Simulating behavioural interventions for developmental deficits: When improving strengths produces better outcomes than remediating weaknesses

Hala Alireza (halalireza@yahoo.co.uk)
Developmental Neurocognition Lab, Birkbeck, University of London
Malet Street, London WC1E 7HX, UK

Anna Fedor (fedoranna@gmail.com)
2MTA-ELTE Theoretical Biology and Evolutionary Ecology Research Group
Budapest, H-1117, Hungary

Michael S. C. Thomas (m.thomas@bbk.ac.uk)
Developmental Neurocognition Lab, Birkbeck, University of London
Malet Street, London WC1E 7HX, UK

Abstract

Computational models of cognitive development have been frequently used to model impairments found in developmental disorders but relatively rarely to simulate behavioural interventions to remediate these impairments. One area of controversy in practices of intervention is whether it is better to attempt to remediate an area of weakness or to build on the child's strengths. We present an artificial neural network model of productive vocabulary development simulating children with word-finding difficulties. We contrast an intervention to remediate weakness (additional practice on naming) with interventions to improve strengths (improving phonological and semantic knowledge). Remediating weakness served to propel the system more quickly along the same atypical trajectory, while improving strengths produced long-term increases in final vocabulary size. A combination yielded the best outcome. The model represents the first mechanistic demonstration of how interventions targeting strengths may serve to improve behavioural outcomes in developmental disorders. The observed effects in the model are in line with those observed empirically for children with word-finding difficulties.

Keywords: artificial neural networks; developmental disorders; intervention; vocabulary development; word-finding difficulties

Introduction

Theories of deficits versus theories of intervention

In the field of developmental disorders, there are extensive theories about the causes of behavioural deficits. However, these have played a relatively small role in intervention practices. Indeed, theories of treatment have often developed relatively independently of theories of deficit (Laws et al., 2008; Michie & Prestwich, 2010). The gap between a mechanistic understanding of the causes of deficits and everyday therapeutic practice exists for a number of reasons. Most obviously, the primary focus of intervention is on behavioural outcomes, which do not in themselves necessitate an understanding of underlying cause. In addition, there are diverse real-world constraints influencing the interventions that are selected. And it is difficult to apply causal principles to the complex therapeutic situation involving treatment of the whole child via a social interaction with the therapist. Nevertheless, it remains an important ambition to narrow the gap between theories of deficit and practices of intervention.

Improve strengths or remediate weaknesses?

One area of controversy in practices of intervention is whether it is better to attempt to remediate an area of weakness or to build on the child's strengths. For example, in the field of developmental language disorders, Leonard (2014) argues that generally, therapists prefer to work on developing compensatory strategies through targeting the child's strengths rather than trying to improve his or her area of weakness (see also Bishop, Nation & Patterson, 2014).

To take a more specific case, where children have difficulties in producing words that they already know (so called word-finding difficulties; WFD), therapists may simply require the child to spend more time practicing naming, the area of weakness. Alternatively, they have the option of targeting children's knowledge of word sounds (phonology) or word meanings (semantics). Therapists have found that interventions that elaborate the semantic aspect of words (e.g., McGregor & Leonard, 1989) or interventions that focus on the phonological component of word finding (e.g., Best, 2005) both alleviate WFD to some extent. In a survey, Best (2003) asked therapists what kind of difficulties they found most often co-occurring with WFD in the children they saw. Phonological problems were reported to co-occur 46% of the time, while semantic problems co-occurred only 13% of the time. Nevertheless, when asked which types of intervention they most often used, therapists reported more often using techniques to improve semantic knowledge than phonology (79% of the time compared to 54%). In this case, then, therapists frequently sought to buttress areas of strength to improve naming skills.

One explanation for the tendency of therapists to work less on areas of weakness and more often on areas of strength is to improve the child's confidence in a domain where he or she is struggling. However, from a theoretical perspective, one might ask through what mechanisms could improving a strength serve to remediate a behavioural impairment in a developing cognitive system?

