Incorporating Learning Mechanisms into the Dual-Route Cascaded (DRC) Model of Reading Aloud and Word Recognition

Stephen C. Pritchard BA/BE

ARC Centre of Excellence in Cognition and its Disorders

Department of Cognitive Science

Macquarie University, Sydney, Australia

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

December 2012

Table of Contents

Summary

The dual-route cascaded (DRC) model of reading aloud and word recognition has achieved considerable success. Despite this, it has faced ongoing criticism for being a static model of skilled reading that does not describe reading acquisition. This PhD research focused on incorporating learning mechanisms into DRC. Work was divided into two broad areas: orthographic learning within DRC's lexical route, and grapheme–phoneme correspondence learning in DRC's sublexical route.

To model orthographic learning, a "learning DRC" (L-DRC) was created. L-DRC provides a computational account of the self-teaching hypothesis, and in accordance with this, models orthographic learning as being self-driven via phonological recoding, with context supporting irregular word learning. L-DRC effectively modelled self-teaching and orthographic learning, and suggested mechanisms for the difficulties children may face when self-teaching difficult words like potentiophones or heterophonic homographs.

To model sublexical learning, a grapheme–phoneme correspondence (GPC) Learning Model was created and tested. This model effectively demonstrated GPC learning, especially when trained on an input corpus limited to mono-morphemic words presented once each. However, it experienced difficulties when trained on more realistic input corpuses. The model's performance suggests that sublexical route learning is sensitive to morpheme structure, and to type-based rather than token-based features in written material.

The investigation of sublexical-route learning was preceded by a comparison of the sublexical routes of two competing dual-route models, the DRC and connectionist dual-process (CDP+) models. These were assessed against new empirical data on how people pronounce nonwords. While neither model provided a good match to the human data, DRC performed significantly better than CDP+, or its successor CDP++.

Declaration

The work in this thesis is my original work. It has not been submitted for a higher degree in any other university or institution. All of the work reported in this thesis was undertaken during the time I was enrolled as a PhD student at Macquarie University, under the supervision of Professor Max Coltheart, Professor Anne Castles, and Dr Eva Marinus. Ethics approval for work with human participants conducted as part of this thesis was obtained from Macquarie University's Ethics Committee, Reference No. HE24NOV2006-R04946C.

Chapter 3 of this thesis is a reproduction of an article published in the *Journal of Experimental Psychology: Human Perception and Performance*. The research presented in Chapters 2 and 4 is expected to be submitted for publication shortly after submission of this PhD thesis.

Pritchard, S. C., Coltheart, M., Palethorpe, S., & Castles, A. (2012). Nonword reading: Comparing dual-route cascaded and connectionist dual-process models with human data. *Journal of Experimental Psychology: Human Perception and Performance, 38*(5), 1268-1288. doi: http://dx.doi.org/10.1037/a0026703

Signed:

Stephen C. Pritchard

14 December, 2012.

Acknowledgements

I wish to thank and acknowledge my supervisors Max Coltheart, Anne Castles and Eva Marinus for their advice, encouragement, time, and reassurance over the course of my research, through many fantastic discussions, countless emails, and several episodes of doubt. In particular, Max for giving an ex-electrical engineer a chance at being a cognitive scientist, Anne for her many “fireside chats”, and Eva for her enthusiasm and cheerful advice. Max, Anne and Eva have adeptly supported and challenged me in such a way that my confidence as an independent researcher grew, for which I am extremely grateful.

I would also like to thank Steve Saunders for his generous time and advice and always having a minute when I asked, Sallyanne Palethorpe for her generous time in transcribing voice recordings, Samantha Baggott for many good discussions and proofreading assistance, Craig Richardson for technical assistance, and Lesley McKnight, Katie Webb and the rest of the CogSci admin team for organisational assistance.

Thanks to my parents Ian and Margaret, my sister Catherine and brother Andrew for your decades of love and support, without you I would not be who I am nor where I am today. Thanks to Luca for many happy distractions and helping me stay focused on the important stuff.

Lastly and most importantly, thankyou Laura, for your love, patience, encouragement and sacrifices, your humour, and your strength. If not for you, I would never have had the wherewithal to start this PhD, let alone finish. I owe it all to you, and dedicate this work to you. Thankyou.

CHAPTER 1.
Introduction

Introduction

The dual-route cascaded (DRC) model of reading aloud and word recognition simulates the cognitive mechanisms involved in skilled reading. It is a highly successful model, one that has proven able to simulate a wide variety of effects observed in human readers (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). Despite this success, it has been criticised for being a static, non-learning model of skilled reading that is incapable of modelling the development of reading skill. This thesis aims to introduce learning to the dual-route cascaded (DRC) model of reading aloud and word recognition, retaining the existing strengths of the DRC model while incorporating this new capacity.

In this introductory chapter, I first argue why computational modelling of cognitive processes is a worthwhile endeavour, before describing the dual-route theory of reading aloud and the DRC model of reading aloud and word recognition. Following this, several other models of reading are considered, all of which include a learning mechanism. After discussing these approaches to modelling learning, I present a framework for understanding and comparing the aims and priorities inherent in each model. Next, several challenges to the suitability of the learning approach of each of these models is identified. The final sections of this introduction will introduce the approach taken in implementing learning in DRC, covering separately an orthographic word learning mechanism based on the self-teaching hypothesis of Jorm and Share (1983), and a sub-lexical learning mechanism for learning grapheme-phoneme correspondences (GPCs).

Why computational modelling?

Before considering the computational modelling of learning to read, it is worthwhile asking why computational modelling is useful in cognitive science in general. What makes

the computational instantiation of a theory any better than a verbal theory alone? I identify three broad reasons why computational modelling is useful in cognitive science.

Firstly, computational models can help reveal the ways in which a theory is incomplete. Theories may initially start out simple and intuitive, but as they are developed they can become quite complex. For example, the dual-route theory as it is embodied in the DRC model includes claims about the representation of words in a lexicon, how word frequency information is stored, how graphemes are parsed, where serial processing and parallel processing are each involved, and how the activation of a certain phoneme might decay should the grapheme parsing system receive additional information and decide that the relevant grapheme-phoneme correspondence (GPC) no longer applies. That is just to name a few of the intricacies. In covering a large amount of detail, verbal theories run the risk of overlooking crucial yet subtle aspects. Constructing a computational model, however, requires completeness, since a computational model that is only partially built will not run on a computer. The need to specify a computational model in sufficient detail for it to execute sets a minimum requirement on the level of detail required in a theory. This strength of computational modelling is identified by Lewandowsky and Farrell (2011), who, in their book on the subject of computational modelling in cognition, argue that even verbal theories that are intuitively highly plausible may turn out to be incoherent. Norris (2005) makes the same point, and discusses how hidden assumptions might be dealt with only superficially by a verbal theorist seeking to adequately explain experimental data, if they fall short of the rigour required by modelling. Coltheart et al. (2001) also acknowledges this when arguing the benefits of computational modelling of reading aloud and word recognition. Thus, computational modelling forces the researcher to consider greater detail to identify these hidden assumptions.

Secondly, computational modelling allows for effective adjudication between models and the theories they represent. Computational models are used to conduct simulations, and these simulations generate something that is akin to experimental data. Having these data allows for a rigorous and *quantitative* comparison between what the theory and model predict will happen, and what people actually do. This is a comparison where differences between theory and reality are *measured*. Lewandowsky and Farrell (2011) capture the significance of this when they state that it is typical for several theories or models to vie for the leading role as explainers of a set of data. Coltheart et al. (2001) also highlights this advantage. The literature on computational modelling of reading certainly attests to this vigorous and quantitative measuring of computational model performance versus experimental benchmarks (see Perry, Ziegler, and Zorzi (2007) for an extensive list of empirical data benchmarks that a model of reading would ideally satisfy to be regarded as an accurate model).

Finally, computational models are straightforwardly falsifiable. That is, model simulations are like specific predictions, and differences between empirical data and simulation results can constitute a falsification of the theory embodied in the model. Verbal theories can certainly be used to generate predictions which can then be assessed through comparison to empirical data. However, the greater rigour implicit in a quantitative analysis of computational model results allows for a finer degree of falsification than could be provided by a verbal model. For example, several models of reading aloud each purport to account for the frequency effect—where higher frequency words are named more quickly than low frequency words. However, Perry et al. (2007) are able to conduct an item-level variance analysis using the quantitative results of modelling simulations to argue that the CDP+ model provides the best account of the way reading-aloud latencies vary from item to item. A verbal theory of reading aloud may suggest broadly that some types of words can be named quicker than others, but without simulated data for reading-aloud latency of specific

words, a verbal theory would not be amenable to the item-level variance analysis conducted by Perry et al.

The DRC model of reading aloud and word recognition

The dual-route theory of reading aloud

The dual-route theory of reading aloud (Forster & Chambers, 1973; Marshall & Newcombe, 1973) holds that two mental mechanisms are involved in reading aloud: a lexical process and a sub-lexical process. The sub-lexical mechanism involves serially constructing a phonological representation through knowledge of how constituent parts of the word (e.g., graphemes) correspond to meaningful sounds. In contrast, the lexical process involves the automatic recognition of whole written words, without needing to parse the constituent parts of the word or recognise phonology beforehand. The recognised whole written word is known to correspond to a spoken word, and the spoken word is then uttered.

While originally a verbal theory, often described with the aid of a box-and-arrow diagram similar to that shown in Figure 1, the dual-route theory has also been developed into a computational model, the dual-route cascaded (DRC) model of word recognition and reading aloud (Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart et al., 2001; "Dual-Route Cascaded Model 1.2.1," 2009). Following is a brief description of the DRC architecture.

The computational DRC model

In accordance with dual-route theory (and Figure 1), the DRC computational model (or just “DRC”) simulates two non-semantic cognitive mechanisms involved in reading: a sub-lexical route and a lexical route. Visual feature identification and letter identification are common to both routes at the input end of the model, and phoneme identification is common to both routes at the output end of the model.

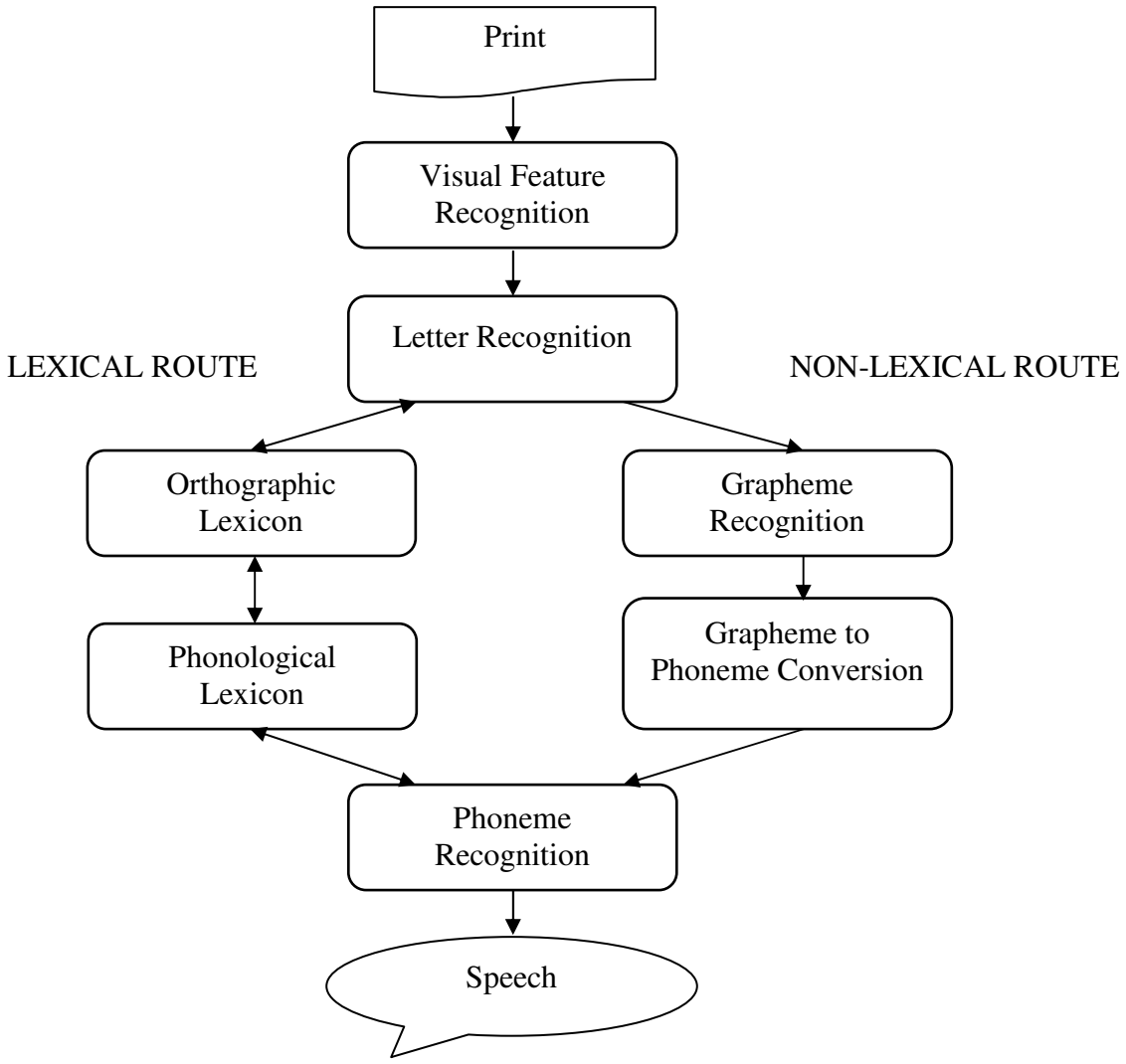

Figure 1 – The dual-route model of reading aloud and word recognition

Slot-based position coding

These common levels (visual feature, letter and phoneme) are divided into slots, with each slot corresponding to either a letter in the stimulus, at the input end of the model, or a phoneme in the output. So, for example, the stimulus CAT will have C active in the first letter slot, A in the second, and T in the third. In the phoneme layer, /k/[1] will be activated in the first slot, /{/ in the second slot, and /t/ in the third.

[1] A list of phonemic symbols used in this introduction and most of this thesis is provided in Appendix A. Note that in chapter 3, different phonemic symbols are used, which are described within that chapter. These

Representing time in the DRC model

DRC approximates the passing of time by dividing processing up into cycles, with each cycle representing a small interval of time. Activation across the model is calculated anew as the model progresses from cycle to cycle. By making the cycles small enough, the digital operation of the model approximates continuous real time. The total number of cycles taken by the model to perform a task is a measure of the response time or latency for performance of that task. While the model approximates real time operation across the presentation of one stimulus, it deviates from real-time simulation from stimulus to stimulus. This is because the network is typically reset (e.g., all activations reset to zero) at the start of each new stimulus, rather than retaining activations (a "memory") of events from one stimulus to the next.

DRC's lexical route

DRC's lexical route consists of an orthographic lexicon layer and a phonological lexicon layer, in addition to the layers shared with the sub-lexical route. Each layer has connections in both directions to adjacent layers as represented in Figure 1 by the bi-directional arrows linking each box. The visual feature layer is an exception and does not receive feedback from the letter layer[1]. The lexical route has the following characteristics:

Connectionist network: the DRC lexical network consists of layers of nodes, with weighted connections from one node to another. Each node has an activation level, represented by a real-number value between 0.0 and 1.0. Whenever a node has an activation level greater than 0.0, it contributes activation to other nodes to which it is connected. While there are a broad variety of connectionist architectures, many very different to DRC, the DRC lexical route still retains the basic features of a connectionist network. It is based on the well-

different symbols are used since chapter 3 details an experiment in which the utterances of Australian English speakers were transcribed, and the phonemes used were chosen to describe Australian English.

[1] This is expected to change in DRC 2.0 (yet to be published) with the introduction of interactivity between the visual feature layer and the letter layer.

known interactive-activation (IA) connectionist network architecture (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982).

Parallel computation: as with most connectionist networks, DRC's lexical route is a form of parallel computing. Each letter receives activation from, and contributes activation to the lexical route at the same time. This parallelism continues throughout the lexical route: the inputs to each node in each layer, and the activations of each node in each layer, are all computed in parallel.

Despite this parallel computing architecture, the DRC model is run on computers that execute instructions serially. The serial operation of the computer is a lower level consideration, and does not alter the fact that, at a higher level of description, the DRC lexical route is *functionally* computed in parallel. Another way to think of this is that the serial computer is emulating a virtual parallel model. The activity at each node is calculated anew every cycle, and the model does not advance to the next cycle until each node has been calculated, thereby simulating each node operating simultaneously, in parallel.

Graded representation: one way to model the process of a mind recognising a word is to use a binary representational scheme—either the word is recognised or it is not recognised. Another way to model recognition is to use a graded system, where a continuous level of activation from 0.0 to 1.0 is used to model recognition, such that an activation of 0.0 indicates no recognition at all, an activation of 1.0 suggests perfect recognition, and intermediate values such as 0.1 or 0.3 indicate partial recognition. The magnitude of activation suggests how strong that recognition is. DRC's lexical route uses a graded system like this, where nodes can have activation values ranging from 0.0 to 1.0. The sole exception to graded representation is that a threshold activation value is used to complete the simulation of a word. Once active phonemes in each slot all reach this threshold level, the simulation is considered complete,

and the stimulus named. There is no graded condition of completion, the simulation is either ongoing, or else it has ended.

Local representation: each node in the DRC lexical route represents specific cognitive knowledge. For example, each node in the orthographic lexicon corresponds to a particular printed word, and each node in the letter level corresponds to a particular letter in a particular position of the input stimulus. This is in contrast to many other kinds of connectionist network, where representations of specific cognitive facts are distributed across many nodes (e.g., see Rumelhart & McClelland, 1986).

Interactivity: Considering Figure 1, it might at first seem that information flow must be only in one direction, from the input at the visual feature level, down through successive levels to the phoneme output level. This is not the case. The DRC lexical route includes connections that carry information in the opposite direction too. For example, activated orthographic word nodes will contribute activation back to the letter level, and activated phonemes will contribute activation back to word nodes in the phonological lexicon. McClelland and Rumelhart (1981) demonstrate how this allows top-down expectations to support the recognition of symbols at lower levels, allowing their interactive activation model (and also the DRC model) to produce a word superiority effect, and giving the model some capacity to process degraded input. For example, if the degraded stimulus BRAK?[1] is presented, where the final letter is obscured and represented by a question mark, feedback from the word level could help to identify that the unclear letter must be an E, to form the word BRAKE. Interactivity also allows the sub-lexical route to interact with the lexical route, via feedback from the phoneme level to the phonological lexicon. This interactivity plays an

[1] Throughout this thesis we will use formatting where orthographic representations are presented in capital letters, while phonological representations will either be presented by providing the print equivalent in inverted commas and lower case, or else using phonemic symbols within forward slashes. For example, DOG (orthographic), and “dog” or /dQg/ (phonological).

important role in Chapter 2 of this thesis, to allow self-teaching, and is also mentioned in Chapter 3, when discussing lexical responses to nonword stimuli.

Excitatory and inhibitory connections: DRC employs inhibitory connections as well as excitatory connections. This means that in addition to activating compatible nodes in adjacent layers, an active node will repress incompatible nodes. For example, the letter C in the first slot has an excitatory connection to the printed word node CAT in the orthographic lexicon, but has an inhibitory connection to the printed word node DOG, so if this letter is activated, it will excite the node for CAT, but actively repress the DOG node from being excited.

Lateral inhibition: In addition to excitatory/inhibitory connections between layers, DRC includes lateral inhibition from node-to-node *within* a layer. For example, if the node for the word CAT is activated in the orthographic lexicon, then this active node will contribute inhibitory signals to all other word nodes in the orthographic lexicon.

Cascaded activation: Some models (such as the logogen model proposed by Morton, 1969) suggest that active nodes will not output signals until their level of activation reaches some threshold level. With this architecture, activation would spread through the network in a stage-by-stage fashion. Letters would need to reach threshold activation before they contribute signals to the orthographic lexicon, orthographic lexicon nodes would need to reach threshold before they contributed signals to the phonological lexicon, and so on. DRC does not use this approach. Instead, activation in DRC is *cascaded.* As soon as a node receives some level of activation, it begins contributing an output signal with a strength proportional to its level of activation. In this way, activation spreads, or cascades, through the network rapidly and smoothly, with the strength of activation growing continuously over time, rather than in a stage-by-stage fashion as would be the case if thresholds were used.

Again, the threshold to determine when reading aloud has occurred is the one exception to the cascaded approach.

"At-each-node" frequency knowledge: The DRC orthographic lexicon nodes and phonological lexicon nodes each have a resting activity that is proportional to the log of the frequency of the word represented by that node (printed frequency for orthographic nodes, and spoken frequency for phonological nodes). This enables DRC to model frequency effects. Words that are higher frequency have higher resting activity, so if a stimulus corresponding to this word is presented to the model, the word node will activate more rapidly, and reading aloud or lexical decision can occur with a shorter response time than if the word had been low frequency.

Static, pre-programmed architecture: DRC's lexical route is pre-programmed. DRC's modellers pick the words that will be in DRC's vocabulary, and dictate how nodes are connected. Connection strengths and frequency-related resting activities are chosen by the programmer, and do not change from simulation to simulation unless the experimenter changes them manually. DRC's lexical route is thus a model of a skilled reader's cognitive mechanisms for reading. By manually changing parameters, acquired impairments may be modelled (and perhaps also strategic effects), but not the developmental processes related to learning to read. Since this knowledge is pre-programmed rather than learned, DRC is considered a *static* model.

DRC's sub-lexical route

The sub-lexical route shares the visual feature layer, letter layer, and phoneme layer with the lexical route. The aspects of this route that are independent of the lexical route include a grapheme parsing mechanism and a store of knowledge governing how graphemes correspond to phonemes. Chapter 4 of this thesis describes a computational model for

learning the grapheme–phoneme correspondences section of the sub-lexical route. The sub-lexical route has the following characteristics:

Rule-based: In contrast to the connectionist structure of the lexical route, knowledge of how graphemes correspond to phonemes in the sub-lexical route is stored as a list of rules, typically referred to as grapheme–phoneme correspondence rules (GPCs). In addition to conventional GPCs that each relate one grapheme to one phoneme, DRC's sub-lexical route also includes knowledge of several rules that are not strictly GPCs. These include a multi-phoneme rule for the grapheme X (it corresponds to two phonemes, /ks/), context rules, and output rules. Context rules manipulate the GPCs that apply for a particular grapheme based on other aspects of the stimulus, (e.g., if the grapheme C is followed by an E, I or Y, then it corresponds to the /s/ phoneme, otherwise, it corresponds to the /k/ phoneme). Output rules are included to account for phonotactic and morphophonemic constraints. Output rules are applied after GPCs are applied, and alter the pattern of phonemes selected (e.g., one output rule replaces /n/ with /N/, if the /n/ phoneme precedes a /k/ phoneme).

Serial computation: Letters from each slot are not made available to the grapheme parser simultaneously in parallel as is the case for the lexical route. This is despite the letters themselves being activated in parallel within the letter layer. Instead, letters are made available to the grapheme parser one-by-one, from left-to-right, serially. This approach suggests that the sub-lexical route employs some form of attention mechanism, where attention is allocated serially to each letter. Introducing this serial element to DRC enables DRC to account for a number of effects observed in human readers that seem to demonstrate serial processing, such as the "whammy effect" (Coltheart et al., 2001; Rastle & Coltheart, 1998).

In previous versions of DRC (i.e., prior to DRC-1.2.1), letters were made available to the sub-lexical route at set intervals of time, measured in cycles. This was changed in DRC-1.2.1, after deficiencies with this approach with respect to the simulation of masked onset-priming effects were identified (Mousikou, Coltheart, Finkbeiner, & Saunders, 2010; Mousikou, Coltheart, Saunders, & Yen, 2010). Now, the introduction of a new letter to the sub-lexical route is triggered each time the rightmost activated phoneme in the phoneme layer achieves a certain activation, governed by the *GPCPhonemeExcitation* parameter. So, for the stimulus CAT, first the C will be made available to the sub-lexical route, which will contribute to the /k/ phoneme being activated. Once the /k/ phoneme reaches the appropriate level of activation, the A will be made available to the sub-lexical route, causing the /{/ phoneme to begin to receive activation from this route. Once the /{/ phoneme is sufficiently activated, the T will be made available, and the /t/ phoneme will begin receiving activation via the sub-lexical route.

Static and pre-programmed: Like the lexical route, the sub-lexical route does not learn. The logic behind grapheme parsing, and the full list of GPC rules known to DRC are all coded directly by the programmer, such that DRC simulates skilled knowledge. DRC can be made to simulate acquired phonological dyslexia by removing some of the rules from the DRC sub-lexical route, but DRC-1.2.1 has no mechanism to simulate the developmental acquisition of rules.

DRC and learning

In addition to lexical and sub-lexical route knowledge being static and pre-programmed, all other aspects of DRC are also manually coded, rather than learned. DRC is pre-programmed with knowledge of letters, knowledge of phonemes, and knowledge of the way visual features correspond to particular letters. That DRC does not model the learning process has been the focus of ongoing criticism. For example, Dufau et al. (2010) have

criticised static models (such as DRC) in general, because they do not explain the dynamic acquisition of frequency information, and instead knowledge of word frequency is manually inserted into the model by its creators. Perry et al. (2007) argue that the absence of learning specifically in the DRC model is a "major shortcoming" (p. 276). Seidenberg and Plaut (2006), also specifically criticise DRC for not including a learning mechanism. Having to explain how a skilled model has come to be skilled provides a realistic constraint on the design of the model. As a hard-wired model of skilled learning, the design of DRC was not subject to this constraint, and therefore it is arguable that DRC's structure is not necessarily realistic (Davis, 1999).

In response to this criticism, Coltheart et al. (2001) counter that constraining a model in this way is only an advantage if the constraint is realistic. That is, if the learning mechanism is not realistic, then its operation might not allow the model to ever come to approximate realistic, skilled reading or the cognitive architecture and functions associated with skilled reading. A hard-wired model of skilled reading avoids this problem by avoiding the inclusion of a potentially unrealistic learning mechanism. Ideally though, a complete model would be able to accurately account for both the true structure and functions of a skilled reader, and also realistically model the developmental processes that lead to these.

DRC remains a successful model of reading aloud, and is able to model a wide variety of benchmark effects identified experimentally in human readers (Coltheart et al., 2001), despite the specific criticism that it offers no account of learning. For this reason, a worthwhile research project is to investigate the design, construction and testing of a "learning-DRC" (L-DRC) model.

Nested modelling

Rather than putting DRC aside to build a new model of reading skill acquisition from scratch, the approach adopted here is to work with the existing DRC model and augment it. This approach is in accordance with the "nested-modelling" approach to model construction, as described in Jacobs and Grainger (1994). This modelling philosophy holds that new model iterations should be able to account for the same capabilities and effects as existing models. The benefits of a new model are diminished if, in developing the capacity to simulate a new effect, the capacity to simulate an existing effect is lost. Given DRC's success, it is sensible to work with this model to retain its existing capacities while introducing new ones. More than merely addressing criticism of DRC, the introduction of learning will increase DRC's value as an investigative tool to understand the developmental processes involved in learning how to read, while retaining DRC's existing advantages.

Computational models of reading that learn

Before deciding on the approach to learning to be taken with DRC, it is worthwhile to first consider the approach to learning taken in a variety of existing models of reading aloud or word recognition. The models considered are the triangle model family of models of reading aloud (Harm & Seidenberg, 1999, 2004; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989), the connectionist dual process (CDP) family of models of reading aloud (Perry et al., 2007; Perry, Ziegler, & Zorzi, 2010), the Self-Organising Lexical Acquisition and Recognition (SOLAR) model of visual word recognition (Davis, 1999), and the Adaptive Resonance Theory (ART) account of orthographic word form learning presented by Glotin et al. (2010).

The triangle family of models

The implicit theory of the triangle model structure is that a separate, rule-based sub-lexical mechanism is not required to generalise from word to nonword stimuli. This is in contrast to DRC, which embodies the theory that there are separate lexical and sub-lexical routes, with the sub-lexical route being crucial for the processing of nonword stimuli. That the triangle models describe only one non-semantic cognitive route between orthography and phonology is one of the most noteworthy differences between the triangle models and DRC. There are many other differences in both architecture and operation between the triangle models and DRC, and—importantly—differences in modelling philosophy between the research groups working with each model. However, I will restrict discussion here to the features of the model related to learning.

The triangle models all use some form of the *back-propagation* algorithm (Rumelhart, Hinton, & Williams, 1986) to permit learning in the model. Back-propagation first received widespread exposure after the twin-volume publication *Parallel Distributed Processing: Explorations In The Microstructure Of Cognition* (Rumelhart & McClelland, 1986), though it was independently developed much earlier than this by other researchers (e.g., Bryson Jr. and Ho (1969), as cited in Russell and Norvig (1995)). This algorithm allows networks with multiple layers of processing units (or nodes) to be effectively trained, and such networks are more powerful than their older and simpler two-layer predecessors. Multi-layer networks employ intermediate layers of so-called "hidden units", which are not accessible to the external world. They receive activation from layers closer to the input, and feed activation forward to layers at the output.

The type of learning achieved via back-propagation can be described as *supervised learning*. The model needs to be provided with both a stimulus, and the corresponding correct output for that stimulus. The model then compares its current output in response to the

stimulus with the correct output, calculates the error between the two, and propagates information about this error back through the network. This error information is used to modify connection weights to reduce the output error. From initial conditions characterised by random connection weights (i.e., no knowledge), exposure to a great number of stimuli accompanied by the correct responses allows connection weights to be honed using back-propagation. After training, the model is able to produce correct outputs to stimuli in the absence of information about what the correct response should be.

The CDP family of models

The CDP family of models are similar to DRC in that they are based on dual-route theory. The lexical route architecture of these models is almost identical to DRC, though CDP++ has a larger vocabulary that includes multi-syllabic words, and several parameters are given different values. However, the sub-lexical route of the CDP models is markedly different to DRC. Rather than using a list of discrete rules to convert graphemes to phonemes as DRC does, the CDP models use a connectionist, two-layer associative network to determine which phonemes to activate, based on the graphemes that have been identified in the input. DRC is committed to the idea that our cognitive systems include representations of *rules* in the sub-lexical mechanism, which dictate clearly and unambiguously the way graphemes correspond to phonemes. The CDP models are not committed to rules, but rather represent a commitment to a statistical relationship between orthographic word parts and phonological word parts, which allows for significantly more contextual influences in the translation of letters to sounds. Pritchard, Coltheart, Palethorpe, and Castles (2012), which is included as Chapter 3 of this thesis, provides a more detailed characterisation of these differences.

Like the triangle models, the learning approach adopted by the CDP models can be categorised as supervised learning. The CDP models are presented with stimuli matched with

correct outputs, and the CDP model learns by calculating the error between its actual output and the correct output, and then adjusting connection weights to reduce this error. This process of using error signals to modify connection weights does essentially the same thing as back-propagation, but for a two-layer network (Anderson, 1995). In a two layer network, the learning algorithm is known as the *delta rule* (Widrow & Hoff, 1960). As with the triangle models, the rate of learning is slow, and learning takes place over many trials. Perry et al. (2007) report that the CDP+ model received 50 epochs of GPC training and 100 epochs of word training (where an epoch involves presenting each word in the training corpus once). The training corpus for CDP+ consisted of 7,383 unique orthographic patterns.

The SOLAR model

The SOLAR model (Davis, 1999), like most models of reading aloud and/or word recognition, includes a theoretical account that is partially realised as a computational model. Like DRC, the theoretical account of the SOLAR model includes multiple levels of representation, including a letter level, orthographic level, phonological level, and semantic level. There are also multiple cognitive routes, such as print–letter recognition–orthography, or print–letter recognition–phonology–orthography, which is also similar to DRC. However, the computational implementation of the SOLAR model (which is what I am referring to hereafter with the term "SOLAR model") only covers the orthographic level, and is what I focus on in my account here. The SOLAR model can perform visual word recognition and lexical decision, but given that it only includes an orthographic level, it does not simulate reading aloud.

The SOLAR model learns to associate an ordered list of letters in a stimulus with a cognitive representation of an orthographic word. The model architecture consists of "item" nodes representing the individual letters in a stimulus, and "list" nodes, representing chunked representations of ordered lists of letters. SOLAR possesses multiple levels of orthographic

representation, so that it can represent parts of words in addition to complete words by doing additional chunking. For example, the word CATALOGUE is comprised of nine letters, and one level might chunk these letters into representations for CAT, A, and LOGUE, thereby identifying some of the sub-words occurring within the whole word. A subsequent level might then chunk these word parts to identify the whole word CATALOGUE.

Unlike the triangle and CDP models, the SOLAR model uses an *unsupervised* approach to learning. Rather than adjusting connection weights to minimise the error between the current output and a supplied correct output, the SOLAR model instead adjusts connection weights based on the critical features it has independently identified in various stimuli. It does this in the absence of any feedback as to what is correct. Davis designed this approach building on earlier work on Adaptive Resonance Theory (Carpenter & Grossberg, 1987) and the SONNET model (Nigrin, 1993). These features are used by the model to differentiate between different stimuli, and cluster similar stimuli. After a particular stimulus type (e.g., a particular word) has been presented enough times, the model will form a "unitised" representation of the stimuli, meaning that it has adjusted connection weights in a way that results in a single list node representing the stimuli being activated in response to that stimuli. Unlike most other models of either word recognition or reading aloud (except CDP++) SOLAR is able to process multi-syllabic stimuli, while still recognising the sub-words that may be contained within larger words. Focussing only on orthographic learning and avoiding phonology may also have enabled this, since the phonological complexities of multi-syllable words such as allocation of stress are avoided.

Adaptive Resonance Theory

The adaptive resonance theory (ART) (Carpenter & Grossberg, 1987) model of orthographic learning of Glotin et al. (2010) (hereafter referred to as simply the "ART model",) is similar to the SOLAR model in that it is an unsupervised, orthographic learning

model with no contribution from phonology. The similarity between the ART and SOLAR models is not surprising given that ART formed part of the inspiration for the development of the SONNET network (Nigrin, 1993) and in turn the SOLAR network (Davis, 1999).

Approach to modelling

Each of the models considered above has the capacity to learn, and therefore to simulate aspects of the development of reading skill. This is a facet of cognition that DRC does not yet possess. In this section, however, the modelling philosophy I have adopted is described, and this philosophy raises challenges for each of these models as an account of learning how to read.

Levels of explanation

In describing how I will proceed to analyse and characterise models of learning to read, I will introduce the idea of *levels of explanation* in cognitive science. This idea of levels hinges on a mechanistic understanding of cognition (cf. Bechtel (2008) in which levels of explanation and cognitive mechanisms are discussed in some detail).

The levels of explanation considered by researchers examining how we read could extend from the behavioural (the externally observed actions involved in reading, uttering words etc.), all the way down to low-level biological explanations (the organisation and operation of neurons and chemical processes in the brain). There is a continuum of levels in between, which I will broadly classify as either *micro-cognitive*—for levels of cognition closer to the lower-level biology, or *macro-cognitive*—for levels of cognition closer to the behavioural level. An example of an account of reading at the macro-cognitive level might be the notion that there are two cognitive mechanisms involved in reading (i.e., the fundamental proposition of dual-route theory). An example of a micro-cognitive theory might be the idea

that words are represented in the mind across a distributed network of nodes, rather than locally at a single node or "address" in memory (a micro-cognitive feature of the triangle model of reading).

None of the models described so far are biological models that attempt to simulate the behaviour of actual neurons in the brain, even though it is fair to say that the kinds of connectionist structures found in all of these models might have found some inspiration from the way neurons form networks (see Chapter 1 of Volume 1 Rumelhart and McClelland (1986) for a discussion of whether parallel distributed programming models such as the triangle model are cognitive science or neuroscience, where the authors describe themselves as cognitive scientists who enjoy neural inspiration). Each of the models of reading previously described purports to model the upper levels of cognition and behaviour, in that they are capable of specific tasks such as word reading aloud or lexical decision, and include decisions about macro-cognition, such as whether or not there are multiple cognitive mechanisms involved in reading. The models also (whether explicitly or inadvertently) model aspects of micro-cognition, since this kind of detail is frequently demanded in computational modelling in order to have a complete, executable model.

However, the models differ in where they place focus within the cognitive levels. The DRC and CDP modellers are more heavily focussed on macro-cognitive functions, structures and behaviour when compared to the triangle models, for example. The dual-route hypothesis that underpins these models is a theory about how cognition is organised at the macro level. Coltheart et al. (2001) describe their approach with the DRC model as being interested in the "functional architecture" rather than "network architecture" (p. 205), which suggests something of their intent to avoid focussing overly on low-level structural detail. The DRC and CDP+ modellers additionally focus heavily on being able to reproduce many benchmark behaviours associated with reading with their models.

The triangle modellers also seek to have their model reproduce these benchmark effects. However, their focus is much more heavily on developing general "computational principles that capture how neural activity gives rise to cognition" (Seidenberg & Plaut, 2006, p.35). They have conceived of a micro-cognitive architecture (i.e., the parallel distributed processing architecture, trained using the back-propagation algorithm) that they argue can be generalised across multiple cognitive domains, and the focus of their research seems to be to investigate this particular micro-cognitive architecture. They will therefore always retain key features of this architecture as a first priority, rather than making accurate simulation on all behavioural or psychological benchmarks the first priority.

Focus of my work: starting macro, progressing towards micro

In building on DRC, I focus on macro-cognitive structure, and am less concerned—though not unconcerned—with particular theories of micro-cognitive structure. It is also my intention to implement an approach to learning that is psychologically plausible at the highest levels of cognition. For example, if children learn to read by being directly instructed in the correct pronunciation of each word they are to learn, then that is the type of high-level fact about the way we learn to read that I would want to capture in the model. Low-level biological facts[1] related to learning to read, such as how connections between actual neurons are formed in response to exposure to print, is not the focus of this work. Micro-cognitive structure, such as the particular choice of algorithm to adjust connection weights in response to a learning experience, is also not the primary or initial focus. Nevertheless, building a computational model provides the impetus to progress from a macro-cognitive theory down

[1] The advent of brain imaging techniques such as MRI, fMRI, MEG, etc. has meant that the biology of cognition can be examined at higher levels than the level of individual neuron activity. Such research can serve to motivate the conception of new macro-cognitive theories of cognition. However, it is debateable whether or not such models thus created are models of brain biology if they don't also involve modelling brain biology at the lowest levels.

into the micro-cognitive detail, since computational modelling requires attention to this detail in order to create an executable program.

In focussing on high level cognitive and psychological plausibility when assessing models, it is clear that models with a micro-cognitive focus will not be viewed in their best possible light. I do acknowledge that these models may also be instructive when viewed with a different research focus to my own, one that is more intent on finding and testing general micro-cognitive principles. But that is not the focus here.

The project of fully characterising and comprehending the mind and brain is one of the broadest and most complex areas of research undertaken, at the frontiers of human knowledge. Perhaps a complete characterisation must involve an account that provides detail at all levels of explanation, from the lowest biological levels up to the behavioural level. But such an account is well beyond the scope of most individual research projects and outside the scope of a single PhD thesis. I will not attempt to argue that the "macro-cognitive level first" approach is the best one. Instead, I claim that the macro-cognitive level is a legitimate level to focus upon, and practical limitations on the scope of work are sufficient justification for focussing on this level.

A verbal theory can be conceived at the macro-cognitive level, but in implementing this theory as a computational model, many micro-cognitive decisions may need to be made. The approach to modelling I take in this work is to treat the macro-cognitive level as the core level to simulate, while the micro-cognitive decisions made to construct the computational model are hypotheses about lower level structure. Modelling and experimentation will likely falsify many of these micro-cognitive hypotheses without falsifying the theory at the macro-cognitive level. So, different micro-cognitive arrangements can be attempted while staying

true to the general macro-cognitive theory, and in this way, micro-cognitive levels can gradually be explored from the macro-cognitive base.

Assessing existing learning models

In this section, the aforementioned existing models of learning to read are critiqued at the macro-cognitive level, which is the level most significant to my work. The criticisms I propose include: i) a learning approach that only includes supervised learning is implausible; ii) learning that can only occur gradually over many trials due to the need to avoid *catastrophic interference* is implausible, and that iii) orthographic learning is not just an unsupervised classification task.

Implausibility of supervised learning as the sole learning mechanism

Both the PDP models and the CDP models use forms of supervised learning, where the model needs to be provided with both a stimulus and the correct response to that stimulus in order to learn. This is analogous to a beginning reader receiving direct instruction, such as a teacher explaining how to pronounce each novel word or character. However, it is implausible that children can learn so many words via direct instruction. Share (1995) argues that direct instruction cannot be the principal means by which children acquire new orthographic knowledge. Nagy and Herman (1987) as cited in Share (1995) found that 5th graders learn approximately 10,000 new words per year, and it is unlikely children acquire words at this rate through direct instruction. These findings suggest that, at the macro-cognitive level, a computational modelling approach to learning new orthographic words that only involves supervised learning is not plausible. This is perhaps less of a challenge for the CDP modellers than the triangle modellers, since the learning mechanism in the CDP models is focussed on learning the sub-lexical relationships between graphemes and phonemes, not the learning of orthographic lexical knowledge. It is perhaps more likely that this sub-lexical knowledge is

learned via direct instruction. I note however that the CDP training regime includes extensive training with whole words as input, rather than only explicit phonics training, and this amount of direct instruction in whole word pronunciation may be unrealistic.

Implausibility of long-duration training

Back-propagation has also been criticised as psychologically implausible due to the amount of training that the triangle model of reading aloud needs to undergo to become skilled (e.g., Norris (2006)). This slow training over many trials is made necessary by the risk of *catastrophic interference* (McCloskey & Cohen, 1989). This is where initially learned information can be lost if the connection weights that accurately reflect the initial information are modified by subsequent learning experiences in a way that is not compatible with the initial learning. Both the triangle models and also the sub-lexical, trained component of the CDP models are prone to catastrophic interference. To minimise its impact, triangle models are typically trained by presenting the network with an entire corpus of words to learn from, with the learning rate (the rate of change of connection strengths) set very low, so that individual learning experiences are not likely to cause either complete learning, or catastrophic interference. The training corpus is typically presented for many epochs. For example, the model of reading described in Plaut et al. (1996) was trained over 300 epochs.

While the fully trained model can deliver nice results as a model of skilled reading, the lengthy process of training seems unlike the way beginning readers learn to read. Human readers do not experience catastrophic forgetting, nor do they need to be presented with each word hundreds of times to learn to read that word and other words correctly. Indeed, beginning readers can acquire new orthographic words in as few as five exposures (Salasoo, Shiffrin, & Feustel, 1985), or perhaps even a single exposure (Nation, Angells, & Castles, 2007; Share, 2004), and learning new words does not jeopardise knowledge of previously learned words. Thus, PDP networks trained using back-propagation may have explanatory

value to investigators of micro-cognitive levels, but they do not offer the kind of macro-cognitive account of how learning to read occurs that I am focussed on modelling.

For similar reasons, the two-layer associative network used in the CDP models with its supervised learning via the delta rule is not suitable for my aims. Delta rule training also suffers from catastrophic interference (Lewandowsky & Li, 1995), and requires many trials using a low rate of learning to minimise this interference.

Implausibility of orthographic learning as a classification task

Both the SOLAR model and the ART model use a form of unsupervised learning to model the beginning reader undergoing independent learning, in contrast to the supervised learning approaches of the triangle and CDP models. Also, the number of presentations required for learning to occur is much less than what is required for either the triangle or CDP models. Davis (1999) reports that the majority of words presented to the SOLAR model are learned in one or two presentations, with only a small number of items taking considerably longer. These seem like promising approaches in that they avoid the issues previously identified with supervised learning and long-duration training. But what exactly is being learned?

Both the ART and SOLAR models are types of *classification* network. This type of network learns to classify input patterns by autonomously identifying the critical features that comprise these input patterns, and using these to distinguish input patterns from one another. This type of learning contrasts greatly with *associative learning*, which is the type of learning employed in the triangle and CDP models, where the model is trained to associate a particular input with a correct output. As classification networks, the SOLAR and ART models receive no instruction as to what critical features are important, or any information about whether a particular stimulus is a word or a nonword.

The type of learning that both the SOLAR and ART models describe is strictly orthographic, and the types of critical features identified by each model are not influenced at all by phonology. So for example, when the printed input is CAT, the ART and SOLAR models teach themselves to activate a certain "unitized" list node to represent this word, and this node classifies this sequence of letters. If learning has been successful, this CAT node should not classify different sequences, such as CAN or RAT.

Is this all that orthographic learning is, chunking groups of letters into words with no contribution from phonology or semantics? Davis (1999) certainly includes phonology and semantics in his theoretical model, but has chosen to investigate learning computationally in the absence of these. Is it worthwhile to investigate orthographic learning in the absence of phonology and semantics? I will focus here on SOLAR, though a similar critique would also apply to the ART model, were it to be used to simulate lexical decision.

SOLAR is able to perform lexical decision by first undergoing a training phase, where the model acquires lexical knowledge, after which there is a testing phase. In the testing phase, the trained network is "reloaded" after each trial, to prevent any learning from one particular trial interfering with any subsequent trials. This is the same as switching learning off for the testing phase. In addition, the testing phase includes additional structure that is specific to this phase and not involved in the learning process (i.e., an opponent processes model to perform lexical decision, using activity from the SOLAR model's various orthographic layers as input). So, learning is exclusive to the training phase, while lexical decision (i.e., distinguishing the words it has learned from nonwords or novel words) is exclusive to the testing phase.

It is sensible to investigate model performance by ceasing learning during the testing phase, so that each stimulus can be tested using the same, unaltered model. In principle

though, the model should be able to perform coherently even if it were not reloaded after each trial, and learning were allowed to continue over the testing phase. However, SOLAR's capacity to perform lexical decision is eroded if learning continues throughout testing. If training were to be ongoing during lexical decision experiments, then there is nothing to stop SOLAR from learning to classify the nonwords to which it is exposed. The SOLAR model lacks a mechanism to distinguish nonwords from words, other than that words are the stimuli to which the model has been exposed sufficiently to have unitised a representation in the training phase. Nonwords are identified as such because they are novel. If learning is active over the testing phase, then SOLAR will unitise representations of the nonwords presented, and therefore begin processing them as familiar words. According to the SOLAR account, if an adult, skilled reader were repeatedly exposed to a nonword such as BLERSK, that reader would, within a handful of exposures, unitise a representation of this nonword, and then be unable to distinguish it as a nonword. It would instead be considered a word. Rather than performing lexical decision, it is more accurate to say that SOLAR is distinguishing stimuli (whether words or nonwords) that have been previously presented from stimuli (whether words or nonwords) that are novel. This does not seem like a psychologically plausible account of orthographic learning. This also contrasts with the results reported in Zeelenberg, Wagenmakers, and Shiffrin (2004), which found that lexical decision for nonwords became easier if the nonwords had previously been presented in lexical decision, rather than more difficult or erroneous.

This isn't to say that SOLAR isn't compelling as a model of how orthographic classification might occur should a reader be attempting to process a novel stimulus in the absence of phonological or semantic information; however, it leaves out the additional contribution (whether from semantics or phonology) that must be made for a beginning reader to undergo psychologically plausible orthographic learning. The kind of orthographic learning

that enables lexical decision is not simply a classification task, it must also involve associative learning—we can discriminate nonwords from words not just because they are novel, but because they are not associated with a spoken word equivalent, or with any meaning.

Precisely the same difficulty would be faced by the ART model, were it used to simulate lexical decision (Glotin et al., 2010, describe testing the model on its acquisition of word identities, but do not test it specifically on lexical decision). Words would be distinguished from nonwords through virtue of the words having been previously exposed to the model during training, there is no other mechanism for the ART model to attempt to distinguish a word from a nonword.

A macro-cognitive model of orthographic learning

Having identified shortcomings in a variety of other computational models of learning, I now turn to a verbal theory of orthographic learning, which offers a promising macro-cognitive account of the way people acquire new orthographic knowledge. This approach avoids some of the difficulties faced by the aforementioned models. This theory is known as the *self-teaching hypothesis*, and what follows is an account of this theory. This theory will form the macro-cognitive basis for a computational account of orthographic learning, which will be described in detail in Chapter 2 of this thesis. This computational version of the self-teaching hypothesis has been constructed and tested, with results also reported in Chapter 2.

The self-teaching hypothesis

The self-teaching hypothesis was first proposed in Jorm and Share (1983), who indicate that the idea originates in the unpublished doctoral dissertation of Firth (1972). Further elaboration of the hypothesis is provided in Share (1995), and Share (2011). The self-teaching hypothesis is the notion that children can learn to read new written words without the

instruction of a teacher (i.e., without supervision). Beginning readers achieve this by combining their knowledge of spoken words with knowledge of the way sub-lexical orthography corresponds to phonology. The self-teaching hypothesis focuses on orthographic word learning, and does not consider the acquisition of other cognitive skills required for reading, such as the ability to recognise and represent letter identities, or phonemic awareness. These skills are assumed to be either already present, or acquired concurrently to self-teaching, but self-teaching does not focus on how they are acquired. It is a verbal, macro-cognitive theory, and lacks detail regarding micro-cognitive implementation.

Share (1995) suggests that there are three possible mechanisms by which a beginning reader could acquire an orthographic representation for a novel written word: direct instruction, contextual guessing, or phonological recoding.

Direct instruction involves a teacher, parent or peer instructing a learning reader as to the correct pronunciation of a printed word that is novel to the reader. Share (1995) dismisses this as the plausible primary mechanism of learning because children learn too many new words too quickly to have been directly instructed as to each one.

