When do behavioural interventions work and why?

Towards identifying principles for clinical intervention in developmental language disorders from a neurocomputational perspective

Anna Fedor*[1], Wendy Best[2], Jackie Masterson[3], Michael S. C. Thomas[1]

WORD Project

Introduction

Our questions

- What are the principles that underlie effective interventions for developmental disorders of language and cognition?
- Are the best interventions specific to problem domains, specific to deficit types, and/or dependent on when in development they take place?
- How are atypical internal representations reshaped by alternative training regimes?
- What are the neurocomputational mechanisms of development and intervention?
- Different ways of intervening have not been researched yet by developmental models

Interventions in practice

- Types of interventions employed are diverse and multiple factors feed into clinical decision making
- Studies in the domain of word retrieval conflict in their conclusions regarding optimum intervention (Ebbels et al., 2012)
- Further evidence is needed to distil the active ingredients of interventions with children with language needs (Lindsay et al., 2011)

Model aims

- We require a simple modeling environment to start an investigation of the principles of intervention from a neurocomputational perspective
- Initial model drawn from the field of language development (acquisition of inflectional morphology; Forrester & Plunkett, 1994)
- Aim: Create typically and atypically developing models; expose atypical models to new training environments to attempt to rescue behavior / normalize internal representations

Methods

The model

- Simple neural network model with 50 hidden units (Figure 1)
- Input units represent dimensions in a 2D space
- Output units represent category of items in the input space (e.g., inflectional categories, lexical categories, semantic categories)

Training

- Two learning problems (Table 1):
  - Regular categories: diagonal
  - Idiosyncratic categories: islands
- Input values varied between -0.5 and +0.5 on each input unit
- Input space consisted of 10,000 items
- Training set consisted of 10% of the input space
- Model was trained to learn categories with the backpropagation algorithm

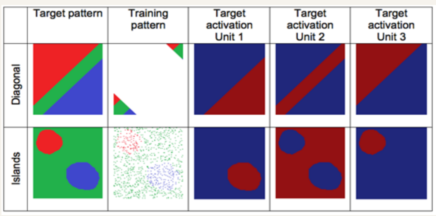

Table 1. Target patterns, training patterns and target activations for the diagonal and the islands problem. In the first and second columns different colors represent different target categories. In the rest of the figures red = active, blue = inactive.

Developmental deficits

- Low connectivity of hidden units (C=30% instead of 100%)
- Insensitive processing units - Shallow sigmoid (temperature, T=0.5 instead of 1)
- We have explored other computational constraints such as: processing noise, low learning rate, and low number of hidden units
- Two-by-two design: deficit vs. learning task (Table 2)

Intervention

- Intervention was modeled as items added to the original training set (intervention complements normal experience)
- Interventions were designed either to add sampling across the input space, to add training in areas that were 'prototypical' or central to each category, or in areas that demarcated category boundaries; see Table 2
- Interventions applied at different time points during development
- 10 replications in each condition

Figure 1. Network architecture (not all hidden units shown)

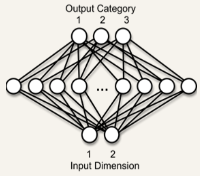

Table 2. Training scenarios

Deficit vs. learning task	Low connectivity	Shallow sigmoid
Diagonal problem	Scenario 1	Scenario 3
Islands problem	Scenario 2	Scenario 4

Results

Typical development (TD)

- The model learnt the categories in both tasks in less than 1000 training epochs
- This means successful generalization beyond the items of the training set (i.e., 10% of the items)
- More training was needed to learn the islands task
- We identified 4 phases of development (Figure 2)

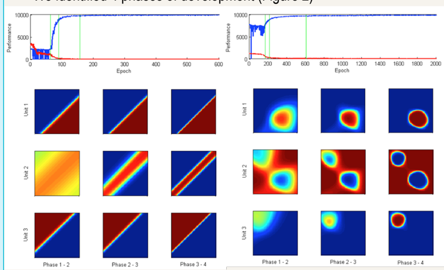

Figure 2. Developmental trajectories and phases for learning the diagonal (left) and the islands (right). Top figure: performance (blue) and mean square error (red) across development. Phase boundaries are indicated by green vertical lines. Second to fourth row of figures: snapshots of activation patterns of Output unit 1 to 3 at phase boundaries. Activation values are color-coded as temperature plots: red and blue indicates activation close to one and zero, respectively.