There are at least three ways that improving a strength could remediate a behavioural impairment. First, the 'strength' could represent an alternative cognitive system or pathway to deliver a similar behavioural outcome. Improving a strength then translates to encouraging a compensatory strategy. Second, the target behavior may be delivered by an interactive system in which multiple sources of knowledge combine to drive behavior. Stronger input from one source might then make up for weaker input from another. Third, the target behavior may require mappings to be learned between representations. Improving the structure of the representations might serve to make learning those mappings easier. In this article, we use computational modeling to investigate the third of these options.

Computational modeling of interventions

Computational modeling, particularly the use of artificial neural networks, has been extensively applied to understanding the mechanisms underlying developmental deficits, in disorders such as dyslexia, Specific Language Impairment, and autism (Thomas & Karmiloff-Smith, 2002). Relatively few models of developmental deficits have been extended to the simulation of behavioural interventions to remediate these deficits, and the framework for doing so has only recently been laid out (Thomas et al., 2017). Two notable exceptions are Harm, McCandliss and Seidenberg's (2003) simulation of an intervention for dyslexia, and Best et al.'s (2015) model of interventions for productive vocabulary deficits. In both cases, a typical model established the developmental trajectory under normal circumstances; an atypical model was created in which a computational constraint limited development; and a behavioural intervention was simulated by adding further input-output mappings to the model's training set for a discrete period, usually relatively early during training. Here, we adapt Best et al.'s model of vocabulary development to contrast the effects of improving strengths versus remediating weaknesses.

Connectionist models of vocabulary development

Sentence production involves a sequence beginning with a planned message, followed by selection of major lexical concepts, assigning syntactic functions, assembling phonologically realized words and morphemes into a sentence frame, and programming articulatory processes (e.g., Bock & Levelt, 1994). Connectionist models of word production have tended to focus on the step involving the retrieval of phonological forms given a semantic specification of the desired lexical item (e.g., Dell, Schwartz, Martin, Saffran, & Gagnon, 1997). Developmental models have simulated the learning of mappings between pre-specified semantic and phonological codes (e.g., Plunkett et al., 1992), or between semantic and phonological representations emerging in self-organizing maps (e.g., Li, Zhao & MacWhinney, 2007; Mayor & Plunkett, 2010). Best et al. (2015) used a similar approach, but implemented the semantic and phonological components of the model via 3-layer autoassociator networks trained using backpropagation (Rumelhart, Hinton & Williams, 1986). The hidden unit representations of the semantic component were then mapped to the hidden units of the phonological component, via an intermediate layer of hidden units, to provide a pathway for the development of naming. A reverse pathway simulated comprehension. Naming behavior began to emerge while the semantic and phonological representations were themselves still developing. By restricting learning in the semantic component, the phonological component, or the pathway between them (for example, by reducing hidden unit numbers or the learning rate), Best et al. were able to capture various patterns of atypical naming development observed in a sample of children with WFD.

Simulations

Simulation design

Using the same architecture, we followed the Best et al. (2015) model in simulating productive vocabulary development in children with WFD by restricting the computational capacity of the pathway mapping between semantic and phonological representations. The semantic and phonological components themselves developed typically. Early in development – when slow vocabulary growth was already detectable – a behavioural intervention was applied for a limited period during training. Five different interventions were contrasted, of three types: (1) remediating the weakness – the model was provided with additional training on the naming pathway; (2) improve the strength – the model was provided with additional training to improve the semantic representations, the phonological representations, or both at once; (3) both types 1 and 2 were combined into an intervention that sought to simultaneously improve strength and remediate weakness. We observed the immediate effect of intervention, in terms of potentially accelerated vocabulary growth, and the eventual outcome, in terms of the largest vocabulary size achieved following each type of intervention.

