Journal of Research in Reading, ISSN 0141-0423
Volume 29, Issue 2, 2006, pp 229–250
DOI: 10.1111/j.1467-9817.2006.00300.x

Does the PMSP connectionist model of single word reading learn to read in the same way as a child?

Daisy Powell

Royal Holloway, University of London
Carnegie Mellon University, USA

David Plaut

Carnegie Mellon University, USA

Elaine Funnell

Royal Holloway, University of London

The Plaut, McClelland, Seidenberg and Patterson (1996) connectionist model of reading was evaluated at two points early in its training against reading data collected from British children on two occasions during their first year of literacy instruction. First, the network's non-word reading was poor relative to word reading when compared with the children. Second, the network made more non-lexical than lexical errors, the opposite pattern to the children. Three adaptations were made to the training of the network to bring it closer to the learning environment of a child: an incremental training regime was adopted; the network was trained on grapheme–phoneme correspondences; and a training corpus based on words found in children's early reading materials was used. The modifications caused a sharp improvement in non-word reading, relative to word reading, resulting in a near perfect match to the children's data on this measure. The modified network, however, continued to make predominantly non-lexical errors, although evidence from a small-scale implementation of the full triangle framework suggests that this limitation stems from the lack of a semantic pathway. Taken together, these results suggest that, when properly trained, connectionist models of word reading can offer insights into key aspects of reading development in children.

Since the publication of Seidenberg and McClelland's (1989) innovative, connectionist account of reading, connectionist learning simulations have had a marked impact on the field of reading research (Brown, 1997; Harm & Seidenberg, 1999; Kello & Plaut, 2003; Plaut, 2001; Plaut, McClelland, Seidenberg & Patterson, 1996; Zorzi, Houghton & Butterworth, 1998a). Although these networks, when fully trained, have provided a good fit to data from skilled readers, less attention has been given to their performance during

Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA

training, and how this performance compares with children in the process of learning to read, an issue which forms the focus of the current research.

Traditional accounts of reading (Coltheart, 1978; Morton & Patterson, 1980) propose that two separate procedures are required for skilled reading, that is, the process of recognising, pronouncing and comprehending written words: a lexical procedure to access stored representations of known words, necessary for reading irregular words, and a sub-lexical procedure to enable novel words to be decoded according to grapheme–phoneme correspondence (GPC) rules.

However, this dual-mechanism account was challenged by Seidenberg and McClelland's (1989) 'triangle' model of reading. The triangle model consists of three sets of simple processing units: a bank of grapheme units representing orthography, a bank of phoneme units representing phonology and a bank of semantic units. Within this model, words are represented as distributed patterns of activity across each set of units. Oral reading is accomplished as follows: a word is presented to the network as a pattern of activity across the grapheme units. This pattern of activity propagates through the network, with the end result that a pattern of activity across the phoneme units, consistent with that word's pronunciation, is generated.

The triangle model comprises two pathways between the written and spoken forms of words: a pathway mapping directly from orthography to phonology, and a second pathway, which maps from orthography to phonology via semantics. However, there are some important differences between this approach and traditional, dual-route accounts of reading (e.g. Coltheart, 1978; Coltheart, Rastle, Perry, Langdon & Ziegler, 2001) in the way in which reading is seen to occur. For example, in the triangle model there is no categorical distinction between the types of words handled by each pathway, whereas dual-route accounts propose that irregular items can only be read by the lexical route and non-words by the sub-lexical route. Furthermore, unlike traditional accounts of reading, which stipulate that different types of processing mechanism underlie the lexical and sub-lexical routes, a single mechanism underlies all processing in the triangle model. Seidenberg and McClelland (1989) implemented part of the triangle model, the direct pathway from orthography to phonology, as a three-layer, connectionist network that learned to 'pronounce' a large corpus of single-syllable orthographic word forms to a level similar to skilled readers, although its non-word reading was poor relative to human readers.

Subsequently, Plaut et al. (1996, henceforth PMSP), reported a series of simulations in which improvements were made to the way in which orthography and phonology were represented. The simplest PMSP network consists of a set of grapheme units, coding all single- and multi-letter graphemes, connected unidirectionally, via a set of hidden units, to a bank of output, phoneme units. This network learned to produce pronunciations for both regular and irregular words and non-words to adult levels, and has successfully simulated a range of phenomena, including the frequency by consistency interaction (Taraban & McClelland, 1987). Like the Seidenberg and McClelland network, this network is a partial implementation of the triangle framework and cannot capture all aspects of reading. For example, when the trained network was subsequently damaged, it was not possible to simulate the pattern of surface dyslexia, in which the reading of irregular words is impaired but non-word reading is spared. However, in a further simulation that included a scaled-down implementation of the triangle model's second, semantic pathway (PMSP simulation no. 4), the hallmark performance pattern of surface dyslexia was achieved.

Although the focus of the PMSP networks was on skilled reading, the triangle framework has been extended to issues relating to reading development by Harm and

Seidenberg (1999). Harm and Seidenberg explicitly investigated the impact of prior exposure to the phonological forms of words on subsequent reading training, acknowledging that unlike children, who possess extensive spoken vocabularies before they start to learn to read, the PMSP (and Seidenberg and McClelland) networks learned to read with no prior exposure to either the phonological or the semantic forms of words. This is partly a limitation imposed by the fact that these networks lack the second, semantic pathway of the triangle model, and thus do not have the capacity to represent word meaning, which is also the case in the Harm and Seidenberg (1999) networks (though see Harm and Seidenberg 2004, for a full-scale implementation of the triangle model). However, Harm and Seidenberg reported a network in which phonology was implemented as an attractor network and which was pre-trained on purely phonological word forms before being trained to produce pronunciations of written words. This network was able to learn the dependencies between phonemes occurring in words: phonotactic knowledge that children can be assumed to possess before the onset of literacy instruction. The network, when fully trained, learnt to read words and non-words to adult levels, and Harm and Seidenberg were further able to investigate aspects of phonological and surface developmental dyslexia by impairing their network.

The various implementations of the triangle model outlined above have provided impressive accounts of skilled reading and its disorders, and in the case of Harm and Seidenberg's network, have tackled aspects of reading development. However, questions still remain about how childlike is the process by which these networks learn. Indeed, there have been a number of criticisms of the approach, relating to ways in which the theoretical underpinning to the triangle model differs from traditional accounts of reading development.

As noted earlier, traditional theories of skilled adult reading propose distinct procedures for the processing of familiar and novel word stimuli. This is discrepant with the theory underlying the triangle model, which posits that a single mechanism, involving simple mappings between various banks of units, can account for the ability to read written words aloud. Developmental theories also emphasise the distinction between a rule-based, decoding mechanism and a lexical retrieval mechanism, although the two procedures are proposed to interact during development. Frith (1985), for example, suggests that orthographic word forms develop from initial logographic word forms through the influence of gradually developing letter identification and phonological awareness skills that operate initially at the single-letter level and later at the level of larger groups of letters.

Stuart and colleagues (Stuart, 1990; Stuart & Coltheart, 1988; Stuart & Masterson, 1992; see also Ehri & Wilce, 1985; Rack, Hulme, Snowling & Wightman, 1994, for a similar view) argue that children make use of letter-sound knowledge and phonological awareness from their earliest attempts to read, and that these skills assist children in developing partial orthographic representations of words. Savage, Stuart and Hill (2001) argue that errors in which the initial letter, or the initial and final letters of words are preserved, errors that they term 'scaffolding errors', provide evidence of such partial representations. Savage et al. (2001) further suggest that scaffolding errors, which reliably predict reading development, indicate that children are making use of alphabetic skills when building early lexical representations.

