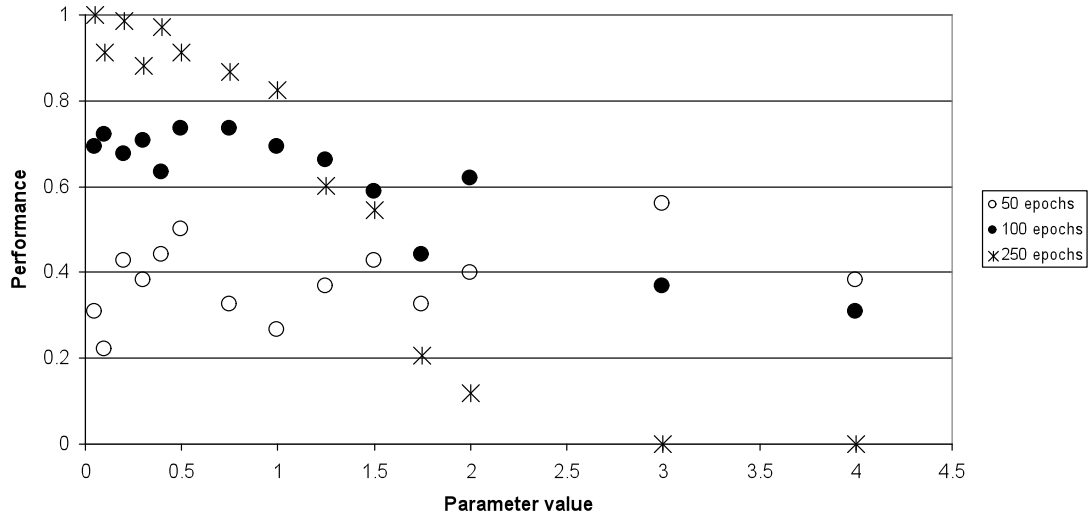
Pruning onset:



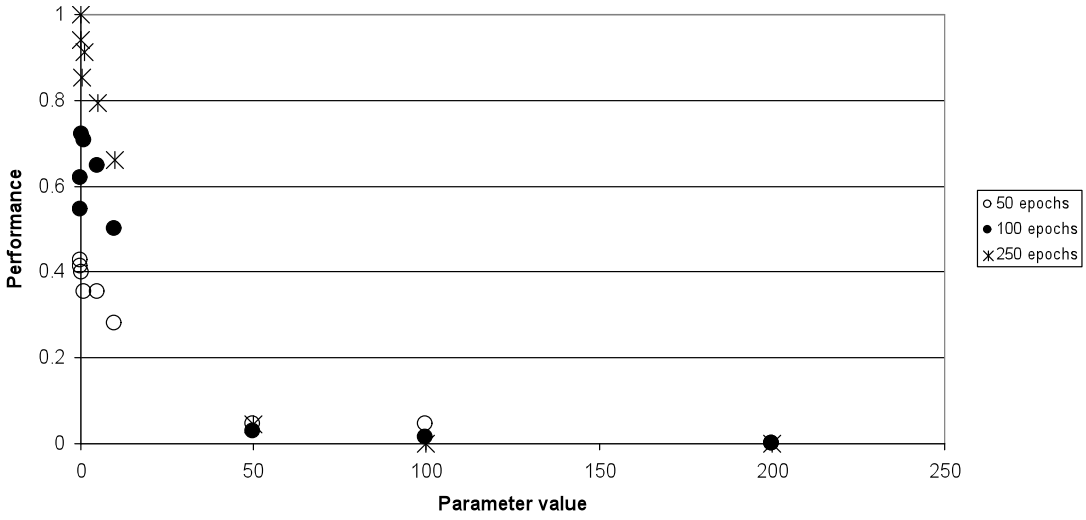
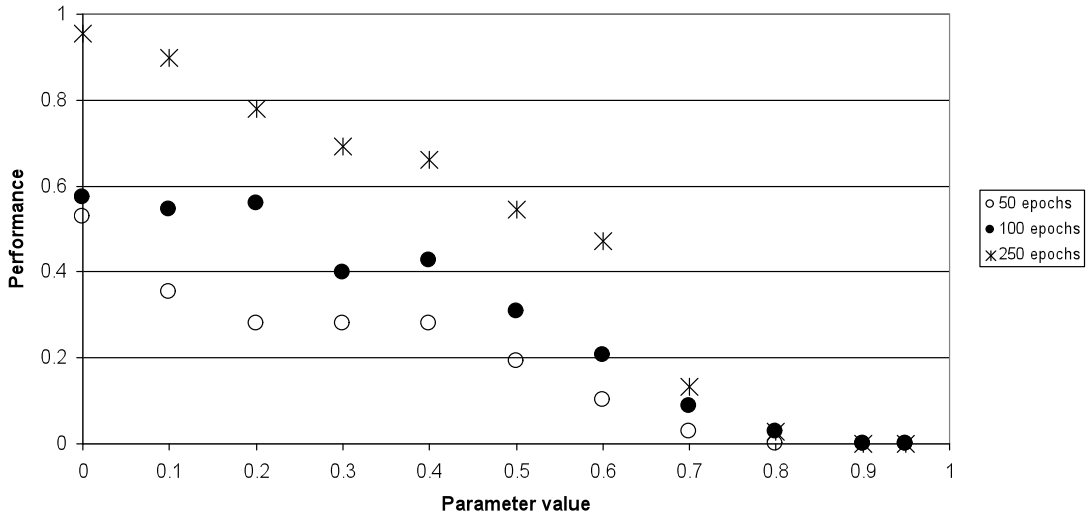
Pruning probability:



Pruning threshold:



Weight decay:



Sparseness of initial connectivity (proportion removed):