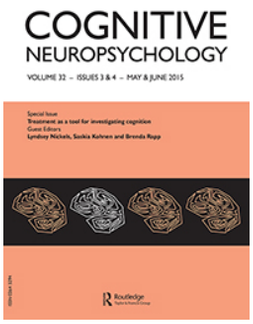
Routledge
Taylor & Francis Group



Cognitive Neuropsychology

ISSN: 0264-3294 (Print) 1464-0627 (Online) Journal homepage: http://www.tandfonline.com/loi/pcgn20

Intervening to alleviate word-finding difficulties in children: case series data and a computational modelling foundation

W. Best, A. Fedor, L. Hughes, A. Kapikian, J. Masterson, S. Roncoli, L. Fern-Pollak & M.S.C. Thomas

To cite this article: W. Best, A. Fedor, L. Hughes, A. Kapikian, J. Masterson, S. Roncoli, L. Fern-Pollak & M.S.C. Thomas (2015) Intervening to alleviate word-finding difficulties in children: case series data and a computational modelling foundation, Cognitive Neuropsychology, 32:3-4, 133-168, DOI: 10.1080/02643294.2014.1003204

To link to this article: http://dx.doi.org/10.1080/02643294.2014.1003204

Published online: 25 Feb 2015.

Submit your article to this journal

Article views: 69

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=pcgn20

Download by: [92.11.185.75] **Date:** 17 September 2015, At: 05:45

Cognitive Neuropsychology, 2015
Vol. 32, Nos. 3–4, 133–168, http://dx.doi.org/10.1080/02643294.2014.1003204

Routledge
Taylor & Francis Group

Intervening to alleviate word-finding difficulties in children: case series data and a computational modelling foundation

W. Best[a*], A. Fedor[b], L. Hughes[a], A. Kapikian[c], J. Masterson[c], S. Roncoli[c], L. Fern-Pollak[d] and M.S.C. Thomas[b]

[a]*Division of Psychology & Language Sciences, University College London, Chandler House, 2 Wakefield Street, London, WC1N 2PF, UK;* [b]*Department of Psychological Sciences, Birkbeck College London, London, UK;* [c]*Department of Psychology and Human Development, Institute of Education, London, UK;* [d]*School of Psychology, Social Work and Human Sciences, University of West London, Brentford, London*

Downloaded by [92.11.185.75] at 05:45 17 September 2015

We evaluated a simple computational model of productive vocabulary acquisition, applied to simulating two case studies of 7-year-old children with developmental word-finding difficulties across four core behavioural tasks. Developmental models were created, which captured the deficits of each child. In order to predict the effects of intervention, we exposed the computational models to simulated behavioural interventions of two types, targeting the improvement of either phonological or semantic knowledge. The model was then evaluated by testing the predictions from the simulations against the actual results from an intervention study carried out with the two children. For one child it was predicted that the phonological intervention would be effective, and the semantic intervention would not. This was borne out in the behavioural study. For the second child, the predictions were less clear and depended on the nature of simulated damage to the model. The behavioural study found an effect of semantic but not phonological intervention. Through an explicit computational simulation, we therefore employed intervention data to evaluate our theoretical understanding of the processes underlying acquisition of lexical items for production and how they may vary in children with developmental language difficulties.

Keywords: intervention; word-finding difficulties; connectionist modelling; phonology; semantics; naming

Introduction

Up to 7% of children have specific language needs, and around 25% of children attending language support services have word-finding difficulties (WFDs; Dockrell, Messer, George, & Wilson, 1998). Difficulty finding words can influence children's relationships, self-esteem, and education. Behaviours characteristic of WFDs include the use of fillers (e.g., *um*), empty words (*thing*), or general verbs (*doing*) instead of more specific words, the use of a similar-sounding response (*canister* for *camera; /grɪɹəl/ for squirrel*), the use of a word with a similar meaning or in the same category (*tiger* for *lion*), hesitation, repetition of words or phrases, rephrasing, the use of gesture (miming

*Corresponding author. Email: w.best@ucl.ac.uk

© 2015 Taylor & Francis

Downloaded by [92.11.185.75] at 05:45 17 September 2015

cleaning teeth for *toothbrush*), and talking about their difficulty (*"I know it, but I can't think of it"*).

WFDs have sometimes been attributed to impairments in the storage of word meaning: For instance, these children may also have problems distinguishing between similar semantic neighbours of a superordinate category, or they may produce impoverished word definitions (Dockrell, Messer, George, & Ralli, 2003; McGregor, Newman, Reilly, & Capone, 2002). However, children may experience difficulties in retrieving word forms even when testing suggests good representation of a word's meaning. This has led to the proposal that WFDs may be caused by problems in phonological processing—that is, in the retrieval or assembly of the component sounds of a word (e.g., Constable, Stackhouse, & Wells, 1997). The true picture may be more complicated, with multiple types of processing difficulty responsible and different children experiencing different sources for their word-finding problem (Best, 2005; Faust, Dimitrovsky, & Davidi, 1997). A similar account has, indeed, emerged in the case of adult aphasia (cf. Nickels, 2002). Nevertheless, a well-developed theoretical account needs to be able to explain what range of deficits might be expected within WFDs, according to the constraints that shape productive vocabulary acquisition and the extent to which these constraints vary in cases of atypical development. Moreover, the range of expected difficulties should also be linked to predictions about the kinds of interventions that should be effective given the underlying causes.

Little research has attempted to relate different profiles of WFDs to the outcome of intervention, and the endeavour is far from straightforward. For example, the outcome in Best's (2005) intervention study using a cueing aid did not differ across the five children with WFDs who took part, meaning it was not possible to meaningfully relate their naming profiles to the outcome of the therapy. Bragard, Schelstraete, Snyers, and James (2012) attempted to relate four individual children's therapy outcomes to their linguistic profiles. Participants' WFDs were characterized as either semantically or phonologically grounded, on the basis of poor performance on picture or spoken judgement tasks. Full assessment results were not reported, but two children with semantically categorized WFDs also presented with severe phonological and/or morphosyntactic difficulties. Each responded better to the phonological intervention, rather than the predicted semantic treatment. There are some methodological concerns with this study (e.g., second pretherapy baseline data were not provided to establish the robustness of the children's naming ability prior to intervention, and treatment sets differed in their pretherapy scores), thereby rendering the findings difficult to interpret.

One methodological approach that aids the advance of theoretical understanding is the construction of implemented computational models of development. Developmental disorders can be captured by altering the constraints under which development takes place, in terms of either the computational properties of the learning system (e.g., its resources or plasticity or level of processing noise) or the information to which it is exposed (Thomas, 2005a, 2005b; Thomas & Karmiloff-Smith, 2002, 2003; Thomas & Knowland, 2014). In principle, implemented models of developmental deficits can then provide the basis to explore the effects of intervention. However, to date, few researchers have extended their models in this way. The greater precision enforced upon theory by implementation is desirable in the case of WFDs, where naming deficits have been attributed to diverse and vaguely specified causes including "a general difficulty accessing semantic information", "a speed of processing deficit", and representations that are "impoverished" or "less developed".

One modelling approach that has had some success in capturing both developmental and acquired disorders of language is the use of artificial neural networks (sometimes called "connectionist" models). Examples include models of developmental dyslexia (Harm, McCandliss, & Seidenberg, 2003), developmental delay in inflectional morphology (Thomas, 2005a; Thomas & Knowland, 2014), aphasia (Foygel & Dell, 2000), and acquired dyslexia (Plaut, McClelland, Seidenberg, & Patterson, 1996). Examples of the parameters that were altered to capture atypical performance include: (a) reducing the number of internal processing units, (b) reducing the connectivity between layers of processing units, (c) reducing the sensitivity of the processing units to changes in input,

Downloaded by [92.11.185.75] at 05:45 17 September 2015

and (d) reducing the learning rate—that is, the amount that connection weights changed in response to learning events.

To our knowledge, there has been only one computational study that has explored the effectiveness of intervention in a model of a developmental deficit: Harm et al. (2003) used a connectionist model of reading to explore why certain classes of interventions are more effective than others to alleviate reading impairments in developmental dyslexia. Models have considered rehabilitation after acquired damage in adulthood. Abel, Willmes, and Huber (2007) sought to show how an adult model of aphasia could guide actual interventions depending on patients' error patterns, while Plaut (1996) explored which training regimes might aid recovery from acquired dyslexia manipulating item typicality. In other work, we have begun to explore the computational foundations of intervening to improve performance in atypically developing connectionist learning systems (Fedor, Best, Masterson, & Thomas, 2013). However, modelling of intervention remains in its early stages.

Importantly in the current context, intervention can be used as a direct test of a model, and to the extent that the model embodies a theory of the cause of a developmental deficit, a test of that theory. This requires the following scenario: We have available one or more children with developmental deficits, characterized by a particular profile of (possibly relative) strengths and weaknesses in the domain of interest; the model is used to capture the atypical profiles of these individuals; a number of interventions have been constructed that can be applied to the model; the model predicts which (if any) of these interventions are most successful for the simulated individuals; actual intervention data are available about the most successful intervention for the individuals (implying, of course, that the children undergo each of the interventions). This is the design we offer in the current article. Specifically, we used a developmental connectionist model of word retrieval to predict the best intervention for two 7-year-old girls with WFDs, who each underwent two interventions aimed at improving their productive vocabulary difficulties. The results of the intervention were used as a test of the model.

Connectionist computational models have been influential in theories of word retrieval, particularly that of Dell et al. (Dell, Faseyitan, Nozari, Schwartz, & Coslett, 2013; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997). This model simulated the retrieval of a phonological form given a word's meaning. The model was "hard-wired" into its adult state and was designed to account for errors in aphasia following damage. It is therefore not best suited to consider developmental mechanisms. A number of computational models have conceptualized lexical acquisition in terms of learning mappings between representations of semantics and phonology. For example, Plunkett, Sinha, Moslashller, and Strandsby (1992) used a connectionist network to associate localist labels with abstract semantic codes and vice versa, focusing on phenomena such as the vocabulary explosion and the comprehension–production asymmetry, as well as under- and over-extension errors. However, for WFDs, a key issue is whether the semantic and phonological representations have developed normally, and therefore these representations should be a product of development rather than specified by the modeller.

Our model therefore embodies the theoretical proposal that word retrieval involves learning the mapping between representations of semantics and phonology, and that each of these representations undergoes its own developmental process. Deficits may occur within the development of the semantic component, within the phonological component, or in the pathway responsible for learning the mapping between the two, and may involve atypical settings of various different computational parameters. A given case of atypical development might involve only one of these deficits, but it might also involve multiple deficits. The proposed model of behavioural impairments therefore considers deficits in different locations (we considered a single location, double location, or triple location) and of different nature (we considered reducing the number of internal processing units, reducing the connectivity between units, and reducing the sensitivity of the processing units to changes in input).

Downloaded by [92.11.185.75] at 05:45 17 September 2015

The DevLex model of Li, Farkas, and MacWhinney (2004), the DevLex II model of Li, Zhao, and MacWhinney (2007), and the early word learning model of Mayor and Plunkett (2010) offered potentially appropriate frameworks upon which to base our word-retrieval model. Each model acquires representations of semantics and phonology in self-organizing maps, before learning associations between the maps via Hebbian links to capture lexical acquisition. Our concern was that by their nature, self-organizing maps enforce a simple two-dimensional feature space on both semantic and phonological representations. However, a richer representation of both semantic and phonological space might be necessary to capture the subtle developmental differences often associated with WFDs. We chose instead to encode these types of information over autoassociative networks developing distributed internal representations, where the internal representational space was a free parameter. This allowed internal representations to develop with (in our case) up to 500 dimensions. Similarly to the DevLex and early word learning architectures, our model then learned associations between semantic and phonological codes, which were themselves at various stages of development.

In the next section, we consider modelling typical and atypical development, detailing the case studies of children with WFDs and how the model captured their profiles. The following section then uses the model to predict interventions, before evaluating those predictions using intervention data.

Modelling typical and atypical development in word retrieval

The initial targets of our computational model were twofold: to capture typical development in word retrieval, and to capture the atypical profile of two children with WFDs. These children were drawn from a larger, ongoing study evaluating interventions for children with WFDs (Best et al., 2013). For the purposes of our simulations, both typical and atypical development were profiled using performance on four core tasks. We first describe these tasks, then our two case studies. We then move on to characterize the typically developing model, and how it was altered to capture the two case studies.

Empirical data

Core tasks

The four core tasks were intended to measure the ability to produce object names, the ability to comprehend object names, semantic knowledge separate from names, and phonological knowledge separate from word meaning, respectively (for full details, see Appendix 1).

In the *confrontation naming task*, children were required to retrieve and produce words in response to a picture. Pictures comprised 72 black and white line drawings of objects. Both accuracy and latency of responses were recorded. Errors were classified according to whether they were semantic (coordinate, superordinate, functional, circumlocution, visual attributes), phonological (nonwords, formal), or mixed semantic and phonological. Explanations and examples of error types can be found in Appendix 2.

In the *word–picture verification task (WPVT)*, children's knowledge of the meaning of words was assessed. Children were presented with a picture on two occasions, one together with the correct word name for the picture, and on a separate occasion accompanied by the name of a close semantic coordinate. Children were asked to decide whether the spoken word corresponded to the picture and to score correct needed to accept the target name and reject the name of the close semantic coordinate. The procedure was carried out for all 72 items presented in the confrontation naming task (after that task was completed). The task was split into two blocks separated by a break, with a picture's two presentations appearing in separate blocks and the order counterbalanced across participants.

In the *picture-judgement task (PJs)*, children's semantic knowledge was assessed. Children were shown three pictures and were required to choose which of two coordinate pictures (e.g., *chair* or *bed*) was associated with a third picture (e.g., *pyjamas*). They were asked to choose which of the two items in the lower part of the screen fitted best with the item at the top (i.e., the correct answer for this practice example was *bed*), responding by using one of two keys on the computer keyboard. The targets were a subset of 20 target

Downloaded by [92.11.185.75] at 05:45 17 September 2015

pictures from the naming task. The PJs task was designed as a developmental analogue of the widely used Pyramids and Palm Trees test (Howard & Patterson, 1992) employed to assess the intactness of semantic knowledge in adults with acquired brain damage. Importantly, no language was used in stimulus presentation and response, so that the children were making judgements based on their knowledge of the semantic relationship between the pictured items. Scores consisted of the proportion of correct trials and the median key press response times for correct items.

The *Children's Test of Nonword Repetition* (CNRep; Gathercole & Baddeley, 1996) was employed to assess the children's phonological abilities in the absence of word meaning. Repetition is a sensitive task as both phonological input and output processing need to be adequate for correct production of the forms. The test consists of 40 nonwords of increasing length and complexity. We report standard scores and percentage correct.

Finally, since two of the preceding tasks required speeded responses, we included a measure of *simple choice reaction time*, to assess possible differences in speeded motor responses. The task was adapted from Powell, Stainthorp, Stuart, Garwood, and Quinlan (2007). Six pictures of animals appeared at random on a screen. Two of these animals (a green dinosaur and an orange dinosaur) were targets. Children were asked to press a key as quickly as they could when either of the targets appeared, with a separate key for each target. We recorded median response times for correct responses.