A central feature of the above accounts of children's reading is the notion of lexical representations, that is, unique, internal representations of familiar whole words. This notion is also a key feature of traditional theories of skilled reading: Morton (1969) for example used the term 'logogen' to describe such lexical representations. It is notable, however, that the PMSP network does not involve static, localist representations of whole words at the level of orthography, phonology, or even semantics. Instead, words are represented as distributed patterns of activity across the sets of grapheme, phoneme and semantic units, and the network learns to read by learning to map from graphemes to phonemes in a way that reflects the context in which graphemes occur.

The above discrepancies between the PMSP network and traditional theories of reading development raise questions about the applicability of such connectionist accounts to reading development. An important step in establishing the validity of the connectionist approach to reading lies in demonstrating that networks do not just simulate skilled reading, but that they achieve this in a way which is consistent with what is known of human reading development. In this study, we chose therefore to focus on the nature of the training received by these networks, and on two key ways in which this differs from the manner in which children are taught to read, as these discrepancies have provoked criticism of the approach (Cassidy, 1990; Coltheart, Curtis & Haller, 1993; Hulme, Snowling & Quinlan, 1991).

First, while British children are taught explicitly the relationships between graphemes and phonemes and, in most cases, also between letters and their alphabetic names, the connectionist networks outlined above are only ever exposed to whole orthographic word forms. Most children can name and even give letter sounds for at least some letters of the alphabet before they start to learn to read. A knowledge of letter identities, particularly when combined with a knowledge of how individual letters map onto the sounds that they represent, has repeatedly been shown to assist progress in reading (Bradley & Bryant, 1983; Muter, Hulme, Snowling & Taylor, 1997; Stuart & Coltheart, 1988) and, as a consequence, plays an integral part in developmental theories of reading. In particular, such alphabetic skills are thought to provide children with a means of decoding unfamiliar words on the basis of grapheme–phoneme correspondences (see Adams, 1990, for a review), and the fact that the networks are not explicitly trained on letter-sound mappings at the sub-word level may give them a disadvantage, particularly in generalising to non-words.

Zorzi, Houghton and Butterworth (1998b) observed that non-word reading appeared to develop slowly in the PMSP simulations. Zorzi et al.'s approach to addressing this issue was to adapt the architecture of their connectionist reading account to incorporate a second, direct route between orthography and phonology that was not mediated via hidden units. They argued that allowing orthography to make direct contact with phonology by incorporating a simple two-layer component to their network would enable regular GPCs to be learnt much more quickly, and thus the ability to generalise to novel items would be acquired more efficiently. Although their dual-route network did learn to generalise very quickly, it is also possible that modifying a network's training to include training on sub-word GPC level mappings would have a similar effect in improving the rate at which the network acquires the ability to generalise.

Second, children are exposed to a small vocabulary of written words that enlarges gradually with proficiency. In contrast, the PMSP network learns to read from exposure to a very large, heterogeneous set of words, which allows the network to extract the statistical regularities of the English language, enabling it to learn not only the words in the training set, but also to generalise to non-words. However, as Cassidy (1990) has pointed out, it is unclear whether this method of learning is consistent with the gradual vocabulary expansion experienced by children. Nevertheless, connectionist models are capable of learning through incremental training regimes, and arguably have been shown to provide a better fit to developmental data than networks trained using regimes where all items are present in the training set from the time training begins (Elman, 1993; Plunkett & Marchman, 1993; but see also Rohde & Plaut, 1999).

The principal aim of this study was to investigate the developmental applicability of the connectionist, triangle model of reading. As connectionist networks have been shown to be able to learn to read to adult performance levels it seems reasonable to investigate how well their performance in the early stages of training matches the performance of children when they are first learning to read, particularly as the developmental applicability of this type of model has been stressed (Seidenberg & McClelland, 1989).

To address this issue, we first collected pertinent data on word and non-word reading from children during the earliest stages of learning to read when few words could be read. This was considered important as our aim was to investigate the properties of beginning reading. The children were then reassessed on the same sets of words and non-words around 6 months later. As well as investigating the children's overall ability to read words and non-words, we also looked at the types of errors the children made. We used these data to evaluate one implementation of the triangle model of reading developed by Plaut et al. (1996). Of the range of possible alternatives, we chose to focus our investigation on one of the PMSP networks, as these networks have provided an account of perhaps the widest range of behavioural data of networks of this type. The results of these explorations led us to conduct further simulations to investigate how modifications to the network's training impacted on its performance.

Experiment 1

Method

Participants

Eleven girls and 12 boys were selected from the reception classes of two schools in a mixed socio-economic area in Middlesex, UK. Teaching methods involved a combination of 'whole language' and phonics approaches so that the children were taught not just whole words, but also letter names and the sounds for letters and letter groups. Words involved in training appeared repeatedly in reading and writing tasks and were also displayed around the classroom so that the learning of familiar words was continually reinforced. As the children were in the Reception year, they were not subject to the formal, closely structured phonics instruction laid out in the UK's National Literacy Strategy. However, in both schools the children in Reception took part in a less formal 'literacy hour' at the beginning of the school day, which often focused on letter knowledge and phonics.

All 23 children were tested during the first school term (Time 1) and 20 children in the group were tested again, about 6 months later (Time 2), during the third school term (three children were no longer available). Children had a mean age of 4 years 10 months (range 4:4–5:2 years) at Time 1, and a mean age of 5 years 5 months (range 4:11–5:10 years) at Time 2.

Materials

A set of 39 monosyllabic concrete nouns of CV or CVC structure was selected. These words, some of which had regular and some exceptional spellings, had low age-of-acquisition ratings (mean = 33.48 months, *SD* = 12.96; excluding seven words for which no ratings were available) and high imageability ratings (mean = 6.3, *SD* = 0.311, excluding three words for which no ratings were available) (Barry, Morrison & Ellis, 1997; Carroll & White, 1973; Morrison, Chappell & Ellis, 1997). This was to ensure, as far as possible, that the words were already in the children's spoken vocabularies. Although no words contained consonant clusters, most words contained digraphs (e.g. yacht, road, horse, shoe). The data reported here were part of a large-scale investigation of children's reading and reading-related skills (Powell, 2001), which included an examination, not reported here, of the role of within-word position in children's ability to read a set of test graphemes. For this reason, the words were selected to ensure that each of a set of 21 critical phonemes was represented once in the initial and once in the final position across the word set (e.g. book/web; thing/bath; dog/bed).

Thirty-three monosyllabic CV/CVC non-words, each beginning or ending with a critical phoneme, were also selected. Digraphs were generally avoided except in those situations where letters rarely occur in isolation at the end of a word (e.g. word-final 'v' and 'c'), in which case an additional non-word with a more conventional usage of the letter (e.g. tove, leck) was added to the test set. An additional three words and three non-words were selected for use as practice items. The words and non-words were handwritten, onto individual cards, in a lower-case script familiar to the children.

