Reading Aloud: Feedback is Never Necessary

by

Serje Marc Robidoux

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Psychology

Waterloo, Ontario, Canada, 2010

Author’s Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Since McClelland and Rumelhart (1981) introduced the concept of interactive activation (IA) to the field of visual word recognition, IA has been adopted by all of the major theoretical models of reading aloud. This widespread adoption of IA has not been met with a close examination of the need for the principle features of this processing approach. In particular, IA assumes feedback from later processing modules to earlier processing modules. Though there exist data that can be explained by such feedback mechanisms, and indeed IA may be an intuitive approach to complex tasks like reading, little effort has been made to explain these same phenomena without feedback. In the present study I apply Occam's razor to the most successful model of reading aloud (CDP+; Perry, Ziegler, & Zorzi, 2007) and test whether feedback is needed to simulate any of the benchmark phenomena identified by Perry et al. (2007) and Coltheart, Rastle, Perry, Langdon and Ziegler (2001). I find that the data currently do not require any feedback mechanisms in reading aloud, and thus conclude that modelers in reading aloud have been too quick to adopt the principles of IA.

Acknowledgments

Graduate work is very rewarding, but also very demanding. Behind every PhD graduate stands a long line of people who share in the achievement, and I am no different. My family, friends, and the numerous colleagues that have come through our labs and the department during my time here have all been instrumental in getting me to this point.

I owe my greatest debt of gratitude to my advisor, Derek Besner. Over the last five years, he has always found the right balance between praise, criticism, patience and pressure. To summarize his contribution to my development both personal and professional would require a second thesis, so instead I'll just ask the words "thank you" to carry the heavy duty.

I am also tremendously indebted to Jenn Stolz. It was she who nurtured my initial interest in cognitive psychology as an undergraduate by giving me a copy of Ramachandran's anosognosia paper (I'd rather not think how many years ago). From there, she gave me my first opportunities as a computational modeler, and inspired my decision to return to graduate school. She guided me through my first years as a graduate student, seeing me through my master's thesis. Less obviously, she has also inspired my love of teaching. As her student so may years ago and then again as her teaching assistant in my first year of graduate school, she provided me with the model I use for my own teaching. (Also, thank you for the laptop. I love my laptop.)

There are dozens of others I could thank for making my graduate experience a positive one, through collaboration, research help, kindness and friendship (and financial support - thanks Mom and Dad). Each and every one of you owns a piece of this degree (no cash value).

Table of Contents

List of Tables

List of Figures

Overview

Theorizing about language processing has a history extending well over a century (Lichtheim, 1885). In contrast, computational modeling of visual word recognition is necessarily younger, but the last 30 years has seen rapid development of increasingly complicated models. In a very influential pair of articles, McClelland and Rumelhart (1981; Rumelhart and McClelland, 1982) set out the principles of a processing approach they called "interactive activation" (IA). In IA models, special purpose processing modules exchange information back and forth in a continuous fashion. Since its introduction, IA has heavily influenced research in psychology: these seminal articles have more than 1,400 citations combined and more than 2,200 records in PsycInfo include the phrase "interactive activation". In particular, IA is ubiquitous in theories of visual word recognition and reading aloud, and has had considerable success there: computational versions of these theories can simulate a wide range of experimental data (Coltheart, Rastle, Perry, Langdon & Ziegler, 2001; Perry, Ziegler, & Zorzi, 2007; Seidenberg & McClelland, 1989; Plaut, McClelland, Seidenberg, & Patterson, 1996; Plaut & Booth, 2000, 2006; Harm & Seidenberg, 1999, 2004; Woollams, Lambon Ralph, Plaut, & Patterson, 2007). Curiously, this success has not been associated with any examination of which properties of IA are strictly necessary for the simulation of human performance (but see Besner, 2006; Reynolds & Besner, 2002, 2004; O'Malley & Besner, 2008; and Norris, McQueen, and Cutler, 2000, who examine the role of feedback in the context of models of speech recognition). This study directly challenges the need for one (possibly the key) feature of IA in arguably the most successful of these models to date (CDP+, Perry et al., 2007). Specifically, it addresses the question of whether or not feedback is ever needed in CDP+ and if so, where in the reading system is it necessary?

The study will begin with a review of the history of computational modeling. Coltheart et al. (2001) have already provided an extensive overview of both the pre-computational and computational history of computational modeling, so I will provide a summary of the history that emphasizes only those aspects of the development directly relevant to the present research. Next I will turn my attention to evaluating the role that feedback plays in reading aloud, first by examining general performance measures such as accuracy in reading aloud words and non-words and the amount of variability in human performance that is captured by these models. The bulk of the study, however, will be concerned with examining the role that feedback plays in CDP+'s ability to simulate a broad range of phenomena identified as benchmarks by Perry et al. (2007) and Coltheart et al. (2001).

Processing in the Reading System

There is little doubt that processing in the reading system is modular in the sense that it consists of a set of interconnected components, each specialized to handle one aspect of visual word recognition (feature detection, letter identification, word identification, semantics). Though the internal structure of these modules and their arrangement is still hotly debated (see the section entitled "Current Models" for the two dominant theories), the idea of modularity itself has been adopted by most researchers and is well-supported by data from patients with brain trauma who have been found to have a wide range of highly specific deficits (for example, phonological dyslexics who have difficulty reading aloud unfamiliar letter strings like `FRANE`, with little or no difficulty reading aloud known words; Derouesné & Beauvois, 1985; Funnell, 1983; but see Dunn & Kirsner, 1988, for a caution in interpreting such dissociations). What is less clear is how information is communicated between the special purpose modules. Three general approaches have been proposed.

Discrete Stages

One possibility is that reading is accomplished in a series of discrete stages. In a staged processing approach, each module completely processes its input and then passes the results of that processing on to subsequent modules. Sternberg (1969; 1998; 2001; see also Roberts & Sternberg, 1993) demonstrated that stages as a general framework could provide a simple and clear interpretation of a broad range of experimental data. Stages of processing approaches also enjoy considerable flexibility in terms of the internal operation of each stage. As far as internal processing dynamics go, each stage is a black box that accomplishes its task however the theorist sees fit. The connections between stages are only relevant for input and output. I know of no currently implemented model of visual word recognition or reading aloud that includes any staged processing.[1]

Cascaded Processing

McClelland (1979) introduced the notion of cascaded processing to visual word recognition (and cognitive psychology more broadly). In cascaded processing, processing is no longer discrete. Rather, activation from each module is continuously passed on to subsequent modules in much the same way that water would flow down a flight of stairs. The internal structure of modules is also more rigidly defined: McClelland formalized the mathematics of nodes that accumulate activation based on the connections between them. It is this activation that "cascades" from one module to another. Though in the earliest moments of processing, the modules may provide weak or ambiguous activation output, the information passed from each module to the next becomes stronger and less ambiguous as processing progresses. McClelland (1979) demonstrated that, like the discrete staged approach, cascaded models could produce the overadditive and additive data patterns observed in many human experiments and provided guidance in interpreting these results (though he did not address underadditive patterns, such as those observed in Besner, O'Malley, & Robidoux, 2010).

Interactive Activation

Shortly after McClelland (1979) introduced cascaded processing, McClelland and Rumelhart (1981; Rumelhart & McClelland, 1982) introduced the Interactive Activation Model (abbreviated to IAM). In their words, the IAM assumes "that 'top-down' or 'conceptually driven' processing works simultaneously and in conjunction with 'bottom-up' or 'data driven' processing to provide a sort of multiplicity of constraints that jointly determine what we perceive." (McClelland & Rumelhart, 1981, p. 378) Simply described, IA models engage in *bidirectional* cascaded processing between modules through the addition of *feedback connections* (see Figure 1). In this way, later modules simultaneously receive activation from, and reinforce processing in, earlier modules (IA also includes within-level inhibitory connections, though these will not be considered further). Though McClelland and Rumelhart (1981) developed their IAM to account for a specific phenomenon (the word superiority effect – more on this later), McClelland (1987) proposed that the IA processing approach is a useful general account of language processing because it allows later stages of processing to assist earlier stages of processing in resolving ambiguities. McClelland and Rumelhart (1981) had a tremendous impact on the field, and IA is now a fundamental component of the most influential models of visual word recognition (Coltheart et al., 2001; Perry, Ziegler, & Zorzi, 2007; Seidenberg & McClelland, 1989; Plaut, McClelland, Seidenberg, & Patterson, 1996; Plaut & Booth, 2000, 2006; Harm & Seidenberg, 2004).

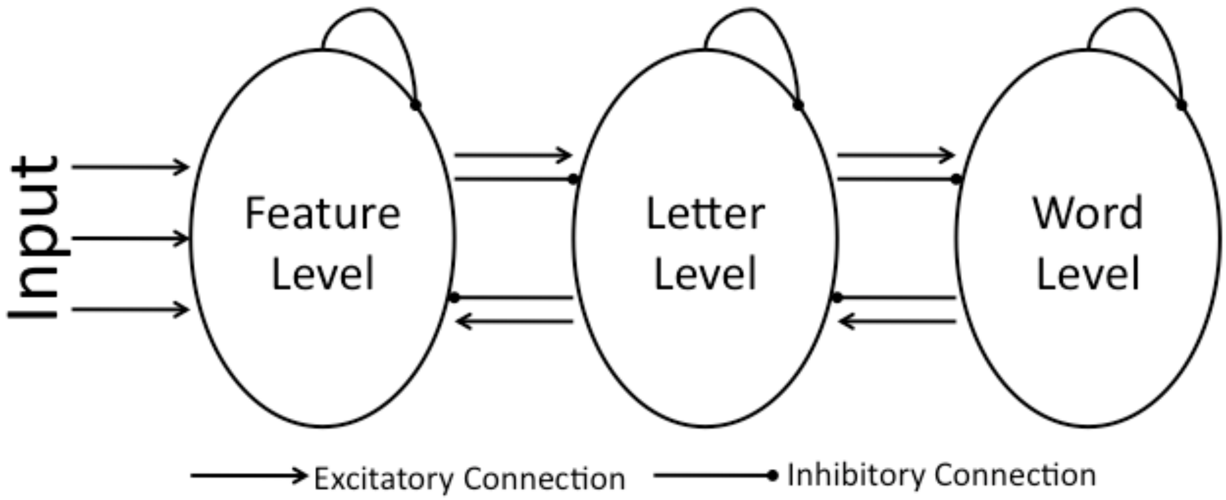

Figure 1. Interactive Activation Model of early visual word recognition processes. (Adapted from McClelland & Rumelhart, 1981)

Whereas McClelland (1987) has since argued that the framework has broader utility, the IAM's promise was initially demonstrated by its ability to simulate a specific phenomenon: the Word Superiority Effect (WSE). I turn now to a consideration of this effect, and the various computational models that account for it.

Accounts of the Word Superiority Effect

Reicher (1969) demonstrated that identification of which of two letters appeared in a brief masked display (e.g., D vs. K) was more accurate when the letter appeared as part of a known word (e.g., WORD) than when it appeared in other contexts [D alone, a display with two letters (G D), or in a scrambled word (OWRD)]. After Reicher's initial demonstration, this effect was closely scrutinized in a number of articles (e.g., Wheeler, 1970; Adams, 1979; McClelland, 1976; Massaro & Klitzke, 1979; Aderman & Smith, 1971; Johnston, 1978) and the effect came to be known as the WSE. As with many phenomena, the details of the WSE became more complex as researchers examined it more closely, but the key finding remained: letters were more easily identified when embedded in words, than when presented in isolation or embedded in nonwords (though see Aderman & Smith, 1971, for evidence of an advantage for letters embedded in pronounceable nonwords over other types of nonword strings). A number of different computational models have been proposed that account for this effect. A closer look at these models follows; with particular emphasis on the role feedback plays in their accounts.

Interactive Activation Model.

In McClelland and Rumelhart's (1981) IAM, the decision about which letter was presented is made based on activation in the letter level. The advantage for letters presented as part of a word arises because the word-level entry for the word is activated, which in turns feeds reinforcing activation back to the letter level. When a letter is presented in isolation (or as part of a nonsense letter string) there is little reinforcing feedback from the word level. This results in the letter level entry accumulating activation more rapidly in the word condition than in the other conditions. This increase in the amount of activation coming into the letter level reduces the ambiguity and results in the WSE: more accurate responses when the letters are embedded in words than when they are presented in isolation or in other nonword letter strings.

In the IAM, identification of which letter is present in the Reicher-Wheeler display relies exclusively on information from the letter level. The WSE is the direct result of feedback from the word level to the relevant letters in the letter level. Thus, under the assumptions of the IAM, feedback is necessary to produce the WSE. However, there are other accounts of the WSE that do not rely on feedback.

Activation-Verification Model

In Paap, Newsome, McDonald, and Schvaneveldt's (1982) Activation-Verification Model (AVM), information flows from the input to the letter-level then on to the word-level just as in the IAM. However, the AVM differs from the IAM in two very important ways: first, the AVM does not directly calculate activations of units. Rather, it relies on a matrix of the probability of confusing any two letters. These probabilities are assumed to be a proxy for the activations of the relevant representations. Second, there is no feedback mechanism in AVM, so that information about which word is present does not directly influence letter processing. The WSE arises because the decision about which letter was present is based on either the alphabetum (letter-level) or the lexicon (word-level) on each trial (but not both). The likelihood of a correct response is thus based on the conditional probabilities of a correct response from each level. The probability of a correct response based on letter-level information does not change with the context (though it might with the specific stimulus), however, the probability of a correct response based on word-level information will be very low if no word is presented, and relatively high if the letter is embedded in a word. This increased accuracy on word trials in which the subject relies on word-level information results in the WSE.[2]

Dual (Multiple) Read-Out Model

Grainger and Jacobs' (1994) Dual Read-Out Model (DROM; updated to the Multiple Read-Out Model, MROM; Grainger & Jacobs, 1996) is based on McClelland and Rumelhart's (1981) IAM, but with a different decision-making process. Instead of relying exclusively on the letter-level for identifying letters, the DROM uses an approach analogous to the one proposed by Paap et al.'s (1982) AVM. Critically, Grainger and Jacobs (1994) reported simulations of their approach both with feedback enabled (as in the original IAM) and with the feedback disabled. Both models were successful at producing a WSE. This result is particularly relevant to the question of interest here: though the IAM was initially designed to simulate the WSE, the Grainger and Jacobs result provides an existence proof that *feedback is not necessary* to achieving this goal even within the IAM. Indeed, McClelland and Rumelhart (1981, p. 404) considered the option of a decision that relied on more than just letter-level information, but argued that a complicated decision making process wasn't necessary due to the feedback in the IAM (they did not consider a model with no feedback). In a sense, the DROM and IAM approaches represent two alternatives to applying Occam's Razor: McClelland and Rumelhart's (1981) IAM opted for a simpler decision-making process at the cost of requiring feedback, whereas Grainger and Jacobs' (1994) DROM showed that feedback can be eliminated at the cost of a more elaborate decision-making process.

Fuzzy Logic Model of Perception

Massaro (1979) and Massaro and Cohen (1991) showed that the Fuzzy Logic Model of Perception (FLMP; Oden & Massaro, 1978), when applied to visual word recognition, can also produce a word superiority effect without need for a feedback mechanism. In the FLMP, pattern recognition (perception) is accomplished by three sequential processes: feature evaluation, feature integration, and decision. During feature evaluation, evidence for features is accumulated and features are deemed to be present to the extent that there is evidence for them (contrasted with the all or nothing view of feature detection in the IAM, AVM, and DROM). In feature integration, the "fuzzy" feature pattern is matched to prototypes to determine which stimuli are likely to be present, and finally, a decision process determines which of the potential items is present. Critical to the present discussion, the FLMP does not include a mechanism for feature integration to influence feature evaluation in the way that feedback allows the word level to influence letter level processing in the IAM. When applied to visual word recognition, the FLMP produces a WSE because the decision process makes use of information from the both the feature evaluation and feature integration processes to determine a response (in the same way that the AVM and DROM rely on multiple sources of information). When the letter is part of a word, the feature integration process provides a better match to the relevant word, and thus responses are more accurate when compared to a nonword, or letter-alone context.

Elementary Perceiver and Memorizer

Richman and Simon (1989) proposed a model of a very different sort. In the Elementary Perceiver and Memorizer (EPAM) decisions are made on the basis of a branching tree of choices that eventually terminates in a leaf that makes the response clear. When a response cannot be made unambiguously, EPAM creates a new discriminating branch that eliminates the ambiguity. Over time, it learns how to accomplish the task it has been given. An EPAM network trained on letter discrimination can be designed to use both word and letter information to accurately make decisions. Though the structures diverge dramatically, this feature of EPAM is analogous to the approaches to letter identification used by the AVM, DROM, and FLMP. As with the other models, by using both word and letter information, EPAM is able to produce a robust WSE. Critically, EPAM does not rely on constructs such as "feed-forward" or "feed-back" in its design.

I have described four very different computational models that are all capable of producing the WSE without the need for feedback (the AVM, the DROM, the FLMP, and the EPAM). What they all share in common is that the decision making process makes use of information from more than one source (i.e., from both feature/letter, and whole-word level information). This contrasts with the IAM, where the assumption is that the decision process relies only on the letter-level, thus requiring feedback to produce the WSE. This is an important distinction in assumptions, but with no particular evidence to favor one view or the other, I am left to conclude that although the IAM provided one account of the WSE, and IA has since found its way into every theory of visual word recognition and reading aloud, feedback is not necessary to explain the WSE. This leaves open the question of whether or not feedback is necessary to explain *any* phenomena in reading aloud or visual word recognition, or if the current theories would do just as well without any feedback. This question is the focus of the remainder of this dissertation.

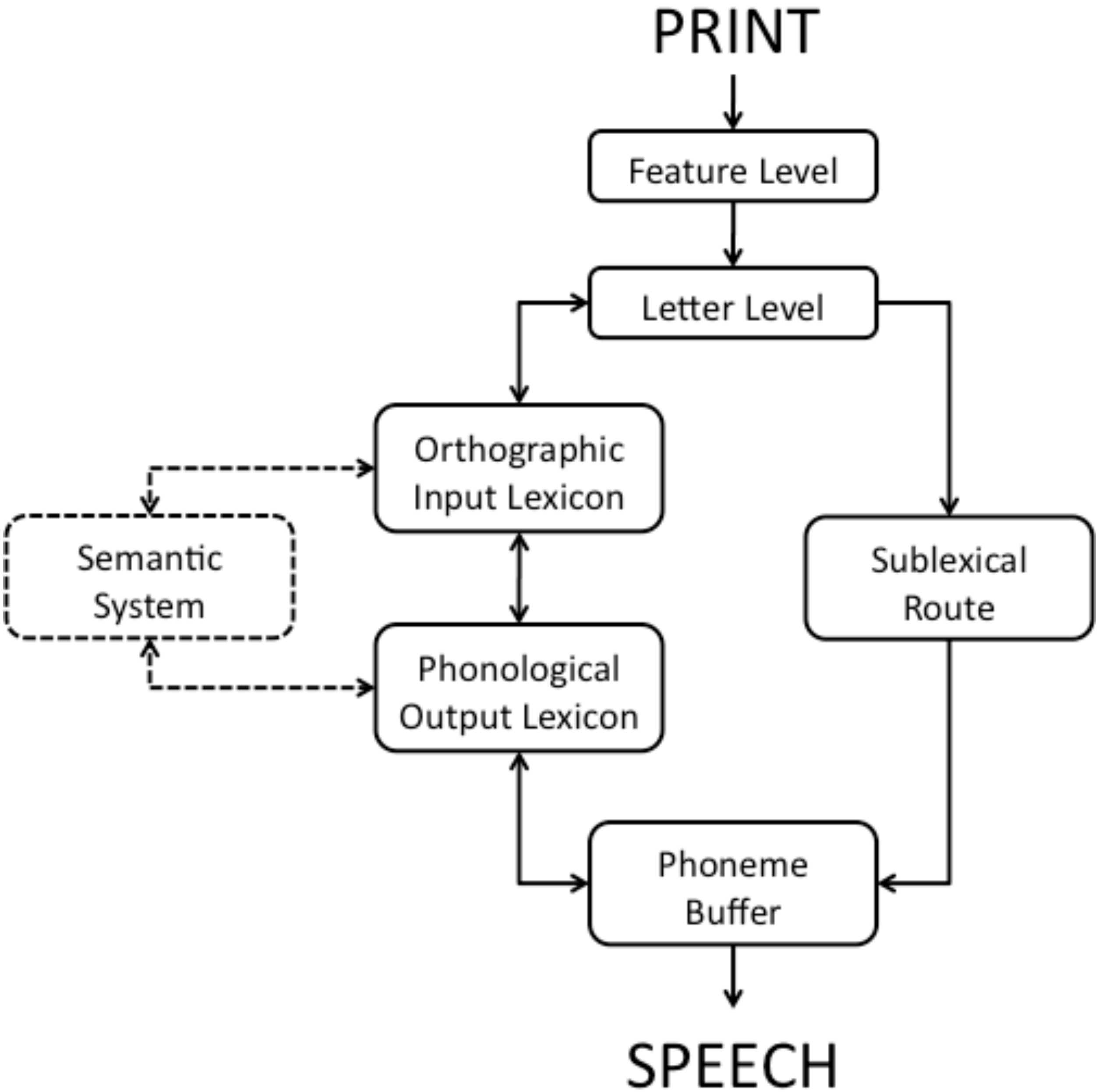

Figure 2. General structure of the dual-route class of models. The semantic system is denoted by dotted lines because no computational dual-route model implements it.