Contextual guessing refers to the possibility that a beginning reader could use the context provided by the words and sentences accompanying a novel word in a text to guess the correct pronunciation. For example, if a child was encountering the written word VASE for the first time, in a sentence such as “The flowers were put in a ????.”, the child might be able to guess that the word they did not recognise orthographically corresponded to “vase”, assuming they are familiar with the spoken form of this word and its semantics, and were able to read the other words in the sentence.

Finn (1977-1978) as cited in Share (1995), and Gough (1983), as cited in Share (1995), are both used by Share to argue that contextual guessing is not sufficiently reliable to

explain the ability of beginning readers to acquire the ability to read novel words without supervision. In particular, predictability is higher for function words (e.g., AND, THE) than content words (e.g., SHIP, QUICK), but it is the content words that are less frequent and most likely novel to a beginning reader.

The self-teaching hypothesis asserts that phonological recoding[1] is the primary mechanism by which beginning readers can teach themselves to read new words, in the absence of a teacher or context. Phonological recoding involves the following mental processes: 1) the identification of sub-lexical orthographic features in the stimulus (e.g., letters, or graphemes, or even larger orthographic units such as bodies), 2) the sub-lexical translation of these sub-lexical features into phonological features (e.g., phonemes, rhymes), 3) the activation of a known phonological word representation, corresponding to the phonological features that have been recognised, and 4) recognition that the orthographic stimulus therefore corresponds to a known spoken word, and then using this recognition as the basis for learning a new whole orthographic word representation for the stimulus.

This process will potentially fail for stimuli that cannot be correctly translated into phonemes matching a spoken word by the sub-lexical orthography-to-phonology translation process (i.e., if the stimulus is an irregular word). However, Share (1995) points out that novel words are not normally encountered in isolation, but in a text, and the constraints imposed by surrounding text (the context) may serve to resolve any ambiguities uncovered in the process of matching the decoded phonemes with a phonological word representation. Rather than context providing all the information required for learning though, context instead offers support to phonological recoding. For example, it is easier to guess that the correct word is "yacht" in the following sentence "I went sailing on a y???t", when the "y" and "t" have been

[1] Jorm and Share, (1983), define two types of phonological recoding, pre-lexical and post-lexical. I focus here on pre-lexical phonological recoding, but will abbreviate to just "phonological recoding."

decoded than if this *partial decoding* had not occurred. It is also easier to identify the word "yacht" when it appears in this sentence than to guess that Y???T without any context is the word YACHT.

The self-teaching hypothesis does not attempt to argue that phonological recoding is the main mechanism by which a skilled reader reads. Rather, Jorm and Share (1983) argue that it is a key mechanism for acquiring orthographic knowledge for *beginning* readers, even if phonological recoding becomes less important for skilled readers who read whole words by sight alone.

Share (1995) makes clear the point at which phonological recoding becomes useful for self-teaching. This is not at the very start of learning to read before the child has learned any words at all. Share points out that orthographic information is acquired quickly, so very high frequency items are likely to be recognised visually or taught via direct instruction with minimal phonological processing in the very earliest stages of reading acquisition. Carroll, Davies, and Richman (1971), as cited in Share (1995), found that approximately 100 items account for about half of all the words appearing in printed school English, and suggests that these items would be acquired visually via direct instruction in the earliest stages of reading acquisition. Phonological recoding is most useful for items that are completely novel to the child, or are as yet still unfamiliar because of their lower frequency of occurrence.

In early work by Jorm and Share (1983), grapheme–phoneme correspondence rules (GPCs) are considered the core of phonological recoding for early readers. They acknowledge the possibility of other forms of phonological recoding, such as the activation-synthesis approach of Glushko (1979), but point out that such an approach would most likely require a lexicon that contains phonemic segments, and these segments would not yet have been learned by an early reader. Therefore, they suggest that there might be a rule-based

segmentation in early reading, gradually shifting to a lexically-derived segmentation as reader's orthographic lexicon improves.

In later work, Share steps back from this commitment, and widens the scope of phonological recoding to include other potential processes, such as a statistical learning mechanism, activation of a connectionist network, an analogical activation-synthesis mechanism, or GPCs. Share explicitly draws the self-teaching hypothesis back from a commitment to any phonological recoding process in particular (Share, 2011).

Share also contrasts the dual-route approach and its focus on regular versus irregular words, with the dualism inherent in the self-teaching approach, which instead focuses on the familiar versus the unfamiliar. The latter "...merges the study of reading with the study of human skill learning…" (Share, 2011, p.49). The dualism in skill learning to which Share is referring is that there is a "…transition from, slow, deliberating, step-by-step unskilled performance to rapid automatized one-step or "unitized" skilled performance…" (Share, 2011, p. 50).

From verbal theory to computational realisation

The self-teaching hypothesis offers a compelling macro-cognitive account of orthographic learning, arguing that phonology is involved in the process of self-teaching. The theory avoids the criticisms made of the learning approaches of the triangle and CDP models, in that it avoids fully supervised learning, and does not imply a need for long duration training. Due to the involvement of phonological knowledge, the self-teaching hypothesis can explain the way a child might be able to distinguish words from nonwords without artificially ceasing learning, thus avoiding the difficulties regarding lexical decision and the nature of orthographic learning identified with the SOLAR and ART models. The self-teaching hypothesis is a form of associative learning, where an orthographic representation is

associated with a phonological one, and yet it is associative learning that requires no external teacher. It is *internally supervised learning* (or self-supervised learning) in that phonological recoding with support from context provides the candidate response needed to guide learning. It seems to be a promising basis for implementing a learning mechanism within the DRC model, and is fully compatible with DRC's dual-route architecture: the sub-lexical route with the aid of context support serves to train the lexical route. However, much micro-cognitive detail needs to be determined to implement this type of learning within a computational model.

One aim of my research then is to construct a new version of DRC that maintains DRC's key strengths, while incorporating a form of orthographic learning compatible with the self-teaching hypothesis' account of orthographic learning. In the spirit of nested modelling, this *learning* DRC ("L-DRC") would come to approximate the performance of the DRC model of skilled reading after it has undergone learning, in addition to including much of DRC's existing architecture. Construction of such a model involves making decisions about micro-cognitive architecture that are not necessarily covered in the self-teaching hypothesis literature. The need to fully specify down to lower levels of explanation is both a constraint and an advantage that must accompany computational modelling, with its requirement for full specification.

The design of this new L-DRC will be described in Chapter 2, offering a full computational account, and some of the design decisions encountered will be discussed. Following this, I will examine: 1) how well the model learns novel words, by seeing whether the model effectively creates orthographic nodes for each word it encounters (type learning), 2) the performance of the model in learning regular words, and 3) the role that context plays in learning irregular words in the model, and the degree to which contextual support is needed to learn such words. Whether the trained model can account for the frequency effect (token-

based learning) is important, but beyond the scope of this thesis, and a prime candidate for future research. In addition to these specific tests, the construction of a new model offers a lot of content for discussion, and this research will also cover changes, future directions, challenges encountered, and comparisons to other architectures.

Modelling sub-lexical learning

Before introducing the direction taken in modelling sub-lexical route learning, I will first describe in greater detail the operation of DRC's sublexical route. DRC's sub-lexical route consists of three broad mechanisms. Firstly, there is a serial processing mechanism for making letters available from the letter level to the sub-lexical route. This mechanism, though simple in operation, should not be mistaken for a mere low-level implementation detail. It is significant in its own right and worth identifying separately. The operation of this mechanism accounts for the strength of any serial effects evident in the output of DRC. Secondly, there is a mechanism for parsing the strings of letters made available to the sub-lexical route by the first mechanism into strings of graphemes. Finally, there is a mechanism that relates the identified graphemes to a sequence of phonemes. All three mechanisms are considered briefly, but only the acquisition of grapheme–phoneme correspondences will be studied in this research.

Serial processing and grapheme parsing

This research does not cover the learning of a serial processing mechanism, or the learning of grapheme parsing. The former is arguably an innate capacity, dictated by working memory constraints that rule out the parallel delivery of all graphemes to the parser at the same time. Grapheme parsing on the other hand is a cognitive ability that must be learned. This is particularly the case in a language like English where letters frequently contribute to the activation of phonemes through being a constituent of a multi-letter grapheme (such as TH

or OUGH), in addition to having an independent single-letter relationship with a different phoneme. Such complex relationships are not innate or trivial to acquire. The acquisition of this grapheme parsing skill is not researched as part of this work.

Grapheme-to-phoneme conversion

For DRC, the mechanism that converts graphemes to phonemes takes the form of a memorised list of discrete rules. Once graphemes have been parsed, it is a straightforward and swift computational task for DRC to consult this list of rules and determine the phonemes to activate based on the activated graphemes and letters. For DRC, this task is also unambiguous: with the exception of a small number of context rules, there is a one-to-one correspondence between graphemes and phonemes. Once the graphemes have been parsed, little logic is required in DRC to determine the phonemes to activate, and there is no uncertainty.

The existence of GPC rules of this nature is a key claim of the DRC model, one that is not universally subscribed to by all cognitive reading researchers. For example, the CDP family of models (Perry et al., 2007, 2010) instead embody the theory that the relationship between graphemes and phonemes in the sub-lexical route is a statistical one, with a many-to-many correspondence between graphemes and phonemes. That is, the activation of particular phonemes can be affected by the activation of multiple graphemes. Chapter 3 of this thesis describes new research investigating and comparing the sub-lexical routes of DRC, CDP+ and CDP++, also published in Pritchard et al. (2012). Since each of these models is able to generate data on how nonwords are pronounced, these data are compared to empirical data on how people pronounce nonwords, to adjudicate between the competing sub-lexical mechanism of each of the models.

Though the outcome of this research lends support to the DRC sub-lexical approach relative to the CDP approach, the CDP architecture affords better performance than DRC on other experimental benchmarks (Perry et al., 2007). However, the focus of my research is on creating a learning DRC. So, in the spirit of nested modelling, I will retain DRC's commitment to GPCs, and consider a learning mechanism that will learn GPCs, rather than use a CDP-like approach to the sub-lexical route, implying a large step away from DRC.

A computational model of sub-lexical route GPC learning

How do people learn GPCs? There seems to be two broad approaches: direct instruction in explicit phonics (i.e., supervised learning of GPCs), or else by implicitly inferring GPCs after exposure to many examples of whole written words matched with their spoken word equivalents, which is how GPCs might be acquired via a pure "whole language" (Goodman, 1989) approach to reading instruction. The two approaches do not need to be mutually exclusive. For example a child may learn many simple letter–sound relationships through direct instruction. Then, as they become skilled at reading, the child might implicitly learn a variety of more complex multi-letter GPCs (e.g., that EIGH corresponds to the phoneme /1/) through exposure to whole texts. This idea is similar to the ideas researched in Fletcher-Flinn and Thompson (2000), where they discuss beginning readers implicitly learning "induced sub-lexical relations" and "invented spelling", after receiving print lexical experience, although Fletcher-Flinn and Thompson distinguish two different phonological recoding mechanisms. A macro-cognitive account of GPC learning would allow either type of learning.

A mechanism similar to the self-teaching hypothesis for lexical orthographic learning, might also work for sub-lexical route learning. A sufficiently advanced lexical route might be able to produce a candidate pronunciation for a word stimulus. Then, using this candidate pronunciation, a form of implicit supervised learning could take place, where the candidate

pronunciation is used by a learning mechanism to infer GPCs. Like the self-teaching hypothesis, this kind of learning would appear to be unsupervised from a vantage point external to the cognitive system, even though, internal to the system, the lexical route might be considered as supervising the learning of the sub-lexical route.

In Chapter 4 of this thesis, an approach to modelling the sub-lexical route acquisition of GPCs is presented. The model developed is a stand-alone model and has not as yet been incorporated into a comprehensive L-DRC model. The model describes a form of learning where either GPCs can be directly taught, or else they can be inferred if a whole written word is presented to the model along with the correct pronunciation of the word. The latter is the focus of Chapter 4's research. Whole-word learning is compatible with externally supervised teaching where the correct pronunciation of the printed word is provided by a teacher, and is also compatible with internal supervision, where a candidate pronunciation for the printed word is provided internal to the cognitive system by the lexical route. However, this simple model does not yet describe the processes involved in the lexical route delivering a candidate pronunciation to the sub-lexical route for this purpose. That is a future research project. The simple model is also very much a macro-cognitive model, and incorporates a variety of arbitrary micro-cognitive decisions.

Summary

Although it offers no account of how people learn to read, the DRC model embodies a compelling theory of the cognitive mechanisms involved in reading aloud. For this reason, it seems sensible to adopt a nested modelling approach, and consider additions and changes to the DRC architecture to further this model, rather than starting afresh. The focus of this research is to build on the DRC model by augmenting it with psychologically plausible learning mechanisms.

Other models of reading aloud such as the SOLAR model and the triangle model have included learning mechanisms. However, these mechanisms do not offer psychologically plausible accounts of learning to read. Our aim is to incorporate a psychologically plausible, macro-cognitive account of orthographic learning into DRC, and then begin adding micro-cognitive detail from this base.

The self-teaching hypothesis offers a psychologically plausible macro-cognitive theory of orthographic learning. It suggests that early readers can teach themselves to correctly recognise and read novel written words by sight, without direct supervision. Early readers do so by using sub-lexical knowledge of how written symbols correspond to sounds (e.g., grapheme–phoneme correspondences) to generate a candidate pronunciation and an opportunity for learning. Supervision is internal to the reading cognitive system, while from a vantage point external to the system, reading is unsupervised. Due to its macro-cognitive plausibility, I have chosen to begin incorporating learning into DRC by developing a computational account of the self-teaching hypothesis. Chapter 2 of this thesis will describe the design and testing of an L-DRC model that attempts to model the self-teaching hypothesis using the general structure of DRC as a starting point.

This research will also examine learning in the sub-lexical route. Specifically, it will examine the competing approaches to representing sub-lexical knowledge embodied in both DRC and also in the CDP family of models. The CDP models already incorporate sub-lexical learning, and this research critically analyses the approach to learning taken in these models. Chapter 3 of this thesis will detail research undertaken to adjudicate between the CDP approach to sub-lexical knowledge and DRC's approach.

Following this, I present a model that embodies the way in which the grapheme–phoneme correspondence rules that comprise DRC's sub-lexical route may be learned. This

work is presented in Chapter 4. Chapter 5 sums up my research and includes a discussion of the promising avenues for future research stemming from this current research.

References

Anderson, J. A. (1995). *An introduction to neural networks*. Cambridge Massachusetts: MIT Press.

Bechtel, W. (2008). *Mental mechanisms : philosophical perspectives on cognitive neuroscience*. New York: Lawrence Erlbaum Associates.

Bryson Jr., A. E., & Ho, Y.-C. (1969). *Applied optimal control: optimization, estimation and control*. Waltham, MA: Blaisdell.

Carpenter, G. A., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics and Image Processing, 37*(1), 54-115.

Carroll, J. B., Davies, P., & Richman, B. (1971). *Word frequency book*. Boston: Houghton Mifflin.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100*(4), 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.

Davis, C. (1999). *The Self-Organising Lexical Acquisition and Recognition (SOLAR) model of visual word recognition.* (PhD), University of NSW, Sydney. Retrieved from http://www.pc.rhul.ac.uk/staff/c.davis/Thesis/

Dual-Route Cascaded Model 1.2.1. (2009) [Computer Software]. Sydney: Macquarie Centre for Cognitive Science, Macquarie University. Retrieved from http://www.maccs.mq.edu.au/~ssaunder/DRC/2009/10/drc-1-2-1/

Dufau, S., Lété, B., Touzet, C., Glotin, H., Ziegler, J. C., & Grainger, J. (2010). A developmental perspective on visual word recognition: New evidence and a self-organising model. *European Journal of Cognitive Psychology, 22*(5), 669-694. doi: 10.1080/09541440903031230

Finn, P. J. (1977-1978). Word frequency, information theory, and cloze performance: a transfer feature theory of processing in reading. *Reading Research Quarterly, 23*, 510-537.

Firth, I. (1972). *Components of reading disability*. Unpublished doctoral dissertation. University of NSW.

Fletcher-Flinn, C. M., & Thompson, G. B. (2000). Learning to read with underdeveloped phonemic awareness but lexicalized phonological recoding: a case study of a 3-year-old. *Cognition, 74*(2), 177-208.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning & Verbal Behavior, 12*(6), 627-635.

Glotin, H., Warnier, P., Dandurand, F., Dufau, S., Lété, B., Touzet, C., . . . Grainger, J. (2010). An Adaptive Resonance Theory account of the implicit learning of orthographic word forms. *Journal of Physiology Paris, 104*(1-2), 19-26.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5*(4), 674-691.

Goodman, Y. M. (1989). Roots of the whole-language movement. *The Elementary School Journal, 90*(2), 113-127. doi: http://dx.doi.org/10.1086/461607

Gough, P. B. (1983). Context, form and interaction. In K. Rayner (Ed.), *Eye movements in reading* (pp. 203-211). New York: Academic Press.

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, 111*(3), 662-720. doi: http://dx.doi.org/10.1037/0033-295X.111.3.662

Jacobs, A. M., & Grainger, J. (1994). Models of Visual Word Recognition-Sampling the State of the Art. *Journal of Experimental Psychology: Human Perception and Performance, 20*(6), 1311-1334.

Jorm, A. F., & Share, D. L. (1983). Phonological recoding and reading acquisition. *Applied Psycholinguistics, 4*(2), 103-147. doi: http://dx.doi.org/10.1017/S0142716400004380

Lewandowsky, S., & Farrell, S. (2011). *Computational Modeling in Cognition: Principles and Practice*. Thousand Oaks, CA: Sage.

Lewandowsky, S., & Li, S.-C. (1995). Catastrophic interference in neural networks: causes solutions and data. In F. N. Dempster & C. J. Brainerd (Eds.), *Interference and inhibition in cognition* (pp. 329-361). San Diego, California: Academic Press, Inc.

Marshall, J. C., & Newcombe, F. (1973). Patterns of paralexia: A psycholinguistic approach. *Journal of Psycholinguistic Research, 2*(3), 175-199. doi: http://dx.doi.org/10.1007/BF01067101

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88*(5), 375-407.

McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. *Psychology of learning and motivation - advances in research and theory, 24*(C), 109-165.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76*(2), 165-178.

Mousikou, P., Coltheart, M., Finkbeiner, M., & Saunders, S. (2010). Can the dual-route cascaded computational model of reading offer a valid account of the masked onset priming effect? *The Quarterly Journal of Experimental Psychology, 63*(5), 984-1003. doi: http://dx.doi.org/10.1080/17470210903156586

Mousikou, P., Coltheart, M., Saunders, S., & Yen, L. (2010). Is the orthographic/phonological onset a single unit in reading aloud? *Journal of Experimental Psychology: Human Perception and Performance, 36*(1), 175-194. doi: http://dx.doi.org/10.1037/a0015795

Nagy, W. E., & 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*. Hillsdale, NJ: Erlbaum.

Nation, K., Angells, P., & Castles, A. (2007). Orthographic learning via self-teaching in children learning to read English: Effects of exposure, durability, and context. *Journal of Experimental Child Psychology, 96*(1), 71-84. doi: http://dx.doi.org/10.1016/j.jecp.2006.06.004

Nigrin, A. L. (1993). *Neural Networks for Pattern Recognition*. Cambridge, Massachusetts: Massachusetts Institute of Technology.

Norris, D. (2005). How do Computational Models Help us Develop Better Theories? In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 331-346). Mahwah, NJ: Lawrence Erlbaum Associates Publishers; US.

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review, 113*(2), 327-357. doi: http://dx.doi.org/10.1037/0033-295X.113.2.327

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review, 114*(2), 273-315.

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the connectionist dual process (CDP++) model. *Cognitive Psychology, 61*(2), 106-151. doi: http://dx.doi.org/10.1016/j.cogpsych.2010.04.001

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.

Pritchard, S. C., Coltheart, M., Palethorpe, S., & Castles, A. (2012). Nonword reading: Comparing dual-route cascaded and connectionist dual-process models with human data. *Journal of Experimental Psychology: Human Perception and Performance, 38*(5), 1268-1288. doi: http://dx.doi.org/10.1037/a0026703

Rastle, K., & Coltheart, M. (1998). Whammies and double whammies: The effect of length on nonword reading. *Psychonomic Bulletin & Review, 5*(2), 277-282.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*(6088), 533-536. doi: http://dx.doi.org/10.1038/323533a0

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89*(1), 60-94.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel Distributed Processing: Explorations In The Microstructure Of Cognition*: MIT Press.

Russell, S., & Norvig, P. (1995). *Artificial intelligence: a modern approach*. Upper Saddle River, New Jersey: Prentice Hall Inc.

Salasoo, A., Shiffrin, R. M., & Feustel, T. C. (1985). Building permanent memory codes: Codification and repetition effects in word identification. *Journal of Experimental Psychology: General, 114*(1), 50-77. doi: http://dx.doi.org/10.1037/0096-3445.114.1.50

Seidenberg, M. S., & McClelland, J. L. (1989). A Distributed, Developmental Model of Word Recognition and Naming. *Psychological Review, 96*(4), 523-568.

Seidenberg, M. S., & Plaut, D. C. (2006). Progress in understanding word reading: data fitting versus theory building. In S. Andrews (Ed.), *From inkmarks to ideas: current issues in lexical processing* (pp. 25-49). Hove, East Sussex, BN3 2FA: Psychology Press.

Share, D. L. (1995). Phonological recoding and self-teaching: sine qua non of reading acquisition. *Cognition, 55*(2), 151-218.

Share, D. L. (2004). Orthographic learning at a glance: On the time course and developmental onset of self-teaching. *Journal of Experimental Child Psychology, 87*(4), 267-298. doi: http://dx.doi.org/10.1016/j.jecp.2004.01.001

Share, D. L. (2011). On the role of phonology in reading acquisition: The self-teaching hypothesis. *Explaining individual differences in reading: Theory and evidence* (pp. 45-68). New York, NY: Psychology Press; US.

Widrow, B., & Hoff, M. E. (1960). Adaptive Switching CircuitsInstitute of Radio Engineers, Western Electronic Show and Convention Record, Part 4 (pp. 96-104). New York:

Institute of Radio Engineers. Retrieved from http://isl-www.stanford.edu/~widrow/papers/c1960adaptiveswitching.pdf.

Zeelenberg, R., Wagenmakers, E.-J., & Shiffrin, R. M. (2004). Nonword Repetition Priming in Lexical Decision Reverses as a Function of Study Task and Speed Stress. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(1), 270-277. doi: http://dx.doi.org/10.1037/0278-7393.30.1.270

Appendix A

Phoneme symbols used in this chapter are those used by DRC model 1.2.1.

Vowels		Consonants	
Symbol	Example	Symbol	Example
1	stay	_	jump
2	sigh	b	buy
3	bird	d	dot
4	boy	f	for
5	goat	g	guy
6	mouth	h	hot
7	beard	j	yell
8	cared	k	kite
9	board	l	low
#	hard / palm	m	my
{	cat	n	no
i	seen	p	pie
u	clue	r	run
E	red	s	stop
I	bid	t	tie
Q	pod	v	vent
U	good	w	west
V	fun	z	zoo
W	few	D	then
		J	chin
		N	hang
		S	shoe
		T	thin
		Z	measure

CHAPTER 2.
Modelling the Self-Teaching Hypothesis with a Learning Dual-Route Cascaded Model of Reading Aloud

Stephen C. Pritchard, Max Coltheart, Eva Marinus and Anne Castles

ARC Centre of Excellence in Cognition and its Disorders, Macquarie University, Sydney

Abstract

The dual-route cascaded (DRC) model of reading aloud and word recognition is a static model of skilled reading that does not include a learning mechanism. As such, it has been previously criticised for offering no account of reading acquisition. We describe a new, learning-DRC (L-DRC), which provides a computational account of orthographic learning within the dual-route framework. L-DRC is based on a well-regarded psychologically plausible verbal account of phonologically-mediated orthographic learning: the self-teaching hypothesis. DRC's sublexical route allows a novel stimulus to be phonologically recoded, and recognition that the written stimulus corresponds to a known spoken word is the trigger for orthographic learning. Contextual support aids in the learning of irregular words that cannot be accurately phonologically recoded. L-DRC also includes a new approach to letter-to-orthographic lexicon excitation, which simulates cognitive resource allocation, and improves L-DRC's ability to learn successfully. L-DRC was able to effectively demonstrate orthographic learning and model the self-teaching hypothesis, though we found that some classes of words, for example, heterophonic homographs, heterographic homophones and potentiophones, are in some cases challenging for L-DRC to learn via self-teaching.

Introduction

Since at least the 1980s, computational modelling has been a popular tool to explore the cognitive mechanisms involved in reading aloud and word recognition. One of the first such computational models of reading was the interactive-activation (IA) model of word recognition (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). It described a network structure that, when presented with an input string of letters, could identify the written word to which these letters corresponded. The model was able to demonstrate how letter perception was improved when the letters were located within the context of a known word, or even in a pronounceable pseudoword. A key characteristic of the IA model was that the knowledge of words and letters embodied within it was pre-programmed by its creators, and the model did not autonomously learn new knowledge or change its own structure. For this reason, we describe it as a *static* model.

More recently, the IA model was used as a basis for the dual-route cascaded (DRC) model of reading aloud and word recognition (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). The DRC model is a computational implementation of the *dual-route theory of reading*, which has been the subject of ongoing research for many years (e.g., Forster & Chambers, 1973; Marshall & Newcombe, 1973). The crux of the dual-route theory is that there are two distinct cognitive mechanisms involved in reading aloud. One mechanism uses sublexical knowledge (such as knowledge of the sounds corresponding to each letter) to build a phonological representation of a written stimulus. The second mechanism uses lexical knowledge, and it is this mechanism that adapts the IA model. Using this mechanism, the model can recognise when a written stimulus is a word without first needing to identify the sounds that comprise this word. Accessing knowledge of the written word allows a representation of the corresponding spoken word to be accessed, and then the spoken word to be uttered.

Like the IA network, the DRC model is a static model. All of the knowledge embodied within DRC, from alphabetic knowledge, to written and spoken vocabulary, to knowledge of how to parse a written stimulus into its constituent graphemes, and knowledge of the phonemes to which those graphemes correspond, is pre-programmed and not autonomously learned by the model. For this reason, DRC is a model of a skilled reader, and does not simulate the development of reading skill, though it can be used to model acquired reading disorders by "lesioning" parts of the model (e.g., Coltheart, Saunders, & Tree, 2010).

That DRC does not explain the way we learn to read has been highlighted as a shortcoming of the DRC model (e.g., Perry, Ziegler, & Zorzi, 2007; Seidenberg & Plaut, 2006). This criticism is reinforced by the existence of competing computational models of reading aloud or word recognition that do learn, such as the triangle model (Harm & Seidenberg, 1999; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989), the connectionist dual-process models (CDP+ and CDP++), (Perry et al., 2007; Perry, Ziegler, & Zorzi, 2010), the self-organising lexical acquisition and recognition model of visual word recognition (SOLAR), (Davis, 1999), the adaptive resonance theory (ART) model developed by Glotin et al. (2010), or the self-organising map (SOM) model developed by Dufau et al. (2010). Coltheart et al. (2001) counter that unless the learning mechanism described by a model is psychologically plausible, then a trainable model of reading is of no greater use in research than a static model such as DRC. Indeed, Coltheart et al. go on to point out that an incorrect learning mechanism may be incapable of producing a trained model that accurately describes the abilities of a skilled reader. If a learning mechanism were therefore implemented in DRC, a focus would be on it describing a psychologically plausible approach to learning.

While other models of reading include learning mechanisms, these often seem to focus on the low-level detail of how the microstructure of a model might be organised, rather than

on higher-level psychological facts regarding learning. For example, the triangle model, which is trained using the *back-propagation* learning algorithm (Rumelhart, Hinton, & Williams, 1986), is often described as having an architecture and approach to learning that is not specific to reading, but can also simulate other cognitive domains, (Seidenberg & Plaut, 2006). The focus of the triangle model researchers seems to be on lower level considerations such as the way "...cognitive processing is shaped and constrained...by properties of the underlying neural substrate." (Seidenberg & Plaut, 2006, p. 35). By "lower level" we mean focussing on the microstructure of cognition and its basis in the brain, rather than on high level cognitive and psychological phenomena. Their goal is to "...formulate a set of computational principles that capture how neural activity gives rise to cognition." (Seidenberg & Plaut, 2006, p. 35). Hence, they place relatively less emphasis on attaining great accuracy in matching higher level empirical behavioural data than some other modellers of reading aloud.

Back-propagation was a revolutionary approach to training feed-forward artificial neural networks, which were then used to create models in various cognitive domains, from reading to memory. However, it is not clear that back-propagation offers a psychologically plausible account of reading skill acquisition. In particular, the triangle model must be trained very gradually, with the entire vocabulary of words to be learned presented to the model several hundred times before all words are appropriately learned. For example, the model described in Plaut et al. (1996) was trained for 300 epochs: that is, each word in the training vocabulary was presented to the model 300 times. This gradual learning over many trials must be done to minimise the impact of catastrophic interference (McCloskey & Cohen, 1989) during the learning process, which is intrinsic to feed-forward networks trained with back-propagation. This slow approach to learning is unlike the way beginning readers learn and indeed, there is research to show that children can acquire new orthographic lexical knowledge in as few as five exposures (Salasoo, Shiffrin, & Feustel, 1985), or perhaps even a

single exposure (Nation, Angells, & Castles, 2007). In addition, the approach to learning taken in the triangle model is a type of *supervised learning.* When a written stimulus is presented to the network during training, the correct pronunciation must be simultaneously provided to the network, so that it can use the information to modify its own output to be more like the correct pronunciation. As we will see later in this article, this is unlike the way children learn to read, because children are able to learn some of the skills and knowledge of reading independently, without direct instruction.

It should be noted that Harm and Seidenberg (1999) and Harm and Seidenberg (2004) intended for these versions of the triangle model to simulate real, psychologically plausible aspects of learning to read, which is contrary to our criticisms. For example, Harm and Seidenberg (1999) identified that prior to learning to read, children have already acquired considerable knowledge about the phonological structure of words from their experience with spoken language. Therefore, in seeking to make the triangle model more psychologically plausible as a model of reading acquisition, Harm and Seidenberg introduced new architecture to allow the model to learn about the phonological structure of words. They preceded reading acquisition training with training the model purely in phonological structure without any exposure to orthography. Harm and Seidenberg were the first to implement such pre-training of phonology, this is something not done in previous iterations of the triangle model, nor in other models of grapheme–phoneme association learning such as that described in Coltheart, Curtis, Atkins, and Haller (1993).

In addition, both Harm and Seidenberg (1999) and Harm and Seidenberg (2004) claim that their models can approximate the *self-teaching hypothesis* of Jorm and Share (1983). This is a curious claim, given that the form of training adopted for the triangle model is strictly supervised learning, and does not involve unsupervised or “self-supervised” training. Harm and Seidenberg (1999) argue that the provision of the correct pronunciation to their model can

either be provided by an external teacher, or, for self-teaching, via knowledge of spoken words. However, their model describes no mechanism for the correct spoken word to be selected to enable self-teaching. Perhaps it would be straightforward to include a new mechanism within the triangle model to enable the retrieval of the correct spoken word from memory. However, the triangle model subscribes to the principle that knowledge and representations are distributed across nodes and there is no orthographic or phonological lexicon, and also to the principle that there is only one direct, non-semantic route from orthography to phonology, in contrast to dual-route theory. It is therefore not clear that the model described by Harm and Seidenberg (1999) provides a means for the correct pronunciation of a word to be internally generated, or that it is compatible with self-teaching.

Harm and Seidenberg (2004) make a more detailed claim to be compatible with the self-teaching hypothesis. They claim that "Connectionist models provide a mechanistic interpretation of this type of learning [i.e., self-teaching]..." (p. 665). Their model explored the relative influence of the direct orthography-to-semantics route versus the indirect orthography-phonology-semantics route in orthographic learning and the acquisition of print-meaning associations. Although their model is trained via a form of back-propagation and thus uses supervised learning, Harm and Seidenberg again argue that the "supervised" correct meanings for each printed word do not necessarily come from an external teacher. They write, "In other cases the child can be thought of as using various strategies to derive a teaching signal rather than using an extrinsically provided one." (p. 665). Harm and Seidenberg raise the idea that contextual support sometimes provides evidence about the correct meaning. At other times, the child may internally generate the correct meaning via the "spoken word recognition pathway" (p. 665), by saying the word to themselves. It is this latter process that Harm and Seidenberg suggest is a version of Jorm and Share (1983)'s self-teaching hypothesis.

Harm and Seidenberg seem to be suggesting that in their print-to-meaning model, the orthography–phonology–semantics route might provide the correct meaning in order to train the orthographic–semantic route. While we accept that this sounds appropriate, it still raises the question: how is the orthography-phonology route trained? The triangle model still seems to suggest that this route is trained via a supervised learning process such as back-propagation. This would imply that a beginning reader learns the correct pronunciation for a printed word via direct external instruction, even if they can then self-teach meanings, which seems contrary to the self-teaching hypothesis, at least for pronunciation.

It is difficult to see how the triangle model could adhere to the principles that define the triangle model while also providing a computational account of self-teaching for correct reading aloud, and for this reason, we reject the claims of Harm and Seidenberg (1999) and Harm and Seidenberg (2004) that their models approximate the self-teaching hypothesis of Jorm and Share (1983). The computational instantiations of the triangle model all use supervised learning, and for this reason deviate substantially away from psychologically plausible learning, and from the self-teaching hypothesis. As will be described later in this chapter, modelling the self-teaching hypothesis is considered a promising approach to developing a psychologically plausible model of orthographic learning.

Other models that learn face similar difficulties, all deriving from a focus on effective low-level network training, rather than close modelling of higher level cognitive and behavioural data. For example, in implementing sublexical route learning, the CDP modellers (Perry et al., 2007, 2010) focus more on the nature of the relationships between graphemes and phonemes, rather than a psychologically plausible account of how these relationships are learned. They suggest that there is a statistical relationship between graphemes and phonemes, in contrast to DRC's adherence to an explicit rule-based conception of the relationship between graphemes and phonemes. CDP+ and CDP++ are both trained using a supervised

learning algorithm, and while they claim some level of psychological plausibility in arguing that the learning algorithm they use (the delta rule) is equivalent to a classical conditioning law (Perry et al., 2007), they provide little evidence that grapheme–phoneme relationships are only taught by direct instruction, or that the lengthy training regime used to train the CDP+/++ models is a good account of the training beginning readers require when learning how to read. Having said that, we note that, since the CDP models are dual-route models, it would not be difficult to implement a mechanism for the lexical route to provide the target pronunciations for the purpose of training the sublexical route, to allow some degree of self-teaching, and reduce the extent to which sub-lexical route learning relies on direct instruction (see Chapter 4).

A number of existing models of word recognition, such as the SOLAR model (Davis, 1999), the ART model of Glotin et al. (2010), or the SOM model of Dufau et al. (2010), seem to regard orthographic learning as simply learning that particular sequences of letters correspond to a written-word representation, without any involvement of phonology (or semantics). In contrast to the triangle model approach to learning or the CDP approach to learning, these models use an unsupervised learning approach. These models all classify input stimuli based on identified critical features, where the critical features are determined by the model itself, not provided by an external teacher. This seems promising from a psychological plausibility perspective. However, we dispute whether or not these models are truly simulating orthographic learning. These models can purportedly simulate lexical decision[1], and are able to do so because stimuli that are words are presented to the model during training, while stimuli that are nonwords are not presented to the model during training. Rather than performing lexical decision, these models are in fact just distinguishing stimuli (whether words or nonwords) that have been previously presented from stimuli that have not

[1] The ART model of Glotin et al. (2010) was not specifically built to simulate lexical decision. Rather it was tested only on whether or not word identities had been acquired, without considering nonword stimuli. However, this model would still face the same difficulty as the SOLAR model or Dufau et al. (2010)'s SOM model in simulating lexical decision, were it used to simulate this.

been previously presented. We do not think this is a satisfactory account of orthographic learning, because readers are able to perform lexical decision based on more than merely prior exposure. A reader may still learn that a stimulus they have not seen before is a word, if the stimulus corresponds to a known spoken word. That is, phonology can come into play in acquiring the skill of lexical decision. Orthographic learning must also involve learning the associations between written words and spoken words, and this is an aspect of orthographic learning not covered in the computational simulations with the SOLAR, ART or SOM models. Further to this, the self-teaching hypothesis (Share, 1995), a well-regarded and plausible account of how orthographic learning occurs, also highlights the central role that phonology plays in orthographic learning, in contrast with purely orthographic models of learning.

Creating a Learning DRC model

Our intent is to work towards creating a DRC model of reading aloud that is capable of learning, an "L-DRC". This will address one of the key criticisms made of DRC, that it does not explain reading acquisition. In pursuing this goal, we will begin by implementing orthographic learning only, and we will adhere to two general principles: nested modelling and psychologically plausible learning. What we mean by these terms is explained below.

Beginning with orthographic learning

Acquiring reading skill, as understood in the context of dual-route theory, involves the learning of a variety of cognitive sub-skills, including knowledge of the alphabet and capacity to perceive letters, phonemic awareness, grapheme–phoneme correspondences (GPCs), a written word vocabulary, a spoken word vocabulary, and the associations between letters, written words, spoken words and meanings. In reality, children most likely learn many of these sub-skills simultaneously. For example, a realistic progression might overlap the learning of letter identities with direct instruction in high frequency simple words; or learning

GPCs might overlap with orthographic learning and with the expansion of the spoken word vocabulary.

It is tempting to try and implement the learning of all of these cognitive facets simultaneously, in order to maximise psychological plausibility. However, this would result in a model so complex it would be challenging to begin exploring the model and testing it. Instead, in the work described here, we made some simplifications, so that our research could proceed in an incremental fashion. The L-DRC that we constructed assumed complete knowledge of letters, GPCs, grapheme parsing, phoneme identities, and a full spoken vocabulary (full as in equal to DRC's, though limited to only monosyllabic words, as is the case for DRC). The idealised reader we modelled is of course unrealistic, but provided a good starting point for making sense of complex cognitive processes.

The main aim of this study is to implement the cognitive skill of orthographic learning as a computational model. By orthographic learning in this context, we mean the creation of a unitised representation of the written word within the reader's mind. This representation is associated with the appropriate semantic and phonological representations for the same word. The existence of this orthographic representation allows the reader to read and understand the word without first needing to recognise the phonological representation, or decode the word on a sub-lexical level. Orthographic learning is what enables a reader to become swift and automatic in their reading, and this capacity to read so effortlessly has been described as the "quintessence" of reading skill, (Share, 2008, p. 34). Because orthographic learning is so central to learning to read, this is where we start our implementation of a learning DRC model.

In focussing on orthographic learning, we will look to produce a computational account that is accurate at a high level of analysis. That is, we focus on high-level cognitive detail, rather than the intricacies of the cognitive processes involved in learning to read. Our

initial model design as described in this article will necessarily be quite simplistic when viewed from a low-level. Subsequent model iterations can be focussed on slowly delving deeper into network architecture and lower-level cognitive considerations, once the high-level account has been tested and found to be robust.

Nested modelling

Despite being a static model, DRC has been highly successful on other measures as a computational model of reading aloud. There are many standard effects seen in data from human subjects reading aloud or doing lexical decision that are successfully simulated by DRC (Coltheart et al., 2001). For this reason, rather than abandoning the DRC model entirely in order to create a learning model, we will seek to retain as much of DRC's structure and performance as we can, and only make changes to introduce or allow learning to take place. This is in accordance with the principle of nested modelling, (Jacobs & Grainger, 1994). Nested modelling holds that new models should retain the explanatory capabilities of existing models, even as they seek to add new capacities. If not, it is unclear whether or not the new model can be said to represent an advance in knowledge.

Psychologically plausible learning: the Self-Teaching Hypothesis

A well-argued verbal model of psychologically plausible orthographic learning exists, and is known as the *self-teaching hypothesis* (Firth, 1972; Jorm & Share, 1983; Share, 2008, 2011). The initial conception of this hypothesis was that beginning readers can teach themselves to read and recognise new written words without receiving explicit instruction from a teacher for each new word, and without relying solely on contextual guessing. They are able to do this through "…the ability to translate unfamiliar printed words into their spoken equivalents ('phonological recoding' or simply 'decoding')", which is "… the central means by which orthographic representations are acquired", (Share, 2008, p. 35). Once the printed word has been translated into a phonological representation, if this candidate

pronunciation corresponds to an already known spoken word, then an opportunity for orthographic learning is created, since the reader will recognise the spoken representation.

These ideas are compatible with DRC. DRC's sublexical route can perform the role of phonological recoding, since it takes letters as input, and generates a string of phonemes as output. A match between this sublexical route output and an item in the phonological lexicon (which, in DRC, already contains a vocabulary of almost 8,000 spoken words) can serve as a trigger for orthographic learning. So rather than a teacher providing the correct reading aloud of a novel written word to supervise learning, DRC can "self-teach", because the sub-lexical route provides the (hopefully correct) reading aloud of the written word, and it is this that is used to train orthographic learning in the lexical route.

It is important to be clear on the concepts hypothesised as part of self-teaching, and those hypothesised as part of DRC's account of reading. Computationally implementing the self-teaching hypothesis within the DRC model necessarily involves committing to a specific cognitive account of phonological recoding: that it is achieved via the use of explicit GPCs to translate from print to speech at a sublexical level. However, the self-teaching hypothesis itself does not specify the precise nature of the cognitive mechanism employed to perform phonological recoding. Share (2008) emphasises that this mechanism may be the application of explicit grapheme–phoneme correspondence rules (GPCs), but it may also be other mechanisms such as analogical mechanism (for example, as suggested by (Glushko, 1979) and computationally implemented in the triangle model (Seidenberg & McClelland, 1989)), or a statistical learning mechanism. So although DRC's sublexical route is compatible with the self-teaching hypothesis, it makes theoretic commitments that are not specifically made as part of the self-teaching hypothesis.

Self-teaching, rather than direct instruction is an important quality of a psychologically plausible account of learning to read, for the following reasons. Share (1995)

argues that externally supervised learning, such as a teacher giving direct instruction, is not a realistic account of the way children acquire orthographic knowledge. This is because children acquire new words at a rate far too high for a teacher to possibly be individually instructing a child on each word. For example, Nagy and Herman (1987), as cited in Share (1995), suggested that a typical 5th grader encounters approximately 10,000 new words per year. Share does point out that initially, perhaps 100 high frequency words are taught via direct instruction, but that self-teaching plays a central and essential role in learning to read the bulk of written words.

Share also considers the possibility that contextual guessing could guide self-teaching, as opposed to phonological recoding. A beginning reader typically encounters novel written words within a whole text, and feasibly they could guess the correct pronunciation for a written word by constraining the potential options based on the context provided by the balance of the text. For example, in the sentence "They climbed aboard the _____ and set sail", it is much more likely that the missing word will be "yacht" than "basketball". Share rejects the possibility of context doing the bulk of the cognitive work involved in self-teaching because natural text is in general not highly predictable (see Finn (1977-1978), as cited in Share (1995)). Also, the words that are most amenable to unambiguous contextual guessing are high frequency and/or function words (e.g., THE, AND, THAT), the kinds of words beginning readers probably already know via early direct instruction. Low frequency content words are precisely the kinds of words that children are likely to have to teach themselves, and yet these are the words that are difficult to unambiguously recognise using context alone. In the previous example sentence, the missing word might be "yacht", but it could just as well be "ship" or "boat" or "catamaran".

A challenge for the self-teaching hypothesis is that many words in a language such as English are irregular or inconsistent, and so the process of phonological recoding will not

always produce a candidate pronunciation that matches with a known spoken word. This is because irregular words are not pronounced according to the sublexical orthographic–phonological relationships that support phonological recoding. Despite this, Share (1995) points out that even highly irregular words such as YACHT have at least some regularity (e.g., the Y and T are pronounced regularly). Learning irregular words can therefore be assisted via "partial decoding" (Share, 1995, p. 166). According to the self-teaching hypothesis, the role of the contextual support provided by, for example, the text accompanying a novel word, is to constrain the possible options for a matching spoken word. A beginning reader can then use partial decoding along with this shortlist of potential spoken words that is constrained by context, to determine what the correct reading aloud of a novel written irregular word should be. Contextual guessing provides support for reading aloud irregular words, but cannot facilitate this on its own without partial decoding to assist. Turning once more to the example sentence, a beginning reader might be able to choose that the unfamiliar word is "yacht" rather than an alternative word for a sea vessel if they are able to partially decode the word and see that it must start with "y" and end with "t".

Summary of our research aims

We aimed to build a psychologically plausible DRC model that includes the capacity for orthographic learning, an L-DRC. Existing approaches to learning in use in other computational models are often focussed on microstructure, and the ability to generalise the microstructure across multiple cognitive domains, at the expense of psychological and behavioural plausibility at higher levels of analysis. Plausibility at these high levels of analysis is the focus of our work. The self-teaching hypothesis is a psychologically plausible, high-level verbal account of how a beginning reader could teach themselves new orthographic word forms. In seeking to incorporate learning into DRC, we will focus on implementing a computational version of the self-teaching hypothesis.

In accordance with the principles of nested modelling, we retained the capabilities of the existing DRC model, even as learning is implemented. To do this, we retained most of DRC's current structure. Changes to DRC's structure were typically only made to allow for learning, or to avoid shortcomings to learning that were unavoidable without changing structure. We strove to avoid making changes that would alter DRC's general performance against empirical data benchmarks.

Architecture of L-DRC

Introduction to the architecture

We intended to model phonologically driven orthographic learning, while retaining as much of DRC as possible. So a logical place to start is with the existing DRC model. A description of DRC's architecture can be found in Coltheart et al. (2001). This describes the first iteration of DRC, 1.0. Our research will use the most current publicly available version, DRC-1.2.1[1] ("Dual-Route Cascaded Model 1.2.1," 2009).

A box-and-arrow model of DRC's basic architecture is shown in Figure 1. The two cognitive routes or mechanisms are shown, and are labelled as the lexical route and the sub-lexical route. Note that the semantic layer has not been computationally implemented in DRC-1.2.1, but a rudimentary semantic system will be included in L-DRC to allow for contextual influence. Some boxes and lines are shown faded. These are the sections of DRC that contain knowledge that L-DRC will learn. These areas consist of the orthographic lexicon, the connections from the orthographic lexicon to the letter level, and connections from the orthographic lexicon to the phonological lexicon. These greyed structures are analogous to a human reader's knowledge of written words (e.g., DOG), their knowledge of the particular letters that comprise each written word (e.g., that DOG is comprised of D, O,

[1] Documents describing subsequent changes to DRC's architecture since the 2001 publication can be downloaded at http://www.maccs.mq.edu.au/~ssaunder/DRC/

and G), and their knowledge of the associations between these written words and spoken words (e.g., that DOG is the written form of the spoken word "dog"). Note that the greyed areas do not include knowledge of actual letters, or knowledge of spoken words. Although we modelled an idealised reader that already knows the alphabet and has a spoken word vocabulary, this beginning reader still needs to learn the associations from the orthographic lexicon to the letter level and to the phonological lexicon.

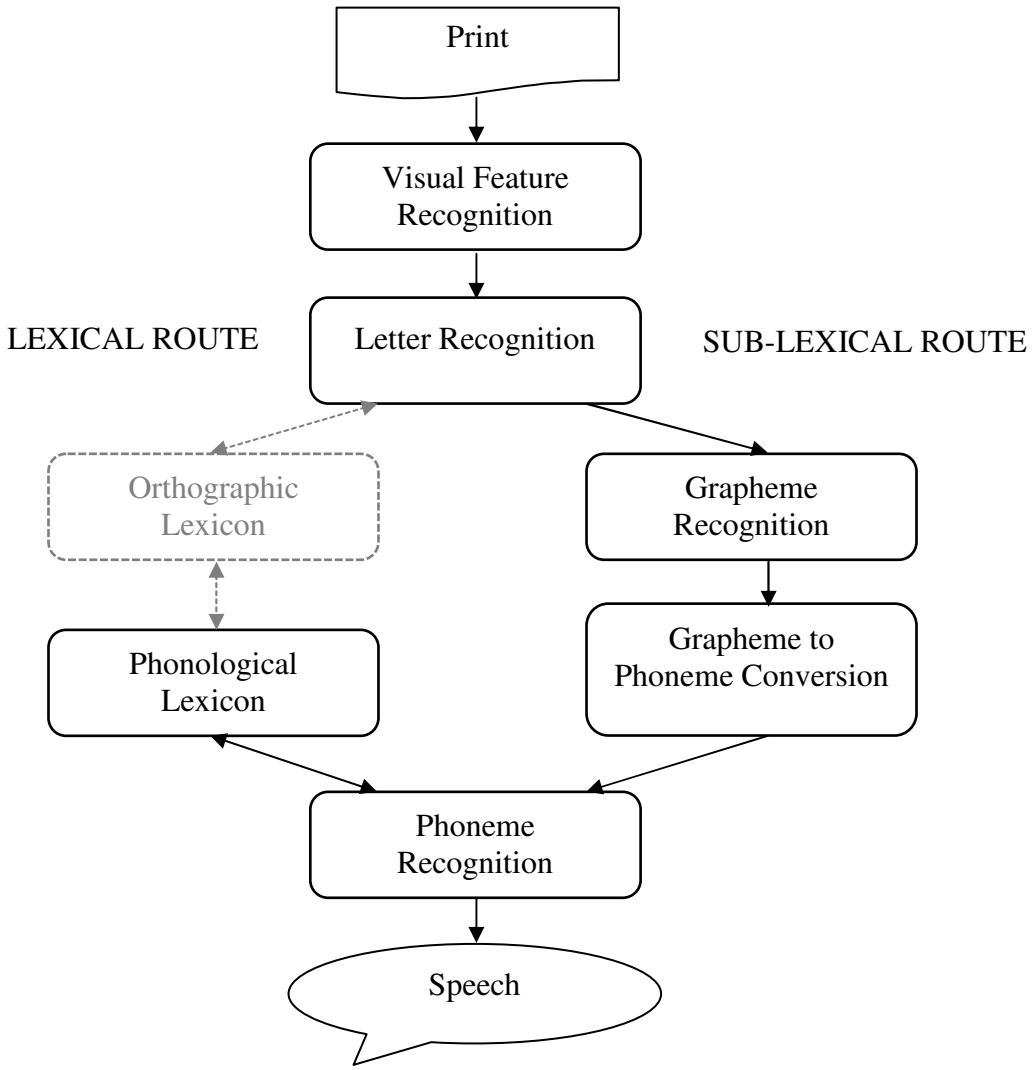

Figure 1 – The DRC model of reading aloud and word recognition. The greyed out features of the diagram are sections of the cognitive mechanism of reading that are to be learned by the L-DRC model.

