Complimentary Author PDF: Not for Broad Dissemination

JSLHR

Research Article

Modeling Mechanisms of Persisting and Resolving Delay in Language Development

Michael S. C. Thomas[a] and V. C. P. Knowland[a]

Purpose: In this study, the authors used neural network modeling to investigate the possible mechanistic basis of developmental language delay and to test the viability of the hypothesis that persisting delay and resolving delay lie on a mechanistic continuum with normal development.
Method: The authors used a population modeling approach to study individual rates of development in 1,000 simulated individuals acquiring a notional language domain (in this study, represented by English past tense). Variation was caused by differences in internal neurocomputational learning parameters as well as the richness of the language environment. An early language delay group was diagnosed, and individual trajectories were then traced.
Results: Quantitative variations in learning mechanisms were sufficient to produce persisting delay and resolving delay subgroups in similar proportions to empirical observations. In the model, persisting language delay was caused by limitations in processing capacity, whereas resolving delay was caused by low plasticity. Richness of the language environment did not predict the emergence of persisting delay but did predict the final ability levels of individuals with resolving delay.
Conclusion: Mechanistically, it is viable that persisting delay and resolving delay are only quantitatively different. There may be an interaction between environmental factors and outcome groups, with individuals who have resolving delay being influenced more by the richness of the language environment.

Key Words: language delay, socioeconomic status, artificial neural networks, population modeling, computational capacity, plasticity

Around 15% of children show delayed language development between the ages of 3 and 4 years (Broomfield & Dodd, 2004). However, in a significant proportion of cases, perhaps as many as two-thirds (Rannard, Lyons, & Glenn, 2005), the delay resolves so that children subsequently fall within the normal range on standardized measures. Other children showing early language delay go on to exhibit persisting deficits that have serious and long-term consequences for their education (Justice, Bowles, Pence-Tumbell & Skibbe, 2009; Young et al., 2002), socialization (see Durkin & Conti-Ramsden, 2010), mental health (Arkkila, Rasanen, Roine, & Vilkman, 2008; Beitchman et al., 2001), and employability (Law, Rush, Schoon, & Parsons, 2009). Little is understood about the developmental mechanisms that differentiate persisting language delay from resolving language delay, with arguments made for the importance of both genetic and environmental factors. In this article, we use computational modeling methods to further a mechanistic understanding of the causes of persisting delay versus resolving delay.

Dale, Price, Bishop, and Plomin (2003) identified a sample of 802 two-year-old children who were at risk for language delay on the basis of parental reports of vocabulary, grammar, nonverbal ability, and the children's use of language to refer to past and future events (also known as *displaced reference*). The authors followed up when these children were 3 and 4 years of age, again using parental measures. At 3 years of age, only 44.1% of these children met the criteria for persisting language difficulties, and at 4 years of age, the proportion fell to 40.2%. Thus, in more than half the cases, the early indicators of language delay had resolved. This pattern has been observed in a number of studies (Bishop & Edmundson, 1987; Paul, 1996; Rescorla, Dahlsgaard, & Roberts, 2000; Whitehurst & Fischel, 1994). For example, Rescorla and colleagues (2000) examined the mean length of utterance of late talkers at 3 years of age and 4 years of age and found that although 41% of the children scored above the 10th percentile at age 3, this figure had risen to 71% by age 4. Rescorla and colleagues found that the resolution of delay continued at slightly older ages. Bishop (2005) assessed 264 children who were identified as being at risk for language impairment at age 4; only one third

[a]Developmental Neurocognition Lab, Birkbeck, University of London
Correspondence to: Michael S. C. Thomas: m.thomas@bbk.ac.uk

Editor: Janna Oetting
Associate Editor: Pamela Hadley
Received August 13, 2012
Revision received February 7, 2013
Accepted June 25, 2013
DOI: 10.1044/2013_JSLHR-L-12–0254

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

Complimentary Author PDF: Not for Broad Dissemination

of these children met psychometric criteria for specific language impairment (SLI) at age 6. By age 7, however, greater stability is apparent in children's developmental trajectories: Conti-Ramsden, St. Clair, Pickles, and Durkin (2012) reported that the language growth trajectories of 242 children with SLI who were followed longitudinally from ages 7 through 17 remained predominantly parallel to—and below—those of children in the normal range. It is notable that the extent to which children can be seen to recover across development depends on the measure that researchers used to assess language acquisition, not only in terms of which levels of language are assessed but also whether parental report or standardized assessment is adopted. Although rate of change may vary depending on the assessment tools, the pattern of findings remains robust across linguistic domains.

For their sample, Dale et al. (2003) explored whether it was possible to predict if children would fall in persisting delay (n = 372) or resolving delay (n = 250) groups on the basis of their profiles at 2 years of age. Children whose delays would persist scored reliably lower across a number of parental rating measures, including vocabulary, grammar, displaced reference, and nonverbal skills, and showed reliably lower maternal education and a greater incidence of ear infection. Nevertheless, the effect sizes were small (.01–.06), and logistic regression analyses found that children's profiles at age 2 offered only modest classification of outcome at age 4, with accuracy rates between 60% and 70% (where chance would be 50%). The derived function failed to detect the majority of children who would show a persisting delay. A substantial minority of children whose delay was predicted to resolve did not show such resolution. Thus although persisting and resolving delay groups differed marginally at diagnosis, it was difficult to predict outcome with any accuracy. Fernald and Marchman (2012) found a similar level of success predicting outcomes earlier in development. The authors grouped children as typically developing (n = 46) or late talkers (n = 36) on the basis of productive vocabulary scores at age 18 months, where late talkers scored below the 20th percentile. Predicting group membership at age 30 months on the basis of vocabulary showed 78% sensitivity and 68% specificity, where sensitivity is the correct prediction of group membership and specificity is the correct prediction of nongroup membership. Fernald and Marchman found that an online measure of the efficiency of language processing at age 18 months provided additional predictive power for language outcome over and above vocabulary. Even given this moderate success in predicting patterns of behavior, the underlying causes of developmental patterns remain unclear, and the variety of outcome measures used makes it difficult to generalize across studies.

Causes of Language Delay

No clear picture has emerged from theoretical accounts of language delay as to why delay should resolve in some cases but not in others. Such accounts tend to differ on two dimensions: (a) whether children with persisting and resolving delay form qualitatively or quantitatively different groups and (b) whether the relevant causal factors are genetic or environmental (see Bishop, Price, Dale, & Plomin, 2003).

On the first dimension, Rice and colleagues (see, e.g., Rice, 2009) have argued that, with respect to specific language impairment (SLI), persisting delay can be traced to the developmental impairment in a specialized system for acquiring morphosyntax and, therefore, form a qualitatively different group. The account assumes that multiple, distinct linguistic domains are involved in language acquisition and that these domains are not well synchronized in children with SLI. However, once language growth begins in a domain, it may proceed in a fashion similar to that seen in typical development. Resolving delay might, then, constitute the bottom of the distribution of normal variation of children who do not have such a specific developmental impairment. This idea has gained suggestive support from genetic analyses implicating a potentially monogenetic cause of SLI (Bishop, 2005). By contrast, researchers such as Leonard (1987) and Rescorla et al. (2000) have suggested that persisting and resolving delay are only quantitatively different: There is a single continuum of individual variation in rates of language development caused by the same kinds of mechanisms, with persisting delay representing a more extreme case than resolving delay. For example, Fernald and Marchman (2012) characterized the dimensional view as assuming that "children with language impairment represent the lower end of multiple continuous dimensions of language skill that are normally distributed" (p. 204). This view is supported by the consistent finding that the strongest predictor of later language outcome—for example, in vocabulary development—is earlier language performance (although, even here, the variance explained is usually less than 40%; see, e.g., Chiat & Roy, 2008; Henrichs et al., 2011). The differences found in later development may be exaggerated versions of those already present in early development.

With respect to the second dimension of nature–nurture, the authors of several studies have implicated environmental factors in rates of language development. Nelson, Welsh, Vance Trup, and Greenberg (2011) examined 336 four–year-old children living in poverty and found that a majority of them exhibited clinically significant language delays. To the extent that poverty is an environmental condition, this implicates environment in causing language delay. Researchers have successfully used socioeconomic status (SES) gradients to explain variation in all aspects of language development, from phonological awareness (McDowell, Lonigan, & Goldstein, 2007) to the comprehension (Huttenlocher, Vasilyeva, Cymerman, & Levine, 2002) and production (Huttenlocher, Waterfall, Vasilyeva, Vevea, & Hedges, 2010) of complex grammatical forms, although vocabulary size may be the language measure most sensitive to the effects of SES. Hart and Risley (1995) observed different rates of language development in children from different SES backgrounds and linked these with large variations in the quantity of language spoken to the child. More recently, Henrichs et al. (2011) reported that maternal education was associated with both patterns of late-onset delay

Complimentary Author PDF: Not for Broad Dissemination

and persisting delay in a population study of 3,759 toddlers. When Anushko (2008) analyzed the development of language skills of 230 children at 15, 27, 37, and 72 months of age, she found that socioeconomic factors, the children's social–emotional competence, and the level of language exposure (as measured by book reading at home) all reliably predicted the rate of increase in expressive and/or receptive language (see also Anushko, Jones, & Carter, 2009). It is interesting to note that when children were initially split between low- and high-performing groups, those who were able to accelerate from low to high groups had significantly more exposure to and experience with language through book-reading activities than their peers who remained in the low-growth group across the time points. A number of longitudinal studies have shown that differences in the richness of linguistic input result in an increasing gap in children's language development (Huttenlocher et al., 2010; Reilly et al., 2010; Rowe, Raudenbush, & Goldin-Meadow, 2012). The consequences of this developmental divergence are persistent and pervasive through the school years (for a review, see Hoff, 2013). Bishop et al. (2003) used behavioral genetic methods with their sample, which comprised monozygotic and dizygotic twin pairs, to explore the etiology of persisting and resolving delay. The results indicated similar and modest heritabilities of .25 for both groups, implicating environmental factors in the cause of delay.

