

Simulating intervention to support compensatory strategies in an artificial neural network model of atypical language development

Juan Yang (jkxy_yjuan@sicnu.edu.cn)

Department of Computer Science, Sichuan Normal University
Chengdu, 610101 China

Michael S. C. Thomas (m.thomas@bbk.ac.uk)

Developmental Neurocognition Lab, Department of Psychological Sciences
Birkbeck, University of London, UK

Abstract

Artificial neural networks have been used to model developmental deficits in cognitive and language development, most often by including sub-optimal input-output representations or computational parameters in these learning systems. The next step is to simulate intervention to alleviate developmental impairments, to inform the mechanistic basis of remediation. Here we used a sample model of atypical language development (in the well-explored domain of past tense acquisition) to investigate the extent to which alternative training regimes may induce short-term or long-term compensatory changes in underlying function, and the extent to which this depends on the timing of intervention. We present a new method to derive ‘intervention’ training sets as a simulation of behavioral interventions, and assess its adequacy in our sample model.

Keywords: language development; developmental disorders; artificial neural network models; intervention; compensation

Introduction

Computational models of development, particularly those employing artificial neural networks (ANN), have provided hypotheses about the mechanistic bases of cognitive and language deficits (Mareschal & Thomas, 2007). For example, in the domain of language, Harm, McCandliss and Seidenberg (2003) demonstrated how limited connectivity in the phonology component of a reading model produced a system with symptoms of dyslexia. In a model of inflectional morphology, Thomas (2005) demonstrated how shallow sigmoid activation functions yielded processing units that were insensitive to small changes in the input, producing networks that exhibited developmental delay. More recently, Thomas and Knowland (2014) considered how multiple changes to intrinsic computational properties and extrinsic environmental factors could produce different types of language delay that either persisted or resolved over developmental time.

Progress of this type motivated Daniloff (2002, p.viii), in his book *Connectionist approaches to clinical problems in speech and language*, to comment ‘ANN theory will ... form the backbone of much of language therapy in the near future’. However, research and practice have yet to repay this optimism (though see Poll, 2011, for renewed attempts to make these links). Only one computational study has systematically explored the efficacy of a single intervention

to address a developmental deficit (in Harm et al.’s 2003 reading model; Harm, McCandliss & Seidenberg, 2003). Slightly more work has used ANN models to investigate remediation following acquired damage. For example, in a model of acquired dyslexia, Plaut (1996) considered the degree and speed of recovery through retraining, the extent to which improvement on treated items generalizes to untreated items, and how treated items are selected to maximize this generalization. Abel et al. (2007) demonstrated how an adult model of aphasia could guide actual interventions depending on patients’ error patterns. Even here, however, the work remains limited.

The computational approach to development has generated a growing understanding of environmental factors that influence learning in typical development (Borovsky & Elman, 2006; Gomez, 2005; Onnis et al., 2005). This includes the importance of factors such as the frequency of training items, their similarity and variability, and the provision of novelty in familiar contexts. However, there has yet to be a consideration of how these factors interact with learning systems containing the sorts of atypical computational constraints that lead to impoverished internal representations and, in turn, behavioral deficits compared to typically developing children. It is yet a further step to link such an understanding with the diverse activities that tend to be used by clinicians in speech and language therapy, including such activities as modeling, forced alternatives, repetition, visual approaches to support oral language, and reducing distractions (Ebbels, 2014; Law et al., 2007).

From the perspective of individual network models simulating development, where development is conceived of as the acquisition of the domain instantiated in the learning environment, it is not obvious that ‘behavioral interventions’ could alleviate a developmental impairment that arises either through inadequate representations or insufficient computational power. Here, we conceive of a behavioral intervention as representing the addition of some further training items to the normal training set of the network model. If the model is unable to learn the training set to a given performance level through limitations in processing capacity, adding further input-output mappings to the training set is unlikely to enhance performance on the patterns comprising the original training set. What one might call *normalization* through behavioral intervention is

therefore difficult if one conceives of developmental deficits as arising from limitations in individual systems. We define normalization as the acquisition of the abilities and knowledge that any typically developing system acquires through exposure to the normal training set.

There are at least three possible responses to this difficulty in achieving normalization. First, the computational properties of the system might be enhanced to enable it to acquire the training set (for instance, for the actual child, by interventions targeting motivation, or by pharmacological means; for the network, by altering one or more of its parameters).