Case studies

Two case studies of children with WFDs were identified based on their performance on the Test of Word Finding Second Edition (TWF–2; German, 2000). The children were referred by the special educational needs coordinators/inclusion managers at their schools. The TWF–2 test assesses a potential disparity between word production and word comprehension. On this test, both children had a word-finding quotient of 60, which was lower than the 1st percentile compared with the TWF–2 standardization sample. Both scored in the normal range on the comprehension component of the test. Neither child had a diagnosis of dyspraxia, autistic spectrum disorder, attention deficit hyperactivity disorder, or global developmental delay. Our consideration of WFDs does not entail that WFDs are the sole language deficit that these children experienced, although for these two, as for many of the children in our larger study, it was the most salient one. On a test of receptive vocabulary (British Picture Vocabulary Scale Third Edition, BPVS–III; Dunn, Dunn, & Styles, 1997), the children scored at the 9th and 3rd percentiles, while on a test of nonverbal ability (Pattern Construction subtest from the British Ability Scales Second Edition, BAS–II; Elliot, Smith, & McCullouch, 1996), the children scored at the 21st and 24th percentiles, respectively.

Case Study 1, Amy,[1] was 7 years 6 months at initial testing. Her family was from White British ethnic background and lived in London. Amy was described by teachers as having "problems with her pronunciation with words", as well as literacy difficulties. She reported feeling "angry" and "annoyed" by her word-finding difficulties because others speak over her at home and at school. Case Study 2, Magda, was 7 years 7 months at initial testing. Her family was from White British ethnic background and also lived in London. Magda had been known to the local Speech and Language Therapy service since 3 years of age. She was originally referred to the Early Years service, due to nursery and parental concerns about delayed language and dysfluency. Magda was described by her mother as frequently using "the wrong word in the wrong place" and having "problems with pronunciation". Her teacher felt that her difficulty in finding words made it "hard for her to work with a partner, as she can't explain her ideas".

Amy and Magda were given the four core tasks, along with the simple choice reaction time task. Their performance was compared against 20 typically developing (TD) children selected from a sample of 100 children participating in the larger study of Best et al. (2013), to form an age-matched comparison group. The 20 TD children ranged in age from 7 years 1 month to 8 years 0

Downloaded by [92.11.185.75] at 05:45 17 September 2015

months (mean = 90.75 months, *SD* = 3.86). They attended schools in London and the surrounding area, within catchments with a similar socioeconomic profile to that of the schools of the two children with WFDs. Background assessments of receptive vocabulary (BPVS–III) and nonverbal ability (Pattern Construction subtest of BAS–II) yielded a mean standard score of 105.35 (*SD* = 12.03) for the BPVS (which has a mean of 100 and *SD* of 15), and 56.95 (*SD* = 10.43) for Pattern Construction (which has a mean of 50 and *SD* of 10).

The two girls with WFDs were also given several other background language tasks, to allow for a richer characterization of their language profiles. These tests included: the Word Discrimination subtest of the Test of Auditory Processing Skills Third Edition (TAPS–3; Martin & Brownell, 2005) assessing their ability to discriminate sounds within words; the BPVS–III (Dunn et al., 1997) to measure receptive vocabulary; four subtests from the Clinical Evaluation of Language Fundamentals Fourth Edition (CELF–4; Semel, Wiig, & Secord, 2003) from which we provide results for Concepts and Directions, to give a measure of language comprehension, and the overall Core Language Score; the Test for Reception of Grammar (TROG; Bishop, 1989), which assesses understanding of different grammatical structures; and the Fluency subtests of the Phonological Abilities Battery (Frederickson, Frith, & Reason, 1997), which require word generation on the basis of either semantic category or initial sound. Although neither of the girls with WFDs was given a formal hearing screening, parents were asked about their child's hearing status, and available test results were requested. There were no indications of hearing deficits with either child, and to be included in the study, the children had to score above a threshold in the TAPS auditory discrimination task (a scaled score of 6), a threshold that both Magda and Amy exceeded.

Results

Table 1 shows the performance of the girls on the four core tasks relative to the performance of the TD children. In line with their performance on the test of word finding and their inclusion in the study, both girls were very poor at confrontation naming relative to TD children. Magda found this task particularly difficult. Appendix 2 shows the error classification scheme and errors in each category made by the girls. Both made semantic errors. However, the number of coordinate errors made by Amy and Magda was not more than 1.5 standard deviations above the mean of the TD children. Magda differed from the TD children in that she produced mixed errors (words both semantically and phonologically related to the target, e.g., *scrape* for *rake*). These are striking because English does not afford many opportunities for such errors. She also produced mixed errors in conversation. These errors indicate both semantic and phonological influence on word finding (Nickels, 1997). Finally, both girls produced phonologically related nonword errors. These were very unusual in the naming attempts of the TD children and tend to be associated with postlexical phonological production difficulties.

On WPVT, which tested comprehension of the target items, Amy's accuracy was almost 1.5 standard deviations below the mean for TD children, while Magda performed well below 1.5 standard deviations from the mean score of the TD children. On the picture judgement task (PJs), which does not require lexical processing, both girls scored 16/20 items correct, which fell 1.5 standard deviations below the mean for the TD group. In addition, Magda performed particularly slowly on this task. Nevertheless, both girls performed comparably to the TD group in the nonlinguistic simple choice reaction time task. Lastly, on nonword repetition (CNRep), both girls performed poorly.

The findings from the background testing are shown in Table 2. Both girls performed well on the word discrimination task (TAPS) suggesting adequate processing of speech input. This implies that the difficulties in CNRep may have stemmed from retrieving, holding, or producing the phonemes, rather than with input processing. Magda showed impaired performance on language comprehension tasks at the single word (BPVS) and sentence level (CELF Concepts and Directions subtest and TROG). Amy had relatively good

Table 1. Performance of Amy and Magda on four core tasks and a measure of general processing speed (choice RT), relative to 20 age-matched TD children.

Task performance	Amy	Magda	TD mean	TD *SD*	1.5 *SD* from TD mean
Naming (accuracy/72)	*21*	*14*	40.40	6.31	30.93
WPVT (accuracy/72)	48	*42*	55.15	5.35	47.12
PJs (accuracy/20)	*16*	*16*	18.65	1.50	16.41
PJs (RT, ms)	2855	*4290*	2886	575	3748
CNRep (standard score)	*51*	*52*	93.68[a]	13.40	73.58
CNRep (% correct)	*22.50*	*25*	66.84[a]	13.04	47.28
Semantic (coordinate) errors	11	7	8.50	3.73	14.10
Mixed (semantic and phonological) errors	0	*5*	0.25	0.55	1.08
Formal (phon. real word) errors	0	0	0.10	0.31	0.56
Phonological (nonword) errors	*2*	*2*	0.15	0.37	0.70
Choice RT task (ms)	589	525	588	140	797

Note: Numbers in italics indicate where case studies differ by more than 1.5 standard deviations from the typically developing (TD) mean. Values in the final column show 1.5 standard deviations below the TD mean for accuracy and standard scores, and 1.5 standard deviations above the TD mean for reaction times and errors. Errors were raw scores out of 72 pictures named; other errors were mostly "don't know" or "no response", with some unrelated or perceptual. WPVT = word–picture verification task; CNRep = Children's Test of Nonword Repetition; PJs = picture-judgement task; RT = reaction time; phon. = phonological.
[a]TD data for the CNRep task are for 19 children rather than 20 as for the other tasks.

Table 2. Background assessments.

Assessment results	Amy	Magda
Test of Auditory Processing Skills Third Edition: Word Discrimination scaled score (percentile)	10 (50)	9 (37)
British Picture Vocabulary Scale Third Edition: standard score (percentile)	80 (9)	71 (3)
Clinical Evaluation of Language Fundamentals Fourth Edition: Concepts & Directions scaled score	11	3
Clinical Evaluation of Language Fundamentals Fourth Edition: Core Language standard score (percentile)	81 (10)	60 (0.4)
Test for Reception Of Grammar: percentile	25–50	10–25
Phonological Abilities Battery: Fluency Test Alliteration standard score (percentile)	95 (37)	87 (20)
Phonological Abilities Battery: Fluency Test Semantic standard score (percentile)	111 (77)	77 (6)

Downloaded by [92.11.185.75] at 05:45 17 September 2015

language comprehension as demonstrated by her performance on these three tasks. On the Phonological Abilities Battery Fluency Task, Magda performed poorly with relatively worse generation of semantic than alliterative items. Amy performed well on this task, although she demonstrated the reverse pattern from Magda with better performance on semantic than alliterative fluency.

Combining these test results, together with clinical observation, the two girls' profiles can be summarized as follows. Amy had relatively good comprehension. Her performance on the tasks involving semantic processing (PJs and WPVT) was around 1.5 standard deviations below the TD mean. In contrast, on tasks requiring phonological output (naming and CNRep) her scores were more than 3 standard deviations below the TD mean. Thus, her naming problem appeared to arise at least in part from difficulties in postlexical phonological assembly for word production. Evidence in support of this view includes poor repetition of nonwords in the context of good auditory discrimination, combined with the production of nonword phonological errors in naming.

Magda had word-finding difficulties in the context of language needs spanning comprehension and expression. Her scores on the background tests suggested wider language impairment beyond her WFDs. Neither her performance on tasks tapping semantic processing nor that on tasks tapping

Downloaded by [92.11.185.75] at 05:45 17 September 2015

phonological processing matched those of typically developing children. Her profile on these tasks matched well with that on our four core tasks. Specifically, she performed very slowly on the PJs task, which required semantic judgements in the absence of linguistic processing, and her accuracy score was more than 2 standard deviations below the TD mean on the WPVT task, where accurate performance required acceptance of the target name and rejection of a close semantic coordinate. Magda also had considerable difficulty with both naming and CNRep, scoring more than 3 standard deviations below the TD mean on both tasks. The pattern across the tasks suggested that her word-finding difficulties may have multiple sources, arising from both semantic and phonological output processing problems, perhaps with a particular difficulty in accessing word forms as indicated by the presence of mixed errors (which are rare in the TD sample) and by her frequent filled pauses (um, er, etc.) before word retrieval in conversation —for example: "OK, um. Well, well . . . my best DVD is Alvin chipmunks."

While we have focused on the girls' patterns of difficulties, they also exhibited considerable communicative strengths. Amy was better able to find words in conversation than in a constrained picture-naming situation and was an enthusiastic communicator and storyteller. Magda was aware of her language difficulties and communicated well—for example, by sometimes holding the conversational floor to avoid questions and saying things in different ways until she got her message across. She used gesture well when unable to find words. Despite these strengths, the girls' everyday communication was influenced by their difficulty in retrieving words, including word-finding behaviours in connected speech and in conversation (see later).

Computational modelling of typical and atypical development of word retrieval

In this section, we first describe the model of typical development, including how it was trained and tested. We then detail how the model was altered to capture atypical development, indicating how deficit types were matched to our case studies.

Typical development

Our typically developing model involved linking the developing representations within a phonological processing component and a semantic processing component. Each component was modelled using an autoassociator—that is, a three-layer artificial neural network trained with the backpropagation algorithm (Rumelhart, Hinton, & Williams, 1986) to reproduce the code applied to its input layer onto its output layer. In doing so, the network had to pass this information through an internal processing layer, thereby requiring it to form internal representational codes of the key features of, respectively, phonological space and semantic space. These two emerging representations were then linked via separate associative pathways. Different mappings between input (either semantic or phonological) and output (semantic or phonological) were used to capture performance on the four core tasks.

To date, models that combine simulation of developmental deficits and intervention are largely absent. For the current model, we wished to start with a relatively simple framework that focused on the implications of the model architecture and the type and location of deficits. We did not emphasize the ecological validity of the training set, and we address this decision in the Discussion. Instead, we followed Plunkett et al.'s (1992) model of vocabulary development, incorporating some basic differences about the nature of phonological and semantic knowledge and the association between them. Phonological representations of words were strings of phonemes encoded using articulatory features; semantics were feature sets with a prototype-based similarity structure; the association between word forms and their meanings was arbitrary.

Simulation details

Lexicon. Words were modelled as randomly paired semantic and phonological representations. The semantic representations were fed into the semantic module, and the phonological representations were fed into the phonological module. The model employed a simplified domain with a lexicon of 100 words. In previous models,

Downloaded by [92.11.185.75] at 05:45 17 September 2015

semantic representations have been considered either in terms of feature sets (either explicitly derived from adult raters or extracted from text corpora) such as in the reading model of Harm and Seidenberg (2004), or as an emergent property of linking features to labels (as in "a bird has wings", "a bird can fly"; see e.g., Rogers & McClelland, 2004). The important characteristic is the existence of separate semantic categories with internal family resemblance structure. We created semantic representations possessing separate categories and family resemblance structure in line with the vocabulary acquisition model of Plunkett et al. (1992). Five prototypes were randomly generated, each consisting of 57 semantic features, 28 active and 29 inactive. Semantic representations for the lexicon were then generated by randomly activating/inactivating units in these prototypes with a probability of .05. The result was five prototype classes, with 20 semantic representations each, where the average Euclidean distance between semantic representations was lower within a prototype class (around 17) than between prototype classes (around 30).

Phonological representations were generated using consonant–vowel templates, where each word was nine phonemes long, and each phoneme was encoded using an articulatory feature based code; there were 42 phonemes—24 consonants and 18 vowels—based on English (Thomas & Karmiloff-Smith, 2003). Similar-sounding phonemes therefore had similar representations, and the Euclidian distance of words that had more phonemes in common was less than that of words that had fewer phonemes in common.

Architecture

The architecture is shown in Figure 1. The model consisted of two components: a semantic component, a phonological component, and two layers in the associative pathways between the components. The semantic and phonological components each had an input layer, an output layer, and a hidden layer. The components were used to input and output the semantic and phonological representations of words, respectively. They also included recurrent connections from the output layers to the input layers. The recurrent connections within each component were employed only during testing to give the model the facility of settling into its "best guess" output given an input by iteratively honing a response, with the number of cycles required to reach this settled state serving as a simulation of reaction time. The associative layers served as pathways to connect the hidden layers of the semantic and phonological components in each direction of activation flow. For the typically developing model, the size of the semantic input and output layers was 57 units, the size of the phonological input and output layers was 171 units, and the size of all hidden layers (semantic, phonological, and both associative layers) was 500 units.

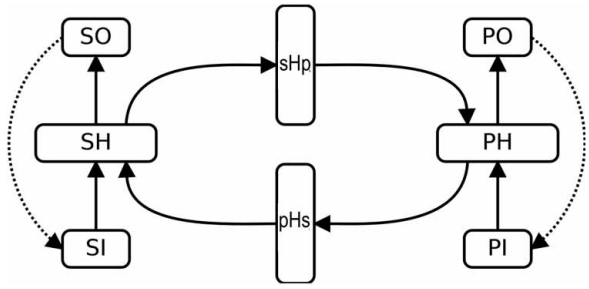


Figure 1. The architecture of the model. Bars represent layers of units; arrows represent layers of weights between these units. Recurrent weights represented by dashed arrows were not trained. Abbreviated layer names stand for: SI = semantic input; SH = semantic hidden; SO = semantic output; PI = phonological input; PH = phonological hidden; PO = phonological output; sHp = associative hidden layer from the semantic to the phonological module; pHs = associative hidden layer from the phonological to the semantic module.

Downloaded by [92.11.185.75] at 05:45 17 September 2015

Adjacent layers were fully connected (i.e., connection density was 1).

Training

The model was trained using the backpropagation learning algorithm (Rumelhart et al., 1986) to perform four tasks, simulating the four core tasks that were used to test the children. Two of the core tasks, picture judgements and nonword repetition, were designed to assess children's semantic and phonological representations, respectively. Since in the model, more direct measures of these representations were available, we did not implement the task designs explicitly (e.g., the use of picture triads in PJs; the use of nonwords in CNRep), instead using the more direct measures. Both phonology and semantics were assessed by performance on the training set, despite generalization of phonological knowledge to novel strings being necessary for nonword repetition. Since such generalization is not required for the semantics task, for consistency we chose to assess performance on training sets across the components, rather than assessing one component on generalization and one on the training set.