By contrast, other researchers have argued that environmental factors play little role in delay, on the basis of the failure of measured environmental variables to predict language levels. In a sample of 1,766 children, Zubrick Taylor, Rice, and Slegers, 2007) found that SES, family, and maternal characteristics did not predict language delay at age 2. The strongest predictors were family history for late language, male gender, and early neurological problems. In a longitudinal study of language growth with a smaller sample size, Rice, Wexler, and Hershberger (1998) did not find any predictive power of maternal education on the growth of morphemes marking tense and agreement in typically developing children and in children with SLI. Moreover, Dale et al. (2003) found that adding level of maternal education to their logistic regression analysis failed to improve their ability to predict whether children's language delays would persist or resolve. Overall, although children who have been diagnosed with early language delay are clearly heterogeneous, the dimensions defining the heterogeneity remain unclear (Desmarais, Sylvestre, Francois, Bairati, & Rouleau, 2008).

Part of the challenge stems from the current limited understanding of the mechanisms that might cause delay. Delay is most often used descriptively rather than mechanistically to recognize the fact that the behavior in some target group resembles that of younger, typically developing children (Thomas et al., 2009). At a mechanistic level, maturational views of delay have been most clearly articulated. These characterize language development as analogous to biological growth and growth rate variations as reflecting differences in (putative) genetically controlled timing mechanisms (see, e.g., Rice, 2009, for such an account in the context of morphosyntax development). A recent proposal by Rice (2012) stipulates that genetic mechanisms operate directly but independently to influence, respectively, the onset, acceleration, and deceleration of language growth trajectories. Experience-dependent views of delay are less frequently articulated, but these entail either a language system that receives fewer learning experiences or a learning system that is less malleable, such that more experience is required to effect a change in behavior. The specificity of causal mechanisms is also unclear. As well as language-specific proposals, some accounts maintain that children with language delay have more general processing limitations. These accounts are motivated by evidence that children with language impairments also show deficits in the speed (Lahey, Edwards, & Munson, 2001; Miller, Kail, Leonard, & Tomblin, 2001) or accuracy (Bavin, Wilson, Maruff, & Sleeman, 2005) with which they perform nonverbal tasks such as mental rotation (Johnston & Ellis Weismer, 1983) and visuospatial memory (Hoffman & Gillam, 2004). However, the exact nature of the generalized deficit remains to be adequately articulated, with authors variously attributing behavioral differences to slower processing speed (Miller et al., 2001), restrictions in working memory capacity (Ellis Weismer, Evans, & Hesketh, 1999), or reduced efficiency of resource allocation across domains (Im-Bolter, Johnson, & Pascual-Leone, 2006).

Using Computational Modeling to Investigate Mechanisms of Delay

One way to address the superficial consideration given to mechanisms of delay is through the use of computational modeling. By virtue of implementation, modeling can advance the detail with which theoretical accounts are specified. Models that embody theoretical proposals can then test the viability of those proposals to account for the observed empirical data. Models can show how a single mechanistic account can unify a range of previously disparate empirical phenomena. And models can generate novel predictions that can then be evaluated against empirical data.

In this article, we consider a computational model that addresses the differences between persisting and resolving language delay. The model takes advantage of a new method called *population modeling* (Thomas, Baughman, Karaminis, & Addyman, 2012). In population modeling, the aim is to simulate a large population of individuals undergoing a developmental process. In this population, multiple intrinsic and extrinsic properties are varied across individuals (where the term *intrinsic properties* refers to the computational abilities of each system and the term *extrinsic properties* refers to the quality of the learning environment to which each system is exposed). In combination, these factors produce a distribution of performance as the population acquires the target behavior. It is possible that atypical conditions can be applied to individuals against this background of variation. The framework permits not only the study of individual differences in rates of development but also the investigation of how cases of qualitatively atypical

Complimentary Author PDF: Not for Broad Dissemination

development may differ from those seen in the typically developing population. In the present study, we employed population modeling to investigate the idea that different limitations in computational processing might be the cause of developmental delays in language acquisition. We then pursued the following four aims: (a) to establish whether a quantitative account of developmental variations in a population was sufficient to generate subgroups that demonstrated persisting delay and resolving delay or whether qualitative differences were necessary; (b) to evaluate whether there were differences in the behavioral profiles of these subgroups when delay was first diagnosed; (c) to assess the role of environmental variation in causing developmental delays or aiding their resolution; and (d) to investigate, in implemented simulations, the mechanisms responsible for producing cases of persisting versus resolving delay.

Method

Simulation Overview

A population of 1,000 artificial neural networks was exposed to the language domain (English past tense), and their developmental trajectories were analyzed. English past tense was used as the language domain because it allows a direct comparison to previously reported empirical data (Bishop, 2005) and because past tense formation is well characterized by previous computational models (Thomas & Karmiloff-Smith, 2003). The model served as an example of a developmental system applied to the problem of extracting the latent structure of a language domain through exposure to a learning environment. Two sources of variation caused individual differences in the models' rates of development. We encoded *extrinsic variation* by altering the amount of information available in the input, which is analogous to the richness of the linguistic environment to which a child is exposed. We encoded *intrinsic variation* by altering the quality of the learning mechanism, inspired by accounts of SLI that propose that language delays may be caused by processing limitations. Implementation involved variations in 16 neurocomputational parameters controlling each artificial neural network's construction, activation, adaptation, and maintenance. Variations in the parameters can be thought of as modulating four broad properties of a network's functioning: its learning capacity, its plasticity, its quality of signal, and its possible regressive events (although some parameters contribute to more than one role). From the population of 1,000 simulated individuals, we used early performance on regular verb acquisition to define a delay group, and we traced their subsequent progress with reference to the population's normal range to identify different possible outcomes.

Simulation Details

Base model. A three-layer, backpropagation network was trained to output the past tense form of a verb from an input vector combining a phonological representation of the verb stem and lexical–semantic information (Joanisse & Seidenberg, 1999). The training set was the "phone" vocabulary from Plunkett and Marchman (1991, p. 70). This comprised an artificial language set constructed to reflect many of the important structural features of English past tense formation. There were 500 monosyllabic verbs, constructed using consonant–vowel (CV) templates and the phoneme set of English. Phonemes were represented over 19 binary articulatory features, 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 five features to represent the affix for regular verbs. The input layer included 500 units to encode the lexical status of each verb in the training set using a localist encoding scheme (Joanisse & Seidenberg, 1999; Thomas & Karmiloff-Smith, 2003). Thus, networks had 557 input units ($[3 \times 19] + 500 = 557$) and 62 output units ($[3 \times 19] + 5 = 62$). There were four types of verbs in the training set: (a) regular verbs that formed their past tense by adding one of the three allomorphs of the +ed rule, conditioned by the final phoneme of the verb stem (e.g., for English: *tame–tamed, wrap–wrapped, chat–chatted*); (b) irregular verbs whose past tense form was identical to the verb stem (e.g., *hit–hit*); (c) irregular verbs whose past tense was formed by changing an internal vowel (e.g., *hide–hid*); (d) 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 2, respectively, of each irregular verb type. We constructed a separate set of novel verbs to evaluate the generalization performance of the network. These verbs could differ depending on their similarity to items in the training set. Generalization in this case was assessed via 410 novel verbs, each of which shared two phonemes with one of the regular verbs in the training set, and was evaluated based on the proportion of these novel verbs that were assigned the correct allomorph of the regular past tense inflection.

Encoding extrinsic variation. Our manipulation of extrinsic variation was motivated by research on SES influences on language development. Each child was assigned a value (henceforth referred to as the *family quotient* [FQ]) to represent the quality of the linguistic environment in which they would be raised. This was a value between 60% and 100%. We used it as a probability to determine whether each verb in the full perfect training set would be included in the child's individual training set. The training set was then fixed. This method captured the reduced quantity of language input experienced by children in lower SES families, including in past tenses (Hart & Risley, 1995), but also the reduced diversity of words (Hoff, 2003). A lower SES family would be modeled as using fewer past tenses overall, fewer types of regular verbs, and fewer types of irregular verbs (see Thomas, Forrester, & Ronald, 2013, for further implementation details). Performance was always assessed against the full training set.

Encoding intrinsic variation. Connectionist networks contain a range of parameters that increase or decrease their ability to learn a given training set. Computational parameters such as learning rate, momentum, and number of

Complimentary Author PDF: Not for Broad Dissemination

hidden units feature in most published simulations. In models of typical or average development, parameters are optimized to achieve best learning (usually in the presence of the perfect training set). In the present model, a number of parameters were simultaneously varied across individual networks, with overall learning ability determined by their cumulative effect. Therefore, the mechanistic variations producing differences in the rates of development were only quantitative. Variations occurred across 16 computational parameters, allowing for more than 2 trillion possible unique individuals.

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 activation*—unit threshold function, processing noise, and response accuracy threshold.
- *Network adaptation*—backpropagation error metric used in the learning algorithm, learning rate, and momentum.

As well as an overall learning rate, there were separate parameters modifying the learning rate between the semantic input units and the hidden units and between the phonological input units and the hidden units, potentially altering the relative balance of these sources of information during learning and, therefore, allowing more lexical or phonological strategies to past tense acquisition. Finally, the parameter of *network maintenance* consisted of weight decay, pruning onset, pruning probability, and pruning threshold.[1]

These parameters can be viewed as serving different types of processing roles within the network, although some parameters contribute to more than one role. Some parameters alter the network's *learning capacity*—that is, the complexity and the amount of information that can be learned. These parameters include the architecture, the number of hidden units, and the initial sparseness of connectivity. *Regressive events* involving pruning of connections can also reduce capacity later in development, implicating the pruning onset, pruning probability, and pruning threshold parameters in predicting learning trajectories (see Thomas, Knowland, & Karmiloff-Smith, 2011). The nature of the learning algorithm determines both what can be learned and also how quickly: The overall speed of learning can be thought of as the network's *plasticity*. Parameters that alter plasticity include the learning rate parameter, the learning rates in semantic and phonological connections, the momentum, the initial range of weight variation, and the unit threshold function. The unit threshold function determines how responsive a processing unit is to variations in its input and, therefore, to some extent, determines the *quality of the signal* propagating through the network. Signal is also affected by the level of processing noise and by the precision required of output units to drive a response.