Procedure

Children were tested individually in one to two sessions lasting around 20 minutes each, both at Time 1 and Time 2. Words and non-words were tested in separate blocks, and the order of task (word or non-word) and of items was counterbalanced across children. Children were given the following instructions at the beginning of the word reading block: 'Each of these cards has a word written on it. I'd like you to try to read each of the words. Here's an example'. The children were then shown the three practice words, and if they had difficulties they were encouraged to try to sound out the letters. Once they started the experimental trials, no further help was given. Before the non-word reading block, the experimenter told the children 'I've made up lots of words and written them on these cards. I'd like you to read them for me and remember they are funny, made-up words so don't worry if they sound a bit strange. Here's an example'. The children were then shown the three practice non-words and corrected when they made mistakes and encouraged to sound out the letter strings. During the experimental trials, the children were neither corrected nor assisted.

Results

At Time 1, the children were significantly more successful at reading words than non-words (mean proportion correct: words = 0.12; non-words = 0.09. Wilcoxon's test, $Z = 2.47$, $p < .01$). However, the data revealed considerable variability in the children's performance. Although the majority of children read, on average, less than 10% of the words and none of the non-words correctly, the performance of three children in the class,

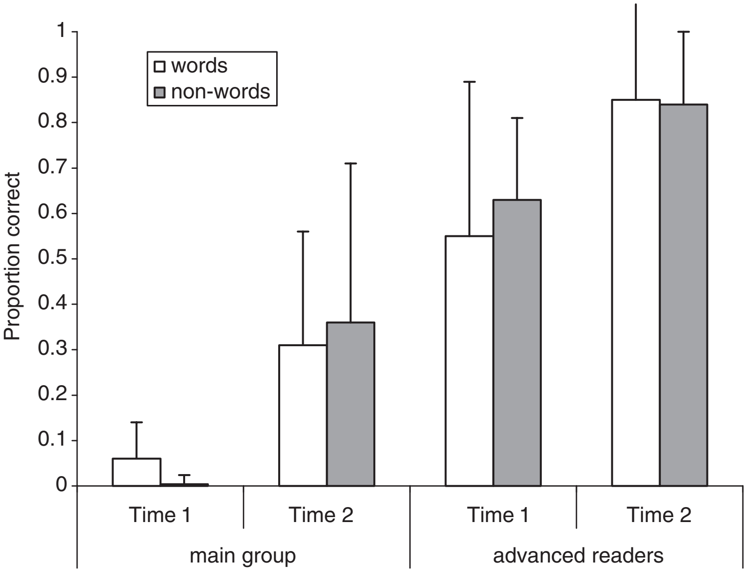

Figure 1. Mean proportion of correct responses to words and non-words made by the three advanced readers (outliers), as well as the remainder of the group, at both Time 1 and Time 2.

each a statistical outlier, far exceeded that of their peers: they read correctly a mean of more than 50% of the test words and the non-words. Figure 1 shows the data from these three children alongside the rest of the group, and the data suggest that they were further along the developmental trajectory than the other children, with no advantage for words even at Time 1. As they were statistical outliers, these three children were removed from the datasets at Time 1 and Time 2, and all following data analyses focused on data from the remaining 20 children (though at Time 2 one child was no longer available for testing, reducing the sample to 19).

As Figure 1 shows, the children read significantly more words than non-words correctly at Time 1 (Wilcoxon's test; $Z = 3.069$, $p < .01$). At Time 2, however, analysis showed that the significant advantage for words had disappeared (mean proportion correct: words = 0.31; non-words = 0.36. (Wilcoxon's test; $Z = -1.090$, $p > .05$). Altogether, these results suggest that the earliest stage of reading favours words in comparison with non-words, but that this difference reduces rapidly as word reading proficiency increases.

Four types of errors dominated the children's responses. These included visually similar lexical errors, which were defined as other words sharing at least one letter with the test item (e.g. the response 'cat' to the test item 'comb'), and visually similar non-lexical errors, defined as non-word responses sharing at least one letter with the test item (e.g. the response 'com' to the test item 'comb'). It should be noted that the criterion used here is more lenient than that used commonly in the literature in which visually similar errors usually share at least 50% of the letters in the target. It is also clearly the case that visually similar errors are also for the most part phonologically similar to their respective test items. Errors also included refusals and unrelated errors, in which the response shared no letters with the target.

The mean number of each error type is reported in Table 1. The children made significantly more lexical than non-lexical errors to words at both Time 1 and at Time 2 (Wilcoxon's test; Time 1: $Z = 3.74$, $p < .001$; Time 2: $Z = 3.35$, $p < .001$) but in non-word

Table 1. Mean proportion of each type of response made by children to words and non-words at Time 1 and Time 2.

Type of response	Words				Non-words			
	Time 1		Time 2		Time 1		Time 2	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Correct responses	.06	.08	.31	.25	.004	.02	.36	.35
Lexical errors	.16	.17	.31	.21	.04	.13	.27	.23
Non-lexical errors	.01	.04	.12	.09	.02	.08	.16	.13
Unrelated errors	.07	.11	.04	.05	.01	.02	.02	.04
Refusals	.70	.02	.22	.26	.93	.24	.19	.33

reading the difference was only significant at Time 2 (Wilcoxon's test; Time 1: $Z = 1.34$, $p > .05$; Time 2: $Z = 2.30$, $p < .05$).

Discussion

The behavioural data described above produced two key findings. First, the children read a higher proportion of words than non-words correctly on the first testing occasion, but by Time 2 the word advantage had disappeared. It should, however, be acknowledged that many of the children read correctly only a very small proportion of the 39 test words presented at Time 1, which showed that they were at the earliest stages of reading development. As our principal aim was to track the earliest stages of reading development these data act as a useful baseline with which to compare later development. Importantly, the children at Time 1 read more words than non-words, a finding that is consistent with the logographic stage of reading (Frith, 1985). By Time 2, this situation had changed, for as the word reading level improved, so too did non-word reading, to the extent that the group overall read non-words slightly more successfully than words. It should be noted, however, that the precise balance of word reading relative to non-word reading is probably a function of the relative complexity of the words and non-words used as test items. The non-words used in this study were mostly composed of three letters (the few exceptions including the graphemes *sh*, *th*, *ng*, *ff* and *ck*), and were of relatively simple CVC construction compared with the test words used, some of which were very complex and had unusual spelling patterns (e.g. yacht). If a set of non-words composed of more complex letter groups had been used instead, we would have expected the extra demands of these non-words upon the children's orthographic and phonological skills to be reflected in relatively poorer performance. However, one of our aims in creating our set of non-words was to make them as easy as possible for these very young children to read. The advantage for word reading found at Time 1 suggests that, at this stage, the children did not possess the necessary alphabetic skills to decode even simple letter strings, and thus that the words which they read correctly were identified as familiar orthographic word forms.

Second, on both testing occasions the children made more lexical than non-lexical errors in word and non-word reading. An explanation for this pattern of errors can be found in the developmental literature. Studies have shown that both segmental phonemic awareness and also letter-sound knowledge take time to develop (e.g. Liberman, Shankweiler, Fischer & Carter, 1974; Seymour & Evans, 1994; Treiman & Baron, 1981).

As a result, it has been suggested (Stuart & Coltheart, 1988) that a partial orthographic representation of a word such as ‘dog’ (that may code only the written letter D) may be built up. Such a partial lexical representation could be shared by a number of words and this may result in lexical confusion errors.

Simulation 1

The aim of this experiment was to compare the PMSP network at an early stage in its training with the children’s data reported above. We report two sets of simulations. First, the original PMSP network was tested against the developmental data and then, in the second set of simulations (Simulation 2), modifications were made to the network to more closely resemble the children’s learning environment.