The four tasks were:

Semantic input–semantic output (SS) task: This task was used to train the semantic component independently of the phonological component. The semantic representation of words was fed into the semantic input (SI) layer, and the network was trained to reproduce the same representation on the semantic output (SO) layer. During testing, performance on this task was used to simulate children's performance on the PJs task.

Phonological input–phonological output (PP) task: This task was used to train the phonological component in isolation, to develop representations of the phonological forms of the words in the lexicon. The phonological representation of words was fed into the phonological input (PI) layer, and the network was trained to reproduce the same representation on the phonological output (PO) layer. During testing, performance on this task was used to simulate children's performance on the CNRep task.

Semantic input–phonological output (SP) task: To simulate lexical retrieval, the model was given a semantic representation on the SI layer and was required to output the appropriate phonological form on the PO layer. During training of this task, the semantic and phonological modules were held constant, and only the weights between semantic hidden (SH) and phonological hidden (PH) layers were trained. (Table 3 indicates weight layers that were altered during training versus those that were held constant in each task.) The intention was to capture the development of lexical retrieval as the learning of associations between emerging semantic and phonological codes. The activation of the PH layer was checked against the activation of the same layer when the input originated from the PI layer in the PP task, to derive error signals for weight change. The objective was to elicit the same hidden representations irrespective of the origin of the input (semantic or phonological). During testing, performance on the SP task was used to simulate performance on confrontation naming, where the

Table 3. Weight layers in the model that were activated during testing for each task and those that were altered during training.

Task	Connection pathways
SS task (PJs)	SI ⇒ SH ⇒ SO
PP task (CNRep)	PI ⇒ PH ⇒ PO
SP task (confrontation naming)	SI → SH ⇒ sHp ⇒ PH → PO
PS task (WPVT)	PI → PH ⇒ pHs ⇒ SH → SP

Note: Thin arrows (→) denote weight layers that were activated during testing, and thick arrows (⇒) denote those that were altered during training. PJs = picture-judgement task; CNRep = Children's Test of Nonword Repetition; WPVT = word–picture verification task; SS = semantics-to-semantics task (simulating the PJs task); PP = phonology-to-phonology task (CNRep); SP = semantics-to-phonology task (confrontation naming); PS = phonology-to-semantics task (WPVT); I = input layer; H = hidden layer; O = output layer; sHp = associative hidden layer from the semantic to the phonological module; pHs = associative hidden layer from the phonological to the semantic module.

Downloaded by [92.11.185.75] at 05:45 17 September 2015

input is a picture, and the output is the phonological form of the verbal label for that picture.

Phonological input–semantic output (PS) task: To simulate lexical comprehension, the model was given a phonological representation on the PI layer and was required to output the appropriate semantic representation on the SO layer. During training of this task, the phonological and semantic modules were held constant, and only the weights between PH and SH layers were trained (see Table 3). The intention was to capture the development of lexical comprehension as a mirror of lexical retrieval—that is, as the learning of associations between emerging phonological and semantic codes. During testing, performance on the PS task was used to simulate performance on the word–picture verification task, where children match a spoken word to a picture.

A training epoch consisted of training the whole lexicon with one of the tasks. Training on the four tasks was interleaved using random selection without replacement, so that in a round of 100 epochs, each task was trained for 25 epochs. Development of normal models was followed until they reached ceiling performance, or until 4000 epochs of training had been completed. The "age" of the model was defined as the number of epochs divided by four. During testing, the outputs of the model were considered as 1 (active) if activation was higher than 0.9 and 0 (inactive) if activation was lower than 0.1. A response was scored as correct if all units were in the required state.

Given the simulated language environment, performance was assessed on the full training set for each task. This obviously contrasts with the empirical case, where experimental tasks use a very limited subset of items compared with the children's vocabulary.

Results

Models with typical parameter settings (TD models) usually learned all four tasks within 3000 training epochs. Figure 2 shows median values averaged over 50 networks with different random seeds. Since the four tasks differed in relative difficulty, the model's rate of acquisition of the four tasks could not be a target of simulation. (Similarly, in the empirical study, no attempt was made to equate naming, word–picture verification, picture judgement, and nonword repetition for difficulty). In addition, the simulated trajectories depict the whole learning process, whereas performance of the 7-to-8-year-old children would match to only an intermediate portion of these trajectories. Because of the way in which the model was

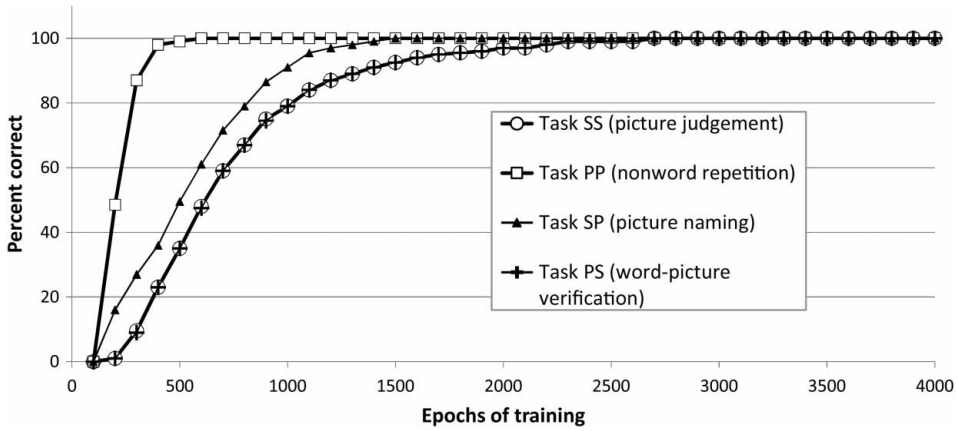


Figure 2. Developmental trajectories of the four core tasks across 4000 training epochs. Trajectories were calculated as medians from 50 typically developing (TD) models. SS = semantic input–semantic output task; PP = phonological input–phonological output task; SP = semantic input–phonological output task; PS = phonological input–semantic output task.

Downloaded by [92.11.185.75] at 05:45 17 September 2015

trained, tasks relying on associations between the phonological and semantic components in either direction were always constrained by the performance within the components themselves, and specifically by the performance of the output component. Thus performance on the PS task could never be higher than performance on the SS task, and, similarly, performance on the SP task could never be higher than performance on the PP task. Our simplified semantic prototype structure had the unintended consequence of making it harder for the network to learn semantic representations (SS task) than phonological representations (PP task). Development of the semantic representations therefore limited development on the lexical comprehension task, causing the SS and PS trajectories generally to overlap. This was a limitation of our simplified TD model. In simulating the lexical retrieval task, where networks produced errors prior to developing ceiling performance, errors were mostly semantic—that is, the name of another item in the same semantic category. This captured the typical preponderance of semantic errors observed in the TD children shown in Table 1.

Simulating atypical development

Before training, the TD model was compromised in three different ways to induce computational deficits. These disturbances included: (a) decreasing the *number of hidden units* in various layers; (b) decreasing the *number of connections* between layers; or (c) using a *shallow sigmoid unit activation function* for the artificial neurons in various components of the model (see Thomas, 2005a, for implementation). The activation function in the processing units of artificial neural networks determines how the units change their activation level given the net excitation and inhibition they receive. The units in the networks we used incorporated sigmoid activation functions, equivalent to a smoothed threshold function. Use of a shallow sigmoid function, induced by reducing a parameter known as the "temperature", alters the response properties of the units to make them less sensitive to changes in the input and therefore less able to discriminate between small changes in the signals they receive. The three types of deficits were always applied prior to the onset of training (Thomas & Karmiloff-Smith, 2002) and could be applied across the whole architecture or to specific parts. We examined the effect of these deficits on the developmental trajectories of the model to establish the single deficit or combination of deficits that best simulated Amy's and Magda's performance on the four core tasks.

Two theoretical points are worth noting in this enterprise. First, case studies of developmental disorders serve a particular role. A case study represents a combination of a developmental deficit, background individual differences, and the individual's history of experience. While the three cannot be definitively disentangled in a single case, even with a detailed case history, the case study can demonstrate what is possible in a given combination of the three factors. Where the pattern is unusual, the case study can show the outer limits of the constraints within which development occurs. Simulations of individual cases should show that the profile of deficits falls within the parameter space of the model (see, e.g., Foygel & Dell, 2000).

Second, one possible criticism of the enterprise of capturing individual cases is that it is an exercise in data fitting. Given that artificial neural networks have many free parameters (the multitude of connection weights), surely a successful fit cannot be informative? The response to this view is twofold. First, alterations to the TD model were highly constrained. The only changes pertained to the computational constraints that shape the developmental process. The connection weights were themselves always the product of a learning system exposed to a structured learning environment. The weights, while driving the behaviour of the model, were not directly altered to bring the system closer to the behaviour that was the target of simulation (that is, the patterns of deficits). Deficits had to emerge from an experience-dependent developmental process in a system with compromised learning abilities. Secondly, the current goal was not solely to capture the profile of the case studies but, with these individualized models in hand, to predict optimal interventions. These predictions were tested empirically.

Downloaded by [92.11.185.75] at 05:45 17 September 2015

Simulation details

Our initial goal was to model the qualitative difference between Amy's and Magda's performance on the four core tasks compared to the range of variation exhibited by the TD children. Both girls were closer to the TD range on the SS (picture judgement) and PS (word–picture verification) tasks, and much poorer on the PP (nonword repetition) and SP (picture naming) tasks. First, we applied alterations to start-state hidden units, connectivity, and activation function temperature one by one in the semantic component (S), the phonological component (P), or the associative layers between the components of the model (A), before considering the possibility that multiple deficits might be necessary to capture the profiles of the case studies.

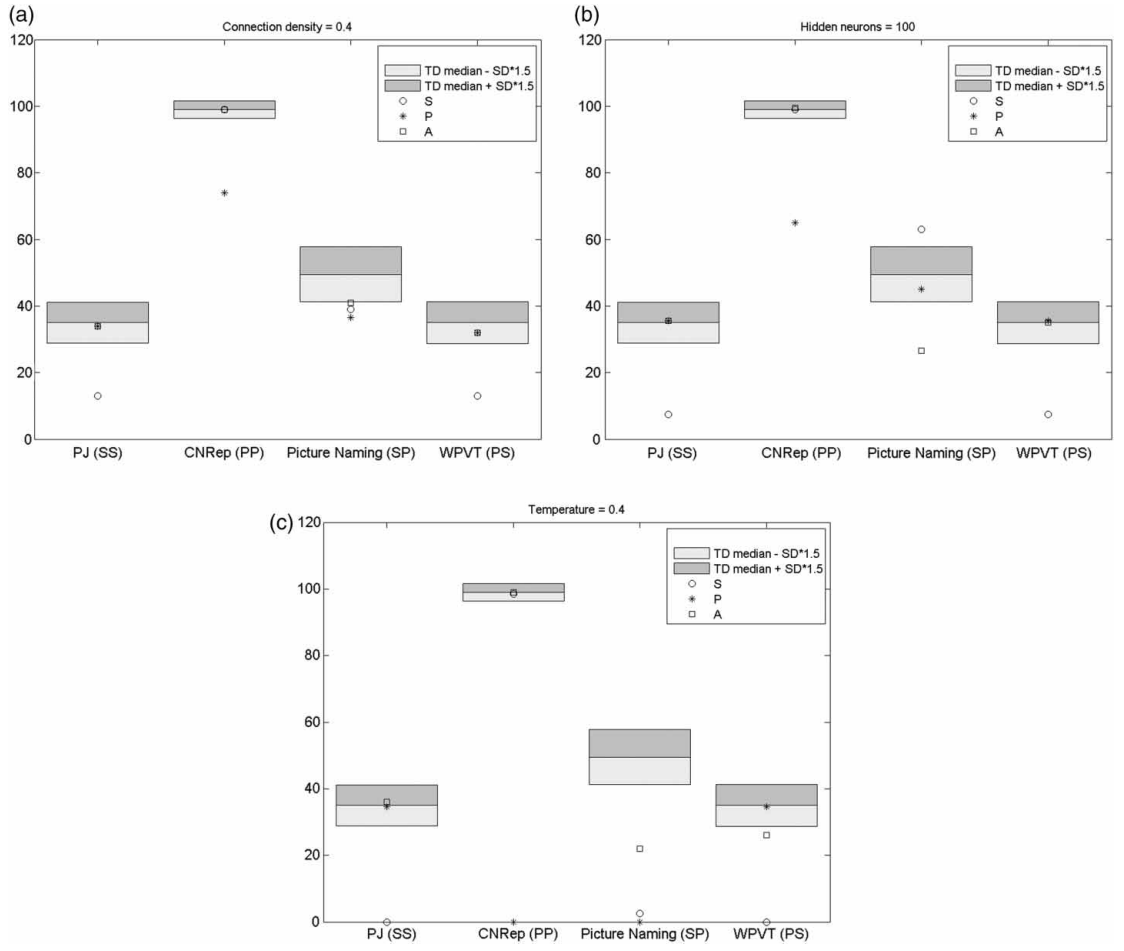


Figure 3. Comparison of typically developing (TD) models and single location deficit models after 500 training epochs. The boxes represent the TD range (median ±1.5 standard deviations) calculated from 50 simulations. The separate data points represent different locations of the deficit calculated as the average of 10 atypical simulations: S = semantic module; P = phonological module; A = associative layers. PJ = picture-judgement task; CNRep = Children's Test of Nonword Repetition; WPVT = word–picture verification task; SS = semantic input–semantic output task; PP = phonological input–phonological output task; SP = semantic input–phonological output task; PS = phonological input–semantic output task. Deficits were (a) lower connection density, (b) lower number of hidden units, or (c) lower temperature of the sigmoid transfer function.

Downloaded by [92.11.185.75] at 05:45 17 September 2015

Results

Figure 3 compares the performance of TD models and atypical models after 500 epochs of training for the three types of deficit, respectively. None of these parsimonious, single-location deficits captured the behavioural patterns produced by Amy and Magda. As expected, deficits in the semantic module usually produced lower performance on the SS (picture judgement) task but did not influence the PP (nonword repetition) task; conversely, deficits in the phonological module resulted in lower performance in the PP task but did not influence the SS task. Both girls performed more poorly than TD children on both SS and PP tasks, implying that, in terms of the model, they had deficits at multiple locations. The effect of single semantic or phonological module deficits on the SP (lexical retrieval) and PS (lexical comprehension) tasks, which involved both modules, varied according to the location and the type of the deficit but also did not yield a good fit.

Exploratory single-deficit simulations. Deficits to the semantic and phonological components affected formation of category boundaries in the respective high-dimensional representational spaces, while deficits to the associative pathways between components altered the ability of the system to learn mappings between those representations. With respect to those mappings, the acquisition of picture naming was little affected by changes in connectivity; changes in hidden units only caused impairments when they occurred in the associative pathways (indeed, when they occurred in the semantic component, performance improved, presumably as a more concise semantic representation was better able to acquire the prototype structure); changes in temperature caused impairments wherever they occurred. It is notable that connection density deficits to single locations did not produce large lexical-retrieval impairments, given that this was the key feature to be simulated. For word–picture verification, changes in connectivity, hidden units, and temperature only had marked effect when they occurred in the semantic component.