[1]Formal specification of the parameters and their value ranges can be found in a technical report available at www.psyc.bbk.ac.uk/research/DNL/techreport/Thomas_paramtables_TR2011–2.pdf

Design. Development was traced across a population of 1,000 simulated individuals, focusing on the rate of acquisition of regular English past tense forms. One thousand sets of the 16 computational parameter values were generated at random, with parameters sampled independently. These sets were instantiated as 1,000 artificial neural networks. An FQ value was generated for each network and was used to create an individualized training set. Each network was trained for 1,000 epochs on its training set, where one epoch corresponded to a single presentation of all the verbs in the individual's training set. At each epoch, performance was measured on the perfect training set. Performance was assessed on regular verbs, on irregular verbs, and on generalization of the past tense rule to novel forms in order to generate a behavioral "profile" for each network. Performance was measured via accuracy levels (percent correct). Early performance on regular verb acquisition was used to define a delay group (see Results section below), and their subsequent progress was then traced with reference to the population normal range.

Results

Defining Delay

Five time points were defined in the development of the population, when the population accuracy for regular verb production was 40%, 50%, 60%, 65%, and 78%. These occurred at 31, 49, 84, 127, and 500 epochs of training, respectively. These points were selected to be representative of the stages of early population development, with the final time point reflecting population asymptote. Figure 1(a) shows the distribution of regular performance for Time 1, at which point individuals were identified as exhibiting developmental delay if their performance fell more than 1 *SD* below the population mean. This corresponded to 28.7% of the population. If the population were to be divided into quartiles based on FQ, the Time 1 delay group comprised 76 individuals from the lowest FQ quartile, 74 and 79 individuals from the middle quartiles, and 58 individuals from the upper quartile. Although the highest FQ quartile had the fewest delayed individuals, the distribution was not reliably different from chance, $\chi^2(3, N = 4) = 3.69$, $p = .297$). Developmental delay was then re-diagnosed at each subsequent time point, shown in Figures 1(b) through 1(e). Figure 2 shows the proportion of the population diagnosed with delay at each time point. A small number of those delayed at the final time point (18) were not delayed at Time 1. Late onset delay is also a prominent pattern reported in the literature (see, e.g., Henrichs et al., 2011; Ukoumunne et al., 2011). Of those delayed at the first time point, 118 showed a delay that persisted through to the final time point, whereas the delay resolved in 169 (or 58.9%) of the cases. By the final time point, only 13.6% of the population was categorized as delayed.

In terms of intrinsic and extrinsic influences on development, simulated individuals differed from one another only quantitatively, yet these quantitative differences were sufficient to produce persisting and resolving delay groups.

Complimentary Author PDF: Not for Broad Dissemination

Figure 1. Performance distribution on regular verbs at each time point, along with the cutoff for defining developmental delay. μ is the mean and σ is the standard deviation at each time point.

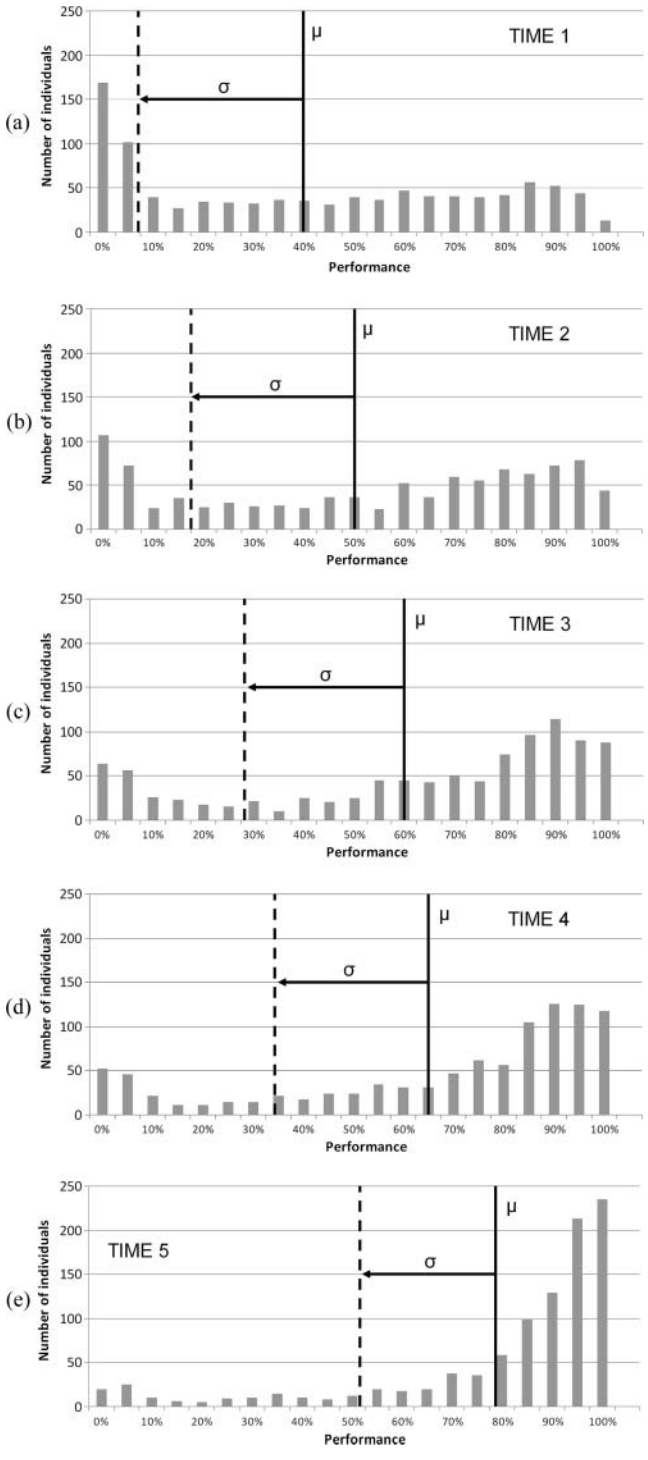
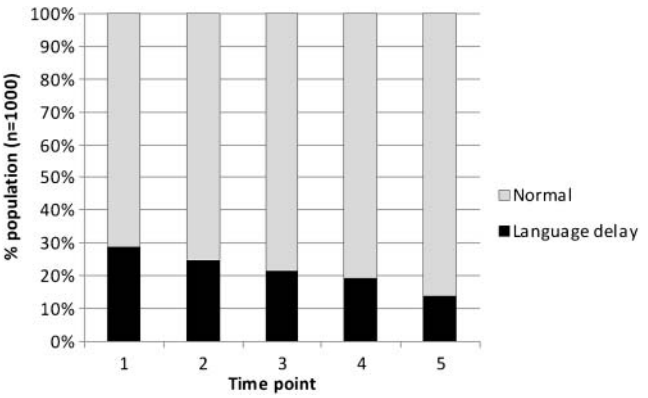


Figure 2. Proportion of simulated population exhibiting language delay at each time point, where delay was defined as falling more than 1 standard deviation below the population mean at that time point.



Indeed, the proportion of cases that resolved was broadly similar to the rates observed in empirical studies (Bishop, 2005; Bishop & Edmundson, 1987; Dale et al., 2003; Paul, 1996; Rescorla et al., 2000); it was, however, less than that observed by Whitehurst and Fischel (1994).

