UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Simulating individual differences in language ability and genetic differences in FOXP2 using a neural network model of the SRT task

Permalink
https://escholarship.org/uc/item/6xh5g88f

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 32(32)

ISSN
1069-7977

Authors
Toscano, Joseph
Mueller, Kathryn
McMurray, Bob
et al.

Publication Date
2010

Peer reviewed

Simulating individual differences in language ability and genetic differences in *FOXP2* using a neural network model of the SRT task

Joseph C. Toscano (joseph-toscano@uiowa.edu)
Dept. of Psychology, University of Iowa, Iowa City, IA 52242 USA

Kathryn L. Mueller (kathryn-mueller@uiowa.edu)
Dept. of Communication Sciences and Disorders, University of Iowa, Iowa City, IA 52242 USA

Bob McMurray (bob-mcmurray@uiowa.edu)
Dept. of Psychology, University of Iowa, Iowa City, IA 52242 USA

J. Bruce Tomblin (j-tomblin@uiowa.edu)
Dept. of Communication Sciences and Disorders, University of Iowa, Iowa City, IA 52242 USA

Abstract

Recent work has shown that individual differences in language development are related to differences in procedural learning, as measured by the serial reaction time (SRT) task. Performance on this task has also been shown to be associated with common genetic variants in *FOXP2*. To investigate what these differences can tell us about the functional properties of language processing, we present a computational model of the SRT task. We varied parameters in the model to observe their effects on performance in the task. We found that the combined effect of several model parameters produced changes in the learning trajectory that were similar to those observed behaviorally.

Keywords: language processing; specific language impairment; *FOXP2*; procedural learning; serial reaction time task; computational modeling; simple recurrent networks

Introduction

The mechanisms that underlie language use emerge over the course of development through the integration of multiple biological and environmental factors (Elman et al., 1997). Much previous research has focused on whether these mechanisms are language-specific or domain-general (Christiansen & Chater, 2008). Regardless of which is the case, we must specify how different factors interact to give rise to language.

One way to study the mechanisms involved in language is to look at individual differences in language ability. Recently, the use of molecular genetics has emerged as a tool for investigating these differences. However, the use of genetics to study complex cognitive processes, like language, presents a challenge: how do we address questions regarding the role of genes when they are so far removed from language processing? Similarly, how do we assess the role of individual genes when it is unlikely that there is a one-to-one correspondence between genes and specific characteristics of language?

As a first step, we need a way to observe the effects of functional properties of language processing on behavior. Computational models offer a tool for doing this. The units in a neural network model, for instance, correspond to functional (rather than structural) units in the system. Thus, computational models may be useful for examining how genetic factors relate to the functional organization of cognitive systems.

The aim of the current paper is to investigate the relationship between individual differences (both differences in language ability and genetic differences) and functional properties of language processing using a computational model of the serial reaction time (SRT) task. The SRT task measures participants' ability to learn pattern sequences. Variation in performance on the SRT task has been associated with both language ability (Tomblin, Mainela-Arnold, & Zhang, 2007b) and genetic differences (Tomblin, Christiansen, Bjork, Iyengar, & Murray, 2007a). Given this, and the fact that sequence processing is a critical component of language use, this task provides a useful paradigm for studying these relationships.

Individual differences in language abilities

One area in which differences in language ability have been extensively studied is specific language impairment (SLI). SLI is a relatively common developmental disorder characterized by difficulty acquiring language in the absence of gross cognitive or sensory impairments, and despite adequate experience and educational opportunities (Tomblin, Records, & Zhang, 1996). Typically, research criteria for SLI classification require that the individual falls 1.15 SD below the mean on a range of standardized assessments of language while falling in the normal range for non-verbal intelligence (Tomblin et al., 1996).

Children with SLI have deficits in various language abilities, such as mopho-syntactic processing, phonological processing, word learning, and spoken word recognition (Leonard, 1998; McGregor, Newman, Reilly, & Capone, 2002; McMurray, Samelson, Lee, & Tomblin, 2010). In many ways, these children demonstrate language abilities associated with typically developing younger peers. They have smaller vocabularies, use shorter, simpler syntactical constructions, and make more morphological errors than would be expected for children their age (McGregor, Friedman, Reilly, & Newman, 2002).