Second, the intervention might target the input and output representations of the system, thereby simplifying the computational problem that the network has to learn. Harm, McCandliss and Seidenberg (2003)’s simulated phonological intervention for developmental dyslexia utilized this method. Best et al. (2015) have recently used a similar approach to simulate behavioral interventions for developmental deficits in productive vocabulary, alternatively targeting phonological or semantic representations that represent the two codes that must be associated in vocabulary development.

Third, one might take the view that what the atypical system needs to learn is not the training set per se (even though this is what typical systems acquire), but a general function implicit in the items comprising the training set. Acquisition of this general function can be assessed by performance on generalization sets rather than the training set. There may then be input-output mappings that can be added to the training set which could improve the network’s ability to learn the general function, even if performance on the original training set did not improve (or even worsened). One might term this approach *compensation*, since the aim is to optimize a subset of behaviors present in the original training set.

In this paper, we investigate possible ‘behavioral interventions’ to encourage compensation (so defined) in a widely used model drawn from the domain of language development, that of English past tense formation. This model has been used to capture developmental trajectories and error patterns as children acquire English verb morphology, but it has also been used as a sample associative system to consider more general issues in development (see, e.g., use of this model to investigate sensitive periods development: Marchman, 1993; Thomas & Johnson, 2006; to investigate population-level individual differences: Thomas, Forrester & Ronald, 2015). We introduce a method to derive ‘intervention patterns’ that are added to the training set of atypical networks for a limited period in development, simulating an intervention given to a child diagnosed with a developmental impairment. We compare the effectiveness of two different intervention sets for improving generalization performance on several possible implicit functions that the network might be able to acquire.

Base Model

We employed a simple simulation of past tense formation (Plunkett & Marchman, 1993) as our base model. This model uses an ANN to learn a quasi-regular mapping problem instantiated in an artificial language design to have many of the properties of English past tense formation. The learning domain is predominantly characterized by a general rule (add *-ed* to a verb stem to form its past tense). However, there exists a minority of exception or irregular verbs which form their past tense in different ways, for instance with arbitrary associations between stem and past tense form (go-went), no-change irregulars where the past tense form is the same as the verb stem (hit-hit), and vowel-change irregulars (sing-sang, ring-rang). The network is required to learn the association of verb stems to their past tense forms. Generalization can be tested on whether the network can apply the past tense rule to novel verbs, or can apply any of the irregular patterns to novel verbs sharing similarity to irregulars existing in the training set.

Simulating atypical development in the base model

Three layered ANNs were used to simulate individual children. All the ANNs had 57 input units in the input layer and 62 output units in the output layer, used to represent tri-phoneme verb stems and their past tense forms. The 57 input units corresponded to the binary encoding of a monosyllabic three-phoneme verb, where each phoneme was represented using 19 binary articulatory features. The output included the same binary encoding of the tri-phoneme string with the addition 5 extra bits to represent the suffix part of the past tense. The encoding is based on one that proposed by Plunkett & Marchman (1991; P&M91) (see Thomas & Karmiloff-Smith, 2003, for more details of the artificial language, including the consonant-vowel templates used to generate the artificial verbs).

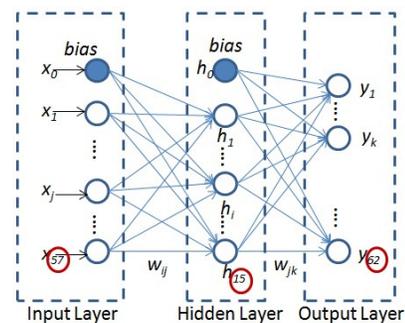


Figure 1: An ‘atypical’ ANN caused by a capacity deficit of reducing the number of hidden units from 50 to 15.

The training set comprised 508 artificial verbs, 410 regular verbs, 20 identical irregular verbs, 10 arbitrary irregular verbs and 68 vowel changed irregular verbs. Developmental trajectories were simulated by 1000 presentations of this training set (epochs). A network with 50 hidden units (back propagation algorithm, learning rate 0.1, momentum 0, temperature 1, initial weights randomized between ± 1) proved able to learn the training set within

approximately 300 epochs. We implemented a developmental deficit by reducing the computational capacity of the network (Thomas & Knowland, 2014). Piloting indicated that a reduction of hidden unit resources to 15 produced a persisting deficit in learning the training set (architecture show in Figure 1).