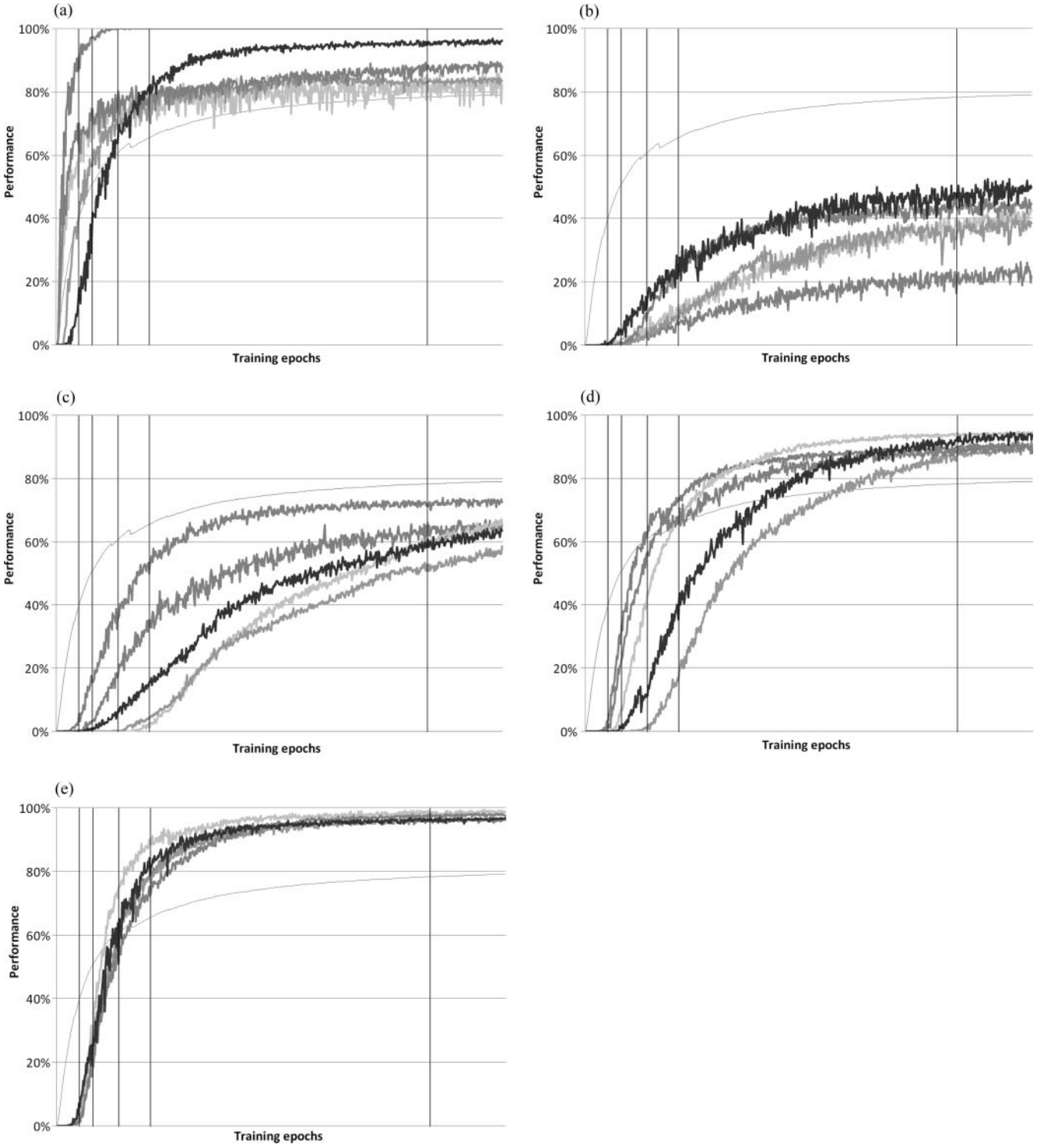
If persisting and resolving delay lie on a strict developmental continuum, one might expect cases of resolving delay to lie toward the bottom of the normal range—that is, resolving cases would slip into the normal range but would still perform relatively poorly and would not, therefore, represent complete resolution of the delay. The population rank orders of individuals in the resolving delay group were examined to evaluate this idea. Of those showing resolving delay, 80% (136 individuals) indeed remained in the bottom 500 of the population. However, in some individuals, performance at Time 5 was somewhat better: 17% (28 individuals) had a rank order in the top 500, and a few (3%; five individuals) even finished in the top 200. Therefore, the outcome of resolving delay was variable. Good final outcomes were possible, suggesting that, in some cases, delay could completely resolve. Examples of individual developmental trajectories for the five groups—typically developing, persisting delay, resolving delay with low outcome, resolving delay with good outcome, and resolving delay with very good outcome—are depicted in Figure 3. Each plot also contains the five time points and the mean developmental trajectory for the population as a whole.

Predicting Persisting Versus Resolving Delay From Time 1 Behavioral Profiles (and FQ)

The behavioral profiles of simulated individuals in the persisting and resolving delay groups were compared at Time 1. The profile initially included nine measures of various aspects of past tense performance on training and generalization sets. Behavior on three of these measures best summarized the pattern: regular verbs, vowel-change irregular verbs, and regularization of novel verbs. At Time 1, the persisting delay group performed reliably worse on

Complimentary Author PDF: Not for Broad Dissemination

Figure 3. Sample developmental trajectories for regular verbs, for each group: (a) typical development, (b) persisting delay, (c) resolving delay with low outcome, (d) resolving delay with good outcome, and (e) resolving delay with very good outcome. Trajectories are shown for the first 600 epochs. The final time point to determine outcome was 500 epochs of training. The thin black line represents the mean trajectory for the whole population, while thicker gray lines represent individual trajectories. Sample trajectories were selected to illustrate the range of variation within each group.



Complimentary Author PDF: Not for Broad Dissemination

regular verbs and novel verbs than did the resolving delay group, but there was no difference on irregular verbs: multivariate analysis of variance (MANOVA) for overall profile difference, $F(3, 283) = 12.22$, $p < .001$, $\eta_p^2 = .115$; individual measures, regular, $F(1, 285) = 36.84$, $p < .001$, $\eta_p^2 = .114$; individual measures, irregular, $F(1, 285) = 2.29$, $p = .131$, $\eta_p^2 = .008$; novel, $F(1, 285) = 28.00$, $p < .001$, $\eta_p^2 = .098$. Although the differences were highly reliable, they were of small effect size.

The Time 1 behavioral differences between persisting and resolving delay groups were small, but were they sufficient to reliably predict outcome group? Logistic regression analyses were used to predict outcome group based on the profile of performance on regular, irregular, and novel verbs. The results are contained in Table 1, which also includes the results of Dale et al.'s (2003) analyses that sought to predict delay outcome based on Time 1 measures of verbal ability, displaced reference, nonverbal ability, and maternal education. Dale et al. found that small initial behavioral differences between groups were not sufficient for accurate classification of delay outcome and that the addition of maternal education did not markedly increase predictive power. Similarly, the simulation data indicate that the small behavioral differences at Time 1 did not produce accurate prediction of outcome, and addition of the FQ variable as a predictor also produced no marked improvement.[2]

The bottom row of Table 1 shows the predictive power when each simulated individual's neurocomputational parameters were added to the logistic regression: Accuracy of classification was now much higher but still somewhat short of 100%. This illustrates the operation of stochastic factors in the model and the nonlinear interactions that occur between parameters in determining overall learning ability. It is noteworthy that addition of the neurocomputational parameters to the analysis improved the ability to predict which networks would have resolving delay from a position of previously overpredicting persisting delay. This suggests that we may find neurocomputational parameters that are markers for the resolution of delay.

Even within a single computational learning system and even when using noise-free performance measures, the correlation between population performance at different time points reduced as the time points became more remote. Time 1 performance on regular verbs predicted only 37% of the variance in Time 5 performance on regular verbs. The result is consistent with the empirical finding that early language performance is the best predictor of later language performance, but the variance explained can be relatively modest (see, e.g., Chiat & Roy, 2008; Henrichs et al., 2011).

Time 1 performance contained some individuals with scores at floor; this may have differentially affected irregular verbs, which are harder to learn, thus producing the lack of a reliable difference between persisting and resolving delay profiles. To evaluate this possibility, we carried out the group comparison at each subsequent time point. Of course, differences must increase on all measures because, by definition, the performance of persisting and resolving delay groups will diverge. However, if the initial overlap in irregular verb performance between the groups was due to floor effects, the effect sizes of the irregular difference should approach those of regular and novel verbs as performance comes off floor in those individuals. Across all time points, the effect sizes were consistently larger for regular and novel verbs than for irregular verbs. Persisting and resolving delay groups, then, were distinguished at Time 1 by small differences in the ability to abstract the regularities of past tense morphology from exposure to the learning environment, with the persisting delay group less able to do so.

Mechanistic Explanations for Persisting Versus Resolving Delay

To explore the mechanistic basis of the distinction between persisting and resolving delay groups, we carried out two complementary sets of analyses using neurocomputational parameter values to predict outcome group, either through the use of MANOVA or multinomial logistic regression (MLR). Table 2 shows the results of a statistical comparison of the mean neurocomputational parameter values for simulated typically developing, persisting delay, and resolving delay groups, using the two statistical methods. Table 3 shows equivalent results for a comparison of the resolving delay group, split by whether the final outcome was low (bottom 500 of population), good (top 500 of population), or very good (top 200 of population). Table 4 incorporates the mean parameter values per group.

Both delay groups differed from the typically developing group across a range of neurocomputational parameters (see Table 2). The strongest effect size for the difference between typical and persisting delay was the power of the learning algorithm. The strongest effect size for the difference between typical and resolving delay was the learning rate. Individually, the delay groups did not differ from the typically developing group on the FQ variable. However, when combined, there was a small difference that approached significance, with delay groups showing lower FQ values, $t(998) = 1.93$, $p = .054$, Cohen's $d = 0.136$; see Figure 4.

The persisting delay group differed from the resolving delay group across a smaller number of parameters. Individuals showing persisting delay tended to have fewer hidden units, a higher pruning threshold (leaving the network at greater risk of connection loss across development), a less powerful learning algorithm, and higher processing noise. In terms of processing roles, the more salient cause of delay was lower computational capacity and poorer signal. By contrast, individuals showing resolving delay had a shallower unit activation function and a lower learning rate in the semantic pathway. In terms of processing roles, the more salient cause of delay in this group was lower plasticity.

Table 3 indicates which parameters predicted the final outcome for individuals showing resolving delay.

[2]The prediction equations for empirical data and model were poor in different ways, with the data equation overpredicting the resolution of delay and the model equation overpredicting the persistence of delay.

Complimentary Author PDF: Not for Broad Dissemination

Table 1. Results from logistic regression analyses, predicting delay group (PD vs. RD).

Predictor	Fit*		% classified correctly	Sens[a] (%)	Spec[b] (%)	PPV[c] (%)	NPV[d] (%)
	df	χ^2					
Empirical data from Dale et al. (2003)							
(a) Vocabulary, displaced reference, nonverbal	3	67.0	65.8	44.6	80.5	61.4	67.7
(b) Add in gender and maternal education	5	91.1	68.5	51.5	80.0	63.8	70.7
Simulation data							
(c) Time 1 behavioral markers	3	39.6	65.5	79.7	55.6	55.6	80.0
(d) Add SES proxy	4	39.7	65.2	79.7	55.0	55.3	79.5
(e) Add neurocomputational parameter set	21	113.9	79.1	72.0	84.0	75.9	81.1

Note. Item (a) contains empirical data from Dale et al. (2003, Table 6), predicting resolving delay at 4 years of age based on parental report measures of vocabulary, displaced reference, and nonverbal ability at 2 years of age. Item (b) contains empirical data from Dale et al. (2003), adding in gender and a measure of mother's education. Item (c) contains simulation data predicting Time 5 delay group based on Time 1 measures of regular verb, irregular verb, and novel verb performance. Item (d) contains simulation data adding in each individual's family quotient parameter. Item (e) contains simulation data, adding in the full set of neurocomputational parameters for each individual.