The original Best et al. (2015) model used a fairly abstract rendition of semantics and phonology and a training set of only 100 items. Here, we used more realistic semantic and phonological representations, and scaled up the training set slightly to around 400 items. The typically developing model was designed in such a way that it reflected salient properties of vocabulary growth, including a comprehension-production asymmetry (Bloom, 1973) and a vocabulary explosion / exponential growth in vocabulary size (e.g., McMurray, 2014).

The model contained three assumptions not in the Best et al. (2015) model. First, phonological representations were required to be more accurate than semantic outputs to drive a behavioural response, under the assumption that phonological output needs to drive motor assemblies, while semantic comprehension only requires that the output fall in the correct attractor basin (Hinton & Shallice, 1991). This assumption generated the production-comprehension asymmetry.

Second, we implemented a sensitive period in the development of the components but not the pathways in the model, through pruning of network connectivity after a given point in development. This created the potential for early training to create enriched lower level representations by utilizing the then-available rich connectivity. Pathways did not experience this pruning, under the view that sensitive periods are characteristic of lower but not higher cognitive systems (Takesian & Hensch, 2013). The effect of timing of intervention was subsequently evaluated.

Third, plasticity was set higher in the pathways than the components (via the learning rate parameter), so that the development of semantic and phonological representations would be the limiting factor on the development of naming. If the semantic and phonological representations were to quickly reach ceiling before naming had developed, interventions targeting phonology and semantics would have no scope to improve naming performance. The effects of both the second and third assumptions were evaluated by also running the model in their absence.

Finally, we explored whether the five types of intervention would enhance performance in a typically developing model, or whether they only had the potential to improve performance in systems exhibiting delayed development.

Simulation details

Architecture: The architecture of the vocabulary development model is shown in Figure 1. It comprised four linked backpropagation networks. The semantics component comprised a 3-layer autoassociator with 1029 input and output units and 45 hidden units. The phonology component was an autoassociator with 456 input and output units and 60 hidden units. The naming pathway linked the semantic hidden units with the phonological hidden units via an intermediate layer of 175 units. Naming constituted activating semantic inputs and measuring phonological outputs. The comprehension pathway ran in the other direction and also contained 175 units. In the atypical model, the number of hidden units in the naming pathway was reduced to 90 prior to training.

Additional parameters: The learning rate in the semantic component was .015 and in the phonological component was .025. In the pathways, the learning rate was .15. Sigmoid activation functions had a temperature of 1.5 in the components and 1 in the pathways. In the components, after epoch 75, any connection weights with an absolute magnitude of less than .5 had a 5% chance of being permanently removed each epoch. Initial weights were given random values via a Gaussian distribution with mean 0 and standard deviation 0.5. Gaussian noise with a standard deviation of .15 was added to the net input of units in the components, and noise with a standard deviation of .05 in the pathways, to provide a stochastic basis for naming errors in normal functioning. Continuous activation values on the phonological output were converted to responses by finding the nearest legal phoneme in each slot and assessing whether the full phoneme string was the correct name. If the average root mean square error between the activation vector for each phoneme and the nearest legal phoneme code exceeded 0.03, that phoneme was coded as no response. A nearest neighbor technique was also used to assess the accuracy of semantic outputs. These parameters were selected to calibrate the typical model.

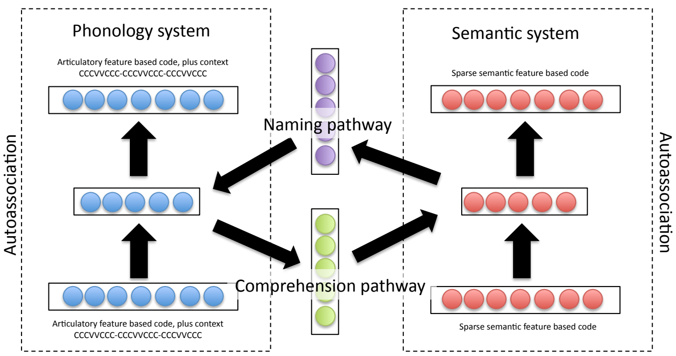

Figure 1: Architecture of the productive vocabulary model, with phonological and semantic components and linking pathways