A range of possible hypotheses for SLI have been proposed, and include deficits in temporal-perceptual processing, generalized slowing, problems with phonological processing, and deficits in working memory (Bishop, North, & Donlan,

1996). Thus, the underlying causes remain unclear, though it is likely that SLI is multiply determined.

Genetic factors and language

Genetics is now commonly employed as a tool for investigating differences in language development. Initial molecular studies centered on the KE family, a multigenerational pedigree that appears to show an autosomal dominant pattern of language impairment (Hurst, Baraitser, Auger, Graham, & Norell, 1990). Affected individuals have been characterized as having apraxia of speech, as well as expressive and receptive language problems (Vargha-Khadem, Watkin, Alcock, Fletcher, & Passingham, 1998). They also have a rare genetic mutation in the *FOXP2* (forkhead box P2) gene (Lai, Fisher, Hurst, Vargha-Khadem, & Monaco, 2001). More recently, Mueller, Bjork, Tomblin, and Murray (in preparation) investigated the role of more common genetic variants in *FOXP2*. These variants were single nucleotide polymorphisms (SNPs), which represent differences in a single base pair in the genome. They examined multiple SNPs in a population with a range of language abilities and found an association between SNPs in the promoter region and language ability as a discrete phenotype. This suggests that these common variants of *FOXP2* also play a role in language development.

FOXP2 is expressed in multiple species as well as several different organs, including the lungs and brain (Shu et al., 2007; Fujita et al., 2008). This has led some to argue that the link between *FOXP2* and language is weak. However, the fact that *FOXP2* is neither species- nor domain-specific means it is likely to play a role in multiple cognitive processes. In addition, since *FOXP2* is a transcription factor (i.e., encodes a regulatory protein that affects gene expression), it is possible to identify other elements of the gene pathway (and therefore the systems) in which it exists (Vernes et al., 2008).

A more general role for *FOXP2* fits with the hypothesis that language itself is shaped by domain-general processes (Christiansen & Chater, 2008). Statistical learning plays an important role in language acquisition (Saffran, Aslin, & Newport, 1996), and it is closely related to procedural learning (Perruchet & Pacton, 2006). *FOXP2* remains a candidate gene involved in language because of its association with procedural learning and the basal ganglia (Enard et al., 2009).

Procedural learning and the SRT task

Given the links between language ability, *FOXP2*, and procedural learning, researchers have examined sequence learning to better understand these relationships and mechanisms associated with language. The SRT task is a sequence learning task designed to measure participants' ability to implicitly learn sequences. Participants are presented with blocks of trials that are either random or repeat in a particular sequence. As sequence processing is fundamental to language and statistical learning provides a useful mechanism for learning language (Saffran et al., 1996), this task allows us to measure some of the key functional properties of language.
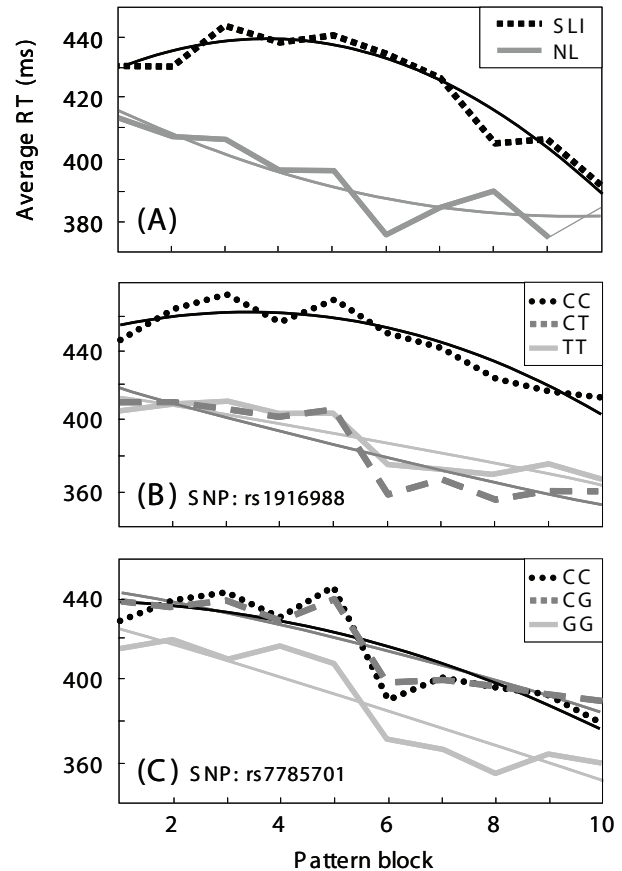