PD = persisting delay; RD = resolving delay; Sens = sensitivity; Spec = specificity; PPV = positive predictive value; NPV = negative predictive value.

*All χ^2 values are significant at $p < .001$.

[a]Proportion of PD whose persisting delay was correctly predicted. [b]Proportion of RD whose resolving delay was correctly predicted. [c]Positive predictive value (PPV) = proportion of predicted PD who had persisting delay. [d]Negative predictive value (NPV) = proportion of predicted RD who had resolving delay.

Outcome depended on the two previously identified plasticity parameters, semantic pathway learning rate, and unit threshold function. A lower semantic pathway learning rate was associated with poorer final outcome, whereas a shallower unit threshold function was associated with better final outcome. As causes of resolving delay, these parameters had differential effects on the potential final level that could be achieved. Most notably, however, the final level of performance was associated with the FQ parameter: the richer the environment, the higher the final level that could be achieved. This pattern emerged despite the relatively weak contribution of FQ in explaining individual differences in the population as a whole (e.g., at Time 5, FQ predicted only 2.2% of the variance in regular verb performance in the full population; in the resolving delay group alone, FQ predicted 8.0% of the variance).

As a minor point, for the results in Table 3, one might ask why the semantic pathway learning rate discriminated between the RL and RG delay groups, rather than the phonological pathway learning rate. This was due to the definition of *delay* according to regular verb performance. Verbs in the training set could be learned by a lexical strategy, facilitated by the lexical–semantic input. By contrast, generalization of the past tense rule was dependent on phonological similarity. Had delay been defined according to regularization of novel verbs, the phonological pathway learning rate would have been the more salient parameter in modifying rates of development.

In summary, a consideration of mechanisms suggests the following picture. Both persisting and resolving delay are caused by a combination of suboptimal learning parameters. Most salient in persisting delay is a limit on the computational capacity of the learning system, which places a ceiling on the highest level that can be achieved. Most salient in the resolving delay is lower plasticity, which reduces the rate of learning but does not place the same ceiling on the highest level that can be achieved. To some extent, the final level is then determined by the richness of the environment in which the learning system is embedded. By contrast, the richness of the learning environment is much less relevant to learning in reduced capacity systems. Although early on, the delay groups are conflated, capacity places a limit on subsequent learning in a way that plasticity does not.

Testing a Novel Prediction of the Model

As indicated above, the model generated a novel prediction that the amount of input available should reliably predict outcome in the resolving delay group but not the persisting delay group. In this section, we test this prediction using a data set from Bishop (2005).[3] Bishop (2005) analyzed data from the large British sample of twins considered in Dale et al. (2003) and Bishop et al. (2003). Bishop (2005) identified a sample of the twins who exhibited language delay risk at 4 years of age. These children, along with a sample of twins not identified as being at risk, were tested at 6 years of age on a test of English past tense production (Rice & Wexler, 2001). At 6 years of age, around one third of the early language delay risk group then met psychometric criteria for SLI, compared with one in 10 of those not identified as being at risk (Bishop, Laws, Adams, & Norbury, 2006). From the sample, three groups of children could be identified: (a) 94 six–year-old children both exhibiting language delay risk at 4 years of age and meeting psychometric criteria for SLI at 6 years of age (persisting delay); (b) 104 six–year-old children exhibiting language delay risk at 4 years of age but not meeting psychometric criteria for SLI at 6 years (resolving delay); and (c) 166 children exhibiting neither language delay risk at 4 years of age nor SLI at 6 years of age

[3]We are grateful to Dorothy Bishop for making the raw data available to us.

Complimentary Author PDF: Not for Broad Dissemination

Table 2. Neurocomputational parameters that reliably discriminated between groups.

Parameter	Role	TD vs. PD		TD vs. RD		PD vs. RD	
		ANOVA	MLR	ANOVA	MLR	ANOVA	MLR
Hidden units	Capacity	**.030****	6.6*	**.005***	4.6*	**.031****	
Architecture	Capacity	**.018****	18.1**	**.013****	17.4**		
Sparseness	Capacity						
Pruning onset	Capacity						
Pruning probability	Capacity						
Pruning threshold	Capacity			**.004***		**.021***	5.4*
Learning algorithm	Capacity/plasticity	**.172****	98.8**	**.012****	24.8**	**.104****	21.3**
Learning rate	Plasticity	**.030****	16.4**	**.044****	33.2**		
Semantic learning rate	Plasticity	**.005***				**.024****	
Phonological learning rate	Plasticity	**.018****	7.0**	**.014****	8.6**		
Momentum	Plasticity	**.006***	4.7*	**.015****	12.8**		
Weight variance	Plasticity			**.009****	11.2**		
Unit threshold function	Plasticity/signal			**.036****	23.0**	**.025****	5.5*
Processing noise	Signal	**.021****	19.1**			**.026****	

Note. TD = typically developing/no delay; PD = persistent delay; RD = resolving delay; ANOVA= analysis of variance; FQ = family quotient; SES = socioeconomic status; MLR = multinomial logistic regression. Results are shown for two complementary statistical analyses. Scores from ANOVA show partial η^2 effect sizes. Scores from MLR show Wald statistic for each parameter as a measure of effect size. Empty cells represent nonreliable differences ($p > .05$). Boldface indicates use of the ANOVA method to compare group parameter means.

MLR model fit: TD versus all delay groups, $\chi(72) = 411.3$, $p < .001$, Nagelkerke $R^2 = .405$; PD vs. RD group, $\chi(18) = 79.9$, $p < .001$, Nagelkerke $R^2 = .328$.

*Effect reliable at $p < .05$. **Effect reliable at $p < .01$.

(typical development). Importantly, SES data, as measured by parental education, occupation, and maternal age at birth, were also available for these children's families. As noted in the introduction to this article, SES has been shown to correlate with language input (Hart & Risley, 1995).

As with the FQ scores in the simulation data, when both early delay groups were combined, they yielded reliably lower SES scores than did the typically developing group, $t(362) = 2.75$, $p = .006$, Cohen's $d = 0.290$. In this case, the persisting delay group was also reliably lower in SES

Table 3. Neurocomputational parameters that reliably discriminated between resolving delay groups.

Parameter	Role	RD-L vs. RD-G		RD-L vs. RG-VG		RD-G vs. RD-VG	
		ANOVA	MLR	ANOVA	MLR	ANOVA	MLR
Hidden units	Capacity						
Architecture	Capacity						
Sparseness	Capacity						
Pruning onset	Capacity						
Pruning probability	Capacity						
Pruning threshold	Capacity						
Learning algorithm	Capacity/plasticity						
Learning rate	Plasticity						
Semantic learning rate	Plasticity	**.053****	6.1*				
Phonological learning rate	Plasticity						
Momentum	Plasticity						
Weight variance	Plasticity						
Unit threshold function	Plasticity/signal	**.064****	8.6**				
Processing noise	Signal						
Response threshold	Signal						
Weight decay	Signal						
FQ (SES)	Environment	**.089****	11.9**	**.095****	12.6**	**.111**$^+$	

Note. RD-L = resolving delay group with low outcome; RD-G = resolving delay group with good outcome; RD-VG = resolving delay group with very good outcome. Results are shown for two complementary statistical analyses. Scores show partial η^2 effect sizes.

*Effect reliable at $p < .05$. **Effect reliable at $p < .01$. $^+$Effect significance, $p = .058$. Scores show Wald statistic for each parameter. Empty cells represent nonreliable differences ($p > .05$). Boldface indicates use of the ANOVA method to compare group parameter means.

MLR model fit: comparison of three resolving delay groups, $\chi(36) = 57.3$, $p < .001$, Nagelkerke $R^2 = .420$; pairwise comparisons: RDL vs. RDG, $\chi(18) = 47.135$, $p < .001$, Nagelkerke $R^2 = .417$; RDL vs. RDVG, $\chi(18) = 4.844$, $p = .999$, Nagelkerke $R^2 = .128$; RDG vs. RDVG, $\chi(18) = 12.951$, $p = .794$, Nagelkerke $R^2 = .567$.

Complimentary Author PDF: Not for Broad Dissemination

Table 4. Mean values for neurocomputational parameters and environment for the groups.