Simulating interventions to encourage compensation in the atypical network

We simulated a behavioral intervention to remediate the developmental impairment in the following way. We assumed that the impairment was diagnosed at some point relatively early in development. At this time, additional patterns were added to the original training set. The intervention set was added to, rather than replaced, the original training set, since we assume that in a clinical setting, interventions take place against the child's continued experience of his or her normal learning environment. In ANNs, replacement would also incur the risk of catastrophic interference. We assumed that the behavioral intervention was much smaller in scale than continued everyday experience, and so limited the intervention set to 10% the size of the original training set (50 input-output mappings versus 508 in the original set). Intervention continued for a limited duration (30 epochs of training), at which point the intervention ceased and training reverted to the original set. Intervention had the goal of encouraging acquisition of the regular past tense rule.

We manipulated the timing of intervention, from 'early' at 50 epochs in steps of 50 up to 'late' at 250 epochs (i.e., 5 stages: 50, 100, 150, 200, and 250) compared to the full training trajectory of 1000 epochs. The importance of early intervention has been stressed within a clinical setting, under the view that plasticity reduces over time. Simple feedforward ANNs have been claimed to capture a reduction in plasticity through entrenchment effects (Marchman, 1993), though a broader set of mechanisms may also produce reductions in plasticity, such as synaptic pruning (Thomas & Johnson, 2006).

We assessed normalization with respect to changes in performance on the original training set. We assessed compensation with respect to changes in performance on five generalization sets. These were:

- A *novel rhyme set*. Each novel verb shared two out of three phonemes with a verb in the training set. There were 410 regular verb rhymes, 20 no-change irregular verb rhymes, 10 arbitrary irregular verb rhymes, and 76 vowel change irregular verb rhymes. Finally, there were 56 novel verbs only shared one phoneme with any verb in the training set. This novel verb set has been used in previous simulations (e.g., Thomas & Karmiloff-Smith, 2003).
- A *shadow training set*. These were novel artificial verbs regenerated using the same consonant-vowel templates as the original training set (P&M91) and in the same proportions: 410 regular verbs, 20 no-change irregular

verbs, 10 arbitrary irregular verbs, and 68 vowel-change irregular verbs.

- A novel set of 508 *arbitrary irregular verbs* generated using the P&M91 templates.
- A novel set of 508 *no-change irregular verbs* generated using the P&M91 templates.
- A novel set of 508 *vowel-change irregular verbs* generated using the P&M91 templates.

For the *novel rhyme* set, generalization was assessed according to accuracy of producing regularly inflected forms. For irregular verbs in the *shadow training* set, and for the three novel irregular sets, generalization was tested according to accuracy of producing the target irregular output form.

We asked two questions. Did the intervention to encourage a compensatory strategy produce any benefit at the immediate end of the intervention period? And if so, did any benefit persist after the intervention ceased so that it was observable at the end of training? For the earliest intervention, performance was therefore assessed at 80 and 1000 epochs. For the latest intervention, performance was assessed at 280 and 1000 epochs. In each case, there were 10 replications of networks with different random seeds.

This leaves the challenge of how to construct an intervention set to encourage a compensatory strategy. In the next section, we propose a method.

A method for generating intervention sets for compensation

The problem we needed to solve is how to choose the most effective 50 intervention verbs among the hundreds of thousands of possible artificial verbs possible within the encoding scheme. The idea is intuitive: if we suppose some of the input units are more important and decisive than others, then the intervention verbs can be chosen based on these features. Now the problem becomes how to identify the key features within the original training set among the original 57 dimensional input space. The extra data set should be able to remedy a disordered ANN in its generalization of the past tense rule. Broadly, the approach we adopted was to take an ANN that had successfully learned the past tense problem. We then varied the activation level of input units and assessed the extent to which this might generate the error on the output. This should indicate the extent which input units encoded key dimensions would influence the performance of learning. Formally, we translated this challenge into an optimization problem, specified as:

$$\text{maximize}(\text{Net}_l(a_i x_i) - Y)^2, \sum_{i \leq 57} a_i = k, a_i = 0 \text{ or } 1(1)$$

