

# Nested Incremental Modeling in the Development of Computational Theories: The CDP+ Model of Reading Aloud

Conrad Perry  
Swinburne University of Technology

Johannes C. Ziegler  
Centre National de la Recherche Scientifique and Université de  
Provence

Marco Zorzi  
Università di Padova

At least 3 different types of computational model have been shown to account for various facets of both normal and impaired single word reading: (a) the connectionist triangle model, (b) the dual-route cascaded model, and (c) the connectionist dual process model. Major strengths and weaknesses of these models are identified. In the spirit of nested incremental modeling, a new connectionist dual process model (the CDP+ model) is presented. This model builds on the strengths of 2 of the previous models while eliminating their weaknesses. Contrary to the dual-route cascaded model, CDP+ is able to learn and produce graded consistency effects. Contrary to the triangle and the connectionist dual process models, CDP+ accounts for serial effects and has more accurate nonword reading performance. CDP+ also beats all previous models by an order of magnitude when predicting individual item-level variance on large databases. Thus, the authors show that building on existing theories by combining the best features of previous models—a nested modeling strategy that is commonly used in other areas of science but often neglected in psychology—results in better and more powerful computational models.

*Keywords:* reading, naming, word recognition, dual-route cascaded model, connectionist models

At least since Huey (1908), experimental and cognitive psychologists have been interested in describing the processes underlying skilled reading in a precise and detailed manner. Early attempts were purely verbal and qualitative, and box-and-arrow models of the reading process were ubiquitous (see, e.g., Morton, 1969).

---

Conrad Perry, Faculty of Life and Social Sciences, Swinburne University of Technology, Melbourne, Australia; Johannes C. Ziegler, Laboratoire de Psychologie Cognitive, Centre National de la Recherche Scientifique and Université de Provence, Marseille, France; Marco Zorzi, Dipartimento di Psicologia Generale, Università di Padova, Padova, Italy.

All authors contributed equally to this work; the order of authorship is alphabetical. An executable version of CDP+ can be downloaded at <http://ccnl.psy.unipd.it/CDP.html>. Data sets for all benchmark effects and all other studies included in this article can also be found at this site. Part of this work was done while Conrad Perry was supported by a University Development Fund grant from The University of Hong Kong. This research was also partially funded by a Procure travel grant (28/03T and 07806VE) between France and Hong Kong and a grant from the University of Padova to Marco Zorzi.

We thank Florian Hutzler for some of the triangle model simulations, Debra Jared for providing the raw data of Jared (2002), Brad Aitken for providing computing facilities, and Max Coltheart for many helpful discussions. Thanks are extended to Dave Balota, Derek Besner, Debra Jared, and David Plaut for very helpful comments.

Correspondence concerning this article should be addressed to Johannes C. Ziegler, Laboratoire de Psychologie Cognitive, Pôle 3C, Case D, CNRS et Université de Provence, 3 place Victor Hugo, 13331 Marseille, Cedex 3, France. E-mail: Johannes C. Ziegler, [ziegler@up.univ-mrs.fr](mailto:ziegler@up.univ-mrs.fr); Conrad Perry, [conradperry@gmail.com](mailto:conradperry@gmail.com); Marco Zorzi, [marco.zorzi@unipd.it](mailto:marco.zorzi@unipd.it)

With the emergence of connectionism, the modeling of aspects of the reading process experienced a quantum leap (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982; Seidenberg & McClelland, 1989). Purely verbal theories were successively replaced by explicit computational models. These models can produce highly detailed simulations of various aspects of the reading process, including word recognition and reading aloud (e.g., Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Grainger & Jacobs, 1996; Harm & Seidenberg, 1999, 2004; Norris, 1994; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989; Zorzi, Houghton, & Butterworth, 1998b). In addition, lesioning the models in various ways made it possible to compare the behavior of the models to that of neuropsychological patients with various reading impairments (i.e., acquired dyslexia; see Denes, Cipolotti, & Zorzi, 1999, for a review). This type of modeling improved understanding of both the fundamental processes involved in reading single words aloud and the causes underlying various reading disorders (see Zorzi, 2005, for a review).

Despite the huge progress in developing computational models, each model has its own fundamental limitations and problems in accounting for the wide range of available empirical data (see Previous Models section). The goal of the present research was to design a new model by building on the strengths of some of the previous models and eliminating their weaknesses. In other sciences, it is standard practice that a new model accounts for the crucial effects accounted for by the previous generations of the same or competing models. This strategy, often neglected in psychology, has sometimes been referred to as *nested modeling*: A

new model should be related to or include at least its own direct predecessors. The new model should also be tested against the data sets that motivated the construction of the old models before it is tested against new data sets (Jacobs & Grainger, 1994). To somewhat preempt the results, we show that building on existing theories by combining the best features of previous models results in a better and more powerful computational model.

The article is organized as follows. We begin by describing previous models of reading aloud, focusing on those that have been shown to account for both normal and impaired performance in written word naming. We then discuss the shortcomings of these models. In the second part of the article, we describe a new connectionist dual process model (the CDP+ model), which addresses the weaknesses of its predecessors while building on their strengths. In developing and testing the new model, we follow a nested modeling approach: That is, the model's architecture was designed in a way that exploited the best features of other models. The model was then tested against a full set of state-of-the-art benchmark effects. In the Results section, we compare the performance of the new model against its main competitors, and we provide an analysis that aims at determining the source of the improved performance.

### Previous Models

The most appealing feature of computational models is their high degree of explicitness and precision. Modelers need to specify input–output representations, learning algorithms, connectivity, and issues of model-to-data connection (i.e., how the data produced by the models can be related to the actual data that is collected). Implemented models thus become testable and falsifiable, and they can be compared with one another in a quantitative way.

Models are typically tested by presenting them with one or several lists of stimuli (words and nonwords) taken from a published study. Dependent measures of the model's performance are error rates and reaction times (RTs). The latter are simply the number of cycles taken by the model to produce a final output for each presented stimulus. Thus, the model's response latencies are collected for each item, and they are subsequently analyzed with appropriate statistics to assess the significance of the effect(s). In some cases, models can also be evaluated at the item level by regressing a model's latencies onto human latencies (Coltheart et al., 2001; Spieler & Balota, 1997; see later discussion).

A number of extant models may have the potential to account for a substantial amount of the empirical data from both skilled reading aloud and neuropsychological disorders of reading aloud following brain damage. Here, we examine three of them: (a) the parallel distributed processing (PDP) model of Seidenberg and McClelland (1989) and its various successors (e.g., Harm & Seidenberg, 1999; Plaut et al., 1996), (b) the dual-route cascaded (DRC) model of Coltheart et al. (1993, 2001), and (c) the connectionist dual process (CDP) model of Zorzi et al. (1998b). The main characteristics of these models are reviewed next (see Zorzi, 2005, for a more comprehensive review).

### The Triangle Model

The computational framework described by Seidenberg and McClelland (1989) has been referred to as the *triangle model* (see

Figure 1). The model assumes the existence of two pathways from spelling to sound: One pathway is a direct mapping from orthographic to phonological representations, whereas the second pathway maps from print to sound by means of the representation of word meanings.

Only the direct orthography-to-phonology pathway was implemented in Seidenberg and McClelland's (1989) seminal work. In the model, the phonology of any given word or nonword is computed from its orthographic representation by a single process. This process is the spread of activation through a three-layer neural network, where the activation patterns over input and output units represent the written and phonological form of the word, respectively. The knowledge about spelling–sound mappings is distributed in the network and resides in the connections that link the processing units.

The back-propagation learning algorithm was used to train the network on a set of nearly 3,000 monosyllabic English words (Rumelhart, Hinton, & Williams, 1986). Training items were presented to the network with a probability reflecting a logarithmic function of Kuçera and Francis's (1967) word frequency norms. Orthographic and phonological representations of words (and nonwords) consisted of activation patterns distributed over a number of primitive representational units following the triplet scheme of Wickelgren (1969). This model was able to account for various facts about the reading performance of normal participants. In particular, it showed the classic interaction between word frequency and spelling–sound regularity (e.g., Paap & Noel, 1991), an interaction that was previously taken as evidence supporting the dual-route model of reading (details of this model are described later).

The model was, however, criticized (see, e.g., Besner, Twilley, McCann, & Seergobin, 1990; Coltheart et al., 1993). Most important, Besner et al. (1990) tested the model on several lists of nonwords and found that it performed very poorly (more than 40% errors). However, Plaut et al. (1996) presented an improved version of the triangle model in which they abandoned the highly distributed Wickelfeature representation in favor of a more localist coding of orthographic and phonological units. This new model

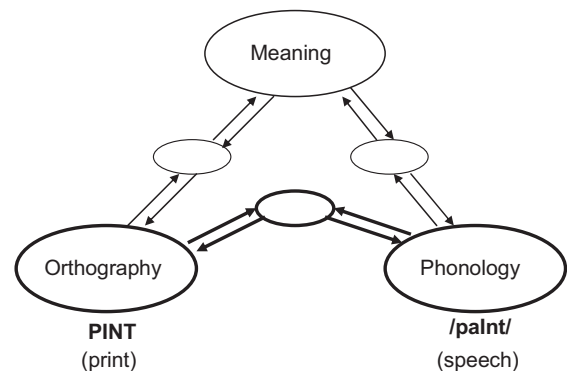


Figure 1. The triangle model. The part of the model implemented by Seidenberg and McClelland (1989) and also by Plaut et al. (1996) is shown in bold. From "A Distributed, Developmental Model of Word Recognition and Naming," by M. S. Seidenberg and J. L. McClelland, 1989, *Psychological Review*, 96, p. 526. Copyright 1989 by the American Psychological Association. Adapted with permission.

overcame some of the limitations of its predecessor; namely, the model was able to read several lists of nonwords with an error rate similar to that of human participants under time pressure (see also Seidenberg, Plaut, Petersen, McClelland, & McRae, 1994), although there were individual items that were pronounced differently from the way people typically pronounce them.

In more recent work, Harm and Seidenberg (2004) implemented a semantic component for the triangle model that maps orthography and phonology onto semantics. This allowed them to simulate a number of effects related to semantics, including homophone and pseudohomophone effects found in semantic categorization and priming experiments (e.g., Lesch & Pollatsek, 1993; Van Orden, 1987). They also added direct connections between orthography and phonology (i.e., not mediated by hidden units; see the CDP model described later), which Plaut et al.'s (1996) model did not have. They suggested that this modification had the effect of further improving the generalization performance (i.e., nonword reading) of the model (see Harm & Seidenberg, 2004, for further details).

### The DRC Model

In response to the single-route model of Seidenberg and McClelland (1989), Coltheart and colleagues (Coltheart et al., 1993, 2001; Coltheart & Rastle, 1994; Rastle & Coltheart, 1999; Ziegler, Perry, & Coltheart, 2000, 2003) developed a computational implementation of the dual-route theory. In this model, known as DRC, lexical and nonlexical routes are implemented as different and representationally independent components (see Figure 2). Moreover, the two routes operate on different computational principles: serial, symbolic processing in the nonlexical route and parallel spreading activation in the lexical route.

The nonlexical route of DRC applies grapheme-to-phoneme correspondence (GPC) rules in a serial left-to-right manner; it can be used on any string of letters and is necessary for reading nonwords. The lexical route, which is implemented as an interactive activation model based on McClelland and Rumelhart's (1981; see also Rumelhart & McClelland, 1982) word recognition model joined with something similar to Dell's (1986) spoken word production model, operates by means of parallel cascaded processing. Processing starts at a letter feature level, and then activation spreads to letters, orthographic word nodes (i.e., an orthographic lexicon), phonological word nodes (i.e., a phonological lexicon), and finally a phonological output buffer (i.e., the *phoneme system*). The lexical route can be used to read known words and is necessary for the correct reading of exception words (also called irregular words). Irregular/exception words contain, by definition, at least one grapheme pronounced in a way that does not conform to the most frequent grapheme-phoneme correspondences (e.g., the pronunciation of *ea* in *head* vs. *bead*).