Method

Network architecture

The children’s data were compared with the fully feed-forward network described by Plaut et al. (1996, simulation 2 with a moderate frequency compression). Full details of the architecture, training and testing procedures are given in Plaut et al. (1996). The architecture of the model consists of three levels of units: an input level of 105 grapheme units, an intermediate level of 100 hidden units and an output level of 61 phoneme units. Each grapheme unit was connected to each hidden unit, by means of a weighted connection, and each hidden unit was in turn connected to each phoneme unit.

Training the network

The training set was composed of the full set of 2,998 monosyllabic words used in original PMSP training corpus, augmented by two additional words, ‘PEAS’ and ‘LION’[1] that appeared in the word-reading tests given to the children. This was necessary because we wanted to make a direct comparison between the children and the network, in terms of performance on the same set of words and non-words. The remaining words used to test the children were all present in the network’s training corpus. As each word was presented to the network, the activity level of the input units for those graphemes occurring in the word was clamped to 1, with the activity of the other grapheme units clamped to 0. Activity then spread forward to the hidden units, and then to the phoneme units, with the resulting pattern of activity of the phoneme units taken as the network’s pronunciation. The learning algorithm used was backpropagation (Rumelhart, Hinton & Williams, 1986): the weights of connections between the different levels of units were adjusted after each word presentation to minimise the discrepancy between the network’s pronunciation of a word and that word’s correct pronunciation. Training continued for 400 epochs, until the error rate had stabilised at its minimum level.

Testing the network

The testing procedure used was the same as that described in Plaut et al. (1996). The network’s phoneme units were organised such that there were three sets of mutually exclusive phoneme groups representing initial consonants, vowels and final consonants separately. The most active consonant phoneme in both initial and final position was included in the network's pronunciation, so long as its activation exceeded 0.5. The vowel that was included in the network's pronunciation was the most active unit in the set of vowel units, regardless of whether its activation exceeded 0.5. At each epoch, data were generated on the model's performance on the entire training set, and the network was also tested on the same word and non-word sets as the children described above.

Procedure for matching the network's performance with the children's at Time 1 and Time 2

Our aim was to find two points in the network's training when its performance could be considered equivalent to that of the children at both Time 1 and Time 2. We used word reading performance, on the same set of words that were used in testing the children, as our benchmark, and identified two epochs in the network's training where it produced correct pronunciations for as many of the test set of words as the children, on both first and second testing occasions. Once these two epochs had been identified, we could then examine the network's non-word reading and we could also investigate the types of errors made in both word and non-word reading.

However, the children's performance, particularly at Time 1, was characterised by a large number of refusals, which are not straightforward to interpret, because when children refuse to respond we cannot assume that they do not know the correct response: on some occasions they may, for example, have a response, either correct or incorrect, but not wish to offer it. The network, though, always produced a response, however incomplete or inaccurate, and there was therefore no natural analogue to the children's refusals. We wanted to find a more sophisticated means of matching the network with the children, which reflected levels of correct responses but also numbers of refusals. We therefore sought a means of classifying a proportion of the network's responses as refusals, which was neutral in terms of accuracy.

The approach we took was to generate a measure of *stress* (Plaut, 1997) for each of the network's responses. Stress is a measure of how binary the network's output units were on each word/non-word presentation, and thus could loosely be described as the degree of confidence the network had in each response, and it is a measure that is entirely independent of the accuracy of the response. Accordingly, we used stress as the basis for determining which of the model's responses could be classified as refusals. Those responses which achieved the lowest stress scores (indicating that the network's output units were least binary) were treated as refusals. We classified as refusals an equal number of the network's responses (regardless of whether they were correct or incorrect) as the mean number of refusals made by the children. As a result, the network's reported performance was generally an underestimate of its actual 'ability', as several correct responses tended to be among those that were classified as refusals and thus treated as errors, but we believe this underestimation provides a fairer point of comparison with the children's performance, which is similarly likely to underestimate their true ability. However, while this stress measure attempts to provide a means of tackling the issue of refusals, we acknowledge that it can only provide a limited reflection of the range of diverse and complex factors that presumably play a part in this aspect of children's behaviour.

Because the children were tested on words and non-words in separate blocks, different stress measures were obtained for words and non-words. For Time 1, the epoch at which the network produced the same number of correct responses as the children (once the appropriate number of responses had been classified as refusals, and thus errors) was considered a match for the children. The same procedure was used for Time 2.

Results

Development of word and non-word reading

As Figure 2 shows, at Time 1 the PMSP network, like the children, read more words than non-words correctly, though this difference just failed to reach significance ($\chi^2(1) = 3.108$, $p = .078$). However, at Time 2 the network read significantly more words than non-words ($\chi^2(1) = 5.939$, $p < .05$), showing a different pattern of performance to the children. Indeed, at Time 2 the children read significantly more non-words correctly than the network ($\chi^2(1) = 9.382$, $p < .01$).

Error analysis

The types of errors made by the children and the PMSP network are summarised in Table 2. To compare the network's errors with the children's, the total number of 'related'

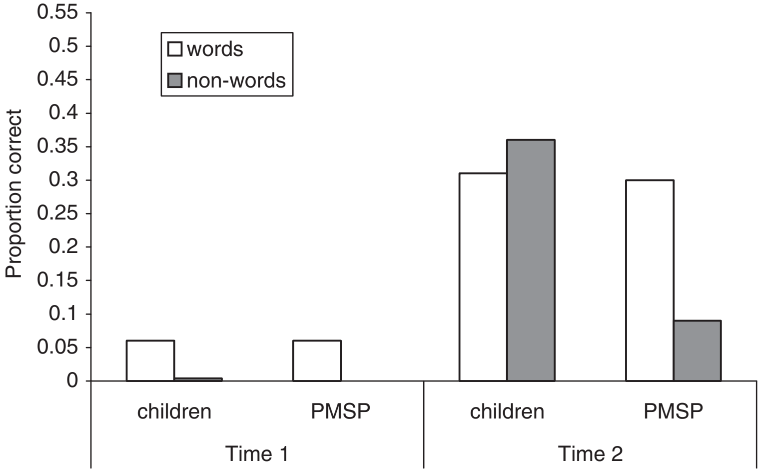

Figure 2. Mean proportion of correct responses to words and non-words produced by the PMSP network and the children at Time 1 and Time 2.

Table 2. Mean proportion of each type of response made by the children and the PMSP network to words and non-words at Time 1 and Time 2.

Type of response	Words				Non-words			
	Time 1		Time 2		Time 1		Time 2	
	Children	PMSP	Children	PMSP	Children	PMSP	Children	PMSP
Correct responses	.06	.06	.31	.30	.004	.00	.36	.09
Lexical errors	.15	.07	.31	.14	.04	.00	.27	.26
Non-lexical errors	.01	.14	.12	.35	.02	.07	.16	.47
Unrelated errors	.07	.00	.04	.00	.01	.00	.02	.00
Refusals	.70	.70	.22	.21	.93	.93	.19	.19