Multiple-deficit simulations to capture behavioural profiles

We next evaluated combinations of deficits to capture the profiles of the two case studies. Both girls performed more similarly to the TD children in semantic output tasks (picture judgement and lexical comprehension) than in phonological output tasks (nonword repetition and lexical retrieval). This suggests that their deficits were more serious in the phonological module than in the semantic module and/or in the links from semantic input to phonological output. Keeping this in mind, we experimented with deficits of different strength in the two modules and identified three "double location deficits" that captured Amy's profile. The modified parameters for these models can be found in Table 4 and the resulting profiles in Figure 4. The double deficits involved a reduction in connectivity in both modules, a reduction of hidden units in both modules, or a reduction in temperature in both modules. The rest of the parameters were set to the same values as in the TD models, and, as before, performance of TD models and atypical models was compared after 500 epochs of training. It is noteworthy that different processing atypicalities generated similar atypical profiles, implying a many-to-one mapping of

Table 4. Parameter settings in the three double-deficit models, simulating Amy's profile.

Deficit type	Deficit at semantic module	Deficit at phonological module
Deficit C at S + P	Connection density of SI–SH = 0.7 Connection density of SH–SO = 0.7	Connection density of PI–PH = 0.3 Connection density of PH–PO = 0.3
Deficit H at S + P	Size of SH = 250	Size of PH = 60
Deficit T at S + P	Temperature of SH = 0.92	Temperature of PH = 0.60

Note: Locations: S = semantic; P = phonological; I = input; H = hidden; O = output (see Figure 1). Deficits: C = connectivity reduction; H = hidden unit reduction; T = unit activation function temperature reduction.

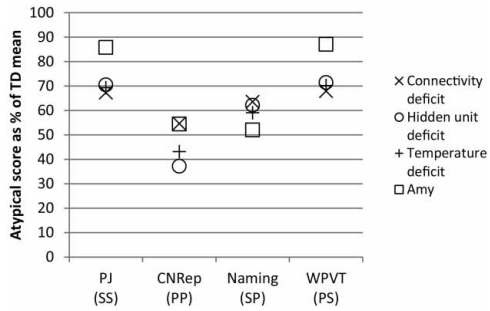


Figure 4. Simulation of Amy's deficit profile. Data show a comparison of the performance of *double-location-deficit* models after 500 training epochs, expressed as a percentage of the performance of typically developing (TD) models at the same point in training. The separate data points represent different types of start-state deficit calculated as the average of 10 atypical simulations, with deficits applied to connection density, number of hidden neurons, or temperature of the sigmoid transfer function. Deficits affected the semantic and phonological modules. Amy's performance is also depicted, once more expressed as a percentage of the mean performance of the TD children. PJ = simulated picture judgement task; CNRep = simulated nonword repetition task; Naming = simulated confrontation naming task; WPVT = simulated word–picture verification task; SS = semantics-to-semantics mapping; PP = phonology-to-phonology mapping; SP = semantics-to-phonology mapping; PS = phonology-to-semantics mapping.

processing deficits to behavioural profile. Figure 4 represents our fit to Amy's deficit. The model somewhat exaggerated the size of the deficit in picture judgement and did not capture the fact that Amy's word–picture verification task performance just fell within the bottom of the normal range.

Turning to Magda, we induced a further deficit. This was based on the view that the girls scored similarly on the within-component tasks (picture judgement and nonword repetition) but that Magda then scored more poorly than Amy in the word-retrieval and word-comprehension tasks. We therefore hypothesized that she might have additional limitations in the links between the semantic and phonological modules, as well as deficits in the semantic and phonological modules themselves, corresponding to a widespread deficit. We considered three methods of inducing the further deficit, parallel to the double-location-deficit conditions. In the connectivity deficit, the connection density of the associative layers was reduced to 0.1; in the hidden unit deficit, the size of the associative layers was much reduced to 30 (associative hidden layer from the semantic to the phonological module, sHp) and 20 (associative hidden layer from the phonological to the semantic module, pHs) units, respectively; and in temperature deficit, the temperature of the associative layers was reduced to 0.5 (sHp) and 0.4 (pHs). The multiple-deficits parameter sets are shown in Table 5. The performance of these models after 500 training epochs is shown in Figure 5. It was the same as the performance of double-location-deficit models on the simulated picture-judgement

Table 5. Parameter settings in the multiple-deficit models, simulating Magda's profile.

	Deficit location		
Deficit type	Semantic module	Phonological module	Associative pathways
Deficit C at S+P+A	Connection density of SI–SH = 0.7 Connection density of SH–SO = 0.7	Connection density of PI–PH = 0.3 Connection density of PH–PO = 0.3	Connection density of SH–sHp = 0.1, sHp–PH = 1, PH–pHs = 0.1, pHs–SH = 0.1
Deficit H at S+P+A	Size of SH = 250	Size of PH = 60	Size of associative layers = 30 (sHp) and 20 (pHs)
Deficit T at S+P+A	Temperature of SH = 0.92	Temperature of PH = 0.60	Temperature of associative layers = 0.5 (sHp) and 0.4 (pHs)

Note: Locations: S = semantic; P = phonological; I = input; H = hidden; O = output (see Figure 1); A = associative layers. Deficits: C = connectivity reduction; H = hidden unit reduction; T = unit activation function temperature reduction. sHp = associative hidden layer from the semantic to the phonological module; pHs = associative hidden layer from the phonological to the semantic module.

Downloaded by [92.11.185.75] at 05:45 17 September 2015

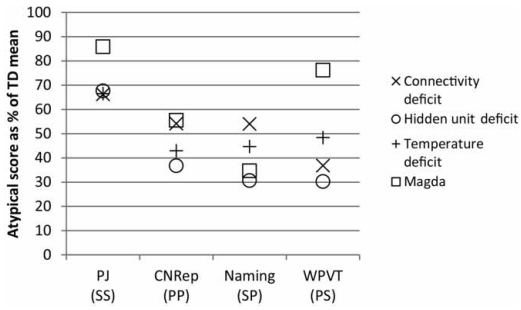


Figure 5. Simulation of Magda's deficit profile. Data show a comparison of the performance of *multiple-deficit* models after 500 training epochs, expressed as a percentage of the performance of typically developing (TD) models at the same point in training. The separate data points represent different types of start-state deficit calculated as the average of 10 atypical simulations, with deficits applied to connection density, number of hidden neurons, or temperature of the sigmoid transfer function. Deficits affected the semantic module, the phonological module, and associative pathways between the modules. Magda's performance is also included, once more expressed as a percentage of the mean performance of the TD children. PJ = simulated picture judgement task; CNRep = simulated nonword repetition task; Naming = simulated confrontation naming task; WPVT = simulated word–picture verification task; SS = semantics-to-semantics mapping; PP = phonology-to-phonology mapping; SP = semantics-to-phonology mapping; PS = phonology-to-semantics mapping.

and nonword repetition tasks, but was now lower on the lexical-retrieval and lexical-comprehension tasks. Figure 5 represents our fit to Magda's deficit. Once more, it somewhat exaggerated the size of the deficit on picture judgement and, while capturing lexical-retrieval deficits in confrontation naming, exaggerated the deficit on the word–picture verification task that tested lexical comprehension. Changes to different processing parameters again yielded similar sorts of profile, implying a many-to-one mapping of processing deficit to behavioural profile.

In sum, Amy was simulated with start-state deficits to semantic and phonological components, while Magda was simulated by start-state deficits to semantic and phonological components and additionally impairments to the pathways linking these components.

Modelling interventions for word-finding deficits

To constrain the simulated interventions we applied to our WFD models, we first considered the literature on successful interventions for WFDs. There are relatively few well-controlled studies investigating therapy for WFDs in children. Studies have focused on comparisons between intervention techniques (Hyde Wright, Gorrie, Haynes, & Shipman, 1993; McGregor & Leonard, 1989; Wing, 1990). The results of such studies are generally positive. Overall, they suggest that therapy can improve word-finding abilities in children. This is the case for both semantic (Ebbels et al., 2012) and phonological approaches (Bragard et al., 2012). In addition, the improvement may be found in children of a wide age range (e.g., Hyde Wright et al., 1993, 8–14 years; Wing, 1990, 6–7 years); it can generalize to untreated words (Ebbels et al., 2012; Hyde Wright, 1993); and it can persist (Bragard et al., 2012; McGregor, 1994).

Nevertheless, the studies conflict as to the most effective approach. For example, Hyde Wright et al. (1993) and Wing (1990) contrasted semantic and phonological interventions. In the former study, with 8–14-year-olds, the semantic techniques appeared to bring about improvements in word finding whilst the phonological techniques did not. In the latter study with younger children (aged 6–7 years) the reverse was found. One reason for this discrepancy may be that different children, for example of different ages, or with different difficulties, respond best to different interventions (e.g., McGregor & Windsor, 1996). A similar finding has emerged from studies on adults with anomia as part of acquired aphasia. It has been established that both phonological components analysis (Leonard, Rochon, & Laird, 2008) and semantic features analysis (Boyle & Coelho, 1995; Coehlo, McHugh, & Boyle, 2000) can improve adults' naming (Van Hees, Angwin, McMachon, & Copland, 2013). However, the relationship between the level of deficit and outcomes of intervention is far from straightforward (Lorenz & Ziegler, 2009).

Another source of constraining evidence is the developing body of research into children's word learning. This has produced mixed evidence on

Downloaded by [92.11.185.75] at 05:45 17 September 2015

Downloaded by [92.11.185.75] at 05:45 17 September 2015

the role of semantic versus phonological cues in influencing children's ability to acquire and retain new words. Gray (2005) found that a group of 24 children with specific language impairment (aged 4;0–5;11 years) comprehended more words in a semantic condition and produced more names accurately when given phonological cues. Meanwhile, the typically developing control group performed similarly in both trials. Zens, Gillon, and Moran (2009) identified an order effect in their study of 19 children with specific language impairment (aged 6;3–8;2 years). Positive treatment effects for producing new words were found for the children who received phonological awareness intervention, *followed by* semantic intervention. There was no improvement in the comprehension of new words for either group.

For our simulated interventions, we chose one intervention that would target the structure of the semantic representations, in isolation from phonological representations, and not in the context of naming or comprehension. This condition exposed the model to further training on semantic distinctions, but retained the same structure of that information. In contrast, a second intervention targeted the phonological representations, once more in isolation from the rest of the system. Our two intervention conditions, semantic and phonological, were applied independently to our models of Amy and Magda, to predict which condition would be more successful in alleviating word-finding problems, in comparison to conditions where development proceeded without intervention.

Simulation details

Intervention was simulated as increased training on one of the tasks, in addition to the four-cycle training that represented experience-driven development in everyday situations. The semantic intervention was modelled by increasing training on the SS task (twice as much as usual), and the phonological intervention was modelled by increasing training on the PP task (also twice as much as usual), while continuing training on all the other tasks to model normal learning. Intervention started after 500 epochs of training (in simulation terms, equivalent to the age of our case studies) and continued until the model reached 100% performance on each task or until the model reached 1000 epochs of training. The ages of these models were calculated according to their nonintervention training epochs; thus, models with intervention received more training on one of the tasks than models of the same age without intervention.

Since the trajectory of each model's development in the absence of intervention was available to us, we employed a target measure that focused on the extent to which the relevant intervention speeded up development. We therefore subtracted the age (in epochs) at which the model reached 90% performance with intervention from the age at which the model reached 90% performance without intervention on the naming task. Positive scores on this metric represent more effective interventions in speeding up the development of lexical retrieval.

Finally, our models of Amy and Magda were specified by deficit location, with different parameter changes yielding similar profiles. We considered in parallel the effects of interventions on systems whose atypical profiles were caused by the different parameter changes.

Results

In the case of simulated-Amy, a system with a double deficit, we observed that the phonological intervention significantly speeded up the development of lexical retrieval whichever deficit (connectivity, hidden units, temperature) was applied. The result is shown in Figure 6. Analyses of variance revealed a main effect of intervention, $F(1, 27) = 22.64$, $p < .001$, $\eta_p^2 = .456$, reflecting the advantage of phonological over semantic intervention, no main effect of deficit type, $F(1, 27) = 0.71$, $p = .500$, $\eta_p^2 = .050$, and a significant interaction reflecting the greater advantage of phonological intervention over semantic in the hidden unit deficit condition, $F(2, 27) = 3.87$, $p = .033$, $\eta_p^2 = .223$. In individual Bonferroni-corrected *t*-tests (shown in Table 6), the semantic intervention was not reliable for any deficit type. For naming, one can conceive of the process as a sending code (semantics), a mapping pathway, and a receiving

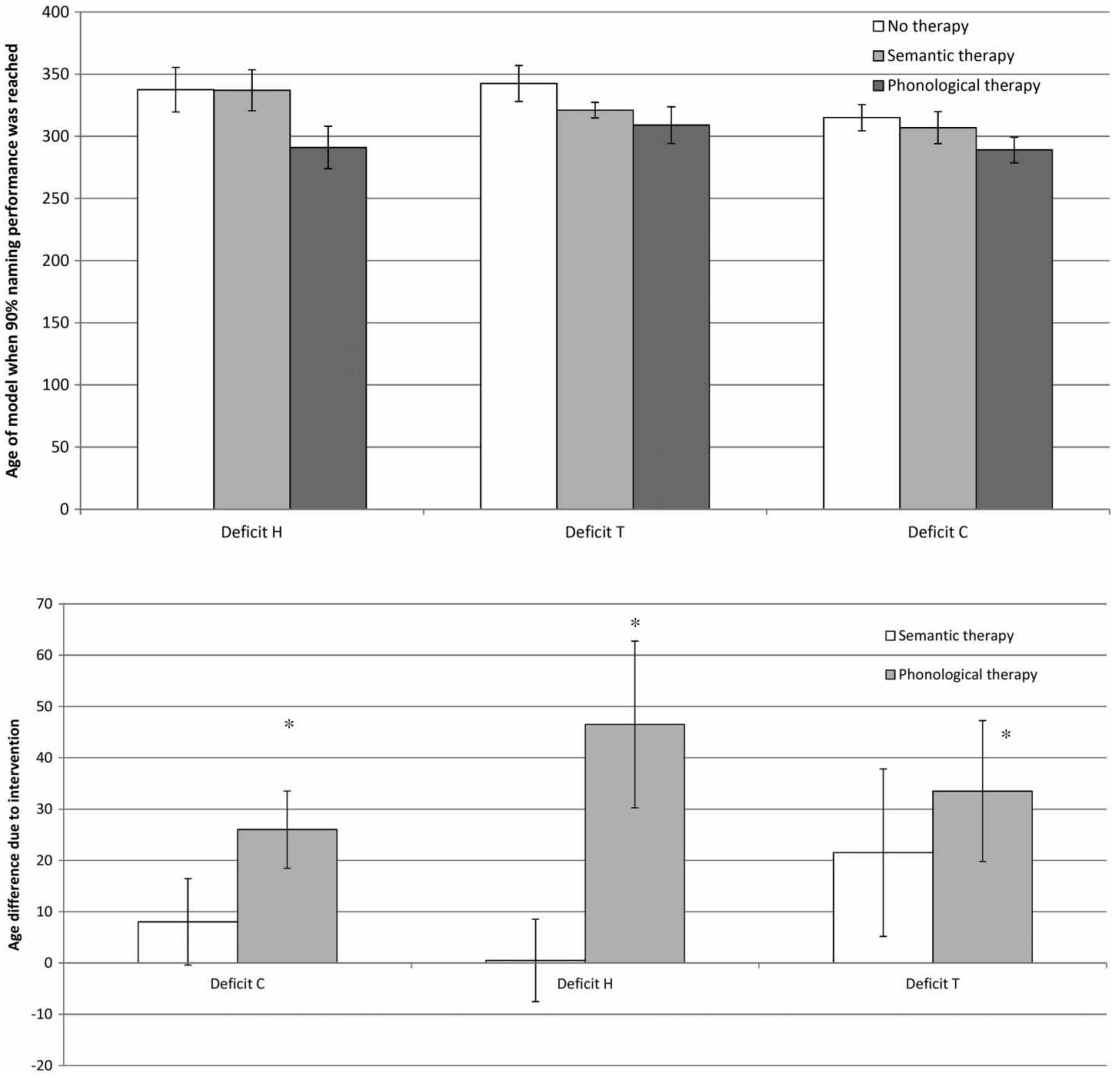


Figure 6. Simulating intervention for Amy. Mean and standard deviation of age difference of models (in epochs) when they reached 90% performance on the lexical-retrieval task with and without intervention, for the three double-deficit groups of models. Asterisks indicate effects that were significantly different from zero after a Bonferroni correction for multiple comparisons.

code (phonology). For the developmental deficits applied, only improvements in the receiving code had a marked effect.