One common type of error produced by the model is known as a regularization error. This type of error occurs if the nonlexical route is used to read an exception word without the lexical route being on, because the nonlexical route generates phonology through rules that specify only the most typical grapheme-phoneme relationships. For example, the nonlexical route would generate a pronunciation that rhymes with *mint* when presented with the word *pint*. Normally, the lexical and nonlexical routes always interact during processing (whether during the processing

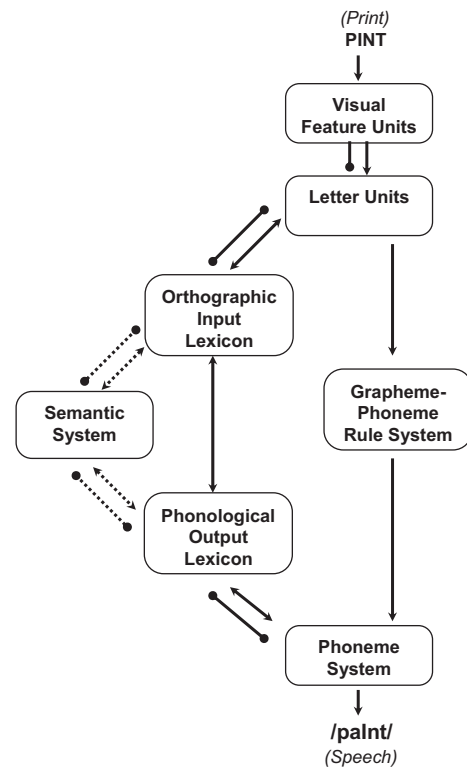


Figure 2. The dual-route cascaded model. The lexical-semantic route of the model is not implemented (dashed lines). From "Models of Reading Aloud: Dual-Route and Parallel-Distributed-Processing Approaches," by M. Coltheart, B. Curtis, P. Atkins, and M. Haller, 1993, *Psychological Review*, 100, p. 214. Copyright 1993 by the American Psychological Association. Adapted with permission.

of words or the processing of nonwords). However, the lexical route runs faster than the nonlexical route, which is why irregular words are usually (but not always) pronounced correctly by DRC, even though the nonlexical route generates a regularized pronunciation.

### The CDP Model

Zorzi et al. (1998b) developed a connectionist model of reading aloud where a dual-route processing system emerges from the interaction of task demands and initial network architecture in the course of reading acquisition. In this model, the distinction between phonological assembly and lexical knowledge is realized in the form of connectivity (either direct or mediated) between orthographic input and phonological output patterns (see Houghton & Zorzi, 1998, 2003, for a similar treatment of the problem of learning the sound-spelling mapping in writing). The model thus maintains the uniform computational style of the PDP models (i.e., connectionist architecture) but makes a clear distinction between lexical and sublexical processes in reading.

Zorzi et al. (1998b) studied in great detail the performance of a fully parallel simple two-layer associative network (i.e., a network without hidden units) trained to learn the mapping between orthography and phonology. Zorzi et al. found that this network acquired properties that are considered the hallmark of a phono-

logical assembly process. In particular, Zorzi et al. showed that the two-layer network of phonological assembly (TLA network) was able to extract the statistically most reliable spelling–sound relationships in English (see also Zorzi, Houghton, & Butterworth, 1998a, for a developmental study of this capacity), without forming representations of the individual training items (such as the exception words). Therefore, the two-layer associative network produces regularized pronunciations (if the input word is an exception word) and is not very sensitive to the base-word frequency of the trained words. Nonetheless, it is highly sensitive to the statistical consistency of spelling–sound relationships at multiple grain sizes (from letters to word bodies), which is reflected by the activation of alternative phoneme candidates in the same syllabic position (especially the vowel). The model’s final pronunciation is produced by a phonological decision system (i.e., a phonological output buffer) on the basis of activation competition, which is a causal factor in naming latencies. The model provides a reasonable match to the nonword reading performance of human participants and can also read single- and multiletter graphemes.

In the full CDP (see Figure 3), the assembled phonological code is pooled online with the output of a frequency-sensitive lexical process in the phonological decision system. Such an interaction allows the correct pronunciation of exception words and produces the latency (and/or accuracy) effects that depend on the combination of lexical and sublexical factors. Zorzi et al. (1998b) discussed the possibility of using either a distributed or a localist implementation of the lexical network, but most of their simulations were based on a localist version that was not fully implemented (i.e., the lexical phonology of a word was directly activated in the model without going through any lexical orthographic processing).

### Shortcomings of Previous Models

Although the three models were able to account for many of the critical effects in single word reading (for direct comparisons of these models, see Coltheart et al., 2001), each model has some

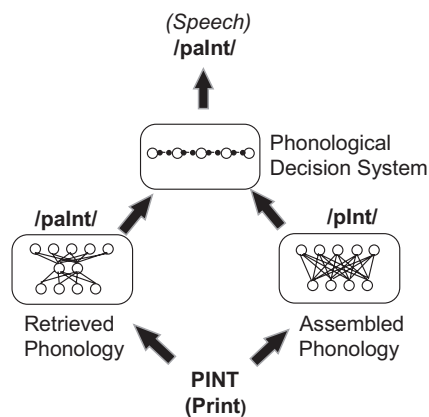


Figure 3. The connectionist dual process model. From “Two Routes or One in Reading Aloud? A Connectionist Dual-Process Model,” by M. Zorzi, G. Houghton, and B. Butterworth, 1998, *Journal of Experimental Psychology: Human Perception and Performance*, 24, p. 1150. Copyright 1998 by the American Psychological Association. Adapted with permission.

fundamental limitations, both qualitatively and quantitatively. These are summarized in Table 1 and are discussed next.

### Learning

A major shortcoming of DRC is the absence of learning. DRC is fully hardwired, and the nonlexical route operates with a partially hand-coded set of grapheme–phoneme rules. These rules include a number of context-specific rules (i.e., rules where a phoneme is assembled on the basis of information greater than a single grapheme) and phonotactic output rules (i.e., rules where the assembled pronunciation is changed to respect phonotactic constraints). The use of such complex rules certainly increases the performance of the model compared with one that would use only grapheme–phoneme rules. In earlier work, Coltheart et al. (1993) showed that a learning algorithm could, in principle, select a reasonable set of rules. This learning algorithm was able to discover not only consistent print–sound mappings, but also inconsistent mappings (i.e., cases where a spelling pattern maps onto more than one pronunciation). However, in the inconsistent cases, the less frequent alternatives were simply eliminated at the end of the learning phase, leaving only the most common mappings in the rule set. Although the learning of the GPC rules was abandoned in the most recent version of the model, the rules still operate in an all-or-none fashion. Because of the absence of learning, DRC cannot be used to simulate reading development and developmental reading disorders.

Both the triangle model and CDP are superior in this respect because the mapping between orthography and phonology is learned (see Hutzler, Ziegler, Perry, Wimmer, & Zorzi, 2004). However, the models use different learning algorithms. CDP uses one of the most simple learning rules, the *delta rule* (Widrow & Hoff, 1960), whereas the triangle model uses *error back-propagation* (e.g., Rumelhart et al., 1986). Although the latter is a generalization of the delta rule that allows training of multilayer networks, supervised learning by back-propagation has been widely criticized because of its psychological and neurobiological implausibility (e.g., Crick, 1989; Murre, Phaf, & Wolters, 1992; O’Reilly, 1998), and it has been argued that the validity of a network’s learning algorithm should be evaluated with respect to appropriate learning laws and learning experiments (e.g., Jacobs & Grainger, 1994). In this respect, CDP has an advantage over the triangle model because delta rule learning is formally equivalent to a classical conditioning law (the Rescorla–Wagner rule; see Sutton & Barto, 1981, for formal demonstration and Siegel & Allan, 1996, for a review of the conditioning law), and it has been widely applied to human learning (see, e.g., Gluck & Bower, 1988a, 1988b; Shanks, 1991; Siegel & Allan, 1996).

### Consistency Effects

The second major problem for DRC is the simulation of graded consistency effects. Glushko (1979) was the first to demonstrate the existence of a consistency effect. He compared two groups of words that were both regular according to grapheme–phoneme correspondence rules but differed in consistency. For example, the pronunciation of a regular inconsistent word, such as *wave*, can be correctly determined by rule. However, *wave* is nevertheless inconsistent because the *-ave* body is pronounced differently in *have*. Using such



Table 1  
*Qualitative Performance of DRC (Coltheart et al., 2001), the Triangle Model (Plaut et al., 1996), and CDP (Zorzi et al., 1998b) Across Different Theoretically Challenging Domains*

Model	Learning	Serial effects	Consistency effects	Database performance
DRC	no	yes	no	mixed
Triangle	yes	no	yes	poor
CDP	yes	mixed	yes	mixed

Note. DRC = dual-route cascaded model; CDP = connectionist dual process model.

items, he showed that inconsistent words took longer to name than consistent words of similar frequency. Subsequently, it was shown that consistency is a graded and continuous variable. In particular, Jared, McRae, and Seidenberg (1990) manipulated the summed word frequency (i.e., token count) of pronunciation *enemies* (i.e., words with the same body spelling pattern but a different pronunciation) and showed that the size of the consistency effect depended on the ratio of the summed frequency of *friends* (i.e., words with the same body and rime pronunciation, including the word itself) to *enemies* (see also Jared, 1997, 2002).

Obviously, DRC runs into problems simulating the consistency effect for regular words because all-or-none regularity rather than graded consistency determines processing in its nonlexical route. One argument against this is that consistency might affect naming latencies through neighborhood characteristics (i.e., the interaction between orthographically and phonologically similar words) of the lexical route (Coltheart et al., 2001). However, as we show later, such influences are too weak to account for the majority of the consistency effects reported in the literature. Also, DRC runs into problems with nonwords that use potentially inconsistent spelling–sound correspondences (e.g., Andrews & Scarratt, 1998). This is because nonword pronunciations are constructed from GPC rules that are selected in an all-or-none fashion. The fact that some graphemes might have alternative pronunciations simply does not enter into the model's computations at all. Thus, graded consistency effects in nonword pronunciations are a major challenge for DRC (see Andrews & Scarratt, 1998; Zevin & Seidenberg, 2006).

Both CDP and the triangle model are able to simulate graded consistency effects (e.g., Plaut et al., 1996; Zorzi et al., 1998b). The triangle model exhibits a body consistency effect, for both high- and low-frequency words (Coltheart et al., 2001; Jared, 2002). The effect depends on neighborhood characteristics and is present in both words and nonwords. It appears, however, that the triangle model is not sensitive to the effect of consistency defined at the level of individual letter–sound correspondences (Jared, 1997; Zorzi, 2000). With CDP, the output of the network reflects the relative consistency of a given mapping. That is, the model delivers not only the most common mapping of a grapheme but also less common mappings, which are activated to a lesser extent (Zorzi et al., 1998b). CDP is thus able to simulate graded consistency effects (also see Zorzi, 1999).

### *Serial Effects*

Accounting for serial effects is the strength of DRC and the weakness of the other two models. Perhaps the strongest evi-

dence for serial processing has come from the examination of length effects. In this respect, Weekes (1997) examined how long people would take to read aloud words and nonwords of different orthographic lengths (i.e., number of letters). His study examined word and nonword naming latencies for 300 monosyllabic items, which were equally divided into high-frequency words, low-frequency words, and nonwords. Within each of these three groups, orthographic length was manipulated by having an equal number of items with 3, 4, 5, and 6 letters. The results showed a systematic length effect for nonwords; that is, the longer the nonword, the longer it took participants to initiate pronunciation. Real words showed fewer systematic length effects (but see the Results section for a reanalysis of Weekes's, 1997, data; indeed, other studies did find systematic length effects with words, e.g., Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Jared & Seidenberg, 1990). Weekes argued that this was evidence for two interacting processes similar to those suggested by DRC. The main idea is that nonwords must be assembled from letters one by one, hence causing a length effect, whereas words are also influenced by a parallel lexical route, hence diminishing serial effects of the assembled phonology mechanism. A similar interaction was found by Ziegler, Perry, Jacobs, and Braun (2001).

The *position-of-irregularity effect* is another effect that suggests serial processing may occur when reading aloud (Coltheart & Rastle, 1994; Rastle & Coltheart, 1999; Roberts, Rastle, Coltheart, & Besner, 2003). The effect was based on the observation that the size of the regularity effect declines as a function of the position of a word's irregular grapheme–phoneme correspondence: Words with an irregular correspondence in the first position (e.g., *chef*) are read aloud more slowly than words with an irregular correspondence in the second position (e.g., *pint*), which are in turn read aloud more slowly than words with an irregular correspondence in the third position (e.g., *blind*). At the third position, words with an irregular correspondence are read aloud at a speed similar to that of words without an irregular correspondence (e.g., Coltheart & Rastle, 1994; Cortese, 1998; Rastle & Coltheart, 1999). This effect was taken to suggest that the generation of nonlexical phonology occurs in a left-to-right (serial) fashion, rather than a parallel fashion. More precisely, it was suggested that if a parallel and a serial route compete with each other at an output level (as they do in DRC), regularized phonology (i.e., phonology assembled incorrectly) should be more harmful early in processing (i.e., phonology assembled in early word positions), because it would allow more time to create conflict with the lexical route.