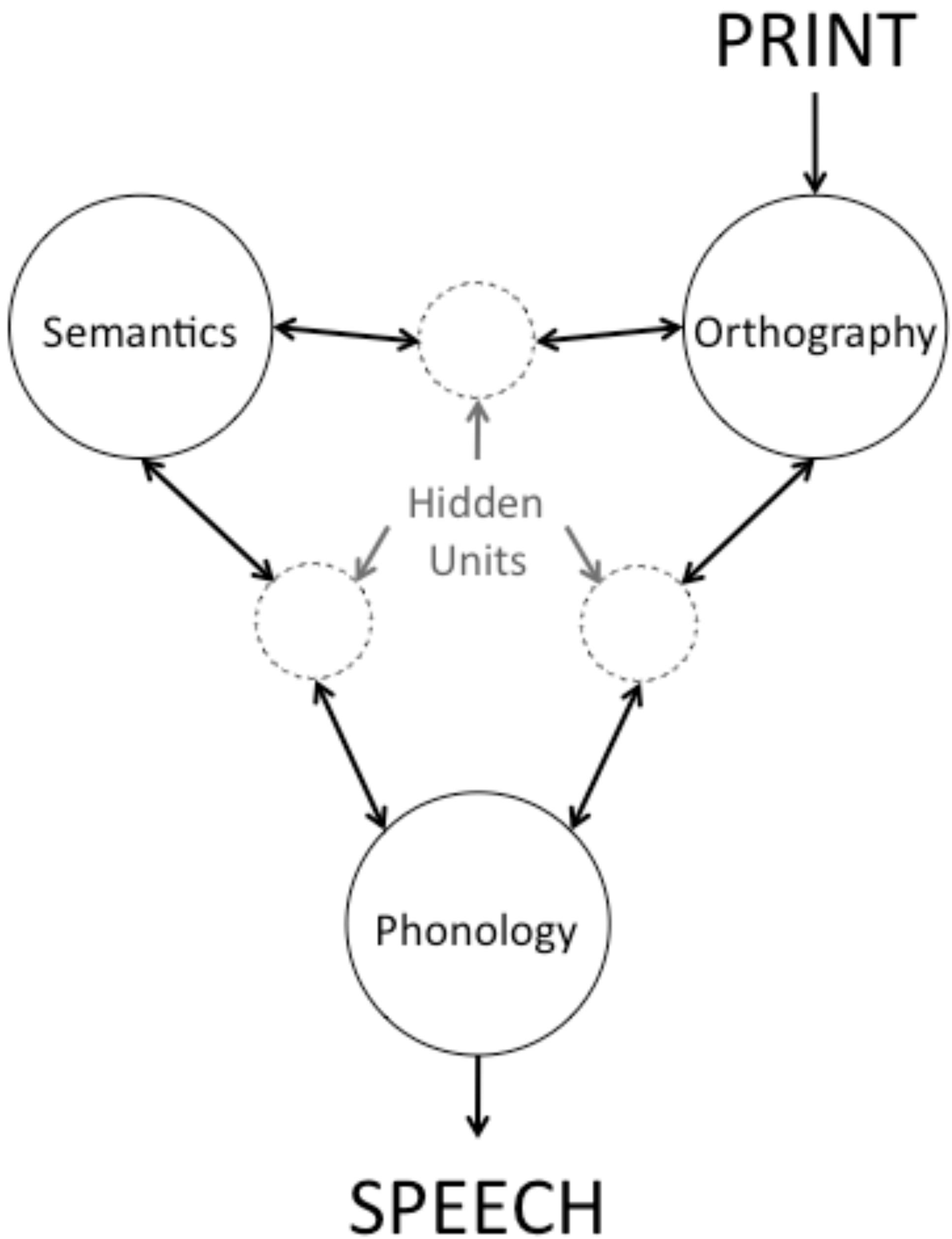

Figure 3. General structure of the Triangle class of models.

Current Models

Currently the field of reading aloud is dominated by two classes of models: dual-route models such as the Dual Route Cascaded model (DRC; Coltheart et al. 2001) and the Connectionist Dual Process model (CDP+; Perry et al., 2007), and models based on the Triangle assumption (Seidenberg & McClelland, 1989; Plaut et al., 1996; Harm & Seidenberg, 1999; Plaut & Booth, 2000; Woollams et al., 2007). These classes of models differ dramatically in both structure (or arrangement of modules; see Figure 2 for a depiction of dual-route architectures, and Figure 3 for the Triangle class of modules) and in how knowledge is represented.

Though the debate between these classes of models continues to rage (see for example the exchange between Patterson & Plaut, 2009; 2010; and Coltheart, 2010), the Triangle Model lags the dual-route models in terms of computational development, in that it does not simulate the same range of effects. In fact, there is no single implementation of the Triangle Model that incorporates all three modules fully inter-connected as the theory proposes (though the implementation by Harm & Seidenberg, 2004, comes close, lacking only the feedback from the phonological and semantic systems to the orthographic system). Consequently, it does not lend itself to further examination here. A second obstacle to testing the Triangle models is their availability: only Woollams et al. (2007) have made their model available on the Internet and it does not include a feedback component.

Dual-Route Models

The DRC and CDP+ models share much in common. Both models begin with a letter identification process (the letter level), which takes features as input and identifies each letter in the presented letter string in parallel. The letter level consists of a set of units that encode the features from one letter position in the string. At this point the, the letter level feeds activation forward in two directions. This "forking" of the processing follows in the steps of Morton's (1980) logogen model, where Morton argued that two independent routes are necessary to account for a double dissociation in acquired dyslexia: that some patients are able to read aloud words, but are impaired at reading aloud nonwords (phonological dyslexia), whereas others are able to read aloud nonwords but are impaired at reading exception words such as PINT (surface dyslexia). This dissociation is taken as evidence that there were at least two independent processes: one that could handle unfamiliar letter strings like FRANE, and another that could handle familiar letter strings such as PINT (but see Dunn & Kirsner, 1988). Both the DRC and CDP+ models trace their ancestry back to Morton's model, and propose two independent routes for converting letter strings into phonology.

Along the "lexical" route, the letter level activates an Orthographic Input Lexicon (OIL) which contains a single node storing the spelling of each word known to the network (e.g., F in the first letter position would activate OIL entries for FROG, FAST, FEAR, etc… and inhibit all other words). These OIL entries in turn feed activation back to the letter level (engaging in IA; e.g., FROG would activate F, R, O, and G in the first, second, third, and fourth letter positions respectively), and forward to the Phonological Output Lexicon (POL), which includes a single node for the pronunciation of each word known to the network. Here again, the model is engaged in IA with the POL feeding activation back to the OIL (e.g., FROG in the OIL would activate the /fɹɔg/[3] entry in the POL and vice-versa). Finally, the POL feeds activation to a Phoneme Buffer where pronunciations are stored until the model is ready to respond. The Phoneme Buffer in turn feeds activation back to the POL, providing the third (and final) locus of IA. This route is capable of reading aloud any letter string that is known to the network (e.g., FROG), but is unable to read aloud unfamiliar letter strings (such as FROP) as such letter strings do not have entries in the OIL. Both the DRC and CDP+ models include a semantic system in theory (see Figure 2), but neither implements semantics in their computational versions.

The lexical routes of both dual-route models (DRC and CDP+) are identical in structure and processing dynamics (though the values of the parameters differ; see Table 1). Where the models diverge significantly is in the operation of the sublexical route. In both the DRC and CDP+ models, the sublexical route takes an input string and converts the string serially into phonology, passing the results on to the Phoneme Buffer (where it is combined with the results from the lexical route to produce a final pronunciation).

The sublexical mechanism through which print is converted into sound differs dramatically between the two models. The DRC uses a set of modeler-specified rules to serially convert the string into phonological codes (e.g., a P in the first position activates /p/ in the first position in the Phoneme Buffer. If it is then followed by an H to produce PH, the sub-lexical route stops supporting the activation of /p/ in the first position, and begins to activate /f/ instead). In contrast to DRC's codified set of rules, CDP+ uses a neural network that has been trained on a combination of rules and the corpus of words in the network, instead of a set of codified rules. Through the learning process, the network discovers the relationships between letter combinations and pronunciations.

In both the DRC and CDP+, the sublexical route is capable of (indeed, necessary for) reading unfamiliar letter strings (e.g., FRAWG or FRAM). It is also capable of reading known words that match the expectations of the sublexical route (either the rules, or the learned associations, as appropriate) (e.g., `MINT`). Where the sublexical route has difficulty is in pronouncing known words that violate the rules (e.g., `PINT`). The sublexical route in DRC virtually never reads these “irregular” words aloud correctly, whereas CDP+’s sublexical route has some success with these words due to its experience with them during the training process. However, reading these words aloud correctly generally relies on the participation of the lexical routes in both models.

In Figure 2, IA is indicated by the double-headed arrows between the letter level and the OIL; the OIL and the POL; and the Phoneme Buffer and the POL. As indicated by the single-headed arrows, the sublexical route is engaged in feed-forward processing only, with no feedback connections from the sublexical route to the letter level, or from the Phoneme Buffer to the sublexical route. Though not clear from the figure, the neural network that underlies CDP+’s sublexical system also consists of only feed-forward processing (i.e., the simplified schematics do not mask any underlying feedback mechanisms).

Table 1. Parameters used by the default version of CDP+ [CDP+ (F), from Perry et al., 2007], the version of CDP+ with no feedback [CDP+ (NF)], and their equivalents in DRC 1.0 (from Coltheart et al., 2001). Bold values are those that differ from the default CDP+ (F) values.

Parameter	CDP+ (F)	CDP+ (NF)	DRC 1.0
Lexical Route			
Feature Level			
Feature-to-letter excitation	0.005	0.005	0.005
Feature-to-letter inhibition	-0.150	-0.150	-0.150
Letter Level			
Letter-to-letter inhibition	0.000	0.000	0.000
Letter-to-orthography excitation	0.075	0.075	**0.070**
Letter-to-orthography inhibition	-0.550	-0.550	**-0.435**
Orthographic Lexicon			
Orthography-to-orthography inhibition	-0.060	-0.060	-0.060
Orthography-to-phonology excitation	1.400	1.400	**0.200**
Orthography-to-letter excitation	0.300	**0.000**	0.300
Phonological Lexicon			
Phonology-to-phonology inhibition	-0.160	-0.160	**-0.070**
Phonology-to-phoneme excitation	0.128	0.128	**0.140**
Phonology-to-phoneme inhibition	-0.010	-0.010	**0.000**
Phonology-to-orthography excitation	1.100	**0.000**	**0.200**
Phonological Output Buffer			
Phoneme-to-phoneme inhibition	-0.040	-0.040	**-0.150**
Phoneme-to-phonology excitation	0.098	**0.000**	**0.040**
Phoneme-to-phonology inhibition	-0.060	**0.000**	**-0.160**
General Parameters			
Activation rate	0.200	0.200	0.200
Frequency scale	0.400	0.400	**0.050**
Phoneme naming activation criterion	0.670	0.670	**0.430**
Sublexical Parameters			
Cycles before route begins	0	0	**10**
Cycles before next letter accessed	15	15	**17**
Sublexical to phoneme activation	0.085	0.085	**0.055**
Letter level threshold for processing	0.210	0.210	**n/a**

Model Choice for Evaluation

The Triangle class of models is not suitable for assessing the utility of feedback for two reasons: most importantly, there is no single computational model that fully implements the theoretical account. Second, the only version of the model that includes feedback connections from the phonological system to the orthographic system (Plaut et al., 1996), is not available in a form that would allow removing feedback (see also the section entitled "Disabling Feedback" in the "General Discussion" for more on the challenges in conducting a similar study with the Triangle Model).

The original DRC model was published in 2001 (Coltheart et al., 2001), and has demonstrated a remarkable ability to simulate a wide range of experimental results. Since that initial version, DRC has undergone considerable assessment and revision, and version 1.0 is no longer readily available. It has been supplanted by two versions (DRC 1.1.4, and DRC 1.2), neither of which has been tested as extensively as DRC 1.0 in a peer-reviewed journal. Furthermore, it is unlikely that testing of these models is forthcoming, as Coltheart and colleagues are currently developing yet another version (DRC 2.0) that will have a very different sublexical process (M. Coltheart, personal communication, July 27, 2010).

Like the DRC, CDP+ (Perry et al., 2007) also simulates a wide range of experimental phenomena. Importantly, CDP+ is capable of simulating consistency effects (which DRC version 1.0 could not simulate), and it captures up to four times as much variance as DRC in large-scale item-level databases (Perry et al., 2007, Table 2). CDP+ is also readily available from the web (http://ccnl.psy.unipd.it/CDP.html) with an agreed upon default parameter set that has been subjected to peer review.[4] Though the two models (DRC and CDP+) continue to have a strong impact on research in reading aloud, because of CDP+'s availability, its performance at the item level, its ability to simulate some phenomena that DRC cannot, and DRC's ongoing development, this study will focus on CDP+ (see Figure 4).

A Simulation Approach

To examine the role that feedback plays in a computational model of reading aloud, I will compare the results from two different versions of CDP+. First, I will confirm the results reported in Perry et al. (2007) using the default parameter sets described in their article. For the most part, Perry et al. (2007) used the parameter settings from Table 1 [see the values for CDP+ (F)], though for a few simulations the parameters were adjusted to reflect special circumstance (e.g., simulating acquired dyslexia). These adjustments (along with any other deviations from the default parameters) will be clearly described in the text where relevant. The second version of the model will have the feedback parameters set to 0 (zero). The changes are indicated in bold in Table 1 [see the values for CDP+ (NF)]. The study will largely be concerned with comparing the performance of CDP+ with feedback (F) to its performance without feedback (NF).

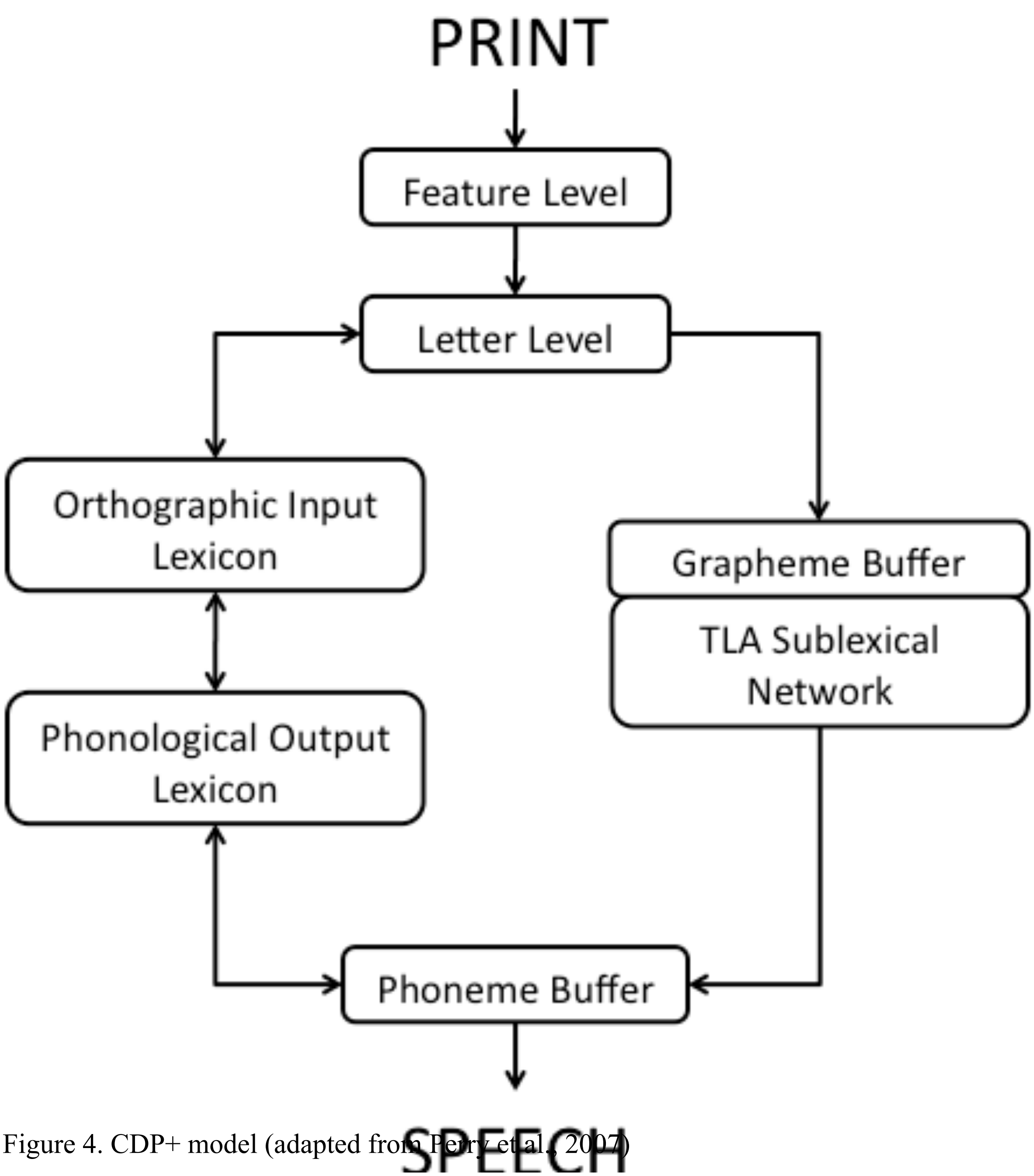

Figure 4. CDP+ model (adapted from Perry et al., 2007)

Nomenclature.

Though it is tempting to refer to the models as being either cascaded or IA versions of CDP+, this is not strictly accurate. Cascaded models can be considered a special case of all of the possible IA models in the following way: cascaded models are IA models in which the feedback and within-level inhibitory connections have been set to zero. The two model specifications that I will be comparing in this dissertation differ only in the presence or absence of feedback, and thus neither are consistent with the cascaded framework proposed by McClelland (1979; because they both preserve the within-level competitive connections that are not included in the seminal work).[5] Rather than confuse the issue by discussing the models in terms of cascaded vs. IA processing, I will describe the models with reference to the key distinction between them and use the terms CDP+ (F) for the default configuration that includes feedback, and CDP+ (NF) for the version in which the feedback has been disabled. This convention will emphasize that the key manipulation is the presence or absence of the feedback component of the default configuration.

Feedback's Contribution to Reading Aloud

The goal of this study is to evaluate the role that feedback plays in a model of reading aloud. Of principal concern is whether CDP+ will be as successful with no feedback present at all, or if there are any currently agreed upon benchmark performance measures that require feedback in the model. There are two reasons this issue deserves attention: the first has to do with the explanatory power of the models. There exists a large body of empirical results in which two experimental factors produce additive effects on reaction time (see Besner, 2006 for a brief summary of these results, and Besner, O'Malley, & Robidoux, 2010, O'Malley &

Besner, 2008, Besner & O'Malley, 2009, for more recent demonstrations of additivity). To date, no IA model has shown reliably additive effects of any two factors, and there is thus far no reason to believe that this is possible (Plaut & Booth, 2000, 2006, claimed to simulate additive effects of word frequency and stimulus quality, but see Besner, Wartak, & Robidoux, 2008, for a critique). On the other hand, cascade models, which do not include feedback or within-level inhibitory connections, have been shown to be mathematically capable of additive effects on mean reaction times, provided certain conditions are met (McClelland, 1979; Ashby, 1982; Roberts & Sternberg, 1993). Thus if there is no demonstrable need for feedback, removing it from activation based models increases the likelihood that they will eventually be successful at simulating additive effects of factors (though there remains the question of within-level inhibition). A second reason for limiting the role of feedback is that it reduces the complexity of the models. Eliminating feedback, entirely or in part, reduces the number of parameters that can be manipulated by researchers.

A Comment On Speed

In general, removing feedback slows the network somewhat. Because there is no a priori method of determining how fast the model should be, this general slowing is of no concern. If one wished to argue that under some circumstances feedback is disabled, and in other contexts it was enabled, then this generally slower performance without feedback would be important: it would imply that in conditions where there is no feedback, people should be slower than in conditions where feedback is operating. In the case of this study, I am not concerned with identifying conditions that require feedback vs. conditions that do not, but rather whether a model without feedback will be able to perform as well as a model with feedback. In such a comparison, the raw values are irrelevant: I need only show that the qualitative patterns match (I will also consider correlations that are insensitive to simple scaling and range effects).

Accuracy

The first, and most obvious benchmark for any model of reading aloud is simply its ability to read aloud the words it knows. Because the goal of this dissertation is to determine whether or not feedback is required in reading aloud, the first test is to confirm whether or not the model can read aloud accurately without feedback. To test the models' accuracy, I submitted the 7,383 unique words that are included in CDP+'s lexicon to both versions of the model [CDP+ (F) and CDP+ (NF)]. With feedback enabled [CDP+ (F)] the model correctly read aloud 7,280 of the 7,383 words (98.60% accurate[6]). With feedback removed [CDP+ (NF)], it read aloud 7,259 words accurately (98.32% accurate). Thus, with or without feedback, the model reads words aloud with a very high degree of accuracy. It should also be noted that the correlation in the cycle times from the two models was extremely high ($r = .982$, $N = 7,257$[7]), suggesting at the outset that feedback's role in reading words aloud is limited.

Skilled readers are also adept at reading aloud letter strings they have never seen before (e.g., nonwords such as `FRANE`). To test the model's ability to read nonwords, Perry et al. (2007) presented CDP+ with the 590 (misstated as 592) nonwords from Seidenberg, Plaut, Petersen, McClelland, and McRae (1994). Of the 590 items in the corpus, 11 are in fact words known to CDP+. For the 579 remaining nonwords, Perry et al. (2007) reported that CDP+ (F) made 37 errors. Determining accuracy when reading aloud nonwords is heavily influenced by the judgment of the person evaluating the model's pronunciations. These evaluations can be influenced by the evaluator's own vocabulary, and especially by his or her dialect and other linguistic influences.[8] As a result, rather than trying to match Perry et al.'s (2007) evaluations on all 579 relevant items, I will examine only those items that resulted in different pronunciations when feedback is removed from CDP+. Those items are summarized in Table 2.

Table 2 makes clear that CDP+'s performance *improves* for nonwords when feedback is removed. Of the 19 items that differed, 16 were errors in CDP+ (F), but only 4 were errors in CDP+ (NF). Thus CDP+ (NF) makes 12 fewer errors (25 errors, or 4.3%) than CDP+ (F) (37 errors, or 6.4%). In particular, CDP+ (NF) corrected a number of CDP+ (F)'s lexicalizations. This result suggests that removing feedback reduces the lexical system's influence in reading aloud of nonwords. Whatever the cause, there is little question that *accuracy in nonword reading favors the model with no feedback* over the one with feedback. As with words, the correlation between the cycle times of CDP+ (F) and CDP+ (NF) was very high (r = .952, N = 579).