Figure 1: Behavioral data for pattern trials in the SRT task. (A) Data from Tomblin et al. (2007b) comparing SLI and NL groups. (B & C) Data from Tomblin et al. (2007a) for individuals with different genotypes of SNPs rs1916988 and rs7785701.

Tomblin et al. (2007b) used an SRT task to examine differences between children with normal language (NL) and children with SLI. In their task, participants were shown four boxes on a computer screen. On each trial, a picture of a cartoon creature appeared in one of the boxes, and the participant's task was to choose the box containing the picture as quickly as possible.

For the first 100 trials, stimuli were presented randomly. Then, 200 trials were presented in which the sequence [1, 3, 2, 4, 4, 2, 3, 4, 4, 2, 4] was repeated (pattern trials). Finally, 100 additional random trials were presented. Participants were not informed which trials were random and which were pattern trials during the course of the experiment. The experiment was divided into blocks of 20 trials each for data analysis (blocks 1-5 were the first set of random trials, blocks 6-15 were pattern trials, and 17-20 were random trials).

Tomblin et al. found that, overall, the SLI group had longer RTs than the NL group (Fig. 1A). During the pattern trials, performance of both groups improved, indicating that they learned something about the sequence. However, the learn-

ing trajectory differed for the two groups. For the NL group, RT decreased rapidly after the first few blocks of pattern trials and then leveled off. In contrast, for the SLI children, RT remained flat (or increased slightly) during the first few pattern blocks before decreasing. The difference between these two learning trajectories can be approximated by a quadratic function (small differences between the two groups at the first and last blocks; large differences in the middle blocks).

In another study, Tomblin et al. (2007a) examined the relationship between multiple SNPs and performance in this SRT task. They found that the CC genotype of SNP rs1916988 (Fig. 1B) and the CC genotype of SNP rs7785701 (Fig. 1C) were associated with slower RTs over the course of the pattern trials. The CC genotype of SNP rs1916988) was also associated with a learning curve that was similar to the SLI children.

These results suggest that both language impairment and genetic variation in *FOXP2* have similar effects on performance in the SRT task. Given previous work showing a link between *FOXP2* and language, these effects may be related to common functional differences evident in language impairment and some variants of *FOXP2*.

Computational model

We used a neural network to examine whether some of the functional properties of procedural learning are related to the differences observed with human participants. In particular, we would like to capture the difference in the shape of the learning trajectory observed between some of the fast RT groups (children with NL [Fig. 1A] and the CC and CT genotypes of SNP rs1916988 [Fig. 1B]) and slow RT groups (SLI children and the CC genotype of that SNP). By exploring the parameter space of the model, we can determine which functional properties are associated with these differences in the learning trajectories.

Model architecture

The model is a simple recurrent network (SRN; Elman, 1990; c.f. Misyak, Christiansen, & Tomblin, 2009, for an adaptation to the SRT task). The network has three layers: an input layer, an output layer, and a hidden layer with recurrent connections. The input and output layers each have four units (corresponding to the four possible stimulus locations). The hidden layer's recurrent connections provide it with information about the state of the hidden units on the previous trial (context units). This allows the network to learn sequences, like those in the pattern trials of the SRT task. Connection weights are updated using backpropagation (Rumelhart, Hinton, & Williams, 1986). Logistic activation functions are used for the hidden and output units.