Training set: The training set comprised 397 words, each with a phonological and a semantic representation. It was generated by combining two sources, a set of 1029 speaker generated semantic feature norms for 456 words collected by Vinson and Vigliocco (2008) from 280 adults; and the Children's Printed Word Database (Masterson et al., 2010), which is an online database of the vocabulary in reading materials used by 5-9 year old children in the UK. The 397 words represent those present in both resources. The semantic representations comprised the 1029 feature set, where a feature was set to 1 if any adult rated it as a characteristic of a given word meaning, and 0 otherwise. The phonological representations used a 19-bit articulatory code for phonemes (Thomas & Karmiloff-Smith, 2003) and a left-justified slot-based CCCVVCCC syllabic scheme to capture words up to three syllables in length, with 3x8x19 = 456 phonological features in total.

Training schedule: Networks were trained for 1000 epochs, with random presentation and pattern update. Training of autoassociators and pathways was interleaved. Weights were updated using the backpropagation algorithm with the cross entropy error measure.

Simulation of interventions: For atypical networks, an intervention began at 100 epochs of training and lasted for 100 epochs. For the main condition, at this point TD models had acquired a productive vocabulary size of 67 words, while atypical models had a vocabulary size of 36 words.

For the intervention, one or more components or pathways were trained with 5 times the frequency of the rest of the system. The extra training could be on the semantics-phonology naming pathway, the semantics component alone, the phonological component alone, both semantics and phonological components, or all of these combined.

Conditions: To test the importance of timing, interventions were compared at 100, 250 and 750 epochs. To test the effect of plasticity assumptions in the model, the first variant removed connectivity pruning from the components. The second variant removed the higher plasticity of the pathways, setting their value to .025.

Replications: All conditions were replicated three times with different random seeds. The full design took approximately 100 days of simulation time. Results graphs are shown averaged over replications; individual data are included in the following tables.

Results

Figure 2 displays developmental trajectories for naming and comprehension in the typical and atypical models. For the typical model, naming lagged behind comprehension, exhibiting the expected comprehension-production asymmetry. Naming itself showed an accelerating rate of development, consistent with a vocabulary explosion. The atypical model with WFD exhibited delayed development in naming but not comprehension (the slightly better comprehension was just a chance difference).

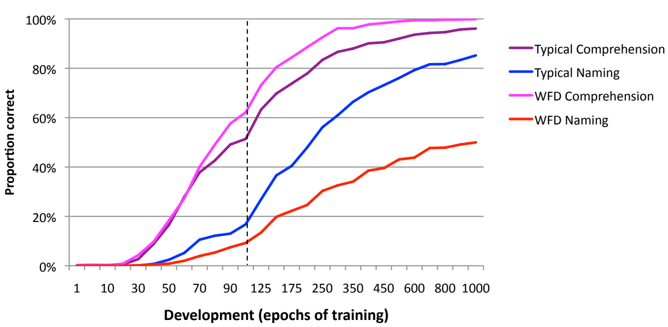

Figure 2: Development of naming and comprehension in the typical and atypical models. The dotting line depicts the point at which the intervention was applied

Figure 3 shows the effects of the intervention to remediate weakness, with extra training on the naming pathway, compared to the impaired model without treatment. The intervention produced accelerated development, but there was no gain in final productive vocabulary level. Figure 4 shows the result of improving strengths – extra training on the otherwise typically developing semantic and phonological representations, which are respectively the inputs and outputs of the naming pathway. These interventions produced slower effects during the intervention itself, but led to long-term (if relatively modest) increases in final productive vocabulary levels. Figure 3 also contains the combined strength-and-weakness intervention. The combined intervention showed the initial immediate gains of the remediation intervention as well as the long-term elevated final level of the strengths intervention. Figure 5 includes the effects of these interventions on typically developing models. Extra practice in naming accelerated development but did not raise the final level. Extra elaboration of semantic and phonological representations by contrast increased the final productive vocabulary size even for the typically developing networks.