L-DRC (prior to any training) consists of a newly coded[1] replica of the standard DRC-1.2.1 model, but with the greyed-out areas removed completely. L-DRC can therefore be understood as a model of an idealised beginning reader, who has acquired a skilled knowledge of printed letter representations, and is able to parse graphemes in strings of printed letters. It has knowledge of all phonemes in English, a skilled knowledge of English GPCs, and a full vocabulary of all monosyllabic spoken words in English. This idealised reader is also able to identify all of the phonemes present in each of the words in their spoken word vocabulary.

Operation of L-DRC learning mechanisms

To illustrate the operation of learning in L-DRC we will describe the step-by-step presentation of a novel stimulus to L-DRC.

Activation of the phonological lexicon via the sub-lexical route

When a novel stimulus is presented to L-DRC, it causes activation in the visual feature and letter layers. As is the case for DRC-1.2.1, the sub-lexical route takes these activated letters and generates a regular pronunciation according to GPCs. Phonemes are thereby activated, corresponding to this regular pronunciation. Activated phonemes excite the phonological lexicon, since the phoneme layer and phonological lexicon layers are interactive. Through this process, the sub-lexical route drives the activation of a spoken word representation without any prior activation of a written word representation in the orthographic lexicon. This is analogous to the concept of "phonological recoding" in the self-teaching hypothesis.

[1] DRC-1.2.1 (the already existing model of skilled reading) was coded in C++. However, L-DRC is coded in a different programming language: C#. There were two reasons for this. Firstly, C# is the language of choice known by the first author. Secondly, reconstructing DRC as a first step on the way to building L-DRC was a highly useful, hands-on process for the first author in becoming intimately familiar with the low-level computational architecture of DRC. The newly coded DRC-1.2.1 replica was tested using both words and nonwords, and found to produce the same responses with the same response times to the original, C++ version.

Spoken-word recognition as a trigger for learning

L-DRC uses a threshold level to initiate learning. When a node in the phonological lexicon reaches a critical level of activation, the written stimulus is considered to have been identified as corresponding to a particular spoken word. This signifies that a learning event can commence. We now introduce the first new parameter included in the L-DRC model: *SpokenWordRecognisedThreshold* (a full list of parameters used in L-DRC is included in Appendix A). This parameter allows the experimenter to set the level of activation required of a node in the phonological lexicon to initiate learning. At the completion of the reading aloud simulation, if a phonological node has been activated above this threshold, then a learning event begins. A phonological lexicon node activated above this threshold models a reader having recognised that the written stimulus corresponds to a known spoken word.

Once learning is triggered by recognition that the stimulus corresponds to a spoken word, there are two possibilities for learning: learning a completely novel written word, which is *type-based* learning, or else improving knowledge of a written word that has previously been seen, which is *token-based* learning.

Learning a novel written word (type-based learning)

If the stimulus is a novel written word, then no orthographic word node should have been significantly activated by the stimulus. While it is possible that neighbours of the stimulus could receive some activation from letters shared with the stimulus, in practice, this will rarely occur since DRC-1.2.1 (and thus L-DRC) include a high level of letter-to-orthographic-lexicon inhibition by default. This inhibition is typically sufficient to prevent the activation of orthographic lexicon nodes where even just one letter differs with the stimulus.

If learning has been triggered by a phonological lexicon node having reached an above-threshold level of activation, and no orthographic lexicon node is sufficiently activated, then learning of a new orthographic word is triggered. We now introduce the second new

parameter exclusive to the L-DRC model: *WrittenWordRecognisedThreshold*. Learning a new written word should only occur if a beginning reader does not already recognise the written stimulus. Therefore, in L-DRC, the learning of a *new* orthographic node is only triggered if there is no orthographic node that is activated *above* this threshold value. This models the beginning reader not recognising the written stimulus. When this event is triggered, L-DRC will undergo the following structural changes:

- A new node is created in the orthographic lexicon
- Bidirectional excitatory and inhibitory connections are created between this node and the relevant nodes in the letter layer. That is, excitatory connections are formed with the most active letter nodes in each slot, and inhibitory connections are formed with other letter nodes.
- Bidirectional excitatory connections are created between the new orthographic node and the most active node in the phonological lexicon, and bidirectional inhibitory connections are created between the new orthographic node and all other phonological lexicon nodes.
- Inhibitory connections are created between this new orthographic node and all other orthographic nodes (lateral inhibition).
- The frequency value attached to this node is set to 1, and then multiplied by *WrittenWordFrequencyMultiplier*, where this new parameter is used to control the speed of learning token-based information about the stimulus.

The type-based learning of a novel stimulus in L-DRC is completed in a single learning event. The connections to and from this newly created node are not formed with low values that will then increase with each subsequent exposure. Rather, the new connections are created at full strength similar to what they would be in DRC-1.2.1 for skilled knowledge of a written word. The strength of skilled connections is set by the experimenter, who can alter

these values by changing the relevant parameters, the same as they would do in DRC-1.2.1. L-DRC's knowledge of a newly learned orthographic word is therefore the same as DRC-1.2.1's skilled knowledge of a very low frequency word. The implications of this approach to type-based learning are covered in the discussion of architecture following.

Increasing knowledge of an already known word (token-based learning)

If a beginning reader encounters a written word they have seen before, they will recognise it, and will not need to learn a new representation for this word. To model this in L-DRC, if the written stimulus has previously been presented to the model and type-based learning has already resulted in the creation of an orthographic node for the written stimulus, then that orthographic node will likely be activated via excitation from activated letters to a value above the *WrittenWordRecognisedThreshold* value. Activation of an orthographic node above this threshold is the way L-DRC models a reader recognising that the written stimulus is a word already known. So, if a learning event has been triggered by a sufficiently activated phonological lexicon node, then, the existence of an orthographic node also above threshold will trigger token-based learning. This type of learning results in the frequency value associated with the most active orthographic node being incremented by the *WrittenWordFrequencyMuliplier* value. This parameter is included for practical reasons, to allow frequency values to grow to a size commensurate with the frequency values employed in DRC-1.2.1, but without needing to present hundreds of thousands of stimuli to achieve this (e.g., the most frequent word, THE, has a frequency value in DRC-1.2.1 of over 100,000). In this way, L-DRC's familiarity with already known written words is increased, so that it is able to generate phonology for such words more rapidly.

When does learning occur in L-DRC?

At present, the computations for learning in L-DRC occur after the reading aloud simulation has completed. That is, L-DRC functions identically to DRC-1.2.1 for the purposes

of reading aloud a word, though obviously with an orthographic lexicon that only contains nodes for words that have been learned. Once reading aloud has been completed (which occurs once phonemes are activated above the threshold set with the *MinReadingPhonology* parameter, see Coltheart et al. (2001) for a detailed account of DRC's reading aloud procedures), then cycle-by-cycle processing and activation changes cease, and the model begins the learning phase, where nodes in the phonological lexicon and orthographic lexicon will be checked to see if any node has reached the threshold level of activation to trigger a learning event.

Future research may explore modifying timing, for example the commencement of learning could be triggered by a node in the phonological lexicon reaching a threshold value, instead of being delayed until after reading aloud has completed.

Snapshot of L-DRC structure after learning

After learning, L-DRC should possess orthographic nodes for all written words to which it has been exposed, and the relevant connections to adjacent layers. It will have internalised token-based information in the form of node-specific frequency parameters, with the frequency value associated with a particular orthographic node being proportional to the number of exposures to the written stimulus represented by that node. That is, L-DRC after training should come close to looking and behaving just like the static, skilled DRC model.

Having provided a general account of the main learning mechanism in L-DRC, we now turn to the many exceptions and challenges that will occur in trying to computationally model orthographic learning, especially for a language with many irregular words, such as English.

Context and irregular words

Activation of phonemes via L-DRC's sub-lexical route is only guaranteed to result in the activation of a node in the phonological lexicon if the written stimulus is a regular word, or a pseudo-homophone (e.g., BRANE, "brain"). If the stimulus is an irregular word, or is a non-word other than a pseudo-homophone, then the sequence of phonemes activated will typically not correspond perfectly to a spoken word, (except in cases where the rule-based translation of an irregular word is itself a different actual spoken word, e.g., COME is "comb" when regularised. Such words are known as *potentiophones* (Friedmann & Lukov, 2008)). By choosing appropriate parameter settings, it is possible that a phonological lexicon node might be activated by phonemes where there is only a partial match. However, DRC-1.2.1's default parameters (and thus also L-DRC) include a high phoneme-to-phonological-lexicon inhibition. This high inhibition mostly prevents activation of any node in the phonological lexicon unless there is a perfect match between the activated phonemes in each slot and a word node in the phonological lexicon. With these default parameters, some other support mechanism will need to assist the sub-lexical route in triggering a learning event for irregular words, and even with parameter changes, this support mechanism might still be required for highly irregular words. How might this be done?

As described in Chapter 1, Share (1995) argues that contextual support can be used to support a partial decoding. Share suggests that it is typically just the vowels in English that contribute irregular pronunciations, while consonants are mostly pronounced regularly. Beginning readers are able to make use of this partial regularity to generate a partially-correct decoding. This would provide some degree of useful information to support the triggering of a learning event, but some additional support is required.

Context is not trivial to model in detail, due to the myriad associations that must comprise the links between concepts, and between concepts and words within the cognitive

systems for reading. This complexity is no doubt a primary reason why detailed semantic representation is frequently omitted from computational models of reading (e.g. neither DRC nor CDP++ computationally model the lexical-semantic route, and nor did early versions of the triangle model, though Harm and Seidenberg (2004) is one example where some of the complexity of the semantic system is computationally implemented in a triangle model). Sometimes when the semantic system is specifically considered, it is modelled in a simple or partial fashion, for example Coltheart, Woollams, Kinoshita, and Perry (1999) included the capacity to activate semantic representations for colour, in order to computationally model the Stroop Effect. Rather than producing a detailed computational model of the semantic system, we also take a simple approach to modelling the impact of context. We only model the provision of activation to the correct word, and do not attempt to model the activation of all of the words that might be relevant to a particular context.

L-DRC's simple semantic system consists of a new layer, containing a node for each word in the L-DRC spoken-word vocabulary. This is shown in Figure 2. Activation of a semantic node corresponds to recognition of the semantic content represented by that word. Each node in this semantic layer has an excitatory connection to the corresponding word node in the phonological lexicon, and inhibitory connections to all other word nodes. When a written stimulus is presented to the model, the visual feature layer is excited, but now also the semantic layer receives direct excitation. This activation represents the activation of concepts due to the action of context. By context, we typically mean the written context provided by the text in which the word is embedded, though conceivably the context could be provided by the direct instruction of a teacher. The excitation of the semantic layer is modulated by the parameter *netInput2SemanticNode*. The value of this node can be altered to represent the confidence and focus that context provides to the identification of a particular word and meaning. Connections from the semantic layer to the phonological lexicon are governed by the parameters *semantic2PhonolexExcitation* and *semantic2PhonolexInhibition*.

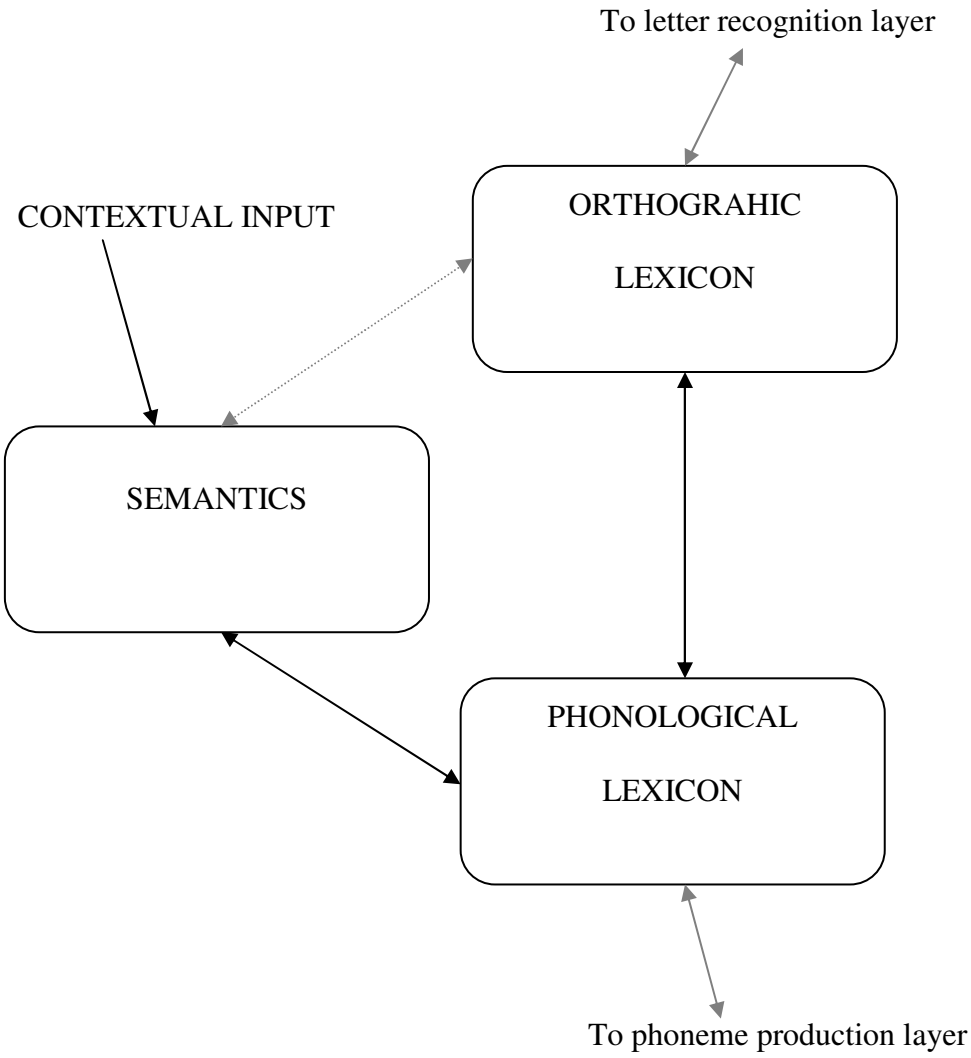

Figure 2 – Including a semantic layer in L-DRC. Note that the associations between the semantic layer and orthographic lexicon are notional, and have not been modelled as yet in L-DRC

Changes to connectivity between letters and words

Early test simulations with L-DRC revealed an incompatibility between DRC's current structure and adequate learning. For this reason, we introduce a quite fundamental change in L-DRC's architecture away from the structure used in DRC. This change and the reasons for it are described as follows.

In DRC-1.2.1, all orthographic nodes are connected to all letter slots. This is the case even for short words that do not take up the full eight slots. Consider the word CAT. The orthographic lexicon node for this word will have excitatory connections from the C node in

the first letter slot, the A node in the second letter slot, and the T node in the 3rd letter slot. It will *also* have excitatory connections from the null character in the remaining slots four to eight. This seems intuitively strange. It means that the orthographic word node CAT receives more excitation from five null characters than it does from the three letters that actually comprise it!

This odd connectivity was no doubt implemented in DRC-1.2.1 to avoid introducing unwanted length effects into the lexical route. If words were only excited by a single null character in addition to the letters they contain, then longer words would receive greater activation than shorter words (more letters to contribute excitation) and thus would be more quickly read aloud. Instead, Weekes (1997) found that there is little effect of length for high frequency words, and for low frequency words, having more letters typically means longer reading aloud latencies, not shorter. Ensuring that all orthographic word nodes, whether these are for long or short words, receive excitation from eight letter slots eliminates this unwanted length effect from the lexical route, even if the means of achieving this seems somewhat counter-intuitive.

This pattern of connectivity, however, poses difficulties for a learning model, especially for learning two close neighbours, or for learning longer words that contain a subset of letters that comprise a shorter word (e.g., AN and ANT). For example, let us say that the word AN has already been learned, and an orthographic node already exists for this word. Now the word ANT is presented but is novel to the model, meaning there is no orthographic node for ANT. The intended outcome is for the phonological lexicon node for “ant” to be activated, but for no node in the orthographic lexicon to be activated above threshold, so that a type-based learning event is triggered leading to the creation of a new orthographic node for ANT. However, with DRC-1.2.1’s pattern of connectivity, there is only one letter slot out of eight difference between A-N-T-+-+-+-+-+ (where ‘+’ is used to represent a null character),

and A-N-+-+-+-+-+-+. So if ANT is the written stimulus, the existing orthographic word node AN will receive seven slots worth of excitation and only a single slot worth of inhibition. There is a good chance that the AN orthographic lexicon node will be inappropriately activated under these conditions, since there is no better matching orthographic node yet in existence to laterally inhibit its activation. If the orthographic node AN is activated above threshold by the stimulus ANT, this would prevent an orthographic node for ANT from ever being learned, and instead "ant" would be learned incorrectly as another pronunciation of AN, and AN would erroneously seem to L-DRC as a heterophonic homograph. To compound this, a kind of "snowball" error can then occur, where the node now associated with "an" and "ant" repeatedly has its token frequency incremented with each exposure, and becomes more likely to incorrectly become associated with even more neighbouring words (e.g., A, ANTE, ANTS, then RANTS).

In addition to being a problem for learning, DRC-1.2.1's connectivity between all letter slots and orthographic word nodes would also become a problem once DRC is augmented to be able to handle much longer multi-syllabic words. A DRC model using the current pattern of connectivity, with enough slots to process the word ANTIDISESTABLISHMENTARIANISM (28 slots) would use 26 null letter slots to excite the 2-letter word AN, in addition to the A and N. Such a model might have a good deal of difficulty distinguishing AN from ANT, since these two words would have 27 letter slots worth of excitation in common, and only a single slot different.

From an intuitive point of view, and to avoid this problem of lexical capture by neighbouring words, it is preferable for orthographic word nodes to only be excited by their constituent letters, plus a single null character. This, however, still leaves an unwanted length effect present. This issue can be overcome by implementing *normalisation* of the excitatory activity from the letter later to the orthographic lexical layer.

Normalisation to avoid the impact of stimulus length effects or for other reasons has been implemented in a variety of other well-known artificial neural network paradigms. For example (Davis, 1999) implemented normalisation of input activity in the self-organising lexical acquisition and recognition (SOLAR) model of visual word recognition. The ART-1 model described in Carpenter and Grossberg (1987) uses normalisation (scaling) to ensure that a small difference in features between an input pattern and a prototype pattern will be treated as noise when the input pattern is complex, while the same small difference when the input pattern is simple will be correctly identified as being significant.

We implement normalisation by ensuring that total letter-to-orthographic-lexicon excitation is equivalent to eight slot's worth of activation. The excitation coming from each letter node to the orthographic lexicon is multiplied by a normalisation multiplier.

The normalisation multiplier is defined as:

$$Normalisation\ multiplier = \frac{total\ number\ of\ slots}{length\ of\ stimulus + 1\ for\ a\ single\ null\ slot}$$

Normalisation is best illustrated with an example. Consider the stimulus STOP. This stimulus will result in S, T, O, and P nodes being activated in letter slots 1 thru 4, and null characters being activated in slots 5 thru 8. The orthographic word node for STOP will only have excitatory connections coming from the first 5 slots (for each of the letters in the word, and a single null character). The excitation contributed by each of these letters will therefore be divided by 5, so that the total excitation is roughly equal to a single slot's worth of activation. Following this, excitation is multiplied by the total number of slots which is 8, giving total excitation roughly equal to the excitation that would have been provided by all

slots contributing excitation. For the stimulus GO, the excitation contributed by the letters G, O and a null character, will be each divided by 3 before being multiplied by 8.

Introducing this normalisation represents a departure from the original DRC structure of version 1.2.1, and so its impact will need to be tested. Our claim is that undertaking this normalisation will avoid unwanted length effects on reading-aloud latency, will avoid lexical capture of novel word stimuli by already learned close neighbours, while still maintaining the strengths of the existing DRC approach.

Theoretical justification for implementing normalisation

Consider three different models: 1) the standard DRC-1.2.1 model, 2) DRC-1.2.1, but with only 1 null slot contributing excitation to the activation of each orthographic lexicon word node, 3) DRC-1.2.1 with only 1 null slot contributing excitation to the activation of each orthographic lexicon word node, and with normalisation of excitation form the letter level to the orthographic lexicon.

One simple justification for the use of normalisation is that it improves the capacity of the model to match the data. While a somewhat ad-hoc model change, the fact that the model can better match empirical results with this change suggests that the previous idea that each letter in a stimulus contributes equal activation regardless of stimulus strength is incorrect, and falsified by empirical data on length effects. Normalisation is not simply tweaking a meaningless parameter to achieve the desired result—tweaking a meaningless parameter does not clearly imply any change of hypothesis. The inclusion of normalisation is a clear statement regarding how information about letter identities must be communicated through the lexical route in order to appropriately model real readers, so in this sense it can be justified. The change embodies a theoretical commitment, which can be empirically tested.

Normalisation can also be understood as a process of cognitive-resource allocation. Shallice and Cooper (2011) discuss the idea of finite cognitive resources being allocated to

achieve particular cognitive tasks (see section 4.5, p. 119, and section 5.3, pp. 159-162). In discussing the interpretation of functional magnetic resonance imaging (specifically, the blood oxygenation level dependent (BOLD) signal), Shallice and Cooper explore the idea of using this BOLD signal as a measure of cognitive resource employed by a particular subject in carrying out a task. They state: "We therefore make the assumption here that the level of activation for a particular [brain] region corresponds monotonically to the amount of [cognitive] resource being employed." (Shallice & Cooper, 2011, p. 162). To borrow from this idea, if the level of activation contributed from the L-DRC letter layer to the L-DRC orthographic lexicon is cognitively constrained by available resource, then it makes sense for the excitation from any particular letter slot to be scaled according to the number of letter slots that are contributing excitation—this normalisation models a constant cognitive resource being deployed and spread across the letter slots. The resource available per letter will be lower if more letters are being computed at once. It seems reasonable to think that there is a finite cognitive resource available to simultaneously contribute information regarding multiple letters to the task of recognising a whole word by sight. This doesn't mean longer words necessarily take longer to recognise and read, just that each letter within that word individually contributes information to completing the task at a slower rate, the more letters there are.

Finally, as is the case with the ART-1 model described in Carpenter and Grossberg (1987), normalisation of letter-to-orthographic lexicon excitation can be considered to represent a beginning reader adjusting their perception of what features are critical in identifying a stimulus, based on the complexity of the stimulus. Normalisation will mean that the letter N in short word like AN will contribute relatively more excitation to the orthographic lexicon than the letter N in a longer word like BRANCH, reflecting the relative importance of the letter N in identifying the word in each case. If the letter N were missing from both words, it would be much easier to guess that BRA-CH is probably the word

BRANCH, than it would be to guess that A- is the word AN. AH, AM, AS, AT and AW are also possible. N is much more important to recognising the short word AN than the longer word BRANCH, which is reflected in the greater excitation coming from this letter in the shorter word, after normalisation.

Simulations: testing the new model

Overview of simulations

In building a new model of reading aloud and orthographic learning, testing on a variety of measures is required. For example, in the spirit of nested modelling, the new model should be tested to ensure it can perform all of the functions that the old model can perform. The scope of this article is, however, not to test the L-DRC model against existing empirical benchmarks. This will be done in subsequent publications. Instead, the focus of the simulations presented here is on testing the basic operation of L-DRC's learning mechanism, to see if and how it works in practice. Should L-DRC prove satisfactory in modelling orthographic learning it will be a necessary next step to test L-DRC against existing benchmarks.

Three simulations are presented as part of this research. Simulation 1 tested the impact of the changes made to letter-to-orthographic lexicon connectivity, and also to investigate the general operation of contextually-supported self-teaching. Simulation 2 investigated the impact of changing the simulation completion criteria to model faster or slower reading speeds, and how this impacted learning. Simulation 3 explored the conditions under which partial decoding works with contextual support to enable the orthographic learning of irregular words. For all simulations, the default set of parameter values listed in Appendix A was used, unless values chosen for a specific simulation are specified.

Simulation 1 – contextual input and normalisation of letter excitation

Aims

- To test whether the model can undergo successful and accurate orthographic learning
- To test how changing the level of contextual support affects the learning performance of the model
- To examine the impact that changing letter-to-orthographic-lexicon connectivity and excitation have on model performance.

Model variations

Fourteen model variations were used. These were created with two structural variations, and a variable tested at seven levels for each structure.

Structure 1: used DRC-1.2.1's approach to excitation from the letter layer to the orthographic lexicon layer. That is, no normalisation of letter-to-orthographic-lexicon excitation, and all eight slots (including null character slots) were connected to each word, regardless of word length.

Structure 2: Each word node in the orthographic lexicon was excited by only its constituent letters and a single null character slot (or none for eight-letter words). Total excitation from the letter layer to the orthographic lexicon layer was normalised by multiplying each excitatory contribution from a letter node to the orthographic lexicon by the normalisation multiplier.

Variable: the contextual input to the semantic layer (referred to from here on as "contextual input") was varied across seven levels: 0.00, 0.02, 0.05, 0.10, 0.25, 0.50 and 1.00.

Reference model variations: in addition to the fourteen variations described above, two reference models were prepared. Each of these variations included a manually-coded orthographic lexicon that contains identical orthographic entries and orthographic frequency

values to the standard DRC-1.2.1 model. One reference variation used Structure 1, and the other used Structure 2. Since they were manually coded, these model variations skip the training step described below.

Training corpus

A training corpus was created comprising 30,220 separate orthographic word tokens, derived from 8,017 word types, with each different pronunciation of a homograph also counted as a separate word type. The corpus was created by presenting each of the word types a number of times equal to its CELEX orthographic frequency divided by 500, rounded to the nearest whole number, then +1 added to each value. The scaling down is necessary to keep the training corpus to a manageable size for quick simulation, while maintaining reasonably accurate frequency information. +1 is added to each value to ensure that every word type is included at least once in the training corpus. In practice, most of the words in the corpus are presented only once: this is all of the words with CELEX orthographic frequency values less than 250. Only 1,620 of the 8,017 words are presented more than once. The word presented the most times was THE, which was presented 2,067 times. Adding +1 will exaggerate the relative importance of low frequency words, but we felt that this approach best dealt with the trade-off between maintaining the accuracy of relative frequency differences between words, whilst ensuring the training corpus was not so large that simulation times became unreasonable. The order of all word tokens was randomised in the training corpus.

Procedure

Each of the fourteen model variations was first presented with all 30,220 stimuli comprising the training corpus, with learning proceeding in the manner previously described. Following this, the orthographic lexicon of each model was analysed by examining whether words from the training set were now represented by a node in the orthographic lexicon. This

was done by direct inspection of model structure by including code to print the orthographic lexicon contents to a text file.

After this training phase, learning was deactivated, contextual input was set to zero, and the model was tested on each of the 8,017 word types from the training set, to assess pronunciation accuracy. Learning was deactivated for this phase to avoid additional learning and order effects potentially impacting the reading aloud of each word following initial training. However, there is nothing requiring training and performance to be separate phases. Indeed, the model is working to generate phonology throughout the training phase, and is clearly intended to learn and read aloud simultaneously. Context was set to zero during this testing phase, so that we could properly assess whether the model had learned orthographic representations in a way that enabled the kind of context-free reading aloud that a skilled reader can do.

Results and discussion of simulation 1

The results of Simulation 1 are presented in Table 1, with results for the Structure 1 variation in the top half of the table, and the Structure 2 results in the bottom half. The first row for each structure provides information for the reference model variation. These indicate how the model will look if it has learned perfectly, and has an orthographic lexicon the same as DRC-1.2.1's.

Table 1 includes data on the results of both the learning phase, and the testing phase. During the learning phase, if an unfamiliar stimulus was recognised as a word and orthographic learning takes place, then a node would have been created in the orthographic lexicon for this word. The results of this learning are presented in the columns labelled as "node created". During the testing phase, our intention is to see whether the model is able to correctly generate the phonology of words. Words that are read aloud correctly are counted in

columns labelled "correct", and errors counted in columns labelled "incorrect". Results for regular words and irregular words are presented in separate sections of Table 1.

Some quick observations about the data are helpful for orientation. Firstly, it is possible that a word might be read aloud correctly even if no node was created for that word. This is most clearly going to happen for regular words, since these words can be pronounced correctly through knowledge of GPCs, even if orthographic lexical knowledge of the word is not developed. It can be seen clearly that for any of the model variations, regular words for which no node was created were still always read aloud correctly. In contrast, irregular words for which no node is created were never read aloud correctly.

It is also possible that despite a node being created for a word, the model might still generate the incorrect phonology for that word. This may happen if a model learns a node for a particular word, but associates that node with an incorrect spoken word (node in the phonological lexicon). Models can also read aloud words incorrectly even if learning has been completely accurate—even the reference drc-1.2.1 variations make errors on some irregular words. This could happen for example if the sublexical route dominates when reading an irregular word, causing a regularisation error.

The final column of Table 1 provides a count of the number of *heterophonic homograph* nodes that exist in a model after learning. When we use the term heterophonic homograph with regards to L-DRC, we mean a written word represented by a single orthographic node that is associated with more than one distinct spoken word representation. For example, the written word BOW can be pronounced as /b5/[1], as in "bow-and-arrow", or /b6/ as in the front section of a ship, and the single orthographic node will excite two phonological lexicon nodes. Heterophonic homographs are another cause of the reference DRC models experiencing errors for some words despite a correct node being present.

[1] The phoneme symbols used in this article are the same as are used in DRC-1.2.1 and L-DRC, and are displayed in Appendix B.

Table 1 – Results for Simulation 1

minReadingPhonology =0.4

Contextual input	Regular words				Irregular words				Heterophonic homograph nodes
	Node created		No node		Node created		No node		
	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	
Structure 1: All nulls slots excite words, no normalisation									
N/A[DRC-121]	**6,658**	**-**	**-**	**-**	**1,286**	**73**	**-**	**-**	**51**
0.00	95	1	6,562	-	-	-	-	1,359	-
0.02	6,568	15	75	-	826	33	-	500	1
0.05	6,426	19	213	-	1,243	54	-	62	1
0.10	6,441	3	214	-	1,228	70	-	61	36
0.25	6,443	1	214	-	1,226	72	-	61	47
0.50	6,443	1	214	-	1,226	72	-	61	47
1.00	6,443	1	214	-	1,226	72	-	61	47
Structure 2: One null slot excites each word, normalised letter-to-orthographic lexicon excitation									
N/A[DRC-121]	**6,658**	**-**	**-**	**-**	**1,286**	**73**	**-**	**-**	**51**
0.00	95	1	6,562	-	-	-	-	1,359	-
0.02	6,630	17	11	-	851	31	-	477	-
0.05	6,508	20	130	-	1,253	53	-	53	-
0.10	6,474	4	180	-	1,237	69	-	53	36
0.25	6,472	1	185	-	1,234	72	-	53	48
0.50	6,472	1	185	-	1,234	72	-	53	48
1.00	6,471	2	185	-	1,235	71	-	53	47

[DRC-121] This row gives reference information—this variation is manually-coded to have a full orthographic lexicon the same as DRC-1.2.1. The full DRC-1.2.1 lexicon was tested separately for both Structure 1 and Structure 2.

If one pronunciation of a heterophonic homograph is intended, and the other pronunciation is produced, this has been counted as an error.

Structure 1 overall performance

The results of simulation 1 provide a good demonstration that L-DRC is able to undergo effective and accurate orthographic learning, provided there is sufficient contextual support to aid learning. For Structure 1 variations where contextual input was at least 0.05, L-DRC failed to create a node for only 213 or 214 out of 6,658 regular words (3.2%), and 61 or 62 out of 1,359 irregular words (4.6%). This indicates a degree of robustness of learning to variations in the level of contextual support. Subsequent to learning, the model was able to generate the correct phonology for almost all words in the absence of contextual support, especially for variations trained with contextual input values of at least 0.05. For example, considering the contextual input equals 0.25 Structure 1 variation, the model correctly generated phonology for 6,657 out of 6,658 regular words (99.98%), and 1,226 out of 1,359 irregular words (90.2%).

Structure 2 overall performance – examining the impact of changes to letter-to-orthographic-lexicon excitation

To recap, structure 2 is the model version with modified letter-to-orthographic lexicon excitatory connections, a different structure to DRC-1.2.1's structure. A first clear indication that Structure 2 is robust is that, when looking at just the reference models, it produces exactly the same responses to words as Structure 1, meaning the change in Structure did not affect DRC's naming accuracy. When considering Structure 2 as it impacts learning, Structure 2 variations also perform very well across a range of contextual input values, and improve on Structure 1. For Structure 2 variations where contextual input was at least 0.05, L-DRC failed to create a node for only 130-185 out of 6,658 regular words (2.0%-2.8%), and failed to create a node for only 53 out of 1,359 irregular words (3.9%). Following learning, the Structure 2

variations also reads aloud most words correctly for contextual input values of 0.05 or greater. For example, considering the contextual input equals 0.25 Structure 2 variation, the model correctly reads aloud 6,657 out of 6,658 regular words (99.98%), and 1,234 out of 1,359 irregular words (90.8%). So despite making substantial changes to the way letter-to-orthographic-lexicon excitation is treated for Structure 2, L-DRC is still able to learn appropriately and correctly generate phonology, even demonstrating a small improvement over Structure 1.

This improvement would most likely be even larger if the very high default value of letter-to-orthographic inhibition were reduced, thereby making it easier for orthographic neighbours to the stimulus to be activated by that stimulus, instead of being completely inhibited by a single letter difference. This inhibition value is set very high in DRC-1.2.1 to prevent involvement of the lexical route in non-word reading. However, this was done on the assumption that the only correct response to a non-word is a regular response. The results presented in Pritchard, Coltheart, Palethorpe, and Castles (2012) (see Chapter 3) suggest that this assumption is invalid, because human readers often produce reading-aloud responses to nonwords that are not the responses specified by standard GPCs.

Regular word learning in the absence of contextual support

We have argued that contextual input is required to support the learning of irregular words. However, regular words should be able to be learned even in the absence of contextual support from accompanying text. For a beginning reader with strong knowledge of GPCs, the application of these GPCs should enable a correct pronunciation to be deduced for an unfamiliar-but-regular written word stimulus. This is the kind of reader we are modelling by having a fully intact sub-lexical route included in each model variation. Despite this, the two model variations for which contextual input equals zero both failed to learn nodes for almost all of the regular words. For both Structure 1 and Structure 2, when contextual input equals

zero, only 96 out of 6,658 (1.2%) of the regular words were learned and had a node created. Ninety-five were subsequently pronounced correctly, with one pronounced incorrectly despite having a node created. As expected, these model variations also failed to learn any irregular word nodes.

We investigated activation levels during potential learning events to explore why learning was not occurring in the model variations where contextual input equalled zero. This revealed a clear cause for the failure to learn regular words. In order for a learning event to occur, a phonological lexicon node needs to be activated above the threshold value for learning. For simulation 1 (and all further simulations reported here), this parameter was set at 0.4. When the stimulus is a regular word, and in the absence of any contextual input, the sub-lexical route should still activate the correct node by first activating the matching phonemes, which then interact with the phonological lexicon. This is certainly occurring, as intended. However, a simulation concludes when phonemes all reach the *MinReadingPhonology* threshold, which is also set to 0.4 as a default value, intended to simulate fast reading responses as typically required in experiments on reading aloud with skilled readers. This is identical to the way DRC-1.2.1 concludes a reading-aloud simulation, and is described in Coltheart et al. (2001). For the contextual input equals zero variations, the phoneme nodes all reach this threshold and the simulation concludes *before* the correct phonological lexicon node has had time to be activated above the learning threshold. Since no node has reached the learning threshold, no learning takes place. In effect, the model seems to be simulating quick reading with little comprehension, where the reader is making use of grapheme–phoneme correspondence knowledge to read aloud the word prior to really thinking about whether the written symbols correspond to a known spoken word.

If the *MinReadingPhonology* parameter was set to a higher value, thereby simulating slower, more careful reading, the phonemes would attain higher activation values, and would

have a greater influence over the phonological lexicon. They would also have more cycles in which to apply that influence. Modifying *MinReadingPhonology* to improve regular word learning in zero context conditions will be explored in simulation 2.

Heterophonic homographs

It can be seen in the rightmost column of Table 1 that contextual support is required for heterophonic homographs to be accurately learned by L-DRC. For the Structure 2 model variations, contextual input of at least 0.25 seems required to maximise the learning of heterophonic homographs. When contextual support is low or absent, heterophonic homographs are not learned. This is because once an orthographic node is created for one pronunciation of the homograph, this orthographic node will be activated whenever that stimulus is presented, and it will in turn activate the one pronunciation for that homograph that it has already learned. In order to learn additional pronunciations for that node, sufficient contextual support is required to ensure that the alternate pronunciation is excited more vigorously than the initial pronunciation will be by the active orthographic node that corresponds to and activates both pronunciations.

That heterophonic homographs are not being learned for low values of contextual input explains the errors on some regular words despite the node having been learned, when contextual input is low. For example, the Structure 2 model variation where contextual input = 0.02 simulated 17 errors on regular words for which a node had nevertheless been accurately learned. All of these errors are on heterophonic homographs, where the model had first learned an irregular pronunciation for the homograph, but had not had sufficient contextual support to subsequently learn the regular pronunciation.

Note that in the rightmost column of Table 1, the reference model variations contain knowledge of 51 heterophonic homographs, while the best that any L-DRC model variation

does is to learn 48. The remaining few that are not learned, (AYE, BASS, and BERTHS are the three not learned with Structure 2, contextual input=0.25), in addition to being heterophonic homographs, are also one of the written representations of a heterographic homophone. Heterographic homophones are challenging for L-DRC to learn, as described in the next section.

Heterographic homophones

Putting the problematic contextual-input-equals-zero variations to one side, we observe that a small decrease in regular-word-node learning occurs as contextual support increases. This occurs for both the Structure 1 and Structure 2 variations. For example, just considering Structure 2, 6,647 regular word nodes are learned for the contextual input = 0.02 variation, decreasing to 6,528 for the contextual input = 0.05 variation, and dropping to 6,474 and lower for higher values of contextual input. This is a very counter-intuitive result—we would expect that learning would improve with increased contextual support.

To explore these results, we investigated two model variations in detail. Both were Structure 2 variations, one with contextual input = 0.02, and the other with contextual input = 0.25. Firstly, all of the nodes and associations learned by the contextual input = 0.02 model are correct, so its higher rate of regular word node creation is not explained by spurious learning. It was observed that all of the regular words learned by the contextual input = 0.25 variation were also learned by the 0.02 variation. The reverse was not the case though—there were 174 orthographic words that were learned by the 0.02 variation that were not learned by the 0.25 variation. On inspection, all of these words were found to be one of the orthographic representations of a *heterographic homophone.* These are spoken words that correspond to more than one written word. For example, the word /skVl/ corresponds to both SKULL and SCULL, so it is a heterographic homophone. An orthographic node for SKULL was learned

by the contextual input = 0.02 model variation, but not by the contextual input = 0.25 variation.

By examining node activation levels over the course of a reading simulation, it is possible to determine the reason why such items are not learned when contextual input is higher. Consider the situation where one of the written forms of a heterographic homophone has already been learned (e.g., SCULL has been learned), and now the model is being exposed for the first time to the second written form (SKULL). When contextual input is high the phonological lexicon node corresponding to both written forms (/skVl/) is activated strongly. This phonological lexicon node will in turn strongly excite any of the written forms to which it corresponds. Since SKULL has not yet been learned, SCULL is the only orthographic node excited. Now, since the stimulus is SKULL, the orthographic node for SCULL will be receiving some inhibition from the 2nd letter slot where the incongruent letter K has been activated by the stimulus. For low values of contextual input, this inhibition seems to be enough to prevent SCULL from being activated above the learning threshold, despite it receiving excitation from the other letter slots in common between the words, and also from the phonological lexical node /skVl/. However, when contextual input is high, the phonological lexicon node contributes enough excitation to ensure that SCULL is activated above the learning threshold. This means that when SKULL is the stimulus, the model is fooled into thinking it has identified SCULL, and increments its learning of this orthographic node, instead of learning a completely new orthographic node for SKULL. In effect, the contextual input provides such strong recognition of a spoken word, that it overrides letter knowledge, causing the written stimulus to be identified incorrectly.

In L-DRC, the semantic layer is presently only connected to the phonological lexicon, not to the orthographic lexicon. If it was also connected to the orthographic lexicon, then contextual input could be applied to inhibit the incorrect orthographic representation SKULL,

thereby allowing SCULL to be learned. It seems appropriate that, for example, in a sentence such as "He continued rowing, scull after scull moving the boat forward." that context would orthographically inhibit activation of SKULL, presuming a beginning reader who read this sentence was already familiar with the word SKULL.

Explaining further errors – more homophones

Turning now to irregular words, we first note that DRC-1.2.1 is unable to read aloud every irregular word without error, despite having a perfectly organised orthographic lexicon. The reference model variations make 73 errors on irregular words. However, 51 of these "errors" are heterophonic homographs, where the model produced the unintended pronunciation (e.g., when the stimulus is BOW and the intended pronunciation is /b5/, the model produces /b6/). These are not strictly errors, given that the reading aloud testing is being done in the absence of context. The remaining 22 errors all involve an irregularity in the first slot. The reference variations regularise this irregularity to produce an error, but generate correct phonology for the balance of the word. For example, GYMS is pronounced as /gImz/ instead of /_Imz/[1]. Due to its serial operation in DRC, the sub-lexical route accepts input from the first letter slot the earliest, (even though the letter slots themselves receive activation in parallel from the visual feature layer). Consequently, this first GPC has the most processing time to exert its influence. So in a handful of cases, the sub-lexical route is able to overpower the lexical route and determine the pronunciation of the first phoneme, even though the lexical route is able to dictate the pronunciation of the rest of the word.

Since the reference model makes these errors despite a well-formed orthographic lexicon, we expect various L-DRC variations to make similar errors, and indeed, all model variations make errors on some irregular words even when the node has been correctly

[1] In fact, 12 of the remaining 22 errors involve a G in the first letter position, followed by an I or a Y, suggesting that perhaps DRC-1.2.1's GPCs relating to G in the first position may need correcting.

learned. Obviously, all model variations, in contrast to the reference model variations, also make errors on irregular words for which no node has been learned. We examined a single model variation—Structure 2 with contextual input = 0.25—to try and determine the nature of these errors.

This model variation made 126 errors, with 125 of these being on irregular words. Of these, 51 were on heterophonic homographs where the model produced the unintended but still correct pronunciation for the written stimulus, so these are not actually errors. The model also produced 53 errors on words where no orthographic node was learned. All of these errors were regularisation errors, which is understandable—without an orthographic lexical node for the stimulus, the lexical route has greatly reduced involvement in the reading of the stimulus, and the sublexical route produces a regular pronunciation without any contention from the lexical route. The remaining 22 errors involved a first position irregularity being regularised, but the rest of the word being pronounced correctly. So the errors made by this model variation on these items are very similar to the errors made by the reference model variation, with the exception of the 53 regularisation errors due to no orthographic node being learned.

In addition to these 53 unlearned irregular word nodes, this model variation did not learn 185 regular words, for a total of 238 unlearned nodes. Why were these nodes not learned? Examining them, we find that each of the unlearned orthographic representations is one of multiple orthographic representations associated with a homophone. As previously described, regular word heterographic homophones are not learned appropriately for higher values of contextual input. As well as explaining the reason why more regular word nodes are learned in low-context conditions, L-DRCs difficulty with heterographic homophones also explains the difficulty that model variations with high contextual input have on irregular words for which no node is learned.

Summary of simulation 1

Simulation 1 tested two different structures of L-DRC each under a range of contextual support conditions differing in strength. This simulation demonstrated the general accuracy of L-DRC in simulating self-teaching to build orthographic knowledge. It also revealed that some items, particularly heterographic homophones, are problematic for L-DRC when contextual support is high, while other items—especially heterophonic homographs—are difficult for L-DRC to appropriately learn when contextual support is low. As expected, L-DRC has difficulty learning irregular words when contextual support is low or zero, however, it was surprising to find that L-DRC also had difficulty learning regular words when contextual support was zero. This was due to the model reading aloud the stimulus before a phonological node was activated above the threshold required to trigger learning. In effect, it seems like the model was simulating a beginning reader with good GPCs who was reading so quickly that they were sounding out the stimuli before it even occurred to them that it might correspond to a spoken word that they already knew. Simulation 2 will investigate improving regular word learning under low contextual support conditions.

Simulation 1 also revealed that structure 2, which features modified letter-to-orthographic lexicon connectivity, performs better than structure 1 that retains DRC-1.2.1's original connectivity.

Simulation 2 – slow reading during learning versus fast reading

Aim

To test whether modifying the MinReadingPhonology parameter to simulate slower reading improves learning, particularly the learning of regular words when there is no support from context for learning.

Model variations

The same model variations were used for Simulation 2 as were used in Simulation 1, except the MinReadingPhonology parameter was changed to 0.9, from the value of 0.4 which was used in Simulation 1. This change was only made for the learning phase. During the testing phase, MinReadingPhonology was again set at 0.4. This was done because we are examining the impact that changing MinReadingPhonology has on learning, not on reading aloud.

Training corpus

The same training corpus as was used for Simulation 1.

Procedure

The same procedure as was used for Simulation 1.

Results and discussion for Simulation 2

The results of Simulation 2 are presented in Table 2, with results for the Structure 1 variation again in the top half of the table, and the Structure 2 results in the bottom half. As for simulation 1, the first row for each structure provides information for the reference model variation.

The impact of changing MinReadingPhonology

Changing *MinReadingPhonology* by increasing it from 0.4 to 0.9 simulates a shift to a slower, more meticulous mode of reading behaviour. By making the change, the model variations where contextual input equals zero were able to effectively learn regular words, though at least one learned node was still pronounced incorrectly in subsequent testing. Both model variations learned nodes for the majority of regular words (6,445 out of 6,658, or 96.8% for Structure 1, and 6,507 out of 6,658, or 97.7% for Structure 2).

Table 2 – Results for Simulation 2

minReadingPhonology =0.9

Contextual input	Regular words				Irregular words				Heterophonic homograph nodes
	Node created		No node		Node created		No node		
	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	
Structure 1: All nulls slots excite words, no normalisation									
N/A[DRC-121]	**6,658**	**-**	**-**	**-**	**1,286**	**73**	**-**	**-**	**51**
0.00	6,444	1	213	-	8	251	-	1,100	2
0.02	6,420	18	220	-	966	76	-	317	6
0.05	6,432	8	218	-	1,228	66	-	65	16
0.10	6,440	1	217	-	1,221	73	-	65	30
0.25	6,439	1	218	-	1,223	72	-	64	48
0.50	6,439	1	218	-	1,223	72	-	64	48
1.00	6,439	1	218	-	1,221	72	-	66	48
Structure 2: One null slot excites each word, normalised letter-to-orthographic lexicon excitation									
N/A[DRC-121]	**6,658**	**-**	**-**	**-**	**1,286**	**73**	**-**	**-**	**51**
0.00	6,506	1	151	-	8	262	-	1,089	-
0.02	6,455	19	184	-	985	76	-	298	5
0.05	6.465	9	184	-	1,241	65	-	53	15
0.10	6,473	1	184	-	1,233	73	-	53	30
0.25	6,471	2	185	-	1,235	71	-	53	47
0.50	6,471	2	185	-	1,235	71	-	53	47
1.00	6,471	2	185	-	1,235	71	-	53	46

[DRC-121] This row gives reference information—this variation is manually-coded to have a full orthographic lexicon the same as DRC-1.2.1. The full DRC-1.2.1 lexicon was tested separately for both Structure 1 and Structure 2.

These variations in fact learn a greater number of regular word nodes than the higher context variations, which is the same result as was seen in simulation 1, and is due to the challenge to learning all of the written forms of a homophone when contextual input is high.

The change in *MinReadingPhonology* that enabled this improved result is justifiable and not ad-hoc, because it is reasonable that a beginning reader would proceed in a slower, more considered fashion when encountering novel words. A higher value for *MinReadingPhonology* reflects a more meticulous reader who allows more time for their level of confidence to grow that the phonemes they are about to utter are satisfactory. A novice reader who was able to construct a pronunciation using sub-lexical information such as that provided by GPCs would arguably also take the time for this pronunciation to jog their memory of known spoken words, rather than just uttering the pronunciation according to GPCs and then moving on without any recognition.

With the exception of the zero contextual support variations, increasing *minReadinPhonology* did not have any impact on the amount of learning for the Structure 2 variations, and had only an extremely small but *negative* impact on learning for the Structure 1 variations. Although small, this negative impact is quite revealing. Examination of errors revealed that all of the Structure 1 variations in simulation 2 learned to associate multiple pronunciations to the same orthographic word node, where most of these associations should not have been made. These incorrect associations were often created to almost absurd degrees. For example, the Structure 1 variation with contextual input equal to 0.25 learned erroneous associations between the orthographic node for THE and all of the following phonological nodes: “the”, “then”, “she”, “they”, “tie”, “toe”, “thew” and “thy”. This is the “snowball” error in action, and again demonstrates the superiority of structure 2 over structure 1.