Simulation procedure

The network was trained on a task based on the one used by Tomblin et al. (2007b). On each trial, a stimulus was presented to the network by activating a particular input unit and setting the rest to zero, and activation flowed to the output

units. Luce choice ratios were computed by dividing each output unit's activation by the total activation. These values were then used to compute an RT for the network according to the equation

$$RT = \frac{1}{C - \frac{\Sigma I}{n-1}} \qquad (1)$$

where C is the activation of the correct output unit, I is the activation of each of the three other output units, and n is the number of output units (four for these simulations). This gives an estimate that is analogous to RT; a lower value corresponds to a lower RT in the SRT task. Thus, when one unit is significantly more active than all the others (i.e., the network is confident in a single response) the RT will be low. When all the units are similarly active (the network is unsure what the response is) the RT will be high.

The correct unit on each trial is the output unit that corresponds to the one that was activated at the input layer. This corresponds to the SRT task in which participants respond by selecting the location containing the stimulus.

For the first 100 trials, a random location was chosen and presented as input. Then, for 200 trials, the sequence [1, 3, 2, 4, 4, 2, 3, 4, 4, 2, 4] was repeated. Finally, an additional 100 random trials were presented. Only trials on which the correct output unit had the highest activation were included in the analysis. The entire simulation run was divided into 20 blocks of 20 trials.

Simulation 1

In the first simulation, we varied several parameters individually to gauge their effect on performance in the SRT task: *context strength*, *input strength*, *learning rate*, *number of hidden units*, and *temperature*.

Context strength determines the strength of the connections from the hidden to context units (i.e., hidden unit activations are multiplied by this amount when setting context unit activations). A lower context strength may have an effect on the network's ability to learn sequences, which could influence learning in the SRT task.

Input strength controls the fidelity of the stimulus presented to the network. The input unit corresponding to the chosen location is set to the value of the input strength and the others are set to zero. A lower input strength makes the stimulus location less distinct from the others.

Learning rate is the value that the weight change term is multiplied by each time the weights are updated. Models with lower learning rates require more trials to learn the task, but may have more stable learning. This could affect the network's ability to learn over the course of the pattern trials.

Number of hidden units affects the amount of information the network can hold about the sequence. If the network has too few, its ability to encode the sequence will be impaired.

Temperature corresponds to the temperature parameter of the logistic activation function. This activation function constrains the hidden and output units to have activations between zero and one. A higher temperature makes the logistic

more nonlinear. Thus, if the correct output unit has the highest activation, a high temperature parameter will make this value more distinct from the values of the incorrect units, resulting in a lower model RT. The temperature parameters for the hidden and output units were varied separately.

Five hundred repetitions of each condition were run.

Results

The network was able to learn the SRT task and showed an overall learning trajectory similar to the ones observed in the behavioral data. The network's performance improved over the course of the simulation and was faster during the pattern trials than the random trials.

Fig. 2 shows the performance of the model on the SRT task for different values of each parameter. A range of values for the parameters were tested to find a set that produced responses similar to those observed for the fast RT groups in the behavioral data. Each parameter was then varied individually, holding the others constant at those values. For example, in Fig. 2A, *context strength* was varied. The other parameters were held constant for both *context strength* conditions at the baseline values (i.e., *learning rate* = 0.10, *hidden units* = 12, *input strength* = 1.0, *hidden unit temperature* = 1.0, *output unit temperature* = 1.0).

Context strength (Fig. 2A) had very little effect on on the network's RT. This suggests that the network can still perform the task with limited information from the previous trial.

Input strength (Fig. 2B) had an effect on overall RT and an effect on the shape of the learning trajectory. Models with a lower input strength showed a small increase in RT at the beginning of the pattern trials, but this did not persist to the middle blocks.

Learning rate (Fig. 2C) also had an effect on the shape of the learning trajectory. This was due to the fact that the network initially shows an increase in RT at the beginning of training. By decreasing the learning rate, this increase was pushed forward in time into the pattern trials. Thus, one reason that some groups show an increase during the pattern trials in the SRT task might be that they are still in this initial learning phase.

Number of hidden units (Fig 2D) had an effect similar to *input strength*. Fewer hidden units resulted in longer overall RTs and a small increase at the beginning of the pattern trials.

Temperature (Figs. 2E & 2F) had an effect on the overall RT at the beginning of the pattern trials, but did not capture the change in the shape of the learning curve.