In Formula (1), Y is the past tense matrix in the training data set, while $\text{Net}_l(a_i x_i)$ is the actual output of the ANN, l is the number of the final layer of the ANN. So, Formula (1) attempts to select out the input units that contribute most to the learning based on the training data set. Net_j is a recursive function defined in Formula (2):

$$Net_j = \begin{cases} \text{sigm}(a_i x_i, w_1), & j = 1 \\ \text{sigm}(Net_{j-1}, w_j), & j > 1 \end{cases} (2)$$

Since this is a combinatorial optimization problem, we used a Genetic Algorithm (GA) approach to find the optimized result. In this algorithm, $k = 25$ features. However, after the GA filtered out these features, a further selection was necessary, since no artificial verbs could fully satisfied the filtered features. In the final step, a subset of 5 or 6 features were chosen to generate two possible intervention data sets.

Key features selected to generate the intervention data sets

After running the GA, two sets of features were constructed. We refer to the first as the *GA feature set*, since it was closest to a shortened version of the original filtered 25 features yet consistent with legal verbs within the P&M91 encoding scheme. Novel verbs each contained 5 selected features shown in Table 1. We refer to the second as the *Voice satisfied feature set*. Novel verbs each contained 6 selected features shown in Table 2. These verbs were more consistent with those present in the original training set in terms of their voicing features. One might think of the first intervention set as optimized but somewhat strange, and the second as slightly less optimized but more natural given the ANNs previous training experiences.

Fifty novel verbs were created for each intervention set. The target output for each novel verb was the regular inflected past tense.

Table 1: GA Featured intervention data set.

Corresponding Unit	Feature location	Meaning
1	first phoneme	sonorant
2	first phoneme	consonantal
5	first phoneme	voiced
21	second phoneme	consonantal
43	third phoneme	voiced

Table 2: Voice Feature Satisfied data set.

Corresponding Unit	Feature location	Meaning
2	first phoneme	consonantal
5	first phoneme	voiced
21	second phoneme	consonantal
24	second phoneme	voiceless
40	third phoneme	vowel
43	third phoneme	voiced

Results

The results of the intervention are shown in Table 3. Beginning with the early intervention condition, no reliable improvement was observed on the original training set at the end of intervention. If anything, intervention caused performance on the training set to worsen. This is in line with the view that normalization is difficult for a network with limited capacity. Compensation was assessed via 5 novel verb sets assessing generalization of different functions that might be extended from the original training set. The first two, *novel rhyme* and *shadow training set*,

contain significant numbers of regular verbs which one might expect to be aided by the intervention. In both cases, statistically reliable benefits of intervention were observed. Three sets considered the possibility of generalizing irregular patterns. Since there is no systematic relation between arbitrary mappings, one would not expect an intervention effect on novel arbitrary verbs, and none was found. However, both novel no-change and novel vowel-change generalization sets showed benefits. This implies that the intervention had better enabled the atypical network to separate regular and irregular mappings within its representational space, and so generalize both types of general function to novel verbs with features that would support these functions.

We ran a series of omnibus ANOVAs to assess broader patterns. To emphasize the possibility that timing of intervention might have an effect, we focused on a comparison between the earliest intervention point (50 epochs) and latest (250 epochs). Figure 2 demonstrates the effect of intervention at intermediate time points. We first examined training set performance, considering factors of group (treated vs. untreated), intervention type (GA vs. V), and timing (50 vs. 250 epochs), separately for the immediate end of intervention and at the end of training. For performance at the immediate end of intervention, there was a reliable effect of the intervention ($F(1,9)=12.96$, $p=.006$), with an effect size of $\eta_p^2=.59$, corresponding to a *worsening* of performance. The intervention effect was not modulated by intervention type, nor by timing of intervention. For performance at the end of training, there was no effect of the intervention at 1000 epochs.