Potentiophones

Potentiophones are irregular words which, when pronounced regularly, correspond to another word (Friedmann & Lukov, 2008). We became aware of the way potentiophones are treated by L-DRC in observing that, for the model variations where contextual input equals zero, the models still seem to learn some irregular word nodes. The Structure 1 variation learned 259 irregular word nodes, and the Structure 2 variation learned 270 irregular word nodes. It struck us as odd that this model variation would learn any irregular word nodes without the support of context.

We investigated node creation in Structure 2 to determine why irregular words were being learned. For the Structure 2 variation with contextual input equal to zero, 256 out of the 270 (94.8%) irregular words for which an orthographic node was created were potentiophones. For example, the model learns an orthographic node for the irregular word LOVES, but incorrectly associates this with the regular pronunciation /l5vs/. This pronunciation corresponds to the spoken form of the word LOAVES.

When the stimulus is an irregular word, the sub-lexical route still generates a regular response. If the stimulus is a potentiophone, this regular response will activate a known spoken word in the phonological lexicon, corresponding to the regular pronunciation of the potentiophone. With no contextual support to provide information that this activation is wrong, the spoken word node will be activated above threshold, triggering a learning event. A node will be created for the written stimulus, but it will be incorrectly associated with the regular pronunciation that has been activated, rather than the correct irregular pronunciation. So even though the orthographic node is created and it seems the word has been successfully learned, the model will produce a regularisation error whenever it attempts to read aloud this word due to an erroneous association having been created between the learned orthographic node and the wrong spoken word node.

For 14 out of the 270 irregular word nodes created that were not potentiophones, there was still a phonological lexicon node sufficiently close to the regular pronunciation that it was activated above the learning threshold. It is not surprising that this was able to happen in simulation 2, given that the longer simulation times associated with the higher value of *MinReadingPhonology* mean that there is more time for activation levels to grow above threshold. For eight of these 14 words, the phonological lexicon node that was activated above threshold happened to be the correct node. The word FRIENDS was one such item. The regular pronunciation is /fr2ndz/, which was sufficiently close to the correct pronunciation /frEndz/ for the correct phonological lexicon node to be activated above threshold and trigger learning of the correct orthographic node. The word STRANGE is an example of a non-potentiophone irregular word for which a node was learned, but an incorrect association learned, resulting in an incorrect pronunciation. This word was associated with the phonological lexicon node /str{nd/. The regular pronunciation of STRANGE is /str{n_/. This is a phonological neighbour of both /str{nd/ and the correct pronunciation /str1n_/. However, the sublexical route operates serially from left to right, so after four phonemes have been activated, the phonological lexicon node for /str{nd/ is as yet receiving no inhibition from activated phonemes, while the correct word /str1n_/ is. The existence of a whammy (Rastle & Coltheart, 1998), in the last few letters of the word STRANGE also delays the phonological lexicon node /str1n_/ receiving its maximum excitation from the regular pronunciation. This is enough for the incorrect word to be able to be activated above threshold by the time the reading simulation is completed.

While it might be true that beginning readers trying to expand their orthographic vocabulary by self-teaching would have difficulties with irregular words, and more specifically, potentiophones, it is also true that they do eventually learn to read potentiophones correctly. It might be that contextual support is required to avoid

potentiophones being incorrectly learned. Simulation 3 explored altering L-DRC's default parameters to try and facilitate partial decoding and improve the learning of irregular words in low context conditions.

Simulation 3 – testing the sub-lexical contribution to irregular word learning

Partial decoding

Share (1995) argues that irregular word learning is still dependent on phonological recoding, which, in L-DRC, has been modelled as the influence of GPC knowledge through the sublexical route on phoneme activation. Some research shows that nonword reading accuracy (which is used as one measure of phonological recoding ability) correlates to a degree with irregular word reading accuracy (e.g., Baron, 1979), though there is a stronger correlation between regular word reading accuracy and nonword reading accuracy. Share cites further research that found that younger readers with poor decoding skill had difficulty learning novel irregular words relative to good readers (e.g., Gough and Walsh (1991), as cited in Share (1995)).

The self-teaching hypothesis proposes that, in natural text, irregular words still have sufficient letter–sound regularity for the correct spoken word to be selected from a set of possible pronunciations. That is, even if context is not sufficient for a reader to unambiguously decide what spoken word corresponds to the novel written stimulus, it may be sufficient for the reader to narrow down the set of spoken words, so that a *partial decoding* of the written stimulus could enable the correct pronunciation to be selected, and learning to occur (Share, 1995). Wang, Castles, Nickels, and Nation (2011) provide empirical evidence that context is more important for children learning novel irregular words than it is for learning novel regular words. Wang, Castles, and Nickels (2012) provide empirical evidence

that, in accordance with self-teaching, it is easier for children to learn novel regular stimuli than irregular stimuli.

From the results of simulation 1 and 2, it is unclear whether or not L-DRC is simulating an interaction between partial decoding and contextual support in learning irregular words. While irregular words were not correctly learned under zero contextual input conditions, and were learned with high values of contextual input, it is not clear that partial decoding contributes anything. Contextual input could be doing all of the work. L-DRC, by default, has a phoneme-to-phonological lexicon inhibition parameter value that is much stronger than its phoneme-to-phonological lexicon excitation parameter value. This would make it difficult for a set of active phonemes to excite a node in the phonological lexicon unless these activated phonemes corresponded exactly to a particular phonological lexicon node. Even one phoneme different would strongly inhibit a word node.

However, it may not be necessary to keep this inhibition at its high default level. Indeed, this inhibition was set lower in the CDP+ model (Perry et al., 2007) that in it is in DRC-1.2.1, and these two models have a similar lexical route structure. It may be that the involvement of partial decoding can be increased by decreasing phoneme-to-phonological lexicon inhibition, and exploring this is the basis of simulation 3.

Aim

To investigate whether L-DRC can better simulate the self-teaching hypothesis and partial decoding, and whether the positive influence of partial decoding on learning in L-DRC can be increased. This was investigated by modifying the phoneme-to-phonological lexicon inhibition, thereby allowing the sub-lexical route to have greater involvement in the learning of irregular words.

Model variations

Twenty-four model variations (all of the structure 2 variety) were used, produced with different levels of two variables:

Variable 1: Contextual input, with values 0.00, 0.001, 0.05, 0.1, 0.15, and 0.25. Note that the 0.00 and 0.001 variations all also used a *MinReadingPhonology* value of 0.9 as per simulation 2, to ensure learning could take place. Variations using higher contextual input values used a *MinReadingPhonology* value of 0.4.

Variable 2: Phoneme-to-phonological-lexicon inhibition, with values 0.00, 0.02, 0.04, and 0.16. The last value is the default value, as used in DRC-1.2.1.

Training corpus

The training corpus is the same as that used in Simulation 1 and 2.

Procedure

The procedure is the same as that used in Simulation 1 and 2. Note that the phoneme-to-phonological-lexicon inhibition values for a variation were maintained over both the learning and testing phases, while the *MinReadingPhonology* value was always set to 0.4 during the testing phase, even if it had been set to 0.9 during learning for the low contextual input variations.

Results and discussion of simulation 3

The aim of this simulation was to see if L-DRC could better simulate the self-teaching hypothesis by ensuring that partial decoding contributed to the learning of irregular words. We investigated whether reducing phoneme-to-phonological-lexicon inhibition would allow the phonemes activated by the sublexical route to activate multiple phonological lexicon nodes, and thereby allow partial decoding to contribute more to the activation of the target spoken word. This is instead of such words being activated and learned exclusively through

the information provided by contextual input. To remain true to the self-teaching hypothesis, contextual input should only play a supporting role in the learning of novel written words, and not be the sole source of information. Results of simulation 3 are presented in Table 3.

The results are mixed. While decreasing phoneme-to-phonological-lexicon inhibition seems to result in improvements to irregular word learning in some cases, it can also result in an increase in erroneous irregular word learning. In some cases, a seeming improvement is a result of the impact of different inhibition values during the testing phase, rather than an impact on learning during the training phase.

Positive impact of partial decoding on irregular word learning under low context conditions

When contextual input equals 0.00 and the default value for phoneme-to-phonological-lexicon inhibition is used, only 270 irregular word nodes are created, and only eight of these resulted in a correct response during testing. That is, there is barely any correct learning of irregular words, and some degree of incorrect irregular word learning, which, as previously discussed in simulation 2, is due to the presence of potentiophones. When inhibition was decreased, there was an increase in correct irregular word node learning. For inhibition = 0.04, 196 irregular words are learned and read aloud correctly. This demonstrates that decreasing inhibition in low context conditions can increase irregular word learning.

The same result applies for the contextual input = 0.001 variations. Only 48 irregular words were learned and correctly pronounced for the default value of inhibition, while more irregular words were learned and correctly pronounced for lower values of inhibition. For inhibition = 0.02, 390 irregular word nodes are learned and correctly read aloud.

Table 3

Phoneme-to-phonological-lexicon inhibition	Regular words				Irregular words				Heterophonic homograph nodes
	Node created		No node		Node created		No node		
	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	
Contextual input = 0.00[a]									
0.00	4,178	1,061	1,419	-	161	834	-	364	85
0.02	5,509	288	861	-	189	845	-	325	154
0.04	6,162	24	472	-	196	790	-	373	130
0.16[b]	6,506	1	151	-	8	262	-	1,089	-
Contextual input = 0.001[a]									
0.00	5,271	419	968	-	378	676	-	305	145
0.02	5,913	74	671	-	390	685	-	284	193
0.04	6,286	2	370	-	387	653	-	319	97
0.16[b]	6,502	-	156	-	48	259	-	1,052	-
Contextual input = 0.05									
0.00	6,503	21	134	-	1,274	32	-	53	-
0.02	6,503	21	134	-	1,276	30	-	53	-
0.04	6,503	21	134	-	1,275	31	-	53	-
0.16[b]	6,508	20	130	-	1,253	53	-	53	-
Contextual input = 0.10									
0.00	6,472	6	180	-	1,259	47	-	53	37
0.02	6,472	6	180	-	1,261	45	-	53	36
0.04	6,474	4	180	-	1,258	48	-	53	38
0.16[b]	6,474	4	180	-	1,237	69	-	53	36
Contextual input = 0.15									
0.00	6,474	1	183	-	1,254	52	-	53	48
0.02	6,474	1	183	-	1,256	50	-	53	48
0.04	6,474	1	183	-	1,255	51	-	53	48
0.16[b]	6,474	1	183	-	1,234	72	-	53	48
Contextual input = 0.25									
0.00	6,472	1	185	-	1,254	52	-	53	48
0.02	6,472	1	185	-	1,256	50	-	53	48
0.04	6,472	1	185	-	1,255	51	-	53	48
0.16[b]	6,472	1	185	-	1,234	72	-	53	48

[a]The variations with contextual input = 0 or 0.001 all used a minReadingPhonology value of 0.9, as per simulation 2, to ensure appropriate learning occurred. All other contextual input cases used the value of 0.4.
[b]The default phoneme-to-phonological lexicon inhibition value is 0.16.

Negative impact of partial decoding on learning

While decreasing inhibition allows more irregular words to be correctly learned when contextual input is zero or very low, it also results in a marked increase in erroneous learning of irregular words. In simulation 2, we saw that potentiophones cause learning errors when contextual input is zero, because the regular pronunciation generated by the sublexical route corresponds to a different spoken word other than the correct, irregular word target. As a result, the regularly-pronounced word is incorrectly associated with the written stimulus. When phoneme-to-phonological-lexicon inhibition is decreased, then a similar kind of error occurs. The regular pronunciation generated by the sublexical route results in multiple phonological lexicon nodes being activated, especially nodes that are phonological neighbours to the regular pronunciation. In many cases, L-DRC will learn to associate one of those pronunciations with the written stimulus, but the pronunciation selected will not be the correct one. For example, in the variation where contextual input = 0.001, and inhibition = 0.02, the word BULL is learned and incorrectly associated with the spoken word /bVt/ ("but"). When this written word is presented to the model, the sub-lexical route generates the regular pronunciation /bVl/. This does not correspond to any spoken word, but under low inhibition conditions, several neighbouring actual spoken word nodes are activated in the phonological lexicon, one of which is /bVt/. This node is activated above the learning threshold, resulting in a learning error. When inhibition is high, this neighbouring node will not be activated.

The same kind of error even occurs with regular words when contextual input is low. For example, when contextual input = 0.001, and inhibition = 0.02, 74 regular words are learned, yet are read aloud incorrectly, because they have been incorrectly associated with the wrong pronunciation. For example, the word BAIT is incorrectly associated with the word /bVt/. Words with high spoken-word frequency that are neighbours to the correct, regular

pronunciation can sometimes be incorrectly associated with a stimulus over the correct pronunciation, if the correct pronunciation is low frequency. The lower the inhibition, the more of these sorts of errors occur, and the number of these sorts of errors is also highest when contextual input is lowest.

Partial decoding when contextual input is high

Erroneous irregular word node learning, while prevalent at low levels of contextual input and low values of inhibition, is greatly reduced at higher levels of contextual input. This is because the influence of context ensures that only correct phonological lexicon nodes will be activated, while incorrect nodes are inhibited. High context effectively renders the impact of partial decoding irrelevant, and the excitation of phonological lexicon nodes as driven by the sublexical route is negligible compared to the strong influence of contextual input on the excitation and inhibition of phonological lexicon nodes.

For contextual input values of 0.05 or greater, there is a small but identifiable positive impact of decreasing phoneme-to-phonological-lexicon inhibition on irregular word learning. However, this seeming improvement is illusory. Considering the contextual input equals 0.05 variations as an example, all variations no matter what level of phoneme-to-phonological-lexicon inhibition were able to learn 1,306 of the 1,359 irregular word nodes (96.1%). However, when phoneme-to-phonological-lexicon inhibition was set to its default value of 0.16, 53 of those words were still read aloud incorrectly, while for an inhibition value of 0.00, only 32 of these words were read aloud incorrectly. It seems as though there are fewer erroneous nodes being learned at lower values of inhibition. Inspecting the model variations, however, reveals that this seeming small improvement in reading-aloud accuracy for irregular words is due to the impact of the change in inhibition during reading aloud testing, not due to any change during learning. Even DRC-1.2.1 makes 73 errors on irregular words, despite having a well formed orthographic lexicon with correct associations between orthographic

nodes and phonological nodes. Decreasing the phoneme-to-phonological lexicon parameter results in an improvement in irregular word accuracy by reducing the number of these errors. However, in the learning phase, the change in inhibition did not have an impact on learning. Both the inhibition = 0.00 variation and the inhibition = 0.16 (default) variation learned the same number of irregular word nodes, and with the same associations.

Summary

As expected, reducing phoneme-to-phonological lexicon inhibition did increase the opportunity for partial decoding. Decreasing this inhibition results in a greater number of irregular words being correctly learned and read aloud, for the cases where contextual input is zero or otherwise very low. The default value for phoneme-to-phonological lexicon inhibition, which is a high value of -0.16, almost entirely prevents the sublexical route from activating irregular word nodes in the phonological lexicon, while decreasing inhibition allows a range of word nodes that are neighbours to the regular pronunciation of a stimulus to be activated. However, relaxing inhibition to allow partial decoding was not without problems. In the absence of significant contextual support, decreasing inhibition resulted in a large increase in incorrect node learning for both regular and irregular words, even if the total number of correctly learned irregular word nodes had increased. These results present a challenge for L-DRC as a model of the self-teaching hypothesis.

General Discussion

A learning DRC that provides a computational account of the self-teaching hypothesis

The dual-route cascaded (DRC) model of reading aloud and word recognition is a successful computational cognitive model, yet it has faced ongoing criticism for not explicitly

modelling a learning process. To address this criticism, we have drawn on the self-teaching hypothesis (Jorm & Share, 1983; Share, 1995) to prepare a "learning" DRC, or L-DRC. According to the self-teaching hypothesis, beginning readers increase their written word vocabulary too quickly for it to be plausible that this learning is achieved via direct instruction from a teacher. This means that beginning readers somehow acquire new written words by phonologically recoding the written word into a candidate pronunciation. If the candidate pronunciation matches a familiar spoken word, an opportunity for orthographic learning is created. Contextual support and partial decoding are assumed to work together to enable self-teaching of irregular words.

The self-teaching hypothesis is a verbal account of how beginning readers build their written word vocabulary, and, as a result, lacks the detail and robustness of a computational model account. So in constructing L-DRC we have sought to explore the self-teaching hypothesis at the finer grain of detail afforded by computational modelling, while also addressing criticisms of DRC that it does not model learning.

Balancing lexical route and sublexical route activity

The standard DRC model 1.2.1 uses a specific set of parameters to achieve the results and match to empirical data that makes it such a successful model. While these parameters can certainly be altered to explore DRC's operation, they are set to certain default values for the purpose of benchmark testing (see Perry et al., 2007 for a list of benchmarks). DRC is set up in a way that the sublexical route has sufficient activity to strongly contribute to nonword reading, and to influence the reading-aloud latency of irregular words relative to regular words, but not so strongly that it overrides the lexical route and causes regularisation errors in more than a handful of cases. All of these parameters are chosen so that the rate at which activation cascades through each route produces a good match to empirical data.

L-DRC introduces a new source of activation: contextual support. This has been modelled as an input of varying strength that excites the semantic layer, which in turn interacts with the phonological lexicon layer. Implementing context in this manner has shed light on the vulnerability of DRC's parameter choices to varying levels of excitation. When contextual support is high, then activation builds up more quickly in the lexical route than when contextual input is very low or zero. This means that the importance of the sublexical route relative to the lexical route in reading aloud varies with the level of contextual support. It seems psychologically plausible that a reader would experience fewer regularisation errors and lower reliance on phonological recoding when the contextual support for a particular lexical representation is strong. However, the challenge becomes one of parameter choice and benchmarking. Rather than just benchmarking L-DRC under one set of conditions, such as the zero-contextual-input condition, the performance of L-DRC must be separately benchmarked under a range of levels of contextual input, to determine whether or not the sublexical route involvement in reading aloud is appropriate at varying levels of contextual input.

The notion of cognitive resource allocation may have some role to play in considering this challenge. Simulations 1 and 2 demonstrated that modelling the allocation of cognitive resources to orthographic processing by normalising the excitation from the letter level to the orthographic lexicon level was useful in ensuring that stimuli of different lengths still contributed roughly the same amount of excitation to the orthographic lexicon. Perhaps a similar model of resource allocation could be applied at the output end of L-DRC. To ensure that the involvement of the lexical route and the sublexical route in exciting phonemes is appropriately balanced despite varying levels of context, one idea might be to model the contribution of each route as being determined by the cognitive resources allocated to each route. This could be modelled by normalising the excitation provided by the phonological

lexicon to the phoneme level, and in turn normalising the excitation of the phoneme level to the phonological lexicon.

Empirical evidence for how beginning readers learn challenging words

Our simulations showed that L-DRC experiences difficulties learning certain types of words, such as heterophonic homographs, heterographic homophones, and potentiophones. Do beginning readers experience similar difficulties? If so, then rather than L-DRC's imperfect performance being a shortcoming, it would instead reflect a real life effect. What kind of experiments might identify whether L-DRC is accurately modelling difficulties that readers face as opposed to merely demonstrating a shortcoming in L-DRCs own operation?

Difficulty learning new words when reading quickly, without contextual support

DRC has difficulty learning any words—even regular words—when contextual support is zero and the *MinReadingPhonology* parameter is set to 0.4, simulating speeded reading. It seems intuitively obvious that beginning readers will have great difficulty trying to learn to recognise and read novel written words when they are reading words quickly and in isolation, without the contextual support of the text accompanying the word. Therefore, it seems highly plausible for DRC to also have difficulty in learning under these conditions. Still, this aspect of the model could be empirically tested by teaching young readers how to pronounce isolated pseudowords (which mimic novel words) under various speed conditions, then assessing the extent of orthographic learning under each condition. Wang et al. (2011) provide examples of how orthographic learning can be measured, including spelling tests, an orthographic choice task (identify the target item when it is presented along with a homophone and two visual distracters), and an orthographic decision task (a variation of the orthographic choice task where the candidate orthographic forms are not presented simultaneously, but are instead presented in random order on flash cards, with the participant having to answer yes or no if the flash card being considered presented a correctly learned

orthographic form). These results could then also be contrasted to children attempting to learn pseudowords embedded in text, without time pressure, a process also described in, for example, Wang et al. (2012).

Heterophonic homographs

L-DRC has trouble learning the multiple pronunciations of heterophonic homographs under low context conditions. Whether beginning readers experience the same difficulty could be investigated empirically in a number of ways. Firstly, early readers could be compared with skilled control readers on pronunciation of heterophonic homographs, and also on pronunciation of matched homophonic homographs. The target pronunciation could be varied for the heterophonic homographs by altering the context in which the stimuli are presented (e.g., present the word BOW in two separate sentences, such as "The captain stood on the BOW of the ship." and "I tied the knot in a BOW." and context could likewise be varied for the homophonic homographs, e.g., "I took my money to the BANK." and "I stood by the BANK of the river." If early readers show a higher error rate for heterophonic homographs, compared to homophonic homographs, this would show that such items are generally challenging for early readers.

A second experiment could involve teaching early readers pseudoword heterophonic homographs, with each pronunciation for a homograph accompanied with a concocted context. The accuracy and speed of learning could be compared for learning pseudoword heterophonic homographs under clear contextual conditions, versus ambiguous contextual conditions, versus no context conditions. This would shed light on the role context may play in assisting early readers to learn the multiple pronunciations and the context in which each applies.

Heterographic homophones

L-DRC has trouble learning the multiple orthographic representations of heterographic homophones under higher context conditions, but was able to do so more successfully when context was low. This result could be empirically tested by presenting early readers with pseudoword heterographic homophones both with and without context, and observing the impact that context plays on the success of learning the multiple orthographic representations. For example, participants could be taught the word SKALL-/sk9l/ in the sentence "I use a skall for cleaning my fishtank." and also taught the word SCALL-/sk9l/ in a sentence like "A scall of waves crashed on the beach.", and their capacity for learning assessed against a control group learning these words and pronunciations without any context.

Potentiophones

L-DRC was prone to incorrectly learning to associate a regular pronunciation with an irregular word, if the irregular word was a potentiophone. Friedmann and Lukov (2008) have already found that surface dyslexics are more error prone when reading aloud irregular words that were potentiophones than reading aloud irregular words that were not potentiophones. In some respects, the idealised beginning reader we are modelling, with an intact sublexical route and an empty orthographic lexicon is the same as an ideal surface dyslexic, so it is not surprising that L-DRC has trouble with potentiophones. To the extent that beginning readers can be regarded as developmental surface dyslexics, L-DRC is accurately reflecting a difficulty for beginning readers in reading and self-teaching potentiophone pronunciations. To more conclusively demonstrate this, the experiments conducted by Friedmann and Lukov could be repeated, but with young readers rather than surface dyslexics, to see whether young readers have similar difficulties with potentiophones to those encountered by the surface dyslexics, and by L-DRC.

Irregular words

L-DRC has difficulty learning irregular words under low context conditions, but improves with the provision of contextual support. This could be easily investigating by investigating the success of teaching early readers irregular pseudowords both with and without context. Such research has already been conducted by Wang et al. (2011), who found that contextual support improves irregular pseudoword learning, but not regular pseudoword learning.

The problem of partial decoding

While L-DRC is able to display a degree of partial decoding for irregular words if phoneme-to-phonological-lexicon inhibition is reduced, it is not able to do so without introducing a good deal of erroneous learning and incorrect reading aloud. The existence of these errors suggests that a permanent change to a lower level of inhibition as a default value would be problematic. Alternatively, temporarily reducing inhibition just to display the capacity for partial decoding for a particular experiment is ad-hoc and hard to justify. Before we make any other changes to the model, a good understanding of the way partial decoding contributes to learning to read across a range of skill levels is required.

The ideal form of learning is that a beginning reader could take advantage of partial decoding as required to provide additional information when attempting to read a novel word, without having partial decoding contribute to error making. Whether or not this ideal form of learning is an accurate account of how people actually learn to read is not clear. What is clear is that whether or not partial decoding contributes to errors, skilled readers eventually do learn how to read correctly. In its current form, L-DRC possesses no mechanism to correct errors. L-DRC will need to be modified to either reduce the potential of partial decoding to cause errors, or else to introduce a mechanism for errors to be corrected if better information is presented.

One promising way to further explore partial decoding in L-DRC is to investigate an alternative approach to simulating varying levels of certainty provided by contextual input. In the present L-DRC model, context provides excitation to only a single phonological lexicon node. We are in effect simulating that contextual input is completely unambiguous in identifying the spoken word to which the written stimulus corresponds, while the speed with which this contextual information contributes to recognition of the spoken word varies depending on the magnitude of the contextual input selected. A different way of simulating contextual input would be to have multiple phonological lexicon nodes activated, with the number of nodes activated being indicative of the level of certainty provided by context. If context and semantic activity provide a clear indication of what the word should be, then maybe only a single node in the phonological lexicon will be excited by the semantic layer. But if context and semantic activity cannot provide a clear indication, then this could be modelled as multiple phonological lexicon nodes receiving excitation from the semantic layer. Rather than simulating the uncertainty in reading the stimulus by varying the strength of context, this uncertainty could instead be simulated by having more phonological lexicon nodes receive excitation from the semantic layer. For example, in the sentence "Red means stop and green means __", the contextual information provided by the sentence suggests that only the spoken word "go" should receive much excitation from the semantic layer in L-DRC. However, for the sentence "My favourite animal to see at the zoo is the ___", many different spoken words could be appropriate, and the uncertainty as to the correct one could be simulating by distributing excitation to all of them. Under this approach then perhaps the influence of partial decoding could be better simulated. With such an approach, it might be possible to choose a phoneme-to-phonological-lexicon value that allows for partial decoding to select the correct spoken word from the multiple phonological lexicon nodes that are

activated by context, without introducing the kinds of erroneous learning that is apparent when inhibition is decreased too much.

It would be prohibitively complex to determine an appropriate "map" of words to be activated for the countless contextual conditions in which various words might be encountered. But we are not interested in accurately modelling semantics at this stage, we are just interested in how contextual support can impact self-teaching. Therefore, to model this, contextual ambiguity can be simulated by just activating random nodes in the phonological lexicon, in addition to the correct node (e.g., if the target word is "yacht", maybe three random other spoken words could be excited. They could be anything, say "dog", "catch" and "raise". Semantically this is nonsensical, but in terms of self-teaching and the operation of the non-semantic lexical and sublexical routes, it is computationally indistinguishable from a case where more meaningful spoken words are activated.)

Black and white learning

L-DRC uses a simple approach to learning. If beginning readers find that a novel stimulus seems to correspond to a known spoken word, they are in a position to learn, and maybe a single exposure will serve to create confidence in a particular association between a written stimulus and a particular pronunciation. However, L-DRC is stark in the choices it makes. If a phonological node is activate above threshold, then a learning event is triggered, which involves the definite creation of an orthographic node with all the requisite connections between this new node and the letter and phonological lexicon layers. So after a single exposure, L-DRC is able to develop strong, unambiguous knowledge about a word, and deliver a confident response when subsequently reading aloud this stimulus. While continued exposure to this word will change its frequency and thereby impact the time taken to read aloud the word, even one exposure is sufficient for naming accuracy, and for the model to simulate orthographic knowledge of the word.

This simplistic approach to learning is ultimately not psychologically plausible. While previous studies indicate that orthographic learning can certainly happen in a small number of exposures (e.g., Nation et al., 2007), L-DRC takes this to an extreme, and does so for *every* learning instance. Research that suggests there is an ongoing impact of increased orthographic vocabulary size, such as work on the lexical tuning hypothesis (Castles, Davis, Cavalot, & Forster, 2007; Castles, Davis, & Letcher, 1999; Forster & Taft, 1994), also runs counter to the absolutism of learning in L-DRC.

As well as learning instantly, L-DRCs approach to learning is final—no mechanism is described that could allow for a node that was learned with incorrect associations to be modified by subsequent learning, such as direct instruction from a teacher. To increase the psychological plausibility of L-DRC's learning mechanism, L-DRC would need to have the capacity to alter or improve existing learning, and re-learn how to read a word correctly after having self-taught an error.

L-DRC's design is a useful starting point for exploring the self-teaching hypothesis while retaining as much of DRC-1.2.1's existing approach as possible. However, to provide a better account of a more gradual approach to learning, one that allows for errors to be made and also provides a mechanism to allow for the correction of these errors, a more advanced structure is required. Such a structure would likely involve the introduction of varying connection strengths, based on how well a node has been learned. This might also facilitate frequency information being embodied in connection strengths, instead of in the node-specific resting activity arrangement currently used in DRC-1.2.1 and L-DRC.

The SOLAR model of visual word recognition might be a useful model to adapt for this purpose. It provides a well-researched and argued mechanism for the gradual acquisition of orthographic knowledge, while also avoiding the lengthy learning processes associated

with other connectionist models such as feed-forward back-propagation trained networks. In the results reported in Davis (1999), orthographic learning typically occurred in only a few trials, but this could vary from a single trial for some words, to several dozen trials for some words. This seems like a more realistic distribution of the speed at which orthographic learning could occur for a range of words than L-DRC's single-presentation learning in every case. One aspect of the computational implementation of SOLAR model that we found incompatible with our goals was that it describes single-route, purely orthographic learning, with no involvement of phonological recoding, and therefore no capacity to simulate self-teaching. This made the computational version of the SOLAR model seem incompatible with the self-teaching hypothesis. Future work could look to adapting the SOLAR model so that the learning and restructuring that it undergoes is initiated by appropriate activation of a phonological lexicon node. In this way, the orthographic nodes that come to classify particular strings of letters according to a SOLAR-type approach to learning could also be associated with a spoken word in the phonological lexicon. Such an approach implies a significant step away from the interactive-activation structure that has to-date been central to DRC's computational account.

Moving from an idealistic model to a realistic model

L-DRC assumes a fully-intact spoken word vocabulary, and a comprehensive knowledge of graphemes and grapheme–phoneme correspondences. This is an idealised model, and in reality, some beginning readers would be self-teaching new orthographic forms with an incomplete knowledge of GPCs, such as just simple knowledge of single-letter–single-phoneme correspondences, and no knowledge of complex graphemes. Some beginning readers might also receive very little direct instruction regarding GPCs, and instead impute their own GPCs based on a strong mix of direct instruction and textual constraints, as may be the case for a child learning to read via a pure "whole-language" method of reading

instruction (Goodman, 1989). A psychologically plausible account of learning to read should include learning mechanisms to handle this type of learning as well.

To develop a complex computational model able to describe realistic learning is a considerable undertaking. One way of incrementally approaching this realistic model by building on L-DRC would be to experiment with lesioning parts of L-DRC. For example, instead of simulating self-teaching with a fully skilled sublexical route, a version of L-DRC could be trained that only has knowledge of a simple set of GPCs, e.g. perhaps only one- and two-letter graphemes, with no context rules and no multi-letter graphemes comprised of three or more letters. This would be straightforward to investigate, and provide a view of what self-teaching might look like with a more rudimentary knowledge of GPCs. Similarly, letter knowledge or phonological lexicon knowledge could also be lesioned to simulate a young, beginning reader trying to learn despite having incomplete letter knowledge, or a limited spoken vocabulary.

Conclusion

L-DRC has successfully introduced learning to the DRC model, and provides a basis for further investigations of the dynamic learning-to-read process within the dual-route framework. In addition, L-DRC provides a starting point for adding lower-level detail to the verbal self-teaching hypothesis. This research has shown the promise of implementing the self-teaching hypothesis within the dual-route framework, and also highlights the nature of the difficulties posed by certain classes of words such as potentiophones, heterographic homophones and heterophonic homographs to a reader trying to self teach.

References

Baron, J. (1979). Orthographic and word-specific mechanisms in children's reading of words. *Child Development, 50*(1), 60-72. doi: http://dx.doi.org/10.2307/1129042

Carpenter, G. A., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics and Image Processing, 37*(1), 54-115.

Castles, A., Davis, C., Cavalot, P., & Forster, K. (2007). Tracking the acquisition of orthographic skills in developing readers: Masked priming effects. *Journal of Experimental Child Psychology, 97*(3), 165-182. doi: http://dx.doi.org/10.1016/j.jecp.2007.01.006

Castles, A., Davis, C., & Letcher, T. (1999). Neighbourhood effects on masked form priming in developing readers. *Language and Cognitive Processes, 14*(2), 201-224. doi: http://dx.doi.org/10.1080/016909699386347

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100*(4), 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.

Coltheart, M., Saunders, S. J., & Tree, J. J. (2010). Computational modelling of the effects of semantic dementia on visual word recognition. *Cognitive Neuropsychology, 27*(2), 101-114.

Coltheart, M., Woollams, A., Kinoshita, S., & Perry, C. (1999). A position-sensitive Stroop effect: Further evidence for a left-to-right component in print-to-speech conversion. *Psychonomic Bulletin & Review, 6*(3), 456-463.

Davis, C. (1999). *The Self-Organising Lexical Acquisition and Recognition (SOLAR) model of visual word recognition.* (PhD), University of NSW, Sydney. Retrieved from http://www.pc.rhul.ac.uk/staff/c.davis/Thesis/

Dual-Route Cascaded Model 1.2.1. (2009) [Computer Software]. Sydney: Macquarie Centre for Cognitive Science, Macquarie University. Retrieved from http://www.maccs.mq.edu.au/~ssaunder/DRC/2009/10/drc-1-2-1/

Dufau, S., Lété, B., Touzet, C., Glotin, H., Ziegler, J. C., & Grainger, J. (2010). A developmental perspective on visual word recognition: New evidence and a self-organising model. *European Journal of Cognitive Psychology, 22*(5), 669-694. doi: 10.1080/09541440903031230

Finn, P. J. (1977-1978). Word frequency, information theory, and cloze performance: a transfer feature theory of processing in reading. *Reading Research Quarterly, 23*, 510-537.

Firth, I. (1972). *Components of reading disability*. Unpublished doctoral dissertation. University of NSW.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning & Verbal Behavior, 12*(6), 627-635.

Forster, K. I., & Taft, M. (1994). Bodies, antibodies, and neighborhood-density effects in masked form priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(4), 844-863. doi: http://dx.doi.org/10.1037/0278-7393.20.4.844

Friedmann, N., & Lukov, L. (2008). Developmental surface dyslexias. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, 44*(9), 1146-1160. doi: http://dx.doi.org/10.1016/j.cortex.2007.09.005

Glotin, H., Warnier, P., Dandurand, F., Dufau, S., Lété, B., Touzet, C., . . . Grainger, J. (2010). An Adaptive Resonance Theory account of the implicit learning of orthographic word forms. *Journal of Physiology Paris, 104*(1-2), 19-26.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5*(4), 674-691.

Goodman, Y. M. (1989). Roots of the whole-language movement. *The Elementary School Journal, 90*(2), 113-127. doi: http://dx.doi.org/10.1086/461607

Gough, P. B., & Walsh, M. A. (1991). Chinese, Phoenicians, and the orthographic cipher of English *Phonological processes in literacy: A tribute to Isabelle Y Liberman* (pp. 199-209). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc; England.

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, 111*(3), 662-720. doi: http://dx.doi.org/10.1037/0033-295X.111.3.662

Jacobs, A. M., & Grainger, J. (1994). Models of Visual Word Recognition-Sampling the State of the Art. *Journal of Experimental Psychology: Human Perception and Performance, 20*(6), 1311-1334.

Jorm, A. F., & Share, D. L. (1983). Phonological recoding and reading acquisition. *Applied Psycholinguistics, 4*(2), 103-147. doi: http://dx.doi.org/10.1017/S0142716400004380

Marshall, J. C., & Newcombe, F. (1973). Patterns of paralexia: A psycholinguistic approach. *Journal of Psycholinguistic Research, 2*(3), 175-199. doi: http://dx.doi.org/10.1007/BF01067101

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88*(5), 375-407.

McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. *Psychology of learning and motivation - advances in research and theory, 24*(C), 109-165.

Nagy, W. E., & 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*. Hillsdale, NJ: Erlbaum.

Nation, K., Angells, P., & Castles, A. (2007). Orthographic learning via self-teaching in children learning to read English: Effects of exposure, durability, and context. *Journal of Experimental Child Psychology, 96*(1), 71-84. doi: http://dx.doi.org/10.1016/j.jecp.2006.06.004

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review, 114*(2), 273-315.

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the connectionist dual process (CDP++) model. *Cognitive Psychology, 61*(2), 106-151. doi: http://dx.doi.org/10.1016/j.cogpsych.2010.04.001

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.

Pritchard, S. C., Coltheart, M., Palethorpe, S., & Castles, A. (2012). Nonword reading: Comparing dual-route cascaded and connectionist dual-process models with human

data. *Journal of Experimental Psychology: Human Perception and Performance, 38*(5), 1268-1288. doi: http://dx.doi.org/10.1037/a0026703

Rastle, K., & Coltheart, M. (1998). Whammies and double whammies: The effect of length on nonword reading. *Psychonomic Bulletin & Review, 5*(2), 277-282.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*(6088), 533-536. doi: http://dx.doi.org/10.1038/323533a0

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89*(1), 60-94.

Salasoo, A., Shiffrin, R. M., & Feustel, T. C. (1985). Building permanent memory codes: Codification and repetition effects in word identification. *Journal of Experimental Psychology: General, 114*(1), 50-77. doi: http://dx.doi.org/10.1037/0096-3445.114.1.50

Seidenberg, M. S., & McClelland, J. L. (1989). A Distributed, Developmental Model of Word Recognition and Naming. *Psychological Review, 96*(4), 523-568.

Seidenberg, M. S., & Plaut, D. C. (2006). Progress in understanding word reading: data fitting versus theory building. In S. Andrews (Ed.), *From inkmarks to ideas: current issues in lexical processing* (pp. 25-49). Hove, East Sussex, BN3 2FA: Psychology Press.

Shallice, T., & Cooper, R. P. (2011). *The Organisation of Mind*. New York: Oxford University Press.

Share, D. L. (1995). Phonological recoding and self-teaching: sine qua non of reading acquisition. *Cognition, 55*(2), 151-218.

Share, D. L. (2008). Orthographic learning, phonological recoding, and self-teaching. *Advances in child development and behavior (Vol 36)* (pp. 31-82). San Diego, CA: Elsevier Academic Press; US.

Share, D. L. (2011). On the role of phonology in reading acquisition: The self-teaching hypothesis. *Explaining individual differences in reading: Theory and evidence* (pp. 45-68). New York, NY: Psychology Press; US.

Wang, H.-C., Castles, A., & Nickels, L. (2012). Word regularity affects orthographic learning. *The Quarterly Journal of Experimental Psychology, 65*(5), 856-864. doi: http://dx.doi.org/10.1080/17470218.2012.672996

Wang, H.-C., Castles, A., Nickels, L., & Nation, K. (2011). Context effects on orthographic learning of regular and irregular words. *Journal of Experimental Child Psychology, 109*(1), 39-57. doi: http://dx.doi.org/10.1016/j.jecp.2010.11.005

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 50A*(2), 439-456. doi: http://dx.doi.org/10.1080/027249897392170

Appendix A

Parameters specific to L-DRC, default values:

SpokenWordRecognisedThreshold	0.4
WrittenWordRecognisedThreshold	0.4
Semantic2PhonolexExcitation	0.3
Semantic2PhonolexInhibition	0
ContextInput2Semantic	0.25
WrittenWordFrequencyMultiplier	10

Parameters common to both L-DRC and DRC-1.2.1, default values:

General parameters	
ActivationRate	0.2
FrequencyScale	0.05
MinReadingPhonology	0.4
Feature Level Parameters	
FeatureLetterExcitation	0.005
FeatureLetterInhibition	0.15
Letter Level Parameters	
LetterOrthlexExcitation	0.07
LetterOrthlexInhibition	0.48
LetterLateralInhibition	0
Orthographic Lexicon Parameters	
OrthlexPhonlexExcitation	0.25
OrthlexPhonlexInhibition	0
OrthlexLetterExcitation	0.3
OrthlexLetterInhibition	0
OrthlexLateralInhibition	0.06
Phonological Lexicon Parameters	
PhonlexPhonemeExcitation	0.09
PhonlexPhonemeInhibition	0
PhonlexOrthlexExcitation	0.25
PhonlexOrthlexInhibition	0
PhonlexLateralInhibition	0.07
Phoneme Level Parameters	
PhonemePhonlexExcitation	0.04
PhonemePhonlexInhibition	0.16
PhonemeLateralInhibition	0.147
PhonemeUnsupportedDecay	0.05
GPC Route Parameters	
GPCPhonemeExcitation	0.051
GPCCriticalPhonology	0.05
GPCOnset	26

Appendix B

Phoneme symbols are those used in both DRC model 1.2.1 and L-DRC.

Vowels		Consonants	
Symbol	Example	Symbol	Example
1	stay	_	jump
2	sigh	b	buy
3	bird	d	dot
4	boy	f	for
5	goat	g	guy
6	mouth	h	hot
7	beard	j	yell
8	cared	k	kite
9	board	l	low
#	hard / palm	m	my
{	cat	n	no
i	seen	p	pie
u	clue	r	run
E	red	s	stop
I	bid	t	tie
Q	pod	v	vent
U	good	w	west
V	fun	z	zoo
W	few	D	then
		J	chin
		N	hang
		S	shoe
		T	thin
		Z	measure

CHAPTER 3.
Nonword Reading: Comparing Dual-Route Cascaded and Connectionist Dual-Process Models with Human Data

Stephen C. Pritchard, Max Coltheart, Sallyanne Palethorpe and Anne Castles

ARC Centre of Excellence in Cognition and its Disorders, Macquarie University, Sydney

Published in the *Journal of Experimental Psychology: Human Perception and Performance*

Pritchard, S. C., Coltheart, M., Palethorpe, S., & Castles, A. (2012). Nonword reading: Comparing dual-route cascaded and connectionist dual-process models with human data. *Journal of Experimental Psychology: Human Perception and Performance, 38*(5), 1268-1288. doi: http://dx.doi.org/10.1037/a0026703

Author Note

Stephen C Pritchard, Max Coltheart, Sallyanne Palethorpe and Anne Castles, ARC Centre of Excellence in Cognition and its Disorders, Macquarie University, Sydney, Australia.

Raw nonword pronunciation data is available online at: http://personal.maccs.mq.edu.au/~spritcha/111020NonwordReading.xls

We thank Conrad Perry for many useful exchanges regarding the design and operation of the CDP+ and CDP++ models, and Eva Marinus for advice on preparation of this article.

Correspondence concerning this article should be addressed to Stephen Pritchard, Macquarie Centre for Cognitive Science, Macquarie University, Sydney NSW 2109, Australia. E-mail: stephen.pritchard@mq.edu.au

Abstract

Two prominent dual-route computational models of reading aloud are the dual-route cascaded (DRC) model, and the connectionist dual-process plus (CDP+) model. While sharing similarly designed lexical routes, the two models differ greatly in their respective non-lexical route architecture, such that they often differ on nonword pronunciation. Neither model has been appropriately tested for nonword reading pronunciation accuracy to date. We argue that empirical data on the nonword reading pronunciation of people is the ideal benchmark for testing. Data were gathered from 45 Australian-English-speaking psychology undergraduates reading aloud 412 nonwords. To provide contrast between the models, the nonwords were chosen specifically because DRC and CDP+ disagree on their pronunciation. Both models failed to accurately match the experiment data, and have deficiencies in nonword reading performance. However, the CDP+ model performed significantly worse than the DRC model. CDP++, the recent successor to CDP+, had improved performance over CDP+, but was also significantly worse than DRC. In addition to highlighting performance shortcomings in each model, the variety of nonword responses given by participants points to a need for models that can account for this variety.

Introduction

The *dual-route theory* of reading aloud was first described in the early 1970s (Forster & Chambers, 1973; Marshall & Newcombe, 1973), and has been the subject of ongoing research since that time. This theory suggests that two separate mental mechanisms, or cognitive routes, are involved in reading aloud, with output of either or both mechanisms contributing to the pronunciation of a written stimulus. One mechanism, termed the *non-lexical route*, is the process whereby the reader can "sound out" a written stimulus by identifying the constituent parts of the stimulus (letters, graphemes) and, through knowledge of how these parts are associated with phonemes, build up a phonological representation and read the stimulus aloud. The other mechanism, termed the *lexical route*, is the process whereby skilled readers can recognize known words by sight alone without first accessing phonological word representations or the phonemes associated with the constituent graphemes. Direct recognition of the entire written word allows the reader to determine the associated spoken word as a whole, and produce this when reading aloud.

Dual-route theory was initially conceived as a verbal model, often supported with box-and-arrow diagramming (e.g., see Marshall & Newcombe, 1973, p. 189, Fig. 1; Patterson & Shewell, 1987, p. 274, Fig. 13.1). See also Figure 1 for an example of how dual-route theory may be represented in this format. However, over the last three decades, it has become commonplace to implement theories about reading aloud as computational models. Computational modelling is useful because it requires completeness and explicitness, and puts theories into a testable format (see also Norris, 2005, on the benefits of computational modeling).

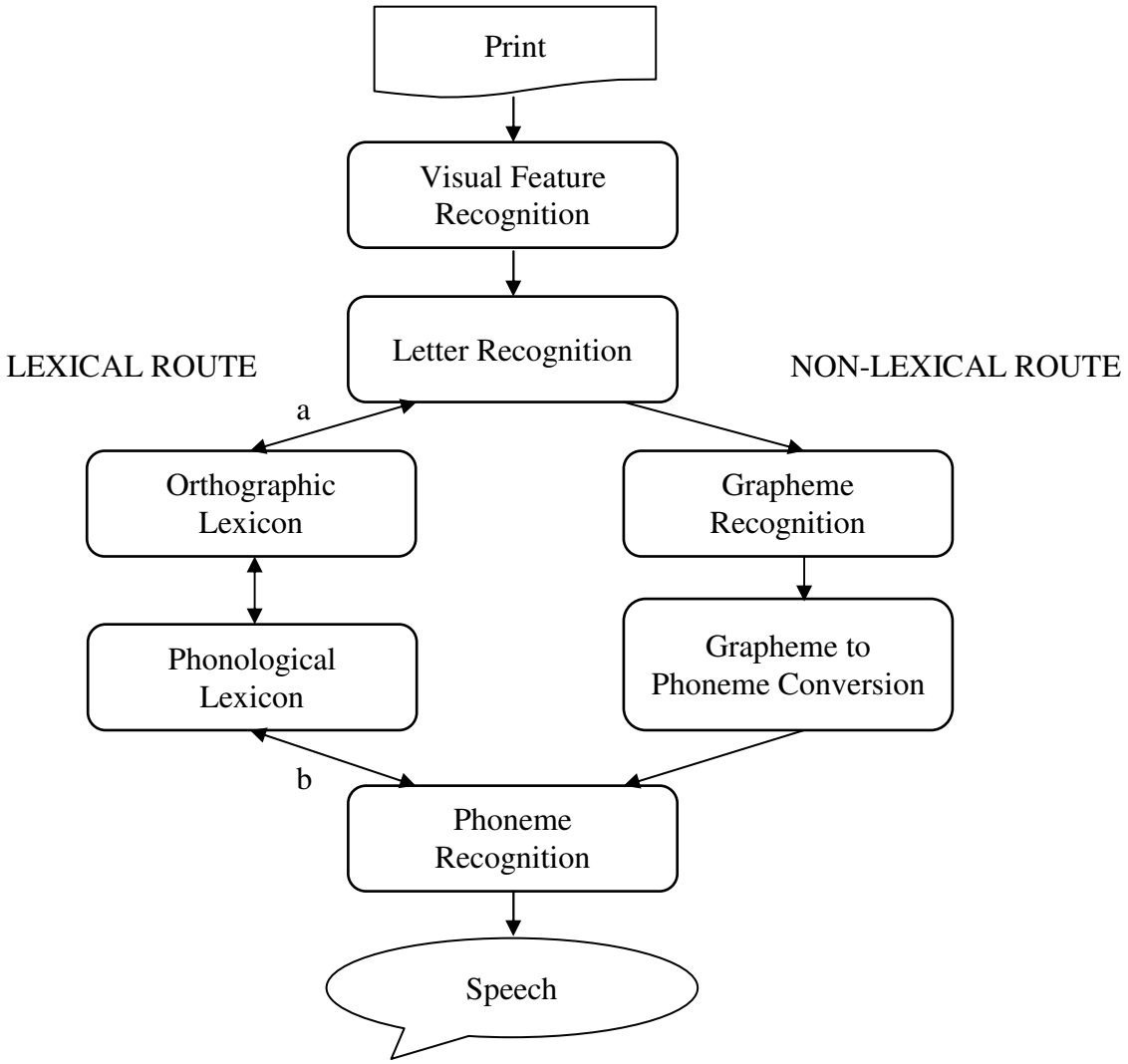

Figure 1 Box-and-arrow depiction of a dual-route model of reading. Note. The dual-route theory also includes a semantic system as part of the lexical route. It is omitted in this diagram for simplicity, because it is not typically implemented in computational models. a) Breaking the letter-to-orthographic-lexicon connections prevents orthographic lexical capture. b) Breaking the phoneme-to-phonological-lexicon connections prevents phonological lexical capture.