Discussion

Several parameters produced changes in the network's median RT and learning trajectory. Changes in *input strength*, *learning rate*, and *number of hidden units* can account for some of the changes in the shape of the learning trajectory observed behaviorally. As discussed above, however, specific SNPs and individual differences in language ability are likely to have multiple functional effects. Thus, we may find a better
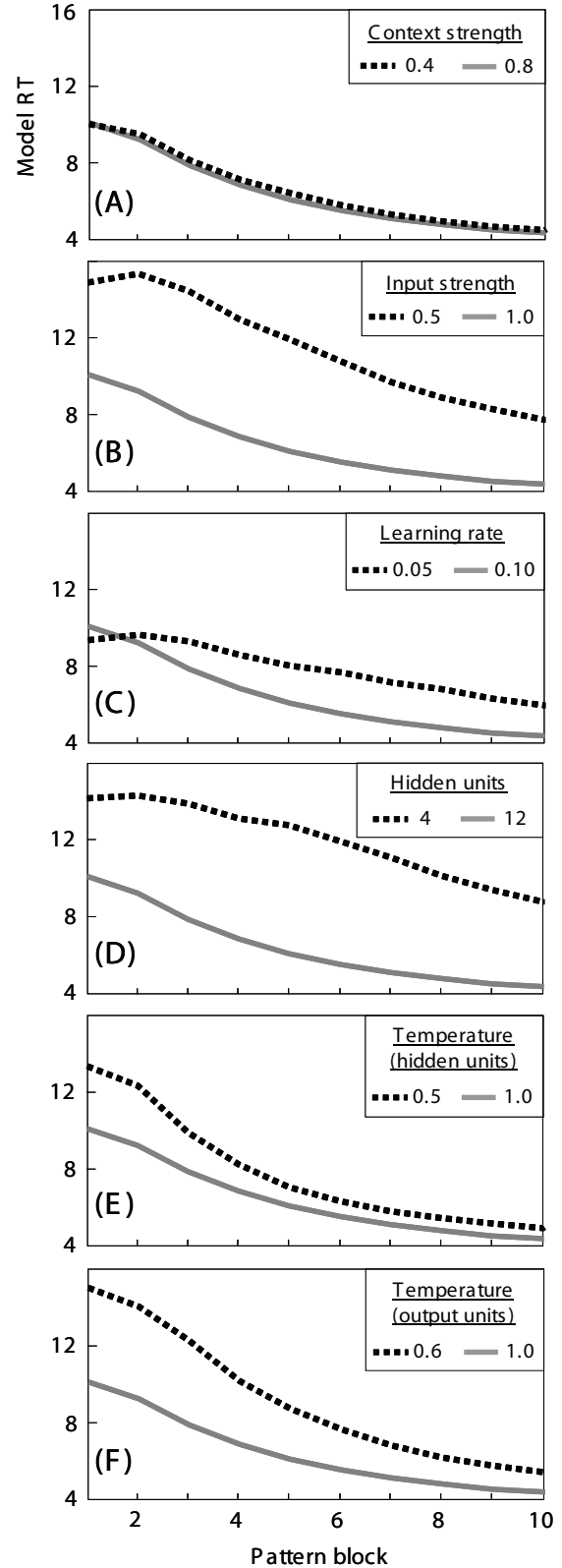

Figure 2: Results of Simulation 1. For each simulation, the set of parameters producing effects similar to those seen in the fast RT groups was used as a baseline (solid lines in figures), and individual parameters were varied (dashed lines).

fit to the behavioral data by examining the combined effects of multiple parameters. This was done in Simulation 2.

Simulation 2

In the second set of simulations, we varied multiple parameters in the model simultaneously, allowing us to explore the parameter space of the network further. Five values were tested for the number of hidden units, and four were tested for each other parameter, yielding a total of 5,120 combinations. The simulation procedure was the same as Simulation 1, except that 50 repetitions of each combination were run.

Results

In order to determine which parameter sets reflected the fast and slow RT groups in the behavioral data, pairwise comparisons were made and the difference scores were fit to quadratic functions (the pattern of the differences in the learning trajectories). Thus, for each comparison there was a set of parameters corresponding to the slow RT groups and a set corresponding to the fast RT groups.