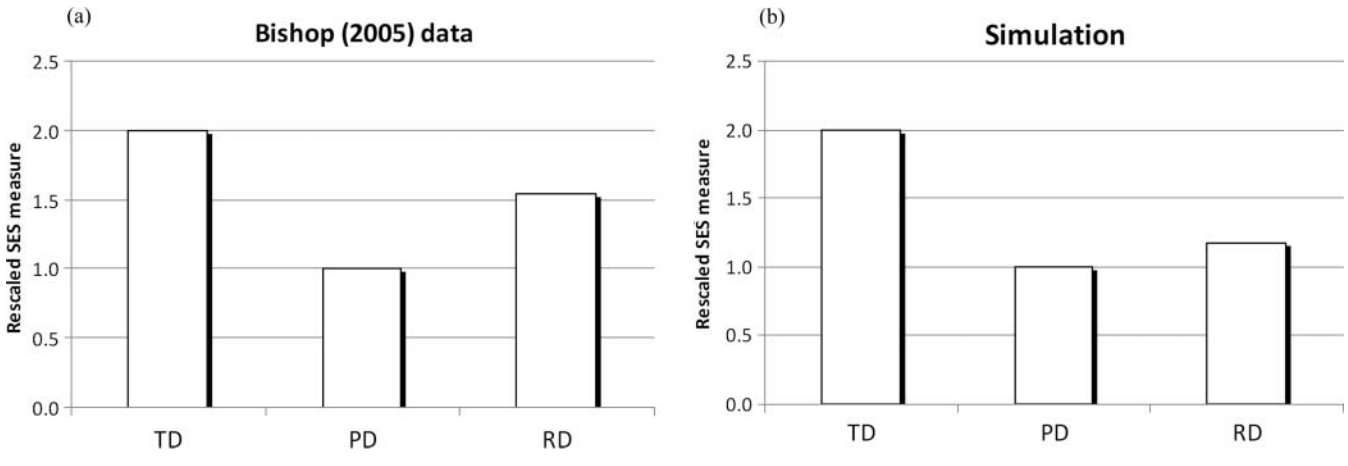
Parameter	Role	Group				
		TD (*n* = 713)	PD (*n* = 118)	RD-L (*n* = 136)	RD-G (*n* = 28)	RD-VG (*n* = 5)
Hidden units	Capacity	31	22	27	29	22
Architecture	Capacity	1.08	.87	.93	.93	.80
Sparseness	Capacity	.06	.07	.06	.05	.10
Pruning onset	Capacity	105	100	104	99	130
Pruning probability	Capacity	.14	.15	.12	.10	.15
Pruning threshold	Capacity	.53	.55	.51	.50	.42
Learning algorithm	Capacity/plasticity	.97	.66	.91	.93	1.00
Learning rate	Plasticity	.13	.11	.11	.12	.13
Semantic learning rate	Plasticity	.55	.48	.55	.74	.65
Phonological learning rate	Plasticity	.38	.26	.28	.32	.32
Momentum	Plasticity	.27	.23	.22	.23	.15
Weight variance	Plasticity	.53	.56	.64	.52	.70
Unit threshold function	Plasticity/signal	1.29	1.24	1.14	.82	1.00
Processing noise	Signal	.60	.83	.59	.63	.65
Response threshold	Signal	.09	.04	.04	.03	.03
Weight decay ($\times 10^{-7}$)	Signal	5.37	15.70	3.57	7.07	4.52
FQ (SES)	Environment	.80	.79	.77	.86	.94

compared with the typical group, but no other differences were reliable: typically developing versus persisting delay, $t(258) = 3.18$, $p = .002$; typically developing versus resolving delay, $t(268) = 1.44$, $p = .152$; persisting delay versus resolving delay, $t(196) = 1.60$, $p = .111$. As with the FQ scores in the simulations, the SES measure demonstrated only a small ability to predict individual differences in regular past tense formation, explaining only 0.5% of the variance across all groups, $F(1, 362) = 1.85$, $p = .175$. In part, this was due to ceiling effects in regular verb scores.

Did SES show a differential ability to predict performance across the delay groups? Table 5 compares the results of linear regressions between SES and past tense performance for the Bishop (2005) sample. The persisting delay group showed no sign of a relationship, whereas for the resolving delay group, there was a weak trend: The higher the SES value, the better the regular past tense performance. A comparison of these two relationships yielded a reliable interaction between persisting and resolving delay groups, SES, and regular verb performance: There was a reliably stronger relationship between SES and performance in the resolving delay group than in the persisting delay group, $F(1, 194) = 4.015$, $p = .047$, $n_p^2 = .020$. Finally, addition of irregular verb performance from the Rice and Wexler (2001) test increased sensitivity, with fewer ceiling scores. In this case, both typically developing and resolving delay groups demonstrated a reliable relationship between SES and past tense performance, whereas the persisting delay group did not.

In sum, a test of the novel prediction of the model through available empirical data produced support for the

Figure 4. (a) Mean SES values for typically developing (TD), persisting delay (PD), and resolving delay (RD) groups for the children in the Bishop (2005) study (*N* = 166 TD, 94 PD, 104 RD). (b) Simulation data showing mean FQ values for *N* = 713 TD, 118 PD, 169 RD. SES and FQ measures have been rescaled to a common range (1 = lowest group mean, 2 = highest group mean).



Complimentary Author PDF: Not for Broad Dissemination

Table 5. Effect size (R^2) of relationship between SES and performance, per group, for the Bishop (2005) sample in English past tense production

	Measure			
	Regular verbs		Regular + irregular verbs	
Group	R^2	*p*	R^2	*p*
TD (*n* = 166)	.002	.540	.044	.007
PD (*n* = 94)	.017	.209	.003	.611
RD (*n* = 104)	.028	.087	.043	.035

Note. Regular verbs, Group (PD vs. RD) × SES interaction: $F(1, 194) = 4.015$, $p = .047$, $n_p^2 = .020$.

Regular and irregular verbs, Group × SES interaction: $F(1, 194) = 2.83$, $p = .094$, $n_p^2 = .014$.

model. Despite the weak overall predictive power of SES on performance (see, e.g., Rice et al., 1998), the resolving delay group showed a stronger relationship with SES in their past tense performance than did the persisting delay group.

Discussion

The current computational model was successful in demonstrating that in a population of developing systems, which varied along continua of intrinsic and extrinsic parameters, early-diagnosed delay resolved in some individuals but persisted in others when exposed to a language domain. The proportion of resolving cases was similar to that observed in empirical studies of early language delay (Bishop, 2005; Bishop & Edmundson, 1987; Dale et al., 2003; Paul, 1996; Rescorla et al., 2000); it was, however, less than that observed by Whitehurst and Fischel (1994). The model captured this pattern through quantitative variations in learning parameters between individuals. Therefore, the model demonstrates the viability of Leonard's (1987) and Rescorla et al.'s (2000) proposal that resolving and persisting delay lie on a continuum of individual variation in rates of language development.

The model qualitatively accorded with six further empirical findings. First, as per Dale et al. (2003), there were small differences in the behavioral profiles of persisting delay and resolving delay groups when delay was first diagnosed. In the model, these were differences in the ability to extract the latent structure of the language domain (the regular past tense "rule"), with the persisting group less able to do so. Second, also as per Dale et al. (2003), these small behavioral differences were not particularly effective in predicting individual outcomes. Third, the model included a manipulation of the richness of language input, equivalent to the variations produced by differences in SES (Thomas, Forrester, & Ronald, 2013); this manipulation accounted for only a small amount of the variance in past tense formation, similar to the findings of Rice et al. (1998) and the larger dataset of Bishop (2005). Fourth, as with Dale et al. (2003), addition of this environmental richness measure to early behavioral differences did not improve the ability to predict delay outcomes. Similar to the data of Bishop (2005), delay groups showed slightly lower FQ scores than did the typically developing group, implicating environmental factors (to some minor extent) in the cause of delay. Fifth, the model simulated cases of late-onset delay, which also have been observed in the literature (Henrichs et al., 2011; Ukoumunne et al., 2011). Sixth, despite the apparently small influence of the environmental input manipulation on the performance of the simulated population, the model generated a novel prediction that the richness of the environment should reliably predict the developmental outcome of individuals with resolving delay but not those with persisting delay; this novel prediction was subsequently supported by the empirical data of Bishop (2005). The model indicated that although the majority of individuals with resolving delay finished in the low–normal range, some individuals in a rich environment finished in the top half or top fifth of the population. Although resolving delay fell between persisting delay and typical development on a mechanistic continuum, it was true that some cases of early delay could completely resolve.

The advantage of a computational model is that one can examine the mechanisms responsible for producing—or reproducing—behavioral patterns. Variations in rates of development were generated by small differences in a relatively large number of neurocomputational parameters (16 total) that influenced processing within an artificial neural network learning system. The majority of these parameters were implicated in causing delay. Variations in these parameters overlapped between persisting and resolving delay groups, but differences could be discerned between delay types. Parameters were identified by broad computational processing roles, including those of *capacity, plasticity, signal,* and *environment*. Persisting delay was more strongly associated with limits in *capacity* as well as noise in the processing *signal*. For these networks, acquisition of the problem domain was restricted as a result of reduced processing resources that placed a limit on the amount and complexity of information that could be learned. Resolving delay was more strongly associated with low *plasticity*—that is, the speed of learning and the responsiveness of the system to inputs—and with *signal* limitations, whereby more precise outputs were required to drive responses, which were not achieved until later in learning. The environment influenced resolving delay outcomes because low plasticity eventually allowed the learning system to take advantage of richer information available in the environment, whereas for persisting delay, capacity limitations made networks insensitive to the information available in richer environments.

Model Evaluation

The model succeeded with respect to its four identified aims: (a) to establish whether a quantitative account of developmental variations in a population was sufficient to generate subgroups that demonstrated persisting delay and resolving delay or whether qualitative differences were necessary; (b) to evaluate whether there were differences in

Complimentary Author PDF: Not for Broad Dissemination

the behavioral profiles of these subgroups when delay was first diagnosed and to evaluate their ability to predict developmental outcomes; (c) to assess the role of environmental variation in causing developmental delays or aiding in their resolution; and (d) to investigate, in implemented situations, the mechanisms responsible for producing cases of persisting versus resolving delay.

The broader aim of the modeling work was to advance a mechanistic understanding of delay, which hitherto has either been ascribed to unspecified maturational processes, alluded to in terms of a vague idea of processing limitations, or remained as a largely descriptive notion. Population modeling was used to address individual differences within a developmental framework. Individual variation in rates of development was caused by a combination of intrinsic influences (neurocomputational parameters) and extrinsic influences (the richness of the input). The result was a population exhibiting a normal distribution of performance, which shifted across development and which was skewed by early floor effects and late ceiling effects (see Figure 1). All parameters, as well as the learning environment, were modeled as varying independently. Delay was caused by an accumulation of suboptimal learning parameters in unlucky individuals. Through nonlinear interactions between parameters, different constellations of poor parameters could lead to subtly different behavioral profiles: Delay was heterogeneous in detail (as illustrated by the persisting vs. resolving patterns) despite its quantitative origins.