Tables 1 and 2 show the final level performance, split by replication, contrasting the intervention targeting weakness (Naming), the intervention targeting strengths (S+P), and the intervention targeting both strengths and weaknesses (Both). Table 1(a) contains the data for the above base condition; 1(b) demonstrates that when plasticity reductions in the components were removed as a model assumption, the same pattern of results held. Table 1(c) shows that without the assumption of greater plasticity in the pathways, the same pattern also held. Table 2 contrasts the effect of interventions at different points in training. The strengths intervention (S+P) diminished in size and disappeared the later in development it was applied. The weakness intervention (Naming) showed the opposite pattern, increasing in size the later it was applied. The combined (Both) showed a uniform effect across development. Within each condition, the three replications demonstrated a common profile.

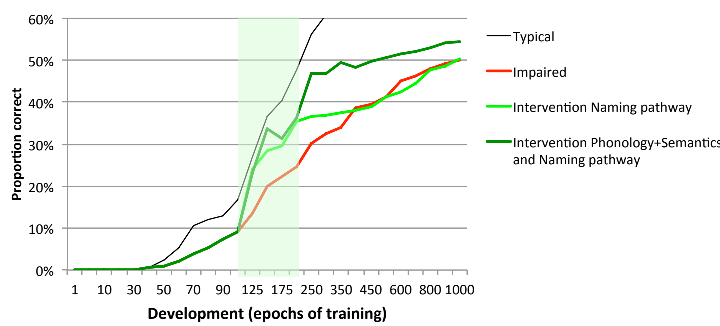

Figure 3: The effect of interventions to remediate weakness on naming accuracy, as well as the combined intervention. Shaded region = period of intervention

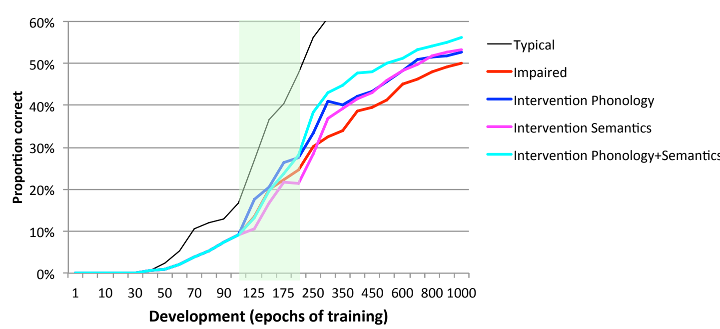

Figure 4: The effect of interventions to improve strengths on naming accuracy. Shaded region = period of intervention

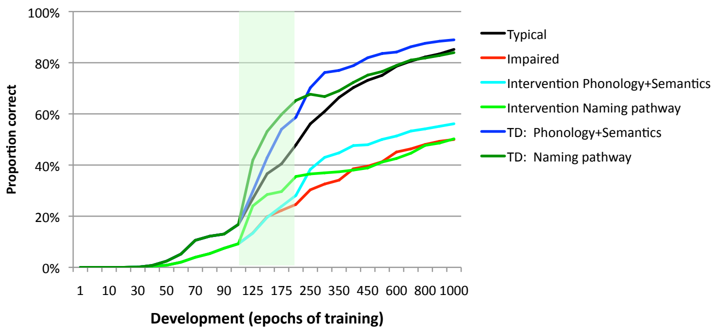

Figure 5: The effect of interventions on naming accuracy for typically developing (TD) and impaired networks. Shaded region = period of intervention

Table 1: Naming accuracy at the end of training for typical (TYP), atypical (ATYP), and atypical intervened networks: (a) the base condition; (b) removing plasticity reduction in the semantic and phonological components; (c) removing greater plasticity in pathways. S+P = strengths intervention. Naming = weakness intervention. Both = combined. Three replications and average are shown