The triangle model does not show the interaction between regularity and position of the irregular grapheme (Coltheart et al., 2001) in the same way that people do. Zorzi (2000), however, showed that CDP was able to simulate the effect even though it is a purely parallel model. He therefore suggested that the position-of-irregularity effect is probably due to the fact that the position-of-irregularity manipulation is confounded with grapheme–phoneme consistency, a variable the CDP model is sensitive to. Subsequently, however, Roberts et al. (2003) replicated the position-of-irregularity effect using a group of stimuli that neither CDP nor the triangle model was able to simulate.

### *Predicting Item-Level Variance*

One of the main advantages of computational models is that they can predict performance at the item level. As Spieler and Balota (1997) pointed out,

This is an important aspect of these models because they have the ability to reflect the more continuous nature of relevant factors (e.g., frequency, regularity) in addition to the categorical manipulations reflected in the designs of typical word recognition studies. (p. 411)

Thus, one of the most challenging tests is to evaluate models at the item level by regressing model latencies onto human latencies (e.g., Coltheart et al., 2001; Spieler & Balota, 1997; Zorzi, 2000).

To facilitate model evaluation and comparison at the item level, Spieler and Balota (1997) collected naming latencies of 2,870 words. They then compared the triangle model's output with the mean naming performance of participants at the item level. They suggested that successful models should pass two critical tests: First, the amount of variance predicted by computational models should be at least as strong as the strongest correlating single factor. Second, the amount of variance predicted by computational models should be similar to the correlation derived from factors that are typically shown to be involved in reading, such as log word frequency, orthographic neighborhood, and orthographic length. These factors accounted for 21.7% of the variance of word naming latencies in the human data. Unfortunately, the three models accounted for much less variance (see Coltheart et al., 2001). CDP accounted for 7.73%, DRC accounted for 3.49%, and the triangle model accounted for 2.54% of the variance of the human naming latencies. Similar numbers were obtained on the Wayne State database (Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995), which contains RTs for all monosyllabic words that have a consonant–vowel–consonant phonological structure. Here, CDP accounted for 4.70% of the variance, whereas DRC and the triangle model accounted for 4.89% and 1.67%, respectively.

For predicting item variance, so far the biggest advantage of DRC over the other two models was obtained on a small-scale database, the so-called length database of Weekes (1997). Whereas the three models did not account for much variance in word reading (<5%), in accounting for variance in nonword reading, DRC showed clearly better performance (39.4%) than the other two models, which accounted for a nonsignificant amount of variance (<2% for both models).

### **CDP+: A Connectionist Dual Process Model of Reading Aloud**

The development of a new model was motivated by an attempt to address the shortcomings of the previous models discussed earlier. The principle of *nested modeling* dictates that the new model should be related to or include, at least, its own direct precursors and that it should also be tested against the data sets that motivated the construction of the old models before it is tested against new data sets. Our approach was to combine the best features of some of the previous models into a single new model. The new model was then tested on the classic data sets and on those data sets that constituted a challenge to the previous models.

The second goal of our modeling approach was *strong inference testing* (Platt, 1964). The core idea of strong inference testing is to devise alternative hypotheses and a crucial experiment with alternative possible outcomes, each of which exclude one or more of the hypotheses. When applied to the area of modeling, strong inference testing requires the implementation of alternative models, which are then tested on a crucial experiment, the outcome of which excludes one or more of the models. Of course, it is not always possible to find effects that exclude an entire class of models. Therefore, we need to be able to compare the *descriptive adequacy* of models, that is, the degree of accuracy of a model in predicting a data set both at the qualitative and quantitative level. Therefore, throughout the article, we compare the performance of the new model with that of the three other models discussed before (i.e., DRC, triangle, and CDP).

### *Architecture of the Model: Combining the Best of CDP and DRC*

The development of a new model should start by considering the weaknesses of previous models with respect to the critical effects discussed in the previous section. From Table 1, it is apparent that CDP offers a good starting point. Indeed, CDP is a learning model that accounts for consistency effects and, to some extent, for hypothesized serial effects in nonword reading aloud (even though it itself is a parallel model). Moreover, it accounts for a fair amount of variance on single word reading aloud (equal or slightly superior to DRC).

One way CDP can be significantly improved is by augmenting it with a fully implemented lexical route. Thus, in the spirit of nested modeling, we implemented a localist lexical route that is as close as possible to that of DRC and is based on the interactive activation model of McClelland and Rumelhart (1981). The advantage of this solution is that it should allow us to capture effects related to orthographic processing and lexical access previously simulated by Coltheart et al. (2001) and Grainger and Jacobs (1996). We discarded the alternative solution of implementing the semantic pathway of the triangle model (e.g., Harm & Seidenberg, 1999) for two main reasons. First, the ability of distributed models to account for lexical decision performance is still hotly disputed (see Borowsky & Besner, 2006, and Plaut & Booth, 2006, for opposing views). Second, the role of semantics in both normal and impaired naming of written words is controversial (see the General Discussion section). On a more practical side, using a symbolic localist lexical route makes it much simpler to evaluate the non-lexical part of the model.

A further problem for CDP, as well as for any connectionist learning model, is the relatively high error rate in nonword reading in comparison with DRC. Indeed, the ability of connectionist models to generalize the knowledge of spelling–sound mappings to novel items (i.e., nonwords) has been a source of concern and controversy since the seminal work of Seidenberg and McClelland (1989; also see Besner et al., 1990; Plaut et al., 1996; Seidenberg et al., 1994). However, one way of reducing the nonword error rate in connectionist models is to improve learning of spelling–sound relationships by using graphemes as orthographic input instead of single letters. This effectively alleviates the *dispersion* problem, which was identified by Plaut et al. (1996) as the main reason for the poor nonword reading performance of the triangle model. In

this case, by using better input and output representations, the frequency at which the same letters map onto the same phonemes is generally increased. This allows the most common statistical relationships in the data to be more easily learned.

A similar approach was taken by Houghton and Zorzi (2003) in their connectionist dual-route model of spelling, which contains a level of grapheme representations. The existence of grapheme representations in their model was explicitly linked to the notion of a *graphemic buffer*, which is endorsed by cognitive models of the spelling system (e.g., Caramazza, Miceli, Villa, & Romani, 1987; Ellis, 1988; Shallice, 1988). Specifically, Houghton and Zorzi assumed that “graphemic representations are syllabically structured, and complex graphemes (e.g., SH, TH, EE) are locally represented” (p. 112). The results of several studies of individuals with a specific acquired disorder of the graphemic buffer (e.g., Caramazza & Miceli, 1990; Caramazza et al., 1987; Cotelli, Abutalebi, Zorzi, & Cappa, 2003; Cubelli, 1991; Jónsdóttir, Shallice, & Wise, 1996) provide the primary motivation for the assumption that the representation is structured into *graphosyllables*, with onset, vowel, and coda constituents (Caramazza & Miceli, 1990; Houghton & Zorzi, 2003). Also, the data from normal readers suggest that graphemes are processed as perceptual units. Rey, Ziegler, and Jacobs (2000), for instance, showed that detecting a letter in a word was harder when this letter was embedded in a multiletter grapheme (e.g., *o* in *float*) than when it corresponded to a single-letter grapheme (e.g., *o* in *slope*). They suggested that this finding supports the idea that graphemes are functional units above the letter level (see also Martensen, Maris, & Dijkstra, 2003; Rey & Schiller, 2005).

We therefore implemented the graphemic buffer of Houghton and Zorzi (2003) as the input level of the sublexical orthography-to-phonology route in the new model. This choice was further motivated by the hypothesis that a common graphemic buffer is involved in both reading and spelling (Caramazza, Capasso, & Miceli, 1996; Hanley & Kay, 1998; Hanley & McDonnell, 1997), which received further support from the recent study of Cotelli et al. (2003). However, this also leads to the problem of how graphemic parsing is achieved. One solution is to perform a serial parsing of the string of letters from left to right and to submit each individuated grapheme to the sublexical network. This has the effect of *serializing* the sublexical route (i.e., the TLA network), which indeed resembles the serial assembly in the GPC route of DRC. Note, however, that we specifically link serial processing to the problem of graphemic parsing and more specifically to spatial attention mechanisms that control attention shifts from left to right over the letter string (Facoetti et al., 2006; see later discussion).

To illustrate the grapheme parsing mechanism, take, for example, the word *check*. On the basis of a syllabic representation, the first largest grapheme encountered, *ch*, should be assigned to the first onset position of the graphemic buffer. The next grapheme among the remaining letters (*eck*) is the vowel letter *e*, which should be assigned to the vowel position. The remaining two letters, *ck*, correspond to a single grapheme that should be assigned to the first coda position (a more detailed explanation is given later). Note that the serial nature of phonological assembly is simply a byproduct of the graphemic parsing process, because assembly in the sublexical network starts as soon as any grapheme is available as input. Nonetheless, the existence of serial operations in the new model should significantly improve the poor correlation

with human nonword naming latencies exhibited by CDP (see Coltheart et al., 2001). Indeed, serial processing in DRC seems to be a major strength with regard to its ability to account for the nonword length effect and for a much larger proportion of variance in nonword reading compared with any other model.

In summary, the new model combines the sublexical network of CDP (updated with graphemic representations and serial graphemic parsing) with the localist lexical network of DRC. The point of interaction between the two routes is the phonological output buffer, a competitive network where lexical and sublexical phonological codes are pooled online to drive the final pronunciation. This component of the model is representationally identical to the phonological decision system of CDP, which uses an onset–vowel–coda organization of the phonemes.<sup>1</sup> We refer to the new model as CDP+. The name acknowledges the greater similarity of the new model to one of its parent models (CDP) as compared with the other (DRC). A summary of the differences between the new model and its two parents is provided at the end of this section. Moreover, as is shown in the Results section, it turns out that the fully implemented lexical route (i.e., the component of the model that is taken from DRC) provides only a minor contribution to the new model’s success in simulating a broad range of phenomena in written word *naming*. The architecture of CDP+ is depicted in Figure 4.

### Implementation of the Model

#### Lexical Route

In our model, we implemented the lexical route of DRC, which is a fully interactive network based on McClelland and Rumelhart’s (1981) interactive activation model. We had to modify the output of DRC’s lexical network to be compatible with the output of our sublexical route, however. There are five levels in the lexical route of DRC (see Coltheart et al., 2001, for implementation details): a letter feature level, a letter level, an orthographic lexicon, a phonological lexicon, and a phonological output buffer. Activation begins at the feature level. The feature level is the same as the interactive activation model of McClelland and Rumelhart, except that instead of four feature detectors (one for each letter position) there are eight. Words of any length up to eight letters can be presented, but the absence of a letter at any of the eight positions must be represented by the activation of a *null letter*, which is coded by having all features of the corresponding detector on. For example, in terms of word coding, the null letter is put on to the end of all words up to the maximum number of letters in the model (which is eight for both CDP+ and DRC). Thus, the orthographic representation of each word should be thought of as the letters in the word plus the null letters (e.g., *cat\*\*\*\*\**, or *stong\*\*\**, where each asterisk is a null letter). Thus all orthographic words in the model are in fact eight letters long if we consider the null letter to be a letter no different from any others (which is what the model does). This coding scheme is used for any string presented to the model, including nonwords. Thus, if a three-letter nonword like *zat* is presented to the model, the last five empty positions are filled with null letters (i.e., *zat\*\*\*\*\**).

<sup>1</sup> Note that in Zorzi et al. (1998b), the term *rime* was used to refer to both the vowel and the coda, whereas here we use the separate terms.

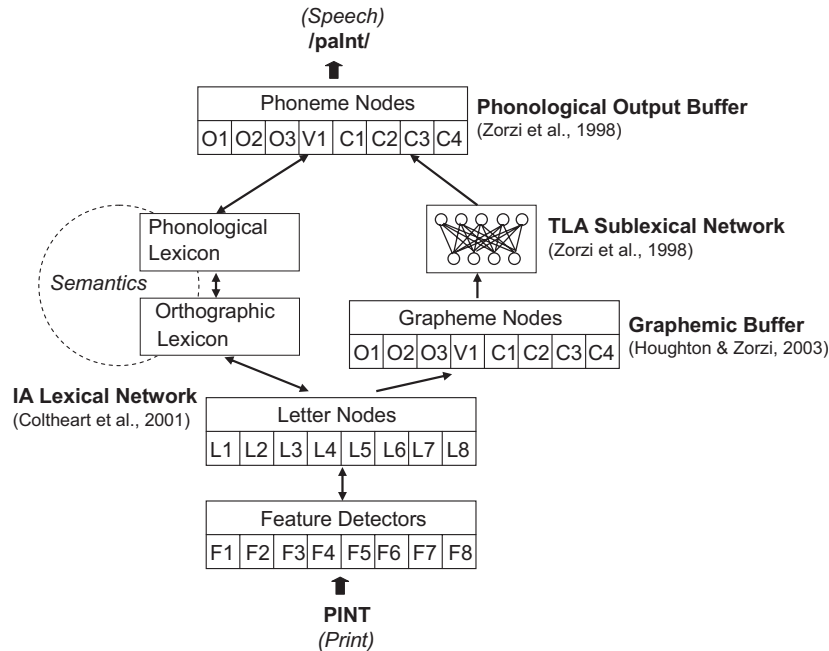


Figure 4. Schematic description of the new connectionist dual process model (CDP+). O = onset; V = vowel; C = coda; TLA = two-layer assembly; IA = interactive activation; L = letter; F = feature.