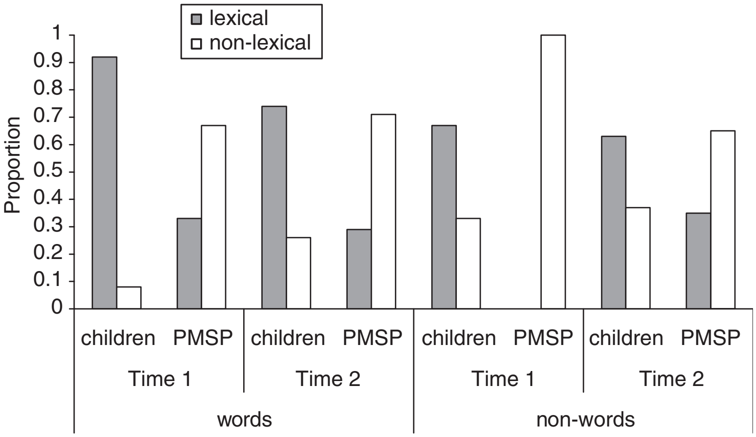

Figure 3. Relative proportion of lexical and non-lexical errors produced by the children and PMSP network at Time 1 and Time 2. Word and non-word reading.

responses (lexical and non-lexical errors) was first calculated, and then the proportion of these related responses that were lexical and non-lexical errors was calculated, both for the network and the children. As Figure 3 shows, in contrast to the children, the PMSP network made more non-lexical than lexical errors to both words and non-words, although in both cases the difference was not significant for words or non-words at Time 1, and only for words at Time 2 (Words: Time 1, $Z = 1.075$; Time 2, $Z = 1.964$, $p < .05$; Non-words: Time 1, $Z = 0.577$, $p > .05$; Time 2, $Z = 1.616$, $p > .05$).

Discussion

The performance of the network diverged from that of the children on both of the performance measures we considered. First, the network did not learn to generalise to novel stimuli as quickly as the children: while by Time 2 the children could read the non-words at least as successfully as the words, the network still read significantly more words than non-words correctly at Time 2. This is consistent with the observation that non-word reading develops slowly in the PMSP simulations (Zorzi et al., 1998a). Children of course have the advantage that they are generally explicitly taught the relationships between graphemes and phonemes as part of learning to read, unlike the network that has to extract these relationships through exposure to whole words alone. In addition, the relative simplicity of the graphemes in the non-words in comparison to those in the words may also have enhanced the level of non-word reading relative to word reading in the children: a factor that the network would not be sensitive to.

The second way in which the network's performance differed from that of the children was in the types of errors it made. The children made mostly real-word errors that shared at least one letter with the target word or non-word, while the majority of the errors made by the network were non-word responses that shared a letter or more with the target. Children are known to be sensitive to the relationships between visually similar words, for they find words that have many visually related neighbours both easier to read and to spell (Laxon, Coltheart & Keating, 1988). It is therefore possible that the children's lexical errors arose when unfamiliar words and non-words, or words with only partial lexical orthographic representations (e.g. Rack et al., 1994; Stuart & Coltheart, 1988), were mistaken for more familiar, visually similar, neighbours. As discussed earlier, the notion of static, whole-word representations is inconsistent with the PMSP network, where contingencies between graphemes occurring in words in the network's training set are learnt gradually. In spite of the lack of localist, whole-word representations, Plaut et al. (1996) provided evidence, such as frequency effects, of sensitivity to word-level knowledge in the fully trained, original PMSP networks. Nevertheless, our data indicate that the network's performance early in training does not capture the lexical effects evidenced in the performance of children learning to read. An additional factor of relevance is the fact that children are trained initially on a small vocabulary, while the model is trained by successive sweeps through a large corpus of words (cf. Cassidy, 1990). Networks trained on small training sets can show rote-learning behaviour, and as a result, if the network were also exposed to a small initial vocabulary, it might also produce more lexical behaviour, with a higher incidence of visually related real-word errors.

Simulation 2

In this experiment, we addressed some of the ways in which the training of the network clearly departs from the learning environment experienced by children. First, we adapted the network's training regime to reflect more realistically the gradual way that new words are introduced to children. Second, we added training on GPCs to the network's training, in keeping with the fact that the children in our sample were taught to associate individual letters and letter groups with the sounds they represent. Finally, we changed the network's training set. Instead of the adult-based corpus on which the original PMSP network was trained, we used a vocabulary of the 2,938 monosyllabic words[2] comprising the Children's Printed Word Database (Masterson, Stuart, Dixon, Desmond & Lovejoy, 2003), which was structured according to the frequency with which words appear in children's reading materials.

Method

Training the network

To approximate the small sight vocabulary that most children bring to the task of learning to read, the network was initially pre-trained, to a 50% accuracy level, on the 20 most frequent words in the Children's Printed Word Database. Raw frequencies were normalised by dividing each word's frequency by the frequency of the most frequent word, which now had a frequency of 1.0. Words were entered into the training set in order of frequency and training was incremental.

After this initial training phase, new items were added to the training set one at a time, whereas previously presented words continue to be reinforced. As the performance of the network improved and thus its reading vocabulary expanded, increasing numbers of words were required to bring the network's overall performance below the criterion mean level of accuracy of 50%, so the size of the increments increased throughout training. The model was also trained simultaneously on GPCs. Single-letter GPCs were assigned a frequency of 3.0 in order to account for the intensive alphabet training received by children early in literacy instruction and to ensure that GPCs entered the training set before the words. Digraph GPCs were given frequencies of 1.5. Consonantal GPCs were trained both in the initial and final positions in words because these are coded by separate sets of grapheme units in the PMSP model.

In summary, the fully adapted network was trained incrementally on the words from the Children's Printed Word Database as well as on single-letter and digraph GPCs.[3]

Control Simulations

To investigate the impact of the three changes we made to the network on the various issues of interest, we also carried out three control simulations, each involving the removal of one of the changes (i.e. the new training vocabulary, incremental training or training on GPCs) we made to the network described above. This allowed the impact of each of the three adaptations to be assessed through a comparison between the fully adapted network and a network that lacked a particular factor. The three control simulations are described below.

1. The *PMSP* vocabulary network: This network was trained incrementally on words and GPCs, but the word set comprised the original PMSP training corpus instead of the words from the Children's Printed Word Database.
2. The *non-INC* network: This network was trained simultaneously on the children's database word corpus and the GPCs. It was not trained incrementally.
3. The *non-GPC* network: This network was trained incrementally on the children's database word corpus but it was not trained on GPCs.

Results

Procedure for matching the network with children at Time 1 and Time 2

The procedure for matching the network with the children at the two testing occasions was the same as that used with the evaluation of the original PMSP network described above. At each epoch, data were generated on the network's performance on not just its training set, but also on the same set of words and non-words used in our assessment of the children's reading.

Development of word and non-word reading

Figure 4 shows the performance of the fully adapted network on words and non-words alongside that of the children. The performance of the network was almost identical to that of the children. At Time 1, the network pronounced more words than non-words correctly, though this difference was only marginally significant ($\chi^2(1) = 3.1084$, $p = .078$). At Time 2, as was the case with the children, the word advantage disappeared, and the network now read slightly, but not significantly, more non-words than words correctly ($\chi^2(1) = 0.8190$, $p > .05$).

Word and non-word reading data for each control simulation, as well as the fully adapted network, are shown in Figure 5. At Time 1, the three control simulations, like the fully adapted network, read no non-words correctly.