Interventions had more diverse results on lexical retrieval in the case of simulated-Magda, the system with a triple location deficit. The data are shown in Figure 7. Here, the response depended to some extent on the nature of the initial computational deficit. Individual Bonferroni-corrected *t*-tests indicated that, in case of connectivity deficits, both intervention types were successful in improving lexical-retrieval performance. In the case of hidden unit deficits, neither of the interventions was successful. In the case of the temperature deficit, only the phonological intervention speeded up development significantly. Analysis of variance, somewhat compromised by the unequal variance between conditions, revealed

Downloaded by [92.11.185.75] at 05:45 17 September 2015

Table 6. Results of *t* tests comparing the effects of simulated semantic and phonological interventions on lexical-retrieval performance for the models of Amy and Magda.

Deficit	Case study	Intervention type	*t*	*df*	*p*	Mean diff.	Low 95% CI	Upper 95% CI
Connectivity	Amy	Semantic	1.50	9	.168	8.0	−4.1	20.1
		Phonological	5.46	9	*.000*	26.0	15.2	36.8
	Magda	Semantic	5.06	9	*.001*	33.5	18.5	48.5
		Phonological	6.63	9	*.000*	63.5	41.8	85.2
Hidden units	Amy	Semantic	0.098	9	.924	0.5	−11.0	12.0
		Phonological	4.53	9	*.001*	46.5	23.3	69.7
	Magda	Semantic	2.73	9	.023	63.0	10.8	115.2
		Phonological	2.26	9	.050	75.0	0.0	145.0
Temperature	Amy	Semantic	2.08	9	.067	21.5	−1.9	44.9
		Phonological	3.85	9	*.004*	33.5	13.8	53.2
	Magda	Semantic	3.11	9	.013	20.5	5.6	35.4
		Phonological	3.82	9	*.004*	40.5	16.5	64.5

Note: Results split by whether the underlying deficit was simulated by connectivity, hidden unit, or temperature manipulations. Six tests were carried out for each case study model. Bonferroni corrections therefore meant that *p*-values below .0083 were considered significant (marked by italics). CI = confidence interval; diff. = difference.

Downloaded by [92.11.185.75] at 05:45 17 September 2015

a main effect of intervention type, $F(1, 27) = 7.40$, $p = .011$, $\eta_p^2 = .215$, once more reflecting the advantage of the phonological intervention, but no main effect of deficit or interaction [$F(1, 27) = 1.34, p = .279, \eta_p^2 = .090; F(2, 27) = 0.47, p = .630, \eta_p^2 = .034$]. Here, with the additional computational restriction to the mapping pathway, improvements in both sending and receiving code could be effective depending on deficit type. Notably, in this second case, the many-to-one mapping of processing deficit to behavioural profile diverged into differential responses to intervention.

In sum, based on the computational model, our predictions for the response of these two children to intervention were that Amy would respond to the phonological intervention, but not the semantic intervention, while for Magda the predictions were less clear and depended on the nature of damage to the model.

An empirical test of the model's predictions of the respective effectiveness of semantic versus phonological interventions for the WFD case studies

The two girls with WFDs entered an intervention study designed to test the relative effectiveness of a semantic versus a phonological therapy. The study used a crossover design, whereby each child received both interventions (with a "washout" period in between), and naming skills were assessed before intervention and again after each intervention. Each girl therefore served as her own control. Both therapies utilized word-webs, where target words are elaborated and augmented with respect either to their meaning or to their component sounds. Therapy protocols were devised taking account of techniques used widely with children with WFDs, as well as approaches used successfully with adults with anomia as part of their aphasia (Boyle & Coelho, 1995; Coehlo et al., 2000; Leonard et al., 2008). To our knowledge, we are the first to publish experimental intervention research using the word-web approach.

Design

The two girls were first given pretherapy assessments that included naming 100 experimental items on three occasions prior to therapy. Multiple pretherapy assessments were employed because naming ability can be variable; it was necessary to establish baseline naming performance prior to intervention. The girls then participated in both therapies in a crossover design with a washout phase between interventions and were followed up to investigate maintenance of any effects. Each phase of the therapy (Therapy 1, washout, Therapy 2, and follow-up) lasted for half a term—

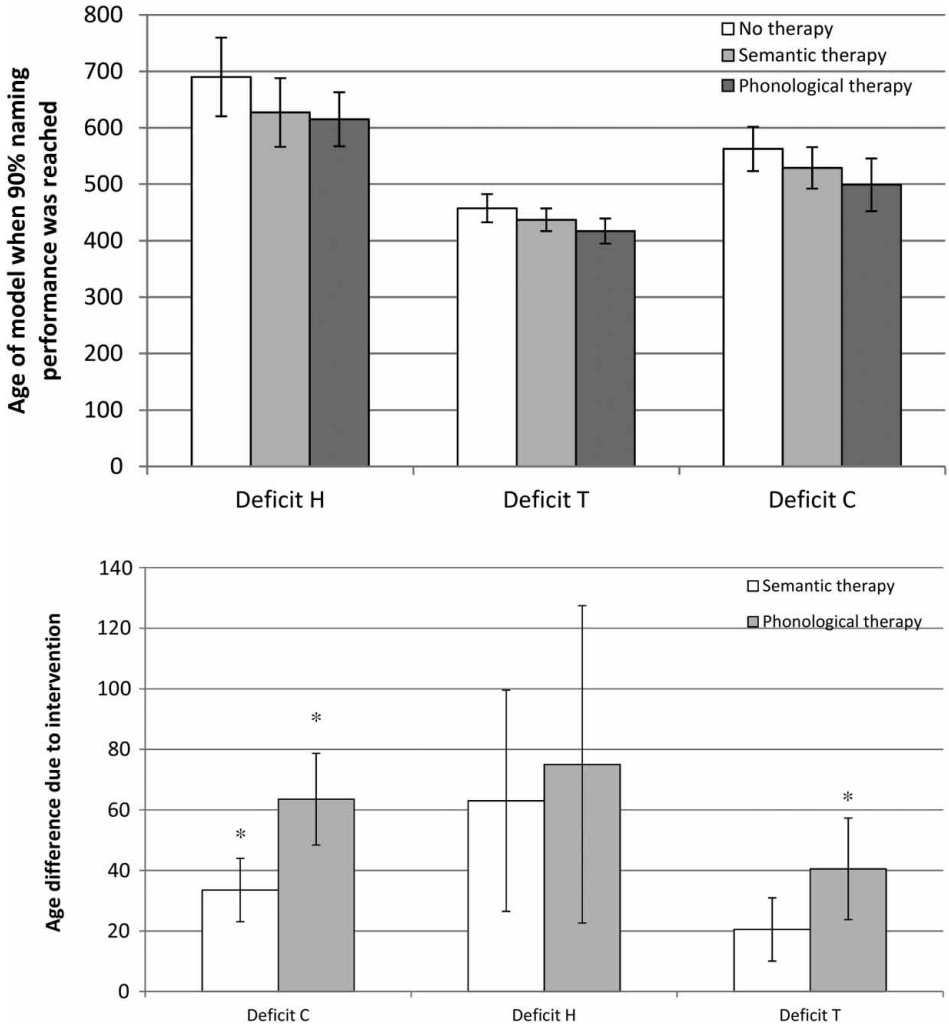


Figure 7. Simulating intervention for Magda. Mean and standard deviation of age difference of models (in epochs) when they reached 90% performance on the lexical-retrieval task with and without intervention, for the three multiple-deficit groups of models. Models in deficit group H never reached 90% so in the case of this group age was measured when the model reached 65% performance instead. Asterisks indicate effects that were significantly different from zero after a Bonferroni correction for multiple comparisons.

6 weeks. The design is illustrated in Figure 8. After each phase of the study the children were reassessed on naming all items. This assessment was carried out by a research associate working at a different institution who remained blind to the phase of the study and the order with which each girl experienced the interventions.

Therapy took place once a week for approximately 30 min, with each intervention block consisting of six sessions. Four sets of 25 words were matched for baseline picture naming accuracy and the following psycholinguistic variables: age of acquisition, log frequency, imageability, and visual complexity. A set of 25 was selected at random and treated in each therapy block, along with a further 6–12 nonexperimental words, which were selected by the children, teachers, and/or carers. Thus each child had different sets of experimental items

Downloaded by [92.11.185.75] at 05:45 17 September 2015

Downloaded by [92.11.185.75] at 05:45 17 September 2015

(within the 100) and of personally chosen items. At the start of each session, prior to therapy, children were asked to try and name pictures representing all of the above items, as well the control items (see below). Four different sequences of presentation were used alternately to control for order effects. The children were invited to press a comedy buzzer to pass on items that they were not able to name. This was to reduce frustration at being asked to repeatedly name items without feedback—especially naming control words, which were not treated.

The therapy blocks were designed to be as similar as possible, albeit one focusing on semantic attributes of the words and the other on phonological attributes. Template semantic and phonological word-webs are provided in Appendix 3, which also provides an overview of the therapy protocol. In the first phase of therapy (which typically covered Sessions 1 and 2), the therapist introduced the appropriate word-web and supported the child to "think around the word" together. The therapist used a series of prompt questions, derived from phonological components or semantic feature analysis, to encourage the child to generate features about an item (for example, a category in the semantic therapy, or number of syllables in the phonological therapy). If the child was unable to produce a target feature within 5 s, or gave vague or inappropriate information, the therapist provided a "forced choice"—for example, "Is it an animal or a vegetable?", or "Does it have 2 or 3 syllables/beats?". If the child was still unable to produce a feature within 5 s, the therapist gave the appropriate spoken information. The therapist wrote this on the word-web unless the child wished to draw or, more occasionally, to write the feature. As sessions progressed, the word-webs were used in games, with a barrier placed between therapist and child, designed to encourage communicative use of the target items. Throughout therapy, emphasis was placed on metalinguistic skills, encouraging the child to consider "what helps you when you can't find the word?", and in the barrier games, "what is the main thing about the word that would help me guess?".

Therapy items were treated in a continuous, cyclical order. Words named correctly at the start of the session were not targeted on that day. For both girls, an average of 4.7 experimental items were treated per session during the phonological therapy. During the semantic therapy, Amy worked on an average of 5.5 experimental items per session and Magda on 6.8. Length of therapy sessions remained constant throughout, regardless of how many items were covered. If a child offered spontaneous information, which was not directly targeted in therapy—for example, drawing the features or writing the word—this was neither inhibited nor encouraged. All sessions were video-recorded.

The primary outcome measure for the intervention was confrontation naming of the pictures. We also collected the girls' views of the intervention by interview and by their completion of a 5-point pictorial Likert scale with a member of research staff who had not been involved in the intervention. Finally, conversations with the girls were collected on three occasions using the guidelines in Appendix 4: twice prior to the start of intervention (approximately 2 months apart) and once at follow-up, after the girls had been involved in both interventions (approximately 8 months later). The conversations were transcribed and scored using the Profile of Word Errors and Retrieval in Speech (POWERS; Herbert, Best, Hickin, Howard, & Osborne, 2013) by team members blind to the date of each conversation. The conversation variable calculated for the present study—content words produced per conversational turn—was predicted to increase as a result of the intervention.

Results

The girls' naming over the course of the study is shown in Figure 9. Statistical analysis of single case and case series experimental designs (SCEDs) is an area of discord, and many authors simply employ visual inspection of the data over the course of the study (for a review see Smith, 2012). We followed both Smith-Lock, Leitao, Lambert, and Nickels (2013) in using the stringent McNemar nonparametric test, which takes into account items moving from correct to incorrect as well as in the desired direction, and Hickin, Best, Herbert, Howard, and Osborne (2002) in using

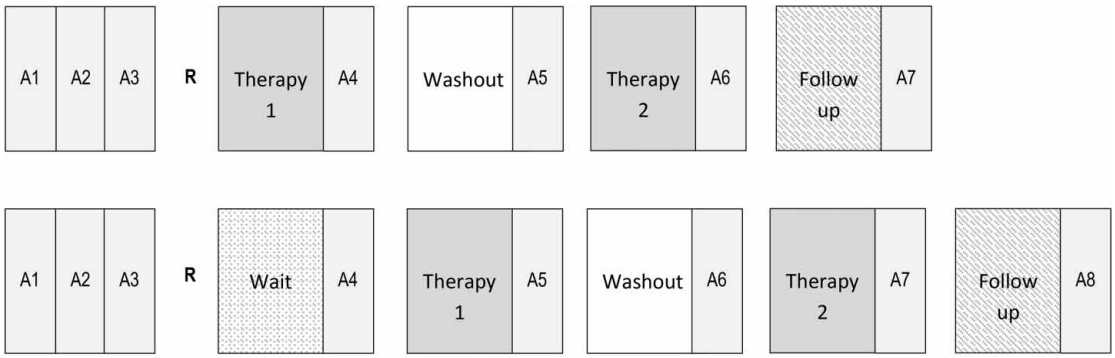


Figure 8. Design of the intervention study. A1 to A8 represent assessments. R denotes randomization. The baseline assessments were carried out over the half term prior to the intervention. As part of the larger intervention study (Best et al., 2013), children were randomly allocated a "wait" period before starting the intervention (as Magda was; see Figure 9) and given an additional baseline assessment immediately prior to the start of therapy. The wait period is not relevant to the current results, but we include it here for consistency with later data. Each phase of the study is represented by a rectangle (wait, therapy, washout, and follow-up) and lasted for 6 weeks (half a school term). The assessment following each phase was carried out as soon as possible thereafter (i.e., on a later day in the final week of half term, in the seventh week of a longer half term, or, less usually, during the school holiday). Both Amy and Magda received the phonological condition for Therapy 1 and the semantic condition for Therapy 2.

Downloaded by [92.11.185.75] at 05:45 17 September 2015

statistics weighted according to the phase of the study to test specific hypotheses about change. The McNemar tests and weighted statistics were used to address different questions. The McNemar tests compared performance at only two time points, while the weighted statistics addressed questions about change across the whole course of the study, with the selected weights testing hypotheses about possible profiles of change.