Activation in the lexical route begins when all the features of the letters in a letter string are turned on. The feature detectors then activate the letter level, where all letters in each of the eight positions are activated on the basis of the overlap they have with the features activated at the feature detector level (including the null letters). These letters then activate orthographic entries on the basis of the letter overlap they have with the word and inhibit other words that do not share the letters. Letters are mapped onto words according to a simple position-specific one-to-one correspondence: The first letter at the letter level activates/inhibits the first letter in a word, the second letter activates/inhibits the second letter, and so on. Entries in the orthographic lexicon then activate entries in the phonological lexicon. There is no inhibition between these levels. Finally, activated entries in the phonological lexicon activate/inhibit phoneme units in the phonological output buffer. One crucial feature of the interactive activation model is that feedback occurs by means of recurrent connections that exist among all levels. Thus, for instance, phoneme units can activate entries in the phonological lexicon, and entries in the orthographic lexicon can activate the letter level.

The lexical route of the new model is identical to that of DRC all the way up to and including the phonological lexicon, excluding the null characters that do not exist in the phonological lexicon of the new model. The equations governing its functioning are fully described in Coltheart et al. (2001, p. 215; note that their equations differ slightly from those of Rumelhart & McClelland, 1982). The phonological output buffer was also changed so that its representation was identical to the phonological decision system of CDP. Thus, instead of the phonemes being aligned as a contiguous string, the phonemes were aligned so that they respected the onset–vowel–coda distinction (see later discussion). In addition, unlike DRC, there was no null phoneme in the output to signify the

absence of a phoneme in a particular position. This was necessary so that phonological activation produced by the two-layer associative network (sublexical phonology) and by the lexical route (lexical phonology) could be integrated, because the sublexical route does not produce null phonemes. Finally, the frequencies of the words in the phonological lexicon were changed so that they were phonological rather than orthographic frequencies (unlike the current implementation of DRC).

Note that the lexical route in CDP was simply simulated as a frequency-weighted activation function of the lexical phonology, an approach similar to that used by Plaut et al. (1996) to model the contribution of the semantic route. In the new model, the dynamics at the output level reflect the online combination of activation from the fully implemented lexical route and the TLA network described later.

### Sublexical Route

*Input and output representation.* As justified earlier, we added an orthographic buffer to the sublexical route. The orthographic buffer was taken from the spelling model of Houghton and Zorzi (2003) and was implemented in the input coding scheme used with the two-layer associative network. Thus, single input nodes do not represent individual letters only, as in CDP, but also complex graphemes such as *ck*, *th*, and so forth. The set of complex graphemes specified by Houghton and Zorzi (see Appendix A) includes 10 onset graphemes, 41 vowel graphemes, and 19 coda graphemes. When letters combine to form one of these graphemes, the grapheme is activated instead of the letters (i.e., conjunctive coding). Note that the complex graphemes are basically the most frequent ones that occur in English (see Perry & Ziegler, 2004, for



a full analysis), although they are by no means the entire set that can be found.

The input representation is constructed by aligning graphemes to a graphosyllabic template (Caramazza & Miceli, 1990; Houghton & Zorzi, 2003) with onset, vowel, and coda constituents. There are three onset slots, one vowel slot, and four coda slots. Each grapheme is assigned to one input slot. If the first grapheme in a letter string is a consonant, it is assigned to the first onset slot, and the following consonant graphemes are assigned to the second and then to the third onset slots. Slots are left empty if there are no graphemes to be assigned. The vowel grapheme is assigned to the vowel slot. The grapheme following the vowel is assigned to the first coda slot, and subsequent graphemes (if any) fill the successive coda slots. Thus, for example, *black* would be coded as *b-l-\*-a-ck-\*-\*.\**, where each asterisk represents a slot that is not activated by any grapheme. Similarly, a nonword like *sloiched* would be coded as *s-l-\*-oi-ch-e-d-\*-.*

The phonological output of the network has a representational structure identical to that described in Zorzi et al. (1998b), except that instead of using three onset, one vowel, and three coda slots, it uses three onset, one vowel, and four coda slots. Thus, when training patterns are presented to the network, the output (phonology) is broken down in a way that respects an onset–vowel–coda distinction. The addition of a fourth coda slot was motivated by the existence of words with four coda phonemes in the database used to train the model. Thus, a word like *prompts* (*/prɒmpts/*) with four consonants in the coda can be handled by the model and would be coded as *p-r-\*-ɒ-m-p-t-s.*

*TLA network.* The sublexical network is a simple two-layer network, identical to the sublexical route of CDP apart from the number of input and output nodes. The input nodes encode the orthography of the word according to the grapheme buffer representation described earlier. Thus, graphemes are encoded over 8 slots in the input layer (3 onset slots + 1 vowel slot + 4 coda slots), where each slot consists of 96 grapheme nodes (26 single letters + 70 complex graphemes). The phonology of the word is encoded at the output layer of the network, which contains 43 phoneme nodes for each of the 8 available slots (3 onset slots + 1 vowel + 4 coda slots). This means that there are 768 input nodes and 344 output nodes (i.e.,  $8 \times 96$  and  $8 \times 43$ ). Replicating an identical coding scheme across slots means that the entire set of orthographic (or phonological) units is potentially available at any position. However, this choice was made only for the sake of simplicity, and it has no practical consequences. Indeed, nodes that are never activated (like codas in onset positions, vowels in coda positions, etc.) are completely irrelevant for the network: That is, nodes that never get any input in training never cause any output (see Equation 2 of Zorzi et al., 1998b, p. 1136), and because of the way the representation is constructed, these irrelevant nodes are never activated. Thus, performance of the network would be identical if we had used a coding scheme based on a slot-specific set of orthographic (or phonological) units divided into three main sections (onset, vowel, coda).

Note that an onset–vowel–coda distinction was also used by Plaut et al. (1996). Their orthographic and phonological representations, however, did not use slots to encode the order of grapheme or phoneme units. Their solution was to arrange the units within each set (onset, vowel, or coda) with an order that allows only orthotactically/phonotactically legal sequences to occur. In the

case of consonant clusters that have two possible orderings (e.g., */ts/* vs. */st/*), an additional node was activated to disambiguate between them. Another difference between CDP+ and the triangle model is that a multiletter grapheme is coded only by the activation of the grapheme unit with CDP+, whereas in Plaut et al., both the grapheme and the individual letters were activated.

The breakdown of input and output nodes into slots does not imply a built-in network structure (e.g., a specific pattern of connectivity); it only reflects the representational scheme used to activate the nodes. Accordingly, input and output nodes are fully connected, with no hidden units between them. Thus, any given input node has the potential to activate any given output. Activation of output nodes is calculated on the basis of the activation of input nodes in a manner identical to the one used by Zorzi et al. (1998b), and indeed the same parameters are used (see Appendix B). The equations used and how they work are described in full detail by Zorzi et al. (pp. 1136–1137) and in Appendix C.

It is worth noting that learning of a given input–output relationship, in any connectionist model, strictly depends on its existence in the training corpus and on how the inputs and outputs are coded. For example, our network cannot learn to generalize relationships from parts of words such as onset consonants to parts of words such as coda positions. Thus, a grapheme does not map to any phoneme if it is never activated in a specific position during learning. Though this is rather uncommon, one example is the consonant *j* in the coda position. Accordingly, a nonword like *jinje* cannot be correctly named by the model. However, the fact that the letter *j* never occurs after the letter *n* in English orthographic codas suggests that the word might be treated as disyllabic at an orthographic level (i.e., *jin-je*; see Taft, 1979, for a discussion of orthographic syllable boundaries). However, regardless of the issue of orthographic syllables, a model learning multisyllabic words would not have any difficulties with *jinje* because the *-nj* pattern does occur in words such as *injure* and *banjo* (see Plaut et al., 1996, for a similar argument).

*Training corpus.* The training corpus was extracted from the English CELEX word form database (Baayen, Piepenbrock, & van Rijn, 1993), and it basically consists of all monosyllables with an orthographic frequency equal to or bigger than one. The database was also cleaned. For example, acronyms (e.g., *mph*), abbreviations, and proper names were removed. Note that we did not remove words with relatively strange spellings (e.g., *isle*). A number of errors in the database were also removed. This left 7,383 unique orthographic patterns and 6,663 unique phonological patterns.<sup>2</sup>

*Network training.* In previous simulation work (Hutzler et al., 2004), we have shown that adapting the training regimen to account for explicit teaching methods is important for simulating reading development. Explicit teaching of small-unit correspondences is an important step in early reading and can be simulated by pretraining a connectionist model on a set of grapheme–phoneme correspondences prior to the introduction of a word corpus.

The two-layer associative network was initially pretrained for 50 epochs on a set of 115 grapheme–phoneme correspondences selected because they are similar to those found in children's

<sup>2</sup> We are grateful to Max Coltheart for providing the cleaned database.

phonics programs (see Hutzler et al., 2004, for further discussion). They consist of very common correspondences but are by no means all possible grapheme–phoneme relationships. The same correspondence (e.g., the correspondence  $L \rightarrow /l/$ ) may be exposed in more than one position in the network where it commonly occurs. The list of correspondences used appears in Appendix D. Note that the total number differs from that of Hutzler et al. (2004) because of the different coding scheme (their simulations were based on the CDP model). Learning parameters were identical to those in Zorzi et al. (1998b; see Appendix C).

After this initial training phase, the network was trained for 100 epochs on the word corpus (randomized once before training). Input (orthography) and output (phonology) of each word were presented fully activated. The number of learning epochs was greater than that reported in Zorzi et al. (1998b). We used a longer amount of training because Perry and Ziegler (2002) noticed that nonword generalization performance slightly increased with longer training times than those reported in Zorzi et al. (see also Hutzler et al., 2004). Performance of the network, in terms of the number of errors on nonwords, had well and truly reached an asymptote at the end of training. Training parameters used were the same as in Zorzi et al., except that we added some frequency sensitivity to the learning. This was done for each word by multiplying the learning rate by a normalized orthographic frequency value (see Plaut et al., 1996, for a discussion of frequency effects in training), which was simply the ratio of the log frequency of the word and the log frequency of the most common word (*the*; frequency counts were augmented by 1 to avoid the potential problem of words with a frequency of 0).

*Graphemic parsing.* Letters must be parsed and segmented into graphemes to obtain a level of representation compatible with the graphemic buffer (i.e., the input of the TLA network). Input to the sublexical route is provided by the letter level of the model (which is activated by the feature level; see the lexical network section). The letter level contains single letters that are organized from left to right according to their spatial position within the string (i.e., a word-centered coordinate system; Caramazza & Hillis, 1990; Mapelli, Umiltà, Nicoletti, Fanini, & Capezzani, 1996). We assume that graphemic parsing relies on focused spatial attention that is moved from left to right across the letters (this issue is taken up in the General Discussion section).

Graphemic parsing begins when a letter in the first slot position of the letter level becomes available. We assume that letters become available to the sublexical route when their activation is above a given threshold (.21 in our simulations). Graphemic parsing—and hence the identification and selection of letters—proceeds from left to right. A fixed number of cycles (15 in the current simulations) elapses between each of the parsings to account for the attention shift that allows for the next letter to be entered into the process and for the processing time required by grapheme identification. This way, each letter slot is examined in a serial fashion, and the single most activated letter in each position is assumed to be identified and selected. This process continues until the most active letter in each of the eight letter positions has been identified and selected or until the most activated letter in a given position is the null letter (see the lexical network section). In this case, no more letters are selected.<sup>3</sup>

Because the biggest complex graphemes have three letters (e.g., *tch*), graphemic parsing can best be conceived of as an attentional

window spanning three letters: the leftmost letter available and the following two letters (if any) at each step. That is, the window is large enough that the biggest existing graphemes can be identified and inserted into the appropriate grapheme buffer slots (e.g., the first consonant grapheme is placed in the first onset slot). This process of finding graphemes in the window continues as the window is moved across the letters and is repeated until there are no letters left. Each step in the graphemic parsing process results in additional graphemes that become available to the sublexical network, and a forward pass of the network activates the corresponding phonemes. Graphemes inserted in the buffer are fully activated and remain active throughout the naming process.