Table 2. Seidenberg et al. (1994) items that CDP+ read aloud differently as a function of the presence (F) or absence (NF) of feedback.

Item	CDP+ (F)	CDP+ (F) Result	CDP+ (NF)	CDP+ (NF) Result
dront	dɹant	Correct	dɹɔnt	Correct
smalse	smo:ls	Correct	smæls	Correct
smead	smed	Correct	smi:d	Correct
fiek	fi:kt	Error	fi:k	Correct
freamt	fɹɔmt	Error	fɹi:mt	Correct
grend	gɹendz	Error	gɹend	Correct
moax	məʉkst	Error	məʉks	Correct
silm	stɪlm	Error	sɪlm	Correct
glebt	gleb	Error	gle	Error
bimpse	blɪmps	Lexicalized (Error)	bɪmps	Correct
foun	fæɔnd	Lexicalized (Error)	fæɔn	Correct
fren	fɹendz	Lexicalized (Error)	fɹen	Correct
poin	poɪnt	Lexicalized (Error)	poɪn	Correct
thout	ðæt	Lexicalized (Error)	ðæɔt	Correct
glarc	gla:s	Lexicalized (Error)	gla:	Error
larf	la:fs	Lexicalized (Error)	la:f	Lexicalized (Correct)
toal	təʉld	Lexicalized (Error)	təʉl	Lexicalized (Correct)
fache	fækt	Lexicalized (Error)	fæɪ	Lexicalized (Error)
rould	ɹæɔnd	Lexicalized (Error)	ɹæɔd	Lexicalized (Error)

Item-Level Variance

Perry et al. (2007) made much of the proportion of variance in human reading aloud times that is capture by the CDP+ model:

> The data showed that CDP+ was far superior to all of its competitors in predicting item-level variance... We consider this to be a major advancement in the area of computational modeling of reading aloud. (p. 292)

To determine the proportion of variance captured by CDP+, Perry et al. (2007) correlated model cycle times with the human reading aloud reaction times (RTs) collected in studies involving large corpora of words. The higher the correlation, the more of the variance in human reading aloud RTs the model is able to explain, ignoring item-specific characteristics or other experimental manipulations. To provide a thorough examination of this benchmark, Perry et al. (2007) compared CDP+'s performance to those of human subjects in four large scale databases: Spieler and Balota (1997), Balota and Spieler (1998), Seidenberg and Waters (1989), and Treiman, Mullennix, Bijeljac-Babic, and Richmond-Welty (1995). In addition to these four databases, I compare CDP+'s performance to the human latencies recorded in the English Lexicon Project (ELP; Balota et al., 2007). The ELP contains a much larger set of items than the other databases used by Perry et al. (2007), and includes reaction time data from hundreds of subjects at six different universities. Thus, ELP should provide the most robust test of item-level benchmarks. Table 3 summarizes the results. For four of the five corpora, removing feedback reduced the proportion of variance explained by a small amount (never more than 1.28 percentage points). In the case of Spieler and Balota (1997), CDP+ (NF) captured slightly more variability than CDP+ (F).

Table 3. Proportion of item-level variance explained in corpora from several large-scale reading aloud studies.

Database	Corpus size[a]	CDP+ (F)	CDP+ (NF)
Spieler and Balota (1997)	2,803	17.27%	17.46%
Balota and Spieler (1998)	2,803	21.59%	20.79%
Seidenberg and Waters (1989)	1,278	9.72%	8.60%
Treiman et al. (1995)	1,276	15.61%	15.27%
ELP (Balota et al., 2007)	5,554	21.60%	20.32%

a. Corpus size here refers to the number of words in the corpus that are also in CDP+'s lexicon.

Limitations of Item-Level Variance as a Benchmark

Perry et al. (2007) argued strongly that the item-level variance benchmark is key for comparing the performance of models of reading aloud: “Setting aside the (often small) discrepancies among models in accounting for specific experimental findings, we agree with Spieler and Balota (1997) that the issue of item variance is the most critical challenge faced by computational models of reading aloud.” (p. 292) There is also little doubt that CDP+ outperformed its competitors on this metric. Compared to DRC (Coltheart et al., 2001), the Triangle model (Plaut et al., 1996), and CDP (Zorzi, Houghton, & Butterworth, 1998), CDP+ explained three to four times as much item-level variance. However, there are two reasons to exercise caution in relying on item-level variance in comparing models.

The first concern is clearly evident in the corpora employed here. The Seidenberg and Waters (1989) and Treiman et al. (1995) studies are based on the same set of items (other than two items not used by Treiman et al., 1995) but resulted in very different estimates of the variance explained by CDP+ (with a 5.9% difference between the two samples). This suggests that there is considerable variance between reported samples[9], and indeed the two human samples are not strongly correlated (r = .394, N = 1,276). Similarly, the Spieler and Balota (1997) and Balota and Spieler (1998) studies used identical item sets, in the same lab, using the same technology, but still produced quite different estimates of the variance explained by CDP+ (with a difference of 4.32% between the samples; correlation between the two samples of human RTs: r = .653, N = 2,819). Though large improvements in the proportion of variance explained by the models is unlikely to be the result of simple inter-sample variance, smaller changes such as the differences in performance between CDP+ (F) and CDP+ (NF) should not be weighted too heavily in comparing models.

A second limitation to the use of item-level databases is the numerous demonstrations in the literature of context influencing reading aloud times in humans. For example, Lupker, Brown, and Colombo (1997) compared blocked and mixed designs and demonstrated that when an experiment intermixes items that differ in their typical response times (e.g., high and low frequency items), the slower items get faster and the faster items get slower when compared to those same items in blocked designs. Thus, the response time to any given item is influenced by the other items that are used in the experimental block. More dramatically, context can also produce *qualitative* changes in the pattern of RTs. For example, O'Malley and Besner (2008) demonstrated that when only words are present in an experiment, word frequency and stimulus quality produce an interaction so that the word frequency effect is larger for dim stimuli than for bright stimuli. However, when nonwords are introduced to the experiment, the interaction disappears for those same word items, so that both bright and dim stimuli show the same magnitude of word frequency effect. Similarly, Besner et al. (2010) demonstrated an influence of the presence of nonwords on the interaction between stimulus quality and regularity (the two factors were additive when nonwords were present, but underadditive so that dim words showed a *smaller* regularity effect than bright words when only words were present), and Ferguson, Robidoux, and Besner (2009) demonstrated that relatedness proportion influences the interaction between stimulus quality and semantic priming so that when related proportion is .5 the two factors are overadditive in reading aloud, but when the proportion is only .25 they are additive (see Stolz & Neely, 1995, for a demonstration of the same pattern in lexical decision). Thus far stimulus quality effects have not been treated as benchmarks by any modeler of reading aloud (but see the exchange between Ziegler, Perry, & Zorzi, 2009 and Besner & O'Malley, 2009; and the "General Discussion" section of this study), but Perry et al. (2007) do specify that base word frequency effects in pseudohomophone reading should represent a benchmark, and researchers have found the presence or absence of these effects to be contextually-dependent (see the section entitled "Pseudohomophones and base word frequency."). Assessments of item-level variance based on large-scale corpora like those examined here will not capture these effects until the corpora themselves include the relevant manipulations.

The Role of Feedback and Parameter Selection

Regardless of the general limitations of item-level databases, it cannot be denied that removing feedback reduced the amount of variance captured by CDP+ in four of the five corpora. There are two reasons I do not believe this provides compelling evidence in favor of the need for feedback. The first is simply that the largest decrease (1.28% in the ELP corpus) is considerably smaller than the between-sample variance observed with both the Spieler and Balota (1997) item set (4.32% difference between Spieler & Balota, 1997 and Balota & Spieler, 1998), and the Seidenberg and Waters (1989) items (5.89% difference between Seidenberg & Waters, 1989, and Treiman et al., 1995). This suggests that CDP+ (NF)'s performance is subsumed within the expected error variance of any given corpus.

A second reason for disregarding the small performance cost of removing feedback is that the parameters in the model are inter-dependent, meaning that dramatic adjustments to one or more parameters can be expected to necessitate compensatory adjustments to other parameters in order to maintain identical performance levels. Removing feedback is a dramatic adjustment to the model; one that likely alters the lexical route's influence in reading. It seems likely, then, that restoring the balance of sublexical and lexical contributions to reading aloud would necessitate changing other parameters in the model (see the later section titled "Pseudohomophone advantage." for further evidence that such an adjustment may be necessary). No such corrections were made in the present analysis of item-level variance, and thus the extremely small changes in the amount of variance explained by the model seem to me to be of little consequence. I now turn to the experimental benchmarks identified by Perry et al. (2007).

The Benchmark Phenomena

Perry et al. (2007) proposed a list of benchmark phenomena that successful models of reading aloud must be able to simulate (see Table 4). These benchmarks are a combination of benchmarks taken from Coltheart et al.'s (2001) simulations with DRC and some new phenomena that CDP+ is capable of simulating. These phenomena will form the basis of the present investigation into the need for feedback in reading aloud.

Data Trimming

In all of the simulations involving RT, errors were removed from the analysis. For experiments involving words, items that were unknown to CDP+ were removed prior to analysis because they are effectively nonwords to the model. Similarly, for experiments involving nonwords, any items that appeared as words in CDP+'s lexicon were removed. After error and "nonword/word" removal, the data were subjected to an outlier trimming procedure that removed any items producing cycle times more than three standard deviations away from the cell means. The results of this trimming procedure are not discussed except where they produced different data sets for the two model versions [CDP+ (F) vs. CDP+ (NF)]. Correlations between individual model latencies and human reading aloud RTs considered all items included in the analysis for that model. Correlations between the latencies of the two model versions considered only items that were present in the analyses of both models.

Table 4. Benchmark phenomena in reading aloud (adapted from Perry et al., 2007). Datasets that are simulated are indicated in parentheses.

Effect	Pattern
Frequency	High frequency words are read aloud faster than low frequency words. (See later sections)
Lexicality	Words are read aloud faster than nonwords. Pseudohomophones are read aloud faster than other nonwords. (See later sections)
Frequency by regularity	Low frequency regular words are faster than low frequency irregular words. There is no regularity effect for high frequency words. (Paap & Noel, 1991; Jared, 2002)
Word consistency	Words with many consistent neighbors are read aloud faster than words with few consistent neighbors if the summed frequency of the consistent neighbors (friends) is larger than the summed frequency of the inconsistent neighbors (enemies). (Jared, 2002)
Regularity by position	The regularity effect only applies for words where the irregularity is early in the letter string. (Coltheart & Rastle, 1994; Roberts, Rastle, Coltheart & Besner, 2003)
Nonword consistency	Nonwords that have regular word-neighbors are more likely to be regularized than nonwords that have no regular word-neighbors. (Andrews & Scarratt, 1998)
Length by lexicality	Nonwords show effects of letter length so that longer nonwords are slower to read aloud. Words do not show this effect. (Weekes, 1997; Ziegler, Perry, Jacobs, & Braun, 2001)
Body neighborhood	Larger body neighborhoods speed response times in reading aloud for both words and nonwords. (Ziegler et al., 2001)
Pseudohomophone advantage	Pseudohomophones are faster to read aloud than nonword controls when all items are intermixed. (McCann & Besner, 1987)
Pseudohomophone base word frequency	The presence/absence of a base word frequency effect depends on the context in which the pseudohomophones are read. (McCann & Besner, 1987)
Masked priming	Significant effects of both onset priming (LEG-LOW) and rhyme priming (TOE-LOW) relative to unrelated primes (RUN-LOW). (Forster & Davis, 1991)
Surface dyslexia	Patients MP & KT show impairment when reading irregular words with more enemies than friends (but not regular words or nonwords). (Patterson & Behrmann, 1997; Taraban & McClelland, 1987)
Phonological dyslexia	Patient LB is impaired at reading nonwords but not words. The impairment is weaker for pseudohomophones. (Derouesné & Beauvois, 1985)
Orthographic N & Phonological N	Words and nonwords with many orthographic neighbors are read aloud faster. For words this is actually a phonological N effect. (Andrews, 1989; 1992; Mulatti, Reynolds, & Besner, 2006)
Whammies	Nonwords with multi-letter graphemes (e.g., PH) are slower to read aloud than controls. (Rastle & Coltheart, 1998)

Evaluating the Benchmarks

The structure for evaluating each of the benchmarks will be as follows: First, I replicate the result reported in Perry et al. (2007) where the feedback is left intact [CDP+ (F) simulations], and comment on any discrepancies. Second, I reproduce the simulations after disabling the feedback processes [CDP+ (NF) simulations]. To disable the feedback processes, I simply set the strength of the feedback connections (from Phoneme Buffer to POL, POL to OIL, and OIL to letter level) to 0, eliminating their influence on the model (see Table 1). All other parameters were left unchanged from the defaults, unless otherwise specified.

In the case of reaction time studies, analysis will emphasize both the underlying qualitative data patterns (significance, non-significance), and the correlation between the cycle times of CDP+ (F) and CDP+ (NF). Very high correlations between the two versions of the model suggest that the influence of feedback on model cycle times is negligible after simple scaling. In most of the simulations, the correlation between the cycle times of the two versions of the model exceeded .99 (exact values are reported for each simulation).

In studies considering accuracy or the proportion of responses of different types as the dependent variable, analysis will focus on changes in pronunciation between the two models. If removing feedback has little effect on the model's ultimate pronunciations, then there is little reason to prefer the feedback version (F) to the no-feedback version (NF). If removing feedback results in many pronunciation differences, it requires a closer look at the patterns to determine which model best described the data (it turns out this is not the case with any of the benchmarks examined here). To preview the results, all of the benchmarks can be simulated by the model whether there is feedback present or not, though the pseudohomophone advantage required some further adjustments to the model's other parameters.

Word frequency.

Perry et al. (2007) did not evaluate CDP+'s ability to produce a word frequency effect in isolation. Rather they demonstrated a word frequency effect as part of simulating other phenomena. Simply described, subjects are faster and more accurate at reading aloud words that occur frequently in print (high-frequency words) compared to words that occur infrequently (low-frequency words). I refer the reader to the sections entitled "Frequency and frequency by regularity.", "Frequency, regularity, and consistency.", and "Lexicality by letter length (by frequency)." for demonstrations that both CDP+ (F) and CDP+ (NF) produce robust word frequency effects. Thus, the word frequency benchmark does not require feedback in order to be simulated.

Regularity.

As with frequency, regularity too was a factor in other simulations. Briefly, it is well established that words that are irregular or exceptional (i.e., that do not follow the typical rules of pronunciation; e.g., `PINT`) are read aloud more slowly than regular words (e.g., `MINT`). This effect is generally only observed for words that are low in printed frequency. I refer the reader to the sections entitled "Frequency and frequency by regularity.", "Frequency, regularity, and consistency.", and "Regularity by position." for demonstrations that the regularity effect does not require feedback in order to be simulated.

Frequency and frequency by regularity.

It is well known that word frequency and regularity each influence response latencies in reading aloud such that responses to high frequency words are faster (and often more accurate) than to low frequency words, and regular words (such as MINT) enjoy a similar advantage over words with unusual pronunciations (e.g., PINT). Paap and Noel (1991) reported that the regularity effect is modulated by the frequency of the words. Low frequency words yield an effect of regularity, whereas high frequency words do not. Perry et al. (2007) submitted the Paap and Noel (1991) stimuli to the CDP+ model to test its ability to simulate these effects.

CDP+ (F).

CDP+ (F) simulated the Paap and Noel (1991) result perfectly (see Table 5). It produced significant effects of word frequency ($F(1, 72) = 124.583$, $MSE = 80.7$, $p < .001$) and regularity ($F(1, 72) = 18.351$, $MSE = 80.7$, $p < .001$). Most importantly, it produced a significant interaction of frequency and regularity ($F(1, 72) = 11.022$, $MSE = 80.7$, $p < .01$). Tests of the simple effects revealed that the regularity effect was significant for low frequency words ($t(34) = 5.135$, $p < .001$), but not for high frequency words ($t < 1$).

CDP+ (NF).

With feedback removed, CDP+ (NF) also produced the correct pattern (see Table 5). There were significant effects of word frequency ($F(1, 72) = 143.264$, $MSE = 83.7$, $p < .001$) and regularity ($F(1, 72) = 21.571$, $MSE = 83.7$, $p < .001$). Most importantly, there was a significant interaction of frequency and regularity ($F(1, 72) = 10.367$, $MSE = 83.7$, $p < .01$). As with CDP+ (F), tests of the simple effects revealed that the regularity effect was significant for low frequency words ($t(34) = 5.627$, $p < .001$), but not for high frequency words ($t(38) = 1.141$).

Table 5. Mean cycles to response (percentage errors) to the Paap and Noel (1991) items for CDP+ (F) and CDP+ (NF), as a function of regularity and word frequency.

	Regular		Exception	
Word Frequency	Cycles	%E	Cycles	%E
CDP+ (F)				
Low Frequency	97.57	(0%)	113.65	(0%)
High Frequency	80.95	(0%)	83.30	(0%)
Difference	*16.62*	*(0%)*	*30.35*	*(0%)*
CDP+ (NF)				
Low Frequency	103.63	(0%)	120.53	(0%)
High Frequency	84.60	(0%)	87.95	(0%)
Difference	*9.03*	*(0%)*	*32.58*	*(0%)*

Model comparison.

Removing the feedback had no effect on errors or outliers, nor on CDP+'s ability to simulate the Paap and Noel (1991) data pattern. Indeed, the correlation between the cycle times produced by CDP+ (F) and CDP+ (NF) was extremely high ($r = .992$, N = 76).

Frequency, regularity, and consistency.

The debate about whether or not the observed regularity effect is the result of regularity (the violation of pronunciation rules) or consistency (whether or not a word's body is pronounced the same way in all of its neighbors) has raged for over 30 years (see Glushko, 1979). It is a difficult problem to disentangle, because all exception words (`PINT`) are, by definition, inconsistent. However, some regular words are consistent and others are not (`HUNT` vs. `HINT`). To delineate consistency from regularity, Jared (2002) constructed lists of test items and control words (such as `BLOWN` [inconsistent exception] vs. `BARGE` [consistent regular], and `BROOD` [inconsistent regular] vs. `BRIBE` [consistent regular]) that allowed disentangling the two closely related concepts. She argued that the regularity effect is, in fact, a consistency effect (a word like `HUNT` is consistent, because all –`UNT` words share the same body pronunciation, whereas `HINT` is not consistent, because at least one of its neighbors, `PINT`, does not share its body pronunciation). Furthermore, she found that this consistency effect is modulated by the ratio of the *summed frequency* of the "friends" (word neighbors that share the word-body pronunciation of the target) to the *summed frequency* of the "enemies" (word neighbors that have alternate word-body pronunciations).[10] For example, `BEAK` has higher frequency friends (`BLEAK`, `CREAK`, `FREAK`, `LEAK`, `PEAK`, `SNEAK`, `SPEAK`, `STREAK`, `TEAK`, `WEAK`, `WREAK`) than enemies (`BREAK`, `STEAK`), whereas `BEAD` has higher frequency enemies

(BREAD, DEAD, DREAD, HEAD, LEAD, READ, SPREAD, STEAD, THREAD, TREAD) than friends (KNEAD, MEAD, LEAD, PLEAD, READ).

In a pair of experiments manipulating frequency, regularity, consistency, and the ratio of friends to enemies (F-E ratio: higher summed frequency of friends than enemies vs. lower summed frequency of the friends than enemies), Jared (2002) reported the following: a significant consistency effect, so that words that have only consistent body-pronunciation neighbors are responded to faster to than words that have neighbors that do not share the pronunciation; there is little evidence of an additional regularity effect when this consistency is controlled for; the consistency effect is modulated by the F-E ratio so that words with a *higher summed frequency* of friends than enemies do not show consistency effects, whereas words with a *lower summed frequency* of friends than enemies do show the consistency effect; and finally, that the regularity by frequency interaction (see previous discussion of Paap and Noel, 1991) is itself modulated by the F-E ratio so that only low frequency words with *lower* summed frequency friends than enemies show a regularity effect.

CDP+ (F) – Jared (2002) experiment 1.

In her first experiment, Jared (2002) reported significant effects of consistency and F-E ratio, and also a significant interaction between consistency and F-E ratio, but no influence of regularity. CDP+ correctly simulated this pattern, producing significant effects of consistency ($F(1, 147) = 17.480$, $MSE = 72.1$, $p < .001$), F-E ratio ($F(1, 147) = 9.986$, $MSE = 72.1$, $p < .01$), and a significant consistency x F-E ratio interaction ($F(1, 147) = 5.109$, $MSE = 72.1$, $p < .05$). Type of inconsistency (regular inconsistent vs. irregular) was not significant, nor did it interact with other effects.

T-tests revealed a complicated interplay between the various factors. For inconsistent exception words, items with *lower summed frequency* friends than enemies were slower than matched controls ($t(35) = 3.080$, $p < .01$), whereas items with *higher summed frequency* friends than enemies were not significantly slower than matched controls ($t(37) = 1.742$, $p = .09$). A similar pattern for inconsistent *regular* words emerged: items with *lower frequency* friends than enemies were slower than matched controls ($t(38) = 2.890$, $p < .01$), whereas items with *higher frequency* friends than enemies were not significantly slower than matched controls ($t < 1$, ns). These patterns matched the human data reported by Jared (2002) nearly perfectly.

CDP+ (NF) – Jared (2002) experiment 1.

Removing feedback produced one additional error (`PINT` regularized to rhyme with `MINT`). Other than this minor difference, the qualitative pattern matched the results from CDP+ (F), producing significant effects of consistency ($F(1, 146) = 17.480$, $MSE = 72.1$, $p < .001$), F-E ratio ($F(1, 146) = 9.986$, $MSE = 72.1$, $p < .01$), and a significant consistency x F-E ratio interaction ($F(1, 146) = 5.109$, $MSE = 72.1$, $p < .05$). Type of inconsistency (regular inconsistent vs. irregular) was not significant, nor did it interact with other effects.