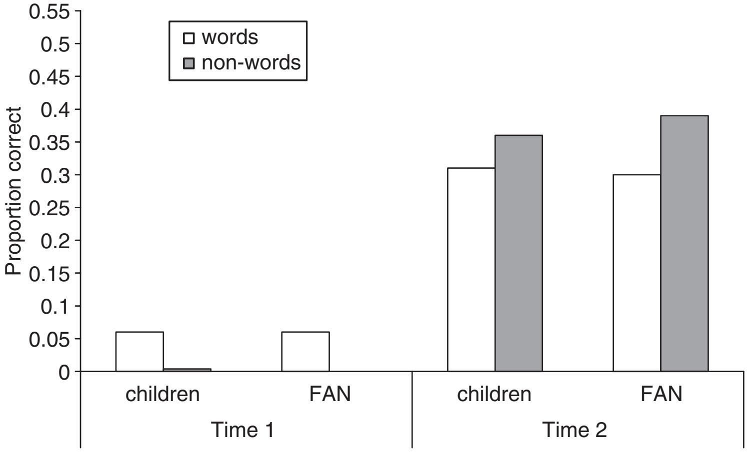

Figure 4. Relative proportion of words and non-words read correctly by the fully adapted network (FAN) and the children at Time 1 and Time 2.

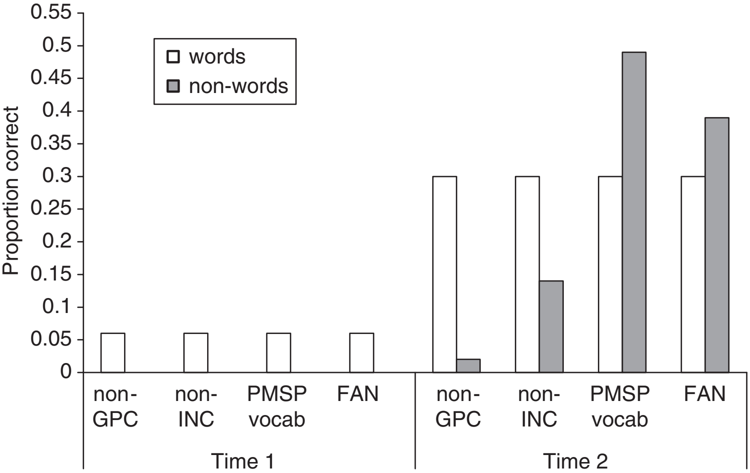

Figure 5. Words and non-words pronounced correctly by the three control simulations and the fully adapted network (FAN).

At Time 2, the non-GPC network was least successful in generalising, reading significantly fewer non-words correctly than the fully adapted network ($\chi^2(1) = 17.987$, $p < .001$). The non-INC network also read significantly fewer non-words correctly than the fully adapted network ($\chi^2(1) = 7.182$, $p < .01$). The PMSP vocabulary network, however, read more non-words correctly than the fully adapted network, but this difference was not significant ($\chi^2(1) = 0.754$, $p > .05$).

Error analysis

Figure 6 shows the proportion of lexical and non-lexical error responses made to words and non-words, at Time 1 and Time 2, by the fully adapted network and the children.

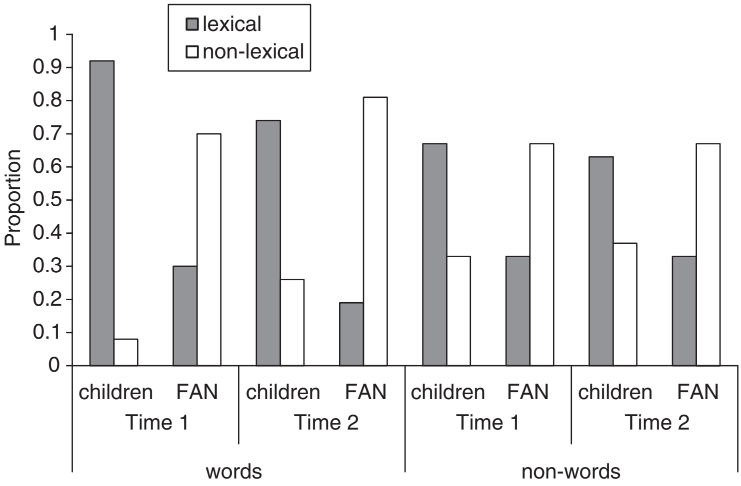

Figure 6. Relative proportion of lexical and non-lexical errors made by the fully adapted network (FAN) and the children at Time 1 and Time 2. Word and non-word reading.

Like the original PMSP network, the fully adapted network produced more non-lexical than lexical errors to the word stimuli, though the difference was only significant at Time 2 (Time 1: $Z = 1.265$, $p > .05$; Time 2: $Z = 2.837$, $p < .01$). The network also made more non-lexical than lexical errors to non-word stimuli, but in this case, the difference did not reach significance either at Time 1 or Time 2 (Time 1: $Z = 0.577$, $p > .05$; Time 2: $Z = 1.414$, $p > .05$).

Examination of the types of errors made by the control networks revealed that none of the changes reliably brought their performance closer to that of the children.

Discussion

The adaptations we made to the network's training substantially improved its ability to generalise. Indeed, at both Time 1 and Time 2, the fully adapted network provided an excellent fit to the children's data on non-word reading. Like the children, this network showed an advantage for words over non-words at Time 1, and a slight advantage for non-words at Time 2. Compared with the fully adapted network, the PMSP Vocabulary Control Network, which received the same training regime but a different training set, read the same proportion of non-words at Time 1 but slightly (and non-significantly) *more* non-words at Time 2. It is possible that the relatively high proportion of function words with irregular spellings among the highest frequency items in the Children's Printed Word Database may have reduced the ability of the fully adapted network to generalise by constraining its capacity to glean the statistical regularities of its training corpus. This demonstrates the value of using a training set structured according to the frequencies with which words occur in children's reading materials: once its training set reflected the properties of words typically encountered by children, its non-word reading performance came closely into line with that of the children.

The two control networks that lacked either incremental training (non-INC) or training on GPCs (non-GPC), also read more words than non-words at Time 1. At Time 2, both the non-GPC network and the non-INC network continued to read significantly fewer non-words correctly than the fully adapted network. Together, these results suggest that both incremental training and training on GPCs are important for generalisation to novel stimuli, and that it is the interaction between incremental training and training on GPCs that is responsible for the dramatic improvement found in the performance of the fully adapted network.

In another respect, the adaptations did not substantially improve the fit of the network to the children's data. Whereas the children's errors that shared at least one letter with the target were most frequently real words, the network tended to respond with non-words that shared a letter with the target. Although the fully adapted network made a smaller proportion of non-lexical to lexical errors than the original PMSP network, it persisted in making more non-lexical than lexical errors to both words and non-words.

One explanation for the poor fit of the network to the children's data on this measure could lie in the fact that children, unlike the fully adapted network, possess an extensive and well-developed spoken vocabulary by the time they start to learn to read. A second explanation concerns the strong effects of orthographic neighbourhood observed in the development of children's reading. In either case, an approximate response to a written word or non-word could result in generalisation to an orthographic or phonological word neighbour.

General discussion

The central aim of the current work was to evaluate the PMSP network of reading against empirical data on early reading acquisition. Results showed that the original PMSP network was slow, relative to the children, to learn to read non-words, nor did it simulate the pattern of reading errors observed. The fact that the network did not provide a match to the behavioural findings is not altogether surprising, as it was not created to answer developmental questions, but rather to model data from skilled readers, which it achieved very successfully.