Two prominent computational implementations of the dual-route theory of reading aloud that are currently the subject of research and debate are the dual-route cascaded (DRC) model of reading aloud and word recognition (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), and the connectionist dual-process plus (CDP+) model (Perry, Ziegler, & Zorzi, 2007). While these two models were constructed according to the broad theory that there are dual routes for reading, there are computational differences between the two models at a finer grain

of analysis. Most significantly, CDP+ adopts a connectionist structure for its non-lexical route that includes the capacity to learn or be trained. This is in contrast to DRC's classical, rule-based approach, which is static and does not model learning. The aim of the current work was to use new empirical data to test these substantial differences. This will contribute to an understanding of the strengths and weaknesses of each model in providing an account of human reading.

Another prominent family of computational models of reading is the parallel distributed processing (PDP) group of models (Harm & Seidenberg, 1999, 2004; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg, 2005; Seidenberg & McClelland, 1989). The PDP models also include two routes from orthography to phonology, one of which is via semantics. Despite this, PDP modellers do not identify their models as embodying the dual-route theory of reading because they include only a single non-semantic route from orthography to phonology. In contrast, dual-route theory and computational dual-route models such as DRC and CDP+ include two routes that are independent of the semantic system. The PDP models are not considered further in this work, since our focus is on examining the difference between two computational accounts of dual-route theory.

Comparison of Model Architecture

Both DRC and CDP+ share a near identical lexical route, structurally the same but with different parameter settings. In particular, the DRC model has a higher ratio of inhibition to excitation in parts of its lexical route than CDP+. Despite the identical lexical route structure, these parameter value differences between the lexical routes of each model can still result in important performance differences. While lexical route parameter settings are more relevant to understanding word naming and reaction time differences than they are to nonword pronunciation differences, they may still impact nonword pronunciation, since

altering some parameters can affect the degree to which the lexical route becomes involved in nonword processing.

The identical structural implementation of the lexical route for each model is described in detail in Coltheart et al. (2001). In constructing CDP+, Perry et al. (2007) sought to build on DRC by developing a new and improved non-lexical route, while retaining many of the capabilities and structure of DRC. In using aspects of the DRC model, the CDP+ modellers were following the philosophy of *nested modelling* (Jacobs & Grainger, 1994), according to which new models should build on the capacities of previous models, and be able to account for all of the empirical effects that previous models can simulate, even as they seek to account for additional effects.

DRC and CDP+ differ most in the structure and operation of their respective non-lexical routes, although some similarities remain. Both models feature a non-lexical route comprised of the same two functions, which are both depicted in Figure 1. These are 1) parsing of an input sequence of letters into graphemes (grapheme recognition), and 2) activation of phoneme representations based on the identified graphemes (grapheme-to-phoneme conversion). The grapheme parsing procedures of each model are quite alike, and both involve a "hard-wired" (that is, no model training is required) rule-based algorithm for choosing graphemes, although the grapheme representations programmed into each model are different. In addition to being coded with different sets of graphemes, each model has been programmed to handle grapheme position differently. DRC processes graphemes as occurring at the beginning, middle, or end of a word. In contrast, CDP+ processes graphemes as occurring in the onset, vowel, or coda of a word. For example, DRC would parse the word THRIFT as TH (beginning), R (middle), I (middle), F (middle), T (end), while CDP+ would parse the same word as TH (onset), R (onset), I (vowel), F (coda), T (coda).

It is in the second function, the activation of phonemes based on the identified graphemes, that the difference between the two models is most acute. DRC's method of selecting phonemes based on the identified graphemes is rule-based, and hard-wired. The DRC model incorporates knowledge of 236 explicit grapheme–phoneme correspondence rules (GPCs), including 27 context-sensitive rules and 8 output rules ("Dual-Route Cascaded Model 1.2.1," 2009). These rules were programmed into DRC by its creators, who derived the rules by choosing the most common phoneme associated with a particular grapheme across the set of English monosyllables containing that grapheme (Coltheart et al., 2001; Rastle & Coltheart, 1999). The rules are chosen based on type frequency rather than token frequency. Having been constructed this way, the DRC non-lexical route therefore includes no knowledge of grapheme-phoneme associations that are not the most common for that grapheme. For example, DRC includes knowledge that the regular pronunciation of the grapheme OO is the phoneme /ʉː/[1], as in TOOL and BOOT. However, DRC does not include any information that OO corresponds to /ʊ/ in many words, such as BOOK and WOOL. The latter is quite common, just not quite as common as the former. The relationship between graphemes and phonemes in DRC could be broadly characterized as one-to-one, since, with the exception of the small number of context-sensitive rules, the GPCs relate individual graphemes to individual phonemes, and only the most common rule—the regular rule—is ever applied. The rules applied by the DRC model to particular stimuli are unambiguous and transparent to the modeller.

In contrast, CDP+'s mechanism for translating graphemes to phonemes is statistical rather than rule-based, and instead of being hard-wired, the knowledge of grapheme–phoneme associations known by the model must be learned by the model. CDP+ employs a connectionist, two-layer associative (TLA) network to compute which phonemes to activate,

[1] 1The phoneme symbols used in this article are those proposed for Australian English by Harrington, Cox, and Evans (1997), and are listed in Table A1 of the Appendix.

given the input graphemes that have been identified. Knowledge of how to activate phonemes given the identified graphemes is spread across the weights of many connections that link every input grapheme node to every output phoneme node in the TLA network. There is a many-to-many relationship between graphemes and phonemes in CDP+, in that each output phoneme may be determined by more than one of the input graphemes, and individual graphemes may contribute to the activation of more than one output phoneme. This structure means the choice of correspondences applied by CDP+ to particular stimuli is somewhat opaque to the modeller, since it is challenging for the modeller to identify how correspondences are applied just through inspection of the 200,000+ connection weights in the TLA network.

This difference in grapheme-to-phoneme conversion is also the reason that the CDP+ model—and indeed other connectionist models such as the PDP model (Plaut et al., 1996)—only include explicit knowledge of a much smaller set of graphemes than DRC. CDP+ includes knowledge of 96 graphemes, less than half the number known by DRC. With its rule-based system and large set of known graphemes, most of the computational work of the DRC non-lexical route is in the parsing of graphemes, and the grapheme-to-phoneme conversion process is a simple table lookup procedure. In contrast, the CDP+ model and other connectionist models spread this knowledge over both the parsing of graphemes and subsequent conversion to phonemes. It is the computational complexity of the connectionist network that seems to drive the CDP modellers to use a smaller number of explicitly known graphemes. The preference to locate knowledge within the connection weights of the network seems to be a connectionist design strategy (see Hinton, 1990; as cited in Plaut et al., 1996). These differences can be clearly understood with an example. Consider the word BAKE. DRC will identify the split grapheme A.E while parsing this word, and once this grapheme has been identified, it is a simple matter of producing the corresponding sound, which is /æɪ/.

When CDP+ undertakes grapheme parsing, it identifies A and E as two separate graphemes, placing A in the vowel slot and E in a coda slot, since the split grapheme A.E is unknown to CDP+ (CDP+ does not recognize any split graphemes). However, the task of correctly producing the long vowel /æɪ/ is performed in the grapheme-to-phoneme conversion process. The complexities of the pattern of connection weights trained for the CDP+ model ensure that the activated A and E grapheme nodes, along with the activation of a consonant node between these two graphemes, work together to produce the correct phoneme /æɪ/. In a sense, the CDP models adopt a more relaxed definition of what a grapheme is, with multiple graphemes potentially contributing to the activation of multiple phonemes. This is compared to DRC's more strictly adhered to definition that a grapheme is a group of letters corresponding to a single phoneme[1].

One significant advantage of the CDP+ connectionist structure over DRC is that it can capture relationships between graphemes and phonemes that are present in words, but are not necessarily the most common, such as the irregular but common OO–/ʊ/ correspondence. While the DRC non-lexical route will always produce regular responses, the CDP+ non-lexical route can deviate away from regularity. It can do this if the pattern of all graphemes present in the input, combined with the learned knowledge embodied in the connection weights, support an irregular correspondence. In effect, the connectionist structure of CDP+ allows its non-lexical route to include knowledge beyond just grapheme-phoneme correspondences. It may also effectively embody knowledge about correspondences between larger orthographic units (such as word bodies or even whole words) and groups of phonemes. This is perhaps the reason why the CDP+ modellers prefer the term "sub-lexical route", rather than "non-lexical route" (Perry et al., 2007). This capacity is advantageous when it comes to modelling responses to nonword stimuli which, according to dual-route

[1] The obvious exception for DRC is the grapheme X, which corresponds to the two phonemes /ks/, rather than a single phoneme.

theory, depend strongly on the operation of the non-lexical route for their pronunciation. Andrews and Scarratt (1998), Glushko (1979) and Treiman, Kessler, and Bick (2003) found that people often produce irregular responses when reading aloud nonword stimuli, and this is particularly the case for nonwords that have bodies that are often or always pronounced irregularly when found in words. For example, the nonword BALF possesses the body -ALF, which is never pronounced regularly as /ælf/ in real monosyllabic words, but is instead pronounced irregularly with no /l/ as /ɐːf/ such as in the word "calf" in Australian English. The DRC non-lexical route is unable to model the often irregular responses people produce to such nonwords, while CDP+ has the potential to do so, if it has learnt to produce irregular pronunciations in just the same way that people do.

Whether or not CDP+ produces nonword pronunciations in the same way that people do has not been adequately determined. When assessing the nonword pronunciation performance of CDP+, Perry et al. (2007) adopted the scoring criterion of Seidenberg, Plaut, Petersen, McClelland, and McRae (1994) (who adopted this criterion after noting that it was also used by Glushko (1979) and McCann and Besner (1987)). According to this criterion, nonword responses are deemed correct if the output phoneme string incorporates only grapheme–phoneme or body–rime relationships that exist in real words. Using this criterion, CDP+ had an error rate of 6.25% when reading aloud the 592 nonwords that Seidenberg et al. reported. Perry et al. argue that this is acceptable performance for CDP+ since it is similar to the human error rate of 7.3% reported in Seidenberg et al. However, it could be argued that this criterion is not an adequate measure of nonword correctness, because many responses are scored correct even though people may in fact not respond in the same way. Consider for example the nonword BONTH. CDP+ pronounces this to rhyme with "month", i.e., pronouncing the O as a U. According to the scoring criterion, this would be considered a correct nonword response. However, it seems unlikely that many—if any—people would

pronounce BONTH in this way. Perry et al. implicitly acknowledge that this method of measuring nonword reading performance may not be sufficiently rigorous, since they label it as a "lenient error scoring criterion" (p. 285).

DRC's nonword pronunciations have also not been appropriately tested. Coltheart et al. (2001) used regularity as the sole measure of correctness when assessing the nonword pronunciation performance of DRC. They found that DRC made 75 errors when tested on 7,000 three- to seven-letter monosyllabic nonwords obtained from the ARC Nonword Database (Rastle, Harrington, & Coltheart, 2002), an error rate of 1.07%. However, it is now well known that because people often produce irregular pronunciations for nonwords (Andrews & Scarratt, 1998), regularity is not an appropriate measure of correctness.

Given that the nonword pronunciation accuracy of the two models has not been adequately tested and reported to date, we sought to perform this assessment. But in order to do so, the first step is to determine acceptable criteria for judging the correctness of model responses to nonwords. We contend that the optimal benchmark for determining nonword pronunciation accuracy for the models is actual human pronunciation. Previous studies (Andrews & Scarratt, 1998; Glushko, 1979; Masterson, 1985) have found that many nonwords are given more than one common pronunciation when read aloud by people, and in light of this, our analysis did not simply take the most frequent human response to a nonword as the only correct response when assessing the models. We also considered the less common nonword responses. With this in mind, we aimed to gather data on the way people pronounce nonwords, in order to then use these data to assess the nonword reading performance of CDP+ and DRC.

Method

The nonwords that are most useful in distinguishing nonword reading performance between the two computational models are those where the models differ on pronunciation. Our approach was therefore to develop a set of nonwords where DRC and CDP+ disagree, test these nonwords with human readers, and use these as a focus of investigation.

While our aim was initially to assess the nonword pronunciation performance of only DRC and CDP+, two other model variations were also considered in our analysis. Perry et al. (2007) suggest that reducing the *phoneme naming activation criterion* parameter of the CDP+ model from its default value of 0.67 down to 0.50 is more appropriate for nonword-only reading. The rationale they provide for this change is that nonwords in the CDP+ model typically produce less activation than words, since they do not have lexical entries and therefore the lexical route contributes little to the activation of output phonemes. Regardless of whether or not such a parameter change could be justified for nonword-only reading, if adjusting the phoneme naming activation criterion in this way does improve nonword reading, then the CDP+ modellers may elect to permanently alter this parameter for all reading, assuming it does not result in CDP+ performance deteriorating on other benchmarks. In light of this possibility, all of the nonwords we ran through DRC and CDP+ were also input to CDP+ with the phoneme naming activation criterion set to 0.50, and we included these model results in our analysis. Results for this variation are labelled as “CDP.50”.

Subsequent to the testing described in this experiment, the CDP++ model of multi-syllabic reading was published (Perry, Ziegler, & Zorzi, 2010) and we also conducted simulations of reading aloud with this model and have included these in our analysis. Perry et al. (2010) indicate that for monosyllabic words, CDP++ should operate similarly to CDP+. However, we have included CDP++ separately in our analysis, because several aspects of

CDP++ may still cause its performance to differ from that of CDP+ on monosyllabic nonword reading. Firstly, CDP++ has been coded to recognize an updated set of graphemes to CDP+, so differences in grapheme parsing may result in CDP++ producing a different pronunciation to CDP+ for monosyllabic stimuli. Secondly, CDP++ was trained on a different corpus of training words to CDP+ (disyllabic words were included). Finally, lexical feedback from disyllabic words in the CDP++ lexicon may still impact on the pronunciation of nonwords in ways that causes CDP++ to produce a different pronunciation to CDP+.

Stimulus Selection

A set of 1,475 nonwords two-to-seven letters in length was randomly selected from a subset of the ARC Nonword Database (Rastle et al., 2002). We did not select nonwords greater than seven letters because DRC is only able to process to completion nonwords of up to seven letters. DRC has eight letter slots, but for nonword processing, the end-of-stimulus character must be present for completion of processing, and takes up the eighth slot.

The nonwords were selected from a subset of the database consisting of only monosyllabic, monomorphemic nonwords. In addition, all nonwords were comprised of existing onsets, existing bodies and legal bigrams. This was done to avoid contentious aspects of nonword reading. Perry et al. (2007) acknowledge that the CDP+ model may have difficulty reading nonwords with extremely uncommon or illegal spelling patterns. However, they argue that the reading of illegal nonwords is not a viable benchmark of reading performance, because they are read aloud in a manner qualitatively different from normal reading. Finally, none of the nonwords selected were pseudohomophones, since the ARC Nonword Database website gives the option of selecting either “nonwords” or “pseudohomophones” but not a mix, and we elected to choose from only the items marked as “nonwords”.

In order to select only those nonwords where DRC and CDP+ differ, the 1,475 nonwords were run through each model, and also later run through CDP.50 and CDP++. The default set of parameters for each model were used. The only exception was to modify the phoneme naming activation criterion for CDP.50. These simulations revealed that DRC and CDP+ differed in pronunciation for 412 of the 1,475 selected nonwords (27.9%). The 412 nonwords became the testing set of nonwords that we used in the experiment. The responses of the four models to these 412 nonwords are tabled in the Appendix.

Participants

Participants were 47 Standard-Australian-English-speaking psychology undergraduates from Macquarie University. Each participant was awarded course credit for their involvement. Two participants' data were discarded, one because the recording quality had been too low for transcription, and the other because the participant had an American accent (DRC, CDP+ and CDP++ all primarily derive their pronunciation knowledge from the CELEX Lexical Database (Baayen, Piepenbrock, & Gulikers, 1995), based on British English. While pronunciation might vary, Australian English and British English, as non-rhotic accents, use almost identical phoneme sets and grapheme-phoneme correspondences. General American English, being rhotic, is somewhat different). This left 45 participants (eight male) for our analysis.

Procedure

Participants were informed that they would be viewing a series of "nonsense words". The nonwords viewed by the participants were presented in uppercase white 36pt Times New Roman font on black background. The order of nonwords was randomized for each participant. The nonwords were presented one at a time, in the centre of a CRT computer screen using the software *DMDX* (Forster & Forster, 2003). Participants were asked to read

each nonword out loud, and their responses were recorded using the audio-capture capabilities of DMDX. The experiment was an untimed experiment. Participants had up to 10 seconds to respond, although responses were typically given in under two seconds.

A voice key advanced the experiment to the next nonword after each utterance, to avoid unnecessary delay should the participant answer in a much shorter time than 10 seconds. Participants had the opportunity to rest after every 44 words. The experiment duration was typically around 50 minutes.

Transcription

Recordings were transcribed into written phonemic representations separately by two transcribers. The first transcriber (the third author) is an experienced speech transcriber who worked with the aid of spectral analysis using the *EMU Speech Database System* and associated speech analysis tools (Cassidy & Harrington, 2001). The first transcriber had no access to the pronunciations generated by each of the computational models. The second transcriber (the first author) did the transcription by ear, to assist in detecting errors.

Responses where the two transcribers disagreed were revisited, with agreement usually reached on the transcription. In the event of continued disagreement (<5% of responses), the transcription of the first, experienced transcriber would typically be used.

Data Cleaning

On occasion, multiple answers were recorded by a single participant in response to the one stimulus, for example because the participant spoke too quietly to initially trigger advancement to the next item and the participant repeated themselves, or because the participant had quickly attempted to alter their response before recording ceased. Multiple answers that were identical were treated as a single response. Where the multiple answers differed, however, they were discarded (0.54% of responses). 1.7% of recordings were

discarded because non-vocal noise prematurely triggered a progression to the next item resulting in a truncated recording that included either no response or else a partial response without a vowel. In total, 2.3% of the recordings were discarded. As a result, the dataset used in our analysis consisted of 18,118 valid recordings, coming from 45 participants tested on 412 words.

Results and Discussion

Participant responses were collated to identify frequency of response to each item (see the Appendix), which includes the three most common pronunciations that the participants gave for each nonword). Table 1 provides descriptive statistics, including both by-item and by-subject statistics, detailing the degree of match between each of the models. It is clear from these results that none of the models does an adequate job of reproducing the human data. No model matches the most frequent participant responses sufficiently well to be an adequate model of an average reader, and no model properly accounts for the large variety of responses produced for many of the items tested. Comparing the models, the results clearly indicate—even without statistical analysis—that the DRC model produced a significantly better match to the participant pronunciations than did any of the CDP model variations. Figure 2 charts the degree of match between each of the models and the participant responses, separately showing how often each model matches the most frequent response to a nonword, the second, third, and subsequent most frequent responses, and also the percentage of the 412 nonwords for which each of the models fails to match any of the participants.

As can be seen in the first row of Table 1, and in Figure 2, DRC matched the most frequent participant response far more often than any of the CDP models. CDP++ also matched the most frequent human response less than DRC, but more than the two CDP+

model variations, indicating that this new iteration of the CDP structure offers improvements over its predecessors that go beyond its headline capacity to process multi-syllabic words.

Table 1. Comparison of Model and Participant Pronunciations, By-Item and By-Subject

Descriptive statistics	DRC	CDP+	CDP+.50	CDP++
Percentage of nonwords in which a model matches:				
The most frequent participant response	73.5	12.1	19.9	37.6
None of the participants	1.5	49.0	40.0	26.9
By-subject: percentage of nonwords for which a participant matches a model				
Minimum	29.3	4.9	8.1	17.8
Maximum	68.2	16.2	23.5	38.5
M	53.0	11.3	17.5	30.1
SD	9.0	2.5	3.0	4.4
Mdn	52.7_a	11.4_b	17.5_b	30.8_b
By-item: percentage of participants who match a model for a given nonword				
Minimum	0.0	0.0	0.0	0.0
Maximum	100.0	100.0	100.0	100.0
M	52.8	11.2	17.4	30.0
SD	28.3	18.5	25.3	32.3
Mdn	53.3_a	2.2_b	4.4_b	15.7_b

Note. DRC medians (a) both by-subject and by-item are significantly greater than the medians of each CDP model (b), $p < .001$

However, examining only the most frequent responses is insufficient because these nonwords were usually given several different responses amongst participants. The mean number of different pronunciations given for a nonword was 8.4 (*SD* = 4.5), and the mean percentage of participants giving the most frequent response to a nonword was only 61.2% (*SD* = 21.4%), suggesting a lot of variety on participant responses. There were only five nonwords out of 412 where the 45 participants were unanimous: ENT, FOZ, GERT, OL and

SPRA. This result is in accordance with previous nonword studies that have found a variety of common pronunciations to some nonwords (Andrews & Scarratt, 1998; Glushko, 1979; Masterson, 1985).

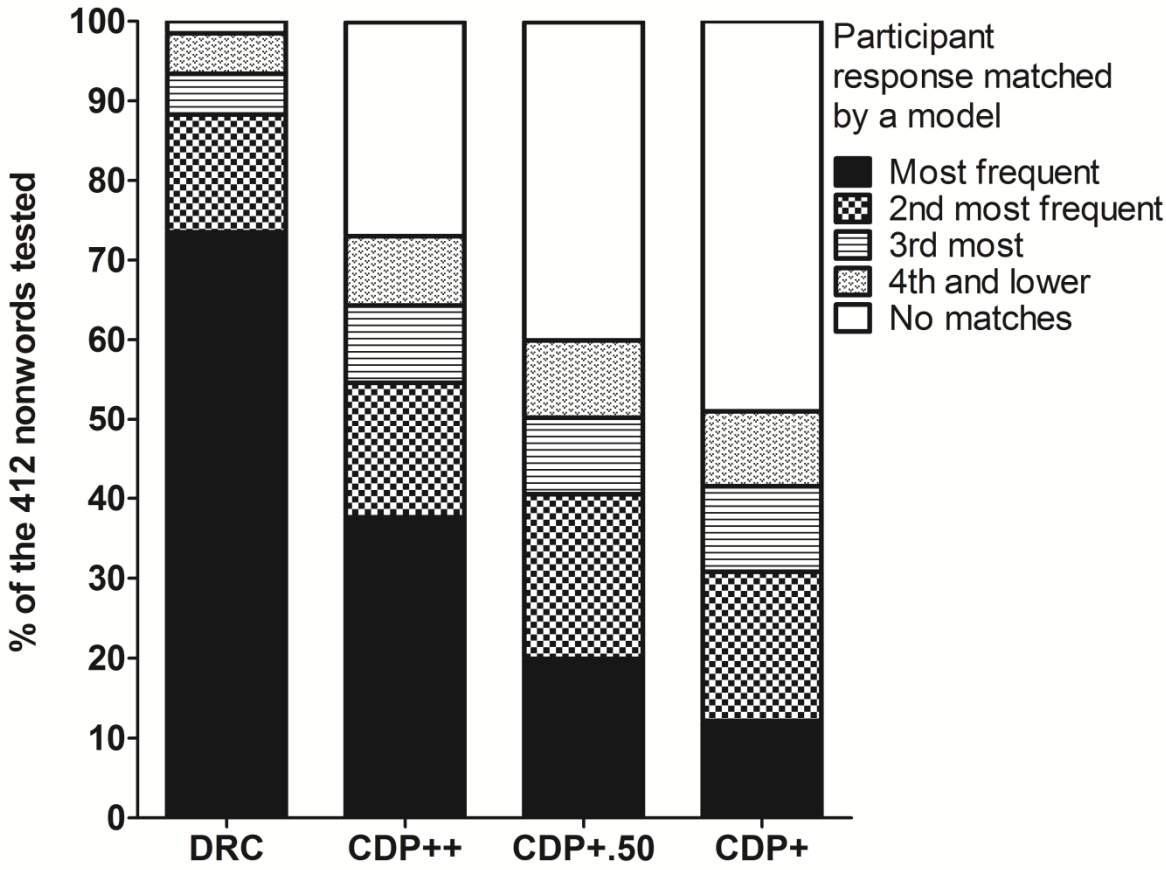

Figure 2. How well each model matches participant responses. Each of the model responses to each nonword is classified as one of: a) matching the most frequent participant response to that nonword, b) matching the 2nd most frequent, c) matching the 3rd most frequent, d) matching the 4th most frequent or some other less common participant response, or e) failing to match any participant. This chart displays the percentage of items that fall into each classification for each model.

Moving beyond most frequent responses to consider all responses, we calculated the mean and median percentage of participants who agreed with each model on pronunciation over the 412 nonwords, and the mean and median percentage of the 412 nonwords for which a participant would agree with each model. The median is a more instructive measure of model performance for this analysis, due to the highly non-normal distribution of the by-items

data, and so we used non-parametric methods (Wilcoxon Signed–Rank) to test for the significance of any difference in median performance between the models. By subjects, the median percentage of the 412 nonwords for which DRC agreed with a participant was significantly greater than the median percentage of nonwords where CDP+ agreed with a participant, $z = 5.8$, $p < .001$, $r = .61$, significantly greater than the median percentage for CDP.50, $z = 5.8$, $p < .001$, $r = .61$, and significantly greater than the median percentage of nonwords for CDP++, $z = 5.8$, $p < .001$, $r = .61$. The by-subjects data minimums and maximums also indicate that even the participant who agreed with DRC the least (*min* = 29.3% of nonwords), still matched DRC on more nonwords than any participant matched CDP+ (*max* = 16.2% of nonwords) or CDP.50 (*max* = 23.5% of nonwords). By items, the median percentage of participants agreeing with DRC on a given nonword was significantly greater than the percentage agreeing with CDP+, $z = 14.6$, $p < .001$, $r = .51$. It was also significantly greater than the median percentage of participants agreeing with CDP.50, $z = 13.5$, $p < .001$, $r = .47$, and significantly greater than the median percentage of participants agreeing with CDP++, $z = 10.6$, $p < .001$, $r = .37$.

The poorest outcome for a computer model on any given nonword is that it fails to match any of the participant responses for that nonword. For almost half of the 412 nonwords (49.0%), the CDP+ model produced a response that did not match any of the 45 participant responses. CDP+.50 and CDP++ also regularly output pronunciations that did not match any participant response, although both variations improved over CDP+, CDP++ greatly so. In comparison, DRC rarely generated pronunciations that did not match any participant (only six nonwords, or 1.5 % of the 412 nonwords). The six nonwords for which DRC matched none of the participants are CHIEL, FRYMPH, GEECH, GERT, QUE, and RHUKE. Interestingly, one of these items (GERT) was pronounced unanimously by the participants, which provides

strong evidence that DRC's modellers may need to modify the rule that causes the mispronunciation.

Analysis of Nonword Responses

To better understand why each of the models did not match the experiment data, we examined the types of responses given by the experimental participants and the by the models. We took two approaches to this task. Firstly, we identified response that were lexicalizations (a word response to a nonword stimulus), and compared the rate of lexicalization between each of the models and the experiment participants. Secondly, we developed a simple computational algorithm to perform a classification of responses. This analysis enabled us to search for any systematic type of response that might be causing each of the models to pronounce nonwords differently from people.

Lexicalizations

In order to be an accurate model of human reading performance, a model needs to a) produce approximately the same percentage of lexicalizations as people do, b) produce these in response to the same nonwords as people do, and c) produce the same word response to each of these nonwords as people. Having implicitly addressed the latter two points in the main results section where it was clear that each of the models is typically not well matched to the participant data, we focus here on the first point of whether the models produce the same percentage of lexicalizations as people. We identified whether a response was a word or not by searching for the phonemic response string in the CELEX Lexical Database (Baayen et al., 1995). This enabled us to calculate the overall percentage of lexicalizations for each of the models and for the experimental participants across the 412 nonwords, which were: experiment participants 8.5%, DRC 0.0%, CDP+ 26.5%, CDP.50 19.9%, and CDP++ 16.5%.

DRC, the CDP models and the experiment participants all produced strikingly different rates of lexicalization. Unlike the experiment participants, DRC did not produce any lexicalizations. In contrast, all of the CDP model variations produced high rates of lexicalization, which was also unlike the experiment participants. CDP+ in particular produced word responses to over a quarter of the 412 nonwords, more than triple the rate of lexicalization evident in the experimental data.

To understand the different lexicalization behaviour of the models, we considered the mechanisms by which dual-route models can produce word responses to nonword stimuli. There are three operations by which a dual-route computational model such as DRC or the CDP models could generate a word response:

1. Orthographic lexical capture: activation cascading from the letter level to the orthographic lexicon level may result in orthographic neighbours to the nonword stimulus receiving high activation. This activity could continue on through the lexical route and may eventually influence the choice of phonemes at the output, resulting in a word response. This is analogous to a reader seeing a written nonword, but confusing it with an orthographically similar written word, e.g. confusing NATCH with WATCH.

2. Phonological lexical capture: in both the DRC and CDP models, activation of the phoneme level by the non-lexical route can potentially back-activate spoken word representations in the phonological lexicon level, which, once activated, would then feed back to the phoneme level and alter the activation of phonemes. This is a purely phonological interaction, not involving the orthographic lexicon. It is analogous to a reader preparing a response to a nonword through sounding out the nonword grapheme by grapheme, but, while preparing the response, the reader comes to confuse the potential nonword speech utterance with a similar, actual spoken word (e.g. tentatively reading the nonword BLIGN as “bline”

(rhyming with "line"), but being influenced by the similarity of this utterance to the spoken word "blind", and uttering this instead).

3. Regular and irregular pseudohomophones: the general operation of the non-lexical route may inadvertently result in a word response, completely independently of the lexical route. For DRC, this only happens if the nonword stimulus is a regular pseudohomophone—a nonword that happens to sound the same as a word when pronounced regularly. For example, the nonword BRANE is a regular pseudohomophone since it is pronounced as the word "brain" according to regular grapheme–phoneme correspondences. However, for each of the CDP models, the non-lexical route converts graphemes to phonemes in a categorically different manner to DRC. The CDP non-lexical route can independently produce a spoken word response, even for irregular nonwords. From the perspective of a CDP modeller, these might be regarded as *irregular pseudohomophones*. For instance, the regular pronunciation of BREKE is "breek", but the CDP+ non-lexical route, even when operating completely independently of the lexical route, outputs the lexical response "brake" to this nonword.

DRC produces no lexical responses for two reasons. Firstly, inhibition parameters in DRC are set such that the lexical route will not contribute to the production of a nonword response at all. Letter-to-orthographic lexicon (L-to-O) inhibition and phoneme-to-phonological lexicon (P-to-P) inhibition are set sufficiently high that orthographic and phonological word neighbours to a nonword stimulus will not be activated. Even one letter or phoneme different contributes sufficient inhibition to prevent activation. These parameter settings were chosen in the context of irregular nonword responses being deemed incorrect by the DRC modellers, a position that is now seen as inappropriate based on the data in the current paper. As a result, DRC will not experience orthographic lexical capture or phonological lexical capture. Secondly, none of the nonwords selected from the ARC nonword database was a regular pseudohomophone. The experiment results seem to point to a

need for adjustment of lexical-route parameters for DRC. In producing no lexical responses, DRC is very much unlike the experimental participants, and an easing of inhibitory parameters in the lexical route to allow more lexical route involvement in nonword pronunciation may allow DRC to produce lexical responses in the way people do.

To investigate which of these three mechanisms were involved in the generation of lexical responses in the CDP+ model, which had the highest rate of lexical response of all the models, we prepared two test variations of CDP+. The first of these involved setting the L-to-O excitation parameter to zero. By breaking this L-to-O connection, the CDP+ model is unable to produce lexicalizations via orthographic lexical capture. The second variation involved setting the P-to-P excitation parameter to zero. By breaking this P-to-P connection, the CDP+ model is unable to produce lexicalizations via phonological lexical capture. The location of these connection and parameter changes are noted on the diagram of the full DRC model displayed in Figure 1.

The 412 nonwords were processed using these two CDP+ variations, and the percentage of lexicalizations recorded for each. CDP+ with the L-to-O connection broken produced almost the same degree of lexicalization as the default CDP+ model (in fact, there was only one nonword out of 412 for which they differed). This suggests that CDP+ is not producing lexicalizations via orthographic lexical capture. With P-to-P excitation set to zero, CDP+ produced a lower percentage of lexicalizations (20.9%) than the default CDP+ (26.5%). This suggests that the difference—approximately one fifth of the lexicalizations produced by the default CDP+ model—is due to phonological lexical capture. The remaining four-fifths of lexicalizations produced by CDP+ must therefore be a product of the training and structure of the connectionist CDP+ non-lexical route. To the CDP+ model, these nonwords are irregular pseudohomophones.

Classification by Algorithm

We also sought to categorize the responses of each of the models and of the experimental participants in depth, looking to identify types of responses that might be the main contributors to the poor performance of each model. Model and participant responses were organized into several specific categories, using a classification algorithm coded specifically for this task. The algorithm screens the strings of phonemes produced by a model or transcribed for a participant, and places the strings into categories. The categories are:

1. Regular: regular pronunciation (identical to DRC's response).

2. Vowel difference: the vowel produced in the response differs from the regular pronunciation, all consonants are the same as the regular pronunciation.

3. Dropped phoneme: almost the same as the regular pronunciation, but for a single phoneme omitted.

4. Extra phoneme: almost the same as the regular pronunciation, but for the inclusion of an additional phoneme(s).

5. s/z coda difference: differs from the regular response in that one uses an /s/ in the coda, while the other uses a /z/ (e.g. if the regular pronunciation of BLISE is /blɑes/ then /blɑez/ would be classed as an s/z coda difference.

6. Consonant difference: the same as the regular pronunciation, but for a single disagreeing consonant (excluding the s/z coda differences).

7. *Other:* some combination of the above differences.

The classification algorithm was applied to the responses to the 412 nonwords for which DRC and CDP+ disagreed, since these were the nonwords that were also tested with

the experiment participants. Results of the categorization process are given in Table 2 for each of the models and the experiment participants.

We first consider regularity. All of the DRC model's responses were regular responses, in that they adhered to the discrete set of GPC rules coded into DRC. However, only 53.0% of participant responses were regular, far less than DRC. This strict adherence to regularity for all nonword responses is an unambiguous shortcoming of the DRC model, one that is also responsible for the poor performance of DRC on the nonword consistency benchmark (that people don't always use the most common grapheme–phoneme correspondences when reading nonwords) reported in Perry et al. (2007). The failure of DRC to produce irregular responses to nonwords while people often do is well known and has been previously noted (Treiman et al., 2003) and is an issue that DRC modellers will need to remedy to achieve tolerable performance on the fundamental benchmark of nonword pronunciation.

Table 2. Classification of Pronunciations Given by Experiment Participants and by Each Model

Classification	Exp[a]	DRC	CDP+	CDP.50	CDP++
Regular	53.0	100.0	0.0	9.0	32.5
Vowel different	20.3	0.0	36.7	35.7	31.6
Dropped phoneme	1.2	0.0	11.4	5.1	2.4
Extra phoneme	4.0	0.0	7.3	5.1	2.9
s/z coda difference	4.0	0.0	10.9	11.9	9.2
Consonant different	5.1	0.0	16.7	20.1	9.2
Other	12.4	0.0	17.0	13.1	12.1

[a]Experiment participants

In contrast, CDP+ produced only irregular responses to the nonwords used in the experiment. While CDP+ would routinely be capable of producing regular responses to nonwords, the basis of selecting nonwords in this experiment was that DRC and CDP+ disagreed on pronunciation. Since DRC is always regular, it follows that none of CDP+'s responses to these particular nonwords could have been regular, since then the response would have agreed with DRC. The fact that CDP+ did not produce a regular response to these nonwords is also quite unlike the experiment data. CDP+.50 and CDP++ both produced some regular responses (9.0% for CDP.50 and 32.5% for CDP++). Considering the wider set of 1,475 nonwords, CDP+ produced a regular response for 71.7% of the 1,475 nonwords; CDP.50 produced a regular response for 73.2%, and CDP++ 75.1%. To the extent that CDP.50 can be regarded as more suited to reading nonwords than CDP+, and CDP++ regarded as an updated version of CDP+, there is a progression towards increased production of regular responses with CDP model iterations, greater agreement with DRC, and greater agreement with the experiment participants.

Turning now to irregular responses, the category accounting for the biggest share of irregular responses produced by each of the CDP models was the *vowel difference* category. This is not surprising, given that a key advantage claimed for the CDP models over DRC is that the CDP models are able to produce irregular-but-consistent nonword responses (Perry et al., 2007), and such responses will typically involve a different vowel to the regular response (e.g., pronouncing DOOK to rhyme with "book", instead of with the regular vowel sound as in "pool"). The experimental data show that when deviating from regularity, people also most often produce a *vowel difference* response. However, the participants produced this type of response for fewer nonwords than did any of the CDP models, and a produced a greater percentage of regular responses.

The CDP models produced a range of other types of irregular response, as did the experimental participants. The data in Table 2 indicate that the CDP models provide a distribution of response classifications that is in closer agreement with the participants than DRC, due to DRC's adherence to regularity. CDP++ in particular seems to parallel the participant distribution of responses across categories better than the other models. However, though they seem to produce broadly the same kind of responses as the participants, the CDP models do not match the participants on an item-by-item basis. We examined the performance of each model within category to ascertain whether some types of response were more problematic for the CDP models than others.

Table 3 displays the match between the types of responses and the experiment data for each of the CDP models. It is clear that one reason CDP.50 and CDP++ match the participant data better than CDP+ is because they sometimes produce regular responses. Both CDP.50 and CDP++ are in agreement with the participants for over 60% of nonwords pronounced regularly. For over 80% of items where each of these models gave a regular response, they matched the most frequent human response. CDP.50 or CDP++ mostly matched at least some participants for those items where they produced a regular response.

Responses from some categories seem more problematic for the CDP models than others. The *vowel difference* type responses were much less problematic than most other types of response. Almost one fifth of the vowel difference responses corresponded to the most frequent participant response for each of the CDP models. While still a low percentage, this is better than the average match to most frequent participant response across all categories. However, vowel difference responses also failed to match any participants on almost a quarter of the occasions when this type of response is made, for each of the models.

Table 3. CDP+, CDP+.50 and CDP++ Model Performance within Each Nonword Response Classification

Classification	% of responses	% match[a] Mean (*SD*)		% match most frequent[b]	% match none[c]
CDP+					
Regular	0.0				
Vowel different	36.7	17.3	(20.1)	19.9	23.8
Dropped phoneme	11.4	0.5	(1.2)	0.0	83.0
Extra phoneme	7.3	3.6	(13.9)	6.7	93.3
s/z coda difference	10.9	27.1	(21.7)	28.9	4.4
Consonant different	16.7	6.4	(17.5)	4.3	72.5
Other	17.0	3.3	(7.3)	2.9	67.1
All	100.0	11.2	(18.5)	12.1	49.0
CDP.50					
Regular	9.0	66.0	(28.3)	86.5	8.1
Vowel different	35.7	16.6	(19.3)	18.4	23.8
Dropped phoneme	5.1	0.4	(0.9)	0.0	81.0
Extra phoneme	5.1	5.1	(16.5)	9.5	90.5
s/z coda difference	11.9	27.0	(21.6)	28.6	4.1
Consonant different	20.1	7.8	(18.9)	6.0	66.3
Other	13.1	4.0	(8.3)	3.7	63.0
All	100.0	17.4	(25.3)	19.9	40.0
CDP++					
Regular	32.5	61.8	(27.5)	83.6	3.0
Vowel different	31.6	17.5	(21.6)	18.5	24.6
Dropped phoneme	2.4	0.7	(1.1)	0.0	70.0
Extra phoneme	2.9	8.9	(21.4)	16.7	83.3
s/z coda difference	9.2	27.7	(21.9)	26.3	2.6
Consonant different	9.2	10.9	(25.2)	13.2	63.2
Other	12.1	4.1	(8.4)	4.0	66.0
All	100.0	30.0	(32.3)	37.6	26.9

[a]mean % of participants matched by the model over all of the model responses in a class
[b]% of nonwords where the model matches the most frequent participant response
[c]% of nonwords where the model matches none of the participants

The *s/z coda difference* responses were also less problematic for the CDP models than other types of response. The CDP models rarely failed to match any participant when they

produce this kind of response. However, each of the CDP models produced this type of response more often than the participants.

CDP.50 produced less than half the number of *dropped phoneme* type of differences that CDP+ produced, supporting the original argument of Perry et al. (2007) that lowering the phoneme naming activation criterion for nonword-only reading would reduce missing phoneme errors. The benefit of reducing such errors is evident in Table 3, where it is clear that for each of the CDP model variations, dropped phoneme type responses never matched the most frequent response given by the experimental participants, and in most cases, failed to match any of the responses given by the participants. Even ignoring item-by-item matches and focusing instead on the general type of response, the experimental participants typically did not make dropped phoneme type responses (only 1.2% of experimental responses were classed as dropped phoneme).

The participants made *extra phoneme* type responses on some occasions (4.0% of responses), similarly to the CDP models. However, on an item-by-item level, the extra phoneme responses are highly problematic for the CDP models. Roughly 90% of the extra phoneme responses made by CDP+ and CDP.50, and over 80% of extra phoneme responses made by CDP++, failed to match any of the participants. So even though people sometimes seem to include an extra phoneme when pronouncing nonwords, it seems the CDP models are not including extra phonemes for the same nonwords and in the same way that people do.

Like the dropped phoneme and extra phoneme type responses, responses of the CDP models that were classed as *consonant difference* or else as *other* were typically not well matched to the experiment data. Such responses rarely matched a large percentage of participant responses, rarely matched the most frequent participant response, and often (approximately two-thirds of the time for both categories) failed to match any participant

response for each of the CDP model variations. Of particular note within the Consonant different category are nonwords that begin with the grapheme TH. Such nonwords comprise 9.7% of the 412 nonwords, suggesting that pronunciation of this grapheme in the initial position is a key source of difference between DRC and CDP+. DRC always uses the regular, unvoiced pronunciation (e.g., the same pronunciation as occurs in the word "thick") for this grapheme at the start of a stimulus. CDP+ uses the voiced TH pronunciation (e.g., the same as occurs in the word "then") for 50% of these nonwords, and drops the phoneme corresponding to TH altogether on another 20%. CDP.50 avoids dropping phonemes, but otherwise uses the voiced TH often, like CDP+. CDP++ is more likely than the CDP+ to use the regular, unvoiced TH, but still uses the voiced TH for 32.5% of the nonwords beginning with TH in the experiment dataset. In comparison, the experiment participants almost always used the regular, unvoiced pronunciation, similar to DRC. There were only six uses of the voiced TH out of 1,765 valid participant utterances for nonwords beginning with TH. This finding is in agreement with the results reported by Campbell and Besner (1981), who also found that people tend to use the unvoiced version of TH when reading aloud individual nonwords. As a result, the CDP models—particular CDP+—perform poorly on the nonwords that begin with TH, and typically do not match any participants when using the voiced TH.

CDP++ produced fewer irregularities than the other CDP models across all categories, and of its irregular responses, a greater proportion were classified as vowel difference (46.8% of irregular responses) than was the case for CDP+ or CDP+.50, where vowel difference responses made up 36.7% and 39.2% of irregular responses respectively. This suggests that CDP++ has been honed so that it produces less irregular responses to nonwords, and when it does deviate from regularity, it does so by producing more irregular-but-consistent (vowel difference) responses, while avoiding the more problematic dropped phoneme or extra

phoneme type of responses. Overall, CDP++[1] seemed to match the participant data better than either CDP+ or CDP.50, pointing to progress in the right direction for the CDP structure in general.

General Discussion

The purpose of this research was to use empirical nonword pronunciation data to assess the nonword pronunciation performance of DRC, CDP+ and CDP++. We focused on this benchmark because DRC and the CDP models differ mainly in the design of their non-lexical routes, so nonword reading highlights these differences. In addition, the nonword pronunciation performance of each of the models has not been adequately tested (i.e., compared with actual human responses) to date. None of the models accurately match the experimental data, although DRC was in agreement with participants significantly more often than any CDP model variation.

DRC produces no lexicalizations to nonword stimuli, unlike the experiment participants. DRC does not give word responses because its lexical route parameter settings minimize lexical route involvement in nonword processing. Lowering either L-to-O inhibition or P-to-P inhibition or both in the DRC model would enable it to produce lexical responses, however further analysis of the mechanisms by which people produce lexicalizations to nonwords is required to ensure that any parameter changes can be properly justified.

In contrast, the CDP models produce lexical responses far more often than the participants did. Analysis of the CDP+ model revealed that even though the CDP+ lexical

[1] The CDP++ model generated one isolated but noteworthy multisyllabic response: it output the word "breakfast" in response to the stimulus BREC. In the whole dataset of 1,475 nonwords, this was the only instance where CDP++ produced a disyllabic output to the monosyllabic stimulus. The nonword BREC was re-simulated with CDP++ after first setting phoneme-to-phonology excitation to 0. Doing so resulted in the output changing to "brek", indicating that it is involvement of the lexical route via interaction between the phonological lexicon and phoneme layers that results in the lexicalization to "breakfast". The DRC and CDP+ models do not produce multisyllabic output since they are monosyllable-only models.

route contributes partly to the high number of word responses generated by this model, it is the trained non-lexical route that is solely responsible for most of the lexicalizations. Any improvements made to CDP+ would therefore necessarily involve modifications to the non-lexical route and its training, rather than simple lexical route parameter changes. That is, the parts of CDP+ rendering it distinct from DRC are those that would need to be modified.

DRC only produces regular nonword responses, which is clearly unlike the experimental data. The CDP models produce irregular responses in roughly the way people do when considered broadly, but on an item-by-item basis, they have a poor match to the experimental data. Some types of irregular response seem more problematic than others, with dropped phoneme and extra phoneme type responses being especially problematic for the CDP models.

These results also suggest that the lenient criteria for measuring the accuracy of nonword pronunciations used by Perry et al. (2007) and others are not without shortcomings as a measure of nonword reading performance, since they include too many pronunciation possibilities that readers simply do not consider. It is clear that using regularity as the only benchmark of correctness as was done for DRC (Coltheart et al., 2001) is also inappropriate.

Improving the computational models

The nonword responses of participants suggest that people favour regular responses, but will sometimes produce irregular responses. However, the number of irregular responses is not as high as might be expected from the patterns in the accepted pronunciations of words. For DRC, changes must be motivated by the need for a greater number of irregular responses in order to better match participants, while for the CDP models, the number of lexicalizations and other irregular responses is too high.

For DRC, lexical route parameter changes may produce more lexicalizations, but it seems clear that non-lexical route changes will also be required. There are a variety of options for modifying the non-lexical route to allow for some percentage of irregular pronunciation, including: introducing different rule strengths for different GPC rules, an idea that was raised in Rastle and Coltheart (1999). This could be combined with the inclusion of multiple rules of varying strengths for individual graphemes that might correspond to more than one phoneme, depending on context. This may allow some rules to dominate other rules in particular circumstances, leading to irregular pronunciations. Another option is to include non-lexical route rules for larger orthographic units such as bodies, which might override grapheme-phoneme correspondences for relevant stimuli (e.g. pronounce OO as /ʊ/ when followed by a K, as in the body –OOK).

The CDP models experience a variety of problematic response types, which might require different modifications to avoid. For example, we have seen that dropped phoneme responses are unwanted, and can be reduced by changing the criteria for processing completion. Another change to the CDP model might be to modify the training regime, on the assumption that perhaps some problematic responses could be avoided with better model training. It is likely that at least some of the issues with CDP model nonword pronunciation arise from catastrophic interference (see McCloskey and Cohen (1989), as cited in French (1999)), where more recent learning erodes previous learning, a well known challenge for many types of connectionist learning model. Perhaps concurrently training CDP+ on GPCs and whole words, instead of sequentially training first on GPCs only, before then switching to training on words, may improve performance, particularly on words starting with TH, where it seems clear that the CDP+ nonlexical route does not retain information on using unvoiced TH after GPC training concludes and word training commences. The CDP models use a limited set of graphemes compared to DRC, and it might be that using an expanded set of

graphemes (with a corresponding increase in input nodes and network size) would allow the CDP models to produce better nonword responses. Lastly, it could also be the case that the two-layer associative network lacks the computational sophistication to adequately learn the pattern of regular and irregular pronunciations in English, and altering network architecture to enable it to perform more complex operations than a two-layer network might improve nonword naming.

The variety of responses provided to these nonwords point to the need for models that can account for this variety, as highlighted by Zevin and Seidenberg (2006). Many computational models of reading—including DRC, CDP+ and CDP++—have been employed to date as static models of an archetypal skilled reader, and thus only produce a single, ideal response to each nonword. However, there is no such thing as the average or ideal pronunciation of a nonword. If some nonwords are given multiple, different, non-outlier pronunciations by human readers, computational models need to include mechanisms that can account for this variety. The DRC model may be able to account for variety if researchers were to produce multiple model instances, each with a different array of parameter settings, and different sets of known GPCs. To avoid the appearance that these parameter adjustments are ad-hoc model fitting, DRC researchers could either demonstrate that the parameter changes reflect meaningful differences between the cognitive mechanisms of readers, or else investigate and evaluate the theoretical implications of making particular parameter adjustments. For example, if all participants pronounce GERT as /gɜːt/ while DRC pronounces this as /ʤɜːt/, DRC's context rule that an initial G followed by an E should be pronounced as /ʤ/ could be omitted, provided additional benchmark testing was done to see if DRC's overall pronunciation of words starting with GE were not adversely affected, and if response-time related benchmark effects were not negatively impacted.

The CDP models may be able to account for this variety if researchers produce multiple model instances that differ due to having undergone different training regimes (different words, or different orders of exposure). Such a process could also avoid the appearance of being ad-hoc if, for example, it were demonstrated that the variety of training regimes used are derived from plausible accounts of the different learning experiences of people. At present, it is not clear that the training regime used for even the currently published CDP models is more than loosely based on the learning experiences of children (Perry et al., 2007, 2010).