The pattern of simple effects also matched that of CDP+ (F). For inconsistent exception words, items with *lower summed frequency* friends than enemies were slower than matched controls ($t(34) = 3.122$, $p < .01$), whereas items with *higher summed frequency* friends than enemies were not significantly slower than matched controls ($t(37) = 1.691$, $p = .10$). A similar pattern for inconsistent regular words emerged: items with *lower summed frequency* friends than enemies were slower than matched controls ($t(38) = 3.075$, $p < .01$), whereas items with *higher summed frequency* friends than enemies were not significantly slower than matched controls ($t < 1$, ns).

Model comparison.

Though the pattern described by Jared (2002) is complicated, the key question here is how feedback influences CDP+'s performance. The results of the simulations indicate that feedback had no influence on CDP+'s performance with this benchmark. The default model [CDP+ (F)] produced two errors (`DOUGH` and `TRAIT`) and one outlier (`HASTE`). The model with feedback removed [CDP+ (NF)] produced one additional error (`PINT`). However, this new error had no effect on CDP+'s ability to simulate Jared's (2002) Experiment 1. As with the simulations of Paap and Noel (1991), the correlation between the cycle times for CDP+ (F) and CDP+ (NF) was extremely high (r = .993, N = 154). Table 6 summarizes the simulation results for Jared's experiment 1.

Table 6. Mean cycles to response (percentage errors) to the Jared (2002) Experiment 1 items for CDP+ (F) and CDP+ (NF).

	Exception/Inconsistent		Control	
Condition	Cycles	%E	Cycles	%E
CDP+ (F)				
Exception (F < E)	112.83	(5.2%)	102.42	(0.0%)
Exception (F > E)	105.05	(0.0%)	100.47	(5.0%)
Regular Inconsistent (F > E)	108.55	(0.0%)	101.25	(0.0%)
Regular Inconsistent (F < E)	101.42	(0.0%)	100.60	(0.0%)
CDP+ (NF)				
Exception (F < E)	120.24	(10.5%)	108.47	(0.0%)
Exception (F > E)	111.65	(0.0%)	106.84	(5.0%)
Regular Inconsistent (F > E)	116.20	(0.0%)	107.40	(0.0%)
Regular Inconsistent (F < E)	107.74	(0.0%)	106.65	(0.0%)

CDP+ (F) – Jared (2002) experiment 2.

In her second experiment, Jared (2002) extended the Paap and Noel (1991) results to examine the role of F-E ratio. This experiment added high frequency words (not present in Experiment 1) and manipulated frequency, regularity, and F-E ratio. She reported a significant regularity effect; *only a marginal* regularity x frequency interaction; and a marginal three-way interaction between regularity, frequency, and F-E ratio. Though Jared (2002) did not report the statistics, visual inspection of her Figure 2 (p. 733) strongly suggests a significant effect of frequency. In a series of planned tests, Jared reported that items that had lower frequency friends than enemies showed a regularity effect, and that this was true of both high and low frequency words. She also reported a regularity effect for high frequency words with higher frequency friends than enemies, and a marginal regularity effect for low frequency words with lower frequency friends than enemies.

CDP+ matched Jared's (2002) human data in that it produced significant effects of regularity ($F(1,145) = 19.5$, $MSE = 73.6$, $p < .001$) and frequency ($F(1,145) = 146.8$, $MSE = 73.6$, $p < .001$), and no interaction between regularity and frequency [F<1; in Jared's Experiment 2, this interaction was marginal whereas CDP+ (F) did not produce even a hint of it]. It differed from Jared's data, however, in that it produced a marginal effect of F-E ratio ($F(1,145) = 3.6$, $MSE = 73.6$, $p = .06$), and a significant regularity by F-E ratio interaction ($F(1,145) = 5.2$, $MSE = 73.6$, $p < .05$). These data are summarized in Table 7.

The key test of CDP+'s ability to simulate this pattern rested in the planned tests, where it closely matched the results reported by Jared (2002). CDP+ produced significant regularity effects for all words with lower frequency friends than enemies (high frequency words:

$t(37) = 2.748$, $p < .01$; low frequency words: $t(35) = 3.080$, $p < .01$), and a marginal regularity effect for low frequency words with higher frequency friends than enemies ($t(37) = 1.742$, $p < .10$). It failed, however, to produce the significant regularity effect for high frequency words with higher frequency friends than enemies that Jared reported ($t(37) = 1.151$, $p > .25$).

CDP+ (NF) – Jared (2002) experiment 2.

CDP+'s ability to simulate Jared's (2002) experiment 2 was unaffected by the removal of feedback. As with experiment 1, CDP+ (NF) produced one additional error (`PINT`). It produced significant effects of regularity ($F(1,144) = 21.5$, $MSE = 89.1$, $p < .001$) and frequency ($F(1,144) = 147.6$, $MSE = 89.1$, $p < .001$), and no interaction between regularity and frequency (F<1). As with CDP+ (F), CDP+ (NF) differed from Jared (2002) in that it produced a marginal effect of F-E ratio ($F(1,144) = 3.55$, $MSE = 89.1$, $p = .06$), and a significant regularity by F-E ratio interaction ($F(1,144) = 5.76$, $MSE = 89.1$, $p < .05$).

The planned tests also corresponded perfectly to those of CDP+ (F). CDP+ (NF) produced significant regularity effects for all words with lower frequency friends than enemies (high frequency words: $t(37) = 3.060$, $p < .01$; low frequency words: $t(34) = 3.122$, $p < .01$), and a marginal regularity effect for low frequency words with higher frequency friends than enemies ($t(37) = 1.691$, $p < .10$). It too failed to produce the significant regularity effect for high frequency words with higher frequency friends than enemies ($t(37) = 1.478$, $p > .14$).

Model comparison.

The default model produced two errors (`DOUGH` and `TRAIT`) and one outlier (`BREADTH`). As with experiment 1, CDP+ (NF) produced one additional error (`PINT`). Critically, this error had little influence on the model's ability to simulate Jared (2002)'s

Experiment 2. The correlation between the response latencies of the two models was extremely high ($r = .992$, N = 152).

Table 7. Mean cycles to response (percentage errors) to the Jared (2002) Experiment 2 items for CDP+ (F) and CDP+ (NF).

	Exception		Control	
Word Frequency	Cycles	%E	Cycles	%E
CDP+ (F)				
Low Frequency (F < E)	112.83	(5.2%)	102.42	(0.0%)
Low Frequency (F > E)	105.05	(0.0%)	100.47	(5.0%)
High Frequency (F > E)	92.58	(0.0%)	84.50	(0.0%)
High Frequency (F < E)	88.74	(0.0%)	87.37	(0.0%)
CDP+ (NF)				
Low Frequency (F < E)	120.24	(10.5%)	108.47	(0.0%)
Low Frequency (F > E)	111.65	(0.0%)	106.84	(5.0%)
High Frequency (F > E)	98.53	(0.0%)	88.40	(0.0%)
High Frequency (F < E)	93.79	(0.0%)	91.42	(0.0%)

Regularity by position.

Coltheart and Rastle (1994), Rastle and Coltheart (1999), and Roberts, Rastle, Coltheart, and Besner (2003) have all demonstrated that not all irregularities are equal. The regularity effect is generally attributed to competition between the lexical and sublexical routes; for irregular words like PINT, the two routes arrive at different pronunciations that must be reconciled before a response can be produced (see Robidoux & Besner, 2010, for a more thorough examination of this view). This process of resolving conflict leads to a slower response time. However, in order for competition to arise, the sublexical route must encounter the irregularity before the lexical route is able to complete processing. Because the sublexical route generates phonology serially from left-to-right, irregularities at the beginning of words are more likely to be encountered before the lexical route has produced the correct pronunciation. Later irregularities are unlikely to be encountered soon enough to produce the competition necessary for the regularity effect. This view has been confirmed in several experiments (Coltheart & Rastle, 1994; Rastle & Coltheart, 1999; and Roberts et al., 2003). In the most carefully controlled of these experiments, Roberts et al. (2003) tested words with irregularities in the 2nd phonemic position (e.g., CHAFF) to words with irregularities in the 3rd position (e.g., PLAID) and found that only 2nd position irregularities produced a response time cost when compared to matched controls. Roberts et al. (2003)[11] partially simulated this effect with the DRC model: after controlling for letter length and neighborhood size, they reported a significant regularity by position interaction so that the regularity effect was weaker for late irregularities than for early irregularities. However, the effect was still significant for late irregularities, which is inconsistent with the human data. They also found that the CDP model (a precursor to CDP+ that did not include any serial processing; Zorzi et al., 1998) could not correctly simulate this effect.

CDP+ (F).

CDP+ (F) produced one error (DOUCHE) and no outliers.[12] It produced a robust effect of regularity ($F(1, 97) = 39.629$, $MSE = 81.7$, $p < .001$), no effect of position ($F < 1$), and no evidence for the regularity x position interaction ($F(1, 97) = 1.134$, $MSE = 81.7$, $p > .20$). Despite the lack of a significant interaction, there was a trend towards a weaker effect for the late irregularity items (see Table 8). Planned t-tests found significant regularity effects for both the 2nd position items ($F(1, 63) = 31.872$, $p < .001$) and 3rd position items ($F(1, 32) = 8.440$, $p < .01$).

CDP+ (NF).

With feedback removed, CDP+ also produced one error (DOUCHE) and no outliers. CDP+ (NF) produced a pattern identical to the one produced by CDP+ (F) (see Table 8): a robust regularity effect ($F(1, 97) = 39.975$, $MSE = 107.4$, $p < .001$), no effect of position ($F < 1$), and no evidence of a regularity by position interaction ($F < 1$). As with CDP+ (F), planned t-tests revealed regularity effects for 2nd position items ($F(1, 63) = 32.576$, $p < .001$) and 3rd position items ($F(1, 32) = 8.243$, $p < .01$), and the same trend toward weaker effects for 3rd position items produced by CDP+ (F).

Table 8. Mean cycles to response (percentage errors) to the Roberts et al. (2003) items as a function of regularity and position of the irregularity, for CDP+ (F), CDP+ (NF), and DRC both with and without feedback.

	Position of the Irregularity			
	Position 2		Position 3	
Regularity	Cycles	%E	Cycles	%E
CDP+ (F)				
Exception	118.73	(2.9%)	113.89	(0.0%)
Regular control	105.26	(0.0%)	106.83	(0.0%)
Difference	*13.47*	*(2.9%)*	*7.06*	*(0.0%)*
CDP+ (NF)				
Exception	126.88	(2.9%)	123.00	(0.0%)
Regular control	111.82	(0.0%)	114.06	(0.0%)
Difference	*15.06*	*(2.9%)*	*8.94*	*(0.0%)*
DRC (default)				
Exception	90.32	(0.0%)	79.72	(0.0%)
Regular control	78.26	(0.0%)	77.56	(0.0%)
Difference	*12.06*	*(0.0%)*	*2.16*	*(0.0%)*
DRC (feedback disabled)				
Exception	94.50	(0.0%)	88.00	(0.0%)
Regular control	85.21	(0.0%)	85.67	(0.0%)
Difference	*9.29*	*(0.0%)*	*2.33*	*(0.0%)*

Model comparison.

Though CDP+ (F) underperforms DRC slightly, in that it does not produce significant interactions of regularity and position, it does show the correct trend for the interaction and matches DRC's performance otherwise. The main concern here, though, is whether or not the presence or absence of feedback influences CDP+'s performance. CDP+ (NF) performed nearly identically to CDP+ (F). Unsurprisingly, the correlation between the latencies in the two models was again extremely high ($r = .993$, N = 103) for this dataset.

DRC with feedback and without.

Because DRC 1.0 more accurately captured this effect (Coltheart et al., 2001), showing a significant regularity by position interaction (though it too shows a significant effect for 3rd position irregular items), here I examine whether DRC 1.0's ability to simulate the pattern is influenced by the presence of feedback. As with CDP+, I presented the Roberts et al. (2003) items to DRC version 1.0 with the default parameter set, and then again with the feedback parameters set to 0 (see Table 1). There were no errors or outliers in either version. Both versions of DRC produced robust effects of regularity (Feedback: $F(1, 98) = 124.009$, MSE = 15.63, $p < .001$; No Feedback: $F(1, 98) = 119.087$, MSE = 10.35, $p < .001$), position (Feedback: $F(1, 98) = 48.158$, MSE = 15.63, $p < .001$; No Feedback: $F(1, 98) = 20.740$, MSE = 10.35, $p < .001$), and a robust interaction between regularity and position (Feedback: $F(1, 98) = 38.531$, MSE = 15.63, $p < .001$; No Feedback: $F(1, 98) = 27.923$, MSE = 10.35, $p < .001$). Planned comparisons found that both models produced significant effects of regularity for both the 2nd (Feedback: $F(1, 64) = 110.764$, MSE = 22.32, $p < .001$; No Feedback: $F(1, 64) = 105.630$, MSE = 13.90, $p < .001$) and 3rd (Feedback:

F(1, 32) = 13.544, MSE = 3.119, $p < .001$; *F*(1, 32) = 13.249, MSE = 3.698, $p < .001$) position irregularities. Thus, simulating this effect in DRC does not rely on the presence of feedback.

Nonword consistency effects.

Andrews and Scarratt (1998) were interested in whether nonwords would be read aloud according to spelling-sound rules, or by analogy to known words. In particular, they examined whether the likelihood of irregular pronunciations (pronunciations that violated the spelling-sound rules) would be influenced by the consistency of a nonword's word neighbors. In a first experiment, they found that nonwords that had regular body neighbors (such as `BIVE` which has `HIVE` as a neighbor) were much more likely to be read according to spelling-sound conversion rules than were nonwords that had no regular body-neighbors (such as `VALK`, which has `WALK`, `TALK`, `CHALK`, etc… as neighbors, none of which have an audible /l/ as spelling-sound rules would require). In a second experiment they found that nonwords with no regular analogies but many body neighbors (such as `VALK` above) were also much less likely to be read according to conversion rules than words with no regular analogies, but few or no body neighbors (such as `REALT` or `JOURT`).

CDP+ (F) & CDP+ (NF) – Andrews and Scarratt (1998) experiment 1.

As can be seen in Table 9, CDP+ (F) and CDP+ (NF) produced virtually identical patterns of regularization. In fact, the models produced different pronunciations to only six items, three of which resulted in errors from both models (see Table 10).

Table 9. Percentage (%) of items pronounced according to regularization rules for Andrews and Scarratt (1998) Experiment 1.

Condition[a]	A & S (1998)[b]	CDP+ (F)[13]	CDP+ (NF)
CV/body analogs	92.2	100.0	100.0
CV/no body analogs	86.8	85.7	85.7
No CV/body analogs	94.0	100.0	100.0
No CV/no body analogs	86.9	90.9	91.1
Few regular analogs	32.3	25.0	25.0

a. CV refers to items that share initial consonant cluster-vowel structures with regular words. “body” refers to items that share a body (ignoring initial consonant cluster) with regular words.
b. As reported in Andrews and Scarratt (1998).

Table 10. Pronunciations that differed between CDP+ (F) and CDP+ (NF) for Andrews and Scarratt (1998) Experiment 1.

Item	Condition	CDP+ (F)		CDP+ (NF)	
goop	CV/body analogs	gɹʉːps	Error	gʉːp	Regular
vack	CV/body analogs	vækt	Error	væk	Regular
wose	No CV/no body analogs	wɔz	Error	wəʉz	Regular
beart	Few regular analogs	bat	Error	biːt	Error
kneart	Few regular analogs	nɔt	Error	nɜːt	Outlier
searn	Few regular analogs	sens	Error	sɜːn	Outlier

Table 11. Percentage (%) of items pronounced according to regularization rules in Andrews and Scarratt (1998) Experiment 2.

Condition[a]	A & S (1998)[b]	CDP+ (F)[14]	CDP+ (NF)
Consistent	92.5	82.9	83.3
Inconsistent	87.4	70.3	70.3
No regular analogs (many bodies)	19.4	10.5	10.5
No regular analogs (few bodies)	41.2	52.1	52.1

a. Consistent/Inconsistent refers to whether an item’s word neighbors were consistently regular, or if the word had a mix of regular and irregular body neighbors.
b. As reported in Andrews and Scarratt (1998).

Table 12. Pronunciations that differed between CDP+ (F) and CDP+ (NF) for Andrews and Scarratt (1998) Experiment 2.

Item	Condition	CDP+ (F)		CDP+ (NF)	
welf	No regular analogs (many bodies)	wel	Error	wæɪ	Error
vack	Consistent	vækt	Error	væk	Regular

CDP+ (F) & CDP+ (NF) – Andrews and Scarratt (1998) experiment 2.

In simulating the items from Andrews and Scarratt's (1998) experiment 2, feedback status influenced only two items. Table 11 summarizes the results and Table 12 summarizes the pronunciation differences.

Model comparison.

CDP+ (F) and CDP+ (NF) produced identical pronunciations to nearly every item in Andrews and Scarratt's (1998) experiment 2. There is thus no evidence that nonword consistency effects rely on the presence of feedback. Though not directly relevant, because this benchmark uses proportion of regularizations as the dependent variable, the correlation in cycle times between the two models was very high (r = .930, N = 191).

Lexicality by letter length (by frequency).

Weekes (1997) reported a study of the effects of letter length (using strings consisting of 3, 4, 5, and 6 letters) on reading aloud times. He examined the influence of letter length on both words and nonwords, and on words of differing frequencies (high vs. low). Weekes found that RT increased as letter length increased, and that this effect was larger for nonwords than for words. Though he argued that low-frequency words showed a stronger letter length effect than high frequency words, subsequent reanalysis of the Weekes data found that this interaction is not reliable (Coltheart et al., 2001). Perry et al. (2007) set out to simulate the following pattern: a significant effect of letter length, a significant letter length by lexicality interaction, but no frequency by letter length interaction in the word data. Data from the simulations can be found in Table 13.

Table 13. Mean cycles to response (percentage errors) to the Weekes (1997) items for CDP+ (F) and CDP+ (NF).

	High Frequency		Low Frequency		Nonword	
Length	Cycles	%E	Cycles	%E	Cycles	%E
CDP+ (F)						
3	77.80	(0.0%)	94.17	(0.0%)	120.00	(8.0%)
4	86.96	(0.0%)	100.92	(0.0%)	131.14	(12.0%)
5	91.08	(0.0%)	106.08	(4.0%)	150.45	(12.0%)
6	98.26	(0.0%)	110.63	(4.0%)	160.88	(0.0%)
CDP+ (NF)						
3	81.84	(0.0%)	100.29	(0.0%)	120.78	(4.0%)
4	91.72	(0.0%)	107.72	(0.0%)	132.52	(12.0%)
5	95.79	(0.0%)	112.92	(4.0%)	153.64	(0.0%)
6	102.96	(0.0%)	117.58	(4.0%)	161.88	(0.0%)

CDP+ (F).

After removing words unknown to the model (BLEST and BRUNCH), CDP+ made no word errors (of 198 words), and 8 errors to the 198 nonwords (THA, CAS, THUN, WILK, GEND, SPONT, GRITE, FRUND). Outlier identification resulted in two additional words (SPA, BRAWN), and three nonwords (TOB, COLM, SQUATE) being removed from analysis. When all three factors are included in the analysis, CDP+ (F) produced main effects of lexicality ($F(1, 271) = 1817.132$, $MSE = 68$, $p < .001$), letter length ($F(3, 271) = 145.595$, $MSE = 68$, $p < .001$), and frequency ($F(1, 271) = 146.918$, $MSE = 68$, $p < .001$), as well as a significant letter length by lexicality interaction ($F(3, 271) = 27.021$, $MSE = 68$, $p < .001$). The interaction between letter length and frequency was not significant ($F < 1$). As nonwords do not have a frequency, the terms for Lexicality × Frequency and Length × Lexicality × Frequency are necessarily omitted from the ANOVA model. When considering only the word data, CDP+ again produced significant effects of letter length ($F(3, 186) = 86.449$, $MSE = 34.9$, $p < .001$) and frequency ($F(1, 186) = 289.986$, $MSE = 34.9$, $p < .001$), but no interaction between them ($F < 1$).

Perry et al. (2007) also evaluated CDP+'s ability to account for variance in the item-level human RTs. To do so, they considered both the human RTs provided by Weekes (1997), and the mean RTs reported in the English Lexicon Project (ELP; Balota et al., 2007). For words, CDP+ (F) explained 8.5% of the variance in the Weekes data, and 21.5% of the variance in the ELP RTs for these items. For nonwords, CDP+ (F) explained 30.8% of the variance in the Weekes RT data (the ELP does not contain values for nonwords).

CDP+ (NF).

Words unknown to the model were removed once again (`BLEST` and `BRUNCH`). The CDP+ (NF) produced no word errors and only four nonword errors (`CAS`, `THUN`, `WILK`, and `GEND`). Thus CDP+ without feedback was once again *more accurate at nonword reading* than CDP+ with feedback. Outlier identification removed the same items as in the analysis for CDP+ (F) (words: `SPA`, `BRAWN`; nonwords: `TOP`, `COLM`, `SQUATE`). Including length, lexicality, and frequency in a single ANOVA analysis produced main effects of all three factors (length: $F(3, 275) = 130.764$, $MSE = 81$, $p < .001$; lexicality: $F(1, 275) = 1302.134$, $MSE = 81$, $p < .001$; frequency: $F(1, 275) = 161.824$, $MSE = 81$, $p < .001$), and a significant interaction between length and lexicality ($F(3, 275) = 24.136$, $MSE = 81$, $p < .001$). Once again, there was no evidence for a length by frequency interaction ($F < 1$). Considering only the word items, CDP+ (NF) again matched CDP+ (F)'s performance. There were main effects of length ($F(3, 186) = 71.149$, $MSE = 46$, $p < .001$) and frequency ($F(1, 186) = 289.553$, $MSE = 46$, $p < .001$), but no interaction between them ($F < 1$).

As for item-level variance, CDP+ (NF) explained 8.2% of the variance in the word data for the Weekes human RTs, and 21.5% of the variance in the ELP RTs. For the nonwords, the CDP+ (NF) explained 24.8% of the variance in the Weekes human RTs.