Atypical development

- Low connectivity of hidden units:
  - Marked deficit (see Figure 3) – never reaches TD performance
  - High individual variability depending on the location of the missing connections
- Shallow sigmoid:
  - Developmental delay – slower learning but usually reaches TD performance
  - Lower individual variability

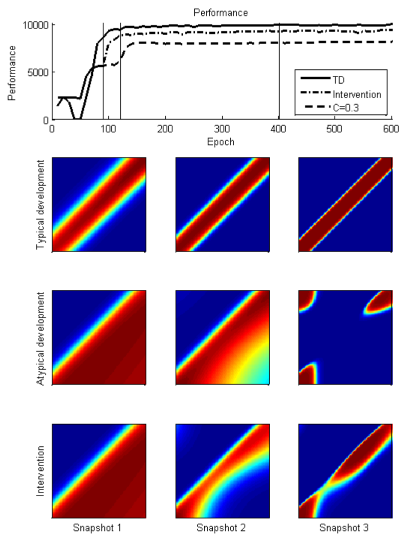

Figure 3. Developmental trajectories and internal representations in a typical case, an atypical case with low connectivity and the same atypical case with intervention. Top figure: Developmental trajectories. Vertical lines show epochs at which snapshots were taken. Colored figures: snapshots of the activation pattern of Unit 2 in the three cases.

Intervention

- Improvement score: improvement in performance due to intervention (model with deficit compared to same model with intervention)
- Type of deficit:
  - Shallow sigmoid: Intervention usually increased the speed of learning
  - Low connectivity: Heterogeneity in response to intervention: intervention did not help in many cases, but in some cases it increased performance
- Timing of intervention:
  - In phase 4 (the 'adult' state), most interventions had no effect
  - Before the adult state, some interventions for some deficits were more effective at earlier phases, but this was not uniform
- Type of intervention (Table 3)
  - Generally the best: random items (Intervention 1) and transect (Intervention2)
  - Generally the worst: separate patches of the categories (Intervention 3 and 4)
  - Deficit-specific intervention: Items from around the boundaries of the categories (Intervention 5) increased performance in the shallow sigmoid case in both problems
  - Task-specific intervention: Bigger corners helped learning the diagonal with both deficits

Table 3. Summary of the results. Numbers in bold represent phases in which a particular intervention was successful according to the t-tests. Below these, numbers lists the phases in which more than 7 networks improved.

Intervention	1	2	3	4	5	6
Diagonal interventions						
Islands interventions						
Scenario 1 C = 0.3	**2, 3** 1,2,3	**3** 1,2,3	**-** -	**-** -	**-** 3	- 1,2
Scenario 2 C = 0.3	**-** 1,2,3	**-** 2	**-** 2	**-** -	**-** -	**-** -
Scenario 3 T = 0.5	**1,2,3** 1,2,3	**1,2** 1,2,3	**1,2** 1	**-** -	**1,2,3** 1,2,3	**1,2,3** 1,2,3
Scenario 4 T = 0.5	**1,2,3** 1,2,3	**1,2,3** 1,2,3	**-** 3	**2** 1,2,3	**1,2,3** 1,2,3	**-** -

Conclusions & Discussion

- Does timing of interventions matter?
  - The effect of the timing of intervention was not uniform across conditions before the adult phase;
  - Interventions were generally ineffective in the adult phase.
- Are there interventions that generally work?
  - Best interventions across deficits and tasks: random items and items from the transect – both provide representative sample from all categories
  - Worst interventions across deficits and tasks: separate patches
- Are there interventions that are especially effective to improve a certain deficit?
  - Deficit-specific intervention: items from the boundaries of the categories for the shallow sigmoid deficit (helps sharpening the boundaries)
- Are there interventions that are especially effective to improve performance in a certain task?
  - Task-specific intervention: bigger corners for the diagonal task – provides more of the same kind of information
- Modeling can elucidate the principles that guide clinical interventions by aiding our theoretical understanding of the key issues
- Future work: scale up model and apply to more realistic rendition of language acquisition tasks

References

Ebbels, S. H. et al. (2012). Effectiveness of semantic therapy for word-finding difficulties in pupils with persistent language impairments: a randomized control trial. IJLCD, 47(1), 35-51.

Forrester, N. & Plunkett, K. (1994). Learning the Arabic plural: the case for minority default mappings in connectionist networks. In A. Ramand & K. Eiselt (Eds.), *Proceedings of the 16th Annual Conference of the Cognitive Science Society* (pp. 319-323). Mahwah, NJ: Erlbaum.

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. *IJLCD, 43*, 245–63.

Lindsay, G., Dockrell, J., Law, J., & Roulstone, S. (2011). *Better communication research programme: 1st interim report*. Technical report, UK Department for Education.

Affiliations & Acknowledgements

* Corresponding author: fedoranna@gmail.com
[1] Developmental Neurocognition Lab, Department of Psychological Sciences, Birkbeck College London, Malet Street, London WC1E 7HX, UK
[2] Division of Psychology & Language Sciences, University College London, Chandler House, 2 Wakefield Street, London WC1N 1PF, UK
[3] Department of Psychology and Human Development, Institute of Education University of London, 20 Bedford Way, London, WC1H 0AL, UK

This work was supported by ESRC grant RES-062-23-2721.