(a) Naming accuracy at the end of training

	TYP	ATYP	Intervention		
			S+P	Naming	Both
R1	84%	49%	57%	50%	52%
R2	86%	52%	57%	50%	57%
R3	86%	49%	54%	51%	55%
Avg	85%	50%	56%	50%	55%

(b) Without plasticity reduction in S and P components

	TYP	ATYP	Intervention		
			S+P	Naming	Both
R1	94%	49%	54%	48%	51%
R2	95%	52%	55%	51%	53%
R3	95%	51%	55%	50%	53%
Avg	95%	51%	55%	50%	53%

(c) Equalized plasticity in pathways and components

	TYP	ATYP	Intervention		
			S+P	Naming	Both
R1	84%	55%	61%	54%	60%
R2	81%	55%	64%	53%	60%
R3	82%	54%	60%	50%	59%
Avg	82%	55%	61%	52%	60%

Discussion

We used an artificial neural network model of impaired vocabulary development to explore the relative merits of a behavioural intervention to remediate weakness versus one to improve strengths. The two interventions yielded contrasting patterns. The intervention to remediate the weakness – more practice on naming itself – produced an immediate improvement in naming accuracy, but did not raise the ceiling vocabulary size that could be attained by the model. Intervention had served to propel the model more quickly along the same atypical trajectory. This is because the (lower) ceiling level of performance was constrained by the reduced computational capacity of the naming pathway. By contrast, either improving the semantics representations or the phonological representations – which were otherwise developing typically – produced slower changes during the intervention period, but then long-term gains in the size of the productive vocabulary that the model could acquire. Improving both semantic and phonological representations together gave the largest gains. These gains occurred because semantic and phonological representations became more delineated (or less confusable) through additional training, so that a pathway with limited capacity could achieve higher accuracy. Combining intervention on weakness and strengths gave both immediate gains during intervention and a long-term improvement in the vocabulary size that could be attained.

Table 2: Effects of timing: (a) Phonological + Semantic intervention, (b) Naming intervention, (c) Combined intervention at 100, 250, and 750 epochs

(a) Phonological+Semantic intervention

	TYP	ATYP	100	250	750
R1	85%	50%	57%	53%	50%
R2	85%	51%	57%	54%	52%
R3	87%	51%	54%	55%	52%
Avg	86%	51%	56%	54%	51%

(b) Naming intervention

	TYP	ATYP	100	250	750
R1	85%	50%	50%	50%	53%
R2	85%	51%	50%	51%	55%
R3	87%	51%	51%	51%	55%
Avg	86%	51%	50%	51%	54%

(c) Combined intervention

	TYP	ATYP	100	250	750
R1	85%	50%	52%	54%	52%
R2	85%	51%	57%	55%	55%
R3	87%	51%	55%	55%	55%
Avg	86%	51%	55%	55%	54%

We included assumptions about plasticity in the model – that there would be sensitive periods in the components but not pathways, that the lower plasticity of the components would be the limiting factor on naming development – but neither proved essential for producing the above effect. We also altered the timing of intervention, and here showed that improving strengths yielded greatest gains early in development, while remediating weaknesses yielded the greatest gains late in development. Combining both produced a uniform effect across development.

This model represents the first mechanistic demonstration of how working on strengths may serve to improve behavioural outcomes in developmental disorders. The observed effects of improving semantic and phonological representations are in line with those observed empirically for children with WFD (Best, 2005; McGregor & Leonard, 1989). Two further methods by which improving strengths might improve behavioural outcomes remain to be explored: encouraging compensatory mechanisms through intervention, and bolstering convergent sources of information in interactive systems.

The model nevertheless demonstrated relatively modest accuracy gains through intervention – certainly there was no elimination of the deficit (it was reduced from 35% to 29%). This is in line with general arguments made by Thomas et al. (2017): with some exceptions, where deficits arise through neurocomputational constraints in developing systems, behavioural interventions alone are unlikely to be successful in fully alleviating deficits. The conditions of optimal outcome are, however, a fruitful avenue for computational investigations, in the wider context of narrowing the gap between mechanistic theories of deficit and clinical practices of intervention.