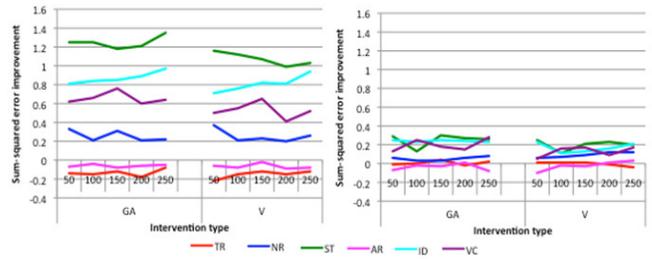


Figure 2: improvements produced by interventions, in terms of reductions in sum-squared error, for interventions occurring at different epochs between 50 and 250. Left panel: effects immediately after the end of intervention. Right panel: effects at the end of training (GA=GA feature intervention set, V=voice satisfied intervention set; TR=training set performance, NR=novel rhyme set, ST=shadow training set, AR=arbitrary, ID=identity, VC=vowel change).

We then considered generalization, with the same design but adding a 5-level factor of generalization set. For performance at the immediate end of intervention, performance was reliably improved by the intervention, ($F(1,9)=478.49$, $p<.001$), with an effect size of $\eta_p^2=.98$; the effect differed between intervention sets, with the GA featured set having the larger effect ($F(1,9)=31.68$, $p<.001$,

$\eta_p^2=.78$); improvement depended on the generalization set used ($F(1,9)=7.64$, $p=.022$, $\eta_p^2=.46$); and the intervention effect was not modulated by timing of intervention. The results at the end of training were similar, but with a reduced intervention effect size ($\eta_p^2=.83$), and now no modulation depending on the type of intervention used.

In sum, in line with our expectations, compensatory strategies were effectively encouraged via the addition of an intervention set. Intervention sets did not achieve

Table 3: Intervention results. UN=untreated, GA=GA feature intervention set, V=voice satisfied intervention set. Scores show performance of the network prior to intervention, immediately following an intervention lasting 30 epochs, and at the end of training of 1000 epochs, for untreated networks and networks treated with each intervention set. Performance is measured by sum-squared error, where lower numbers represent better performance and higher numbers represent worse performance. Reliable treatment effects are marked. Interventions were at five time points, 50, 100, 150, 200 and 250 epochs.

Test data sets		Average RMS errors returned by ANNs' intervened at different time point														
		50 th			100 th			150 th			200 th			250 th		
		UN	GA	V	UN	GA	V	UN	GA	V	UN	GA	V	UN	GA	V
508 verbs (training data)	UN	1.50	1.24	0.66	1.17	1.07	0.61	0.98	0.93	0.56	0.95	0.86	0.57	0.88	0.90	0.61
	GA	1.50	1.38*	0.67	1.17	1.22*	0.61	0.98	1.05*	0.52	0.95	1.04*	0.59	0.88	0.98	0.59
	V	1.50	1.46*	0.65	1.17	1.22*	0.60	0.98	1.05*	0.55	0.95	1.01*	0.58	0.88	1.02	0.65
572 novel rhymes	UN	6.41	6.41	6.65	6.34	6.35	6.64	6.38	6.41	6.63	6.28	6.30	6.53	6.28	6.31	6.56
	GA	6.41	6.08*	6.59	6.34	6.14*	6.61	6.38	6.10*	6.60	6.28	6.09*	6.47	6.28	6.09*	6.48
	V	6.41	6.04*	6.59	6.34	6.14*	6.57	6.38	6.18*	6.54	6.28	6.10*	6.41	6.28	6.05*	6.44
508 shadow training set	UN	7.34	7.39	8.22	7.40	7.50	8.26	7.56	7.67	8.21	7.68	7.70	8.27	7.81	7.84	8.33
	GA	7.34	6.14*	7.93+	7.40	6.25*	8.13+	7.56	6.49*	7.91*	7.68	6.49*	8.00*	7.81	6.49*	8.07*
	V	7.34	6.23*	7.97+	7.40	6.38*	8.15	7.56	6.60*	8.00	7.68	6.71*	8.04*	7.81	6.81*	8.13*
508 arbitrary irregular verbs	UN	18.47	18.56	18.77	18.58	18.62	18.81	18.60	18.69	18.88	18.64	18.66	18.86	18.70	18.74	18.88
	GA	18.47	18.63	18.84	18.58	18.66	18.83	18.60	18.77	18.91	18.64	18.72	18.85	18.70	18.79	18.96
	V	18.47	18.62	18.87	18.58	18.70	18.83	18.60	18.71	18.91	18.64	18.75	18.85	18.70	18.82	18.85
508 identical irregular verbs	UN	5.36	5.38	6.00	5.43	5.45	6.00	5.53	5.68	5.98	5.47	5.52	5.96	5.63	5.69	6.00
	GA	5.36	4.57*	5.75	5.43	4.61*	5.77+	5.53	4.83*	5.73	5.47	4.63*	5.72*	5.63	4.72	5.77
	V	5.36	4.67*	5.77	5.43	4.69*	5.89	5.53	4.86*	5.85	5.47	4.71*	5.80*	5.63	4.75	5.79
508 vowel changed irregular verbs	UN	8.15	8.24	9.04	8.29	8.35	9.11	8.42	8.47	8.93	8.45	8.48	8.98	8.57	8.52	8.91
	GA	8.15	7.62*	8.91	8.29	7.69*	8.86+	8.42	7.71*	8.75	8.45	7.88*	8.83	8.57	7.88*	8.63+
	V	8.15	7.74*	8.99	8.29	7.80*	8.95	8.42	7.82*	8.76	8.45	8.07*	8.89	8.57	8.00*	8.74