The network was adapted to more closely resemble the learning experience of children, by training the network incrementally on a vocabulary based on the frequency of words in children's early reading books and by including grapheme–phoneme correspondences in the training. These modifications resulted in a network that generalised to produce a near perfect match to the children's non-word reading performance on both testing occasions. Generalisation ability is known to be crucial to children's literacy acquisition (Share, 1995), and was shown to be poor in the original Seidenberg and McClelland (1989) implementation of the triangle network. Furthermore, Zorzi et al. (1998b) pointed out that while the PMSP network showed good non-word reading at the end of training, inspection of the network's performance during training indicated that the ability to generalise to non-words appeared to develop too slowly to be consistent with what is known of children's reading development, though this has not until now been directly assessed. More generally, questions have been raised about whether the statistical learning that characterises connectionist networks could effectively model children's literacy acquisition, where new words are encountered gradually (Cassidy, 1990).

It is significant therefore that in this case, a gradually expanding training corpus, which reflected the frequencies of words appearing in children's early reading materials, combined with GPC training, greatly improved the ability of the network to generalise. This finding is interesting for theoretical reasons, as it contradicts Zorzi et al.'s (1998b) suggestion that the apparently slow development of generalisation ability in the PMSP networks was due to the fact that the network lacked a dedicated, sub-lexical pathway between orthography and phonology. Hutzler, Ziegler, Perry, Wimmer and Zorzi (2004) have also recently demonstrated that the PMSP network fails to show the strong advantage, early in training, of a regular orthography (German) over an irregular orthography (English), a result which they attribute to the fact that the PMSP network lacks a separate, sub-lexical pathway, and to factors relating to teaching methods. Our results, however, further predict that a training regime that included incremental training combined with GPC training may well result in an early acceleration in ability to read regular items, similar to that shown by the German children in the Hutzler et al. study, without the necessary addition of an additional pathway. Our finding also has practical implications, in terms of literacy acquisition more generally, providing an explicit demonstration of the value of training on the relationships between orthography and phonology at the sub-word level, early in literacy training.

The adaptations had little effect in reducing the discrepancy between the types of errors made by the network and the children. This suggests that the means by which the network was deriving pronunciations for words was not the same as those adopted by the children.

Overall, the fully adapted network only provided a partially successful fit to the children's data. However, our adaptations were constrained by the structure of the PMSP network, which lacks the semantic representations and the vocabulary of phonological word forms that are in place when children begin to read. The theoretical model on which the PMSP network is based (Seidenberg and McClelland, 1989) incorporates a bank of semantic units, connected bi-directionally to both orthographic and phonological units. This architecture provides a second pathway, in which orthography is mapped to phonology via semantic representations, and indeed further allows mappings from orthography to phonology to semantics and then back to phonology. It is possible that if such a network was instantiated and pre-trained on the spoken forms of words, contributions from the semantic pathway might introduce lexical influences into the network's phonological responses to the written forms of words.

We have obtained some support for this proposal through a preliminary investigation of the early performance of a small-scale 'triangle' model incorporating semantic representations, presented by Kello and Plaut (2003).[4] The network learned to map orthographic representations of 470 monosyllabic English words to their corresponding phonological representations both directly and via semantics. Analogous to children's acquisition of a spoken vocabulary before literacy instruction starts, the network was initially trained just on the (bi-directional) mappings between phonology and semantics. Once this learning was established, the network was trained to map from orthography to phonology.

We examined the errors made by the Kello and Plaut network, at a stage in its training when it produced pronunciations to the words in its training set at a level equivalent to the networks described above, at Time 1 and Time 2, and as predicted the proportion of lexical errors it made both to words and non-words was much higher. At Time 1 and Time 2, lexical errors accounted, respectively, for 75.8% and 73.3% errors to words, and for 74.7% and 74.1% errors to non-words. This performance pattern is much closer to the children we tested, who produced 81.7% and 74.6% lexical errors to words at Time 1 and Time 2, respectively, and 59.5% and 62.1% errors to non-words. This provides strong initial support for the prediction that additional contributions from a semantic pathway should give rise to lexical influences and would provide a closer match to the children's data presented in this paper.[5]

To conclude, we have demonstrated that a combination of incremental training, training on GPCs and a training corpus that reflects the words found in children's reading materials is important in accelerating the capacity of the network to generalise to novel letter strings, a skill crucial to human reading development. The discrepancies between the children's errors and those produced by the PMSP network, as well as the fully adapted network, raise interesting challenges for the future.

Acknowledgements

The first author was supported by a Royal Holloway University of London Studentship, and by grants from the Experimental Psychology Society and the British Psychological Society to support a study visit to Carnegie Mellon University and during her postdoc by NIH Grant MH55628 (D. Plaut, P.I.). The second author was supported by NIH Grant MH55628 and NSF Grant BCS-0079044. We thank the Forum for Research in Language and Literacy in the UK and the PDP Group at Carnegie Mellon University for helpful comments, and the children and teachers at the schools where the empirical data were collected.

Notes

1. The PMSP did not contain a phoneme unit for the medial diphthong in 'lion', or a grapheme unit for the digraph 'IO', so the word 'lion' was essentially unreadable for the network. For this reason, we replaced it with the visually similar word 'loin', which was, however, given the same frequency for the purposes of training the network as the real-world frequency of the word 'lion'.
2. All of the monosyllabic words in the Masterson et al. (2003) database were used, with the exception of items with unconventional spellings such as 'zzz'. It should be noted that this represents only a subset of the Masterson et al. database, which is not restricted to monosyllabic words.
3. In addition to these changes, a minor modification to the learning procedure was used that made it more appropriate for incremental training. Specifically, Plaut et al. (1996) used a procedure for adapting connection-specific learning rates, known as delta-bar-delta (Jacobs, 1988), that depends on the stability of the training corpus over the course of training. Accordingly, only momentum-descent was applied in the current context.
4. Kello and Plaut (2003) also examined an alternative, 'integrated-pathway' architecture in which orthography maps into the same group of hidden units that mediated semantic–phonological interactions. A simulation based on that architecture exhibited qualitatively equivalent behaviour to the 'triangle' simulations discussed here.
5. Harm and Seidenberg (2004) have recently described a full-scale triangle network, which offers the future potential to more directly test the prediction that a network incorporating semantic representations, which was trained on a more realistically sized vocabulary and tested on the same words and non-words as the children, would produce a higher rate of lexical relative to non-lexical errors, and thus a closer fit to the children's data.

References

Adams, M.J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Barry, C., Morrison, C.M. & Ellis, A.W. (1997). Naming the Snodgrass & Vanderwart Pictures: Effects of age of acquisition, frequency and name agreement. *Quarterly Journal of Experimental Psychology*, 50a(3), 560–585.

Bradley, L. & Bryant, P. (1983). Categorising sounds and learning to read: A causal connection. *Nature*, 301, 419–421.

Brown, G.D.A. (1997). Connectionism, phonology, reading and regularity in developmental dyslexia. *Brain and Language*, 59, 207–235.

Carroll, J.B. & White, M.N. (1973). Age-of-acquisition norms for 220 picturable nouns. *Journal of Verbal Learning and Verbal Behaviour*, 12, 563–576.

Cassidy, S. (1990). When is a developmental model not a developmental model? *Cognitive Systems*, 2–4, 329–344.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing*. London: Academic Press.

Coltheart, M., Curtis, B.P.A. & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, 100, 589–608.

Coltheart, M., Rastle, K., Perry, C., Langdon, R. & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108(1), 204–256.

Ehri, L.C. & Wilce, L.S. (1979). The mnemonic value of orthography among beginning readers. *Journal of Educational Psychology*, 71, 26–40.

Ehri, L.C. & Wilce, L.S. (1985). Movement into reading: Is the first stage of printed word learning visual or phonetic? *Reading Research Quarterly*, 22, 47–65.

Elman, J.L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, 48, 71–99.