Model comparison.

The CDP+ (F) and CDP+ (NF) models were both able to correctly simulate the qualitative pattern of results for this data set, and the latencies correlated very strongly (words: $r = .996$, N = 194; nonwords: $r = .991$, N = 89). At first glance, it appears that CDP+ (F) was more successful at explaining item-level variance in the nonword data. CDP+ (F) explained 30.8% of the variance, whereas CDP+ (NF) explained only 24.8%. This is a significant drop in the proportion of variance explained (a paired test of the underlying correlations using an inter-correlation of .991 results in a t-value of 7.71, $p < .001$), however it can almost entirely be attributed to CDP+ (NF)'s increased accuracy in reading aloud nonwords. There are four items (THO, SPONT, GRITE, FRUND) that are included in the correlation for CDP+ (NF) but not for CDP+ (F) because the latter produced incorrect responses. When these items are removed from the correlation analysis for CDP+ (NF), the variance explained improves to 30.2%, which is no longer significantly different from CDP+ (F)'s 30.7% ($t < 1$).

Body neighborhood.

Body neighborhood is defined as the number of orthographic neighbors that share the same body as the target stimulus. For example, the homophone nonwords VEAP and VEEP have very different body neighborhoods. There are only 5 monosyllabic words ending in -EAP, but there are 13 that end in -EEP. Ziegler, Perry, Jacobs, and Braun (2001) manipulated lexicality (word vs. nonword), string length (3, 4, 5, or 6 letters)[15], and body neighborhood (high vs. low). They reported that body neighborhood influenced reading aloud times for both words and nonwords, but did not interact with length. They also replicated the Weekes (1997) interaction between lexicality and length. This pattern is important because DRC (version 1.0) produced no body neighborhood effect (Perry et al., 2007). Simulation results are summarized in Table 14.

Table 14. Mean cycles to response (percentage errors) to the Ziegler et al. (2001) items for CDP+ (F) and CDP+ (NF).

	Words				Nonwords			
	Low Body N		High Body N		Low Body N		High Body N	
Length	Cycles	%E	Cycles	%E	Cycles	%E	Cycles	%E
CDP+ (F)								
3	82.90	(0.0%)	83.80	(0.0%)	125.22	(10.0%)	118.22	(10.0%)
4	91.60	(0.0%)	86.70	(0.0%)	132.10	(0.0%)	127.10	(0.0%)
5	93.40	(0.0%)	90.70	(0.0%)	156.80	(0.0%)	150.10	(0.0%)
6	96.60	(0.0%)	96.10	(0.0%)	170.30	(0.0%)	175.33	(10.0%)
CDP+ (NF)								
3	87.50	(0.0%)	88.70	(0.0%)	123.60	(0.0%)	118.90	(0.0%)
4	97.10	(0.0%)	91.10	(0.0%)	132.60	(0.0%)	127.00	(0.0%)
5	97.90	(0.0%)	95.20	(0.0%)	154.90	(0.0%)	152.10	(0.0%)
6	101.10	(0.0%)	100.60	(0.0%)	170.00	(0.0%)	176.10	(0.0%)

CDP+ (F).

CDP+ with feedback made three errors to nonwords (`SIL`, `LAN`, `FRATCH`) and produced one nonword outlier (`SCRAST`). With these items removed, there were main effects of lexicality ($F(1, 140) = 880.33$, $MSE = 123$, $p < .001$), body neighborhood ($F(1, 140) = 5.46$, $MSE = 123$, $p < .05$), and length ($F(3, 140) = 51.79$, $MSE = 123$, $p < .001$), as well as a significant lexicality by length interaction ($F(1, 140) = 20.00$, $MSE = 123$, $p < .001$). This pattern matches the pattern reported for the human data in Ziegler et al. (2001).

CDP+ (NF).

Without feedback, CDP+ made no errors, and produced one nonword outlier (`SCRAST`). With only that item removed, there were main effects of lexicality ($F(1, 143) = 775.45$, $MSE = 118$, $p < .001$) and length ($F(3, 143) = 52.65$, $MSE = 117$, $p < .001$); a marginal effect of body neighborhood ($F(1, 143) = 3.51$, $MSE = 117$, $p = .063$); and a significant lexicality by length interaction ($F(1, 143) = 23.01$, $MSE = 117$, $p < .001$). The difference in the effect of body neighborhood (significant in CDP+ (F), only marginal in CDP+ (NF)) is driven entirely by the reduction in errors when feedback is removed. If the three items that produced errors in CDP+ (F) are removed (`SIL`, `LAN`, `FRATCH`), CDP+ (NF) produces a significant effect of body neighborhood ($F(1, 140) = 4.32$, $MSE = 117$, $p < .05$).

Model comparison.

On the surface, it appears that CDP+ (F) was slightly more successful at simulating the human data from Ziegler et al. (2001). In particular, CDP+ (F) produced a significant effect of body neighborhood ($p = .02$), where CDP+ (NF) produced only a marginal effect ($p = .06$). There are two reasons why I don’t consider this a strong case for the need to include feedback in the model. First, p = .05 is a widely accepted, but still relatively arbitrary, cut off point for significance. Whereas CDP+ (NF) does not meet this criterion, it is not far from achieving it and a larger stimulus set would likely increase the power enough to achieve it. Second, closer examination revealed that this discrepancy is the result of changes in the error rates – whereas CDP+ (F) made three errors to nonword items, CDP+ (NF) read all nonwords aloud correctly. If the three items CDP+ (F) named incorrectly are removed from the analysis for CDP+ (NF), the body neighborhood effect is now significant.

The correlation in latencies between the two models was very high (r = .994, N = 156). Even within lexical classes, the correlations remained very high (Words: r = .993, N=80; Non-words: r = .981, N=76). Both versions were also equally capable of capturing item-level variance (CDP+ (F): 55.2%; CDP+ (NF): 56.0%).

<u>Pseudohomophone advantage.</u>

McCann & Besner (1987) reported that nonwords that can be pronounced to sound like words (e.g., `BRANE`) are read aloud faster than matched control nonwords (e.g., `FRANE`). They argued that this effect was the result of "the assembly process [making] contact with existing whole-word representations in the phonological lexicon." (pp. 19-20) In dual-route models, this point of contact would rely on the feedback connections from the Phoneme Buffer to the POL. Because this account relies on feedback from the Phoneme Buffer to the POL, it predicts that CDP+ (NF), which does not include those feedback connections, should not be able to produce a pseudohomophone advantage; prima-facie, this effect may be one that simply cannot be simulated without recourse to feedback.[16] Results for the pseudohomophone advantage simulations can be found in Table 15.

Table 15. Mean cycles to response (percentage errors) to the McCann and Besner (1987) items for all versions of CDP+: (F), (NF), no lexical route, minimal feedback (from the Phoneme Buffer to the POL), and (NF) with lexical enhancement (LE).

	Pseudohomophones		Control Nonwords	
Model	Cycles	%E	Cycles	%E
CDP+ (F)	137.65	(13.9%)	145.02	(8.5%)
CDP+ (NF)	143.30	(11.1%)	147.14	(7.0%)
CDP+ no lexical route	143.30	(11.1%)	147.14	(7.0%)
CDP+ minimal feedback	137.95	(15.3%)	145.27	(9.9%)
CDP+ (NF) with LE	138.02	(13.9%)	145.78	(9.9%)

Table 16. Parameters used to simulate the pseudohomophone advantage (1987) for CDP+ with minimal feedback, CDP+ with no lexical route, and CDP+ (NF) with lexical enhancement (LE). Only values that differ from the default CDP+ (F) values are indicated.

Parameter	CDP+ (F)	minimal feedback	no lexical route	(NF) with LE
Lexical Route				
Feature Level				
Feature-to-letter excitation	0.005			
Feature-to-letter inhibition	-0.150			
Letter Level				
Letter-to-letter inhibition	0.000			
Letter-to-orthography excitation	0.075		0.000	
Letter-to-orthography inhibition	-0.550		0.000	-0.316
Orthographic Lexicon				
Orthography-to-orthography inhibition	-0.060			0.000
Orthography-to-phonology excitation	1.400			
Orthography-to-letter excitation*	0.300	0.000	0.000	0.000
Phonological Lexicon				
Phonology-to-phonology inhibition	-0.160			0.000
Phonology-to-phoneme excitation	0.128		0.000	
Phonology-to-phoneme inhibition	-0.010		0.000	
Phonology-to-orthography excitation*	1.100	0.000		0.000
Phonological Output Buffer				
Phoneme-to-phoneme inhibition	-0.040			
Phoneme-to-phonology excitation*	0.098		0.000	0.000
Phoneme-to-phonology inhibition*	-0.060		0.000	0.000
General Parameters				
Activation rate	0.200			
Frequency scale	0.400			
Phoneme naming activation criterion	0.670			
Sublexical Parameters				
Cycles before route begins	0			
Cycles before next letter accessed	15			
Sublexical to phoneme activation	0.085			
Letter level threshold for processing	0.210			

* indicates parameters that are set to zero for CDP+ (NF)

CDP+ (F).

Perry et al. (2007) used only 144 of the 160 items in the McCann and Besner (1987) corpus. It is not clear why some items were excluded from the simulation, however, to ensure comparability, the present simulations used the same 144 items originally reported in Perry et al. (2007). Of these 144 items, Perry et al. removed one nonword (VOLE) from the analysis because it is in fact a word (though it is unknown to the model). To match their analyses, I did the same here. With feedback present, CDP+ (F) produced errors to 16 items from the McCann and Besner (1987) stimuli. In addition, 3 items were identified as outliers. However, CDP+ (F) did produce significantly faster response times to pseudohomophones than to other nonwords ($t(122) = 2.14$, $p < .05$).[17]

CDP+ (NF).

As predicted by McCann and Besner's (1987) account, CDP+ (NF) is unable to simulate this effect. Without feedback from the Phoneme Buffer to the lexical route, pseudohomophones are no longer significantly faster than other nonwords ($t < 1$, ns; 13 errors, 2 outliers).

Model comparison (CDP+ (F) vs. CDP+ (NF)).

CDP+ (NF) was unable to simulate this effect, and that failure is reflected in the correlations between the cycle times. Compared to the other inter-model correlations reported here, the correlation in cycles between the CDP+ (F) and CDP+ (NF) was very low ($r = .895$, N = 124). This low correlation is seen for both the pseudohomophones ($r = .906$, N = 64) and nonwords ($r = .888$, N = 60). Curiously, despite CDP+ (NF)'s inability to simulate the basic pseudohomophone advantage, the two versions performed equivalently at modeling item-level variance (CDP+ (F): 4.96%; CDP+ (NF): 5.15%; the difference is not significant).

This is the first (and only) of the phenomena that could be simulated by CDP+ (F) but not by CDP+ (NF). On first glance this would seem to be because the pseudohomophone advantage requires that the non-lexical pronunciation find its way into the lexical system in order for its lexical status to influence reading aloud. In CDP+, this can only be accomplished by feedback from the Phoneme Buffer to the POL. There is another possibility that I will revisit shortly, but first, a closer look at this feedback account of CDP+ (NF)'s failure.

CDP+ with minimal feedback.

The failure of CDP+ (NF) to produce a significant pseudohomophone effect is consistent with the view that feedback plays a role in producing an advantage for reading aloud pseudohomophones in CDP+. However, this does not imply that feedback need appear everywhere that it exists in CDP+ (F). IA between the Phoneme Buffer and the POL should suffice, without a need for feedback along the rest of the lexical route. Indeed, with feedback from the Phoneme Buffer to the POL restored (without restoring any other feedback connections, see Table 16), the model once again produces a significant pseudohomophone advantage ($t(120) = 2.10$, $p < .05$; 18 errors, 3 outliers).

Model comparison (CDP+ (F) vs. CDP+ with minimal feedback).

A modified version of CDP+ (NF), in which only feedback from the Phoneme Buffer to the POL is intact, entirely restores the model's ability to simulate the pseudohomophone advantage. CDP+ with minimal feedback also improves on the inter-model correlation, increasing it to .9996 (N = 122). This nearly perfect correlation suggests that the minimal feedback included here captures nearly all of the contribution of feedback to reading aloud the

McCann and Besner (1987) stimulus set. Here again, both models capture the same proportion of item-level variance (CDP+ (F): 4.96%; CDP+ minimal feedback: 5.03%; the difference is not significant).

This result limits the role of feedback considerably. Because no other benchmark requires feedback at all, and the pseudohomophone advantage can be simulated with only a minimal amount of feedback, there is still no demonstrated need for feedback from the POL to the OIL, or from the OIL to the letter level. That said, it is too early to conclude that feedback is actually necessary: even with no feedback at all, there is a trend towards a pseudohomophone advantage with CDP+ (NF) (see Table 15). A closer look at that trend follows.

CDP+ (NF) with lexical emphasis (LE).

Though the pseudohomophone advantage can be observed with a minimal amount of feedback, it is informative that CDP+ (NF), with no feedback at all, still produces a trend towards a pseudohomophone advantage. This trend turns out to be driven entirely by the sub-lexical processes: a simulation with a version of the model in which the lexical system has been completely severed produced exactly the same results (down to the cycle times and responses to individual items, see Table 15). This suggests a number of things. First, the residual advantage is in no way lexically mediated, which means that some of the pseudohomophone effect in CDP+ (F) is purely non-lexical in nature. Whatever this advantage represents, it is not a lexically based orthographic confound. Second, it suggests that feedback in CDP+ may not simply be the way the sub-lexical route makes contact with the lexically stored pronunciations. With the default parameters, feedback appears to be necessary in order for the lexical system to have *any influence at all* on reading these items aloud. In essence, because removing feedback eliminated any lexical participation in reading these pseudohomophones aloud, it remains unclear whether it is the feedback per se that produces the pseudohomophone advantage, or whether it is lexical participation that produces the advantage. In CDP+ (F), lexical participation and feedback are confounded. To determine whether or not the pseudohomophone advantage could be attributed to lexical participation rather than feedback, I tested a third alternative model where feedback is eliminated but the lexical route is strengthened so that it still contributes to reading aloud. To enhance the lexical route's contribution, I adopted Reynolds and Besner's (2005) approach of reducing the amount of inhibition along the lexical route (the approach adopted here is similar to the GAS approach described in Reynolds and Besner, 2005; see Table 16), which has the effect of allowing the lexical route more influence in reading aloud. The result was a significant pseudohomophone advantage without the need for any feedback ($t(122) = 2.25$, $p < .05$; 17 errors, 2 outliers).

Model comparison (CDP+ (F) vs. CDP+ (NF) with LE).

Modifying CDP+ (NF) to allow an influence from the lexical route had the same effect as adding feedback. CDP+ (NF) with LE produced a robust pseudohomophone advantage. CDP+ (NF) with LE also improves on the inter-model correlation, increasing it to .976 (N = 119) from .895. Importantly, CDP+ (NF) with LE outperforms CDP+ (F) in the proportion of item-level variance explained (CDP+ (F): 4.96%; CDP+ LE: 7.28%; $t(117) = 2.52$, $p < .05$).

Most importantly, it is clear that the pseudohomophone advantage can be simulated without recourse to any feedback. In CDP+, feedback is a mechanism that allows the lexical system to influence reading of non-words, but it is this lexical participation that is responsible for the pseudohomophone advantage, and not the feedback per se. Reducing inhibition offers an alternative, feedback-free approach to enhancing the lexical participation.

Pseudohomophones and base word frequency.

McCann and Besner (1987) also found that reading aloud of pseudohomophones (e.g., `BRANE`) was not influenced by the frequency of the base words (e.g., `BRAIN`), despite the base words themselves showing robust frequency effects. The absence of a base word frequency influence on reading aloud of pseudohomophones places an important constraint on computational modelers. However, subsequent research (Borowsky, Owen, & Masson, 2002; Borowsky, Phillips, & Owen, 2003; Grainger, Spinelli, & Ferrand, 2000; Reynolds & Besner, 2005) has found the base word frequency effect is very context dependent. There is no base word frequency effect when pseudohomophones and nonwords are intermixed within a block of trials, but when the pseudohomophones are presented on their own, a base word frequency influence emerges (see Reynolds and Besner, 2005, for a thorough review of this literature). To further complicate the situation, both Borowsky et al. (2003) and Reynolds and Besner (2005) found an effect of the order of presentation of the nonwords and pseudohomophones, so that when the pseudohomophones were read aloud first there was a large base word frequency effect, but when the nonwords preceded the pseudohomophones, the base word frequency effect was absent. Importantly, when nonwords and pseudohomophones are read aloud in separate blocks, the pseudohomophone advantage observed in typical mixed-list experiments becomes a pseudohomophone *disadvantage* under blocked conditions.

Clearly, then, the influence of base word frequency on reading aloud pseudohomophones is sensitive to context. Currently, computational models of reading aloud do not track contextual information. The general strategy employed by computational modelers faced with this conundrum is to demonstrate that the different data patterns that are found under different contexts can be simulated by assuming that the context influences the configuration of the system in some plausible way (see Reynolds & Besner, 2005). In this way modelers provide an existence proof that the network could simulate all of the necessary patterns, and lacks only a mechanism for detecting the context and making the necessary adjustments.

In the case of pseudohomophony and base word frequency, models must accommodate the fact that one context produces a pseudohomophone advantage with no base word frequency effect (mixed lists), another context produces a pseudohomophone *disadvantage* with a base word frequency effect (pseudohomophones read aloud first in a pure list followed by a pure list of nonwords), and finally another context produces a pseudohomophone disadvantage in conjunction with the *absence* of a base word frequency effect (nonwords followed by pseudohomophones in pure lists).

A comment on word frequencies.

McCann and Besner (1987) used the Kucera and Francis (1967) word frequency norms in their analysis of the influence of base word frequency on human RTs. These frequency norms are a proxy for what is thought to be “typical” experience of subjects. In the case of CDP+, the exact frequencies are clearly specified by the modeler in the lexicon file. Thus, Perry et al. (2007) used the frequencies that were embedded in the CDP+ orthographic system, with a few exceptions. Items `GANE` (`GAIN`), `WAIJE` (`WAGE`), and `WEAD` (`WEED`) were associated with frequencies 6,342, 5,911, and 5,634 respectively in the correlations. However, within the CDP+ lexicon (and thus during training of the non-lexical system) the frequencies for the base words were 861, 445, and 167 respectively. It is not clear where the erroneous values were obtained (C. Perry, personal communication).[18] The present analysis corrects this discrepancy, with very important consequences for CDP+ (F)'s ability to simulate the base word frequency phenomenon.

CDP+ (F).

To demonstrate the different patterns of base word frequency effects and pseudohomophone advantages/disadvantages, Perry et al. (2007) proposed that subjects adjust their response criterion depending on the context. They implemented this mechanism in the CDP+ model by varying the Minimum Naming Activation Criterion (MNAC), to show that as subjects become more conservative in responding, the base word frequency effect increases so that at low values of the MNAC, there is no frequency effect, whereas at higher values the effect emerges. By default, the MNAC is set to .67, which has already been shown to produce a pseudohomophone advantage (see previous section). Perry et al. reported that the default model also produced a significant base word frequency effect (using log transformed word frequencies), and indeed the present simulation supports that claim ($r = -.334$, N = 60, $p < .01$). Thus, by default the model produces both a significant pseudohomophone advantage, and a significant base word frequency effect – a pattern not yet observed in humans under any condition.

According to Perry et al. (2007), subjects lower their response criterion in order to reduce the influence of the lexical route with the result that base word frequencies do not influence latencies (but see Reynolds & Besner, 2010, for an argument against this view). In CDP+, the challenge is to reduce the MNAC enough to eliminate the influence of base word frequency while preserving the pseudohomophone advantage (to simulate the mixed-list conditions). Though Perry et al. (2007) reported that setting the MNAC to .64 achieved this aim, their success relied on the use of the three incorrect word frequencies described earlier. Correcting those frequencies produced only a marginal pseudohomophone effect (Cycles: 133.26 vs. 139.50; $t(123) = 1.78$, $p = .077$; 16 errors, 2 outliers), but continued to produce a robust base word frequency effect ($r = -.305$, $N = 61$, $p < .05$). Although reducing the MNAC did reduce the influence of base word frequency, the pseudohomophone advantage was lost before the base word frequency effect was eliminated. It is difficult to see how only manipulating the MNAC can rectify this problem. Simply manipulating the response criterion is not enough to simulate the complicated relationship between context, pseudohomophony, and base word frequency.

Reynolds and Besner's (2005) inhibition account.

Reynolds and Besner (2005) proposed an alternate method of simulating the complicated data pattern. They argued that subjects modulate how they used lexical information depending on the context. Generally speaking, the lexical route will produce many candidate responses to a given stimulus, and the breadth of the lexical activation is modulated by competitive inhibition of the weaker candidates by the stronger candidates. Reynolds and Besner (2005) argued that when pseudohomophones are presented within a nonword context (either mixed with nonwords, or when a list of nonwords precedes the pseudohomophone list), subjects allow activation of a broader range of candidates (which they termed the General Activation Strategy, or GAS). When the pseudohomophones are presented without the non-word context (i.e., in a pure-list before being exposed to any nonwords), subjects rely more on specific word knowledge (which they termed the Specific Activation Strategy, or SAS). When pseudohomophones are presented after a list of nonwords, the options become broader (see Reynolds and Besner, 2005 for a complete treatment) but from a simulation perspective they adopted an intermediate level of inhibition (which they named the Intermediate Activation Strategy, or IAS). They further argued that when nonword controls were read aloud alone or in a mixed-list with pseudohomophones, subjects adopted the GAS.

Broadly speaking, Reynolds and Besner (2005) adjusted the strength of the competitive inhibition in DRC along the lexical route and found that when inhibitory connections are weakened (allowing a larger set of activated lexical entries), the base word frequency effect was diminished. Strengthening the inhibitory connections from the letter level to the OIL (and thus giving the model the ability to hone in more accurately on the most relevant lexical entry) increased the base word frequency effect. Though Reynolds and Besner (2005) demonstrated the success of this strategy in the context of the DRC model, the similar structures of the two models suggest a similar approach might be successful in CDP+.