* Independent t-test treated vs. untreated $p < .05$
+ Independent t-test treated vs. untreated $p < .10$

normalization and indeed the compensatory strategy (while effective) initially caused performance to further diverge from the typical trajectory. Benefits of intervention were possible across a wide stretch of the developmental trajectory, with little indication of reductions in plasticity across the range of timing of interventions we considered. However, early interventions showed dissipating effects across development once the intervention was discontinued, with the exact type of intervention becoming less relevant.

Discussion

In this work, we have sought to build on successful research using ANNs to simulate atypical cognitive and language development, to consider implications for behavioral interventions to remediate developmental deficits. We focused on the domain of past tense formation, which has been a target of intervention for children with grammatical deficits (Ebbels, 2007; Kulkarni et al., 2014; Seeff-Gabriel et al., 2012). Rather than a realistic model of these interventions, our goal here was more preliminary: to explore methods for deriving possible intervention sets, to assess their impact on different areas of performance, to assess the influence of timing of intervention, and to assess the extent to which any gains were sustained following the cessation of intervention. However, we followed one of the broad tenets of an intervention called grammar facilitation, one of the most widely investigated methods for intervening to address grammar deficits in school age children. In grammar facilitation, the aim is to make target forms more frequent, which is hypothesized to help the child identify grammatical rules and give them practice at producing forms they tend to omit (Ebbels, 2014). In line with this view, our intervention added information to the training set

of an ANN model for a fixed period, to increase the salience of certain regularities in the problem domain.

Our results demonstrated that, where a language deficit arises due to limitations in processing capacity, compensation (optimization on a subset of the problem domain) is more readily achievable than normalization (improvement on the whole problem domain), and the particular training items chosen to effect the compensation can alter the size of the effect. Within the intervention window we considered, we found no reductions in receptiveness of the ANN to remediation, indicating no entrenchment or reductions in plasticity. However, benefits did dissipate once the intervention had ceased.

Returning to the target phenomenon, in reality, behavioral interventions to remediate developmental disorders of language and cognition are multi-faceted. They are usually interactional and social, and involve emotional and motivational factors in the child, as well as cognitive factors. There are myriad causes of variability in children's abilities, be they biological, psychological, environmental, or social – factors that must be considered in planning preventions or interventions (Beauchaine, Neuhaus, Brenner & Gatzke-Kopp, 2008). Clinical practice is driven by a range of principles including the emerging evidence base and the therapeutic setting, as well as the child and family's goals. Within approaches targeting speech and language needs

directly, the clinician may form a hypothesis as to (i) the nature of the difficulty and (ii) what will be optimally effective for a child. The results of intervention will further refine these hypotheses.

Nevertheless, the quality of neurocomputational mechanisms of learning and development is a key constraining factor, given that these mechanisms underlie behavior, and given that their plasticity is crucial in achieving remediation. We believe there is value in computational modeling work to further understand the mechanistic basis of atypical development and how deficits might be remediated by behavioral means.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (61402309) and UK ESRC grant RES-062-23-2721.