McNemar tests

We first tested whether the girls' naming of the items improved with each type of therapy. This was done separately for the treated items ($n = 25$) and for the untreated items ($n = 75$), in each case making a comparison between naming just prior to and immediately after each intervention. We used one-tailed tests as we predicted improvement, employing a cut-off of $p < .05$. For treated items, Amy showed significant benefit from the phonological therapy but not from the semantic approach. (*treated set*: phonological 0.56→0.88, $p = .011$, significant; semantic 0.56→0.76, $p = .063$, not significant). There was no significant change on untreated items following either intervention with Amy (*untreated set*: phonological 0.53→0.57, $p = .254$, *ns*; semantic 0.68→0.65, $p = .363$, *ns*). Magda showed no significant benefit from the phonological intervention but naming of the treated set benefited significantly from the semantic approach (*treated set*: phonological 0.44→0.48, $p = .5$, *ns*; semantic 0.60→0.92, $p = .004$, significant; *untreated set*: phonological 0.41→0.47, $p = .172$, *ns*; semantic 0.48→0.45, $p = .377$, *ns*).

We also tested for improvement over the course of both interventions together, again using one-tailed tests, by comparing the final pretherapy baseline score with naming immediately after the second phase of therapy on all items. Both girls made significant progress (Amy 0.54→0.68, $p = .001$, significant; Magda 0.42→0.6, $p = .002$, significant). Lastly, we compared final posttherapy naming performance with follow-up half a term later. In this case we used two-tailed tests as naming may have continued to improve or dropped off after interventions ended. Neither of the girls showed significant drop-off post therapy (Amy 0.68→0.62, $p = .146$, *ns*; Magda 0.57→0.60, $p = .774$, *ns*).

Weighted statistics

We weighted the girls' naming at each assessment to test four different hypotheses (Howard, Best, & Nickels, 2015). The weightings differ for the two

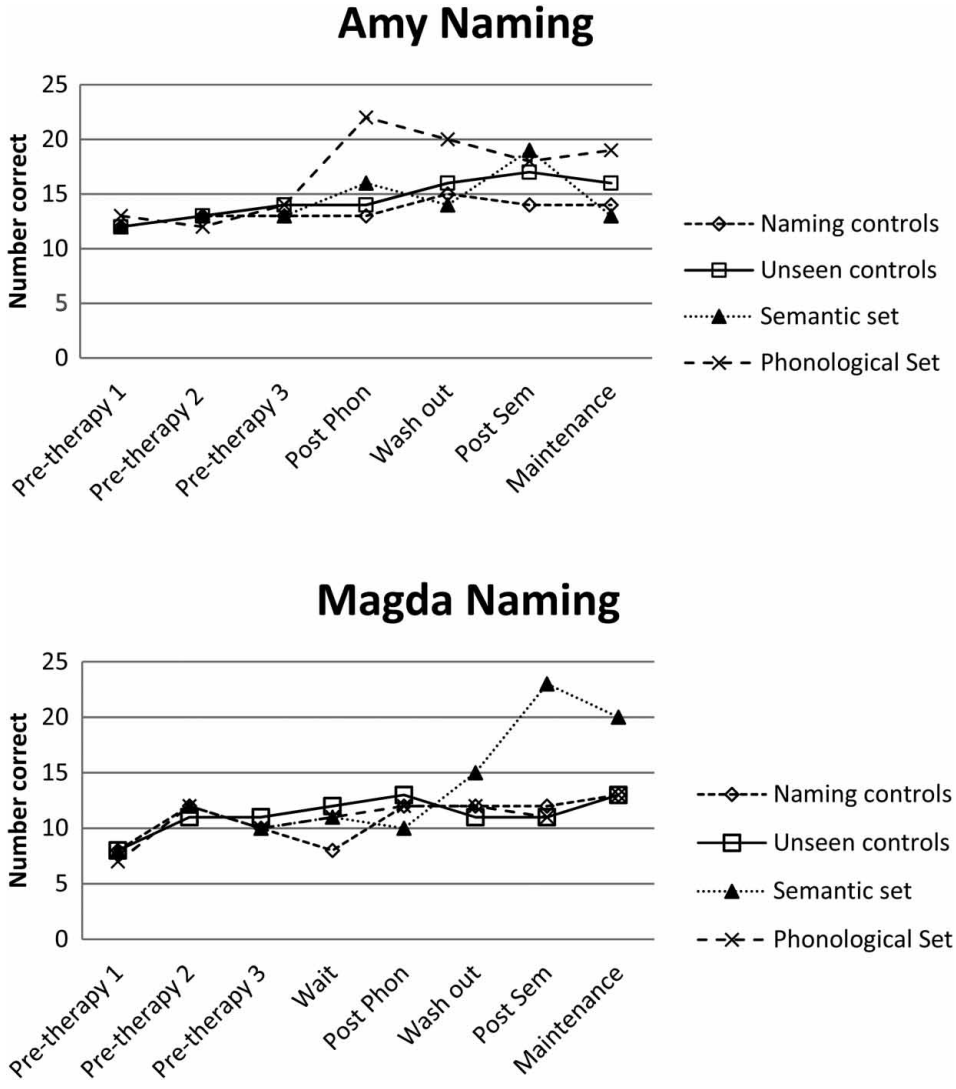


Figure 9. Naming over the course of the study. The girls' picture naming accuracy on the four experimental sets of 25 items at each assessment. Note that Magda has four pretherapy baselines as she was randomly assigned to the "wait" condition.

girls as Magda was assigned to the "wait" condition prior to starting therapy and thus had four pretherapy baselines, whereas Amy had three pretherapy baselines (see Figure 9). We used one-tailed tests throughout; the full weightings are provided in Appendix 5.

Hypothesis 1: First we looked for an overall trend for improvement over the course of involvement in the study. Both girls' naming demonstrated this ($n = 100$, Amy, $t = 4.31$, $p < .001$, significant; Magda, $t = 5.23$, $p < .001$, significant). However, this change may simply have reflected development and not be due to the interventions.

Hypothesis 2: We tested whether there was greater change during the therapy phases of the study than over the other phases (baseline, washout, and follow-up), both for the whole set and for treated items only. Neither girl showed significantly greater change during intervention on the

Downloaded by [92.11.185.75] at 05:45 17 September 2015

Downloaded by [92.11.185.75] at 05:45 17 September 2015

whole set ($n = 100$, Amy $t = 1.04$, $p = .151$, *ns*; Magda $t = 0.76$, $p = .225$, *ns*), while both showed significantly greater change on the treated items during intervention phases of the study than the remainder ($n = 50$, Amy, $t = 2.56$, $p = .007$, significant; Magda, $t = 2.05$, $p = .023$, significant).

Hypothesis 3: We also tested whether there was a different effect of the two treatments on all items. There was a significant difference for Amy, with improvement following phonological but not semantic therapy, but interestingly no significant difference for Magda ($n = 100$, Amy $t = 2.40$, $p = .009$, significant; Magda, $t = 0.70$, $p = .244$, *ns*).

Hypothesis 4: Finally we tested whether there was greater change during therapy for the subsets of items used in the different interventions. The findings support those from the McNemar tests. For Amy there was greater change during therapy than during nontherapy phases for the set treated with the phonological intervention, and this was not the case for the set given semantic therapy ($n = 25$, phonological $t = 3.29$, $p = .002$, significant; semantic $t = 0.44$, $p = .333$, *ns*). For Magda the reverse was true ($n = 25$, phonological $t = –0.15$, $p = .559$, *ns*; semantic $t = 3.09$, $p = .003$, significant). Furthermore, for Amy there was significantly greater improvement from phonological than semantic intervention for the treated sets ($n = 25$, $t = 2.02$, $p = .025$, significant). For Magda, there was significantly greater improvement from the semantic than phonological intervention for the treated sets ($n = 25$, $t = 2.34$, $p = .012$, significant).

In sum, Amy showed a significant change in naming of both treated items and the whole set following the phonological therapy but there was no significant difference following the semantic therapy. Magda showed a significant change in naming of the treated items following the semantic therapy, though her improvement on all items fell just outside statistical significance; she did not show significant gains from the phonological therapy.

Two wider outcomes of the therapy were assessed: the children's own views of the therapies, and the effects of intervention on the children's word finding with a conversational context rather than the artificial situation of picture naming tasks. With respect to the girls' own views, the children were asked to rate which aspects of the research interventions they perceived as most helpful to them using a 5-point pictorial Likert scale, with 5 at the positive end of the scale. The results are summarized in Table 7. The girls were asked how helpful it was to think about the meaning in words (i.e., semantic therapy), versus the sounds in words (i.e., phonological therapy).

Notably, some results were in the opposite direction to the effectiveness of the interventions on the children's naming. The most effective therapy was reported as least helpful, and vice versa. However, fine-grained responses were more consistent: When asked "What helps you most when you are stuck in finding words?", Amy cited a strategy worked on during phonological therapy (the most effective of the two interventions for her). She described this as "chunking it out"—that is, breaking down longer words into shorter, more memorable parts. Meanwhile, Magda's response "I show someone the action" cites an

Table 7. Children's views of the intervention and outcome for them.

Question	Amy	Magda
How much did you enjoy taking part in WORD?	5	5
How helpful was it to think about the MEANING of words?	4	3
How helpful was it to think about the SOUNDS in words?	3	5
What helps you most when you are stuck?	Chunking out; doing the actions; sometimes spelling.	I show someone the action. . . . Tell a teacher or friend.
Do you think finding words is easier now?	At the beginning 1 and now it is 3.	A little bit easier

Note: Scores on a 5-point pictorial Likert scale, with 5 at the positive end of the scale.

Table 8. Conversation, scored using POWERS.

	Amy			Magda		
	Pretherapy		Posttherapy	Pretherapy		Posttherapy
Conversational measure	C1	C2	C3	C1	C2	C3
Content words per child's turn	3.52	7.56	13.19	5.80	5.32	11.00

Note: POWERS = Profile of Word Errors and Retrieval in Speech (Herbert et al., 2013).

Downloaded by [92.11.185.75] at 05:45 17 September 2015

idea not directly targeted in therapy, but which formed part of the semantic intervention and which she used spontaneously with some success to help get her message across during conversation. Both children rated their enjoyment of the project as 5 (the maximum score). Magda stated that finding words was "a little bit easier" at the end of the study, while Amy spontaneously used numerical ratings to illustrate her perceived progress: "At the beginning it was 1, and now it is 3".

With respect to word finding within conversation, the data exploring word retrieval are provided in Table 8. Amy showed a gradual increase in the number of content words (mainly nouns and verbs) that she produced per conversational turn from the first pretherapy to second pretherapy to posttherapy sessions—that is, there was a trend to change over the study and no clear change that could be attributed to intervention. For Magda the pretherapy conversations were at a similar level to one another. The posttherapy conversation showed a dramatic change, with around twice as many content words produced per turn.

Comparison to model predictions

For Amy, the model fitted to her behavioural profile on the four core tasks predicted that an intervention focusing on phonology would be effective, but an intervention focusing on semantics would not. This was confirmed by the subsequent intervention study. For Madga, the model fitted to her behavioural profile differed in its prediction depending on the underlying processing deficit. The results of the behavioural intervention study showed that Magda benefited from the semantic, but not the phonological, intervention.

Discussion

We used a computational model to generate predictions at the case study level on appropriate interventions for lexical-retrieval deficits and, by testing these predictions on the actual case studies, to evaluate the model. Using an artificial neural network architecture, we sought to model vocabulary acquisition as a process of learning to map between emerging internal representational codes of semantic and phonological knowledge of words, and atypical vocabulary acquisition as the operation of this process under atypical computational constraints. We fitted two atypical models to two case studies of 7-year-old girls with word-finding difficulties and then simulated interventions on these models, targeting the improvement of either semantic or phonological knowledge and assessing which intervention would be more effective in improving naming ability. When the effectiveness of the respective interventions was evaluated on the girls themselves, the prediction of the model was borne out in one case, but not in another. In this section, we consider the limitations of the model, the innovations and limitations of the intervention study, and the implications of our approach for the use of intervention studies to advance theory.

Evaluating the model

We can evaluate the model in two ways. First, how good was it as a model of intervention? Second, what could we infer about the model and its assumptions based on the test of its predictions regarding which intervention would be more successful for each case study?

The aim of using a computational model not only to simulate the individual profiles of children

Downloaded by [92.11.185.75] at 05:45 17 September 2015

with a developmental disorder but also to predict the outcomes of intervention was an ambitious one, since little computational work to date has investigated interventions in atypical development (though see Harm et al., 2003, for a notable exception). Our model was therefore fairly simplified with respect to the ecological validity of the vocabulary on which it was trained: The vocabulary size was small (100 words), semantics was restricted to abstract sparse binary features embodying categories with a prototype structure, and phonology was restricted to strings of phonemes without realistic phonological neighbourhoods. The model focused on the developmental process of associating emerging representational codes (in contrast to the naming model of Dell et al., 1997) and permitted these representational codes to be high dimensional (in contrast to, e.g., the naming models of Li et al., 2004, 2007; Mayor & Plunkett, 2010). This is because our theoretical focus was the underlying cause of WFDs and to clarify hypotheses that these might involve atypical formation of category boundaries in either semantics or phonology. We explored how different types of computational processing deficit applied to different locations within the model's architecture prior to development would impact on the emergence of representations and the behaviour that they drove, as well as the subsequent response to intervention.

Two notable theoretical findings resulted. First, the model was able to qualitatively simulate the profiles of the two case studies, Amy and Magda, as instances of atypical development, producing broadly similar profiles between model and children on four core tasks (picture naming, picture–word verification, nonword repetition, and picture matching). Each model required multiple deficits to produce the relevant profile, consistent with the view that WFDs do not involve highly circumscribed deficits. Different computational processing deficits, including changes to connectivity, changes to the number of internal processing units, and changes to the activation function (sensitivity) of processing units, all produced similar atypical profiles. In that sense, there was a many-to-one mapping between underlying processing deficits and behaviour profiles. Second, in the case of simulated-Amy, the response to intervention of the different processing deficits was similar; but in the case of Magda, the response to intervention diverged across the different processing deficits. The implication is that behavioural profiles may not be uniquely predictive of response to intervention.

The model also had two significant shortcomings. First, in our simplification of the vocabulary, the difficulty of acquiring semantic and phonological knowledge was not sufficiently closely matched, meaning the development of the model was limited by the emergence of semantic knowledge. This will be addressed in our future work in moving to a larger, more realistic training vocabulary.

Turning to the empirical evaluation of the model's predictions on which interventions would be more effective for each case study, the model was successful in predicting the effects of intervention for simulated-Amy (that the phonological intervention would be effective but the semantic would not). However, it was not successful for simulated-Magda. The model made different predictions, depending on processing deficit. Two of the simulated deficits predicted that the phonological intervention would be effective. In reality, only the semantic and not the phonological intervention was successful. The key question is what this tells us about the model. For Amy, the model indicated that the performance of the phonological component was the limiting factor on lexical retrieval. When a deficit in this component was alleviated by extra training, performance improved. The empirical data supported this view. For Magda, the model indicated that phonology was again a limiting factor, but that improvement in semantics could also help, as the pathway between the components was compromised—improvements in both sending and receiving codes could help overcome the limitations of the associative pathway. By contrast, the empirical data for Magda indicate that phonology was not the limiting factor on her word retrieval, despite deficits in nonword repetition. Rather, the limiting factor was in semantic processing, a deficit that did not greatly impact on semantic picture judgement and word–picture verification accuracy to the

Downloaded by [92.11.185.75] at 05:45 17 September 2015

expected extent. This implies that the semantic code served more poorly as the input to a process (the semantics-to-phonology mapping required in lexical retrieval) than it did as the output of a process (the phonology-to-semantics mapping required in lexical comprehension) or in a task requiring only semantic input processing (picture judgement). One property that can make codes particularly poor as inputs is that they are too similar to (or confusable with) each other. This might explain the greater response times Madga demonstrated in the picture judgement task, despite her similar accuracy level to Amy. Under this interpretation, then, in contrast to our simulations, Magda may have a different type of deficit in semantics from Amy, a type of deficit that limits lexical retrieval, and while Magda has phonological deficits, these in turn do not appear to limit lexical retrieval to the same extent. We note this is one interpretation of the data, which could be further investigated by both behavioural testing and computational modelling. Finally, then, the implication of the intervention results for the model is that the current version was unable to capture the potential confusability of emerging semantic representations in the case of Magda, due either to the lack of ecological validity of the training set, or to the manipulations used to create atypical representations. These must be the target of an improved model.