To test the Reynolds and Besner (2005) account with CDP+, I first applied their parameter changes proportionately to the analogous parameters in CDP+ (see Table 17). With the GAS parameter set (meant to simulate mixed-list conditions), CDP+ failed to produce a significant pseudohomophone advantage (Cycles: 141.66 vs. 143.89; $t < 1$, n.s.; 14 errors, 2 outliers), but still produced a significant base word frequency effect ($r = -.255$, N = 64, $p < .05$). When the nonwords are read using the GAS parameter set, but the pseudohomophones are read using the SAS parameter set (meant to simulate the conditions of a pure-list of pseudohomophones read aloud before the nonwords), CDP+ failed to produce the pseudohomophone disadvantage (Cycles: 142.33 vs. 143.89; $t < 1$, n.s.; 14 errors, 2 outliers), and no longer produced a base word frequency effect ($r = -.199$, N = 63, $p = .117$). With the IAS parameter set (meant to simulate a pure-list of pseudohomophones read aloud *after* a list of nonwords), the results are similar to those for the SAS set (Cycles: 142.30 vs. 143.89; $t < 1$, n.s.; 14 errors, 2 outliers; $r = -.203$, N = 63, $p = .111$). Thus this approach, successful in DRC, fails with CDP+.

CDP+ relies more heavily on the non-lexical route in reading aloud than does the DRC (Robidoux & Besner, 2010). Reynolds and Besner (2005) vastly reduced the lexical route's influence on naming in DRC by reducing the output from the POL to the phonemic buffer by 90%. They did this to avoid lexical capture when competitive inhibition is reduced (where non-words are read aloud as words that share similar orthographies). It's possible that this is not necessary in CDP+ due to the weaker influence of the lexical route. To test this possibility, I conducted the same series of simulations described above (GAS, SAS, and IAS) but with the POL to Phoneme Buffer connection strengths left unchanged from the defaults – the results were a resounding failure. All three strategies (GAS, SAS, IAS) produced robust pseudohomophone advantages but also robust base word frequency effects.

In summary, it appears that Perry et al.'s (2007) success at simulating the base word frequency effect relied on errors in their base word frequency values. When these erroneous frequencies are corrected, CDP+ can no longer simulate the complicated pattern of pseudohomophone and base word frequency effects.

Table 17. CDP+ parameters used to mimic Reynolds and Besner's (2005) General Activation Strategy (GAS), Specific Activation Strategy (SAS), and Intermediate Activation Strategy (IAS). (Only parameters that differ from the default CDP+ (F) values are indicated).

Parameter	CDP+ (F)	GAS	SAS	IAS
Lexical Route				
Feature Level				
Feature-to-letter excitation	0.005			
Feature-to-letter inhibition	-0.150			
Letter Level				
Letter-to-letter inhibition	0.000			
Letter-to-orthography excitation	0.075			
Letter-to-orthography inhibition	-0.550	-0.316	-0.550	-0.487
Orthographic Lexicon				
Orthography-to-orthography inhibition	-0.060	0.000	0.000	0.000
Orthography-to-phonology excitation	1.400			
Orthography-to-letter excitation*	0.300			
Phonological Lexicon				
Phonology-to-phonology inhibition	-0.160	0.000	0.000	0.000
Phonology-to-phoneme excitation	0.128	0.0128	0.0128	0.0128
Phonology-to-phoneme inhibition	-0.010			
Phonology-to-orthography excitation*	1.100			
Phonological Output Buffer				
Phoneme-to-phoneme inhibition	-0.040			
Phoneme-to-phonology excitation*	0.098			
Phoneme-to-phonology inhibition*	-0.060			
General Parameters				
Activation rate	0.200			
Frequency scale	0.400			
Phoneme naming activation criterion	0.670			
Sublexical Parameters				
Cycles before route begins	0			
Cycles before next letter accessed	15			
Sublexical to phoneme activation	0.085			
Letter level threshold for processing	0.210			

* indicates parameters that are set to zero for CDP+ (NF)

CDP+ with minimal feedback.

Because CDP+ (NF) did not produce a significant pseudohomophone advantage, there is little point to assessing its ability to simulate the base word frequency effects. However, adding a minimal amount of feedback (between the Phoneme Buffer and the POL) entirely restored the pseudohomophone advantage. As with CDP+ (F), CDP+ with minimal feedback produced a significant effect of base word frequency ($r = -.302$, N = 59, $p < .05$). Setting MNAC to .64 the pseudohomophone advantage weakens somewhat (Cycles: 133.42 vs. 140.55; $t(122) = 1.934$, p = .055; 16 errors, 3 outliers), but the base word frequency effect remains significant ($r = -.284$, N = 60, $p < .05$). Thus the pseudohomophone advantage is lost before the base-word frequency effect with this version of the model as well. In short, CDP+ with minimal feedback also fails to produce the appropriate pattern of pseudohomophone and base word frequency effects.

CDP+ (NF) with lexical emphasis (LE).

It is possible to simulate a pseudohomophone advantage with CDP+ (NF), provided the lexical route is given more influence in reading aloud than it has by default. However, as with the other two models that produced pseudohomophone advantages, CDP+ (NF) with LE also produced a robust base word frequency effect ($r = -.265$, N = 60, $p < .05$). With the MNAC set to .64, the pseudohomophone advantage is only marginally significant (Cycles: 133.43 vs. 137.72; $t(119) = 1.74$, p = .085; 18 errors, 4 outlier), but the base word frequency effect remains marginally significant as well ($r = -.234$, N = 60, $p = .073$). To make the pseudohomophone effect significant required an MNAC of .66 (which is close to the default of .67). With MNAC set to .66, the pseudohomophone advantage is significant (Cycles: 136.13

vs. 142.84; $t(121) = 1.74$, $p = .0432$; 17 errors, 3 outlier), but the base-word frequency is very nearly significant ($r = -.249$, $N = 60$, $p = .055$). Certainly, this is not the sort of result that would lead a researcher to conclude that base word frequency is no longer influencing CDP+'s response times.

It seems that, with or without feedback, CDP+ is unable to simulate the conjunction of pseudohomophone advantages/disadvantages and base word frequency effects that are simulated by DRC. It remains to be seen whether future versions of CDP+ will be able to produce all of the necessary patterns (a pseudohomophone advantage with no base word frequency effect, a pseudohomophone disadvantage and a base word frequency effect, and a pseudohomophone disadvantage with no base word frequency effect).

Masked priming.

In a masked priming experiment (where subjects are assumed to be subjectively unaware of the primes) Forster and Davis (1991) reported significant advantages for words when they were preceded by other words that shared the same onset (e.g., `LEG-LOW`), but not when they were preceded by words with the same rhyme (`TOE-LOW`) relative to when they are preceded by unrelated primes (`RUN-LOW`). Later, Montant and Ziegler (2001) found that rhyme priming can be obtained if the onset of the prime is masked with a # symbol (e.g., `#AKE` primed `MAKE` even though `FAKE` does not), so it appears that both onset and rhyme primes facilitate reading aloud even when the primes are masked. Perry et al. (2007) simulated this pattern with CDP+ using the Forster and Davis (1991) stimuli.[19]

CDP+ (F).

Among the targets in Forster and Davis' (1991) stimuli, the word `CLUE` does not appear in CDP+'s lexicon, and was removed from analysis. One of the primes (`YUK`, used as the onset prime for YOU) is also a nonword in CDP+. To ensure a balanced design, all trials containing YOU as a target were removed. In the case where the models made errors, or responses were identified as outliers, all occurrences of those targets were removed from analysis (i.e., if CDP+ incorrectly pronounced an item in any of the prime conditions, it was removed from all conditions before analysis).

CDP+ (F) produced no outliers, but made two errors on unrelated trials (OAR, HUE), three errors in the onset-priming condition (BOW, KNEE, and OAR), and three in the rhyme-priming condition (BOW, LOW, and HUE). Thus, in addition to YOU and CLUE, the items OAR, HUE, BOW, KNEE, and LOW were removed. With the remaining items, there was a significant advantage of onset priming relative to unrelated primes ($t(18) = 11.40$, $p < .001$), and a significant advantage of rhyme priming relative to unrelated primes ($t(18) = 4.24$, $p < .001$).

CDP+ (NF).

Without feedback, CDP+ made no errors, and produced no outliers, so only CLUE and YOU were removed from analysis. As with CDP+ (F), CDP+ (NF) produced a significant advantage of onset priming relative to unrelated primes ($t(23) = 4.84$, $p < .001$), and a significant advantage of rhyme priming relative to unrelated primes ($t(23) = 4.16$, $p < .001$).

Model comparison.

CDP+ (NF) matched CDP+ (F)'s performance on this benchmark (see Table 18). It was successful with Forster and Davis' (1991) stimuli. For the Forster and Davis (1991) items, the correlation in latencies between the two models was very high ($r = .998$, N = 57). Within priming conditions, the correlations remained very high (Unrelated: $r = .999$, N = 19; Onset Prime: $r = .999$, N = 19; Rhyme prime: $r = .998$, N = 19).

Table 18. Mean cycles to response (percentage errors) to the Forster and Davis (1991) items for CDP+ (F) and CDP+ (NF).

	Unrelated prime		Onset prime		Rhyme prime	
Model	Cycles	%E	Cycles	%E	Cycles	%E
CDP+ (F)	82.21	(8.0%)	79.89	(12.5%)	80.37	(12.0%)
CDP+ (NF)	87.38	(0.0%)	85.54	(0.0%)	85.92	(0.0%)

Surface dyslexia.

Patterson and Behrmann (1997) tested a surface dyslexic subject (MP) who tends to provide regularized pronunciations to exception words (e.g., MP would be more likely to pronounce `PINT` to rhyme with `MINT` than would a non-dyslexic reader). Patterson and Behrmann (1997) reported that MP was not only more likely to regularize exception words, but this tendency towards regularization was influenced by the consistency of the word so that MP produced more regularizations to words that were inconsistent than to words that were consistent. They operationalized consistency as the ratio of regular words to exception (or irregular) words and found that if an exception word had many more regular neighbors than exception neighbors; MP was more likely to provide a regularized pronunciation. This operationalization is similar to the F-E ratio employed by Jared (2002) in examining consistency effects in skilled readers.

In addition to the consistency effects, two separate studies on two separate patients have examined the effect of word frequency on regularizations in surface dyslexia. Behrmann and Bub (1992) tested patient MP (the same patient described in Patterson and Behrmann, 1997), whereas McCarthy and Warrington (1986) tested patient KT. In both cases, the patients showed strong influence of frequency on regularizations of exception words, despite very high accuracy with regular words. This interaction between frequency and regularity was much stronger in KT than in MP.

Table 19. CDP+ parameters used to mimic MP and KT's surface dyslexia. (Only parameters that differ from the default CDP+ (F) values are indicated).

Parameter	CDP+ (F)	MP	KT
Lexical Route			
Feature Level			
Feature-to-letter excitation	0.005		
Feature-to-letter inhibition	-0.150		
Letter Level			
Letter-to-letter inhibition	0.000		
Letter-to-orthography excitation	0.075		
Letter-to-orthography inhibition	-0.550		
Orthographic Lexicon			
Orthography-to-orthography inhibition	-0.060		
Orthography-to-phonology excitation	1.400		
Orthography-to-letter excitation*	0.300		
Phonological Lexicon			
Phonology-to-phonology inhibition	-0.160		
Phonology-to-phoneme excitation	0.128	0.055	0.055
Phonology-to-phoneme inhibition	-0.010		
Phonology-to-orthography excitation*	1.100		
Phonological Output Buffer			
Phoneme-to-phoneme inhibition	-0.040		
Phoneme-to-phonology excitation*	0.098	0.000	0.000
Phoneme-to-phonology inhibition*	-0.060		
General Parameters			
Activation rate	0.200		
Frequency scale	0.400	0.800	1.000
Phoneme naming activation criterion	0.670		
Sublexical Parameters			
Cycles before route begins	0		
Cycles before next letter accessed	15		
Sublexical to phoneme activation	0.085		
Letter level threshold for processing	0.210		

* indicates parameters that are set to zero for CDP+ (NF)

To create a surface dyslexic version of CDP+, Perry et al. (1997) reduced the influence of the lexical system by eliminating excitatory feedback from the Phoneme Buffer to the POL, and reducing the excitation from the POL to the Phoneme Buffer to 0.055 from 0.128. The net result is that reading aloud relies more heavily on the sublexical route. They also followed Coltheart et al.'s (2001) strategy of increasing the influence of word frequency to reflect the sensitivity to frequency found in both MP and KT. In the case of MP, they increased the frequency parameter from 0.4 to 0.8, and to reflect the larger effect of frequency for KT they increased it from 0.4 to 1.0. These parameter settings are summarized in Table 19.

Model comparison for patient MP (CDP+ (F) vs. CDP+ (NF)).

As modeling of surface dyslexia relies on error rates, the principal concern is whether or not removing feedback from CDP+ will affect the pronunciations. MP has been tested extensively by a number of researchers. In particular, Patterson and Behrmann (1997) tested her using stimuli that manipulated consistency and regularity, and Behrmann and Bub (1992) tested her using stimuli that manipulated regularity and word frequency. Perry et al. (2007) presented CDP+ (F) with the two stimulus sets and found that CDP+ (F) matched MP's performance.

Removing feedback had little effect on the responses provided by CDP+ to the Patterson and Behrmann (1997) stimuli. For the low (24 regular, 24 irregular words) and medium consistency words (29 regular, 30 irregular words; SOUR is excluded from analysis because it does not appear in the lexicon), both models produced identical responses. In the case of high-consistency exception words (12 regular; 12 exception, all of which began with "WA" or "WO") only one item's pronunciation differed: when presented with WAR, CDP+ (F) pronounced it correctly (/wo:/, to rhyme with BORE), whereas CDP+ (NF) regularized it (/wa:/ to rhyme with BAR). Thus, of 131 words presented, only one item's pronunciation (<1%) was affected by the removal of feedback.

For the Behrmann and Bub (1992) items, feedback had no effect on the pronunciations produced by the model. All 223 items known to CDP+ produced identical pronunciations whether there was feedback or not. Five items are not known to CDP+: CHOIR, SOUR, GAUGE, PLOW, and MIL. They have been removed from the analysis, though it wasn't necessary: both models produced identical pronunciations to all five items.

In summary, CDP+'s success at simulating MP's performance is not dependent on the presence of feedback. Of the 307 unique items in the Behrmann and Bub (1992) and Patterson and Behrmann (1997) stimulus sets, only WAR's pronunciation was influenced by feedback.

Model comparison for patient KT (CDP+ (F) vs. CDP+ (NF)).

KT was tested on a list that manipulated word frequency and regularity (McCarthy & Warrington, 1986). Perry et al. (2007) chose a similar list from Taraban and McClelland (1987) and presented it to CDP+ (F) using the parameters meant to simulate KT's deficits. Of the 96 items in this stimulus list, PLOW is unknown to CDP+. Of the remaining 95 items, CDP+ (F) and CDP+ (NF) produced identical pronunciations to 91 (95.8%). The remaining four are summarized in Table 20. Two of the four items are errors in both models. In the case of LOSE, CDP+ (F) produces a regularization, whereas CDP+ (NF) simply gets it wrong (pronouncing LOW instead). Both models mispronounce BUS (a regular word). Of the remaining two, one (ARE) is correctly read aloud by CDP+ (F), and regularized by CDP+ (NF), and the other (WAND) is regularized[20] by CDP+ (F), and read correctly by CDP+ (NF). The net result is a slight shift in the proportions of high frequency and low-frequency exception words read aloud correctly, but the pattern still closely matches that of KT.

Table 20. Pronunciation differences between CDP+ (F) and CDP+ (NF) to the Taraban and McClelland (1987) items, using parameters meant to simulate surface dyslexic patient KT.

Item	Condition	CDP+ (F)		CDP+ (NF)	
are	High Frequency, Exception	eː	Correct	aː	Regular[a]
lose	Low Frequency, Exception	ləʉz	Regular	ləʉ	Error
wand	Low Frequency, Exception	wænd	Error	wɔnd	Correct
bus	Low Frequency, Regular	baz	Error	ba	Error

a. “Regular” indicates an exception word that was regularized.

Phonological dyslexia.

Phonological dyslexia is characterized by selective impairment of nonword reading (e.g., `FRANE`) with preserved ability to read words (e.g., `BRANE`). Coltheart et al. (2001) provide a thorough review of phonological dyslexia, so I will provide only a brief review here. Derouesné and Beauvois (1985) tested a patient, LB, who showed the characteristic pattern of severely impaired nonword reading, with near-normal reading of words. However, LB found pseudohomophones (e.g., `BRANE`) easier to read aloud than other nonwords, particularly if the pseudohomophones were orthographically similar to their base word (e.g., `SEAD` is similar to `SEED`, whereas `PHOCKS` is not similar to `FOX`). Coltheart et al. (2001) directly attributed this effect to feedback in the DRC: "For [pseudohomophones], this abnormally weak excitation is boosted by top-down interactive activation from the entry in the phonological lexicon of the [pseudohomophone's] parent word; that is the source of the [pseudohomophone] advantage in our simulation of phonological dyslexia." (p. 243)[21]

To simulate LB's phonological dyslexia, Perry et al. (2007) reduced the contribution of the sub-lexical route by reducing the strength of activation from the sublexical route to the Phoneme Buffer, and reduced all inhibitory connections along the lexical route by 50%. Though they argued that this second set of changes served to "[increase] the amount of noise in the model because it allows competing representations to be activated that would otherwise have been suppressed by means of inhibition" (p. 295), the changes are similar to the ones I made that increased lexical contributions to reading aloud in simulating the pseudohomophone advantage in intact readers (see the section entitled "Pseudohomophone advantage."). The parameter set used is summarized in Table 21.

Table 21. CDP+ parameters used to mimic LB's phonological dyslexia. (Only parameters that differ from the default CDP+ (F) values are indicated).

Parameter	CDP+ (F)	LB
Lexical Route		
Feature Level		
Feature-to-letter excitation	0.005	
Feature-to-letter inhibition	-0.150	
Letter Level		
Letter-to-letter inhibition	0.000	
Letter-to-orthography excitation	0.075	
Letter-to-orthography inhibition	-0.550	-0.275
Orthographic Lexicon		
Orthography-to-orthography inhibition	-0.060	-0.030
Orthography-to-phonology excitation	1.400	
Orthography-to-letter excitation*	0.300	
Phonological Lexicon		
Phonology-to-phonology inhibition	-0.160	-0.080
Phonology-to-phoneme excitation	0.128	
Phonology-to-phoneme inhibition	-0.010	-0.005
Phonology-to-orthography excitation*	1.100	
Phonological Output Buffer		
Phoneme-to-phoneme inhibition	-0.040	-0.020
Phoneme-to-phonology excitation*	0.098	
Phoneme-to-phonology inhibition*	-0.060	-0.030
General Parameters		
Activation rate	0.200	
Frequency scale	0.400	
Phoneme naming activation criterion	0.670	
Sublexical Parameters		
Cycles before route begins	0	
Cycles before next letter accessed	15	
Sublexical to phoneme activation	0.085	0.060
Letter level threshold for processing	0.210	

* indicates parameters that are set to zero for CDP+ (NF)

Model comparison for patient LB (CDP+ (F) vs. CDP+ (NF)).

As with most research on acquired dyslexia, assessing LB's nonword naming emphasized accuracy in naming various word-types rather than speed. Perry et al. (2007) presented the network with 160 nonwords: 40 pseudohomophones that were visually similar to their base words (e.g., FORSE), 40 pseudohomophones that were visually distant from their base words (e.g., SHAIK), and 80 matched control nonwords. The question of interest here is whether feedback modifies the pronunciations that CDP+ produces to the stimulus set. Both the CDP+ (F) and CDP+ (NF) versions of the models produced identical pronunciations to every item. Feedback played no role in CDP+ (F)'s ability to simulate LB's phonological dyslexia. Indeed, taken together with the results in simulating pseudohomophone effects in skilled readers, this result further supports the view that pseudohomophone advantages can be simulated without recourse to feedback, and that this is true whether we examine error rates in acquired dyslexics, or reaction time effects in skilled readers. It is particularly encouraging that both effects are simulated without feedback using similar approaches: reducing the inhibition along the lexical route, which I argue serves to increase the lexical influence on naming.

Other Phenomena

Perry et al. (2007) discussed several other phenomena that they argued were not clearly enough established to be considered benchmarks. These included results from orthographic and phonological neighborhood (where CDP+ either matched or outperformed DRC 1.0), and results related to whammies (where DRC outperforms CDP+). Perry et al. (2007) and Coltheart et al. (2001) differed in whether they considered these phenomena key for computational modelers to simulate. To be consistent with Perry et al. (2007), I separate them from the discussion of the agreed upon benchmarks, but feel they are important enough to warrant examination nonetheless.

Orthographic and phonological N.

Whereas body neighborhood effects are included in Perry et al.'s (2007) list of benchmarks, there are other measures of neighborhood size (or N). The most widely used measure is orthographic N: the number of words that can be formed by changing only a single letter. Letter strings with more orthographic neighbors are read aloud faster (and often more accurately) than those with fewer neighbors, and this is true for both nonwords and words (Andrews, 1989; 1992; 1997). Importantly, Mulatti, Reynolds, and Besner (2006) showed that the orthographic N effect is confounded with phonological N – the number of words that can be formed by changing a single *phoneme* (e.g., PINT and MINT are orthographic, but not phonological neighbors; whereas PHONE and FOAM are phonological, but not orthographic neighbors). Mulatti et al. (2006) found that when phonological N is controlled for, there is no orthographic N effect. Conversely they found that a phonological N effect persisted even when orthographic N was controlled for – evidence that the N effect is in fact phonological, and not orthographic, in nature. Mulatti et al. found that no computational models were able to capture this phenomenon at the time. Evidence that the N effect is phonological has found more recent support in Adelman and Brown (2007) who found similar results in an examination of a series of mega-studies: the time to read aloud words is influenced by the number of phonological, but not orthographic, neighbors.

Table 22. Mean cycles to response (percentage errors) to the Mulatti et al. (2006) items for CDP+ (F) and CDP+ (NF).