This research also adds to the discussion of the relative merits of statistical learning models such as CDP+ and CDP++ versus a hard-wired model such as DRC. The connectionist learning mechanism incorporated into the CDP models gives them additional sensitivity to the statistical variety in the English language over the DRC model. However, as is shown by the consideration of nonwords beginning with TH in our experiment, the sensitivity to this variety can mean that the model learns features of the language that people do not use when generalizing to novel stimuli. People almost never seem to use the voiced TH when reading aloud nonwords, despite the occurrence of the voiced TH being common in the English language and therefore a correspondence that the CDP models all learn. The generally poor item-by-item performance of each CDP model in matching empirical nonword pronunciation data also indicates that the learning undergone by these models seems unlike the way people learn to read aloud.

References

Andrews, S., & Scarratt, D. R. (1998). Rule and analogy mechanisms in reading nonwords: Hough dou peapel rede gnew wirds? *Journal of Experimental Psychology: Human Perception and Performance, 24*(4), 1052-1086.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX Lexical Database [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Campbell, R., & Besner, D. (1981). THIS and THAP--constraints on the pronunciation of new, written words. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 33A*(4), 375-396.

Cassidy, S., & Harrington, J. (2001). Multi-level annotation in the Emu speech database management system. *Speech Communication, 33*(1-2), 61-77. doi: 10.1016/s0167-6393(00)00069-8

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.

Dual-Route Cascaded Model 1.2.1. (2009) [Computer Software]. Sydney: Macquarie Centre for Cognitive Science, Macquarie University. Retrieved from http://www.maccs.mq.edu.au/~ssaunder/DRC/2009/10/drc-1-2-1/

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning & Verbal Behavior, 12*(6), 627-635.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers, 35*(1), 116-124.

French, R. M. (1999). Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences, 3*(4), 128-135. doi: http://dx.doi.org/10.1016/S1364-6613%2899%2901294-2

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5*(4), 674-691.

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, 111*(3), 662-720. doi: http://dx.doi.org/10.1037/0033-295X.111.3.662

Harrington, J., Cox, F., & Evans, Z. (1997). An acoustic phonetic study of broad, general, and cultivated Australian English vowels*. *Australian Journal of Linguistics, 17*(2), 155-184. doi: 10.1080/07268609708599550

Hinton, G. E. (1990). Mapping part-whole hierarchies into connectionist networks. *Artificial Intelligence, 46*(1-2), 47-75.

Jacobs, A. M., & Grainger, J. (1994). Models of Visual Word Recognition-Sampling the State of the Art. *Journal of Experimental Psychology: Human Perception and Performance, 20*(6), 1311-1334.

Marshall, J. C., & Newcombe, F. (1973). Patterns of paralexia: A psycholinguistic approach. *Journal of Psycholinguistic Research, 2*(3), 175-199. doi: http://dx.doi.org/10.1007/BF01067101

Masterson, J. (1985). On how we read nonwords: data from different populations. In K. E. Patterson, J. C. Marshall & M. Coltheart (Eds.), *Surface Dyslexia: Neuropsychological and Cognitive Studies of Phonological Reading* (pp. 289-299). London: Lawrence Erlbaum Associates Ltd.

McCann, R. S., & Besner, D. (1987). Reading pseudohomophones: Implications for models of pronunciation assembly and the locus of word-frequency effects in naming. *Journal of Experimental Psychology: Human Perception and Performance, 13*(1), 14-24. doi: http://dx.doi.org/10.1037/0096-1523.13.1.14

McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. *Psychology of learning and motivation - advances in research and theory, 24*(C), 109-165.

Norris, D. (2005). How do Computational Models Help us Develop Better Theories? In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 331-346). Mahwah, NJ: Lawrence Erlbaum Associates Publishers; US.

Patterson, K., & Shewell, C. (1987). Speak and spell: Dissociations and word-class effects. In M. Coltheart, G. Sartori & R. Job (Eds.), *The cognitive neuropsychology of language* (pp. 273-294). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc; England.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review, 114*(2), 273-315.

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the connectionist dual process (CDP++) model. *Cognitive Psychology, 61*(2), 106-151. doi: http://dx.doi.org/10.1016/j.cogpsych.2010.04.001

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.

Rastle, K., & Coltheart, M. (1999). Serial and strategic effects in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 25*(2), 482-503. doi: http://dx.doi.org/10.1037/0096-1523.25.2.482

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 4*(4), 1339-1362. doi: http://dx.doi.org/10.1080/02724980244000099

Seidenberg, M. S. (2005). Connectionist models of word reading. *Current Directions in Psychological Science, 14*(5), 238-242. doi: 10.1111/j.0963-7214.2005.00372.x

Seidenberg, M. S., & McClelland, J. L. (1989). A Distributed, Developmental Model of Word Recognition and Naming. *Psychological Review, 96*(4), 523-568.

Seidenberg, M. S., Plaut, D. C., Petersen, A. S., McClelland, J. L., & McRae, K. (1994). Nonword pronunciation and models of word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 20*(6), 1177-1196.

Treiman, R., Kessler, B., & Bick, S. (2003). Influence of consonantal context on the pronunciation of vowels: A comparison of human readers and computational models. *Cognition, 88*(1), 49-78. doi: http://dx.doi.org/10.1016/S0010-0277%2803%2900003-9

Zevin, J. D., & Seidenberg, M. S. (2006). Simulating consistency effects and individual differences in nonword naming: A comparison of current models. *Journal of Memory and Language, 54*(2), 145-160. doi: 10.1016/j.jml.2005.08.002

Appendix

Table A1. Phoneme symbols used in this study are those proposed for Australian English published in Harrington et al. (1997).

Vowels		Consonants	
Symbol	Example	Symbol	Example
iː	seen	p	pie
ɪ	bid	b	buy
e	red	t	tie
æ	cat	d	dot
ɐː	hard / palm	k	kite
ɐ	fun	g	guy
ɔ	pod	tʃ	chin
oː	board	dʒ	jump
ʊ	good	m	my
ʉː	clue	n	no
ɜː	bird	ŋ	hang
æɪ	stay	f	for
ɑe	sigh	v	vent
æɔ	mouth	θ	thin
əʉ	goat	ð	then
oɪ	boy	s	stop
ɪə	beard	z	zoo
eː	cared	ʃ	shoe
		ʒ	fusion
		h	hot
		r	rot
		j	you
		w	wet
		l	lie

Table A2. Model and Participant Responses to the 412 Nonwords Used in the Experiment

Note. Only the three most frequent participant responses to each nonword are included. The complete dataset of model responses to all 1,475 nonwords, and all participant responses can be found at: http://personal.maccs.mq.edu.au/~spritcha/111020NonwordReading.xls

Nonword	DRC	CDP+	CDP+.50	CDP++	Participant responses					
					1st	n	2nd	n	3rd	n
BAFF	bæf	bɐːf	bɐːf	bæf	bæf	44	bɐːf	1		
BASP	bæsp	bɐːsp	bɐːsp	bæsp	bæsp	35	bɐːsp	7	bælsp	1
BLAUCE	bloːs	bloː	bloːs	bloː	bloːs	14	blæɔs	7	blæɔtʃ	4
BLAUCHE	bloːʃ	bloː	bloː	bloːtʃ	bloːʃ	10	blæɔʃ	9	blæɔtʃ	7
BLEASE	bliːs	bliːz	bliːz	bliːz	bliːz	28	bliːs	16	bəliːz	1
BLIGN	blɪn	blɑend	blɑen	blɑen	blɑen	34	blɪŋ	2	bliːn	1
BLINE	blɑen	blɑend	blɑen	blɑen	blɑen	41	blɑe	1	bliːn	1
BLIRGE	blɜːdʒ	blɜː	blɜː	blɜːdʒ	blɜːdʒ	28	blɜːʒ	4	blɜːrdʒ	3
BLISE	blɑes	blɑez	blɑez	blɑes	blɑes	15	blɑez	12	bliːs	7
BLUISE	blʉːs	blʉːz	blʉːz	blʉːz	blʉːz	19	blʉːs	10	blʉːwiːz	3
BLYNCH	blɪntʃ	blɑentʃ	blɑentʃ	blɪntʃ	blɪntʃ	30	blɑentʃ	11	blɪŋθ	1
BOUCHE	bæɔʃ	bæɔtʃ	bæɔtʃ	bæɔtʃ	bʉːʃ	20	bæɔtʃ	7	bæɔʃ	4
BRASK	bræsk	brɐːsk	brɐːsk	brɐːsk	bræsk	28	brɐːsk	14	bræɪsk	2
BREC	brek	bre	brek	brekfəst	brek	43	briːk	1	bretʃ	1
BREKE	briːk	bræɪk	bræɪk	bræɪk	briːk	21	brek	12	bræɪk	5
BRETE	briːt	bræɪt	bret	bræɪt	briːt	20	bret	15	bræɪt	4
BROLK	brɔlk	brəʉlk	brəʉlk	brəʉlk	brɔlk	39	brəʉk	1	brəʉlɔk	1
BROR	broː	broːt	broː	broː	broː	32	broːr	8	brɔr	2
BRORE	broː	broːt	broː	broː	broː	26	broːr	12	brɜːr	2
BROS	brɔs	brɔz	brɔz	brɔz	brɔs	23	brəʉz	17	brɔz	2
BUKE	bjʉːk	bjʊk	bjʊk	bjʊk	bjʉːk	26	bʉːk	17		
BUNE	bjʉːn	bɐn	bɐn	bɐn	bʉːn	22	bjʉːn	19	bəʉn	1
CANC	kæŋk	kæn	kænk	kænk	kæŋk	30	kæns	3	kæn	2
CEB	seb	keb	keb	seb	seb	33	keb	10	kep	1
CEBB	seb	keb	keb	seb	seb	27	keb	15	tʃeb	2
CELK	selk	kek	kelk	selk	selk	23	kelk	15	tʃelk	3
CERM	sɜːm	kɜːm	kɜːm	sɜːm	sɜːm	25	kɜːm	13	krem	2
CES	ses	kez	kez	ses	ses	27	kes	7	sez	5
CESH	seʃ	keʃ	keʃ	seʃ	seʃ	20	keʃ	19	kleʃ	1
CHACH	tʃætʃ	tʃæk	tʃæk	tʃætʃ	tʃætʃ	19	kætʃ	4	tʃæk	3
CHIEL	tʃɑel	tʃiːl	tʃiːl	tʃiːl	tʃiːl	27	ʃiːl	5	tʃɪl	3
CHONGE	tʃɔndʒ	tʃɔnʒ	tʃɔnʒ	tʃəʉndʒ	tʃɔndʒ	18	kɔndʒ	5	tʃɔŋg	3
CHUILT	tʃʉːlt	tʃɪlt	tʃɪlt	tʃɪlt	tʃɪlt	9	tʃʊlt	9	tʃəlt	5
CHYNCH	tʃɪntʃ	tʃɑentʃ	tʃɑentʃ	tʃɪntʃ	tʃɪntʃ	18	sɪntʃ	5	ʃɪntʃ	4

Nonword	DRC	CDP+	CDP+.50	CDP++	Participant responses					
					1st	n	2nd	n	3rd	n
CICE	sɑes	kɑes	kɑes	sɑes	sɑes	22	siːs	8	kɑes	2
CID	sɪd	kɪd	kɪd	kɪd	sɪd	36	kɪd	5	sɪds	1
CILTH	sɪlθ	kɪlθ	kɪlθ	kɪlθ	sɪlθ	24	kɪlθ	9	klɪθ	5
CLALF	klælf	klɐːs	klɐːf	klɐː	klælf	16	klɐːf	9	klæf	6
CLALVE	klælv	klɐː	klɐːv	klɐː	klælv	21	klɐːlv	6	klɐːv	4
CLOME	kləʉm	kləʉmz	kləʉm	kləʉm	kləʉm	40	klɔm	2	kləʉ	1
CLUGUE	klʉːg	klɐg	klɐg	klɐg	klʉːg	15	klʉːʒ	7	klʉːʤ	6
COWTH	kæɔθ	kəʉθ	kəʉθ	kæɔθ	kæɔθ	39	kæɔ	2	kəʉθ	1
CRICHE	krɪʃ	krɑetʃ	krɑetʃ	krɑetʃ	kriːʃ	18	kriːtʃ	6	krɪʃ	3
CRUSQUE	krɐsk	krʉːsk	krɐsk	krɐsk	krʉːsk	23	krɐsk	11	krʉːskə	1
DANGE	dændʒ	dæɪnʒ	dænʒ	dæɪndʒ	dændʒ	20	dæɪndʒ	15	dæɪŋg	2
DAUCHE	doːʃ	doː	doː	doː	doːʃ	18	dæɔʃ	8	doːtʃ	6
DECHE	deʃ	de	de	diː	diːʃ	11	deʃ	8	diːtʃ	5
DONGE	dɔndʒ	dɔnʒ	dɔnʒ	dɔndʒ	dɔndʒ	36	dəʉndʒ	2	dɔŋg	2
DREVE	driːv	drev	drev	dræɪv	driːv	33	drev	6	driː	1
DRICHE	drɪʃ	drɑe	drɑetʃ	drɑetʃ	driːʃ	14	drɑeʃ	7	driːtʃ	6
DRIECE	driːs	drɑes	drɑes	drɑes	driːs	23	drɑes	7	driːʃ	3
DROSE	drəʉs	drəʉz	drəʉz	drəʉs	drəʉz	25	drəʉs	14	drɔs	1
DWAL	dwæl	dwoːl	dwoːl	dwoːl	dwoːl	13	dwæl	10	dwɐːl	10
DWALP	dwælp	dwoːlp	dwoːlp	dwoːlp	dwɔlp	16	dwælp	15	dwoːlp	7
DWARB	dwɐːb	dweːb	dweːb	dweːb	dwoːb	27	dwɐːb	16	dvɐːb	1
DWARN	dwɐːn	dwoːn	dwoːn	dwoːn	dwoːn	28	dwɐːn	13	dwoː	1
DWAS	dwæs	dwʉːz	dwʉːz	dwɔs	dwæs	21	dwɔz	10	dwɐs	3
DWAWSE	dwoːs	dwoːz	dwoːz	dwoːz	dwoːs	16	dwoːz	7	dwæɪz	5
DWEKE	dwiːk	dwæɪk	dwæɪk	dwæɪk	dwiːk	30	dwek	11	dwiː	1
DWI	dwɑe	dwɪ	dwɪ	dwɪ	dwɑe	23	dwiː	21		
DWOU	dwæɔ	dwʉː	dwʉː	dwʉː	dwæɔ	17	dwʉː	14	dwəʉ	6
DWOUSE	dwæɔs	dwʉːs	dwʉːs	dwʉːs	dwæɔs	22	dwæɔz	8	dwʉːs	6
DWUDD	dwɐd	dwʉːd	dwʉːd	dwʉːd	dwɐd	38	dwʉːd	3	dwʊd	2
DWUP	dwɐp	dwʉːp	dwʉːp	dwʉːp	dwɐp	35	dwʉːp	5	dwʊp	2
ELCH	eltʃ	weltʃ	weltʃ	eltʃ	eltʃ	33	elk	9	ɪltʃ	1
ENGE	endʒ	enʒ	enʒ	endʒ	endʒ	38	eŋg	3	ɪndʒiː	1
ENT	ent	went	ent	ent	ent	44				
FAC	fæk	fækt	fækt	fæk	fæk	41	fɐːk	2	fæ	1
FASP	fæsp	fɐːsp	fɐːsp	fæsp	fæsp	33	fɐːsp	11	fwɐːsp	1
FATH	fæθ	fɐːθ	fɐːθ	fɐːθ	fæθ	43	fæɪθ	1	fɐːθ	1
FENE	fiːn	fen	fen	fen	fiːn	27	fen	11	feniː	2
FLALSE	flæls	flæɪls	flæɪls	flæls	flɔls	15	flæls	12	flælz	3
FLES	fles	flez	flez	flez	fles	26	flez	12	fliːs	3

Nonword	DRC	CDP+	CDP+.50	CDP++	Participant responses					
					1st	n	2nd	n	3rd	n
FLOLL	flɔl	fləʉl	fləʉl	fləʉl	flɔl	43	floːl	1	fɔl	1
FLOUGHT	floːt	flæɔ	flæɔ	flɐt	floːt	21	flæɔt	19	flæɔ	1
FLOZ	flɔz	fləʉz	fləʉz	fləʉz	flɔz	42	fɔlz	1	fləʉs	1
FOLF	fɔlf	fəʉlf	fəʉlf	fɔlf	fɔlf	42	fləlf	1	fɔlɔf	1
FOOSH	fʉːʃ	fʊʃ	fʊʃ	fʊʃ	fʉːʃ	36	fʊʃ	7	fʉːs	1
FOZ	fɔz	fəʉz	fəʉz	fɔz	fɔz	45				
FRA	frɐː	fræm	fræ	fræ	frɐː	43	fræ	2		
FRAK	fræk	frækt	fræk	fræɪk	fræk	41	frɐːk	2	fræɪk	1
FRAUSE	froːs	froːz	froːz	froːz	froːz	15	fræɔs	9	fræɔz	8
FREAR	frɪə	frɪəm	frɪəm	frɪə	frɪə	26	frɪər	11	freː	3
FREBE	friːb	freb	freb	freb	friːb	30	freb	6	fræɪb	2
FRECH	fretʃ	frentʃ	frentʃ	fretʃ	fretʃ	27	freʃ	6	frek	5
FRIME	frɑem	frɔm	frɑem	frɑem	frɑem	34	frɪm	5	friːm	2
FROCHE	frɔʃ	frɔm	frɔm	frəʉ	frəʉʃ	17	frɔʃ	7	frəʉtʃ	7
FRONGE	frɔndʒ	frɔmʒ	frɔnʒ	frɔdʒ	frɔndʒ	36	frɐndʒ	3	frɔŋ	2
FROOR	froː	froːm	froːm	froː	froː	24	frʉːr	7	froːr	7
FROSE	frəʉs	frəʉz	frəʉz	frəʉz	frəʉz	31	frəʉs	9	frə	1
FRUGUE	frʉːg	frɐg	frɐg	frɐg	frʉːdʒ	15	frʉːg	11	frʉːʒ	7
FRUILT	frʉːlt	frɪlt	frɪlt	frɪlt	frɪlt	11	frʊlt	10	frʉːt	5
FRUR	frɜː	frɜːm	frɜː	frɜː	frɜː	15	frɜːr	9	frʉːr	7
FRYMPH	frɪmf	frɔmf	frɔmf	frɪmf	frɪmpf	28	frɪmp	6	frɑempf	5
GANC	gæŋk	gænd	gænd	gæn	gæŋk	36	gæns	2	gænəs	1
GEECH	dʒiːtʃ	giːtʃ	giːtʃ	giːtʃ	giːtʃ	33	giːʃ	7	gek	2
GERT	dʒɜːt	gɜːt	gɜːt	gɜːt	gɜːt	45				
GESK	dʒesk	esk	gesk	gesk	gesk	39	geks	1	dʒesk	1
GEVE	dʒiːv	giːv	giːv	giːv	giːv	22	dʒiːv	14	gev	3
GHELCH	geltʃ	eltʃ	geltʃ	geltʃ	geltʃ	37	gelk	2	geleʃ	1
GHEP	gep	ep	ep	ep	gep	36	gelp	2	kelp	2
GHESH	geʃ	eʃ	geʃ	geʃ	geʃ	30	giːʃ	6	gwiːʃ	1
GHETE	giːt	get	get	get	giːt	16	get	15	gæɪt	3
GHIGN	gɪn	gɑen	gɑen	gɑen	gɑen	16	gɪn	5	gɪndʒ	4
GHIMP	gɪmp	ɪmp	gɪmp	gɪmp	gɪmp	39	dʒɪmp	1	gɪm	1
GHOW	gæɔ	gəʉ	gəʉ	gəʉ	gæɔ	30	gəʉ	5	gləʉ	1
GHURR	gɜː	gɜːd	gɜːd	gɜːd	gɜː	27	gɜːr	14	grʉːr	1
GHUTE	gjʉːt	gʉːt	gjʉːt	gʉːt	gʉːt	30	gjʉːt	6	dʒʉːt	2
GINT	gɪnt	ɪnt	gɪnt	gɪnt	gɪnt	39	gɑent	2	dʒɪnt	2
GIS	gɪs	gɪz	gɪz	gɪs	gɪs	29	dʒɪs	8	gɪz	5
GLALF	glælf	glɐːl	glɐːl	glɐːlf	glælf	22	glæf	6	gælf	4
GLASK	glæsk	glɐːsk	glɐːsk	glɐːsk	glæsk	22	glɐːsk	19	gɐːsk	1

Nonword	DRC	CDP+	CDP+.50	CDP++	Participant responses					
					1st	n	2nd	n	3rd	n
GLAUK	glo:k	glɐ:k	glɐ:k	glo:k	glo:k	24	glæɔk	15	glɔk	2
GLIEF	glɑef	gli:f	gli:f	gli:f	gli:f	34	glɑef	3	glʉ:jef	1
GLUIT	glʉ:t	glɪt	glɪt	glɪt	glʉ:t	23	glʉ:wɪt	12	glʉ:wi:t	3
GNALPH	nælf	næl	næl	nælf	nælf	24	gnælf	11	nɔlf	2
GNANC	næŋk	nænk	nænk	nænk	næŋk	11	næk	9	gnæk	4
GNEUTH	nʉ:θ	njʉ:θ	njʉ:θ	njʉ:θ	nʉ:θ	14	njʉ:θ	7	gnʉ:θ	6
GNOMB	nɔm	nəʉm	nəʉm	nəʉmbl	nəʉm	18	nɔm	13	gnɔm	3
GNOOSH	nʉ:ʃ	nʊʃ	nʊʃ	nʊʃ	nʉ:ʃ	20	gnʉ:ʃ	10	nʊʃ	6
GNOSE	nəʉs	nəʉz	nəʉz	nəʉz	nəʉz	22	nʉ:s	6	nəʉs	4
GNUSE	njʉ:s	nʉ:s	nʉ:s	nʉ:s	nʉ:s	14	nʉ:z	9	gnʉ:z	6
GNYTH	nɪθ	nɑeθ	nɑeθ	nɪθ	nɑeθ	10	gɪnθ	7	nɪθ	6
GRACH	grætʃ	græ	græ	grætʃ	grætʃ	19	græʃ	7	græk	5
GRAUNT	gro:nt	grænts	gro:nts	gro:nt	gro:nt	36	græɔnt	5	grɔnt	1
GRELCH	greltʃ	grel	greltʃ	greltʃ	greltʃ	37	grelk	4	grelʃ	2
GROUGHT	gro:t	græɪt	gro:t	grɐt	græɔt	23	gro:t	14	grʉ:t	3
GROZ	grɔz	grəʉz	grəʉz	grəʉz	grɔz	39	grəʉz	4	grɐz	1
GRUKE	grʉ:k	grɐk	grɐk	græɪk	grʉ:k	44	grəʉk	1		
GWALF	gwælf	gwo:l	gwo:l	gwo:lf	gwælf	17	gwɔlf	17	gwo:f	3
GWANK	gwæŋk	gwo:ŋk	gwo:ŋk	gwæŋk	gwæŋk	37	gwɔŋk	3	gweŋk	2
GWARN	gwɐ:n	gwo:n	gwo:n	gwo:n	gwo:n	25	gwɐ:n	18	gwo:	1
GWENE	gwi:n	gwen	gwen	ʤwen	gwi:n	28	gwen	8	gwene	2
GWI	gwɑe	gwɪ	gwɪ	gwɪ	gwi:	35	gwɑe	8	ʤwi:	1
GWIEL	gwɑel	gwɪl	gwɪl	gwɪl	gwi:l	33	gwɑel	3	gwɪl	2
GYNCH	gɪntʃ	ʤɪntʃ	ʤɪntʃ	ʤɪntʃ	gɪntʃ	28	gɑentʃ	5	gɪn	2
HACE	hæɪs	hæs	hæs	hæɪs	hæɪs	38	hæs	2	hi:s	1
HALC	hælk	hæl	hæl	hɐ:l	hælk	33	hɔlk	4	hɐlk	4
HASE	hæɪs	hæz	hæɪz	hæɪs	hæɪz	26	hæɪs	14	hæs	2
HAUVE	ho:v	hæv	ho:v	ho:v	ho:v	24	hæɔv	7	həʉv	4
HESE	hi:s	hi:z	hi:z	hes	hi:s	15	hi:z	15	hes	4
HIECE	hi:s	hɑes	hɑes	hɑe	hi:s	20	hɑes	12	hɑek	3
HOLL	hɔl	həʉl	həʉl	həʉl	hɔl	42	həʉl	2		
JEICH	ʤæɪtʃ	ʤæɪ	ʤæɪk	ʤæɪtʃ	ʤi:ʃ	8	ʤi:tʃ	7	ʤæɪtʃ	6
JIEVE	ʤi:v	ʤɑev	ʤɑev	ʤi:v	ʤi:v	28	ʤɑev	6	ʤɪv	2
JIS	ʤɪs	ʤɪz	ʤɪz	ʤɪs	ʤɪs	33	ʤɪz	10	ʒɪs	1
KNOL	nɔl	nəʉl	nəʉl	nəʉl	nɔl	36	knɔl	5	nəʉl	3
KNOUCH	næɔtʃ	nəʉtʃ	nəʉtʃ	nɔtʃ	næɔtʃ	14	nʉ:ʃ	6	gnæɔtʃ	3
KUNGE	kɐnʤ	kɐnʒ	kɐnʒ	kjɐnʤ	kɐnʤ	25	kʉ:nʤ	8	kʊnʤ	3
LARCE	lɐ:s	lɐ:st	lɐ:st	lɐ:s	lɐ:s	39	lɐ:	2	lɐ:rs	1
LASP	læsp	lɐ:sp	lɐ:sp	lɐ:sp	læsp	35	lɐ:sp	9	lɐsp	1

Nonword	DRC	CDP+	CDP+.50	CDP++	Participant responses					
					1st	n	2nd	n	3rd	n
LOOTH	lʉːθ	lʉː	lʉː	lʉːð	lʉːθ	43	lʉːð	1	θʊlθ	1
MOLF	mɔlf	məʉlf	məʉlf	məʉlf	mɔlf	43	moːlf	1	mʊlf	1
NACH	nætʃ	næ	næk	nætʃ	nætʃ	19	næʃ	12	næ	2
NALK	nælk	noːkk	noːkk	noːkk	nælk	17	noːk	10	nɔlk	9
NENGE	nendʒ	nenʒ	nenʒ	nendʒ	nendʒ	37	neŋ	3	neŋg	1
NIS	nɪs	nɪz	nɪz	nɪz	nɪs	41	nɪz	3	nʉːs	1
NOF	nɔf	nɔt	nɔt	nɔf	nɔf	43	nəʉf	1	nʊf	1
NOOSH	nʉːʃ	nʊʃ	nʊʃ	nʉːʃ	nʉːʃ	37	nʊʃ	7	nɜːʃ	1
NUNE	njʉːn	njɐn	njɐn	njʉːn	nʉːn	30	njʉːn	12	nʉː	1
NYTH	nɪθ	nɑeθ	nɑeθ	nɑeθ	nɪθ	19	nɑeθ	15	niːθ	6
OL	ɔl	əʉl	əʉl	əʉl	ɔl	45				
OOSH	ʉːʃ	ʊʃ	ʊʃ	ʊʃ	ʉːʃ	31	ʊʃ	13	ʉːtʃ	1
PHEASE	fiːs	fiːz	fiːz	fiːz	fiːz	38	fiːs	3	piːz	1
PHLAUCE	floːs	floː	floː	floː	floːs	13	flæɔs	12	flæɔk	2
PHLERSE	flɜːs	flɜːz	flɜːz	flɜːz	flɜːs	29	flɜːz	2	fɪləreseː	1
PHLEUCE	flʉːs	flʉː	flʉːs	flʉːs	flʉːs	24	plʉːs	3	floːs	1
PHLOLT	flɔlt	fləʉlt	fləʉlt	fləʉlt	flɔlt	29	fɔlt	6	flɔt	3
PHLOMB	flɔm	fləʉm	fləʉm	fləʉm	flɔm	23	fləʉm	8	flɔmb	5
PHLOSE	fləʉs	fləʉz	fləʉz	fləʉz	fləʉz	23	fləʉs	12	flʉːs	2
PHLOTH	flɔθ	fləʉθ	fləʉθ	fləʉθ	flɔθ	27	plɔθ	5	fləʉθ	3
PHOIN	foɪn	foɪnd	foɪnd	foɪn	foɪn	35	fəʉn	3	fɑejɔn	1
PHOL	fɔl	fəʉl	fəʉl	fəʉl	fɔl	42	fəʉl	1	pɔl	1
PHOLK	fɔlk	fəʉk	fəʉkk	fəʉlk	fɔlk	38	fəʉk	3	pɔlk	2
PHOMP	fɔmp	frɔmp	fɔmp	fɔmp	fɔmp	38	pəmp	1	fɔm	1
PHONK	fɔŋk	fəʉŋk	fəʉŋk	fəʉŋk	fɔŋk	41	pɔŋk	2	fɔŋ	1
PHOZ	fɔz	fəʉz	fəʉz	fəʉz	fɔz	40	fəʉz	2	fʉːz	2
PHRALPH	frælf	fræɪl	fræɪl	frælf	frælf	30	frɐːlf	2	rælf	2
PHROOK	frʉːk	frʊk	frʊk	frʊk	frʉːk	22	frʊk	18	prʊk	2
PHUGE	fjʉːdʒ	fjɐdʒ	fjɐdʒ	fjɐdʒ	fʉːdʒ	21	fʉːʒ	8	fjʉːdʒ	6
PHUISE	fʉːs	fʉːz	fʉːz	fʉːz	fjʉːz	10	fʉːz	8	fʉːs	5
PLALL	plæl	ploːl	ploːl	ploːl	plæl	16	ploːl	12	flæl	8
PLANGE	plændʒ	plæɪnʒ	plænʒ	plæɪndʒ	plændʒ	23	plæɪndʒ	6	flændʒ	5
PLAUCHE	ploːʃ	ploː	ploːtʃ	ploːtʃ	ploːʃ	11	plæɔʃ	11	plæɔtʃ	4
PLENGE	plendʒ	plenʒ	plenʒ	plendʒ	plendʒ	36	flendʒ	5	plɪndʒ	1
PLU	plʉː	plɐs	plɐ	plɐ	plʉː	42	flʉː	2	plɐ	1
PRAUGH	proː	prɐːf	prɐːf	proː	proː	10	præɔ	8	prɐːg	6
PREACE	priːs	priːsts	priːs	priːs	priːs	31	priːtʃ	3	priːæɪs	2
PREBE	priːb	preb	preb	preb	priːb	29	præɪb	4	preb	4
PRUDD	prɐd	prʉːd	prʉːd	prʉːd	prɐd	34	prʉːd	8	prəd	2

Nonword	DRC	CDP+	CDP+.50	CDP++	Participant responses					
					1st	n	2nd	n	3rd	n
PSAISE	sæɪs	pæɪz	pæɪz	pæɪz	sæɪz	15	psæɪz	10	sæɪs	7
PSAMB	sæm	pæm	pæm	pæmb	sæm	16	sæmb	7	sɐ:m	7
PSAR	sɐ:	pɐ:	pɐ:	pɐ:	sɐ:	27	psɐ:	10	pəsɐ:	3
PSAUGE	so:ʤ	po:ʤ	po:ʤ	po:ʤ	so:ʤ	13	so:ʒ	5	sæɔʤ	4
PSAUNCH	so:nʧ	po:nʧ	po:nʧ	po:nʧ	so:nʧ	24	pso:nʧ	8	sæɔnʧ	2
PSAWP	so:p	po:p	po:p	po:p	so:p	14	swɔp	7	pso:p	4
PSEEF	si:f	pi:f	pi:f	pi:f	si:f	26	psi:f	8	sef	4
PSELSE	sels	pels	pels	sels	sels	17	psels	4	selz	2
PSEN	sen	pen	pen	pen	sen	25	psen	14	pəsen	1
PSEUCE	sʉ:s	pjʉ:s	pjʉ:s	pjʉ:s	sʉ:s	15	sjʉ:s	5	pəsʉ:s	2
PSICH	sɪʧ	pɪʧ	pɪʧ	pɪʧ	sɪʧ	15	si:tʃ	3	sɑek	3
PSIRP	sɜ:p	pɜ:p	pɜ:p	sɜ:p	sɜ:p	20	psɜ:p	7	psɜ:rp	3
PSIZ	sɪz	pɪz	pɪz	pɑez	sɪz	19	psɪz	10	fɪz	2
PSOATH	səʉθ	pəʉθ	pəʉθ	pəʉθ	səʉθ	25	psəʉθ	6	sæɔθ	3
PSONGE	sɔnʤ	pɔnʒ	pɔnʒ	pɔnʤ	sɔnʤ	19	psɔnʤ	6	sɔn	3
PSOOSH	sʉ:ʃ	pʊʃ	pʊʃ	pʊʃ	sʉ:ʃ	20	psʉ:ʃ	4	psʊʃ	4
PSOOTH	sʉ:θ	pʉ:	pʉ:ð	pʉ:θ	sʉ:θ	22	psʉ:θ	8	sʉ:ð	5
PSORB	so:b	po:b	po:b	po:b	so:b	32	pso:b	11	pso:rb	1
PUISE	pʉ:s	pʉ:z	pʉ:z	pʉ:z	pjʉ:z	12	pʉ:z	7	pjʉ:s	6
PUSQUE	pɐsk	pʉ:sk	pʉ:sk	pʉ:sk	pʉ:sk	10	pɐsk	9	pjʉ:sk	4
QUE	kwi:	kwe	kwe	kwe	kjʉ:	36	ke	4	kwe	3
RALL	ræl	ro:l	ro:l	ro:l	ro:l	20	ræl	19	rɐ:l	4
RENGE	renʤ	renʒ	renʒ	renʤ	renʤ	39	rɔnʤ	1	ri:nʤ	1
RHAFF	ræf	rɐ:f	rɐ:f	rɐ:f	ræf	38	hræf	3	rɐ:f	3
RHAWSE	ro:s	ro:z	ro:z	ro:z	ro:s	16	ro:z	10	ræɔs	5
RHETE	ri:t	ræɪt	ræɪt	ræɪt	ri:t	22	ret	9	ræɪt	2
RHINGE	rɪnʤ	rɪnʒ	rɪnʒ	rɪnʤ	rɪnʤ	36	rɪŋ	1	rɪn	1
RHOUSE	ræɔs	ræɔz	ræɔz	rʉ:z	ræɔs	19	rʉ:s	10	ræɔz	5
RHUKE	rjʉ:k	rɐk	rɐk	rʉ:k	rʉ:k	39	hrʉ:k	5	rɔk	1
RINGE	rɪnʤ	rɪnʒ	rɪnʒ	rɪnʤ	rɪnʤ	35	rɪŋ	3	rɑenʤ	2
ROLT	rɔlt	rəʉlt	rəʉlt	rəʉlt	rɔlt	41	wɔlt	1	rəlt	1
ROUCHE	ræɔʃ	ræɔtʃ	ræɔtʃ	rəʉtʃ	rʉ:ʃ	22	ræɔtʃ	5	ræɔʃ	4
ROWSE	ræɔs	rəʉz	rəʉz	rəʉz	ræɔz	19	ræɔs	18	rəʉz	5
RUILT	rʉ:lt	rɪlt	rɪlt	rɪlt	rʊlt	12	rɪlt	11	rʉ:lt	6
RURSE	rɜ:s	rɜ:z	rɜ:z	rɜ:s	rɜ:s	18	rɜ:rs	4	rʉ:rs	4
SALM	sælm	sɐ:m	sɐ:m	sɐ:m	sælm	23	sɐ:m	14	sɐ:lm	2
SARR	sɐ:	sɐ:d	sɐ:d	sɐ:d	sɐ:	37	sɐ:r	7		
SCAQUE	skæɪk	skæk	skæk	skæɪk	skɐ:k	10	skæk	9	skæɪk	4
SCILTH	sɪlθ	skɪlθ	skɪlθ	skɪlθ	skɪlθ	28	sɪlθ	8	sɪlk	1

Nonword	DRC	CDP+	CDP+.50	CDP++	Participant responses					
					1st	n	2nd	n	3rd	n
SCINE	sɑen	skɑen	skɑen	skɑen	skɑen	18	sɑen	12	siːn	6
SCRALK	skrælk	skræk	skrækk	skrækk	skrælk	14	skrɔlk	14	skroːlk	8
SCRALL	skræl	skroːl	skroːl	skroːl	skroːl	22	skræl	10	skrɐːl	7
SCRIPE	skrɑep	skrɑept	skrɑept	skrɑep	skrɑep	36	skrɑeb	3	skrep	2
SCROLK	skrɔlk	skrəʉl	skrəʉlk	skrəʉlk	skrɔlk	40	skɔlk	1	skrɔlɔk	1
SCROME	skrəʉm	skriːm	skrəʉm	skrəʉm	skrəʉm	38	skrɔm	2	skrə	1
SCROSE	skrəʉs	skrəʉz	skrəʉz	skrəʉz	skrəʉz	22	skrəʉs	14	skrɔs	2
SCRUKE	skruːk	skrɐk	skrɐk	skrɐk	skruːk	39	ʃruːk	2	skrɐŋk	1
SCRYM	skrɪm	skrɑem	skrɑem	skrɑem	skrɪm	24	skrɑem	16	skriːm	2
SCUTE	skjʉːt	skɐt	skɐt	skʉːt	skʉːt	27	skjʉːt	15	ʃʉːt	1
SHALSE	ʃæls	ʃæɪls	ʃæɪls	ʃæls	ʃæls	21	ʃɔls	8	ʃælz	6
SHECHE	ʃeʃ	ʃiː	ʃiː	ʃiː	ʃeʃ	6	ʃiːʃ	6	ʃiːtʃ	4
SHESE	ʃiːs	ʃiːz	ʃiːz	ʃiːs	ʃiːs	17	ʃiːz	11	ʃes	5
SHINC	ʃɪŋk	ʃɪnk	ʃɪnk	ʃɪnk	ʃɪŋk	35	ʃɪntʃ	3	sɪntʃ	2
SHIS	ʃɪs	ʃɪz	ʃɪz	ʃɪz	ʃɪs	21	ʃɪz	14	ʃɪʃ	4
SHOS	ʃɔs	ʃɔz	ʃɔz	ʃɔz	ʃɔs	25	ʃɔz	7	ʃəʉz	5
SHOULE	ʃæɔl	ʃʉːl	ʃʉːl	ʃəʉl	ʃʉːl	16	ʃæɔl	5	ʃoːl	5
SHOWTH	ʃæɔθ	ʃəʉθ	ʃəʉθ	ʃəʉθ	ʃəʉθ	21	ʃæɔθ	14	ʃɔθ	3
SHRAQUE	ʃræɪk	ʃrɐːk	ʃrɐːk	ʃræɪk	ʃræk	17	ʃrɐːk	8	ʃræɪk	2
SHRAS	ʃræs	ʃrɐːz	ʃrɐːz	ʃræz	ʃræs	18	ʃræz	18	ʃræ	1
SHRAUK	ʃroːk	ʃræk	ʃræk	ʃroːk	ʃroːk	21	ʃræɔk	11	ʃrɐŋk	1
SHRIC	ʃrɪk	ʃrɪ	ʃrɪ	ʃrɪk	ʃrɪk	38	ʃriːk	3	ʃɜːk	2
SHRIRR	ʃrɜː	ʃrɜːd	ʃrɜːd	ʃrɜːd	ʃrɪər	18	ʃrɪə	6	ʃrɜːr	5
SHRUICE	ʃruːs	ʃruː	ʃruː	ʃruː	ʃruːs	22	ʃruːwɪs	4	skruːs	2
SHRUKE	ʃruːk	ʃrɐk	ʃrɐk	ʃræɪk	ʃruːk	40	ʃruː	2	ʃrɐk	1
SHRUNGE	ʃrɐndʒ	ʃrɐnʒ	ʃrɐnʒ	ʃrɐndʒ	ʃrɐndʒ	21	ʃruːndʒ	10	srɐndʒ	3
SHUGE	ʃjʉːdʒ	ʃɐdʒ	ʃɐdʒ	ʃɐdʒ	ʃʉːdʒ	15	ʃʉːg	8	ʃʉːʒ	6
SILGE	sɪldʒ	stɪldʒ	stɪldʒ	sɪldʒ	sɪldʒ	35	sɪlg	3	sɪlʒ	3
SILN	sɪln	stɪl	stɪln	sɪln	sɪln	35	slɪn	3	sɪl	2
SKALC	skælk	skæl	skæl	skɐːl	skælk	29	skɐlk	4	skɔlk	3
SKARCE	skɐːs	skɐː	skɐːs	skɐː	skɐːs	33	skɐːk	4	skɐːrs	2
SKECHE	skeʃ	skiː	skiː	skiː	sketʃ	12	skiːtʃ	9	skeʃ	7
SKUBE	skjʉːb	skɐb	skɐb	skɐb	skʉːb	32	skjʉːb	10	skɐb	1
SLARR	slɐː	slɐːd	slɐːd	slɐːd	slɐː	35	slɐːr	8	splɐː	1
SLULE	slʉːl	slɐl	slɐl	slɐl	slʉːl	25	slʉː	4	slɐl	3
SLUS	slɐs	slɐz	slɐz	slɐz	slɐs	27	slʉːs	8	slɐz	3
SLYS	slɪs	slɑez	slɑez	slɑez	slɑes	12	slɪs	12	slɑez	12
SMEKE	smiːk	smæɪk	smæɪk	smæɪk	smiːk	26	smek	16	smɪk	1
SMEPE	smiːp	smep	smep	smep	smiːp	23	smep	10	smæɪp	4

Nonword	DRC	CDP+	CDP+.50	CDP++	Participant responses					
					1st	n	2nd	n	3rd	n
SMOCHE	smɔʃ	smənʉ	smənʉ	smənʉ	smʉːtʃ	15	smənʉʃ	8	smɔtʃ	6
SMOUTH	smæɔθ	smæɔ	smæɔ	smæɔθ	smæɔθ	34	smʉːθ	5	smæɔð	3
SMYNC	smɪŋk	smɑek	smɑenk	smɪŋk	smɪŋk	30	smɑeŋk	7	smɪntʃ	3
SMYNCH	smɪntʃ	smɑentʃ	smɑentʃ	smɑentʃ	smɪntʃ	33	smɑentʃ	5	smɪŋk	2
SMYS	smɪs	smɑez	smɑez	smɑez	smɑes	12	smɪs	11	smɑez	10
SNENGE	snendʒ	sniːnʒ	snenʒ	snendʒ	snendʒ	31	snedʒ	3	snɪndʒ	1
SNESE	sniːs	sniːz	sniːz	sniːs	sniːz	21	snes	7	sniːs	5
SNICHE	snɪʃ	snɑetʃ	snɪtʃ	snɑetʃ	sniːʃ	11	snɪtʃ	8	snɑetʃ	4
SNONGE	snɔndʒ	snɔnʒ	snɔnʒ	snəʉn	snɔndʒ	32	snəʉndʒ	3	snɐndʒ	1
SNOWL	snæɔl	snəʉl	snəʉl	snəʉl	snæɔl	25	snəʉl	13	snɔl	5
SOOSE	sʉːs	sʉː	sʉːz	sʉːs	sʉːs	29	sʉːz	12	sʉːʒ	1
SPEINT	spæɪnt	spent	spent	spent	spɑent	15	spiːnt	12	spæɪnt	5
SPEVE	spiːv	spev	spev	spev	spiːv	35	spev	7	spɪ	1
SPEWTH	spjʉːθ	spʉːθ	spjʉːθ	spjʉːθ	spjʉːθ	23	spʉːθ	9	splʉːθ	2
SPLACH	splætʃ	splæ	splæ	splætʃ	splætʃ	26	splæʃ	6	splæk	6
SPLALM	splælm	splɐːm	splɐːm	splɐːm	splɐːm	12	splælm	8	splæm	7
SPLANC	splæŋk	splænk	splænk	splænk	splæŋk	34	splɐːŋk	4	splɔŋk	2
SPLEASE	spliːs	spliːz	spliːz	spliːz	spliːz	36	spliːs	6	spiːz	1
SPLICHE	splɪʃ	splɑetʃ	splɪtʃ	splɑetʃ	spliːʃ	13	splɪtʃ	7	spliːtʃ	6
SPLOL	splɔl	spləʉl	spləʉl	spləʉl	splɔl	43	slɔl	1		
SPLOURT	sploːt	ploːts	ploːt	ploːt	ploːt	18	splæɔrt	4	splæɔt	4
SPLOWSE	splæɔs	spləʉz	spləʉz	splæɔz	splæɔs	26	splæɔz	6	spləʉs	6
SPOLK	spɔlk	spəʉl	spəʉlk	spəʉlk	spɔlk	43	spəʉk	1	spʊlk	1
SPRA	sprɐː	spræ	spræ	spræ	sprɐː	44				
SPRARR	sprɐː	sprɐːd	sprɐːd	sprɐːd	sprɐː	23	sprɐːr	15	spɐːr	4
SPRARSE	sprɐːs	sprɐːz	sprɐːz	sprɐːz	sprɐːs	26	spɐːs	4	spɐːrs	2
SPRAUK	proːk	sprɐːk	sprɐːk	proːk	proːk	24	spræɔk	11	sprɐːk	3
SPREN	spren	sprend	spren	spren	spren	42	spriːn	3		
SPRURSE	sprɜːs	sprɜː	sprɜːs	sprɜːs	sprɜːs	16	sprɜːz	4	sprʉːs	3
SPUBE	spjʉːb	spʉːb	spɐb	spʉːb	spʉːb	24	spjʉːb	14	spʉː	1
STAISE	stæɪs	stæɪz	stæɪz	stæɪz	stæɪz	24	stæɪs	12	stræɪs	3
STAITCH	stæɪtʃ	stæɪts	stæɪts	stæɪtʃ	stæɪtʃ	29	stræɪtʃ	4	stætʃ	2
STAUSE	stoːs	stoːz	stoːz	stoːz	stoːs	14	stæɔs	13	stoːz	7
STILN	stɪln	stɪl	stɪln	stɪln	stɪln	35	stɪlən	5	stɪl	2
STOARSE	stoːs	stoːz	stoːz	stoːz	stoːs	29	stroːs	3	stoːwɐːs	2
STOLK	stɔlk	stəʉk	stəʉk	stəʉl	stɔlk	41	stoːlk	2	stəʉk	1
STRASE	stræɪs	stræɪz	stræɪz	stræɪz	stræɪs	16	stræɪz	10	strɐːs	7
STRATH	stræθ	strɐːθ	strɐːθ	strɐːθ	stræθ	37	strɐːθ	4	stræ	1
STREPE	striːp	strep	strep	strep	striːp	25	strep	10	strɪp	2

Nonword	DRC	CDP+	CDP+.50	CDP++	Participant responses					
					1st	n	2nd	n	3rd	n
STRIQUE	strɪk	strɑeks	strɑeks	strɑek	striːk	21	strɪk	7	strɑek	6
STRONGE	strɔnʤ	strɔŋʒ	strɔŋʒ	strəʉn	strɔnʤ	20	strəʉnʤ	5	strɔŋ	5
STROOK	strʉːk	strʊk	strʊk	strʊk	strʉːk	32	strʊk	12		
STROW	stræɔ	strəʉ	strəʉ	strəʉ	strəʉ	33	stræɔ	10	ʃtrəʉ	1
STUKE	stjʉːk	stɐk	stɐk	stɐk	stʉːk	28	stjʉːk	10	strʉːk	4
SUILE	sʉːl	skʉːl	skʉːl	sʉːl	sʉːl	13	sjʉːl	8	swiːl	8
SWATT	swæt	swɔt	swɔt	swɔt	swɔt	36	swæt	7	swɐt	1
SWIC	swɪk	swɪs	swɪs	swɪk	swɪk	39	swɑek	1	zwɪk	1
SWIEL	swɑel	swiːl	swiːl	swoːl	swiːl	35	swæɪl	2	swɑel	1
SWOUNGE	swæɔnʤ	swæɔnʒ	swæɔnʒ	swæɔnʤ	swæɔnʤ	28	swʉːnʤ	7	swɐnʤ	2
SWUS	swɐs	swɐ	swɐz	swʉːz	swɐs	27	swʉːs	6	swʊs	5
THAC	θæk	ðæk	ðæk	θæk	θæk	40	θwæk	2	ðæk	1
THAG	θæg	ðæg	ðæg	θæg	θæg	41	θɐːg	2	ðæg	1
THAK	θæk	ðæk	ðæk	θæk	θæk	38	θɐːk	2	θælk	1
THALC	θælk	ðæl	ðæl	ðoːl	θælk	35	θoːlk	2	θɔlk	2
THANCH	θæntʃ	ðæntʃ	ðæntʃ	ðæntʃ	θæntʃ	28	θæŋk	5	θoːntʃ	2
THAQUE	θæɪk	ðæk	ðæk	θæɪk	θæk	18	θɐːk	4	θoːk	4
THECHE	θeʃ	ðiː	ðiː	ðiː	θeʃ	9	θetʃ	7	θiːʃ	6
THEDGE	θeʤ	ðeʤ	ðeʤ	ðeʤ	θeʤ	44	θeʤə	1		
THEEL	θiːl	ðiːl	ðiːl	θiːl	θiːl	40	θwiːl	1	tiːl	1
THEIL	θæɪl	ðel	ðel	θel	θiːl	34	θel	2	θɪl	2
THEL	θel	ðel	ðel	ðel	θel	40	θiːl	3	θwel	1
THELK	θelk	elk	θelk	θelk	θelk	42	θælk	1	θwelk	1
THELM	θelm	ðelm	ðelm	ðelm	θelm	44	θwelm	1		
THERP	θɜːp	ðɜːp	ðɜːp	ðɜːp	θɜːp	38	θwɜːp	2	θrep	2
THESK	θesk	esk	ðesk	ðesk	θesk	42	θiːsk	2	desk	1
THESS	θes	ðes	ðes	ðes	θes	36	θiːs	5	θres	1
THET	θet	ðet	ðet	ðet	θet	41	θiːt	3	θe	1
THETCH	θetʃ	ðetʃ	ðetʃ	ðetʃ	θetʃ	40	θe	1	θretʃ	1
THINGE	θɪnʤ	θɪnʒ	θɪnʒ	θɪnʤ	θɪnʤ	35	θɪŋg	3	θɪŋ	2
THITE	θɑet	ɑet	θɑet	θɑet	θɑet	35	θiːt	6	θiːθ	1
THODD	θɔd	ðɔd	ðɔd	θɔd	θɔd	43	θrɔd	1	θɔ	1
THOLVE	θɔlv	ðɔlv	ðɔlv	θɔlv	θɔlv	37	θwɔlv	4	θʊlv	1
THRALC	θrælk	θræl	θræl	θroːl	θrælk	29	θrɔlk	4	θrelk	1
THRANC	θræŋk	θrænk	θrænk	θræŋk	θræŋk	36	θrɔŋk	2	θwæŋk	1
THREAR	θrɪə	θreː	θreː	θrɪə	θrɪər	17	θrɪə	16	θreː	7
THROUSE	θræɔs	θrʉːz	θrʉːz	θrʉːs	θræɔs	26	θræɔz	9	θrʉːs	5
THRUME	θrʉːm	θrɐm	θrɐm	θrʉːm	θrʉːm	38	θrʉːmb	1	θrɔm	1
THUBE	θjʉːb	θjɐb	θjɐb	θjʉːb	θʉːb	35	θjʉːb	5	tjʉːb	1