There are in fact two choices to be made with respect to the parsing of graphemes in the moving window. The first is whether only a single grapheme is ever assembled in the window or whether all potential graphemes are assembled. In the latter, which is the one we used, all letters in the window at any given time are segmented into graphemes.<sup>4</sup> The second choice we needed to make was when to start graphemic processing. We did this as soon as a single letter occurs in the window. Thus, it was assumed the right edge of the window begins at the first letter of the word rather than the third, and this causes very minor processing differences. We did not explicitly simulate grapheme identification; however, it would be relatively straightforward to implement a network that takes three letters as input and produces as output the biggest grapheme that is formed by the leftmost letters together with the following letters.

Another important aspect of the assembly process is that it assumes that the graphemic parsing process causes inputs into the

<sup>3</sup> Note that tying the onset of the grapheme parsing process to the first letter in the string does not have important consequences for the results, and it could be replaced, for instance, with a criterion based on the activation of any letter position or the summed activation of all active letters. Further empirical work is clearly needed to differentiate between these alternative assumptions. In principle, there is the possibility that the activation of the first letter may drop below the activation criterion needed to start the grapheme parsing process or that the most active letter at one slot in the letter level may change from one cycle to the next. In the simulations reported here, this never happens on any words or nonwords that are presented in isolation, but it does happen when words are primed. Whenever this occurs, it is assumed that the assembly processes begin again at the first letter. In terms of the operations of the attentional window, resetting the parsing process can be conceived of as if an exogenous attention shift had occurred because of the detection of a sudden change in the stimulus.

<sup>4</sup> This means that the model can incorrectly segment graphemes as the window moves, because the window is limited to three letters; thus, if, say, the first letter is used to create a grapheme, the two potential letters left may then be incorrectly assigned to the wrong grapheme because the number of letters left in the window is not enough to construct the correct grapheme. For example, take the word *match*; if the window initially contained the letters *[atc]*, it would assign the graphemes *-a*, *-t-*, and *-c* to the network. However, after moving the window across the letter string, the new letters would become *[tch]*, and thus the grapheme *-tch* would be chosen. In this case, it is assumed that the earlier graphemes incorrectly assigned to the network are revised and that this should be possible because they are still in the window. The alternative to this scan-all-and-revise type of process would be to stop identifying graphemes once a single one has been found and then to start processing again once the window had moved fully over the letters that were used in the grapheme that was identified.

sublexical network to be fully activated, even if they are not activated when taken from the letter level. This is not the case for DRC, where the activation of letters in the nonlexical route is identical to that of the letter level in the lexical route. The latter assumption is problematic, as pointed out by Reynolds and Besner (2004; see also Besner & Roberts, 2003). The DRC model predicts that if the activation buildup at the letter level is slowed because of the presentation of degraded stimuli, nonword length and stimulus quality should have underadditive rather than additive effects, which is contrary to what is found in human participants. Reynolds and Besner offered thresholded processing at the letter level as a potential solution to accommodate such a finding with DRC. The sublexical part of CDP+ essentially operates as if thresholded processing were used at the letter level, and it therefore produces an additive effect of stimulus quality and nonword length. Thresholded processing would certainly fix the DRC model, but the assumption that letters above a specific threshold are given full activation before entering the GPC system seems somewhat post hoc. Our proposal, instead, is that letters above threshold are submitted to a graphemic parsing process that is controlled by focused spatial attention. Each grapheme is fully active when inserted into the graphemic buffer simply because it represents the output of grapheme identification taking place within the attended portion of the letter string.

As a note of caution, we should point out that a literal translation of our implementation of the serial parsing process into operations of a spatial attention mechanism is potentially misleading. The empirical data are extremely sparse, and they currently do not permit any principled assumption regarding, for instance, the span of the attention window or how many attention shifts are needed to parse a string of letters of a given length. Nonetheless, our explicit proposal points to an area where much research is needed (also see the General Discussion section).

### *Combining the Networks*

Activation from the sublexical and lexical routes of the model is combined by summing the activation from both the lexical and the sublexical networks at the phonological output buffer. This activation is used as raw input before being transformed by means of the standard interactive activation equations. The parameters and equations used for calculating the activation of a given unit in the sublexical network were identical to those reported in Zorzi et al. (1998b). The network was updated at each cycle.

The way our model determines reading aloud latencies had to be different from Zorzi et al.'s (1998b) method simply because our output scheme combines activation from a fully interactive lexical network with that of a serialized associative network. In addition, it had to be different from that of DRC, because our model does not use a null character, which tells DRC when to finish processing. We therefore used a settling criterion, which is commonly used in recurrent networks to terminate processing (e.g., Ackley, Hinton, & Sejnowski, 1985). According to such a settling criterion, processing is stopped once nothing interesting is happening at the phonological output buffer (i.e., the network has settled). Thus, processing was stopped once the activation of all phonemes below the phoneme naming activation criterion did not change from one cycle to the next. Such a settling criterion was also used in one of the attractor network simulations of Plaut et al. (1996). Because no

activation is produced at the phonological output buffer at the beginning of each word presentation, the settling criterion was operative only after one phoneme had risen above the phoneme naming activation criterion.

More specifically, the activation produced by the network in the phonological output buffer is examined at each cycle. Processing is terminated (a) when at least one phoneme is above the phoneme naming activation criterion *and* (b) when, in all phoneme positions where no individual phoneme is above the phoneme naming activation criterion, the absolute difference in the activation between the current cycle and the previous cycle of all individual phonemes is below a small constant (.0023 in all simulations reported here). Note that an identical scheme could be used for DRC, which would allow the removal of the end null character. Such a change would, however, modify the activation dynamics of the network.

After processing stopped, the pronunciation produced was determined by taking the most highly activated phoneme in each position of the phonological output buffer. The activation had to be above the phoneme naming activation criterion, except for the vowel. For the vowel, the phoneme with the highest activation was chosen. Note that although it makes little difference, we allowed such a consonant–vowel difference because all words in English must contain a vowel. Zorzi et al. (1998b) and Plaut et al. (1996) used the same logic.

### *Parameter Set*

When trying to get CDP+ to produce values that closely resemble the human data, it is not possible to systematically search through the parameter space because of the large number of parameters in the model. Our strategy was to manipulate a small number of theoretically important parameters and to choose the current set on the basis of how well the model performed on a small number of data sets (see later discussion). These parameters control (a) the time course of phonological assembly, (b) the strength of lexical phonology, and (c) the strength of feedback from the phonological output buffer to the phonological lexicon.

The first step was to determine the appropriate balance between lexical and sublexical phonology, which in turn largely depends on the speed at which the serial process of grapheme parsing occurs. These parameters need to be chosen together, because slower grapheme parsing speeds reduce the amount of sublexical phonology in the model, and faster speeds increase it. Performance on irregular words provides a particularly important benchmark for parameter setting. In this respect, the study of Jared (2002) provided the most reliable set of experimental stimuli because it also controlled for body neighborhood characteristics (see the *Consistency Effects* section). Therefore, we manipulated the impact of sublexical phonology in the model until we obtained a marginal, nonsignificant latency cost for low-frequency irregular words with more friends than enemies, as well as a nonsignificant effect for low-frequency inconsistent words with more friends than enemies (Jared, 2002, Experiment 1).

The second step was to manipulate the strength of feedback from the phonological output buffer to the phonological lexicon. Weak or absent feedback results in the lack of a pseudohomophone effect when reading a set of pseudohomophonic nonwords (McCann & Besner, 1987). On the other hand, excessively strong feedback results in the activation of spurious lexical items, which

causes many lexicalization errors on nonwords (i.e., nonwords are read aloud using phonologically similar word names).

We did not try to optimize the parameters for any other experiments. Importantly, we did not optimize the parameter set to boost item-level correlations on large-scale databases. All parameters of the model appear in Appendix B. The results reported in the text refer to this unique set of parameters. Any parameter changes are noted in the text, and they were implemented only to simulate the effect of brain damage (acquired dyslexia) or strategy manipulations.

#### *Summary of Differences Between CDP+ and DRC*

1. The database was changed so that phonological word frequencies were used in the phonological lexicon instead of orthographic frequencies.
2. The network does not use a null character to signify when to stop running.
3. The network terminates processing on the basis of the settling criterion described earlier.
4. The phonological output buffer is not a linear string. Rather, it uses an onset–vowel–coda distinction.
5. The lexical route was changed so that the phonological component used the same output representation as the one used by CDP.
6. The grapheme–phoneme rule system was replaced by a new sublexical orthography–phonology network (similar to the TLA network of CDP).
7. Individual graphemes fed to the graphemic buffer (and hence to the sublexical assembly network) are always fully activated, unlike DRC's rules.
8. Graphemes begin to be fed to the graphemic buffer after the first letter rises above a given threshold.
9. Different parameters are used.

#### *Summary of Differences Between CDP+ and CDP*

1. The lexical route, up to and including the phonological lexicon, was replaced with that from DRC.
2. The network terminates processing on the basis of the settling criterion described earlier.
3. The input layer of the sublexical orthography–phonology network is a graphemic buffer with the same graphosyllabic structure as that described in the spelling model of Houghton and Zorzi (2003).
4. Input into the sublexical network was serialized as a result of the graphemic parsing process.
5. Different parameters are used for the integration of lexical and sublexical activation.

## Results

Two methods were used to evaluate the goodness of fit between model and data: (a) the factorial method and (b) the regression method. The former consists of analyzing RTs and/or errors of the model through analyses of variance (ANOVAs) to evaluate whether effects found to be statistically significant in the human data are also significant in the model data (conversely, effects that are not significant in the human data should not be significant in the model data). The second method is related to the issue of predicting item-level variance (Spieler & Balota, 1997; see also Besner, 1999). It consists of computing the proportion of variance ( $R^2$ ) in human RTs that is accounted for by the model at the level of single items. This method is a particularly tough test when applied to large word corpora, such as Spieler and Balota's (1997) database, which contains almost 3,000 words. The regression method is also very stringent when applied to the items of one specific small-scale experiment.

All results from Zorzi et al.'s (1998b) original CDP model were calculated with a new network trained on the CELEX (Baayen et al., 1993) database for 20 cycles. All results for the triangle network were calculated with a newly trained network set up in a manner identical to that of the log10 frequency trained feedforward network in Plaut et al. (1996). As in the original model, Kuçera and Francis's (1967) database was used to train the model because we did not want to change the input and output representations of the network, which would have been necessary if the CELEX database had been used. Error scores were calculated with a cross-entropy measure. All data from the models were examined with a 3 standard deviation ( $SD$ ) cutoff criterion. The attempt to account for a large amount of empirical data derived from many different studies calls for a stringent (i.e., conservative) strategy in statistical testing. Thus, when we compare the model with effects that are significant in the human data, unless otherwise stated, we do so only for the effects that were significant in the item data according to a standard between-groups comparison.

The nested modeling approach dictates that a new model should be tested on the data sets that motivated the development of its predecessors before being tested on new data. We therefore simulated all the naming studies reported by Coltheart et al. (2001), excluding those that have been superseded by more controlled experiments (e.g., pseudohomophone effects; Reynolds & Besner, 2005). However, to avoid distracting the reader with a long list of effects that do not necessarily adjudicate among different models (e.g., effects of frequency and lexicality), we postpone the presentation of these results to the *Other Phenomena in Word Naming* section. To anticipate our main findings, CDP+ was able to successfully simulate all benchmark effects reported by Coltheart et al. (2001) except for the whammy effect whose empirical robustness is still a matter of debate (see later).

The Results section is organized in the following way: First, before embarking on simulations of particular reading phenomena with any computational model of reading, we show the global performance of CDP+ on word and nonword naming. Second, we focus on the effects that were able to adjudicate among different models: (a) consistency effect for words and nonwords, (b) serial effects, and (c) performance on large-scale databases. In the spirit of strong inference testing, for all effects, we present a comparison of CDP+ performance with that of the other three models. For the



sake of clarity in the data presentation, we focus on specific subsets of data in the main text. However, all results of our simulations (both with CDP+ and with the other models) are fully reported in the Appendixes. Finally, in the *Credit Assignment and Componential Analyses* section, we assign credit for the components that are responsible for the superior performance of the CDP+ model.

### *Overall Performance on Words and Nonwords*

In terms of word naming, when confronted with the 7,383 words of its lexicon, CDP+ gets 98.67% correct. Heterophonic homographs were considered correct if they yielded any of the potential pronunciations. These simulations were obtained with the standard parameter set (see Appendix B). Of the errors, 70% were either regularization errors or alternative readings of the word that used commonly found grapheme–phoneme relationships. Accuracy for reading aloud words can be easily increased by changing the parameters of the lexical route. In fact, when we increased the amount of letter-to-orthography inhibition (to stop more than one word from ever being activated at a time) while turning off the sublexical route, the model made only four errors (0.05% errors).