The model is consistent with proposals that language delays arise from processing limitations. It adds detail about the form that the limitations may take and how particular processing limitations have different effects, even if their trajectories look similar early in development. The model cannot, however, address the specificity of processing limitations because only one system was simulated.

The population modeling approach presented here contrasts with previous computational models of atypical language development, which have simulated disorders by manipulating a single parameter while holding other parameters constant. For example, Joanisse (2004) simulated deficits associated with SLI in inflectional morphology by the addition of processing noise to phonological representations in a connectionist network, whereas Thomas (2005) captured similar empirical data by altering the unit threshold function in a past tense model (see Karaminis, 2011, for a more general consideration of how neurocomputational processing limitations can lead to behavioral symptoms of SLI in inflectional morphology, syntax comprehension, and syntax comprehension in a cross-linguistic context). By contrast, in simulating language delay as a chance accumulation of many suboptimal computational properties, the current simulations are more consistent with the behavioral genetics approach known as *quantitative trait loci* (QTL), which argues that normal and abnormal behavior lie on the same continuum of genetic variation and that with the exception of known genetic mutations of large effect, many behaviorally defined disorders represent the chance accumulation of a large number of common gene variants, each carrying a small risk for disorder—accumulations that will inevitably occur in large populations (Kovas, Haworth, Dale, & Plomin, 2007; Plomin, DeFries, McClearn, & McGuffin, 2008). However, the model framework is not committed to the assertion that processing properties are solely under genetic influence. Variations in the processing properties of language systems may also arise from other environmental influences, including caregiver input. For example, Hurtado, Marchman, and Fernald (2008) found that measures of mothers' speech heard by children at age 18 months predicted children's language processing efficiency at age 24 months, and Fernald, Marchman, and Weisleder (2013) observed that differences in language processing efficiency across SES groups were already present in children at age 18 months.

Model Limitations

As with any implemented model, there were limitations. First, our intention was to use the model as a notional developing system applied to a structured language domain rather than as a specific model of the acquisition of inflectional morphology. This is because the target empirical phenomenon considered disparate measures of language ability for the early and later markers of language delay. In many cases, delay was initially diagnosed with respect to vocabulary, or even parents' global concerns regarding their children's language development, with later assessments considering a wider range of measures that sometimes included morphosyntax. Used as an illustrative system, the model evaluated the hypothesis that patterns of persisting and resolving delay could emerge in the development of a single system. One could argue that a disconnect exists between the single system explored in the modeling and the multiple aspects of language addressed in the clinical consideration of language delay. Nonetheless, a novel prediction generated by the model was subsequently tested against and supported by empirical data from inflectional morphology.

The past tense model itself involved a number of simplifications necessitated by population modeling (Thomas, Forrester, & Ronald, 2013). For example, the model used an artificial past tense–like problem domain rather than a full-scale English verb corpus. Therefore, its manipulation of input to simulate the effects of SES could only be qualitative rather than a reflection of the real input data experienced by children in different family types. In addition, changes in the intrinsic properties of the artificial neural networks were restricted to learning properties rather than to the structure of the input and output representations. Some models have considered how differences in phonological or semantic information that is supplied to learning systems can cause variations in developmental trajectories, perhaps even simulating SLI (Hoeffner & McClelland, 1993; Karaminis, 2011; Thomas & Karmiloff-Smith, 2003). If the key distinctions that are required to learn the latent structure of a language domain are not present in the inputs and outputs of a learning system, then it is clear that acquisition cannot be

Complimentary Author PDF: Not for Broad Dissemination

successful. Alternatively, if those distinctions are simply encoded less saliently (i.e., with smaller activation differences), learning may ultimately be successful but may take longer. In other words, it is plausible that a distinction between persisting and resolving delay would occur if the manipulation of input and output representations were added to the parametric variations. Of course, this remains to be demonstrated with further simulations.

A further simplification of the model was that intrinsic and extrinsic parameters were sampled independently. This means that the model did not consider the possibility that intrinsic and extrinsic parameters could be correlated—for example, the possibility that there might be a correlation between individuals who possess suboptimal computational parameters in their language-learning systems and these individuals' exposure to a poorer language environment. Although no definitive empirical evidence has demonstrated a correlation between, for example, children possessing gene variants that place them at risk of language disorders and being raised in a low SES family, such a correlation is implied now or in the future by the combined findings that (a) language disorders are heritable and (b) as adults, individuals with speech and language disorders have lower SES scores (see, e.g., Bishop, North, & Donlan, 1995; Felsenfeld, Broen, & McGue, 1994). It is logical to presume that the children of such individuals would be more likely to inherit both genes for language disorders and a lower SES environment. Were the model to include a correlation between intrinsic and extrinsic parameters across individuals, this would have a greater impact on resolving delay (where a richer environment can produce good final outcomes) and a lesser impact on persisting delay (where the principal limitation stems from an intrinsic reduction in computational capacity). However, the possibility also exists that parents whose own (heritable) language delay as a child subsequently resolved may not have suffered such an impact on their SES as adults and may offer more enriched language input to their children, compared to parents with persisting language delay. Further evidence is required.

The simulations in this study considered only a single learning system. Modeling should be expanded to consider multiple systems within language, thereby (a) allowing early delay to be diagnosed on different behaviors than those considered later in development and (b) enabling the effects of SES on different levels of language to be explored (see, e.g., Noble, Norman, & Farah, 2006). The impact of variation in input levels may be even more robust for a model of vocabulary development, as suggested by the empirical data considering SES effects on different levels of language development (for a summary, see Hoff, 2013).

Lastly, despite the empirical effects that were successfully simulated by the model, it could, of course, turn out to be wrong: Implemented models serve to demonstrate only the viability of theoretical proposals; they cannot demonstrate the truth of those proposals. It could be that persisting and resolving delay are qualitatively different—or that some instances of persisting delay are qualitatively different. It could also be that a proportion of the cases of resolving delay arise from the way in which early delay is diagnosed, so that early assessments of vocabulary bring together a heterogeneous sample of children, only some of whom will have subsequent deficits in morphosyntax that are characteristic of SLI. Should there be wider (or different) causes of vocabulary delay than grammar delay, and should diagnoses of delay weigh heavily on the ability that is being measured, then the resolution of delay may turn out to be a measurement artifact. In that case, the current approach of tracing developmental trajectories in a single model system would not be appropriate.

Implications

The main findings of the computational modeling work are twofold. First, mechanistically, it is viable that persisting and resolving delay both represent limitations in the processing properties of the language system and are only quantitatively different. Second, there may be an interaction between environmental factors and outcome groups, with individuals with resolving delay being influenced more by the richness of the language environment.

The first finding places the model in accord with theories that have explained language delay by appealing to notions such as slower processing speed (Miller et al., 2001), restrictions in working memory capacity (Ellis Weismer et al., 1999), or reduced efficiency of resource allocation across domains (Im-Bolter et al., 2006). Implementation demonstrated how a number of low-level neurocomputational parameters could lead to delay, sometimes acting in concert. We characterized these as having broad effects on *capacity, plasticity, signal,* and *regressive events*.

The interaction of environmental influence with outcome group accords with Anushko's (2008) finding that in a sample of 230 children whose language trajectories were followed between 15 months and 6 years of age, those children accelerating from the lower performing group to the higher performing group had significantly more exposure to and experience with language through book-reading activities, compared with peers in the low-growth group. The model suggests a clearer framing of the role of environmental input: It is not the cause of early language delay—delay is the result of the intrinsic property of low plasticity; however, in cases where low plasticity is the cause of delay, greater experience with language can maximize subsequent outcomes.

The cause of variation in processing properties could be under genetic or environmental control. A twin study carried out by Bishop et al. (2003) found lower heritability in language scores for cases of resolving delay, higher heritability for language scores for cases of persisting delay in those children whose parents went on to seek professional help, and significant environmental involvement in both cases. This finding is consistent with the greater environmental influence in cases of resolving delay, but it also implies that environmental influences affect processing properties (as observed, for instance, by Hurtado, Marchman, & Fernald, 2008). Environmental variation may also have a different influence depending on the absolute level being considered. Thomas, Forrester, and Ronald (2013) argued that from

Complimentary Author PDF: Not for Broad Dissemination

lower middle class upward, the influence of SES on language development may operate more via the information content of the environment, but as poverty and deprivation increase, SES may operate more via an influence on neurocomputational processing properties. This would explain why Nelson et al. (2011) identified so many cases of clinical language delay in the sample of children in poverty, whereas Zubrick et al. (2007), with a more representative sample, determined that persisting delay was best predicted by family history of late language, male gender, and early neurological growth rather than by environmental factors.

An important clinical priority is to identify markers that can predict the outcome of early-diagnosed language delay. Measures of language processing efficiency hold promise in this regard (see, e.g., Fernald & Marchman, 2012). It remains to be seen whether such early measures can distinguish between processing differences in capacity versus differences in plasticity, which the computational model argues are crucial for predicting long-term outcomes in language development. That is, environmental factors may only impact processing properties at low SES levels. Poor processing properties at higher SES levels may tend to stem from genetic variations.

Acknowledgments

This research was supported by U.K. Economic and Social Research Council (ESRC) Grant RES-062-23-2721, awarded to Wendy Best (University College London), Jackie Masterson (Institute of Education, London), and Michael S. C. Thomas (Birkbeck College London), and by a Leverhulme Study Abroad Fellowship held at the University of Chicago, awarded to Michael S. C. Thomas (Birkbeck College London).

References

Anushko, A. E. (2008). *Multi-domain predictors of trajectories of language development in early childhood* (Unpublished doctoral dissertation). Fordham University, New York.

Anushko, A. E., Jones, S. M., & Carter, A. S. (2009, April). *Multi-domain predictors of trajectories of language development in early childhood*. Poster presented at the meeting of the Society for Research in Child Development, Denver, CO.