Acknowledgments

This research was supported by UK Economic and Social Research Council (ESRC) grant no. RES-062-23-2721.

References

Best, W. (2003). Word-finding in children: Finding the right approach. *Royal College of Speech Language Therapists Bulletin*, September 2003, 5-6.

Best, W. (2005). Investigation of a new intervention for children with word-finding problems. *Int. Journal of Language Communication Disorders*, *40(3)*, 279-318.

Best, W., Fedor, A., Hughes, L., Kapikian, A., Masterson, J., Roncoli, S., Fern-Pollak, L., & Thomas, M. S. C. (2015). Intervening to alleviate word-finding difficulties in children: Case series data and a neurocomputational foundation. *Cognitive Neuropsychology*, *32(3-4)*, 133-168

Bishop, D., Nation, K., & Patterson, K. (2014). When words fail us: insights into language processing from developmental and acquired disorders. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1634), 20120403.

Bloom, L. (1973). *One word at a time: The use of single word utterances before syntax*. Mouton: The Hague.

Bock, K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 945-984). London: Academic Press.

Dell, G., Schwartz, M., Martin, N., Saffran, E., & Gagnon, D. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*, 801–838.

Harm, M., McCandliss, B. & Seidenberg, M. (2003). Modeling the successes and failures of interventions for disabled readers. *Scientific Studies of Reading 7*, 155-182.

Hinton, G. E. & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review*, *98(1)*, 74-95.

Laws, J., Campbell, C., Roulstone, S., Adams, C. & Boyle, J. (2008). Mapping practice onto theory: the speech and language practitioner's construction of receptive language impairment. *International Journal of Language Communication Disorders*, *43(3)*, 245-263.

Leonard, L. B. (2014). *Children with Specific Language Impairment*. Cambridge, MA: MIT Press.

Li, P., Zhao, X., & MacWhinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cognitive Science*, *31*, 581-612.

Masterson, J., Stuart, M., Dixon, M., & Lovejoy, S. (2010). Children's printed word database. *British Journal of Psychology*, *101*, 221–242.

Mayor, J., & Plunkett, K. (2010). A neuro-computational account of taxonomic responding and fast mapping in early word learning. *Psychological Review*, *117(1)*, 1-31.

McGregor, K. K., & Leonard, L. B. (1989). Facilitating word-finding skills of language-impaired children. *Journal of Speech and Hearing Disorders*, *54*, 141-147.

McMurray, B. (2014). Defusing the childhood vocabulary explosion. *Science*, *317*, 631.

Michie, S., & Prestwich, A. (2010). Are interventions theory-based? Development of a theory coding scheme. *Health Psychology*, *29(1)*, 1-8

Plunkett, K., Sinha, C., Møller, M. F. & Strandsby, O. (1992). Symbol grounding or the emergence of symbols? Vocabulary growth in children and a connectionist net. *Connection Science*, *4*, 293-312.

Rumelhart, D. E., Hinton, G. E., & Williams, J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel Distributed Processing: Explorations in the micro-structure of cognition (Vol. 1)* (pp. 318-362). Cambridge, MA: MIT Press.

Takesian, A. E. & Hensch, T. K. (2013). Balancing plasticity/stability across brain development. *Progress in Brain Research*, *207*, 3–34.

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

Thomas, M. S. C. & Karmiloff-Smith, A. (2003). Modeling language acquisition in atypical phenotypes. *Psychological Review*, *110(4)*, 647-682.

Thomas, M. S. C., Fedor, A., Davis, R., Yang, J., Alireza, H., Charman, T., Masterson, J. & Best, W. (2017). *Computational modeling of interventions for developmental disorders*. Submitted for publication.

Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods*, *40(1)*, 183-190.