Nonword	DRC	CDP+	CDP+.50	CDP++	Participant responses					
					1st	n	2nd	n	3rd	n
THUPE	θjʉːp	θʉːp	θjʉːp	θʉːp	θʉːp	36	θjʉːp	4	θɜːp	1
THUSE	θjʉːs	ðʉːs	ðjʉːs	ðʉːs	θʉːz	25	θʉːs	11	θjʉːz	5
THWALC	θwælk	woːl	θwoːl	θwoːl	θwælk	22	θwɔlk	15	θwoːk	2
THWAZZ	θwæz	θwoːz	θwoːz	θwoːz	θwæz	21	θwɔz	8	θwæ	2
THWEB	θweb	web	θweb	θweb	θweb	41	θwiːb	1	θwe	1
THWELVE	θwelv	welv	welv	θwelv	θwelv	29	twelv	8	θwelf	3
THWINGE	θwɪnʤ	θwɪnʒ	θwɪnʒ	θwɪnʤ	θwɪnʤ	34	twɪnʤ	2	twɪns	1
THWOLVE	θwɔlv	wɔlv	θwɔlv	θwɔlv	θwɔlv	28	θwʊlv	6	twɔlv	3
THWONCH	θwɔntʃ	wɔntʃ	θwɔntʃ	θwɔntʃ	θwɔntʃ	34	twɔŋk	2	θwɐnɐtʃ	1
THWOS	θwɔs	ðwɔz	ðwɔz	ðwɔz	θwɔs	26	θwɔz	5	θwəʉs	4
THWOWN	θwæɔn	θwɔn	θwɔn	θwæɔn	θwəʉn	21	θwæɔn	13	θwɔn	3
THWUILT	θwʉːlt	θwɪlt	θwɪlt	θwʉːlt	θwɪlt	18	θwɔlt	6	θwʊlt	5
TOWSE	tæɔs	təʉz	təʉz	təʉz	tæɔs	16	tæɔz	13	təʉs	3
TRASS	træs	trɐːs	trɐːs	trɐːs	træs	32	trɐːs	7	træz	2
TREESE	triːs	triːz	triːz	triːz	triːs	24	triːz	15	θriːz	2
TROW	træɔ	trəʉ	trəʉ	trəʉ	trəʉ	18	træɔ	18	θrəʉ	4
TRURE	troː	trʉː	trʉː	trʉː	trʉːr	14	trʉː	5	trɜːr	4
TRURSE	trɜːs	trɜːz	trɜːz	trɜːs	trɜːs	18	trʉːs	6	trʉːz	4
TUISE	tʉːs	tʉːz	tʉːz	tjʉːz	tʉːz	7	twiːs	5	twiːz	5
TUME	tjʉːm	tjɐm	tjɐm	tjɐm	tjʉːm	22	tʉːm	18	tɐm	2
TWALPH	twælf	twæɪl	twæɪl	twoːlf	twælf	20	twɔlf	10	θwælf	5
TWARK	twɐːk	twoːk	twoːk	tweːk	twoːk	18	twɐːk	15	θwoːk	6
TWERE	twɪə	tweː	tweː	tweː	twɜː	20	θwɜː	6	tweː	4
TWOLE	twəʉl	twʉːl	twʉːl	twəʉl	twɔl	35	θwɔl	3	twəʉl	2
TWORE	twoː	toː	toː	toː	twoː	31	θwoː	6	twoːr	4
TWOWN	twæɔn	twʉːn	twʉːn	twæɔn	twəʉn	11	twæɔn	11	twɔn	5
TWUG	twɐg	twʉːg	twʉːg	twɐg	twɐg	33	θwɐg	6	twʊg	2
TWUSQUE	twɐsk	twʉːsk	twʉːsk	twʉːs	twʉːsk	10	twɐsk	9	θwɐsk	3
VACHE	væʃ	væɪ	væɪ	væɪ	væʃ	10	væɪʃ	8	vætʃ	6
VANC	vænk	vænd	vænk	vænk	vænk	33	vɔŋk	2	vɐŋk	2
VAS	væs	væz	væz	væz	væs	36	vɐːs	4	væz	3
VEESE	viːs	viːz	viːz	viːz	viːs	27	viːz	12	viːsiː	2
VIGN	vɪn	vɑen	vɑen	vɑen	vɑen	17	vɪgən	5	vɪnʤ	4
WAICE	wæɪs	wæɪ	wæɪs	wæɪ	wæɪs	31	wɑes	4	wæɪk	3
WALC	wælk	wel	wel	woːl	wɔlk	18	wælk	14	woːk	4
WAUCE	woːs	woː	woː	woː	woːs	20	wæɔs	7	wɐs	2
WEICH	wæɪtʃ	wæɪk	wæɪk	wæɪtʃ	wiːtʃ	18	wɑek	4	wæɪtʃ	4
WHA	wɐː	wʉː	wʉː	wæɪ	wɐː	37	wæ	4	woː	1
WHALL	wæl	woːl	woːl	woːl	woːl	24	wæl	5	wɐːl	4

Nonword	DRC	CDP+	CDP+.50	CDP++	Participant responses					
					1st	n	2nd	n	3rd	n
WHAUCHE	woːʃ	woː	woːtʃ	woːtʃ	woːtʃ	10	woːʃ	8	wæɔtʃ	7
WHAZZ	wæz	wɔz	wæz	wæɪz	wæz	32	wɐz	3	wɔz	3
WHOLT	wɔlt	wɑːlt	wɑːlt	wəʉlt	wɔlt	29	hɔlt	6	hwɔlt	5
WHONE	wəʉn	wɑːn	wɑːn	wɑːn	wəʉn	26	hwəʉn	9	həʉn	5
WHOS	wɔs	wɔz	wɔz	wɔz	hɑːz	27	wɔs	4	hɑːs	3
WHUMB	wɐm	wɑːm	wɑːm	wɐmbl	wɐm	21	wɐmb	8	hwɐmb	2
WIS	wɪs	wɪz	wɪz	wɪz	wɪs	32	wɪz	11	vɪs	1
WOUGE	wæɔʤ	wɔʤ	wɔʤ	wæɔʤ	wɑːʒ	10	wɑːʤ	10	wæɔʤ	5
WRAWSE	roːs	roːz	roːz	roːz	roːs	14	roːz	8	ræɔz	2
WREWTH	rɑːθ	rjɑːθ	rjɑːθ	rjɑːθ	rɑːθ	18	reθ	8	riːθ	4
WRICHE	rɪʃ	rɑetʃ	rɑetʃ	rɑetʃ	rɑetʃ	10	riːʃ	9	riːtʃ	7
WROUNGE	ræɔnʤ	ræɔnʒ	ræɔnʒ	ræɔnʤ	ræɔnʤ	33	rɑːnʤ	5	rəʉnʤ	2
YALK	jælk	joːk	joːkk	joːkk	jælk	21	jɔlk	12	joːlk	7
YEC	jek	jes	jes	jek	jek	42	zek	1	jetʃ	1
YEESE	jiːs	jiː	jiːz	jiːz	jiːs	33	jiːz	6	jiːsiː	3
YIVE	jɑev	jɪv	jɪv	jɑev	jɑev	33	jiːv	4	jɪv	4
YONT	jɔnt	jɔn	jɔn	jɔnt	jɔnt	43	jʊnt	1	jəʉnt	1
YOUNGE	jæɔnʤ	jɔnʒ	jɔnʒ	jɑːnʤ	jæɔnʤ	16	jɐnʤ	7	jɑːnʤ	7
ZALPH	zælf	zæɪl	zæl	zælf	zælf	41	zoːlf	2	zɔlf	1
ZAQUE	zæɪk	zæk	zæk	zæɪk	zæk	23	zɐːk	5	zæɪk	5
ZARSE	zɐːs	zɐːz	zɐːz	zɐːz	zɐːs	35	zɐːz	6	zɐːrs	2
ZAUSE	zoːs	zoːz	zoːz	zoːz	zoːs	15	zæɔs	15	zoːz	10
ZENGE	zenʤ	zeʒ	zeʒ	zenʤ	zenʤ	33	zeŋ	3	zeŋg	3
ZI	zɑe	zɪ	zɪ	zɪ	ziː	36	zɑe	6	zɪ	2
ZOOK	zɑːk	zʊk	zʊk	zʊk	zɑːk	31	zʊk	14		
ZOOSE	zɑːs	zɑːz	zɑːz	zɑːz	zɑːs	37	zɑːz	8		
ZOS	zɔs	zəʉz	zəʉz	zɔz	zɔs	31	zɔz	10	zəʉs	2

CHAPTER 4.
Modelling the Acquisition of Grapheme–Phoneme Correspondences

Stephen C. Pritchard, Max Coltheart, Eva Marinus and Anne Castles

ARC Centre of Excellence in Cognition and its Disorders, Macquarie University, Sydney

Abstract

Coltheart, Curtis, Atkins, and Haller (1993) described an algorithm for learning grapheme–phoneme correspondences (GPCs). When presented with written words paired with their correct pronunciations, the algorithm could deduce GPCs. This algorithm was not comprehensively tested by its creators. In the present study, we programmed a GPC Learning Model that was based on the earlier work of Coltheart et al., and tested it more comprehensively than the earlier model was tested. Results show that the GPC Learning Model is able to learn GPCs, but that it experiences a range of difficulties. These include that it is more prone to error when trained with multi-morphemic words, and also when single-letter and multi-letter rules are learned in the same training phase. Its performance also deteriorates when trained with a realistic, token-based input corpus, as opposed to a type-based corpus where each word is presented only once. Despite these challenges for the GPC Learning Model, its operation raises interesting possibilities regarding the interaction of morphemic structure and GPC learning, and regarding whether GPC learning should be sensitive to type-based or token-based information.

Introduction

The dual-route cascaded model of reading aloud and word recognition (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; "Dual-Route Cascaded Model 1.2.1," 2009), is a *static model* of skilled reading. By “static”, we mean that DRC does not model the acquisition of reading skill, just the cognitive mechanisms involved in skilled reading. DRC does not improve or learn with additional exposure to print. It has been pre-programmed by its creators with knowledge relevant to reading, such as a written and spoken word vocabulary, knowledge of letters, knowledge of graphemes and how the graphemes correspond to phonemes.

That DRC does not explain the way people learn to read has been a point of ongoing criticism (e.g., Perry, Ziegler, & Zorzi, 2007; Seidenberg & Plaut, 2006). Despite this criticism DRC can account for a large range of empirical data regarding the way people read (Coltheart et al., 2001), and through examination of how the model performs when parts of it are lesioned, the DRC model offers an account of various types of acquired dyslexia (Coltheart, 2006; Coltheart, Saunders, & Tree, 2010). Due to its success in modelling these phenomena, DRC is regarded as a highly successful model of reading aloud and word recognition (e.g., Adelman, Marquis, Sabatos-DeVito, & Estes, in press; Protopapas & Nomikou, 2009; Sprenger-Charolles, Siegel, Jimenez, & Ziegler, 2011).

Due in part to its success, a sensible approach to computationally modelling reading skill acquisition is to introduce learning to the DRC model, rather than start afresh in creating a new model. This article examines one possible approach to introducing learning to DRC, focussing specifically on the sublexical route. This approach is a model for learning grapheme identities (e.g., knowing that SH is a grapheme that corresponds to a single phoneme, and is present in a word like WISHED) and grapheme–phoneme correspondences (GPCs) (e.g., that

SH corresponds to /S/[1]). We define a grapheme as being either a letter or sequence of letters that corresponds to a single phoneme.

Coltheart, Curtis, Atkins and Haller (1993) previously described an algorithm for learning graphemes and GPCs. This algorithm took a *supervised learning* approach to GPC learning, where the inputs to the algorithm are written words accompanied by their correct spoken pronunciation. The written word was input to the model as a string of alphabetic letters, and the corresponding spoken word was input to the model in the form of a string of phonemic symbols (e.g. /pEt/ for the word "pet"). The task performed by the algorithm was to analyse the written and spoken word strings, and determine which letters or groups of letters corresponded to each phoneme. It was trained using a corpus 2,897 words that was developed by Seidenberg and McClelland (1989) to train their connectionist model of reading. Seidenberg and McClelland indicate that "morphologically complex words" (p. 530) were removed from this corpus, suggesting it contained only mono-morphemic words.

It is important to recognise that Coltheart et al. (1993)'s algorithm was not simply learning GPCs. It was also learning grapheme identities. Before a reader can make use of GPCs, they must decide what letters constitute graphemes. Their algorithm did not attempt to model the learning of grapheme parsing, and instead was pre-programmed with knowledge of which graphemes to apply and under what circumstances, similar to the way graphemes are parsed in DRC (Coltheart et al., 2001). Their algorithm also assumed knowledge of letter identities, and phoneme identities. Unfortunately, this algorithm is no longer available, having been programmed almost two decades ago.

The aim of this research was to re-examine the learning algorithm proposed in Coltheart et al. (1993). This involved re-programming a version of the algorithm (hereafter

[1] A list of phonemic symbols used in this article is included in Appendix A.

referred to as the "GPC Learning Model", or simply "the model"), based on the account provided in Coltheart et al. While we attempted to accurately recreate the original algorithm, this was challenging to do with only the written description available, so there were some differences between the present model and the original algorithm.

The output of this new model is a set of GPC rules. These rules will be in a form that can be used in the existing most-current DRC model ("Dual-Route Cascaded Model 1.2.1," 2009). We assessed the capacity of the GPC Learning Model in appropriately and plausibly learning GPCs via five measures: 1) comparing the accuracy of using the newly learned GPCs in DRC to the accuracy of DRC with its default GPCs, in naming words; 2) testing the GPCs learned against the empirical nonword naming data reported in Pritchard, Coltheart, Palethorpe, and Castles (2012); 3) testing whether the learned GPCs allow better nonword naming than DRC, or the connectionist dual process models (CDP+/CDP++) (Perry et al., 2007; Perry, Ziegler, & Zorzi, 2010); 4) by inspecting the GPCs learned and observing whether any obviously incorrect GPCs are learned, and 5) by considering the psychological plausibility of the model, especially in light of previous criticism of Coltheart et al.'s original algorithm published in Andrews and Scarratt (1998).

Design of the GPC learning algorithm

The GPC Learning Model is based on an intuitively plausible, high-level psychological account of GPC learning: if a beginning reader is exposed to printed words, while also being provided with, or having knowledge of, the spoken words to which these printed words correspond, the reader can learn to recognise graphemes and deduce GPCs, *if* the reader already has knowledge of phoneme identities. For example, if a beginning reader is presented with the written word CAT, and also has knowledge that this corresponds to the spoken word /k{t/, then they might identify the possibility that C corresponds to /k/, A to /{/

and T to /t/. After encountering many other words using these graphemes and phonemes, the beginning reader might increase their confidence that these three GPCs are reliable, to the point where the reader is able to apply these GPCs to sound out other words in which they are used, without supervision. This account of implicit GPC learning describes how children might acquire GPCs if they were taught under a *whole language* (Goodman, 1989) teaching regime, where there is less focus on explicit phonics instruction.

Knowledge of the spoken word to which each written word corresponds is typically assumed to come via direct instruction, such as a teacher reading with a child and voicing the spoken word while pointing to the printed word. However, we make the point that the correct pronunciation of written words that is required for learning GPCs could also be obtained independently by the beginning reader via a separate cognitive mechanism, such as the lexical route proposed as part of the dual-route theory of reading (Baron & Strawson, 1976; Marshall & Newcombe, 1973), a route that is implemented in the DRC model. For example, a child who has already learned that the word CAT corresponds to /k{t/ and has memorised this written-word–spoken-word correspondence could then use this knowledge to deduce sublexical relationships such as GPCs.

While this high-level account of GPC learning seems plausible, it is not sufficiently detailed to fully specify a computational model. A good deal of lower level detail is required to adequately computationally model an activity such as “deducing GPCs”. The GPC Learning Model described here provides one computational account of this lower-level detail. We now present a summary of how the new GPC Learning Model functions.

The model’s operation consists of two stages, an “information gathering” stage, and a “rule consolidation” stage. In the information gathering stage the model is presented one by one with many inputs (the “input corpus”), where each input consists of a written-word–spoken-word pair (e.g., CAT–/k{t/). All words used in the input corpus were monosyllabic in

the simulations conducted for this research. The model attempts to identify graphemes and GPCs in each input, and keeps track of the number of times it identifies a GPC as it proceeds through the input corpus. After the full corpus has been presented, the model will thus have developed a list of candidate GPCs, with a number associated with each one indicating the number of times that GPC was identified in the input corpus of word pairs.

In the rule consolidation stage, the model examines its knowledge of GPCs, and makes changes, including: dropping (or forgetting, or ignoring) GPCs that have only been infrequently identified; modifying GPCs that apply to the same grapheme (e.g., if both C–/k/ and C–/s/ have been learned) by converting them to "context" rules, thereby specifying the circumstances under which each GPC should be applied; and extrapolating some rules so that they can be more broadly applied, if this will not result in a contention. These processes are now described in more detail.

Information gathering stage

The GPC Learning Model is "rule-based", rather than connectionist (see Perry et al., 2007 for an example of a connectionist approach to learning sublexical knowledge). The model identifies each input as being one of three types, and applies a particular procedure for each type. We now describe the three types of input and how each is processed. Note that while the processing steps describe parsing the stimuli to search for graphemes, these parsing procedures are exclusive to the GPC learning process, and are different to the pre-programmed parsing procedures used in DRC to model reading aloud.

Type 1: number of letters equals number of phonemes

For inputs where the number of letters equals the number of phonemes, the model assumes that only single-letter graphemes are present, and that there is a one-to-one correspondence between graphemes and phonemes. For example, for CAT–/k{t/, the model

will identify this word-pair as the GPCs: C–/k/, A–/{/, T–/t/. There are three single-letter graphemes and three phonemes.

Type 2: number of phonemes is less than the number of letters

When there are fewer phonemes in the spoken word than there are letters in the written word part of the input, the model will deduce that a multi-letter grapheme must be present, and will search for it. This is done by identifying and incrementing the count for already known single-letter GPCs, and taking them out of the input. After this is done, if the remainder consists of only one phoneme and some letters, the model will deduce that the letters correspond to the phoneme. For example, if the model is presented with the input THAT–/D{t/, and has already seen A–/{/ and T–/t/ (perhaps from previously encountering CAT–/k{t/), the model will identify and remove the A and T GPCs, leaving an orthographic remainder of TH, and a phonological remainder of /D/. It will deduce that TH corresponds to /D/. A more detailed computational account of how this is done is provided in Coltheart et al. (1993). Note that the GPC Learning Model can cope with multi-letter graphemes that are also split graphemes (e.g., A.E–/1/ in the word BAKE), without requiring any logic specific to split graphemes. For such items, once all of the single-letter GPCs have been identified, the remainder still consists of a multi-letter grapheme and a single phoneme, and the split grapheme is learned as corresponding to the single phoneme. The GPC Learning Model can also cope with silent letters. If, after matching off all of the single-letter GPCs there is a remainder letter but no remainder phoneme, the model will deduce that a silent letter must be present. The remainder letter will be paired off with an adjacent letter to form a digraph, corresponding to the phoneme to which the adjacent letter corresponds. For example, when learning from the input KNIT–/nIt/, the N, I, and T are matched to the phonemes, leaving the K as an orthographic remainder with no phonemic remainder. The K is then grouped with the adjacent letter N, and the GPC KN–/n/ will be learned. This process also applies for learning geminate consonants, such as FF–/f/ in the word PUFF.

Note also that the GPC Learning Model does not necessarily learn from every input. If the model has not already learned the single-letter graphemes present in an input containing a multi-letter grapheme, it will be unable to reduce the remainder to a single phoneme. For example, if T–/t/ had not yet been identified, the remainder for THAT–/D{t/after taking out the already known A–/{/ GPC would be THT–/Dt/, which has more than one phoneme. Under such circumstances, the model would not be able to learn from this input. Similarly, if there is more than one multi-letter grapheme the model will not learn from the input. For example, for an input such as CHAIN–/J1n/, the model will not be able to remove known single-letter GPCs in a way that will result in a single phoneme in the remainder (after removing N–/n/, CHAI–/J1/ remains). Although the model does not learn from such inputs, it does not seem reasonable that beginning readers would fail to learn anything from such words. This is a shortcoming that will hopefully be corrected in future iterations of the GPC Learning Model, but for now, we hypothesise that there will be sufficient content in the input items that do not contain more than one multi-letter grapheme for the model to learn.

In summary, type 2 inputs, where there are fewer phonemes in the spoken word than letters in the written word, will only result in a learning experience if all of the single-letter GPCs in the input are already known, and if the input does not contain more than one multi-letter grapheme.

Type 3: words containing the letter X

Nearly all inputs will be either type 1 (where the number of letters equals the number of phonemes) or type 2 (where there are more letters than phonemes). However, there is also the comparatively uncommon occurrence of input word pairs where there are more phonemes than letters. These will all involve the letter X. Inputs that include X are a special case, since X is the only grapheme in English that corresponds to more than one phoneme (X–/ks/). This means that words containing X might actually contain more phonemes than letters, (e.g.,

BOX–/bQks/ has three letters and four phonemes). The GPC Learning Model needs to identify whether the input contains the letter X, and, if so, it will process the input only if the spoken word has exactly one more phoneme than there are letters in the input. If this is the case, all of the known single-letter GPCs are identified and removed, and the remainder should be the single letter X and two phonemes, which are then learned as a GPC. For inputs that contain the letter X and at least one multi-letter grapheme (e.g. COAX–/k5ks/), no learning will occur, and inputs where not all of the other single letter GPCs are as yet learned will also not result in any learning. We note that searching the input for the letter X happens prior to identifying whether the input is of type 1 or type 2. This means that inputs such as COAX which include the letter X and where the number of letters equals the number of phonemes will be identified as type 3 inputs, and not as type 1 inputs.

Rule consolidation stage

After the information gathering stage, the GPC Learning Model will have in memory a list of GPCs that it has identified, with a count next to each GPC of how many times that particular GPC has been identified over the course of the information gathering stage. This list comprises the candidate set of GPCs, which are now refined. The following refinement steps occur: a) delete low frequency rules; b) form context-sensitive rules, and c) extrapolate rules. These steps are explained below.

Delete low frequency rules

Written English is characterised by a high number of irregular words. By this, we mean that, while many words in English seem to be pronounced according to clear grapheme–phoneme correspondences, there are many exceptions to this. These are termed “irregular words”. For example, the grapheme CH is typically pronounced as /J/, but in the word CHEF it is pronounced as /S/. The GPC Learning Model will identify many such pronunciations as it proceeds through the information gathering stage. If a particular GPC is identified in less-

than-a-threshold number of instances (referred to as "low frequency" GPCs here on in) during the information gathering stage, then the GPC is dropped from the set. The threshold number of instances to avoid being dropped is set by the experimenter as a parameter choice.

Form context-sensitive rules

Even after the low frequency rules have been deleted, the draft list of GPCs may still contain contentious rules. These are rules that apply to the same grapheme but indicate a correspondence to different phonemes. For example, there are many words where C corresponds to /k/ (e.g., COT and CAT) and many words where C corresponds to /s/ (e.g., CELL and CITE). So it is possible that both correspondences will have been identified with sufficiently high frequency to avoid being dropped. During the formation-of-context-sensitive-rules step, the GPC Learning Model will systematically go through its draft list of GPCs, and group the GPCs into small sets, where each set consists of GPCs all applying to the same grapheme. For each small set, the GPC that will be regarded as the default rule is the one that was observed the most number of times during the information gathering stage. For the other GPCs in each set, if their frequency relative to the GPC with maximum frequency is less than a value determined by the experimenter and specified as a parameter choice, then they are dropped from the list. If any of the non-highest-frequency rules in a set are above the relative frequency cut-off, then the model will take the one with the highest frequency, and attempt to form a context-sensitive rule for it. To do this, the model will go back and loop through the full list of inputs in the input corpus, looking for the instances when the GPC under consideration was identified. Whenever it finds an input where this is the case, it will take note of the letters that precede and follow the letter comprising the GPC, and record how many times it notices the GPC with particular preceding and following letters. After doing this, the model will have a list of preceding and following letters, with a frequency count for each letter. If one particular letter is seen to precede or follow the rule with a frequency that dominates the frequency of the other preceding or following letters (with "dominance"

meaning it has a frequency relative to the frequency of other preceding/following letters greater than the value set by the modeller via a parameter value choice) then a context-sensitive rule is formed.

The formation of context rules is best illustrated with an example:

Suppose we choose the following parameter values:

Absolute frequency threshold: 2

Relative context frequency threshold: 0.15

Context letter dominance threshold: 2

Suppose the algorithm learned the following rules in the information gathering stage, and then grouped them into a set during the form-context-rules step because they all apply to the single-letter grapheme C:

C–/k/, frequency 160

C–/s/, frequency 130

C–/J/, frequency 40

C–/S/, frequency 16

C–/I/, frequency 1

The rule C–/I/ has a frequency less than the absolute frequency threshold, so it will be discarded, leaving four GPCs in the set.

The algorithm will try and form a context sensitive rule for all rules apart from the one with highest frequency, which is considered a default rule.

C–/k/ is the highest frequency rule, so it is treated as the default rule.

C–/S/ is fairly low frequency, and its frequency relative to the highest frequency rule is less than the relative context frequency threshold. So it, too, will be discarded. That leaves C–/s/, and C–/J/ as rules that may have context sensitivities.

The model will now re-examine the complete input corpus for any words that contain the rule C–/s/, and compiles a list of all the letters that precede or follow the letter C on inputs where this rule was identified. Assume that the model notices that the most frequent letter either preceding or following the GPC C–/s/ is that the C is followed by an E (as in the word CELL–/sEl/), and that this is seen 80 times. The model also notices that the rule is followed by an I 30 times (as in the word CITE–/s2t/).

Since the E is seen following the rule the most times, it is the prime candidate for creation of a context-sensitive GPC. The E dominates the I by more than the context letter dominance parameter, and so the following context GPC is formed: C[E]–/s/. This notation means that when a C is followed by an E, pronounce it as /s/.

The same process is applied for C–/J/, and a context rule will also be formed for this rule, if a dominant context is discovered. If no dominant context is discovered, then no context rule will be formed, and the rule will be discarded.

Note: even though the following rules should also be formed C[I]–/s/ and C[Y]–/s/, the model will only learn the context sensitivity for the E, the following contextual letter that dominates the other potential contextual letters I or Y. This is a clear shortcoming of the model that we hope to correct in future iterations. One way to correct this would be to dispense with the idea of contextual dominance, and instead make a rule for any contextual conditions that occur a sufficient number of times.

Extrapolate rules

In order to explain the extrapolate rules step, we must first discuss GPC position, which we have avoided til now for simplicity. The DRC model understands GPCs as being the correspondence between a grapheme and phoneme, where the grapheme occurs *at a particular position* within the word. DRC identifies graphemes in three separate positions: beginning of a word (the first grapheme), end of a word (the last grapheme), and the middle of a word (all other positions)[1]. The same grapheme corresponding to the same phoneme but in a different position is considered a separate GPC. For example, B–/b/ in the beginning position is one GPC, and B–/b/ in the end position is another GPC[2].

During the extrapolate rules step, the model will identify graphemes for which there are GPCs in two positions, but no GPC for that grapheme in the 3rd position. For any such graphemes that are found, if the grapheme corresponds to the same phoneme in the two positions where GPCs have already been identified, then this correspondence will be extrapolated to the 3rd position. For example, if OO–/u/ was identified in the middle and end positions during the information gathering stage, then, assuming that no rule was previously identified for OO in the beginning position, OO–/u/ will be extrapolated to the beginning position.

Learning single-letter and multi-letter graphemes in different phases

Our description of the learning of multi-letter graphemes makes clear that, in order to learn a multi-letter grapheme from a particular input item, the single-letter graphemes in that input item must have first been identified. This raises the question of timing—should the

[1] There is some ambiguity about the position of split graphemes, since they include parts that seem to be in multiple positions (e.g., A.E in the word BAKE seems to occupy the middle position with the A, but the end position with the E. This ambiguity is resolved by taking the position as corresponding to the location of the phoneme to which the split grapheme corresponds. For example, in BAKE–/b1k/ the split grapheme corresponds to the /1/ phoneme, which is in the middle of the word. Therefore the applicable grapheme is A.E in the middle position, not the end position.

[2] Perusing DRC's "gpcrules" file will reveal that, where the same grapheme corresponds to the same phoneme in the beginning, middle and end positions, the relevant entries in the file will be combined into a single entry listed as an "all positions" rule. Even though it is listed as a single line item in DRC's gpcrules file, this single line can still be understood as three separate GPCs, one for each position.

model even attempt to learn multi-letter rules before it has been well-trained on single-letter rules? It seems plausible that beginning readers might learn single-letter GPCs first, but there would likely still be some overlap. For example, TH–/T/ is plausibly acquired prior to X–/ks/.

One approach to timing would be to allow the model to proceed once through the entire input corpus. It will of course fail to learn any multi-letter graphemes initially, but, as more and more single-letter graphemes are identified, it will be able to slowly begin learning multi-letter graphemes. After proceeding through the input corpus one or more times (we will refer to each presentation of the corpus form hereon as an "epoch", which is the term typically used in connectionist model training), a single rule consolidation stage is then performed.

An alternative approach is to divide learning into two phases. During the first phase, the model might be presented with each written-word–spoken-word pair in the input corpus for one or more epochs, and only attempt to learn single letter rules, and then perform rule consolidation on these single-letter GPCs. Following this, a second phase might occur, where the model again receives one or more epochs of the input corpus, this time with the capacity to also learn multi-letter GPCs, and a second rule consolidation stage occurs for this second phase. Very similar to this approach is the idea of dividing the input corpus into words that only contain words with simple single-letter graphemes, which are presented in one phase, and words that contain at least one multi-letter grapheme, which are presented in the second phase. That is, the input corpus is altered from phase to phase, instead of modifying what the model learns from phase to phase. However, doing this is almost identical to the model only learning single-letter rules from the entire input corpus, since the model will ignore any words that contain multi-letter rules while doing this. For this reason we will not separately consider the case of dividing the input corpus, and will only consider an approach where the model learns single-letter GPCs from the one entire corpus in the first phase, and multi-letter GPCs from the same corpus in the second phase.

The alternative approach of dividing learning into two phases seems intuitively to offer greater potential for effective GPC learning, but it may be unnecessary and may also raise questions of psychological and behavioural plausibility. Coltheart et al.'s original algorithm was only tested using a two-phase approach, though they suggested their algorithm should still work appropriately if single-letter and multi-letter GPC learning were combined into a single phase. We will test the impact of trying to learn all GPCs in a single phase versus learning single-letter and multi-letter GPCs in separate phases as part of this research. Now we turn to testing the newly programmed GPC Learning Model.

Word-reading simulations

Our aims in conducting simulations with this new version of the GPC learning algorithm were as follows. Firstly, to determine whether the new model could accurately and appropriately learn GPCs, and match the results reported in Coltheart et al. (1993) for the previous model on which it is based. Secondly, to investigate how altering the training corpus might affect learning. We test a training corpus with multi-morphemic words and one with only mono-morphemic words, and we also test a type-based training corpus where each word is presented only once, versus a token-based training corpus where each word is presented a number of times according to its frequency. And finally, we aim to examine the impact on performance of dividing learning up into different phases.

A number of preliminary, exploratory simulations were first conducted to determine a working set of parameter values. The best results (in terms of the percentage of words named correctly using the learned GPCs) were obtained by using the following parameter settings:

Minimum absolute frequency threshold: 2

Minimum relative frequency threshold: 0.09

Minimum contextual dominance: 2

These parameter settings were used in all simulations, unless otherwise indicated.

Simulation 1: performance with multi-morphemic words

This simulation aims to test the performance of the GPC Learning Model using a monosyllabic input corpus that includes multi-morphemic words (e.g., words such as STACKED).

Input corpus

The input corpus for simulation 1 is comprised of all the 8,017 monosyllabic words in DRC-1.2.1's vocabulary (nine 9-letter words were removed, since they are too long to be processed by an 8-letter-slot model such as DRC, and DRC-1.2.1 ignores them). This includes the multiple pronunciations of homographs and a large number of monosyllabic but multi-morphemic words, such as EATS, CHECKED, and YIELDS. The order of words in the corpus was randomised before being used to train the GPC Learning Model.

Procedure

The GPC Learning Model was trained using a two-phase process: an information gathering stage where only single-letter GPCs are learned, and the input corpus is presented once. Following this a rule consolidation stage is performed. This is the first phase. In the second phase, an information gathering stage is performed where the input corpus is again presented once, but only multi-letter GPCs are learned. This is followed by a second and final rule consolidation stage. This is a similar modelling procedure to that used in the previous model of Coltheart et al. (1993).

The output of the learning model is a set of GPC rules. This set of GPC rules is then used in the DRC model, in place of DRC's default, pre-programmed set of GPC rules. The DRC model with the newly generated GPC rules is used to model *sublexical-route-only*

responses to all of the words in the input corpus. By "sublexical-route-only" we mean that DRC's lexical route is not used in the simulations, and the output to each stimulus is generated solely by the sublexical route[1]. This is because we are testing for differences in sublexical route output as a result of different GPC rules, and lexical route involvement would obscure these differences. The output of DRC with the generated set of GPC rules is compared to: 1) the correct pronunciations for the words in the corpus, and 2) DRC's sublexical-route-only output to the corpus while using its original set of GPC rules.

Results and discussion of simulation 1

The results for simulation 1 (in addition to the results for simulations 2, 3 and 4) are presented in Table 1. This table also includes results for testing DRC with its default set of GPCs, and also the results reported in Coltheart et al. (1993) for their model, to aid in benchmarking performance. Note that the comparison to Coltheart et al.'s earlier model should not be considered too precise, due to the use of different word corpuses.

We first note that none of the model results presented in Table 1, not even for DRC-1.2.1 using its default GPCs, gets 100% of the words correct. This is simply because English is irregular, and it is central to dual-route theory that the sublexical route should be unable to generate correct pronunciations for every word. Using its default GPCs, DRC-1.2.1 does pronounce all of the regular words correctly, and none of the irregular words. This is because we define "regular" words as those that DRC-1.2.1's sublexical route can name correctly.

The GPC Learning Model does not perform well when the input corpus includes multi-morphemic words (Simulation 1). It performs substantially worse than the Coltheart et al. model, and substantially worse than DRC's default GPCs. It seems clear that this difference is more than just random order effects, due to the magnitude of difference in performance. So what explains these differences?

[1] DRC-1.2.1 has a built-in mechanism to quickly test non-lexical-route only responses, using the "- -reg" argument when initiating a simulation

Table 1- Simulation results for simulations 1 to 4.

Model	Input corpus	Testing corpus	% correct			% match to DRC-1.2.1's default GPCs
			all words	regular words	irregular words	
DRC-1.2.1 default GPCs	8,017 multimorph.	8,017 multimorph.	83.0	100.0	0.0	-
DRC-1.2.1 default GPCs (no outrules)	8,017 multimorph.	8,017 multimorph.	58.9	70.9	0.1	71.9
DRC-1.2.1 default GPCs	3,540 monomorph.	3,540 monomorph.	83.7	100.0	0.0	-
Coltheart et al. (1993)	2,997 monomorph.	2,997 monomorph.	78.2	-	-	-
Simulation 1	8,017 multimorph.	8,017 multimorph.	52.7	59.4	19.6	55.1
Simulation 1 (+ output rules)	8,017 multimorph.	8,017 multimorph.	59.1	66.6	22.6	61.5
Simulation 2	3,540 monomorph.	3,540 monomorph.	77.0	88.2	19.8	82.9
Simulation 2 (+ phonotactic outrules)	3,540 monomorph.	3,540 monomorph.	77.3	88.6	19.8	83.2
Simulation 2 (reordered)	3,540 monomorph. (reordered)	3,540 monomorph.	78.0	88.9	22.0	83.4
Simulation 3, Variation 1	3,540 monomorph.	3,540 monomorph.	61.8	69.4	22.7	65.2
Simulation 3, Variation 2	3,540 monomorph.	3,540 monomorph.	71.2	80.7	22.5	76.2
Simulation 4	102,574 tokens, monomorph.	3,540 monomorph.	62.5	69.0	29.6	62.8
Simulation 4 (alt. freq. cutoff)	102,574 tokens, monomorph.	3,540 monomorph.	62.1	69.1	26.2	63.4

The new GPC Learning Model, when trained with the large 8,017 word input corpus, learns many strange rules. For example, rules learned include <end position> PED - /t/ or <end position> WED - /d/. Counter-intuitive rules like these are acquired from multi-morphemic words in the input corpus like STOPPED - /stQpt/ or ROWED - /r5d/, but would cause problems when reading aloud words like SPED or WED. The reduced corpus of words used by Coltheart et al. to train the earlier algorithm omitted morphologically complex words, so presumably the earlier GPC learning algorithm avoided learning strange rules as a result.

Another reason for its seemingly poor performance is that DRC-1.2.1's default GPCs include "output" rules. These include both phonotactic constraint rules (e.g., when an end-position /d/ is preceded by a /p/, /k/, /S/ or /J/ it is pronounced as /t/), and morphophonemic rules (e.g., when a final /s/ is preceded by an /n/, it should instead be pronounced as /z/) (Coltheart et al., 2001). These rules will have a great impact on the successful pronunciation of multi-morphemic rules, which are comprised largely of plural or past-tense words ending in -/d/ or -/s/. When DRC-1.2.1's output rules were added to the set of learned GPCs and this new set tested, performance was greatly improved, and when DRC-1.2.1's default rules were tested with all output rules removed, its performance dropped substantially. These results highlight the importance of these output rules in correctly reading aloud many of the multi-morphemic words in the monosyllabic input corpus.

To further test whether the complications introduced by multi-morphemic words were indeed responsible for the overall low percentage of words read correctly, we created and tested a monomorphemic monosyllabic corpus, as described in Simulation 2.

Simulation 2: performance with monomorphemic words

This simulation aimed to test the performance of the GPC Learning Model using an input corpus consisting of only mono-morphemic monosyllabic words, to avoid the issues observed with simulation 1. The impact of changing word order in the input corpus was also investigated, as was the inclusion of phonotactic constraints.

Input corpus

The input corpus for Simulation 2 was comprised of 3,540 mono-morphemic monosyllabic words obtained from CELEX (Baayen, Piepenbrock, & Gulikers, 1995), and using British English pronunciations compatible with DRC-1.2.1.

Procedure

The same two-phase training process as in simulation 1 was used, training single-letter GPCs first. The performance of the learned GPCs was again tested by using these GPCs in conjunction with the DRC model. The words tested were the 3,540 monomorphemic words used in the input corpus for training. That is, as well as being trained on a different corpus to simulation 1, the output of the GPC Learning Model for simulation 2 was also tested on this different corpus.

Results and discussion of simulation 2

The results (presented in Table 1) show the GPC Learning Model performs similarly to the results reported in Coltheart et al. (1993) in terms of overall word pronunciation accuracy, when trained and tested on the mono-morphemic and mono-syllabic input corpus. This result suggests that the poor performance of the GPC Learning Model in Simulation 1 was a result of the use of multi-morphemic words in the input corpus, which lead to dysfunctional GPCs being learned. Despite the improved result, beginning readers are certainly not limited to mono-morphemic words as was the case in this simulation, so the challenge of appropriate learning in a multi-morphemic environment remains.

In considering the performance of the GPC rules determined by the GPC Learning Model relative to DRC's static, pre-programmed GPCs, we again considered DRC's output rules. It is justifiable that the phonotactic constraint rules (though not the morpho-phonemic rules) could be included with the rules learned by the GPC learning algorithm, since phonotactic constraints aren't learned, but are a product of our articulatory anatomy. For this reason, we also tested the GPC rules learned in Simulation 2 with phonotactic output rules from DRC-1.2.1 included. This, as expected, resulted in a small improvement in performance. The results of this additional simulation are also included in Table 1.

To explore the impact of changing the order of words in the input corpus, we also re-ran simulation 2 using the same input corpus, but with the order of words in the corpus randomly altered. The resulting set of rules generated performed slightly better than the rules generated with the initial ordering. These results suggest that the order of rule acquisition is somewhat dependent on the order of words in the input corpus, and changing the order of words will change rules acquired, even if by only a small amount.

Simulation 3: testing simultaneous learning of single- and multi-letter GPCs

Coltheart et al. (1993) suggested that learning single-letter GPCs in a separate pass prior to learning multi-letter GPCs was not essential. However, they did not test this approach. The aim of simulation 3 was to test whether the GPC Learning Model can maintain performance if multi-letter GPCs are learned in the same phase as single-letter GPCs.

Input corpus

The same 3,540 mono-morphemic, monosyllabic input corpus as was used in simulation 2.

Procedure

Two training variations were tested:

Variation 1: Single-phase learning, one rule-consolidation step: the model conducted one information gathering stage, where the input corpus was presented for two epochs, and the model attempted to learn both single-letter and multi-letter GPCs during this one stage. Following this, a single rule-consolidation stage was performed.

Variation 2: “Repeated-phase” mixed learning, with two information gathering stages, and two rule consolidation stages: the model was presented with the input corpus for a single epoch in an information-gathering stage, and attempted to learn both single-letter and multi-letter GPCs simultaneously during this stage. After this, a rule-consolidation stage was

performed, which removes infrequent rules, extrapolates rules, and creates context rules. A second information gathering stage was then performed for one epoch, and again, both single-letter and multi-letter GPCs could be learned in this stage. Finally, this was followed by a second rule-consolidation stage. In this variation, the model had the opportunity to consolidate rules after the first information gathering phase, to see if this would aid the learning of additional rules in the repeated information gathering phase.

Both variations were tested similarly to simulation 2, where the generated GPC rules were used in conjunction with the DRC model to generate pronunciations for the 3,540 mono-morphemic and monosyllabic words in the input corpus.

Results and discussion of simulation 3

Both variations incurred a drop in performance relative to the two-phase training approach used in simulation 2, though variation 1 was much worse than variation 2. This demonstrates that, for the GPC Learning Model, there is benefit in breaking learning up into different phases. Even if single-letter and multi-letter GPCs are learned concurrently, learning is improved if the model has an opportunity to consolidate its learning from the first pass through the input corpus, extrapolate some rules and learn context rules, before proceeding through a second pass of the input corpus. These results are in partial disagreement with the suggestion in Coltheart et al. (1993), that single-letter and multi-letter GPCs could be learned concurrently. GPCs can certainly be learned when single-letter and multi-letter GPCs are learned concurrently, but the number of words pronounced correctly with the GPCs learned decreases.

Simulation 4: token-based learning

In the previous simulations each word was presented a single time. We characterise this as “type-based” learning, because the model is exposed to different word types, without receiving any information about which words occur more commonly in standard printed texts.

However, beginning readers are not exposed to each word once when learning. Instead, they may see some words many times (e.g. a very frequent word such as THE) and some words they may see only very rarely (e.g., VAUNT). We characterise learning from words based on frequency as "token-based" learning. This kind of learning would present the model with some indication of the frequency of each word, perhaps by presenting frequent words multiple times.

The nature of the input corpus has implications for the GPC Learning Model, because the GPC Learning Model discards rules that are infrequently observed. Also, when two rules conflict, the GPC Learning Model will take the rule it has observed most often as the default rule, and attempt to make a context rule out of the other. So a rule that is observed infrequently with a type-based input corpus might get discarded, but that same rule might end up being retained if the input corpus is token-based. For example, TH in the beginning position most frequently corresponds to the unvoiced /T/ (as in a word like THIN) when considering words by type, but for token-based learning, the high frequency of function words using voiced TH (as in words like THE, THEN and THAT) may mean that TH–/D/ would be seen more often (Campbell & Besner, 1981).

The aim of Simulation 4 was to investigate whether there are any differences in performance when a token-based input corpus was used relative to a type-based input corpus.

Input corpus

Starting with the 3,540-word mono-morphemic input corpus from Simulation 2, we prepared a token-based corpus. This was done by including each word a number of times equal to its orthographic frequency (as reported in CELEX (Baayen et al., 1995)), divided by 100, rounding up, and adding 1. These manipulations ensure that the number of tokens in the final corpus would not be impractically huge, while ensuring that each of the 3,540 words was still presented at least once. The rounding and adding of 1 slightly distorted the frequency

distribution, mainly by making extremely infrequent words seem slightly more frequent, but this was necessary to keep the corpus to a manageable size, while still ensuring that each word was presented at least once. There were 102,574 word tokens in the final corpus, derived from the 3,540 word types. The order of these 102,574 word tokens was randomised. The word presented the most times was THE, which was presented 10,330 times. One-hundred-and-sixty of the 3,540 words types were presented only once (e.g., SCULPT was presented only once). The mean number of presentations for any particular word was M = 29.0, and SD = 280.3. The median was two presentations.

Procedure

Two variations were simulated, both using two-phase learning with single-letter GPCs learned first, and multi-letter GPCs learned in the second phase, as per Simulations 1 and 2.

Variation 1: uses the default GPC Learning Model parameters, as for previous simulations.

Variation 2: used an increased *minimum absolute frequency threshold* value of 4. We tested this variation since the large number of words presented meant that the model was seeing many rules more often than in the previous simulations, and so it is possible that increasing this minimum threshold could help to remove problematic low frequency rules that are observed more than twice within the large input corpus.

While the variations were trained on a token-based input corpus, they were tested on each word just once using the 3,540 word type-based input corpus.

Results and discussion

Both variations performed worse than the type-based learning model of Simulation 2. Variation 2 performed slightly worse than Variation 1, indicating that even some of the very low frequency rules learned by Variation 1 but removed for Variation 2 were useful in correct

reading aloud. These results suggest that GPC knowledge may not be best learned by a mechanism that is sensitive to word frequency. Rather, it seems that GPCs might be better derived from word types.

Nonword reading

The results reported in the simulations all focus on the word-reading performance of DRC's sublexical route only, using a set of GPCs learned by the GPC Learning Model. However, given the dominance of DRC's lexical route in determining the pronunciation of word stimuli, the differences highlighted between the various model variations may possibly not result in problems with word reading aloud accuracy when both cognitive routes contribute to the response. A clearer measure of GPC acquisition accuracy would be to assess the performance of the model variations against empirical data on how people pronounce nonwords. This would better isolate the action of the sublexical route, since the lexical route in dual-route models is typically not heavily involved in nonword pronunciation.

Empirical data on nonword pronunciation was recently published in Pritchard et al. (2012). They presented 412 nonwords to 45 participants for the purposes of obtaining data to adjudicate between the nonword naming performance of DRC-1.2.1, and the connectionist dual-process models, CDP+ and CDP++ (Perry et al., 2007, 2010).

We used the GPC rules from simulation 2 (without adding the phonotactic rules, and with the 1st randomised order, rather than the 2nd randomised order) to simulate reading of these 412 nonwords, then compared the results to those reported for the models in Pritchard et al. Results are displayed in Table 2.

Table 2 – Nonword reading performance. *Note.* DRC medians both by subject and by item (a) are significantly greater than the medians of the GPC Learning Model (d), $p < .001$. GPC Learning Model medians both by subject and by item (d) are significantly greater than the medians of the CDP+ model (b), $p < .001$. The GPC Learning Model median (d) by subject is significantly greater than the CDP++ model median (c), $p < .001$, however, by items, there is no significant difference between the GPC Learning Model median (d), and the CDP++ median (c). The Wilcoxon signed-rank non-parametric test was used for significance testing, due to the non-normality of the data.