To examine the overall nonword reading performance of CDP+, we tested the model on the only large-scale nonword reading database that is currently available (Seidenberg et al., 1994). In this database, 24 participants named 592 nonwords. Using the lenient error scoring criterion proposed by Seidenberg et al. (1994), according to which a nonword response was correct if the phonology given by the model corresponded to any grapheme–phoneme or body–rime relationship that exists in real words, the model made 37 errors (6.25%). The model's error rate is therefore very similar to the human error rate reported by Seidenberg et al. (7.3%). Of these errors, 16 (43%) displayed the pattern where a phoneme that should have been activated was indeed activated but not enough to get above the phoneme naming activation criterion. In these cases, the underactivated phoneme was not taken into account, and hence the final pronunciation looked as if it was missing that phoneme (e.g., /wʌl/ for *wulch*). Given that nonwords in the model generally produce far less activation than words (because they do not have lexical entries), we reduced the phoneme naming activation criterion from .67 to .50 to simulate a lower threshold that people may use when reading aloud lists of nonwords only. When this was done, the model produced only 17 errors (2.87%). In the *Consistency Effects in Nonword Reading* section, we go beyond this rather superficial assessment of the model's nonword reading ability by investigating whether the model's responses actually look like those that people give when asked to read certain types of nonwords.

### *Consistency Effects*

The issue of whether consistency or regularity (or both) best characterize the difficulty associated with naming a word has been controversial since the seminal work of Glushko (1979). More recently, the debate has been polarized by the fact that only regularity is a critical variable in DRC, whereas consistency is more relevant in the triangle model. In this section, we examine the effect of consistency in three different ways: First, we examine the effect of consistency on word naming latencies (Jared, 2002), and

then we investigate the consistency effect on nonword naming (Andrews & Scarratt, 1998). Finally, we look at the effect of consistency on the impaired performance of an individual with surface dyslexia (Patterson & Behrmann, 1997).

### *Consistency Effects in Word Reading*

In a recent influential study, Jared (2002) argued quite convincingly that there were a number of potential confounds in almost all of the data sets examining regularity and consistency used in previous studies of word naming. Jared's stimuli therefore represent probably the best controlled stimulus set for examining consistency and regularity effects. The main results of Jared's study can be summarized as follows. First, there was clear evidence for an effect of word-body consistency, regardless of whether the word was regular or irregular. There was little evidence of a regularity effect over and above the effect of consistency. Second, the effect of consistency was modulated by the neighborhood characteristics of the stimuli, that is, the ratio between friends and enemies. Only words with more enemies than friends showed a latency cost. Third, there was very little evidence for a Frequency  $\times$  Regularity/Consistency interaction when neighborhood characteristics were controlled.

Jared (2002) showed that different aspects of these data challenge DRC and the triangle model. DRC was challenged in two ways. First, it did not show any hint of a consistency effect for regular words. Second, the latency cost for irregular words (i.e., regularity effect) was not modulated by the friend–enemy ratio. The triangle model correctly predicted the existence of a consistency effect for both regular and irregular words, but it also produced a strong interaction with frequency, which was not present in the human data. We simulated all four experiments reported by Jared with CDP+ and with the three other models. In the text, we focus on some of the results that clearly adjudicate between the different models. The full set of simulations (all experiments and all models) can be found in Appendix E.

Jared's (2002) Experiment 1 investigated whether there are effects of GPC regularity independent of effects of consistency for low-frequency words. Naming performance on exception (e.g., *breast*) and regular inconsistent words (e.g., *brood*) was compared with performance on matched regular consistent words. A further manipulation was performed on the first two groups: Half of the words in each group had more friends than enemies, whereas the other half had more enemies than friends. Results showed longer naming latencies for both the exception and regular inconsistent words, but only for the groups with more enemies than friends (see Figure 5A).

CDP+ produced a pattern almost identical to that of the human data (see Figure 5B). In particular, there were main effects of consistency/regularity combined,  $F(1, 147) = 17.73$ ,  $MSE = 1,281$ ,  $p < .001$ . There was no Consistency/Regularity  $\times$  Group interaction (i.e., exception vs. regular inconsistent words),  $F(1, 147) = 2.69$ ,  $MSE = 194$ ,  $p = .10$ . However, there was a Regularity/Consistency  $\times$  Friend–Enemy Ratio interaction,  $F(1, 147) = 5.09$ ,  $MSE = 367$ ,  $p < .05$ . Pairwise comparisons showed that both groups with more enemies than friends were significantly slower than their controls: 10.41 cycles for low-frequency exception words,  $t(35) = 3.08$ ,  $SE = 3.38$ ,  $p < .005$ , and 7.30 cycles for low-frequency inconsistent words,  $t(38) = 2.89$ ,  $SE = 2.52$ ,  $p < .01$ .

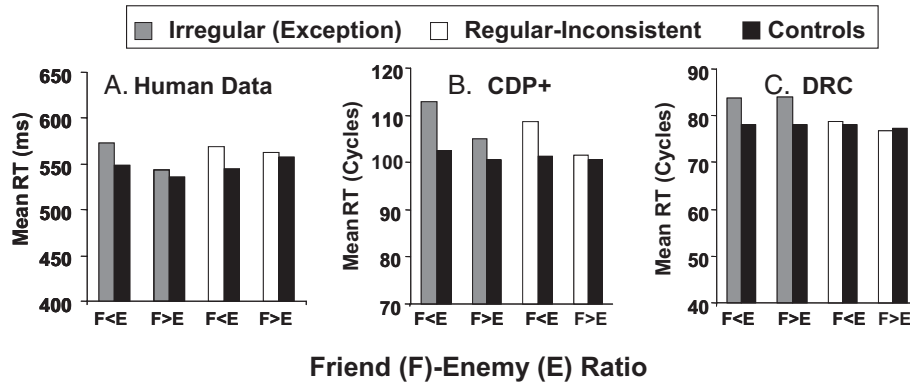


Figure 5. Human data (in milliseconds) and CDP+ and DRC simulations of Jared’s (2002) Experiment 1. RT = reaction time; CDP+ = new connectionist dual process model; DRC = dual-route cascaded model.

.01, whereas both groups with more friends than enemies were not slower than their controls: 4.58 cycles for low-frequency exception words,  $t(37) = 1.74$ ,  $SE = 2.63$ ,  $p = .09$ , and 0.82 cycles for low-frequency inconsistent words ( $t < 1$ ). As discussed earlier, DRC showed a completely different pattern of results (see Figure 5C). The striking difference between CDP+ and DRC can be also seen in the amount of item-level variance accounted for by the two models: CDP+ accounted for 24%, whereas DRC accounted for just 1% of the variance. The triangle model accounted for 7% (for details see Appendix E).

Before Jared (1997), one of the long-held beliefs about regularity and consistency effects in English was that they only occurred with low-frequency words. Jared (1997) argued that this may not be due to word frequency per se, but rather may be due to the fact that most high-frequency words also have fewer enemies of a higher frequency. In Experiment 2, Jared (2002) therefore investigated the Regularity  $\times$  Frequency interaction in words matched for neighborhood characteristics. The human data showed that there was a main effect of regularity but no reliable Regularity  $\times$  Frequency interaction. The effect of regularity was 13 ms for high-frequency words and 19 ms for low-frequency words. Additionally, the size of the regularity effect was modulated by the friend–enemy ratio, with a reliable effect only for words with more enemies than friends (see Figure 6A).

When confronted with the items of Jared (2002, Experiment 2), CDP+ displayed a pattern that was similar to the human data (see Figure 6B). Statistical analyses confirmed significant main effects of regularity,  $F(1, 144) = 20.43$ ,  $MSE = 1,493$ ,  $p < .001$ , and frequency,  $F(1, 144) = 145.44$ ,  $MSE = 10,628$ ,  $p < .001$ , but no interaction between the two ( $F < 1$ ). The Neighborhood (friend–enemy ratio)  $\times$  Regularity interaction was significant,  $F(1, 144) = 4.59$ ,  $MSE = 336$ ,  $p < .05$ . As in the human data, pairwise comparisons showed that the regularity effect was significant in both low- and high-frequency words for the groups with more enemies than friends,  $t(35) = 3.08$ ,  $SE = 3.38$ ,  $p < .005$ , and  $t(37) = 2.75$ ,  $SE = 2.94$ ,  $p < .01$ , for low- and high-frequency words, respectively. This was not the case for words with more friends than enemies, however, with neither the low- nor high-frequency groups being significant,  $t(37) = 1.74$ ,  $SE = 2.63$ ,  $p = .09$ , and  $t(35) = 1.04$ ,  $SE = 1.94$ ,  $p = .31$ .

As discussed earlier, the triangle model does not fully capture the human data because it shows a strong Regularity  $\times$  Frequency interaction (see Figure 6C). The advantage of CDP+ over the triangle model is further testified to by the striking difference in the amount of item-level variance accounted for by the two models: CDP+ accounted for 40%, whereas the triangle model accounted for less than 1%. Although DRC correctly captured the

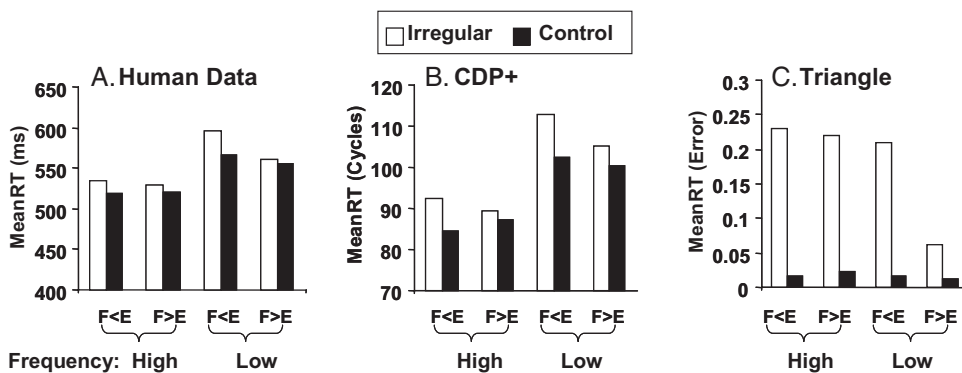


Figure 6. Human data (in milliseconds) and CDP+ and triangle model simulations of Jared’s (2002) Experiment 2. RT = reaction time; CDP+ = new connectionist dual process model; F = friends; E = enemies.

absence of a Frequency  $\times$  Regularity interaction, it accounted for only 9% of the variance (for additional details, see Appendix E).

The two final experiments of Jared (2002) investigated the Frequency  $\times$  Regularity  $\times$  Consistency interaction on words with more enemies than friends (Experiments 3 and 4 differed only with respect to a blocked vs. mixed list composition) and confirmed the overall pattern of results. As for the previous experiments, simulations with CDP+ provided a very good match to the human data. The model accounted for a striking proportion of item-level variance (52% for Experiment 3 and 46% for Experiment 4), a proportion at least three times larger than the variance accounted for by DRC and the triangle model (for full simulation details, see Appendix E).

### Consistency Effects in Nonword Reading

Consistency manipulations are possible not only on real words but also on nonwords. The main issue, debated since the study of Glushko (1979), is whether nonword reading relies on the use of grapheme–phoneme correspondences or whether larger spelling units, like word bodies (orthographic rimes), play a role. Inconsistent nonwords, that is, letter strings containing a word body with more than one common pronunciation (e.g., *ove* may be pronounced as in *stove*, *wove*, *grove*, etc. or as in *love*, *glove*, *shove*), often produce much more varied distributions of answers (i.e., multiple different pronunciations) compared with consistent nonwords (see Seidenberg et al., 1994).

Andrews and Scarratt (1998) performed two experiments examining the extent to which people use grapheme–phoneme and larger sized spelling–sound relationships when reading aloud nonwords. In Experiment 1, they examined the consistency of the initial consonant–vowel cluster (CV consistency) and body-rime consistency. All of the nonwords in the groups shared a body with at least one other word that could be pronounced with a regular pronunciation (i.e., a word whose phonology could be derived by means of the use of the most common grapheme–phoneme rules). Thus, for instance, the nonword *bive* has a body that can be given an irregular pronunciation (e.g., *give*), but at least one word that it shares an orthographic body with does have a regular pronunciation (*hive*). Thus, a regular analogy can be made by using the body-rhyme correspondence from the word *hive*. They also examined how people would read a group of stimuli that had no regular analogies, that is, nonwords that shared a body only with words that have an irregular pronunciation (note that their definition of *regularity* was not that of Coltheart et al., 1993, but was very similar). Thus, for example, the nonword *valk* has the body *-alk*. However, all words that have that body (e.g., *walk*, *talk*, *chalk*) do not use a regular body pronunciation, which would be /ælk/.