Use of four core tasks

We assessed 20 TD children close in chronological age to the two girls with WFDs. This approach means that the relationship between performance on the different tasks can be compared with confidence, as the same TD children were tested rather than different samples for each task. There was considerable variability within the TD sample, and this is important as it reflects the children's different points of their developmental trajectories.

The inclusion of two bespoke tasks, PJs (picture judgement) and WPVT (word–picture verification), provided new and appropriate measures of aspects of processing relevant to lexical retrieval in children that have not been measured sensitively in the past. They are particularly useful as WFDs frequently occur in the presence of wider language difficulties including expressive language. In contrast with other tasks, such as providing definitions for items the children are unable to name, neither task required spoken output, meaning they can be employed to find areas of relative strength (such as Amy's typical response times for PJs). They enabled us to demonstrate that WFDs may have multiple causes within a single child, and this understanding was supported by the findings from the model, which was unable to match the girls' performance by disrupting just one module.

Evaluating the intervention study

We carried out a tightly experimentally controlled intervention study that demonstrated an effect of therapy over development on Amy's and Magda's word finding as measured by picture naming. The intervention was based on techniques commonly used clinically with children and with adults with anomia as part of their aphasia (Boyle & Coelho 1995; Leonard et al., 2008). There are several clinical resources, which employ related approaches with children (e.g., Commtap Speech and Language Therapy Activities, n.d.; Word Whizzer, n.d.), but this is the first study to our knowledge to test the use of word-webs, focusing separately on semantic and phonological features, experimentally with children.

Strengths of the intervention study included the guidance of experienced therapists in intervention selection and development, the use of a cross-over design where each child served as her own control, assessment by a researcher blind to assignment, pretherapy baseline matching specific to each child's naming, relatively large item sets, and the inclusion of naming controls and personally chosen items to supplement the experimental set. The main weakness of the design was that within the intervention, possible order effects of the phonological intervention preceding the semantic intervention could not be discounted, though the girls showed a differential response to the same order.

The main finding was greater change during therapy than during the other phases of the study,

Downloaded by [92.11.185.75] at 05:45 17 September 2015

demonstrating an effect of intervention over and above development. On treated items, Amy, who had particular difficulty with assessment tasks requiring phonological output, benefited from the intervention highlighting phonological properties of target words but did not benefit from the semantic intervention. In contrast, Magda, with wider language difficulties including with semantic and phonological output processing, benefited from semantic intervention, but her naming did not benefit significantly from the phonological intervention. The effects of intervention were largely specific to treated items, although Amy did improve on the set as a whole after the phonological intervention. The effects of therapy maintained for at least half a term. Overall, the results fit with related research indicating that both semantic (Ebbels et al., 2012) and phonological (Wing, 1990) approaches can be used successfully in helping children with WFDs.

One implication for clinical decision-making could be that therapy resources may be best directed at areas of *need*, rather than areas of relative *strength*, for children with WFDs. This would fit well with Amy benefiting from the phonological intervention and Magda from the semantic approach. However, given that Magda also had difficulty with phonological output tasks, on this account, we would need to explain why she did not benefit from the phonological therapy (as, indeed, the model predicted she would). One possibility is that in order to benefit from the phonological intervention, Magda would first need to have better established semantic representations for the target items. The sequential design of our behavioural intervention study did not allow this to be investigated.

One key aspect of the therapy approach is to encourage the children to use strategies that they can employ when learning new words in the future. The demands for new word learning and retrieval are considerable with children. Once a child enters school in the UK, she or he is exposed to about 10,000+ new words each year and adds approximately 3000 of these words per year to their productive vocabulary (Nagy & Anderson, 1984; Nagy & Herman, 1987). While the effects were clear for treated items, in line with the intervention literature, the evidence for wider generalization to untreated items and to conversation remains unclear. Interestingly, both girls reported that word finding was easier after the study and, in line with the persisting nature of language difficulties, were aware of strategies that helped them. The increase in Magda's use of content words in conversation is suggestive of wider changes but this cannot be attributed unequivocally to the intervention rather than development.

Lastly, regarding their own views of involvement and change, when the girls rated how helpful each intervention was for them, neither chose the approach that most improved their naming. A possible explanation for this is that both favoured the activities they found easiest, rather than those targeting their core difficulty. When asked what helped her retrieve words, Amy suggested a strategy that was part of the phonological intervention. We concurred that this was a useful tool for her, based on her response to therapy tasks. Magda suggested gesture, which she used in conversation prior to therapy and which was part of the semantic approach that aided her word retrieval. It is encouraging that both children rated their enjoyment of the project at the top of the scale, given the relative severity of their word-finding difficulties and the frequent requirement for them to attempt to name hard-to-find words. Both accurately reflected that they had made some progress with their naming post therapy, while acknowledging the persistent nature of their difficulty.

Implications of using computational modelling as a bridge between intervention and theory

Theories of the causes of developmental deficits typically stem from correlational data. In the case of WFDs, these include the association of WFDs with poorer performance on tasks testing phonological knowledge or testing semantic knowledge. Two main methods go beyond correlation to test causality: the use of longitudinal designs, to establish, for instance, that developmentally, phonological or semantic deficits preceded WFDs; or the use of intervention designs, to establish, for instance, that alleviating phonological or semantic deficits serves to improve WFDs. The latter was pursued

Downloaded by [92.11.185.75] at 05:45 17 September 2015

here, with a computational model of vocabulary acquisition serving as a bridge between the causal theory and the intervention study.

Computational models provide two clear benefits in the current context. First, they more appropriately conceptualize the behavioural impairments in terms of the emerging consequence of a developmental process that is taking place under atypical processing constraints, rather than as deficits applied directly to a static model. This is essential in theories of developmental disorders (Thomas & Karmiloff-Smith, 2002). Second, by implementation, they force much needed clarification on theoretical explanations and predictions. For example, what exactly happens according to your theory when you intervene? How much does the content or detailed nature of the information presented in the domain of intervention matter? When researchers suggest that semantic or phonological representations might be impoverished or poorly specified in children with WFDs, this presumably implies that the boundaries between representations of different word meanings or word sounds are inadequate. But what is a category boundary like in semantic or phonological space, typical or atypical, and how responsive is it to changes at different ages? Moreover, where representational codes differ in atypical cases, implementation focuses consideration on the task the codes are to be used for: Representations may be adequate to drive one behaviour but not to drive another.

The greater clarity forced by implementation does, however, come at the cost of simplification. In the current case, one prediction of the model regarding the effect of intervention on a given atypical model was not supported. This implies that the theory that the model embodies cannot be correct and must be altered. However, simplification means one must also determine the extent to which the model embodies the theory that it is implementing, or whether the disparity stems from simplifying assumptions (e.g., in the current case, the ecological validity of the training set). Moreover, these concerns are separate from broader simplifications regarding the therapy process (including aspects of social interaction, attention, motivation) and the child's potential response to it.

However, the potential benefits of a successful computational model make it an enterprise worthy of pursuit. As the model becomes able to predict more reliably the interventions that best remediate impairments in children with different profiles, so confidence increases that the model may be used as a tool to facilitate understanding of the intervention process.

Acknowledgements

We acknowledge the important contributions of our wider advisory group (including an educational psychologist and a young person with speech and language difficulties) on the design of the behavioural study, and of our Clinical Advisory Group (including Susan Ebbels, Kathleen Cavin, and Sarah Simpson) for shaping the therapies used and for more general advice. Mike Coleman constructed the reaction time (RT) version of the Picture Judgement Task. The randomization and the weighted statistics were kindly provided by David Howard, Newcastle University. University College London (UCL) MSc student Elisabeth Salt contributed to the analysis of the conversation data.

We are greatly indebted to the children who took part so willingly in the study and to their parents, teachers, and schools and to speech and language therapist Vivien Gibson who carried out some of the therapy.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by Economic and Social Research Council (ESRC) [grant number RES-062-23-2721].

Note

1. Names have been changed for the purposes of anonymity.

References

Abel, S., Willmes, K., & Huber, W. (2007). Model-oriented naming therapy: Testing predictions of a connectionist model. *Aphasiology*, *21*(5), 411–447.

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

Best, W., Masterson, J., Thomas, M. S. C., Hughes, L., Fedor, A., Roncoli, S., & Kapikian, A. (2013, June 24–25). The WORD Project: A case series study on intervention for word-finding difficulties. Presented at the Child Language Seminar 2013, Manchester.

Bishop, D. (1989). *Test for the reception of grammar (TROG)*. Oxford: Pearson.

Boyle, M. (2004). Semantic feature analysis treatment for anomia in two fluent aphasia syndromes. *American Journal of Speech-Language Pathology*, *13*(3), 236–249.

Boyle, M., & Coehlo, C. A. (1995). Application of semantic feature analysis as a treatment for aphasic dysnomia. *American Journal of Speech-Language Pathology*, *4*, 94–98.

Bragard, A., Schelstraete, M.-A., Snyers, P., & James, D. G. H. (2012). Word-Finding intervention for children with specific language impairment: A multiple single-case study. *Language, Speech, and Hearing Services in Schools*, *43*, 222–234.

Coehlo, C., McHugh, R. E., & Boyle, M. (2000). Semantic feature analysis as a treatment for aphasic dysnomia: A replication. *Aphasiology*, *14*(2), 133–142.

Commtap Speech and Language Therapy Activities. (n.d). Retrieved January 23, 2015, from http://en.commtap.org/

Constable, A., Stackhouse, J., & Wells, B. (1997). Developmental word finding difficulties and phonological processing: The case of the missing handcuffs. *Applied Psycholinguistics*, *18*, 507–536.

Dell, G. S., Faseyitan, O., Nozari, N., Schwartz, M. F., & Coslett, H. B. (2013). Voxel-based lesion-parameter mapping: Identifying the neural correlates of a computational model of word production. *Cognition*, *128*(3), 380–396.

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

Dockrell, J. E., Messer, D., George, R., & Ralli, A. (2003). Beyond naming patterns in children with WFDs—Definitions for nouns and verbs. *Journal of Neurolinguistics*, *16*, 191–211.

Dockrell, J. E., Messer, D., George, R., & Wilson, G. (1998). Notes and discussion children with word-finding difficulties-prevalence, presentation and naming problems. *International Journal of Language & Communication Disorders*, *33*(4), 445–454.

Druks, J., & Masterson, J. (2000). *Object and action naming battery*. Hove: Psychology Press.

Dunn, L. M., Dunn, D. M., & Styles, B. (1997). *British picture vocabulary scale III*. Windsor: NFER-NELSON.

Ebbels, S. H., Nicoll, H., Clark, B., Eachus, B., Gallagher, A. L., Horniman, K., … Turner, G. (2012). Effectiveness of semantic therapy for word-finding difficulties in pupils with persistent language impairments: A randomized control trial. *International Journal of Language & Communication Disorders*, *47*(1), 35–51.

Elliot, C. D., Smith, P., & McCullouch, K. (1996). *British ability scale II edition*. Windsor: NFER-NELSON.

Faust, M., Dimitrovsky, L., & Davidi, S. (1997). Naming difficulties in language-disabled children: Preliminary findings with the application of the tip-of-the-tongue paradigm. *Journal of Speech, Language, and Hearing Research*, *40*, 1026–1036.

Fedor, A., Best, W., Masterson, J., & Thomas, M. S. C. (2013, July 31–August 3). *When do behavioural interventions work and why? Towards identifying principles for clinical intervention in developmental language disorders from a neuro-computational perspective*. Poster presented at *35th* Annual Meeting of the Cognitive Science Society, Berlin, Germany. Retrieved from http://www.psyc.bbk.ac.uk/research/DNL/personalpages/Fedor_Cogsci_Berlin13pdf.pdf

Forster, K. I., & Forster, J. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Reasearch Methods, Instruments, & Computers*, *35*(1), 116–124.

Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production.

Downloaded by [92.11.185.75] at 05:45 17 September 2015

Downloaded by [92.11.185.75] at 05:45 17 September 2015

Journal of Memory and Language, *43*(2), 182–216.

Frederickson, N., Frith, U., & Reason, R. (1997). *Phonological assessment battery*. Windsor: NFER Nelson.

Funnell, E., Hughes, D., & Woodcock, J. (2006). Age of acquisition for naming and knowing: A new hypothesis. *The Quarterly Journal of Experimental Psychology*, *59*(2), 268–295.

Gathercole, S., & Baddeley, A. (1996). *The children's test of nonword repetition*. London: Psychological Corporation.

German, D. (2000). *Test of word finding* (2nd ed. (TWF-2)). San Antonio: Pearson.

Gray, S. (2005). Word learning by preschoolers with specific language impairment: Effect of phonological or semantic cues. *Journal of Speech, Language and Hearing Research*, *48* (6), 1452–1467.

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.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*, 662–720.

Herbert, R., Best, W., Hickin, J., Howard, D., & Osborne, F. (2013). *POWERS profile of word errors and retrieval in speech*. Guildford: J & R Press.

Hickin, J., Best, W., Herbert, R., Howard, D., & Osborne, F. (2002). Phonological therapy for word finding difficulties: A re-evaluation. *Aphasiology*, *16*(10), 981–999.

Howard, D., Best, W., & Nickels, L. (2015). Optimising the design of intervention studies: Critiques and ways forward. *Aphasiology*. Advance online publication. doi:10.1080/02687038.2014.985884.

Howard, D., & Patterson, K. (1992). *The pyramids and palm trees test*. Oxford: Pearson.

Hyde Wright, S. (1993). Teaching word-finding strategies to severely language-impaired children. *International Journal of Language and Communication Disorders*, *28*, 165–175.

Hyde Wright, S., Gorrie, B., Haynes, C., & Shipman, A. (1993). What's in a name? Comparative therapy for word-finding difficulties using semantic and phonological approaches. *Child Language, Teaching and Therapy*, *9*, 214–229.

Leonard, C., Rochon, E., & Laird, L. (2008). Treating naming impairments in aphasia: Findings from a phonological components analysis treatment. *Aphasiology*, *22*(9), 923–947.

Li, P., Farkas, I., & MacWhinney, B. (2004). Early lexical development in a self-organizing neural network. *Neural Networks*, *17*, 1345–1362.

Li, P., Zhao, X., & MacWhinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cognitive Science*, *31*(4), 581–612.

Lorenz, A., & Ziegler, W. (2009). Semantics vs. word-form specific techniques in anomia treatment: A multiple single-case study. *Journal of Neurolinguistics*, *22*(6), 515–537.

Martin, N., & Brownell, R. (2005). *Test of auditory processing* (3rd ed. (TAPS-3)). Novato: Academic Therapy Publications.

Massaro, M., & Tompkins, C. A. (1994). Feature analysis for treatment of communication disorders in traumatically brain-injured patients: An efficacy study. *Clinical Aphasiology*, *22*, 245–256.

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. (1994). The use of phonological information in a word-finding treatment for children. *Journal of Speech and Hearing Research*, *37*, 1381–1393.