Summary statistics	DRC default	CDP+	CDP++	GPC Learning Model
Percentage of nonwords in which a model's response matches:				
The most frequent participant response	73.5	12.1	37.6	40.5
None of the participants	1.5	49.0	26.9	42.2
By Subject: Percentage of nonwords for which a participant matches a model				
Minimum	29.3	4.9	17.8	18.5
Maximum	68.2	16.2	38.5	42.2
M	53.0	11.3	30.1	32.9
SD	9.0	2.5	4.4	4.9
Mdn	52.7_a	11.4_b	30.8_c	33.1_d
By Item: Percentage of participants who match a model for a given nonword				
Minimum	0.0	0.0	0.0	0.0
Maximum	100.0	100.0	100.0	100.0
M	52.8	11.2	30.0	30.7
SD	28.3	18.5	32.3	34.6
Mdn	53.3_a	2.2_b	15.7_c	13.6_d

The results indicated that the learned GPC rules were less suitable for nonword naming than DRC's default rules, but achieved better nonword naming accuracy than the CDP+ model, and comparable—if not slightly better—accuracy when compared to CDP++. By subjects, the median percentage of the 412 nonwords for which DRC with its default rules agreed with a participant was significantly greater than the median percentage of nonwords where DRC using the GPC Learning Model's rules agreed with a participant, $z = 5.8$, $p < .001$, $r = .61$. However, DRC using the GPC Learning Model's rules had a significantly

higher by-subjects median than either the CDP+ model ($z = 5.8$, $p < .001$, $r = .61$), or the CDP++ model ($z = 4.1$, $p < .001$, $r = .43$). The by-subjects minimums and maximums also indicate that even the participant who *least* agreed with the GPC Learning Model's rules (agreed on 18.5% of nonword pronunciations) still matched the GPC Learning Model more than any participant matched the CDP+ model (the participant who most agreed with CDP+ matched its pronunciation for 16.2% of the nonwords). By items, the median percentage of participants agreeing with DRC with its default rules on a given nonword was significantly greater than the percentage agreeing with DRC using the GPC Learning Model's rules, $z = 11.5$, $p < .001$, $r = 0.40$. However, DRC using the GPC Learning Model's rules had a significantly higher by-items median than the CDP+ model ($z = 8.1$, $p < .001$, $r = .28$). There was no significant difference between the by-items medians of the CDP++ model and the GPC Learning Model rules.

General discussion

Exploring the rules learned

We explored the rules learned by the model in Simulation 2 (without phonotactic rules, and on the first randomised input corpus ordering). A full list of these rules is provided in Appendix B. Most of the rules seem straightforward and uncontroversial. However, it is easy to see why the GPC Learning Model seems to do reasonably well when its output is used to pronounce words, compared to nonwords. It learns some strange rules, for example <beginning position> O[N]–/w/. Presumably, this rule is acquired from word-pairs such as ONE–/wVn/ and ONCE–/wVns/. Given there are only three words beginning with O followed by N in the mono-morphemic and monosyllabic corpus used to test this model, it is clear that this odd rule is "regular" in a domain restricted to this corpus. However, when pronouncing nonwords such as ONT or ONCH, it seems unlikely that a reader would employ

this rule to produce /wnt/ and /wnJ/ or even /wVnt/ and /wVnJ/ (the latter two are not pronounced by the model because it learns to always pronounce N as /n/, and the occurrence of N–/V/ in words like ONE is not frequent enough to avoid being dropped.) Similarly, <middle position> AL–/#/ seems a strange rule, acquired from words such as CALM–/k#m/ and PALM–/p#m/. While this might lead to some words being incorrectly pronounced (e.g., SALT is pronounced using the model's rules as /s#t/), many words will still be pronounced correctly. Again, while this might be regarded as the model's idea of regularity when considering words, the problem becomes clearer when nonword pronunciation is examined. For example, using this rule, DRC using the generated GPCs pronounces DWALP as /dw#p/, a pronunciation that none of the 45 participants produced. Another example of a strange rule is <middle position> O.B–/u/, presumably from a words like TOMB and WOMB.

The learning of counter-intuitive rules like these may reflect a shortcoming in the GPC Learning Model. This shortcoming might be in the type of training (word-based training, rather than training in explicit phonics), or it may be a shortcoming in the procedures of the algorithm itself. Increasing the cut-off frequency to drop these strange rules (which are typically only observed infrequently) is no solution, since many other low frequency rules learned by the model are valuable, and if these rules were dropped, performance would be negatively impacted. It seems that, if we are to accept that the cognition of reading involves learning GPCs in a manner analogous to the GPC Learning Model, people must retain certain GPCs despite them being infrequently used, while still rejecting other GPCs of similarly low frequency.

Another interesting aspect of the learned set of rules concerns middle-position rules and multi-morphemic words, as illustrated with the following example. The GPC Learning Model acquires <middle position> IE–/i/, and <end position> IE–/2/. In comparison, DRC's default rules include only <all positions> IE–/2/. Presumably, DRC includes only the one

pronunciation in all positions because there are many plural and past tense words in DRC's vocabulary where IE in the middle of the word is pronounced as /2/ (e.g., DIES, CRIED, LIED, PIES). So DRC regards words such as SHRIEK, THIEF, and SHIELD as irregular. However, when trained on a purely mono-morphemic input corpus, the GPC Learning Model perceives that IE corresponds to /i/ more often than /2/ when found in the middle of the word, and, according to the GPC Learning Model's rules, words like SHRIEK, THIEF and SHIELD are regular, not irregular. These results suggest the possibility that GPCs might be learned and applied with morphology in mind. A similar situation occurs with other GPCs (e.g., DRC uses the rule EAR–/7/ in all positions, and regards multi-morphemic words like SEARS and FEARED as regular, and regards mono-morphemic words such as HEARD, LEARN and PEARL as irregular. Yet the GPC Learning Model when trained on a mono-morphemic input corpus learns <middle position> EAR–/3/ and <end position> EAR–/7/).

Psychological Plausibility

The simulations show that the GPC Learning Model can successfully acquire GPCs. However, being able to acquire GPCs does not necessarily mean that the model is an effective model of cognitive processes, or that it can provide insight into the cognition of learning to read. Nor does it mean that the procedures for training the GPC Learning Model are an accurate account of the procedures a beginning reader undertakes to acquire GPCs. So how does the model fare in these respects?

From a behavioural viewpoint, the way the GPC Learning Model is trained seems akin to implicitly learning phonics via a "whole language" (Goodman, 1989) method of instruction. Rather than presenting the model with explicit relationships between individual letters/graphemes and phonemes, the GPC Learning Model is presented with whole-word correspondences, and left to try and deduce GPCs on its own. So while the simulations conducted for this research may provide an account of the cognitive processes involved in

learning GPCs via a whole-language type instruction program, they do not accurately portray the learning of GPCs under an explicit phonics program. This is unlike what was done for the CDP+ model, which was pre-trained on explicit grapheme–phoneme relationships (Perry et al., 2007). This is important, because it may explain some of the difficulty the GPC Learning Model has in acquiring GPCs—it is straightforward that the model, or a reader, who is not explicitly trained in phonics will be less likely to have internalised a knowledge of GPCs that maximises reading performance, or is characterised by rigid set of standard GPCs. The GPC Learning Model could be trained in a way more in keeping with explicit phonics training if its input corpus included grapheme–phoneme pairs in addition to written-word–spoken-word pairs. This might also improve performance. Even in a whole-language learning environment, it seems likely that most readers would receive some direct instruction in explicit phonics, especially for single letters and common digraphs, if not from teachers then from parents and other family members, or education programming on television.

More generally, the previous Coltheart et al. (1993) algorithm was criticised on psychological plausibility grounds (Andrews & Scarratt, 1998). These criticisms included that the model incorporated arbitrary design and parameter decisions rather than theoretically-motivated decisions, that the separation of single-letter GPC and multi-letter GPC learning into distinct phases was not realistic, and that the success of rule acquisition might be dependent on the order in which words are presented in the input corpus, in a way that doesn't occur with human readers. While none of these criticisms can be completely dismissed, we do make a number of points in response.

Firstly, on the question of arbitrary design choices, we argue that the current GPC Learning Model (and also Coltheart et al.'s previous model) is based on a high-level hypothesis regarding the way beginning readers learn GPCs when presented with printed words matched with spoken words—that they can deduce GPCs by examining the way

orthographic and phonological word forms are comprised, and how they relate to one another. To implement this computationally, many lower-level choices were made. To some extent, these are decisions made simply to complete a fully-executable model, rather than important choices regarding theory. If some of these low-level decisions were to greatly impact performance, they would be considered hypotheses about the cognitive processes involved in learning GPCs, which can be tested against empirical data. For example, Andrews and Scarratt suggested it was arbitrary for low frequency rules to be deleted. But this computational procedure models a plausible action on the part of a beginning reader: that candidate GPCs that do not seem to apply very often should not play a role in subsequent reading. A beginning reader upon first seeing the word CORPS–/k9/ might notice that ORPS seems to correspond to the phoneme /9/, and is a possible grapheme. However, since this correspondence will be rarely seen, it seems plausible that they might "ignore" it in future reading. The choice of a particular parameter value to decide the cut-off frequency may seem arbitrary, but it is not. It seems plausible that a beginning reader would settle on a means of ignoring low-frequency correspondences that maximises their reading performance.

Secondly, with regards to the separation of single-letter and multi-letter grapheme learning and the separation of rule consolidation and information gathering phases: we agree that it is artificial to so completely separate the learning of each of these. However, it is probably not controversial to suggest that children generally begin to learn single-letter correspondences before more complex correspondences, particularly split graphemes, or graphemes with more than two letters, perhaps even the less common digraphs. In the context of implicit grapheme learning where the beginning reader is attempting to deduce correspondences without direct instruction or phonics training, it is even more plausible that a beginning reader would focus on simple relationships before they were able to make sense of complex ones. We demonstrated that, while performance does deteriorate if all graphemes are learned in the same phase, most GPCs are still learned and performance largely maintained.

These results are promising for development of a future model that preferences the early learning of single-letter correspondences without having a clear-cut boundary between single-letter and multi-letter correspondence learning. With regards to the separation of the rule-consolidation phase (e.g., learning context rules) from the information gathering phase, this seems like a clear shortcoming of the GPC Learning Model, one that perhaps could be addressed in future model iterations.

Finally, Andrews and Scarratt questioned whether it would be realistic if the order of presentation of words to the model greatly impacted on the GPCs that were learned. In Simulation 2, we demonstrated that performance is affected by the order of acquisition, which justifies this criticism. However, to temper the criticism, performance might have changed, but it was not by much (the first ordering resulted in 77.0% of words being named correctly, while the second ordering resulted in 78.0% correct), and arguably there would be minor order of acquisition differences between beginning readers who are not explicitly taught phonics.

Beyond the criticisms of Andrews and Scarratt, there are other areas that stretch the psychological plausibility of the GPC Learning Model. In particular, the treatment of the letter X, and the way context rules are learned. The GPC Learning Model checks each input for the presence of the letter X, to determine whether or not to apply a procedure specific to this letter. While it is plausible that X is a challenging letter for beginning readers, since it corresponds to two phonemes, and it is less frequently used than many other letters, it is hard to justify treating it as a special case.

With respect to context rule learning, the GPC Learning Model implies that beginning readers first identify when learned GPCs provide conflicting pronunciations at a frequency too high to ignore, and then re-examine the entire input corpus to deduce the context rules that can solve the contention. Alternatively, beginning readers keep constant track of the context

in which each GPC occurs (quite a feat of memory), so that they can identify a context rule if the need arises due to conflicting GPCs being identified. Both alternatives seem somewhat questionable in terms of psychological plausibility.

Multi-morphemic words

The outcomes of Simulation 1 are especially problematic for the GPC Learning Model. This simulation showed that the GPC Learning Model performed considerably worse when the input corpus included multi-morphemic words. This sensitivity to choice of input corpus might indicate that the GPC Learning Model is fundamentally flawed. Another possibility is that GPC Learning Model contributes evidence that beginning readers identify morphemic structure when attempting to implicitly learn GPCs, in a way that the GPC Learning Model currently does not handle. Perhaps readers do not attempt to apply GPCs across morpheme boundaries. This is a hypothesis that is easily tested against empirical data.

The nature of learning in a realistic, multi-morphemic input environment has not been adequately explored even by other models of reading that learn. The triangle models (e.g., Seidenberg & McClelland, 1989) were trained on word corpuses less than 3,000 words that did not seem to include any morphologically complex items. While the CDP models (Perry et al., 2007, 2010) were trained on large word corpuses that did include multi-morphemic words, these models also received explicit pre-training on individual grapheme–phoneme correspondences which may compensate for any deleterious learning due to multi-morphemic words. And even with this pre-training Pritchard et al. (2012) found that the CDP models experience significant problems with nonword naming accuracy when assessed against human data.

The DRC model can handle multi-morphemic words and performed significantly better against the empirical data in Pritchard et al. than the CDP models. However, it is a static model that does not explain GPC acquisition. DRC-1.2.1 ("Dual-Route Cascaded Model

1.2.1," 2009) includes a number of context-sensitive GPCs as well as output rules that seem to be tailored specifically to aiding the pronunciation of multi-morphemic words and nonwords with a similar structure (e.g., DRC-1.2.1 includes the rule that when -ED is found at the end of a word and is preceded by vowel followed by 1-3 consonants, such as in a word like BANGED, then it should be pronounced as /d/. This ensures that BANGED is pronounced as /b{Nd/ by DRC's sublexical route, and not as /b{NEd/, with two syllables). It is not evident how such rules could be acquired, and the GPC Learning Model certainly does not acquire complex context rules like this.

The GPC Learning Model possesses a fairly basic approach to developing context-sensitive rules, and we earlier highlighted shortcomings with learning some context rules. Future iterations might look to make improvements to this aspect of the Model. Such improvements may look to ensure that more—and more complex—context rules are learned, so that useful GPCs can be learned from multi-morphemic training items, and fewer errors made when subsequently pronouncing multi-morphemic words.

Type versus token learning

If the psychological plausibility of the GPC Learning Model is debateable, of what use is this model? Coltheart et al. (1993) seemed to be using their earlier implementation of the model to demonstrate that the vocabulary on which it was trained contained sufficient information for comprehensive sub-lexical knowledge to be derived. They did this while critiquing the reading model of Seidenberg and McClelland (1989)—this model was found to be poor at reading nonwords accurately, while its creators argued that this was due to the impoverished set of words used to train the model (Seidenberg and McClelland (1990), as cited in Coltheart et al. (1993)). Coltheart et al. sought to demonstrate that this argument was incorrect. Does the current model offer anything new?

A new insight provided by the present model concerns whether beginning readers learn according to type or according to token. The sensitivity of readers to word frequency is well known (e.g., Forster & Chambers, 1973), suggesting that learning is sensitive to words being seen multiple times. However, in our simulations we found that the performance of the GPC Learning Model was higher if the model is trained only once or twice on each word type, rather than with multiple word tokens dependent on frequency. This strongly suggests that sublexical knowledge as proposed in dual-route models of reading might be learned according to type, rather than token, and challenges the token-sensitive learning approach to sublexical knowledge of the GPC Learning Model. It also challenges the token-sensitive learning approach of many connectionist models, such as the CDP models (Perry et al., 2007, 2010) and the triangle models (Harm & Seidenberg, 1999; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989). Even if beginning readers are exposed to words by token (which they obviously are) they seem to acquire knowledge that is type-based. Our findings add to the evidence for type-based sublexical learning already gathered by Pritchard et al. (2012), who pointed out that one reason the CDP models performed worse than DRC was because they were sensitive to token frequency, especially on the pronunciation of nonwords beginning with TH. Most experiment participants pronounce such nonwords with an unvoiced /T/, which is more common by type, while the CDP models were more likely to pronounce these nonwords with a voiced /D/, which is more common by token, due to the high frequency of words such as THIS and THAT. This idea is also in keeping with results published in Campbell and Besner (1981), who also found that people are more like to pronounce nonwords starting with TH using the unvoiced phoneme /T/ instead of the voiced /D/. This is unless the nonword is used in a sentence in place of a function word, in which case it is more likely to be pronounced voiced, again suggesting an interaction between learning morphology and learning and applying GPCs.

Future research

There are clear opportunities to conduct further research using the GPC Learning Model. These include: a) investigating explicit phonics training for the model, b) adding the capacity to learn grapheme parsing, rather than just learning GPCs, c) creating a new model where the rule-consolidation stage and the information gathering stage are merged into the one stage, to make the model more psychologically plausible, and d) incorporating the GPC Learning Model into a wider model of reading aloud. Exploring this last idea could be used, for example, to investigate how sublexical route learning might influence or be influenced by the concurrent learning of other facets of the cognitive systems involved in reading, such as orthographic learning.

References

Adelman, J. S., Marquis, S. J., Sabatos-DeVito, M. G., & Estes, Z. (in press). The unexplained nature of reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Andrews, S., & Scarratt, D. R. (1998). Rule and analogy mechanisms in reading nonwords: Hough dou peapel rede gnew wirds? *Journal of Experimental Psychology: Human Perception and Performance, 24*(4), 1052-1086.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX Lexical Database [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Baron, J., & Strawson, C. (1976). Use of orthographic and word-specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance, 2*(3), 386-393.

Campbell, R., & Besner, D. (1981). THIS and THAP--constraints on the pronunciation of new, written words. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 33A*(4), 375-396.

Coltheart, M. (2006). Acquired dyslexias and the computational modelling of reading. *Cognitive Neuropsychology, 23*(1), 96-109. doi: http://dx.doi.org/10.1080/02643290500202649

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100*(4), 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.

Coltheart, M., Saunders, S. J., & Tree, J. J. (2010). Computational modelling of the effects of semantic dementia on visual word recognition. *Cognitive Neuropsychology, 27*(2), 101-114.

Dual-Route Cascaded Model 1.2.1. (2009) [Computer Software]. Sydney: Macquarie Centre for Cognitive Science, Macquarie University. Retrieved from http://www.maccs.mq.edu.au/~ssaunder/DRC/2009/10/drc-1-2-1/

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning & Verbal Behavior, 12*(6), 627-635.

Goodman, Y. M. (1989). Roots of the whole-language movement. *The Elementary School Journal, 90*(2), 113-127. doi: http://dx.doi.org/10.1086/461607

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

Marshall, J. C., & Newcombe, F. (1973). Patterns of paralexia: A psycholinguistic approach. *Journal of Psycholinguistic Research, 2*(3), 175-199. doi: http://dx.doi.org/10.1007/BF01067101

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review, 114*(2), 273-315.

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the connectionist dual process (CDP++) model. *Cognitive Psychology, 61*(2), 106-151. doi: http://dx.doi.org/10.1016/j.cogpsych.2010.04.001

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.

Pritchard, S. C., Coltheart, M., Palethorpe, S., & Castles, A. (2012). Nonword reading: Comparing dual-route cascaded and connectionist dual-process models with human data. *Journal of Experimental Psychology: Human Perception and Performance, 38*(5), 1268-1288. doi: http://dx.doi.org/10.1037/a0026703

Protopapas, A., & Nomikou, E. (2009). *Rules vs. lexical statistics in Greek nonword reading.* Paper presented at the 31st Annual Conference of the Cognitive Science Society, Amsterdam, Netherlands.

Seidenberg, M. S., & McClelland, J. L. (1989). A Distributed, Developmental Model of Word Recognition and Naming. *Psychological Review, 96*(4), 523-568.

Seidenberg, M. S., & McClelland, J. L. (1990). More words but still no lexicon: Reply to Besner et al. (1990). *Psychological Review, 97*(3), 447-452. doi: http://dx.doi.org/10.1037/0033-295X.97.3.447

Seidenberg, M. S., & Plaut, D. C. (2006). Progress in understanding word reading: data fitting versus theory building. In S. Andrews (Ed.), *From inkmarks to ideas: current issues in lexical processing* (pp. 25-49). Hove, East Sussex, BN3 2FA: Psychology Press.

Sprenger-Charolles, L., Siegel, L. S., Jimenez, J. E., & Ziegler, J. C. (2011). Prevalence and reliability of phonological, surface, and mixed profiles in dyslexia: A review of studies conducted in languages varying in orthographic depth. *Scientific Studies of Reading, 15*(6), 498-521. doi: http://dx.doi.org/10.1080/10888438.2010.524463

Appendix A

Phoneme symbols used in this study are those used by DRC model 1.2.1.

Vowels		Consonants	
Symbol	Example	Symbol	Example
1	stay	_	jump
2	sigh	b	buy
3	bird	d	dot
4	boy	f	for
5	goat	g	guy
6	mouth	h	hot
7	beard	j	yell
8	cared	k	kite
9	board	l	low
#	hard / palm	m	my
{	cat	n	no
i	seen	p	pie
u	clue	r	run
E	red	s	stop
I	bid	t	tie
Q	pod	v	vent
U	good	w	west
V	fun	z	zoo
W	few	D	then
		J	chin
		N	hang
		S	shoe
		T	thin
		Z	measure

Appendix B

GPCs learned by the GPC Learning Model, trained with two-phases, with single-letter GPCs learned in the first phase, and multi-letter GPCs in the second phase. Parameter settings were: *Absolute frequency threshold:* 2, *Relative context frequency threshold:* 0.09*, Context letter dominance threshold*: 2. Note, for position: A = all positions, b = beginning, m = middle, and e = end. Context rules are denoted with use of square brackets. The letter within the brackets provides the context for the grapheme outside of the brackets.

Position	Grapheme	Phoneme
A	augh	9
e	ough	6
A	eigh	1
A	igh	2
e	dge	_
e	rsh	S
m	ear	3
e	are	8
e	eer	7
e	tch	J
e	rch	J
A	oar	9
e	oor	9
e	ear	7
e	ach	J
e	ier	7
A	our	9
A	air	8
e	che	S
e	ase	z
e	ece	s
m	eig	1
b	ear	3
m	hoo	u
e	urr	3
e	ore	9
e	ath	T
e	ugh	f
e	ech	J
e	rge	_

Position	Grapheme	Phoneme
m	uoi	4
e	ere	7
e	ege	_
m	ie	i
m	y.e	2
A	sh	S
A	a.e	1
A	ch	J
A	aw	9
A	ou	6
m	i.e	2
A	oo	u
A	ar	#
A	ng	N
A	o.e	5
e	dd	d
e	oe	5
A	oi	4
e	ck	k
e	ff	f
A	e.e	i
m	ow	6
A	ai	1
A	ea	i
A	oa	5
A	er	3
m	o.b	u
A	th	T
A	ee	i
e	ss	s

Position	Grapheme	Phoneme
e	ey	1
m	al	#
A	or	9
e	se	s
e	ye	2
e	ze	z
b	wh	w
m	o.l	5
e	ll	l
A	a.l	9
m	u.l	U
e	ie	2
e	zz	z
A	ur	3
A	ae	1
e	ay	1
A	ue	u
m	au	9
e	ce	s
A	ew	W
e	ow	5
e	tt	t
m	ei	1
e	ge	_
e	ve	v
m	ui	u
A	ir	3
e	mb	m
e	oy	4
b	bu	b
A	ph	f
e	mn	m
m	a.s	#
b	gu	g
b	wr	r
m	ha	#
e	nn	n
m	u.k	w
b	kn	n
m	ol	5
A	x	ks
m	n[k]	N

Position	Grapheme	Phoneme
m	[q]u	w
m	a[s]	#
m	o[l]	5
b	g[e]	_
b	o[n]	w
e	[n]s	z
A	s	s
m	e	E
A	n	n
A	d	d
A	t	t
A	r	r
A	u	V
A	m	m
b	h	h
m	i	I
A	l	l
A	f	f
m	a	{
A	b	b
m	o	Q
A	c	k
A	k	k
A	p	p
A	g	g
A	v	v
A	w	w
b	j	_
b	a	{
e	e	i
A	q	k
e	y	2
e	i	i
m	y	I
b	o	Q
b	y	j
e	o	5
b	i	I
A	z	z
e	a	#
b	e	E

CHAPTER 5.
Discussion

Introduction

In this general discussion, I will provide a short overview of the research presented in this thesis, and describe the contributions made by my research to the study of the cognition of reading. This will be followed by a discussion of the future avenues for research that could build on my work.

In conducting this research, my primary aim was to introduce learning to the dual-route cascaded (DRC) model of reading aloud and word recognition (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). As a static model of skilled reading, DRC has been quite successful, but the absence of any account of reading acquisition is often regarded as a shortcoming of the model. In introducing learning to the DRC model, my approach involved focussing on computationally modelling a high-level psychologically plausible account of reading acquisition. This is as opposed to emphasising the consideration of lower-level micro-cognitive or biological plausibility.

In researching the introduction of learning to DRC, I divided the task into two broad areas: orthographic learning in the non-semantic lexical route of DRC which was described in Chapter 2, and learning in DRC's sublexical route which was described in Chapter 4. Separate modelling exercises were conducted for each area. I did this to be both practical and systematic: it was sensible to keep my investigation initially simple by separately examining cognitive sub-mechanisms involved in reading, with a view to incorporating knowledge about each sub-mechanism into a unified model of reading as future work. The modelling work on sublexical route learning was also supported by new empirical research as detailed in Chapter 3, which sought to adjudicate between two competing accounts of sublexical route structure and operation: DRC's grapheme–phoneme correspondence (GPC) rule-based account, and the connectionist account of the connectionist dual-process (CDP) models of reading aloud (Perry, Ziegler, & Zorzi, 2007, 2010).

Lexical-route learning

To investigate incorporating learning into DRC's lexical route, I designed, programmed, and tested a "learning-DRC" (L-DRC) model, which provides a computational account of orthographic learning. L-DRC makes use of the existing DRC framework so that DRC's functions of reading aloud and word recognition are maintained in the new model. Maintaining DRC's capabilities while investigating new ones is in accordance with the idea of "nested modelling" (Jacobs & Grainger, 1994).

L-DRC's approach to orthographic learning is a computational account of the *self-teaching hypothesis* (Share, 1995, 2011), a well-regarded proposal for how children acquire the capacity to quickly and automatically read written words. According to the self-teaching hypothesis, beginning readers acquire new orthographic knowledge without comprehensive direct instruction as to the correct pronunciation of each and every word. They do this by using phonological recoding as a means of generating candidate pronunciations and thereby self-teaching, a process that is aided by contextual support when reading irregular words that cannot be fully decoded via a sublexical mechanism like phonological recoding.

The L-DRC model of orthographic learning and reading aloud was broadly successful in simulating the self-teaching hypothesis, while its learning mechanism also resulted in a structure and performance that approaches that of the static, skilled DRC-1.2.1 model ("Dual-Route Cascaded Model 1.2.1," 2009), post learning. Despite this success, a variety of challenges and problems were still exposed.

This research contributed on a number of levels to the pool of knowledge regarding the cognitive mechanisms of reading, and the computational modelling of reading aloud, word recognition, and reading skill acquisition. These contributions are now summarised.

Demonstrated DRC is compatible with learning

The approach to learning developed as part of the L-DRC model is compatible with DRC's general structure. This is significant because previous analysis of DRC's static structure questioned whether it was possible to learn such a structure. The development of L-DRC addresses this criticism, and improves the claim that DRC is a successful model of the cognitive mechanisms involved in reading.

A computational account of the Self-Teaching Hypothesis

The self-teaching hypothesis is a verbal theory, and lacks both detail and a means of being rigorously tested. In creating L-DRC I have produced a computational implementation that can be tested, and is thus more subject to falsification than a verbal theory. A computational account also commits to providing detail at finer levels than is provided by a purely verbal theory, such as the nature of the phonological recoding mechanism, or the way that partial decoding and contextual support might interact to enable irregular word learning.

Challenges for self-teaching

My research provided detail on how self-teaching might be problematic for certain word types: L-DRC finds it challenging to learn heterographic homophones (e.g., SALE/SAIL), heterophonic homographs (e.g., BOW), and potentiophones (e.g., BEAR, when pronounced regularly, becomes "beer"), suggesting that perhaps children will have similar difficulty when self-teaching. While it is not a surprise to suggest that children might have difficulty with such words, L-DRC goes a step further by providing specific explanations and hypotheses as to exactly why these words come to be challenging, and suggests a promising empirical research program to investigate how children learn such words.

Self-supervised learning, versus supervised or unsupervised learning

As a new computational model of orthographic learning, L-DRC can be contrasted with other computational models of reading that learn, and other theories of learning to read. Specifically, in building on the self-teaching hypothesis, L-DRC adopts a self-supervised (or internally-supervised, or semi-supervised) approach to learning, using the sub-lexical route to assist in training the lexical route. This contrasts sharply with the computational implementation of learning presented in the triangle model which uses back-propagation, a supervised learning approach. It seems implausible that beginning readers acquire the bulk of their reading skill via direct instruction regarding the correct pronunciation of every word, as is implied in the triangle model approach to learning. While Harm and Seidenberg (1999) and Harm and Seidenberg (2004) both argue that the triangle model is compatible with self-teaching, they do not adequately describe a computational mechanism—either verbally or computationally—that could achieve this type of learning within the triangle model framework.

L-DRC is also contrasted with clustering algorithms that are completely unsupervised in their approach to orthographic learning, such as the computational implementation of the Self-Organising Lexical Acquisition and Recognition (SOLAR) model of reading (Davis, 1999) or the Adaptive Resonance Theory (ART) model of Glotin et al. (2010). These models do not learn to associate orthographic word forms with phonological word forms, and we argued in Chapter 2 that this kind of associative learning is central to orthographic learning, in order for lexical decision to be possible. The SOLAR model which is presented as being able to learn to do lexical decision, is in fact not performing lexical decision—it is instead just learning to distinguish orthographic stimuli that have been previously presented (whether words or nonwords) from orthographic stimuli that are novel (whether words or nonwords). L-DRC differs from this in that it can learn which of the orthographic stimuli it sees are words and which are nonwords, even if these words and nonwords are seen equally often. It does

this because orthographic learning is guided by the recognition that a particular orthographic stimulus corresponds to a spoken word. While it was not explicitly tested on the lexical decision task, the ART model of Glotin et al. would experience the same difficulty as SOLAR, as would the self-organising map (SOM) model of orthographic learning described in Dufau et al. (2010).

A new approach to avoid length effects in DRC

DRC avoids unwanted length effects when simulating reading aloud by having any stimulus—regardless of length—contribute eight slots worth of excitation to compatible orthographic word nodes. DRC achieves this by having null (or blank) letters contribute excitation to make up eight slots worth of excitation for words that are shorter than eight slots (e.g., the orthographic word node for DOG will be excited by D, O, G, and 5 null characters). This seems implausible. For example, the node for the word AN in the orthographic lexicon should not be excited by six null letters, which would provide more excitation than the two letters actually comprising the word. Similarly, a completely different two-letter orthographic word such as GO should not receive six slots worth of excitation from null letters if AN is the stimulus. This approach to eliminating length effects when reading words is also likely to be problematic for a future DRC model that attempts to handle longer, multi-syllabic words. A model that could handle both of the words ENCYCLOPAEDIA and AN, for example, would need to contribute at least 11 slots worth of null-letter excitation to the AN orthographic node for this node to receive the same excitation from letters that the ENCYCLOPAEDIA orthographic node would receive. Otherwise, the longer word will be activated more rapidly, and named in shorter time.

In L-DRC, we introduced an alternative approach that involved orthographic word nodes being excited by only a single, end-of-word null letter each, with length effects being avoided by normalising the excitatory signals from the letter layer to the orthographic lexicon.

We demonstrated that this new mechanism performs as well as DRC's original mechanism during skilled reading, and this approach also improved performance when learning, by avoiding lexical capture associated with the excitation coming from null-letter slots.

A computational model of the interaction between partial decoding and contextual support

L-DRC attempts to model the combination of partial decoding and contextual support in aiding the self-teaching of irregular words. It was shown that in its current form, L-DRC has difficulty simulating partial decoding in a way that does not result in erroneous pronunciations being learned. Both L-DRC and DRC function by strongly inhibiting phonological lexicon nodes that are somewhat incompatible with the phonemes that have been activated. If this inhibition is too strong, partial decoding is not possible in L-DRC, because the correct irregular spoken-word representation will not be activated, even if the regular phonemes activated by the sub-lexical route comprise a near neighbour. But if this inhibition is too weak, then errors become more frequent, because lots of neighbouring word nodes will be excited.

In the general discussion of chapter 2, I suggested an alternative means of simulating context that may result in a computational model (L-DRC 2.0?) that is better able to simulate the interaction of partial decoding and context to learn irregular words with fewer errors.

Sub-lexical route learning

To explore sub-lexical route learning, I constructed a GPC Learning Model which was based on an earlier GPC learning algorithm described in Coltheart, Curtis, Atkins, and Haller (1993). I tested this new GPC Learning Model more comprehensively than the earlier model had been tested.

Analysis of the new GPC Learning Model highlighted the extent to which this model aimed for high-level psychological plausibility, while also discussing its deviation from psychological plausibility at lower levels of analysis. Like the L-DRC model of orthographic learning, the GPC Learning Model achieves some success. However, my research also highlighted several problems, which were extensively discussed in Chapter 4.

Research on the GPC Learning model, in addition to the empirical research described in Chapter 3, resulted in several contributions to the pool of knowledge on the cognitive mechanisms of reading, which are now described.

A new dataset of nonword pronunciations

Chapter 3 describes the development of a new corpus of nonwords, accompanied by a comprehensive dataset of empirical, human pronunciations for these nonwords. This dataset provides new insight into the variety of real-life pronunciations people produce to nonwords, and is potentially another important empirical benchmark for assessing both verbal and computational models of reading.

Improved assessment of the nonword naming accuracy of DRC, CDP+ and CDP++

Chapter 3 also included a critique of the nonword naming accuracy of DRC, CDP+ and CDP++. Previous methods of assessing the nonword naming accuracy of these models were deemed unsatisfactory. and assessing the nonword pronunciations of these models against an empirical dataset of human pronunciations considered a better measure. This chapter also includes data/analysis to inform the debate over whether sublexical knowledge is rule-based or statistical (connectionist) in nature.

Type-based versus token-based sublexical knowledge

My research provided both evidence and argument that sub-lexical knowledge may be type-based rather than token-based. Results reported on both the models comparison of Chapter 3 and the GPC Learning Model described in Chapter 4 show that type-based training seems to afford some benefits over token-based training. This might mean that, when learning sub-lexical orthographic–phonological associations, people are more sensitive to type than token. This is a challenge not just for the GPC Learning Model but also for the CDP models, and for other models that employ statistical, token-sensitive training regimes, such as the triangle models.

Learning sublexical knowledge from multi-morphemic words

The GPC Learning Model performs worse when trained on a large corpus that includes many multimorphemic words, as compared to a smaller corpus comprised of only monomorphemic words. Additionally, there is a slight change in performance when the order of items in the training corpus is altered. This result may offer one avenue to understanding and improving the nonword naming accuracy of CDP+ and CDP++. The sublexical route of these models was trained on word corpuses including multimorphemic words. If these training corpuses were limited to monomorphemic words, then perhaps the performance of CDP+ and CDP++ in naming nonword as reported in Pritchard, Coltheart, Palethorpe, and Castles (2012) may be improved. In sum, my research raises questions regarding how morphology and the learning of sublexical route knowledge may interact.

Future directions

There are a number of very clear paths forward to build upon the research described in this thesis. A first clear task is to test a variety of trained instances of the L-DRC model against the full gamut of empirical benchmarks (e.g., as listed in Perry et al. (2007)), to ensure

that, in the spirit of nested modelling, L-DRC can account for the same range of behavioural phenomena that DRC-1.2.1 can. Additional avenues for future research are as follows.

An alternative way of modelling contextual constraints

In Chapter 2, I described one mechanism for modelling the way context may act to aid the processes of self-teaching and reading aloud. This mechanism involved a single node (for the target word) being activated in the semantic layer, with the strength of this activation being a reflection of the degree to which contextual information (e.g., from the text that might accompany the stimuli) works to assist in identifying the spoken word corresponding to the written stimulus. This method seemed to show difficulties for L-DRC in modelling the way partial decoding might interact with context to facilitate irregular word self-teaching. By modelling weak context as being low excitation from the semantic layer to the phonological lexicon, more work is placed on the way phonemes excite the phonological lexicon, for the correct word to be excited. Using this approach to modelling context, we have seen that it is not possible to choose phoneme-to-phonological-lexicon inhibition in a way that allows partial decoding to occur without resulting in errors.

At the end of Chapter 2, I alluded to an alternative means of modelling context. Instead of the level of support provided by context being represented by a single semantic node receiving more or less excitation, an alternative way to model the usefulness of contextual support would be to activate greater or fewer nodes in the semantic layer. For cases where context provides a strong, unambiguous indication of the identity of the written stimulus, then a single node in the semantic layer can be activated. For example, only the node GO would be activated in the semantic layer when trying to read the missing word in the sentence ‘RED MEANS STOP AND GREEN MEANS ___’. When context is less clear about the identity of the written stimulus, then multiple nodes (most likely including the correct node) could be activated in the semantic layer, and since we are not attempting to model

semantic accuracy, only the impact context might have on self-teaching, the nodes activated other than the correct one could be randomly chosen. The number of additional random nodes activated in addition to the correct one could be used as a parameter to control the level of contextual ambiguity in identifying the written stimulus. More words activated equates to a more ambiguous contextual constraint.

This alternative approach may be more amenable to modelling the use of partial decoding with ambiguous/weak context. In the current L-DRC approach, when phoneme-to-phonological-lexicon inhibition is set low enough, phonemes activated by the sublexical route activate multiple neighbouring word nodes in the phonological lexicon, and the weak action of context tries to choose the correct one, but often fails, since weak context is simulated as weak excitation. The alternative approach would see weak context activating multiple nodes in the phonological lexicon, with strong activation from the phonemes activated by the lexical route serving to choose the correct word node from the context-constrained shortlist of possibilities.

Investigating the L-DRC approach to learning at the micro-cognitive level

Being focussed on macro-cognitive ideas such as the self-teaching hypothesis meant that I adopted a simple approach to node creation and connection creation, once a learning event had been triggered. This simulates a “black-and-white” learning process, where novel words are self-taught in a single exposure, or else unable to be learned correctly (e.g., a potentiophone). While subsequent exposures influence the frequency value attached to a specific word, this only impacts reaction time, and doesn’t really reflect a beginning reader gradually acquiring orthographic knowledge over more than one exposure. An obvious task for future development of L-DRC would be to retain self-teaching and the dual-route structure at the macro-cognitive level, but introduce more complexity at the micro-cognitive level in

order to more realistically simulate variety in the number of exposures required to acquire orthographic knowledge of a new word. How might this be done?

In Chapter 2's discussion we highlighted the value of the Self-Organising Lexical Acquisition and Recognition (SOLAR) model in describing—at the micro-cognitive level—an approach to orthographic learning. SOLAR describes an approach whereby individual connection strengths and node representations are modified by learning. The SOLAR model learns most words in only a few exposures, while occasionally some words take many more. This capacity for variety in the number of exposures required for orthographic learning seems more psychologically plausible than the simple approach adopted in L-DRC in this respect. We opted not to base our work on the SOLAR model, however, since it did not involve phonology in orthographic learning, and while it purported to model lexical decision, it is only able to distinguish words from nonwords through virtue of previous exposure. An interesting project for future research would be to attempt to include some of SOLAR's approach within L-DRC. Instead of self-teaching triggering a very simple node creation computation, maybe a more finely tuned account of how learning happens at the micro-cognitive level could be developed by having self-teaching trigger a SOLAR-like node and connection-strength-changing learning process. Doing this would provide L-DRC with more plausibility at the micro-cognitive level to match its plausibility at the macro-cognitive level at which the self-teaching hypothesis has its explanatory value.

On the question of normalisation in other parts of DRC

Chapter 2 included the successful inclusion of a new approach to managing length effects in DRC and L-DRC. The new approach only has one end-of-word letter per word contribute excitation to the orthographic lexicon, while length effects are avoided by normalising this excitation. Should normalisation be limited to just the letter-to-orthographic connections though?

At the moment in DRC-1.2.1 and L-DRC, the interaction between the phoneme layer and the phonological lexicon layer is such that phonological lexicon word nodes only receive activation from the phonemes comprising the word, and just a single end-of-word phoneme. In DRC-1.2.1 and in L-DRC, excitation from the phoneme layer to the phonological lexicon is not normalised. This means that a longer sequence of phonemes corresponding to a word will interact more vigorously with the corresponding word node in the phonological lexicon than a shorter sequence of phonemes, because there are more phonemes to contribute excitation. Since DRC-1.2.1 is a model of reading and not spelling, and since the activation of phonemes in L-DRC's self-teaching processes is dominated by the serial addition of letters to the sublexical route, no unintended length effects due to the lack of normalisation at the phoneme level have yet been observed. A future experiment and simulation may yet be determined that would expose this, but even if not, it seems theoretically inconsistent to normalise letter-to-orthographic lexicon excitation in L-DRC, and not do the same for phoneme-to-phonological lexicon excitation. It is also psychologically implausible that a short spoken word would be more slowly excited by the small number of phonemes that comprise it than a long spoken word.

This is not the only area of DRC or L-DRC where normalisation might be introduced. With the introduction of a semantic layer, the rate at which the lexical route cascades activation from the letter layer through the phoneme layer will change, depending on whether context-based activation of the semantic layer is present or not. This presents a challenge for balancing the relative contributions of the lexical route and the sublexical route when simulating reading aloud. It might be the case that it is sensible for the lexical route to overpower the sublexical route when contextual support is strong, or it might not be the case. If needs be, the relative excitation delivered to the phoneme level by each route could be normalised, to ensure that the presence or absence of context does not change the relative contribution of each route.

Improving the GPC Learning Model

In Chapter 4 we presented the GPC Learning Model, and described its capacity to plausibly model the way a beginning reader might implicitly acquire GPCs after being trained on the correct pronunciation of whole words. In addition, we identified and discussed shortcomings of the GPC Learning Model. Seidenberg (2005) notes that it is easy to falsify a computational model because every model is necessarily limited in scope. However, falsifications and shortcomings are important contributors to the formation of new hypotheses, and can drive new research. For example, although the GPC Learning Model is negatively impacted by being trained on a realistic, token-based word corpus, this result has led to the hypothesis that sublexical knowledge is acquired via sensitivity to word types, not word tokens.

Even so, the GPC Learning Model can be further investigated by developing new versions that seek to address perceived shortcomings. One shortcoming of the existing model is that learning is divided into various stages for reasons of computational practicality, rather than because beginning readers learn different GPCs in discrete stages. An updated version of the GPC Learning Model would look to try and incorporate the information gathering stage and the rule consolidation stage, so that both steps occur concurrently. For example, it seems unlikely that human readers form context rules in a separate stage to other rules. It also seems unlikely that extrapolation of rules would need to occur in a separate phase, rather than being something that the model can accomplish during the information gathering phase as it learns new GPCs.

Another shortcoming of the GPC Learning Model is that the simulations detailed in Chapter 4 involve training the model on only whole written-word-spoken-word correspondences, leaving it to the model to deduce GPCs from these. This effectively simulates a kind of whole-language approach (Goodman, 1989) to GPC knowledge

acquisition. In order to model the acquisition of GPC knowledge via explicit phonics training, the GPC Learning Model could instead be presented with individual grapheme–phoneme correspondences, as was done for CDP+ (Perry et al., 2007). For example, if the GPC Learning Model were first trained on all single-letter graphemes, common two-letter graphemes, and common context rules (such as C before E, I or Y being pronounced as "s"), then, in addition to achieving better results when simulating word and nonword reading, the GPC Learning Model might also be able to subsequently learn more complex GPCs with greater ease. It seems trivial that the GPC Learning Model would perform better if trained with explicit GPCs rather than whole words.

Incorporating the GPC Learning Model into L-DRC

Another challenging, but highly interesting avenue for future research would be include the GPC Learning Model as a part of L-DRC. This would enable a complex investigation of the simultaneous (or at least overlapping) acquisition of multiple sub-skills of reading. Self-teaching could be investigated dynamically, observing how it is affected by an evolving sublexical route. The potential self-teaching of the sublexical route, by having L-DRC's lexical route provide the pronunciations to the GPC Learning Model, could also be investigated.

Conclusion

Computational modelling has proven to be a popular and useful approach to investigating cognition, one that has been pursued vigorously in the study of the cognitive mechanisms involved in reading aloud and word recognition. Computational modelling has been used to develop accounts of organisation of our mental reading apparatus that focus on the micro-cognitive level, such as the triangle model (Harm & Seidenberg, 1999; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989), and accounts

that are more geared towards matching empirical results and the behavioural level and developing a robust macro-cognitive account, such as DRC (Coltheart et al., 2001) or CDP++ (Perry et al., 2010).

DRC has proven a successful model of the macro-cognitive hypothesis that there are two separate cognitive mechanisms involved in reading (e.g., Adelman, Marquis, Sabatos-DeVito, & Estes, in press; Protopapas & Nomikou, 2009; Sprenger-Charolles, Siegel, Jimenez, & Ziegler, 2011), which are termed the lexical route and the sublexical route within DRC. Perceptions of DRC's success have previously been subdued slightly by the observation that DRC provided no theory about reading skill acquisition (e.g., Perry et al., 2007; Seidenberg & Plaut, 2006). It is hoped that this research goes some way towards addressing this concern, and can contribute to improving our understanding of the mind, of how we read, how we learn to read, and how best to assist beginning readers in acquiring this crucial life skill.

References

Adelman, J. S., Marquis, S. J., Sabatos-DeVito, M. G., & Estes, Z. (in press). The unexplained nature of reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100*(4), 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.

Davis, C. (1999). *The Self-Organising Lexical Acquisition and Recognition (SOLAR) model of visual word recognition.* (PhD), University of NSW, Sydney. Retrieved from http://www.pc.rhul.ac.uk/staff/c.davis/Thesis/

Dual-Route Cascaded Model 1.2.1. (2009) [Computer Software]. Sydney: Macquarie Centre for Cognitive Science, Macquarie University. Retrieved from http://www.maccs.mq.edu.au/~ssaunder/DRC/2009/10/drc-1-2-1/

Dufau, S., Lété, B., Touzet, C., Glotin, H., Ziegler, J. C., & Grainger, J. (2010). A developmental perspective on visual word recognition: New evidence and a self-organising model. *European Journal of Cognitive Psychology, 22*(5), 669-694. doi: 10.1080/09541440903031230

Glotin, H., Warnier, P., Dandurand, F., Dufau, S., Lété, B., Touzet, C., . . . Grainger, J. (2010). An Adaptive Resonance Theory account of the implicit learning of orthographic word forms. *Journal of Physiology Paris, 104*(1-2), 19-26.

Goodman, Y. M. (1989). Roots of the whole-language movement. *The Elementary School Journal, 90*(2), 113-127. doi: http://dx.doi.org/10.1086/461607

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, 111*(3), 662-720. doi: http://dx.doi.org/10.1037/0033-295X.111.3.662

Jacobs, A. M., & Grainger, J. (1994). Models of Visual Word Recognition-Sampling the State of the Art. *Journal of Experimental Psychology: Human Perception and Performance, 20*(6), 1311-1334.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review, 114*(2), 273-315.

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the connectionist dual process (CDP++) model. *Cognitive Psychology, 61*(2), 106-151. doi: http://dx.doi.org/10.1016/j.cogpsych.2010.04.001

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.

Pritchard, S. C., Coltheart, M., Palethorpe, S., & Castles, A. (2012). Nonword reading: Comparing dual-route cascaded and connectionist dual-process models with human data. *Journal of Experimental Psychology: Human Perception and Performance, 38*(5), 1268-1288. doi: http://dx.doi.org/10.1037/a0026703

Protopapas, A., & Nomikou, E. (2009). *Rules vs. lexical statistics in Greek nonword reading.* Paper presented at the 31st Annual Conference of the Cognitive Science Society, Amsterdam, Netherlands.

Seidenberg, M. S. (2005). Connectionist models of word reading. *Current Directions in Psychological Science, 14*(5), 238-242. doi: 10.1111/j.0963-7214.2005.00372.x

Seidenberg, M. S., & McClelland, J. L. (1989). A Distributed, Developmental Model of Word Recognition and Naming. *Psychological Review, 96*(4), 523-568.

Seidenberg, M. S., & Plaut, D. C. (2006). Progress in understanding word reading: data fitting versus theory building. In S. Andrews (Ed.), *From inkmarks to ideas: current issues in lexical processing* (pp. 25-49). Hove, East Sussex, BN3 2FA: Psychology Press.

Share, D. L. (1995). Phonological recoding and self-teaching: sine qua non of reading acquisition. *Cognition, 55*(2), 151-218.

Share, D. L. (2011). On the role of phonology in reading acquisition: The self-teaching hypothesis *Explaining individual differences in reading: Theory and evidence* (pp. 45-68). New York, NY: Psychology Press; US.

Sprenger-Charolles, L., Siegel, L. S., Jimenez, J. E., & Ziegler, J. C. (2011). Prevalence and reliability of phonological, surface, and mixed profiles in dyslexia: A review of studies conducted in languages varying in orthographic depth. *Scientific Studies of Reading, 15*(6), 498-521. doi: http://dx.doi.org/10.1080/10888438.2010.524463

ethics amendment for Ref: HE24NOV2006-R04946

ethics secretariat <ethics.secretariat@vc.mq.edu.au> 29 March 2010 16:22
To: scpritchard@gmail.com
Cc: Max Coltheart <max@maccs.mq.edu.au>

Dear Stephen

Re: Recognizing, understanding and naming pictures and words.

Thank you for your recent email. The following amendment has been approved:

1. The addition of Mr Stephen Pritchard as a co-investigator on the project. Mr Stephen will be conducting the study for his PhD.

2. Participants will be drawn from the same subject pool as outlined in the original ethics application.

3. The information and consent form has been updated to reflect the addition of Mr Pritchard to the project.

Please do not hesitate to contact me if you have any questions or concerns.

Kind regards
Fran

****
Office of the Deputy Vice Chancellor (Research)

Ethics Secretariat
Research Office
Level 3, Research HUB, Building C5C
Macquarie University
NSW 2109

Ph: +61 2 9850 6848
Fax: +61 2 9850 4465

Email:
For Enquiries: ethics.secretariat@ro.mq.edu.au

http://www.research.mq.edu.au/