References

- Abel, S., Willmes, K. & Huber, W. (2007). Model-oriented naming therapy: Testing predictions of a connectionist model. *Aphasiology*, 21(5), 411-447.
- Beauchaine, T. P., Neuhaus, E., Brenner, S. L., & Gatzke-Kopp, L. (2008). Ten good reasons to consider biological processes in prevention and intervention research. *Development and Psychopathology*, 20, 745-774.
- 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 computational modelling foundation. *Cognitive Neuropsychology*. Article first published online: 25 FEB 2015, doi: 10.1080/02643294.2014.1003204
- Borovsky, A. & Elman, J. L. (2006). Language input and semantic categories: a relation between cognition and early word learning. *Journal of Child Language*, 33(4), 759-790.
- Daniloff, R. G. (2002). *Connectionist approaches to clinical problems in speech and language*. Erlbaum: Mahwah, NJ
- Ebbels, S. (2007). Teaching grammar to school-aged children with specific language impairment using Shape Coding. *Child Language Teaching & Therapy*, 23, 67-93.
- Ebbels, S. (2014). Effectiveness of intervention for grammar in school-aged children with primary language deficits. *Child Language Teaching & Therapy*, 30(1), 7-40.
- Fedor, A., Best, W., Masterson, J., & Thomas, M. S. C. (2013). Towards identifying principles for clinical intervention in developmental language disorders from a neurocomputational perspective. *DNL Techreport 2013-1* (www.psyc.bbk.ac.uk/research/DNL)
- Gomez, R. L. (2005). Dynamically guided learning. In M. Johnson & Y. Munakata (Eds.) *Attention and Performance XXI* (pp. 87—110). Oxford: OUP.
- Harm, M. W., McCandliss, B. D. & Seidenberg, M. S. (2003). Modeling the successes and failures of interventions for disabled readers. *Scientific Studies of Reading*, 7, 155-182.
- Kulkarni, A., Pring, T., & Ebbels, S. (2014). Evaluating the effectiveness of therapy based around Shape Coding to develop the use of regular past tense morphemes in two children with language impairments. *Child Language Teaching & Therapy*, 30(3), 245-254.
- Law, J., Campbell, C., Roulstone, S., Adams, C. & Boyle, J. (2007). Mapping practice onto theory: The speech and language practitioner's construction of receptive language impairment. *International Journal of Language and Communication Disorders*, 43, 245–63.
- Marchman, V. A. (1993). Constraints on plasticity in a connectionist model of the English past tense. *Journal of Cognitive Neuroscience*, 5, 215-234.
- Mareschal, D. & Thomas M. S. C. (2007) Computational modeling in developmental psychology. *IEEE Transactions on Evolutionary Computation (Special Issue on Autonomous Mental Development)*, 11, 137-150.
- Onnis, L., Monaghan, P., Christiansen, M., & Chater, N. (2005). Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies. In: K. Forbus, D. Gentner & T. Regier (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 1047-1052). Mahwah, NJ: Erlbaum.
- Plaut, D.C. (1996). Relearning after damage in connectionist networks: Toward a theory of rehabilitation. *Brain and Language*, 52, 25-82.
- Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perception: Implications for child language acquisition. *Cognition*, 38, 43-102.
- Poll, G. H. (2011). Increasing the odds: Applying emergentist theory in language intervention. *Language, Speech, and Hearing Services in Schools*, 42, 580-591.
- Seeff-Gabriel, B., Chiat, S., & Pring, T. (2012). Intervention for co-occurring speech and language difficulties. *Child Language Teaching & Therapy*, 20, 123-35.
- Thomas, M. S. C. (2005). Characterising compensation. *Cortex*, 41(3), 434-442.
- Thomas, M. S. C., Forrester, N. A., & Ronald, A. (2015). Multiscale modeling of gene-behavior associations in an artificial neural network model of cognitive development. *Cognitive Science*. Article first published online: 3 APR 2015, doi: 10.1111/cogs.12230
- Thomas, M. S. C. & Johnson, M. H. (2006). The computational modelling of sensitive periods. *Developmental Psychobiology*, 48(4), 337-344.
- Thomas, M. S. C. & Karmiloff-Smith, A. (2003). Modeling language acquisition in atypical phenotypes. *Psychological Review*, 110, 647-682.
- Thomas, M. S. C. & Knowland, V. C. P. (2014). Modelling mechanisms of persisting and resolving delay in language development. *Journal of Speech, Language, and Hearing Research*, 57(2), 467-483