Several pre-processing criteria were used to exclude sets that did not show correct performance on the SRT task (i.e., better performance over the course of the pattern trials) and comparisons that would not yield a pattern consistent with the difference between groups in the behavioral data (i.e., quadratic). The remaining pairs were then fit to quadratic functions using the least squares method, and R^2 was used to determine the goodness of fit.

R^2 values greater than 0.9 were found for 0.47% of the pairs. To determine which parameters drove the effect, we computed the mean parameter values for the slow and fast RT groups for these pairs. The mean values for each parameter for the two groups are shown in Table 1. Some parameters did not differ between the groups, whereas others differed greatly. We found that the parameters in Simulation 1 that produced changes in the learning trajectory (*learning rate*, *number of hidden units*, and *input strength*) had similar effects when varied in conjunction with *temperature*. Fig. 3A shows the responses of the model when these parameters are varied simultaneously.

Adjusting the parameters by hand allowed us to distill the set of parameters down to two, *learning rate* and *temperature*, that accounted for the difference in learning trajectories for the first half of the pattern trials, but not the second half (the
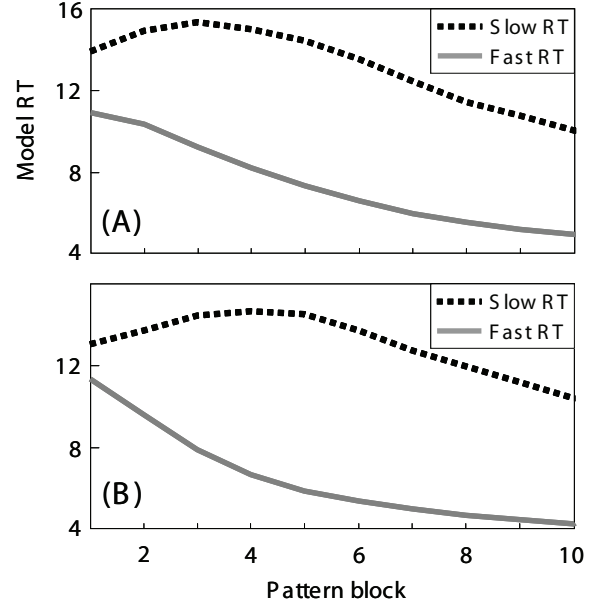
Table 1: Simulation 2 results.

Parameter	Slow RT	Fast RT
Context strength	0.48	0.53
Input strength	0.83	0.90
Hidden units	5.9	9.5
Learning rate	0.13	0.18
Temperature (hidden)	0.39	0.52
Temperature (output)	0.38	0.44

Figure 3: Results of Simulation 2. (**A**) Responses of the model when *learning rate*, *input strength*, *number of hidden units*, and *temperature* are varied simultaneously. (**B**) Responses when *learning rate* and *temperature* are varied simultaneously. Five hundred repetitions of each condition were run to produce the figures.

slow RT model did not reach the same RT by the end of the pattern trials). Fig. 3B shows the responses of the model when these parameters are varied together.

Discussion

The results of this simulation show that the combined effects of several parameters together can better approximate the difference in learning trajectories. This suggests that this approach can be used to determine which combinations of parameters mirror the behavioral data. Additional exploration of the parameter space (i.e., testing a larger range of values) may allow us to find a better fit.

General discussion

The results of these simulations suggest that several functional aspects of sequence processing contribute to the differences in SRT performance observed behaviorally and that by examining multiple factors at the same time, we can get a better estimate of the effects of language impairment and genetic variation. This fits with the notion that genetic differences are likely to have multiple functional consequences.

Recently, McMurray et al. (2010) used a similar approach to determine which parameters in TRACE (McClelland & Elman, 1986) corresponded to differences between NL and SLI children in a spoken word recognition task. They found that variation in the network's decay parameter produced differences similar to those between the SLI and NL groups. This parameter is related to competition. In the SRN used here,

the *temperature* parameter corresponds to competition (e.g., a lower temperature parameter for the output unit activation function leads to greater activation for the competitor units). Thus, these two sets of of simulations, modeling different tasks with different networks, provide converging evidence that competition between internal representations may be a critical mechanism in language processing that produces differences between NL and SLI children.