Arkkila, E., Rasanen, P., Roine, R. P., & Vilkman, E. (2008). Specific language impairment in childhood is associated with impaired mental health and social well-being in adulthood. *Logopedics Phoniatrics Vocology, 33,* 179–189.

Bavin, E. L., Wilson, P. H., Maruff, P., & Sleeman, F. (2005). Spatio-visual memory of children with specific language impairment: Evidence for generalized processing problems. *International Journal of Language and Communication Disorders, 40,* 319–332.

Beitchman, J. H., Wilson, B., Johnson, C. J., Atkinson, L., Young, A., Adlaf, E., … Douglas, L. (2001). Fourteen-year follow-up of speech/language impaired and control children: Psychiatric outcome. *Journal of the American Academy of Child and Adolescent Psychiatry, 40,* 75–82.

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

Bishop, D. V. M., & Edmundson, A. (1987). Language-impaired four-year-olds: Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders, 52,* 156–173.

Bishop, D. V. M., Laws, G., Adams, C., & Norbury, C. F. (2006). High heritability of speech and language impairments in 6-year-old twins demonstrated using parent and teacher report. *Behavior Genetics, 36,* 173–184.

Bishop, D. V. M., North, T., & Donlan, C. (1995). Genetic basis of specific language impairment: Evidence from a twin study. *Developmental Medicine and Child Neurology, 37,* 56–71.

Bishop, D. V. M., Price, T. S., Dale, P. S., & Plomin, R. (2003). Outcomes of early language delay: II. Etiology of transient and persistent language difficulties. *Journal of Speech, Language, and Hearing Research, 46,* 561–575.

Broomfield, J., & Dodd, B. (2004). Children with speech and language disability: Caseload characteristics. *International Journal of Language and Communication Disorders, 39,* 303–324.

Chiat, S., & Roy, R. (2008). Early phonological and sociocognitive skills as predictors of later language and social communication outcomes. *Journal of Child Psychology and Psychiatry, 49,* 635–645.

Conti-Ramsden, G., St. Clair, M. C., Pickles, A., & Durkin, K. (2012). Developmental trajectories of verbal and nonverbal skills in individuals with a history of specific language impairment: From childhood to adolescence. *Journal of Speech, Language, and Hearing Research, 55,* 1716–1735.

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

Desmarais, C., Sylvestre, A., Francois, M., Bairati, I., & Rouleau, N. (2008). Systematic review of the literature on characteristics of late-talking toddlers. *International Journal of Language and Communication Disorders, 43,* 361–389.

Durkin, K., & Conti-Ramsden, G. (2010). Young people with specific language impairment: A review of social and emotional functioning in adolescence. *Child Language Teaching and Therapy, 26,* 105–121.

Ellis Weismer, S., Evans, J., & Hesketh, L. (1999). An examination of verbal working memory capacity in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 42,* 1249–1260.

Felsenfeld, S., Broen, P. A., & McGue, M. (1994). A 28-year follow-up of adults with a history of moderate phonological disorder: Educational and occupational results. *Journal of Speech and Hearing Research, 37,* 1341–1353.

Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late talking toddlers. *Child Development, 83,* 203–222.

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science, 16,* 234–248.

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

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

Henrichs, J., Rescorla, L., Schenk, J. J., Schmidt, H. G., Jaddoe, V. W. V., Hofman, A., … Tiemeier, H. (2011). Examining continuity of early expressive vocabulary development: The Generation R study. *Journal of Speech, Language, and Hearing Research, 54,* 854–869.

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*

Complimentary Author PDF: Not for Broad Dissemination

Research Forum (pp. 38–49). Palo Alto, CA: Stanford University Press.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development, 74,* 1368–1378.

Hoff, E. (2013). Interpreting the early language trajectories of children from low SES and language minority homes: Implications for closing achievement gaps. *Developmental Psychology, 49,* 4–14.

Hoffman, L. M., & Gillam, R. B. (2004). Verbal and spatial information processing constraints in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 47,* 114–125.

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science, 11,* F31–F39.

Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input at home and at school: Relation to child syntax. *Cognitive Psychology, 45,* 337–374.

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology, 61,* 343–365.

Im-Bolter, N., Johnson, J., & Pascual-Leone, J. (2006). Processing limitations in children with specific language impairment: The role of executive function. *Child Development, 7,* 1822–1841.

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

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, J. R., & Ellis Weismer, S. E. (1983). Mental rotation abilities in language-disordered children. *Journal of Speech and Hearing Research, 26,* 397–403.

Justice, L. M., Bowles, R. P., Pence Tumbell, K. L., & Skibbe, L. E. (2009). School readiness among children with varying histories of language difficulties. *Developmental Psychology, 45,* 460–476.

Karaminis, T. N. (2011). *Connectionist modeling of morphosyntax in typical and atypical development for English and Modern Greek* (Doctoral thesis). University of London.

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, 72*(3). Serial No. 188, 1–144.

Lahey, M., Edwards, J., & Munson, B. (2001). Is processing speed related to severity of language impairment? *Journal of Speech, Language, and Hearing Research, 44,* 1354–1361.

Law, J., Rush, R., Schoon, I., & Parsons, S. (2009). Modeling developmental language difficulties from school entry into adulthood: Literacy, mental health, and employment outcomes. *Journal of Speech, Language, and Hearing Research, 52,* 1401–1416.

Leonard, L. B. (1987). Is specific language impairment a useful construct? In S. Rosenberg (Ed.), *Advances in applied psycholinguistics* (Vol. 1, pp. 1–39). Hillsdale, NJ: Erlbaum.

McDowell, K. D., Lonigan, C. J., & Goldstein, H. (2007). Relations among socioeconomic status, age, and predictors of phonological awareness. *Journal of Speech, Language, and Hearing Research, 50,* 1079–1092.

Miller, C., Kail, R., Leonard, L., & Tomblin, B. (2001). Speed of processing in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 44,* 416–433.

Nelson, K. E., Welsh, J. A., Vance Trup, E. M., & Greenberg, M. T. (2011). Language delays of impoverished preschool children in relation to early academic and emotion recognition skills. *First Language, 31,* 164–194.

Noble, K. G., Norman, M. F., & Farah, M. J. (2006). Neurocognitive correlates of socioeconomic status in kindergarten children. *Developmental Science, 8,* 74–87.

Paul, R. (1996). Clinical implications of the natural history of slow expressive language development. *American Journal of Speech-Language Pathology, 5,* 5–30.

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

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multilayered perceptron: Implications for child language acquisition. *Cognition, 38,* 43–102.

Rannard, A., Lyons, C., & Glenn, S. (2005). Parent concerns and professional responses: The case of specific language impairment. *British Journal of General Practice, 55,* 710–714.

Reilly, S., Wake, M., Ukoumunne, O. C., Bavin, E., Prior, M., Cini, E., … Bretherton, L. (2010). Predicting language outcomes at 4 years of age: Findings from Early Language in Victoria study. *Paediatrics, 126,* e1530–e1537.

Rescorla, L., Dahlsgaard, K., & Roberts, J. (2000). Late-talking toddlers: MLU and IPSyn outcomes at 3;0 and 4;0. *Journal of Child Language, 27,* 643–664.

Rice, M. L. (2009). Language acquisition lessons from children with specific language impairment: Revisiting the discovery of latent structures. In V. C. Gathercole (Ed.), *Routes to language* (pp. 287–313). Hove, East Sussex, United Kingdom: Psychology Press.

Rice, M. L. (2012). Toward epigenetic and gene regulation models of specific language impairment: Looking for links among growth, genes, and impairments [Open Access Version]. *Journal of Neurodevelopmental Disorders, 4,* 1–17. doi:10.1186/1866-1955-4-27

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

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

Rowe, M. L., Raudenbush, S. W., & Goldin-Meadow, S. (2012). The pace of vocabulary growth helps predict later vocabulary skill. *Child Development, 83,* 508–525.

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

Thomas, M. S. C., Annaz, D., Ansari, D., Serif, G., Jarrold, C., & Karmiloff-Smith, A. (2009). Using developmental trajectories to understand developmental disorders. *Journal of Speech, Language, and Hearing Research, 52,* 336–358.

Thomas, M. S. C., Baughman, F. D., Karaminis, T., & Addyman, C. (2012). Modeling development disorders. In C. Marshall (Ed.), *Current issues in developmental disorders* (pp. 93–124). Hove, East Sussex, United Kingdom: Psychology Press.

Thomas, M. S. C., Forrester, N. A., & Ronald, A. (2013). Modeling socioeconomic status effects on language development. *Developmental Psychology, 49,* 2325–2343.

Thomas, M. S. C., & Karmiloff-Smith, A. (2003). Modeling language acquisition in atypical phenotypes. *Psychological Review, 110,* 647–682.

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,* 637–654.

Complimentary Author PDF: Not for Broad Dissemination

Ukoumunne, O. C., Wake, M., Carlin, J., Bavin, E. L., Lum, J., Skeat, J., … Reilly, S. (2011). Profiles of language development in pre-school children: A longitudinal latent class analysis of data from the Early Language in Victoria study. *Child: Care, Health and Development, 38,* 341–349. doi:10.1111/j.1365–2214.2011.01234.x

Whitehurst, G. J., & Fischel, J. E. (1994). Early developmental language delay: What if anything should the clinician do about it? *Journal of Child Psychology and Psychiatry, 35,* 613–648.

Young, A. R., Beitchman, J. H., Johnson, C., Douglas, L., Atkinson, L., Escobar, M., & Wilson, B. (2002). Young adult academic outcomes in a longitudinal sample of early identified language impaired and control children. *Journal of Child Psychology and Psychiatry, 43,* 635–645.

Zubrick, S. R., Taylor, C. L., Rice, M. L., & Slegers, D. W. (2007). Late language emergence at 24 months: An epidemiological study of prevalence, predictors, and covariates. *Journal of Speech, Language, and Hearing Research, 50,* 1562–1592.