	Phonological N		Orthographic N	
Neighborhood size	Cycles	%E	Cycles	%E
CDP+ (F)				
Low N	102.00	(0.0%)	99.87	(0.0%)
High N	96.77	(0.0%)	99.14	(0.0%)
Difference	*5.23*	*(0.0%)*	*0.73*	*(0.0%)*
CDP+ (NF)				
Low N	108.23	(0.0%)	106.43	(0.0%)
High N	102.07	(0.0%)	105.75	(6.7%)
Difference	*6.16*	*(0.0%)*	*0.68*	*(-6.7%)*

Perry et al. (2007) reported that CDP+, like DRC, was able to capture the orthographic N effect in a set of nonwords (taken from Coltheart et al., 2001) but failed to capture the effect in two sets of words (taken from Andrews, 1989; 1992). However, with Mulatti et al.'s (2006) carefully constructed items, CDP+ produced a significant phonological N effect, but no orthographic N effect (accurately simulating the human data).

CDP+ (F).

I found that with the default settings, CDP+ produced a marginal effect of N with Coltheart et al.'s (2001) set of nonwords ($r = -.114$, $p = .091$, $N = 224$).[22] With respect to words, my simulations confirmed Perry et al.'s (2007) finding that CDP+ (F) does not produce a significant N effect with the Andrews items (1989 Items: $F(1, 52) = 1.05$, ns; 1992 Items: $F(1, 92) = 2.63$, $MSE = 59.9$, $p > .10$), though it does produce a significant word frequency effect (1989 Items: $F(1, 52) = 36.06$, $MSE = 63.6$, $p < .001$; 1992 Items: $F(1, 92) = 120.18$, $MSE = 59.9$, $p < .001$). There is also no evidence of a N by word frequency interaction ($Fs < 1$ for both the 1989 and 1992 items).

With Mulatti et al.'s (2006) set of carefully unconfounded items, CDP+ (F) produced a significant effect of phonological N ($t(58) = 2.21$, $p < .05$), but no effect of orthographic N ($t < 1$). See Table 22 for these results.

CDP+ (NF).

The no-feedback version of CDP+ produced a significant effect of N in Coltheart et al.'s (2001) corpus of nonwords ($r = -.134$, $p < .05$, $N = 226$). For the Andrews (1989) and (1992) items, CDP+ (NF) produced no N effect (1989 Items: $F(1, 52) = 1.01$, ns; 1992 Items: $F(1, 92) = 2.22$, $MSE = 70.9$, $p > .10$), though it does produce a significant word frequency effect (1989 Items: $F(1, 52) = 43.62$, MSE = 74.7, $p < .001$; 1992 Items: $F(1, 92) = 139.68$, MSE = 70.9, $p < .001$). There is also no evidence of a neighborhood by word frequency interaction (Both *F*s < 1).

With Mulatti et al.'s (2006) set of carefully unconfounded items, CDP+ (NF) matched CDP+ (F)'s performance, producing a significant effect of phonological N ($t(58) = 2.28$, $p < .05$), but no effect of orthographic N ($t < 1$).

Model comparison.

CDP+ produced identical results in both forms (feedback vs. no-feedback) for the Coltheart et al. (2001) set of nonwords, the Andrews (1989; 1992) sets of words, and Mulatti et al.'s (2006) set of words. For Coltheart et al.'s (2001) corpus of nonwords, the models performed similarly, and the latencies from the two models were highly correlated ($r = .947$, $p < .001$, N = 224). For the Andrews (1989) and Andrews (1992) corpora, the model correlations were very high (1989: $r = .990$, $p < .001$, N = 58; 1992: $r = .994$, $p < .001$, N = 96; respectively). The model latencies were also very highly correlated for the Mulatti et al. (2006) items ($r = .996$, $p < .001$, N = 118). Feedback is clearly not necessary to CDP+'s ability to simulate orthographic and phonological N effects.

Whammies.

Rastle and Coltheart (1998) asked whether reading aloud of nonwords would be influenced by the length of the letter string (letter-length), or the number of graphemes embedded in the letter string (grapheme-length; a grapheme is a letter cluster that produces a single phoneme. e.g., PH is a single grapheme because it maps to /f/). To test this question they presented participants with letter strings that were identical in letter-length, but differed in the number of graphemes (e.g., BARCH vs. BREPS). They reasoned that if letter-length was important, than there should be no difference between the two types of words, if grapheme-length influenced reading aloud latencies then the items with fewer phonemes (BARCH) should be read aloud faster than those with more phonemes (BREPS). Their results were inconsistent with both of these predictions – nonwords like BARCH, which require converting two letters into a single phoneme, were read aloud *more slowly* than BREPS, which had no such competition. They hypothesized that this effect (which they called the whammy effect) was due to the serial processing of the letter string by the sublexical processes. When the sublexical process first encounters the "C" in BARCH, the phoneme /k/ is activated in the Phoneme Buffer. When the "H" is subsequently encountered, the correct phoneme /tʃ/ must overcome the activation of /k/ in order for the nonword to be read aloud correctly. This competitive process slows response latencies and results in the whammy effect.

Perry et al. (2007) argued that because Rastle and Coltheart's (1998) own data failed to produce a significant whammy effect in the item analysis, the existence of this effect remains in question. As such, they argued that it does not yet warrant being considered a benchmark. They did, nonetheless, attempt to simulate the pattern using Rastle and Coltheart's (1998) items, but were unable to produce a whammy effect. Though this would seem to negate the need to test the role of feedback in producing the effect, it is possible that feedback is somehow hampering CDP+'s performance, and that removing feedback would allow CDP+ to simulate an effect it had been unable to produce by default. Consequently, I submitted Rastle and Coltheart's items to both versions of CDP+. Neither CDP+ (F) nor CDP+ (NF) was able to simulate the whammy effect ($t < 1$ in both cases; See Table 23 for simulation results).

Table 23. Mean cycles to response (percentage errors) to the Rastle and Coltheart (1998) items for CDP+ (F) and CDP+ (NF).

	Whammies		No Whammies	
Model	Cycles	%E	Cycles	%E
CDP+ (F)	152.35	(0.0%)	153.52	(8.3%)
CDP+ (NF)	152.09	(0.0%)	152.83	(4.2%)

General Discussion

This study has examined the role that feedback plays in CDP+'s ability to simulate a wide range of phenomena in reading aloud. The ultimate result is that feedback is not needed to simulate any of the phenomena identified as benchmarks by Perry et al. (2007) and Coltheart et al. (2001). Even the pseudohomophone advantage, which on the surface would seem to require feedback, can be simulated without feedback by making small adjustments to other parameters to allow a stronger lexical influence. Based on the simulation results presented here, it would seem that the widespread adoption of IA is premature, and that the assumptions that underlie IA warrant closer scrutiny. It remains to be seen if any phenomenon in reading aloud will challenge this view. In this section I address a few other issues that bear further consideration.

Other Potential Benchmarks

Coltheart et al. (2001) and Perry et al. (2007) identified a range of benchmarks that they felt were important for any computational model to simulate. Reviewing these phenomena, they generally meet the following three criteria: they are well-established, having been replicated in the literature, usually in multiple laboratories and studies; the computational models offer ways of testing the underlying theories (e.g., acquired dyslexia), or otherwise incorporate the effect into the model directly (e.g., word frequency); and the models are successful at simulating them, though in some cases one or other of the two models (DRC and CDP+) outperforms the other. Some phenomena that are within the purview of the theoretical models are not included because the computational models are not yet able to simulate them. For example, the theoretical dual-route models include a semantic system, but the computational models have yet to implement one, thus although semantic priming and imageability effects are well established, they are not included in any of the benchmarks.

Computational modelers are, understandably, biased towards selecting benchmarks that their models are capable of simulating. I believe that there are advantages to expanding the benchmarks to other phenomena where the models have yet to succeed. The first is that it will make researchers more aware of the limitations of the models. Authors necessarily focus on the strengths of their models, and it is our responsibility as colleagues to ensure that researchers making use of these models do not lose sight of their limitations. The second advantage is that by identifying a phenomenon as a benchmark before it is simulated, modelers are provided with an incentive to direct their efforts at the phenomena of most interest to the field at large. Here I propose a few phenomena that I believe are worthy of benchmark status.

Stimulus quality and word frequency.

Word frequency is already a benchmark that the models handle easily. Joint manipulation of word frequency and stimulus quality, however, has been ignored in compiling benchmarks. In lexical decision, this manipulation has a long history (beginning with Stanners, Jastrzembski, & Westbrook, 1975), where the two factors are found to be additive: word frequency effects are equivalent for low-quality stimuli and high-quality stimuli. In reading aloud, the history is shorter but more complex: using different stimulus quality manipulations, Yap and Balota (2007) and O'Malley, Reynolds, and Besner (2007) found that whereas word frequency and stimulus quality are additive in lexical decision, they are overadditive in reading aloud such that low-quality stimuli show a larger frequency effect than do high-quality stimuli. More recently, O'Malley and Besner (2008) demonstrated that the qualitative difference in the pattern is related to the context, and not the task. When nonwords are intermixed with the words (as they must be in lexical decision, but usually are not in reading aloud), stimulus quality and word frequency are once again additive, *even in reading aloud*. It is this latter result that presents a particular challenge to computational modelers. We know of no demonstrations of additivity to date between stimulus quality and any other factor in a computational model of reading aloud (but see Ziegler et al., 2009; Besner & O'Malley, 2009).

Stimulus quality and regularity.

Besner et al. (2010) jointly manipulated stimulus quality and regularity, both in the presence of nonwords, and with only words present. The goal was to provide further support for the additive effects observed in O'Malley and Besner (2008) by demonstrating a similar pattern with a different lexical characteristic, and to strengthen the evidence for a contextual influence on the reading system. The additivity when nonwords are present was confirmed (when nonwords were present, stimulus quality and regularity were additive on both reaction times and errors, as was the case with word frequency), but with an additional surprise: when only words are present, regularity and stimulus quality are *underadditive*. Exception words were *less* affected by the stimulus quality manipulation than were regular words. Besner et al. (2010) provide both a demonstration of the underadditivity and a replication. These two patterns (additivity and underadditivity) present a significant challenge to computational modelers (but see Besner et al., 2010 for one way to simulate the underadditivity in CDP+).

The question of additivity is one that seems likely to become more and more important in computational models. To date, modelers have ignored experiments that produce such patterns, but the evidence that additivity is real continues to accumulate (see Besner, 2006, for a summary of numerous demonstrations of additivity that precede those describe here). At some point, modelers will be forced to accept these effects as benchmarks, and the sooner they address the issue, the better.

Semantic priming and imageability.

The benchmarks simulated in this study are based on dual-route models and selected by their proponents (Coltheart et al., 2001; Perry et al., 2007). Thus far, dual-route modelers have ignored semantic phenomena, likely because they have yet to implement a semantic system. Proponents of the Triangle model, on the other hand, have addressed the issue of the semantic system to some extent. Plaut and Booth (2000) implemented the orthographic and semantic systems in a model designed to simulate the lexical decision task. In reading aloud, Plaut et al. (1996) approximated the semantic system's influence on phonology but without a true semantic system. In the most complete version of the Triangle model to date, Harm and Seidenberg (2004) implemented all three modules (orthographic, phonological, and semantic) but not all of their inter-connections: the semantic-phonological pathway was fully implemented, but the orthographic system acted only as an input to both the semantic and phonological systems; it did not receive connections from those modules (see Figure 5). Nevertheless, Harm and Seidenberg (2004) demonstrated that their model was sensitive to both imageability (Simulation 5: words high in imageability are named faster than words low in imageability), and semantic priming (Simulation 16: words are read aloud more rapidly when preceded by semantically related items, e.g., FROG–TOAD, than by unrelated items, e.g., NOSE–TOAD).

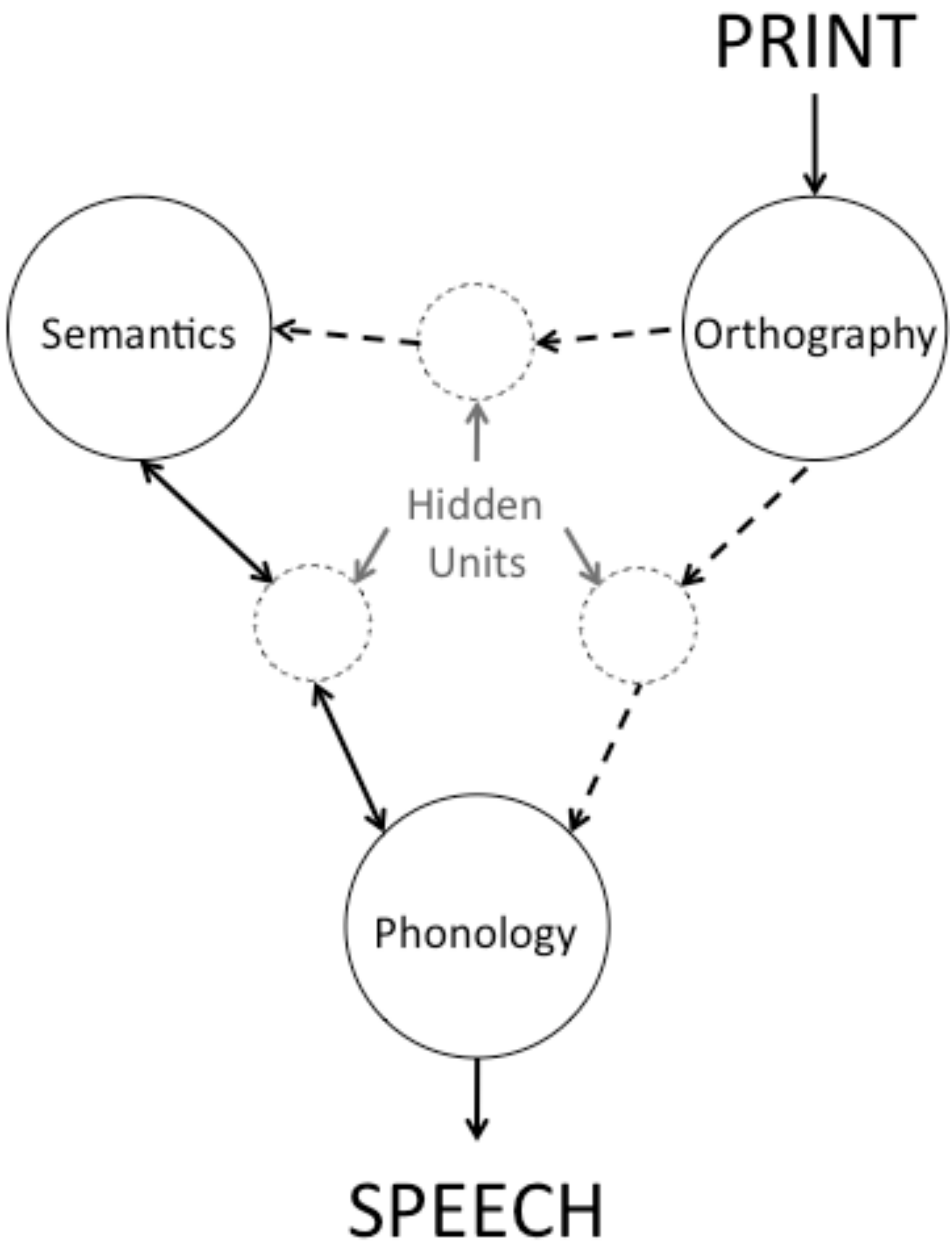

Figure 5. Structure of the implemented version of the Triangle model from Harm and Seidenberg (2004). Dotted arrows indicate activation flowing in only one direction, whereas the solid, double-headed arrows indicate interactive activation.

For the two classes of models (Triangle and dual-route) to be compared, the semantic system and its influence on reading aloud should, in all fairness, be included in any evaluations. Thus it is my view that, imageability and semantic priming should be included as benchmarks.

Disabling Feedback

The present work provides an existence proof for a set of parameters with no feedback in CDP+ that is capable of producing all of the identified benchmarks with as much success as the default model (with feedback). I am fortunate that arriving at that set of parameters required nothing more than simply setting the feedback parameters to 0 (with the exception of modeling the pseudohomophone advantage), however, this will not always be the case. The parameters in these complex models are inter-dependent, such that changes to one parameter will often necessitate changes to other parameters in order to compensate (e.g., see the simulations of the pseudohomophone advantage). Thus future models may require more effort to find a suitable parameter set with no feedback.

In particular, it seems very unlikely that the simple approach employed here would work in a trained PDP network such as the Triangle Model. Network connection strengths in PDP models are learned through a complex and computationally intensive training process until the network's "knowledge" is directly embedded in the pattern of connection weights. Simply removing all of the feedback in such a model would likely decimate any knowledge the model gained in training. In PDP models, carrying out a study such as the one reported here for CDP+ would require training a model with no feedback connections and comparing its performance to that of a network trained with feedback connections. The training process itself is subject to several variables that can be manipulated which creates yet another challenge: it may be that a model with no feedback requires somewhat different training conditions in order to match the performance of a model with feedback.

Woollams et al. (2007) do provide one example of a feed-forward only version of a Triangle model applied to surface dyslexia. They simulated surface dyslexia (in conjunction with semantic dementia) in a version of the Triangle model that included only an orthographic to phonology pathway (with an approximation of a semantic system influence on phonology). Though there is debate about its success (see Coltheart, Tree, & Saunders, 2010a, 2010b; Woollams, Lambon Ralph, Plaut, & Patterson, 2010a, 2010b), it is notable that the Woollams et al. (2007) model included no feedback connections, and thus they have demonstrated that feedback from phonology to orthography is unnecessary in the Triangle model for at least one benchmark.

Conclusion

Since McClelland and Rumelhart's seminal work (1981; Rumelhart & McClelland, 1982) introducing the concept of interactive activation, those principles have been incorporated into all of the major theoretical models of reading aloud. Despite its popularity, the assumptions built into the IA framework have yet to be closely examined. This study represents a first pass at evaluating one of the key assumptions underlying IA: the presence of feedback. Using CDP+ (Perry et al., 2007), I evaluated the role that feedback played in a set of benchmark phenomena that have been identified as key to evaluating computational models of reading aloud, first with the DRC (Coltheart et al., 2001) and then with CDP+. In every case, the data were equally well described by a model with no feedback, and, in the case of nonword reading, removing feedback actually improved performance.

The title of this paper is borrowed from Norris, McQueen, & Cutler (2000), so I will turn to them again to summarize my conclusions. In testing the role of feedback in speech recognition, Norris et al. (2000) stated that "…although the assumption of interaction fits with many people's intuitions about the nature and complexity of the speech recognition process, it is certainly not forced by the data." (p. 301) and "We cannot prove the null hypothesis that no interaction takes place. In our view, however, this remains one of the best arguments for adopting autonomous theories as the default option in this field: Occam's razor dictates that we do so." (p. 324) In the speech recognition literature, "autonomous" is the term used for models that do not include feedback. Consequently, these statements could just as well be made about processes in reading aloud. Modelers in reading aloud have been too quick to adopt the IA framework, and, for the moment at least, Norris et al.'s title applies just as well here: Feedback is never necessary.

Endnotes

[1] CDP+ (Perry et al., 2007) does have a modified form of staged processing between the letter level and the sublexical system (specifically, the grapheme parser). Letters are only made available to the grapheme parser once a given threshold of activation is attained. Once that level is reached, the relevant letter is "turned on" in the grapheme parser. In effect, this means that the letter level passes information on to the grapheme parser in a discrete way once it has identified a letter unambiguously. This activation threshold is not the only factor that determines the grapheme parser's access to the letters, though. To mimic an attentional system that scans the letters from left to right, the grapheme parser can only see one letter for every 15 cycles (i.e., only the first letter is initially available, subject to the activation threshold; after 15 cycles, the first two letters are available, subject to the activation threshold; and so on). Sternberg (1969) did not consider this combination of an activation-based threshold and a time-based threshold in his treatment of discrete stages.

[2] Paap et al. (1982) formalize the decision about which level of information to use as well, but that level of detail is not needed for our discussion. It suffices to note that on some trials the decision is made at the word level, whereas on others it is made at the letter-level.

[3] This notation is based on the International Phonetic Alphabet. The output from CDP+ is based on a customized notation system using only basic alphanumeric and punctuation characters. The mapping between the symbols used in the output of the models and those specified in the IPA are available in Appendix A.

[4] There is a newer version (CDP++) that has been extended to disyllabic words (Perry, Ziegler, & Zorzi, 2010). This newest version is not yet available for public use, and is considerably slower to run than CDP+ (C. Perry, personal communication, May 22, 2010).

[5] In my view, Dual-Route *Cascaded* is a misnomer for the DRC model because it ignores that the model includes feedback connections and within-level inhibitory connections along the lexical route. To avoid ambiguity, this thesis will only use the term cascaded to describe processing that is consistent with the framework defined in the original work by McClelland (1979).

[6] Perry et al. (2007) reported 98.67% accuracy, suggesting that they found the model correctly named two items more than my replication suggests. I have no explanation for the discrepancy.

[7] This correlation considers only the items that were correctly named by both CDP+ (F) and CDP+ (NF). No effort has been made to remove outliers.

[8] To give a sense of the importance of this issue, CDP+'s lexicon is based on DRC's, which was developed with an Australian dialect in mind. The authors of CDP+, however, have RP English, French (France), and Italian as their mother tongues. I have French (Canada) as my mother tongue, though English (Canada) is my primary language.

[9] There are likely many sources for the variance. In addition to simply variance between subjects and subject pools (which can't be controlled for), the variance could also arise from other differences in experimental designs (e.g., timing of stimulus presentation, nature of fixation points, brightness of the display, etc…), or in differences in the equipment used to

make the reaction time measurements (e.g., relying on voice key triggers vs. analysis of wave recordings).

[10] The *summed frequency* of a set of neighbors is generally going to be confounded with the *number* of neighbors. See Jared, McRae, and Seidenberg (1990) for evidence that the former is the more influential of the two measures.