The simulations presented here provide a first step towards assessing the role of genetic variation and language ability in procedural learning, and they suggest several functional properties that may be influenced by these differences. More broadly, they show that exploring the parameter space of a computational model may offer an approach to studying the effects of genetic factors on cognitive systems.

Acknowledgments

References

Bishop, D. V. M., North, T., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: evidence from a twin study. *J Child Psychol Psychiatry*, *37*, 391-403.

Christiansen, M. H., & Chater, N. (2008). Langauge as shaped by the brain. *Behav Brain Sci*, *31*, 489-509.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179-211.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1997). *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.

Enard, W., Gehre, S., Hammerschmidt, K., Holter, S. M., Blass, T., Somel, M., et al. (2009). A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. *Cell*, *137*, 961-971.

Fujita, E., Tanabe, Y., Shiota, A., Ueda, M., Suwa, K., Momoi, M. Y., et al. (2008). Ultrasonic vocalization impairment of Foxp2 (R552H) knockin mice related to speech-language disorder and abnormality of Purkinje cells. *Proc Nat Acad Sci*, *105*, 3117-3122.

Hurst, J. A., Baraitser, M., Auger, E., Graham, F., & Norell, S. (1990). An extended family with a dominantly inherited speech disorder. *Dev Med Child Neurol*, *32*, 352-355.

Lai, C. S. L., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., & Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, *413*, 519-523.

Leonard, L. B. (1998). *Children with specific language impairment*. Cambridge, MA: MIT Press.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychol*, *18*, 1-86.

McGregor, K. K., Friedman, R. M., Reilly, R. M., & New-man, R. M. (2002). Semantic representation and naming in young children. *J Speech Lang Hear Res*, *45*, 332-346.

McGregor, K. K., Newman, R. M., Reilly, R. M., & Capone, N. C. (2002). Semantic representation and naming in children with specific language impairment. *J Speech Lang Hear Res*, *45*, 998-1014.

McMurray, B., Samelson, V. M., Lee, S. H., & Tomblin, J. B. (2010). Eye-movements reveal the time-course of online spoken word recognition language impaired and normal adolescents. *Cognitive Psychol*, *60*, 1-39.

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2009). Statistical learning of nonadjacencies predicts on-line processing of long-distance dependencies in natural language. In N. Taatgen, H. van Rijn, J. Nerbonne, & L. Schomaker (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (p. 177-182). Austin, TX: Cognitive Science Society.

Mueller, K. L., Bjork, J. B., Tomblin, J. B., & Murray, J. C. (in preperation). *Common genetic variants in FOXP2 are associated with individual differences in language development*.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends Cognitive Sci*, *10*, 233-238.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propogating errors. *Nature*, *323*, 533-536.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926-1928.

Shu, W., Lu, M. M., Zhang, Y., Tucker, P. W., Zhou, D., & Morrisey, E. E. (2007). Foxp2 and Foxp1 cooperatively regulate lung and esophagus development. *Development*, *134*, 1991-2000.

Tomblin, J. B., Christiansen, M. H., Bjork, J. B., Iyengar, S. K., & Murray, J. M. (2007a). Association of FOXP2 genetic markers with procedural learning and language. Poster presented at the Annual Meeting of the American Society of Human Genetics.

Tomblin, J. B., Mainela-Arnold, E., & Zhang, X. (2007b). Procedural learning in adolescents with and without specific language impairment. *Lang Learn Dev*, *3*, 269-293.

Tomblin, J. B., Records, N. L., & Zhang, X. (1996). A system for the diagnosis of specific language impairment in kindergarten children. *J Speech Hear Res*, *39*, 1284-1294.

Vargha-Khadem, F., Watkin, K. E., Alcock, K., Fletcher, P., & Passingham, R. (1998). Praxic and nonverbal cognition deficits in a large family with a genetically transmitted speech and language disorder. *Proc Nat Acad Sci*, *92*, 930-933.

Vernes, S. C., Newbury, D. F., Abrahams, B. S., Winchester, L., Nicod, J., Groszer, M., et al. (2008). A functional link between distinct developmental language disorders. *N Eng J Med*, *359*, 2337-2345.