Frith, U. (1985). Beneath the surface of developmental dyslexia. In K.E. Patterson, J.C. Marshall & M. Coltheart (Eds.), *Surface Dyslexia*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Harm, M.W. & Seidenberg, M.S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, 106(3), 491–528.

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*, 11(3), 662–720.

Hulme, C., Snowling, M. & Quinlan, P. (1991). Connectionism and learning to read: Steps towards a psychologically plausible model. *Reading and Writing*, 3, 159–168.

Hutzler, F., Ziegler, J.C., Perry, C., Wimmer, H. & Zorzi, M. (2004). Do current connectionist learning models account for reading development in different languages? *Cognition*, 91, 273–296.

Jacobs, R.A. (1988). Increased rates of convergence through learning rate adaptation. *Neural Networks*, 1, 295–307.

Kello, C.T. & Plaut, D.C. (2003). Strategic control in word reading: A computational investigation. *Journal of Memory and Language*, 48, 207–232.

Laxon, V.J., Coltheart, V. & Keating, C. (1988). Children find friendly words friendly too: Words with many orthographic neighbours are easier to read and spell. *British Journal of Educational Psychology*, 58, 103–119.

Liberman, I.Y., Shankweiler, D., Fischer, F.W. & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology*, 18, 201–212.

Masterson, J., Stuart, M., Dixon, M., Desmond, L. & Lovejoy, S. (2003). The children's printed word data base, www.essex.ac.uk/psychology/cpwd.

Morrison, C.M., Chappell, T.D. & Ellis, A.W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology*, 50A(3), 528–559.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, 76, 165–178.

Morton, J. & Patterson, K. (1980). A new attempt at an interpretation, or, an attempt at a new interpretation. In M. Coltheart, K. Patterson & J.C. Marshall (Eds.), *Deep dyslexia*. (pp. 91–118). London: Chapman & Hall.

Muter, V., Hulme, C., Snowling, M. & Taylor, S. (1997). Segmentation, not rhyming, predicts early progress in learning to read. *Journal of Experimental Child Psychology*, 65, 370–396.

Plaut, D.C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes*, 12, 767–806.

Plaut, D.C. (2001). A connectionist approach to word reading and acquired dyslexia: Extension to sequential processing. *Cognitive Science*, 23, 543–568.

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. & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition*, 48, 21–69.

Powell, D. (2001) The development of orthographic and phonological representations in children and connectionist networks. *Unpublished PhD Thesis*.

Rack, J., Hulme, C., Snowling, M. & Wightman, J. (1994). The role of phonology in young children learning to read words: The direct mapping hypothesis. *Journal of Experimental Child Psychology*, 57, 42–71.

Rohde, D.L.T. & Plaut, D.C. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition*, 72, 67–109.

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

Savage, R., Stuart, M. & Hill, V. (2001). The role of scaffolding errors in reading development: Evidence from a longitudinal and a correlational study. *British Journal of Educational Psychology*, 71, 1–13.

Seidenberg, M.S. & McClelland, J.L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523–568.

Seymour, P.H.K. & Evans, H.M. (1994). Levels of phonological awareness and learning to read. *Reading and Writing*, 6, 221–250.

Share, D.L. (1995). Phonological recording and self-teaching: *Sine qua non* of reading acquisition. *Cognition*, 55, 151–218.

Stuart, M. (1990). Factors influencing word recognition in pre-reading children. *British Journal of Psychology*, 81, 135–146.

Stuart, M. & Coltheart, M. (1988). Does reading develop in a sequence of stages? *Cognition*, 30, 139–181.

Stuart, M. & Masterson, J. (1992). Patterns of reading and spelling in 10-year-old children related to prereading phonological abilities. *Journal of Experimental Child Psychology*, 54, 168–187.

Taraban, R. & McClelland, J.L. (1987). Conspiracy effects in word pronunciation. *Journal of Memory and Language*, 26, 608–631.

Treiman, R. & Baron, J. (1981). Segmental analysis: Development and relation to reading ability. In G.C. MacKinnon & T.G. Waller (Eds.), *Reading research: Advances in theory and practice*, Vol. III. New York: Academic Press.

Zorzi, M., Houghton, G. & Butterworth, B. (1998a). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology, Human Perception and Performance*, 24(4), 1131–1161.

Zorzi, M., Houghton, G. & Butterworth, B. (1998b). The development of spelling-sound relationships in a model of phonological reading. *Language and Cognitive Processes*, 13(2/3), 337–371.

Received 4 March 2005; revised version received 6 February 2006.

Address for correspondence: Daisy Powell, School of Psychology and Human Development, Institute of Education, University of London, 25 Woburn Square, London WC1H 0AA, UK. E-mail: *d.powell@ioe.ac.uk*

Appendix 1

Table A1

Table A1. Word and non-word stimuli.

Test phoneme	Words		Non-words	
	Initial position	Final position	Initial position	Final position
b	Book (.05/.42)	Web (.00/.37)	Bep (.00/.26)	Wib (.00/.21)
k	Car (.10/.37)	Duck (.00/.47) Book (.05/.42)	Cug (.00/.42) Kib (.00/.26)	Gak (.00/.42) Mec (.00/.42) Leck (.00/.37)
d	Dog (.50/.89)	Bed (.10/.63)	Dat (.00/.32)	Pid (.00/.63)
f	Foot (.00/.11) Fork (.00/.16)	Leaf (.00/.05)	Fom (.05/.47)	Mef (.00/.37) Roff (.01/.37)
g	Gate (.00/.11)	Peg (.00/.42)	Gan (.00/.53)	Yag (.00/.42)
h	Horse (.00/.11) Hat (.35/.79)	–	Hud (.00/.37)	–
l	Lion (.00/.05)	Ball (.50/.58)	Lum (.00/.63)	Rell (.00/.37) Nol (.00/.63)
m	Moon (.00/.42)	Comb (.00/.00) Arm (.00/.11)	Mef (.00/.37)	Fom (.05/.42)
n	Nose (.05/.16)	Pen (.05/.53)	Nus (.00/.37)	Gan (.00/.53)
p	Pig (.10/.74)	Cup (.00/.42)	Pid (.00/.63)	Bep (.00/.26)
r	Ring (.00/.26) Road (.00/.26)	–	Rin (.00/.58)	–
s	Sword (.05/.00) Sun (.10/.74)	Bus (.40/.79)	Sep (.00/.26) Cib (.05/.32)	Tiss (.00/.26) Nus (.00/.37)
t	Tie (.00/.00) Tin (.00/.63)	Foot (.00/.11)	Teg (.00/.37)	Dat (.00/.32)
v	Van (.05/.63)	Cave (.00/.05)	Voz (.00/.37)	Piv (.00/.47) Tove (.00/.05)
w	Wall (.00/.16) Web (.00/.26)	–	Wib (.00/.21)	–
y	Yacht (.00/.05)	–	Yag (.00/.42)	–
z	–	Peas (.00/.00)	–	Rese (.00/.00) Voz (.00/.37)
ʃ	Shoe (.00/.05)	Fish (.05/.63)	Shad (.00/.21)	Pish (.00/.26)
θ	Thumb (.00/.05) Thing (.00/.16)	Bath (.00/.26)	Thib (.00/.16)	Suth (.00/.21)
η	–	King (.05/.42)	–	Ting (.00/.16)

Note: The proportion of children reading each item correctly is shown in parentheses, beside each item, for Time 1 and Time 2.