The results from Andrews and Scarratt (1998) showed that in the regular-analogy group (e.g., *bive*), people tended to read such nonwords according to the most common grapheme–phoneme relationships. That is, they chose the regular pronunciation that rhymes with *hive* and not *give*. They also did not appear to use CV relationships often, as the difference between nonwords with relatively inconsistent CV relationships and nonwords with relatively consistent CV relationships was small. In contrast, in the no-regular-analogy group, where nonwords used a body that was not shared with any other words that used a regular grapheme–phoneme based pronunciation, participants showed a marked pref-

erence for using a body pronunciation that would not be generated by a set of grapheme–phoneme rules.

An ANOVA performed on regular response probabilities of CDP+ showed that nonwords with inconsistent bodies were pronounced with less regular pronunciations than nonwords with consistent bodies,  $F(1, 183) = 10.34$ ,  $MSE = 0.41$ ,  $p < .005$ . However, the effect of onset consistency was weak ( $F < 1$ ). This was similar to the human data. The crucial issue for computational models, however, is the performance in the no-regular-analogy group. That is, if GPC rules are used to read nonwords, as postulated by DRC, little or no deviation from the regular pronunciation is expected (which is not what Andrews & Scarratt, 1998, found). In terms of the no-regular-analogy group, the proportion of regular answers given by CDP+ was very similar to that of the human data. The results appear in Figure 7A along with a comparison of the other models. As can be seen, CDP+ and CDP fit the observed pattern better than DRC or the triangle model, both of which produced too many regular responses (for further details, see Appendix F).

In Experiment 2, Andrews and Scarratt (1998) examined four groups of nonwords. In two of the groups (i.e., the regular-analogy groups), body-rime consistency was manipulated in a dichotomous manner. Words in these groups shared their bodies with at least one other word that had a regular pronunciation. In the other two groups, the word bodies had no regular analogy, but the two groups differed in terms of the size of the body neighborhood (i.e., many vs. few body neighbors). The results showed that nonwords in the regular-analogy groups were generally given regular pronunciations. Nonwords in the no-regular-analogy group were generally *not* given regular pronunciations, and that was particularly true of nonwords that had many body neighbors.

The simulations of this experiment again showed a very good match between CDP+ and the human data (see Figure 7B). In contrast, DRC and the triangle model produced too many regular responses in the no-regular-analogy group with many body neighbors (for further details, see Appendix F). It appears that a different version of the triangle model produces a slightly better fit to the

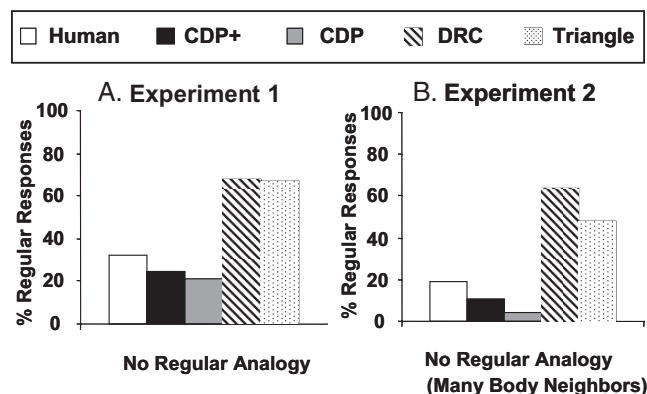


Figure 7. Human data (response probabilities for regular pronunciations) and simulations of all models for no-regular-analogy words (Experiment 1) and no-regular-analogy words with many body neighbors (Experiment 2) of Andrews and Scarratt (1998). CDP+ = new connectionist dual process model; CDP = connectionist dual process model; DRC = dual-route cascaded model.

data (Zevin & Seidenberg, 2006). This is not surprising as connectionist models should be more sensitive to consistency than rule-based GPC models. However, the triangle model used by Zevin and Seidenberg (2006) still overestimates the proportion of regular pronunciations in the no-regular-analogy group, whereas CDP+ does not.

### Consistency Effects in Surface Dyslexia

MP is probably the most well-documented individual surface dyslexic (Bub, Cancelliere, & Kertesz, 1985; see also Behrmann & Bub, 1992). When reading aloud irregular (or inconsistent) words, MP has the tendency to give regularized pronunciations (i.e., words are read as if they were nonwords). There is a very important pattern in MP's errors, which was noted by Patterson and Behrmann (1997). They observed that the number of regularization errors MP made was a function of how inconsistent the word was. MP made more regularization errors on words that were very inconsistent compared with words that were less inconsistent. Indeed, Patterson and Behrmann investigated MP's performance in much the same way as did Jared (1997, 2002). That is, two groups of irregular words were used, one in which the words had more friends than enemies and another in which the words had more enemies than friends. Patterson and Behrmann also used a group in which an inconsistent vowel was made relatively consistent by its preceding onset consonants (typically, words starting with *wa*; see also Treiman, Kessler, & Bick, 2003).

The existence of consistency effects in the naming data of MP is problematic for DRC, which uses a set of rules with relatively little context sensitivity. At least in its current form, the model is not able to capture that effect. In contrast, when the model of Plaut et al. (1996) was "lesioned" by removing the semantic component of the model, the results showed that the model was generally able to simulate this pattern (see Patterson & Behrmann, 1997, for further discussion). This simulation, however, relied on the assumption that surface dyslexia is caused by semantic impairment. In a survey of a number of different cases of surface dyslexia, Coltheart et al. (2001) showed that this explanation may have some problems. They noted that, in spite of the relatively strong association between semantic dementia and surface dyslexia, it was possible to find cases of severe surface dyslexia in patients who had a completely functioning semantic system, and most notably, it was possible to find patients with severe dementia who did not show any signs of surface dyslexia (see the General Discussion section).

Because we agree with Coltheart et al. (2001) that attributing surface dyslexia to a semantic deficit is problematic from a theoretical point of view, our goal was to simulate the surface dyslexic pattern of MP through a simple manipulation of the lexical route (also see Zorzi et al., 1998b). To do this, we simply reduced the contribution of lexical phonology to the final pronunciation by reducing the amount of lexical activation into the phonological output buffer. That was done by setting the inhibition parameter from the phonological lexicon to the phonological output buffer to zero and by setting the excitation parameter from the phonological lexicon to the phonological output buffer to 0.055. This meant that only a comparatively small amount of lexical activation reached the phonological output buffer compared with usual. A similar type of lesion was used by Houghton and Zorzi (2003) to simulate

surface dysgraphia: The underlying assumption was that the asymptotic strength of the lexical output has been reduced following neural damage (e.g., because of lowered excitability of the neural structures encoding lexical knowledge). The size of the frequency parameter was also increased to 0.8 from 0.4 to simulate potential difficulties in lexical access (this second strategy was also pursued by Coltheart et al., 2001). That is, instead of the frequency of items in the orthographic and phonological lexicons being coded between 0 and  $-0.4$ , they were coded between 0 and  $-0.8$  on the basis of the normalization procedure in Coltheart et al. (2001, p. 216). The effect of this was to increase the difference in the amount of input that is needed to activate low- and high-frequency words in the lexicon. Thus, low-frequency words become comparatively harder to activate than high-frequency ones.

We should note here that whether the activation that combines with the phonological assembly route is actually lexical or semantic is not especially relevant for this simulation. Rather, what is important is that the contribution of lexical/semantic activation is comparatively very small. In the same vein, the "semantic" lesion in Plaut et al.'s (1996) simulation of surface dyslexia is implemented by removing the contribution of an external input to the phoneme units of the orthography-to-phonology network. Therefore, the models differ with respect to the functional interpretation of the lesion within the broader lexical system, but they simulate the phenomenon in a very similar way. The issue of the role of semantics in reading aloud is taken up in the General Discussion section.

With the parameter changes detailed earlier, CDP+ displayed a pattern very similar to that of MP. In particular, it correctly predicted that the more consistent irregular words have a greater chance of being named correctly. Of the 32 errors the model made for irregular words, all but 2 were regularization errors, which was also very similar to the performance of MP, who also produced only 2 errors on the irregular words that were not regularizations. Of the 3 errors the model made on regular words, all were pronunciations that could be found in irregular words that shared the same orthographic body. MP also showed that pattern. The performance of CDP+ on irregular words appears in Figure 8.

We also tried to simulate the results of MP with DRC. Following Coltheart et al. (2001), we increased the response threshold to simulate the lack of speed pressure in the task, and we modified the

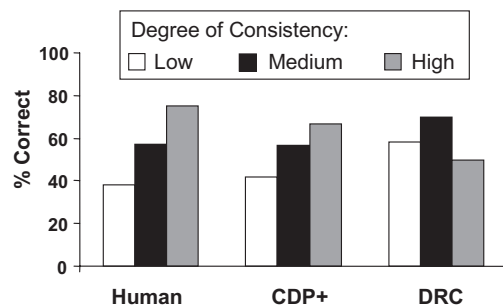


Figure 8. Performance of surface dyslexic patient MP on irregular words together with CDP+ and DRC simulations. The graph shows the percentage of correct responses for three lists of irregular words that differ in their degree of consistency. CDP+ = new connectionist dual process model; DRC = dual-route cascaded model.



frequency scaling parameter in the orthographic lexicon (to simulate difficulty in retrieving the orthographic form of the words). However, no matter what we did, we were not able to cause an effect of graded consistency. The results of DRC are visible in Figure 8 (note that the orthographic frequency scaling parameter was set to 0.20, rather than to the value of 0.25 used by Coltheart et al., 2001, because it gave slightly better results). It is thus the only model that does not simulate MP's results, as both CDP and the triangle model (see Zorzi, 1999, and Patterson & Behrmann, 1997, respectively) can simulate this pattern (see Appendix G for further details).

Another factor that interacts with consistency in surface dyslexia is word frequency. In general, low-frequency irregular words are more likely to be read aloud incorrectly than high-frequency irregular words. This pattern was documented in an earlier study with MP by Behrmann and Bub (1992) and in KT, one of the most extreme surface dyslexic patients reported (McCarthy & Warrington, 1986; see also Patterson, 1990). KT is of particular interest, because he was able to read only 47% of high-frequency and 26% of low-frequency exception words correctly, despite very good performance on regular words (100% and 89% correct for high- and low-frequency regular words, respectively). It is important to show that the model can simulate this effect, because both DRC and the triangle model can (also see Zorzi et al., 1998b, for simulations with CDP).

The Frequency  $\times$  Regularity interaction of both patients (MP and KT) was simulated with CDP+ by making the same parameter changes as in the previous MP simulation, except that we increased the frequency scaling parameter slightly from 0.8 to 1.0. For KT, we examined the model's performance on Taraban and McClelland's (1987) high- and low-frequency regular and exception words. According to Patterson (1990), Taraban and McClelland's exception words are very similar to the words actually presented to KT. For MP, we used the monosyllabic words that were actually presented.

The results showed that CDP+'s error pattern was very similar to that of KT: The model read only 50% of the high-frequency and 29% of the low-frequency exception words correctly, but read 100% of the high-frequency and 96% of the low-frequency regular words correctly. Similarly, CDP+ also produced results very similar to those of MP. In particular, for words from the six categorical frequency bands reported (frequencies: 1–9, 10–19, 20–49, 50–99, 100–199, >199), for which MP made 7, 6, 9, 5, 2, and 19 correct responses, respectively, CDP+ made 9, 6, 9, 6, 4, and 16 correct responses. The fit between our model and MP is actually closer than the fit between DRC and MP reported in Coltheart et al. (2001). Together, these simulations show that CDP+ can simulate some of the most extreme cases of surface dyslexia reported.

### Summary: Consistency Effects

Simulating the detailed pattern of consistency effects in word and nonword naming has turned out to be one of the most damaging but also one of the most challenging areas for previous models of reading aloud. Here, we have successfully simulated seven consistency experiments. CDP+ not only captured the qualitative pattern of the seven experiments, but it also predicted up to 52% of the item-specific variance. In contrast, DRC failed on

almost every data set, both qualitatively and quantitatively. Surprisingly, the triangle model was not much better than DRC quantitatively (i.e., it never accounted for more than 7% of the variance in any of the word consistency experiments), and there were also some qualitative deviations from the human data.

### Serial Effects

Here we focus on two effects that have been widely cited in the literature as evidence for serial processing in reading aloud: (a) length effects in nonword reading (Weekes, 1997; Ziegler et al., 2001) and (b) the position-of-irregularity effect (Rastle & Coltheart, 1999).