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

McGregor, K. K., Newman, R. M., Reilly, R., & Capone, N. C. (2002). Semantic representation and naming in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *45*, 998–1014.

McGregor, K. K., & Windsor, J. (1996). Effects of priming on the naming accuracy of preschoolers with word finding deficits. *Journal of*

Speech, Language, and Hearing Research, *39*, 1048–1058.

Nagy, W., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly*, *19*, 304–330.

Nagy, W., & Herman, P. A. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M. McKeown & M. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 19–59). Hillsdale, NJ: Erlbaum.

Nickels, L. (1997). *Spoken word production and its breakdown in aphasia*. Hove: Psychology Press.

Nickels, L. (2002). Therapy for naming disorders: Revisiting, revising, and reviewing. *Aphasiology*, *16*(10–11), 935–979.

Plaut, D. C. (1996). Relearning after damage in connectionist networks: Toward a theory of rehabilitation. *Brain and Language*, *52*, 25–82.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*(1), 56–115.

Plunkett, K., Sinha, C., Moslashller, 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.

Powell, D., Stainthorp, R., Stuart, M., Garwood, H., & Quinlan, P. (2007). An experimental comparison between rival theories of rapid automatized naming performance and its relationship to reading. *Journal of Experimental Child Psychology*, *98*(1), 46–68.

Protopapas, A. (2007). CheckVocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, *39*(4), 859–862.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical evaluation of language fundamentals* (4th ed). San Antonio: Pearson.

Smith, J. D. (2012). SCEDs: A systematic review of published research and current standards. *Psychological Methods*, *17*(4), 510–550.

Smith-Lock, K. M., Leitao, S., Lambert, L., & Nickels, L. (2013). Effective intervention for expressive grammar in children with specific language impairment. *International Journal of Language & Communication Disorders*, *48* (3), 265–282.

Thomas, M. S. C. (2005a). Characterising compensation. *Cortex*, *41*(3), 434–442.

Thomas, M. S. C. (2005b). Constraints on language development: Insights from developmental disorders. In P. Fletcher & J. Miller (Eds.), *Language disorders and developmental theory* (pp. 11–34). Philadelphia: John Benjamins.

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

Thomas, M. S. C., & Karmiloff-Smith, A. (2003). Modelling language acquisition in atypical phenotypes. *Psychological Review*, *110*(4), 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*, 467–483.

Van Hees, S., Angwin, A., McMachon, K., & Copland, D. (2013). A comparison of semantic feature analysis and phonological components analysis for the treatment of naming impairments in aphasia. *Neuropsychological Rehabilitation*, *23*(1), 102–132.

Wing, C. S. (1990). A preliminary investigation of generalization to untrained words following two treatments of children's word-finding problems. *Language, Speech, and Hearing Services in Schools*, *21*, 151–156.

Word Whizzer. (n.d.). Retrieved January 23, 2015, from http://www.wordwhizzer.com/wordwhizzer.htm

Zens, N. K., Gillon, G. T., & Moran, C. (2009). Effects of phonological awareness and semantic intervention on word-learning in children with SLI. *International Journal of Speech Language Pathology*, *11*(6), 509–524.

Downloaded by [92.11.185.75] at 05:45 17 September 2015

Downloaded by [92.11.185.75] at 05:45 17 September 2015

Appendix 1. Full details of experimental tasks

Picture naming

The materials from the study of Funnell, Hughes, and Woodcock (2006) were used in a confrontation naming task. They consist of 72 black and white line drawings of objects from four categories, with 18 items in each category. Two categories (animals and fruits/vegetables) represented living things, and two (implements and vehicles) represented artefacts. The picture naming task was programmed using the experimental software DMDX (Forster & Forster, 2003) running on a laptop computer with a 15.4″ screen. Naming responses were recorded using an external microphone connected to the laptop. CheckVocal software (Protopapas, 2007) was used to obtain naming latencies. Accuracy of the naming responses was also recorded.

Items were presented in one session divided into three blocks of 24 items each. The child was asked to provide a single word for each picture. The tester moved to the next item as soon as the child named the picture. Naming responses were recorded at the time of testing and were checked later from the recording. Four fixed randomized orders were rotated across children during testing. No more than two objects from the same category appeared in succession, as in the naming study of Funnell et al. (2006). Each trial began with the presentation of a fixation cross in the centre of the screen for 500 ms. Then the picture appeared and stayed on the screen for a maximum of 5000 ms in the case of the TD children and 10,000 ms in the case of children with word-finding difficulty. Three items, not used in the main testing session, were presented for practice. Feedback on accuracy was given during the practice trials but not during the main task.

Word–picture verification task

A word–picture verification task was developed using the pictures from the naming task. It involved presenting one picture at a time on the computer together with a prerecorded spoken word. On one occasion the picture was presented with the matching word, and on another the picture was presented with a semantically related word. The child was asked to decide whether the spoken word corresponded to the picture or not. Knowledge of the word meaning would be represented by accepting the correct word for a picture but rejecting the semantically related word. Seventy-two object names were selected that were semantically related to the objects depicted in the Funnell et al. (2006) pictures. Related object names that were phonologically close to the target picture name or that started with the same phoneme (e.g., picture of a butterfly, semantically related word "bee") were replaced.

The names of the pictures and the semantically related words were recorded by an adult English speaker. The audio files were edited to add 500 ms of silence at the beginning and end of each file. The task was run on a laptop computer with a 15.4-inch screen and was programmed using the experimental software DMDX (Forster & Forster, 2003). There were two testing sessions with individual target pictures appearing once in each session. In one session the picture appeared with its name, and in the other with the semantically related word. The 72 items in each testing session were split in three blocks of 24 items each, with a rest pause between blocks. The child was asked to press the left "Ctrl" button on the keyboard for NO responses and the right "Ctrl" button for YES responses. Buttons were highlighted with stickers. Responses were scored correct only if the child accepted the correct name and rejected the semantically related word. Three practice trials were presented with stimuli that were not included in the main testing session. Feedback was given after practice trials but not during the main session. Four fixed random orders of stimuli were rotated across participants. Each trial began with the presentation of a fixation cross in the centre of the screen for 500 ms. The picture preceded the audio file by 17 ms.

Picture judgement task of associative semantics (PJs)

In this task, three pictures depicting objects were presented —a target together with two pictures underneath. One of the two pictures presented in the lower part of the screen had an associative semantic relationship to the target; the second came from the same semantic category as the first (e.g., *tie* presented with associate *shirt* and distractor *shorts*). Sixty-nine pictures depicting items from the Funnell et al. (2006) and Druks and Masterson (2000) picture sets were selected from the Shutterstock website. The task was administered using a laptop computer with a 15.4″ screen, and it was programmed using Visual Basic software. There were three practice trials using items that did not appear in the main session and 20 trials in the main task. A fixation point appeared at the start of each trial. The child was asked to choose which of the two items in the lower part of the screen fitted best with the item at the top. If it was the item on the left, the child was asked to press the Z button, for the item on the right, the M button. The two buttons were designated with stickers. Feedback on accuracy was given during the practice trials but not in the main task.

Nonword repetition (CNRep)

The widely used Children's Test of Nonword Repetition (Gathercole & Baddeley, 1996) was used to investigate children's repetition of unfamiliar forms. The test consists

of 40 nonwords of increasing length and complexity. The child was asked to repeat each nonword.

Simple and choice reaction time

Computerized tasks of simple and choice reaction time were adapted from Powell et al. (2007) and programmed on a laptop computer with a 15.4″ screen using the DMDX software (Forster & Forster, 2003). The simple reaction time task measured the time taken to make a key press response following the appearance of a target on the screen. Six different coloured drawings of monster characters were the target stimuli. The child's task was to press a key if two out of the six (the green dinosaur or the orange dinosaur) appeared, with separate response keys for each of these two targets. The six pictures and instructions appeared on the welcome screen. The instructions were read aloud to ensure that the child understood the task. There were six items for practice followed by two blocks of 18 trials each. Each trial started with the presentation of a fixation point (a black cross) in the centre of a white screen, followed by a lag and then the appearance of the target stimulus. The duration of the lag varied, to discourage anticipatory responses, and was 300, 600, or 900 ms. The lag times were randomized across trials, and presentation of the six target stimuli was also randomized across trials. The target stimuli remained on the screen for 1500 ms.

In the choice reaction time task, the child was asked to decide which of two stimuli appeared on the computer screen and to press the appropriate response key as quickly as possible. Children were asked to press the left "Ctrl" button as soon as the green dinosaur appeared, or the right "Ctrl" button if the orange dinosaur appeared. Green and orange stickers were placed on the two buttons. As for the simple reaction time task, instructions that appeared on the welcome screen were read aloud by the tester. A mouse press initiated the practice block of six items, with half containing the orange and half the green dinosaur. A black fixation cross appeared in the middle of the white screen for 500 ms followed by the target stimuli. Lag times varied in randomized order, as did appearance of either the orange or the green dinosaur. The lag times were 300, 600, or 900 ms. The target stimulus remained on the screen for 1500 ms. There were two blocks of 18 trials each in the main test session.

Appendix 2. Error categorization and girls' errors

Error type	Error subtype	Description	Amy's errors	Magda's errors
Semantic	Coordinate	Within same semantic category	tapir → *cow* coconut → *pineapple* jet ski → *speedboat* tomato → *apple* lemon → *pear* carrot → *pepper* cheetah → *leopard* milk float → *bus* torch → *light* donkey → *horse* garlic → *plum*	coconut → *lettuce* donkey → *horse* vulture → *duck* parachute → *balloon* ladle → *spoon* sledge → *boat* pelican → *duck*
	Superordinate	Semantic category to which target belongs	ostrich → *bird* barge → *boat* submarine → *boat* windsurf → *boat* tandem → *bike*	barge → *boat* windsurf → *boat* tandem → *bike* yacht → *boat*
	Functional	Functional attributes/ use of target	trowel → *digger*	
	Circumlocution	Multiword descriptive response	grater → something that you grate cheese on ladle → *big spoon*	
	Visual attributes	Similar physical features	scorpion → *crab*	scorpion → *crab*
Phonological	Nonwords	Nonword that shares at least 50% phonemes with target	squirrel → /grɪrəl/ binoculars → /mɪnɒkuːlɜːz/	caravan → /kəraːrə/ aeroplane → /eəlɔpreɪn/

(*Continued*)

Downloaded by [92.11.185.75] at 05:45 17 September 2015

Appendix 2. Continued

Error type	Error subtype	Description	Amy's errors	Magda's errors
	Formal	Real word that shares at least 50% phonemes with target, but not semantically related		
Mixed	Semantic and phonological	Semantically and phonologically related		tractor → *truck* saw → *sword* motorbike → *bike* tank → *truck* rake → *scrape*

Downloaded by [92.11.185.75] at 05:45 17 September 2015

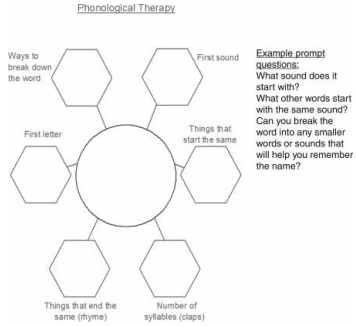
Appendix 3. Summary of therapy protocol

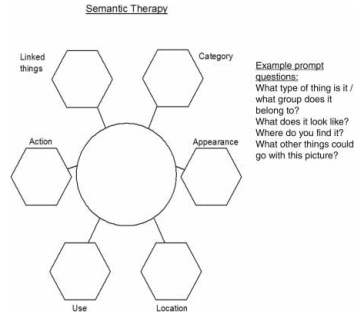
The therapy protocol was based on Boyle and Coelho (1995), Coelho et al. (2000), Boyle (2004), Massaro and Tompkins (1994), and Leonard et al. (2008) and utilized word-webs:

- Sessions occur once a week for 45 minutes (approximately 10 minutes assessment, 5 minutes activity while therapist selected unnamed items and 30 minutes intervention). Six sessions per intervention block.
- 25 items from the experimental set treated each half term, plus a further 6–12 words selected by the children, carers, and teachers.
- Therapy items treated in a continuous, cyclical order. Words correctly named at the start of a session will not be treated on that day.
- Record sheets used, including tick charts for monitoring participants' production attempts and overall response to therapy.

Therapy, first 2/3 sessions:

- Task introduction
- Generation of features, using prompt questions linked to word-webs:





- If unable to generate features a choice of features is provided (e.g., it has 2 or 3 beats (phonological–syllables), it has stripes or spots (semantic–appearance).
- All features are considered in the same order, starting with the hexagon at the top right and proceeding clockwise. Once all features have been generated or chosen, they are reviewed, and the child is asked to say the word. If unable, it is provided by the therapist, and the child is encouraged to say the word.

As sessions continue and according to child's ability:

- Develop metacognitive awareness: Encourage child to reflect on what aspects of word-webs are most helpful to them.

Therapy, Sessions 4/5/6:

- Barrier games: Position a screen between the therapist and child. Using completed word-webs, take turns to describe and guess items covered in previous sessions.
- Review of most useful strategies learnt during therapy, create card to help child remember what helps them when they cannot retrieve a word.

Appendix 4. Guidelines for conversation

Conversation measure:
Start by saying "let's talk a little" . . .

Begin with a topic that is personal to the child, based on their own interests/experience—as reported by the child themselves and/or their parent/teacher.

Standard short set of questions for all children:
Can you tell me:

- about your bedroom?
- What TV programmes do you like to watch?
  Follow-up probes:
  - "Tell me about that one, I haven't seen it."
  - "What happened on the last one you watched?"
  - "Do you ever watch (insert current programmes likely to be of interest)?"
- what you are good at?
- what you would like to be better at?

Hierarchy of cues to help support/scaffold conversation:

Mmn
Uhuh
Tell me more.
Just do your best/you're doing great
I'd like to hear more about that/Tell me what you can.
That sounds interesting
What else?

Nonverbal prompts:

Smiles and eye contact
Nods of affirmation and agreement

Appendix 5. Weightings for statistical comparisons between phases of the study

The weightings are calculated to test specific hypotheses. For example, the second set of weightings "treatment versus no treatment" compares the rate of change across intervention and no intervention (baseline, washout, and follow-up) phases of the study (for details see Howard et al., 2015).

1. Weightings of naming assessments for Amy

Naming assessment	1	2	3	4	5	6	7
	Pre 1	Pre 2	Pre 3	Post phon	Post washout	Post sem	Follow-up
Trend	−6	−4	−2	0	2	4	6
Treatment vs. no treatment	9	−2	−13	4	−7	10	−1
Treatment A vs. Treatment B	2	4	6	−13	−11	5	7
Phon trt	2.24	−4.47	−11.18	13.42	6.71	0	−6.71
Sem trt	7	2	−3	−8	−13	10	5

Note: Phon = phonological; sem = semantic; trt = treated.

2. Weightings of naming assessments for Magda

Naming assessment	1	2	3	4	5	6	7	8
	Pre 1	Pre 2	Pre 3	Pre 4	Post phon	Post washout	Post sem	Follow-up
Trend	−7	−5	−3	−1	1	3	5	7
Treatment vs. no treatment	5	1	−3	−7	1	−3	5	1
Treatment A vs. Treatment B	0	2	4	6	−13	−11	5	7
Phon trt	7	−1	−9	−17	17	9	1	−7
Sem trt	9.04	3.87	−1.29	−6.45	−11.62	−16.78	14.20	9.04

Note: Phon = phonological; sem = semantic; trt = treated.

Downloaded by [92.11.185.75] at 05:45 17 September 2015