[11] Perry et al. (2007) also reported successful simulation of the regularity by position effect using the items from Rastle and Coltheart (1999). I omit the details of this simulation here because it is known that those items included a grapheme consistency confound. The Roberts et al. (2003) stimulus list controlled for this confound and thus provides a more stringent test for this benchmark. Briefly though, the presence or absence of feedback played no role in CDP+'s performance on the Rastle and Coltheart (1999) items.

[12] This does not agree with the values reported in Perry et al. (2007). I have been unable to resolve the discrepancy – however, the results do not differ in any significant way from those reported by Perry et al. (2007).

[13] There are minor discrepancies between the values here, and those reported in Perry et al. (2007). Those discrepancies can be attributed to Perry et al. mistakenly treating the pronunciations of `POOK` and `DOOK` to rhyme with `TOOK` as regular. According to the rules set out by Andrews and Scarratt (1998), the regular pronunciation would pronounce the `OO` as in `BOOT`, not `TOOK`.

[14] The discrepancies between the values reported here, and those reported in Perry et al. (2007) are larger than in the experiment 1 simulations. Though I do not have the Perry et al.

results for individual items, the discrepancy is not important because so few of the pronunciations differed between the two models. No matter how the model responses are evaluated, CDP+ (F) and CDP+ (NF) will produce nearly identical results.

[15] Perry et al. (2007) treated length as a categorical variable. Because the assumption is that cycle times would increase monotonically with increasing length, it might make more sense to treat this variable as continuous. I conducted the analysis using both specifications, and found that it made no difference to the conclusions.

[16] There are other accounts of the pseudohomophone advantage that do not rely on the same mechanisms. For example, Borowsky, Owen, & Masson, (2002) propose a form of lexical checking that could give rise to the effect even in the cascaded version of the model. No such mechanism is implemented in any computational model of visual word recognition, so I ignore this account for the present purposes. (See Reynolds and Besner, 2005, for other accounts of the pseudohomophone advantage).

[17] The cell means and degrees of freedom differ somewhat from those reported by Perry et al. (2007) because there were three differences in our assessments of accuracy. Perry et al. accepted as correct an alternate pronunciation of `GOLPH` (`GOLF`) that is not one stored in CDP+'s lexicon. Similarly, they identified `GOOL` as incorrect, when in fact CDP+ generates the pronunciation that matches its stored pronunciation for `GHOUL`. To remain consistent with CDP+'s dialect, I have identified `GOLPH` as an error, and accepted `GOOL` as correct. Also, whereas in some dialects of English `BRUVE` might be pronounced to rhyme with `LOVE`, in the

dialect known to CDP+, it should rhyme with PROVE. Consequently, only that pronunciation was accepted.

[18] I'd like to thank Conrad Perry for his providing the items for the various simulations, and for his help in clarifying the methods used in Perry et al. (2007) whenever needed.

[19] More recently, Mousikou, Coltheart, Finkbeiner, and Saunders (2010) have taken a closer look at the masked onset priming effect (or MOPE). In a first experiment, they found that two-letter overlap primes (SIF–SIB) provided a greater advantage than one-letter overlap primes (SUF–SIB). In a second experiment, they found the priming effects increased with prime duration. These results are likely to be important in evaluating and comparing models in the future. For the moment, CDP+ (F) fails to capture either the one- or two-letter overlap priming effects with these stimuli, so this data is not considered further.

[20] It's worth noting that this item is subject to the "W" influence (Venezky, 1970) wherein words that begin with "WO" or "WA" typically violate the rules, in the sense that the vowel is frequently pronounced differently than it would be if preceded by another consonant cluster (e.g., WORM vs. FORM, DORM, NORM; or WARM vs. HARM, and FARM).

[21] Though the benchmarks describe some of the broader characteristics of phonological dyslexia (PD), PD is a well-studied disorder. Indeed there is an entire issue of Cognitive Neuropsychology devoted to the topic (volume 13, issue 6). It is also worth noting that whereas Coltheart's account relies on feedback, there is another view promoted by Patterson, Suzuki, & Wydell (1996): that PD is the result of impairment, not in sublexical processes, but in the phonological system of the Triangle Model (but see Caccappolo-van Vliet, Miozzo, & Stern,

2004). In the dual-route models, this could be thought of as damage to the phonemic buffer itself, rather than in the sublexical processes. This view does not rely on IA. Unfortunately, CDP+ does not offer any useful parameter manipulations to test their theory.

[22] Perry et al. (2007) reported that the default version of CDP+ produced a significant correlation (r = -.13, $p < .05$, N = 224). The discrepancy between that result, and the one reported here for CDP+ (F) is likely due to the challenge of identifying errors with nonwords. It is likely that our final sets of accurate responses did not overlap perfectly.

References

Adams, M. J. (1979). Models of word recognition. *Cognitive Psychology, 11*(2), 133-176. doi:10.1016/0010-0285(79)90008-2

Adelman, J. S., & Brown, G. D. A. (2007). Phonographic neighbors, not orthographic neighbors, determine word naming latencies. *Psychonomic Bulletin & Review, 14*(3), 455-459.

Aderman, D., & Smith, E. E. (1971). Expectancy as a determinant of functional units in perceptual recognition. *Cognitive Psychology, 2*(1), 117-129. doi:10.1016/0010-0285(71)90005-3

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(5), 802-814. doi:10.1037/0278-7393.15.5.802

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(2), 234-254. doi:10.1037/0278-7393.18.2.234

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review, 4*(4), 439-461.

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. doi:10.1037/0096-1523.24.4.1052

Ashby, F. G. (1982). Deriving exact predictions from the cascade model. *Psychological Review, 89*(5), 599-607. doi:10.1037/0033-295X.89.5.599

Balota, D. A., & Spieler, D. H. (1998). The utility of item-level analyses in model evaluation: A reply to Seidenberg and Plaut. *Psychological Science, 9*(3), 238-240. doi:10.1111/1467-9280.00047

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*(3), 445-459.

Behrmann, M., & Bub, D. (1992). Surface dyslexia and dysgraphia: Dual routes, single lexicon. *Cognitive Neuropsychology, 9*(3), 209-251. doi:10.1080/02643299208252059

Berndt, R. S., Haendiges, A. N., Mitchum, C. C., & Wayland, S. C. (1996). An investigation of nonlexical reading impairments. *Cognitive Neuropsychology, 13*(6), 763-801. doi:10.1080/026432996381809

Besner, D. (2006). Visual language processing and additive effects of multiple factors on timed performance: A challenge for the interactive activation framework? Retrieved July 16, 2007, from PsyCrit Web site: http://psycrit.com/wikiup/6/6a/Besner2006.pdf

Besner, D., & O'Malley, S. (2009). Additivity of factor effects in reading tasks is still a challenge for computational models: Reply to Ziegler, Perry, and Zorzi (2009). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(1), 312-316. doi:10.1037/a0014555

Besner, D., O'Malley, S., & Robidoux, S. (2010). On the joint effects of stimulus quality, regularity, and lexicality when reading aloud: New challenges. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(3), 750-764. doi:10.1037/a0019178

Besner, D., Wartak, S., & Robidoux, S. (2008). Constraints on computational models of basic processes in reading. *Journal of Experimental Psychology: Human Perception and Performance, 34*(1), 242-250. doi:10.1037/0096-1523.34.1.242

Borowsky, R., Phillips, J., & Owen, W. (2003 June). Phonetic decoding processes: On the generalization of the dissociation of frequency and lexicality effects in pseudohomophone naming. *Paper Presented at the 13th Annual Meeting of the Canadian Society for Brain Behaviour and Cognitive Science,* Hamilton, ON.

Borowsky, R., Owen, W. J., & Masson, M. E. J. (2002). Diagnostics of phonological lexical processing: Pseudohomophone naming advantages, disadvantages and base-word frequency effects. *Memory & Cognition, 30*(6), 969-987.

Caccappolo-van Vliet, E., Miozzo, M., & Stern, Y. (2004). Phonological dyslexia: A test case for reading models. *Psychological Science, 15*(9), 583-590. doi:10.1111/j.0956-7976.2004.00724.x

Coltheart, M. (2010). Lessons from cognitive neuropsychology for cognitive science: A reply to Patterson and Plaut (2009). *Topics in Cognitive Science, 2*, 3-11.

Coltheart, M., & Rastle, K. (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception and Performance, 20*(6), 1197-1211. doi:10.1037/0096-1523.20.6.1197

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

Coltheart, M., Tree, J. J., & Saunders, S. J. (2010). Computational modeling of reading in semantic dementia: Comment on Woollams, Lambon Ralph, Plaut, and Patterson (2007). *Psychological Review, 117*(1), 256-271. doi:10.1037/a0015948

Coltheart, M., Tree, J. J., & Saunders, S. J. (2010). Postscript: Reading in semantic dementia—A response to Woollams, Lambon Ralph, Plaut, and Patterson (2010). *Psychological Review, 117*(1), 271-272. doi:10.1037/0033-295X.117.1.271

Derouesné, J., & Beauvois, M. F. (1985). The "phonemic" stage in the non-lexical reading process: Evidence from a case of phonological alexia. In K. Patterson, J. C. Marshall & M. Coltheart (Eds.), *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (pp. 399-458). London: Erlbaum.

Dunn, J. C., & Kirsner, K. (2003). What can we infer from double dissociations? *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, 39*(1), 1-7. doi:10.1016/S0010-9452(08)70070-4

Ferguson, R., Robidoux, S., & Besner, D. (2009). Reading aloud: Evidence for contextual control over lexical activation. *Journal of Experimental Psychology: Human Perception and Performance, 35*(2), 499-507. doi:10.1037/a0013162

Forster, K. I., & Davis, C. (1991). The density constraint on form-priming in the naming task: Interference effects from a masked prime. *Journal of Memory and Language, 30*(1), 1-25. doi:10.1016/0749-596X(91)90008-8

Funnell, E. (1983). Phonological processes in reading: New evidence from acquired dyslexia. *British Journal of Psychology, 74*(2), 159-180.

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. doi:10.1037/0096-1523.5.4.674

Grainger, J., & Jacobs, A. M. (1994). A dual read-out model of word context effects in letter perception: Further investigations of the word superiority effect. *Journal of Experimental Psychology: Human Perception and Performance, 20*(6), 1158-1176. doi:10.1037/0096-1523.20.6.1158

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103*(3), 518-565. doi:10.1037/0033-295X.103.3.518

Grainger, J., Spinelli, E., & Ferrand, L. (2000). Effects of baseword frequency and orthographic neighborhood size in pseudohomophone naming. *Journal of Memory and Language, 42*(1), 88-102. doi:10.1006/jmla.1999.2675

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review, 106*(3), 491-528. doi:10.1037/0033-295X.106.3.491

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:10.1037/0033-295X.111.3.662

Jared, D. (2002). Spelling-sound consistency and regularity effects in word naming. *Journal of Memory and Language, 46*(4), 723-750. doi:10.1006/jmla.2001.2827

Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language, 29*(6), 687-715. doi:10.1016/0749-596X(90)90044-Z

Johnston, J. C. (1978). A test of the sophisticated guessing theory of word perception. *Cognitive Psychology, 10*(2), 123-153. doi:10.1016/0010-0285(78)90011-7

Kucera, H., & Francis, W. N. (1967). *Computational Analysis of Present-Day American English*. Providence: Brown University Pess.

Lichtheim, L. (1885). On aphasia. *Brain, 7*, 433-484.

Lupker, S. J., Brown, P., & Colombo, L. (1997). Strategic control in a naming task: Changing routes or changing deadlines? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(3), 570-590. doi:10.1037/0278-7393.23.3.570

Massaro, D. W. (1979). Letter information and orthographic context in word perception. *Journal of Experimental Psychology: Human Perception and Performance, 5*(4), 595-609. doi:10.1037/0096-1523.5.4.595

Massaro, D. W., & Cohen, M. M. (1991). Integration versus interactive activation: The joint influence of stimulus and context in perception. *Cognitive Psychology, 23*(4), 558-614. doi:10.1016/0010-0285(91)90006-A

Massaro, D. W., & Klitzke, D. (1979). The role of lateral masking and orthographic structure in letter and word recognition. *Acta Psychologica, 43*(5), 413-426. doi:10.1016/0001-6918(79)90033-7

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:10.1037/0096-1523.13.1.14

McCarthy, R. A., & Warrington, E. K. (1986). Phonological reading: Phenomena and paradoxes. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, 22*(3), 359-380.

McClelland, J. L. (1976). Preliminary letter identification in the perception of words and nonwords. *Journal of Experimental Psychology: Human Perception and Performance, 2*(1), 80-91. doi:10.1037/0096-1523.2.1.80

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86*(4), 287-330. doi:10.1037/0033-295X.86.4.287

McClelland, J. L. (1987). The case for interactionism in language processing. In M. Coltheart (Ed.), *Attention and performance 12: The psychology of reading.* (pp. 3--36). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

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. doi:10.1037/0033-295X.88.5.375

Montant, M., & Ziegler, J. C. (2001). Can orthographic rimes facilitate naming? *Psychonomic Bulletin & Review, 8*(2), 351-356.

Morton, J. (1980). The logogen model and orthographic structure. In U. Frith (Ed.), *Cognitive processes in spelling* (). London: Academic Press.

Mousikou, P., Coltheart, M., Finkbeiner, M., & Saunders, S. J. (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:10.1080/17470210903156586

Mulatti, C., Reynolds, M. G., & Besner, D. (2006). Neighborhood effects in reading aloud: New findings and new challenges for computational models. *Journal of Experimental Psychology: Human Perception and Performance, 32*(4), 799-810. doi:10.1037/0096-1523.32.4.799

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23*(3), 299-370. doi:10.1017/S0140525X00003241

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review, 85*(3), 172-191. doi:10.1037/0033-295X.85.3.172

O'Malley, S., & Besner, D. (2008). Reading aloud: Qualitative differences in the relation between stimulus quality and word frequency as a function of context. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(6), 1400-1411. doi:10.1037/a0013084

O'Malley, S., Reynolds, M. G., & Besner, D. (2007). Qualitative differences between the joint effects of stimulus quality and word frequency in reading aloud and lexical decision: Extensions to Yap and Balota (2007). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(2), 451-458. doi:10.1037/0278-7393.33.2.451

Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation–verification model for letter and word recognition: The word-superiority effect. *Psychological Review, 89*(5), 573-594. doi:10.1037/0033-295X.89.5.573

Paap, K. R., & Noel, R. W. (1991). Dual-route models of print to sound: Still a good horse race. *Psychological Research/Psychologische Forschung, 53*(1), 13-24. doi:10.1007/BF00867328

Patterson, K., & Behrmann, M. (1997). Frequency and consistency effects in a pure surface dyslexic patient. *Journal of Experimental Psychology: Human Perception and Performance, 23*(4), 1217-1231. doi:10.1037/0096-1523.23.4.1217

Patterson, K., & Plaut, D. C. (2009). ‘‘Shallow draughts intoxicate the brain’’: Lessons from cognitive science for cognitive neuropsychology. *Topics in Cognitive Science, 1*, 39-58.

Patterson, K., & Plaut, D. C. (2010). Beyond functional architecture in cognitive neuropsychology: A reply to Coltheart (2010). *Topics in Cognitive Science, 2*, 12-14.

Patterson, K., Suzuki, T., & Wydell, T. N. (1996). Interpreting a case of Japanese phonological alexia: The key is in phonology. *Cognitive Neuropsychology, 13*(6), 803-822. doi:10.1080/026432996381818

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. doi:10.1037/0033-295X.114.2.273

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,* In press.

Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review, 107*(4), 786-823. doi:10.1037/0033-295X.107.4.786

Plaut, D. C., & Booth, J. R. (2006). More modeling but still no stages: Reply to Borowsky and Besner. *Psychological Review, 113*(1), 196-200. doi:10.1037/0033-295X.113.1.196

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. doi:10.1037/0033-295X.103.1.56

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

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:10.1037/0096-1523.25.2.482

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81*(2), 275-280. doi:10.1037/h0027768

Reynolds, M. G., & Besner, D. (2002). Neighbourhood density effects in reading aloud: New insights from simulations with the DRC model. *Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale, 56*(4), 310-318. doi:10.1037/h0087407

Reynolds, M. G., & Besner, D. (2004). Neighbourhood density, word frequency, and spelling-sound regularity effects in naming: Similarities and differences between skilled readers and the dual route cascaded computational model. *Canadian Journal of Experimental Psychology, 58*(1), 13-31.

Reynolds, M. G., & Besner, D. (2005). Basic processes in reading: A critical review of pseudohomophone effects in reading aloud and a new computational account. *Psychonomic Bulletin & Review, 12*(4), 622-646.

Reynolds, M. G., & Besner, D. (2010). Contextual control over the breadth of lexical activation when reading aloud. *Quarterly Journal of Experimental Psychology,* Manuscript submitted for publication.

Richman, H. B., & Simon, H. A. (1989). Context effects in letter perception: Comparison of two theories. *Psychological Review, 96*(3), 417-432. doi:10.1037/0033-295X.96.3.417

Roberts, M. A., Rastle, K., Coltheart, M., & Besner, D. (2003). When parallel processing in visual word recognition is not enough: New evidence from naming. *Psychonomic Bulletin & Review, 10*(2), 405-414.

Roberts, S., & Sternberg, S. (1993). The meaning of additive reaction-time effects: Tests of three alternatives. In D. E. Meyer, & S. Kornblum (Eds.), *This book is based on the papers that were presented at the Fourteenth International Symposium on Attention and Performance held at the University of Michigan, Ann Arbor, Jul 9-13, 1990.* (pp. 611-653). The MIT Press, Cambridge, MA: US.

Robidoux, S., & Besner, D. (2010). When benefits outweigh costs: Reconsidering “automatic” phonological recoding when reading aloud. *Canadian Journal of Experimental Psychology,* In press.

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. doi:10.1037/0033-295X.89.1.60

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*(4), 523-568. doi:10.1037/0033-295X.96.4.523

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. doi:10.1037/0096-1523.20.6.1177

Seidenberg, M. S., & Waters, G. S. (1989). Word recognition and naming: A mega study. *Bulletin of the Psychonomic Society, 27*, 489.

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science, 8*(6), 411-416. doi:10.1111/j.1467-9280.1997.tb00453.x

Stanners, R. F., Jastrzembski, J. E., & Westbrook, A. (1975). Frequency and visual quality in a word-nonword classification task. *Journal of Verbal Learning & Verbal Behavior, 14*(3), 259-264.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica, Amsterdam, 30*, 276-315.

Sternberg, S. (1998). Inferring mental operations from reaction-time data: How we compare objects. In D. Scarborough, & S. Sternberg (Eds.), (pp. 365-454). Cambridge, MA, US: The MIT Press.

Sternberg, S. (2001). Separate modifiability, mental modules, and the use of pure and composite measures to reveal them. *Acta Psychologica, 106*(1-2), 147-246. doi:10.1016/S0001-6918(00)00045-7

Stolz, J. A., & Neely, J. H. (1995). When target degradation does and does not enhance semantic context effects in word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(3), 596-611. doi:10.1037/0278-7393.21.3.596

Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word pronunciation. *Journal of Memory and Language, 26*(6), 608-631. doi:10.1016/0749-596X(87)90105-7

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General, 124*(2), 107-136. doi:10.1037/0096-3445.124.2.107

Venezky, R. L. (1970). *The structure of English orthography*. The Hague, The Netherlands: Mouton.

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, 50*(2), 439-456. doi:10.1080/027249897392170

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology, 1*(1), 59-85. doi:10.1016/0010-0285(70)90005-8

Woollams, A. M., Lambon Ralph, M. A., Plaut, D. C., & Patterson, K. (2007). SD-squared: On the association between semantic dementia and surface dyslexia. *Psychological Review, 114*(2), 316-339. doi:10.1037/0033-295X.114.2.316

Woollams, A. M., Lambon Ralph, M. A., Plaut, D. C., & Patterson, K. (2010). Postscript: SD-squared revisited again. *Psychological Review, 117*(1), 282-283. doi:10.1037/0033-295X.117.1.282

Woollams, A. M., Lambon Ralph, M. A., Plaut, D. C., & Patterson, K. (2010). SD-squared revisited: Reply to Coltheart, Tree, and Saunders (2010). *Psychological Review, 117*(1), 273-281. doi:10.1037/a0017641

Yap, M. J., & Balota, D. A. (2007). Additive and interactive effects on response time distributions in visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(2), 274-296. doi:10.1037/0278-7393.33.2.274

Ziegler, J. C., Perry, C., Jacobs, A. M., & Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science, 12*(5), 379-384. doi:10.1111/1467-9280.00370

Ziegler, J. C., Perry, C., & Zorzi, M. (2009). Additive and interactive effects of stimulus degradation: No challenge for CDP+. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(1), 306-311. doi:10.1037/a0013738

Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance, 24*(4), 1131-1161. doi:10.1037/0096-1523.24.4.1131

Appendix A: Correspondence between CDP+ output symbols and the International Phonetic Alphabet

CDP+	IPA	Example	CDP+	IPA	Example
_ or Z	dʒ	age	j	j	**y**am
{	æ	c**a**n	J	tʃ	bir**ch**
#	aː	**a**ft, **ar**t	k	k	bar**k**
1	æɪ	c**a**ne	l	l	**l**earn
2	aɪ	file	m	m	**m**ean
3	ɜː	f**ir**m	n	n	**n**ame
4	oɪ	p**oi**nt	N	ŋ	si**ng**
5	əʉ	r**o**pe	p	p	**p**ower
6	æɔ	f**ou**nd	Q	ɔ	w**a**nt
7	ɪə	beard	r	ɹ	**r**ing
8	eː	dare	s	s	**s**un
9	oː	f**a**ll	S	ʃ	ra**sh**
b	b	**b**ad	t	t	**t**en
d	d	rai**d**	T	θ	**th**ought
D	ð	**th**at	u	ʉː	b**oo**t
E	e	b**e**d	U	ʊ	t**oo**k
f	f	**f**ox	v	v	**v**ary
g	g	**g**ame	V	a	ssn**u**b
h	h	**h**ouse	w	w	**w**ait
i	iː	d**ee**m	x	x	u**gh**
I	ɪ	film	z	z	zest