### Length Effects

Weekes (1997) used three-, four-, five-, and six-letter words and nonwords to study length effects in reading aloud. The words were of either low or high word frequency. He found a main effect of length that was qualified by a significant Length  $\times$  Lexicality interaction. This interaction reflected the fact that the length effect was stronger for nonwords than for words. Weekes also suggested that low-frequency words produced bigger length effects than high-frequency words. A reanalysis of his item data on words showed, however, that the Frequency  $\times$  Length interaction did not reach significance,  $F(3, 190) = 1.71$ ,  $MSE = 1,451$ ,  $p = .16$ . He further argued that the length manipulation is naturally confounded with orthographic neighborhood (the longer the word, the smaller the number of neighbors). However, even with orthographic neighborhood used as a covariate (as was done in Coltheart et al., 2001), there was no significant Length  $\times$  Frequency interaction,  $F(3, 189) = 1.97$ ,  $MSE = 1,635$ ,  $p = .12$ . Finally, to make sure that we did not miss a potential Frequency  $\times$  Length interaction because of a lack of power, we submitted Weekes's items to the English Lexicon Project (Balota et al., 2002), which contains naming data for thousands of words collected from over 400 participants. This analysis showed that even in this database, Weekes's word set did not produce a significant Length  $\times$  Frequency interaction ( $F < 1$ ). All together then, the correct pattern that needs to be predicted by current models is a main effect of length and a Length  $\times$  Lexicality interaction but no Length  $\times$  Frequency interaction for real words.

The simulation results of Weekes (1997) are presented in Figure 9. CDP+ showed a main effect of length,  $F(3, 274) = 37.06$ ,  $MSE = 3,806$ ,  $p < .001$ , a main effect of lexicality,  $F(1, 274) = 1,191.82$ ,  $MSE = 122,397$ ,  $p < .001$ , and a Length  $\times$  Lexicality interaction,  $F(3, 274) = 18.32$ ,  $MSE = 1,889$ ,  $p < .001$ . As in the human data, there was no significant interaction between the effects of length and frequency when words were examined separately: for length,  $F(3, 186) = 83.84$ ,  $MSE = 2,924$ ,  $p < .001$ ; for frequency,  $F(1, 186) = 289.19$ ,  $MSE = 10,081$ ,  $p < .001$ ; for Length  $\times$  Frequency,  $F < 1$ . Thus, CDP+ predicts the correct overall pattern.

The simulation results of all other models are shown in Appendix H. As previously noted by Coltheart et al. (2001), Weekes's (1997) data were not fully captured by CDP or the triangle model. In brief, CDP produced a length effect but no Length  $\times$  Lexicality interaction, whereas the performance of the triangle model was markedly different from the human data. In terms of item-level

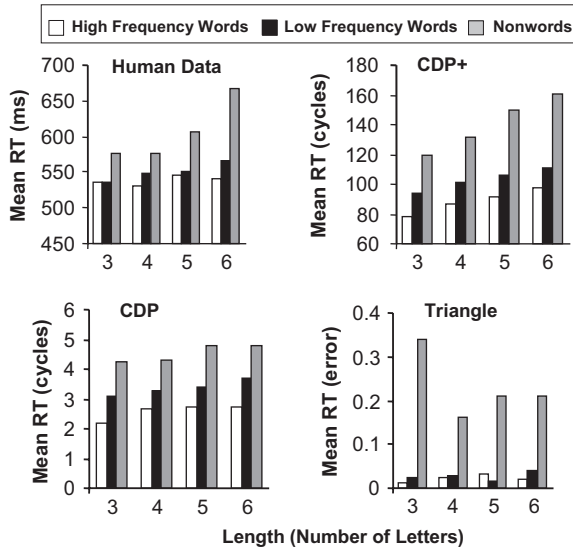


Figure 9. Human data (in milliseconds) and simulations of Weekes's (1997) experiment, which manipulated length, frequency, and lexicity. RT = reaction time; CDP+ = new connectionist dual process model; CDP = connectionist dual process model.

variance, CDP+ accounted for at least twice as much of the variance as the other models on the word data. In contrast, DRC still outperformed all other models on the nonword data; nonetheless CDP+ was greatly superior to the other connectionist models as it accounted for 31% of the variance, whereas CDP and the triangle model accounted for less than 3%.

It is worth noting that there has been a debate about whether length has a unique contribution to naming or whether length effects are simply the result of other confounding variables. Along this line, Seidenberg and Plaut (1998) suggested that length effects might be due to articulatory factors. However, because the length effect disappeared in delayed naming (Weekes, 1997, Experiment 2), it is unlikely that the effect is due to articulation. On the other hand, Weekes (1997) himself suggested that length does not produce a *unique* effect on word naming because length did not account for any unique variance in a regression analysis after partialing out orthographic neighborhood and phoneme length. However, the strong conclusion of Weekes that there are no unique effects of length was contested in a comprehensive study of Balota et al. (2004), who analyzed the unique effects of length in a large-scale database of 2,428 words. Their results clearly show that length has a unique contribution to word naming (for similar results, see also Baayen, Feldman, & Schreuder, 2006). For the interested reader, Balota et al. gave several plausible arguments for why their results differed from those of Weekes.

### Body Neighborhood and Length

As mentioned earlier, the stimulus set of Weekes (1997) presents some confounds that are not related to length. These included orthographic neighborhood size and *body neighborhood* (e.g.,

Brown, 1987; Forster & Taft, 1994; Jared et al., 1990; Ziegler et al., 2001). The effect of body neighborhood size simply reflects the frequency at which orthographic bodies occur, regardless of the way they map onto phonology (Ziegler & Perry, 1998). For instance, take the two nonwords *veap* and *veep*. Both of these nonwords have completely consistent orthographic bodies. That is, the bodies *-eap* and *-eep* are only ever pronounced one way in real words. However, the frequencies at which the bodies occur is markedly different: *-eap* occurs in 5 words whereas *-eep* occurs in 13. This is known as orthographic body neighborhood. When Ziegler et al. (2001) examined this effect in a Lexicality (words vs. nonwords)  $\times$  Orthographic Length (3–6 letters)  $\times$  Body Neighborhood (high vs. low) design, they found that there was an effect of body neighborhood that appeared to occur with both words and nonwords, and there was no significant Body Neighborhood  $\times$  Length interaction. However, there was the standard Length  $\times$  Lexicality interaction. The study of Ziegler et al. therefore provided a highly controlled test of the length effect. It also provided an additional important constraint for strong inference testing because previous models make opposite predictions as to whether body neighborhood should play a role in reading aloud (e.g., the DRC predicts no influence of body neighborhood, whereas the CDP predicts an influence).

CDP+ displayed essentially the same pattern as the human data: It produced significant main effects of length,  $F(1, 140) = 53.59$ ,  $MSE = 6,615$ ,  $p < .001$ , lexicality,  $F(1, 140) = 877.64$ ,  $MSE = 108,327$ ,  $p < .001$ , and body neighborhood,  $F(1, 140) = 4.98$ ,  $MSE = 615$ ,  $p < .05$ . The effects of length and lexicality interacted,  $F(3, 140) = 19.85$ ,  $MSE = 2,500$ ,  $p < .001$ , whereas body neighborhood did not interact with either length or lexicality (both  $F_s < 1$ ). No other model was able to produce this result: DRC failed to produce a body neighborhood effect ( $F < 1$ ), CDP failed to produce a Length  $\times$  Lexicality interaction,  $F(1, 137) = 1.97$ ,  $MSE = 1.27$ ,  $p > .1$ , and the triangle model failed to produce the correct shape of the length effect. The mean results appear in Figure 10. Finally, CDP+ accounted for a much larger proportion of item-level variance in Ziegler et al.'s (2001) set of nonwords than the other models, including DRC (for further details, see Appendix I).

The results suggest that the only model able to capture both the Length  $\times$  Lexicality interaction and the additive body neighborhood effect was CDP+. Thus, the model can show both the effects of simple frequency of occurrence at a subsyllabic level as well as consistency.

### Position of Irregularity

Rastle and Coltheart (1999) found that the cost of irregularity was modulated by the position of the irregular grapheme in the word. That is, the regularity effect was stronger for first position irregular words compared with second position irregular words, and it was stronger for second position irregular words than for third position irregular words. Although the original CDP model can simulate this effect even with completely parallel input (Zorzi, 2000), it remains a challenging effect because it cannot be simulated by at least one of the models (i.e., the triangle model). Our

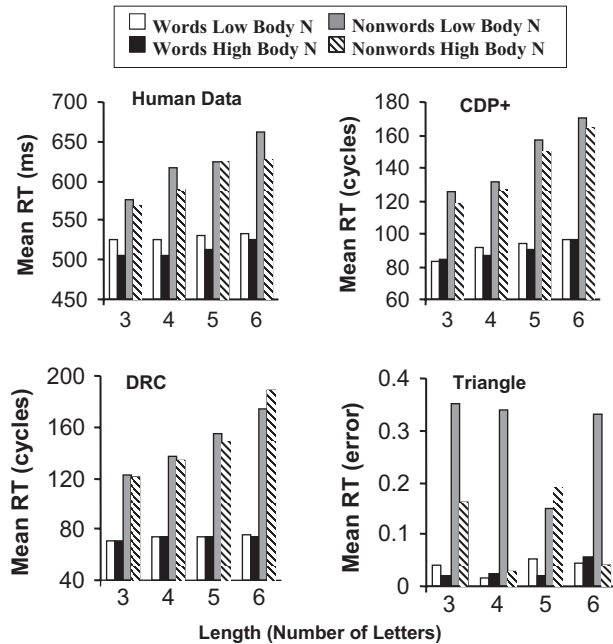


Figure 10. Human data (in milliseconds) and simulations of Ziegler et al.'s (2001) experiment, which manipulated word length, lexicality, and body neighborhood. N = neighborhood; RT = reaction time; CDP+ = new connectionist dual process model; DRC = dual-route cascaded model.

aim here is simply to show that the new model accounts for the effect at least as well as DRC and CDP.<sup>5</sup>

When confronted with the items of Rastle and Coltheart (1999), CDP+ displayed a pattern of results very similar to that of the human data. Planned comparisons showed that the first position irregular words were significantly slower than their controls (9.20 cycles),  $t(38) = 2.13$ ,  $SE = 4.32$ ,  $p < .05$ , and the second position irregulars were significantly slower than their controls (6.47 cycles),  $t(74) = 2.57$ ,  $SE = 2.52$ ,  $p < .05$ , but the third position irregulars were not significantly slower than their controls (2.79 cycles),  $t < 1$ . The human data showed a similar pattern, although the second position irregular items were only marginally significant: for first position, 61.15 ms,  $t(38) = 4.94$ ,  $SE = 12.38$ ,  $p < .001$ ; for second position, 12.87 ms,  $t(75) = 1.74$ ,  $SE = 7.41$ ,  $p = .086$ ; for third position: 0.07 ms,  $t < 1$ .

CDP+ accounted for 21% of the item-level variance, which is almost twice as much as the variance explained by DRC (13%) and CDP (12%; see Appendix J for further details). As noted in Coltheart et al. (2001), the position-of-irregularity effect was not captured by the triangle model.

In response to Zorzi's (2000) demonstration that the results of Rastle and Coltheart (1999) could be simulated by CDP (despite being a purely parallel model) and that these results were likely due to a grapheme consistency confound in their stimuli, Roberts et al. (2003) ran a new study examining the position-of-irregularity effect with a different set of words. These were chosen beforehand such that CDP could not simulate a position-of-irregularity effect with them. Only second and third position irregular words were used.

As previously found by Rastle and Coltheart (1999), Roberts et al. (2003) showed bigger regularity effects with second than with

third position irregular words. However, the Regularity  $\times$  Position interaction was only marginally significant in the item data—even more marginal if a between measures rather than repeated measures comparison was done,  $F(1, 100) = 2.47$ ,  $MSE = 0.83$ ,  $p = .12$ .<sup>6</sup> The absolute RTs also looked a little different, with the second and third position irregular words displaying very similar mean RTs and the mean RT of the control words in the two groups differing. Although Roberts et al. were not able to give a definitive reason why this difference existed compared with Rastle and Coltheart's data, they did point out that it might be due to factors such as differences in the voicing of initial phonemes across the two groups. Thus, RTs without such problems might in fact look more like those they expected.

Despite potential difficulties in interpreting Roberts et al.'s (2003) data, we examined how the different models would simulate the results that were found. CDP+ showed a pattern relatively similar to that of the data. The second position irregular words were significantly slower than their controls,  $t(63) = 4.40$ ,  $SE = 2.77$ ,  $p < .001$ , and although the effect was smaller, unlike the data, so were the third position irregular words,  $t(34) = 2.83$ ,  $SE = 2.49$ ,  $p < .001$ . Just like the data, however, the interaction was not significant,  $F(1, 97) = 1.53$ ,  $MSE = 154$ ,  $p = .22$ .

Further investigation of DRC, CDP, and the triangle model showed that all of them in fact predicted a significant effect on third position irregular items that did not appear in the human data: for DRC,  $t(34) = 3.40$ ,  $SE = 0.64$ ,  $p < .005$ ; for CDP,  $t(32) = 2.36$ ,  $SE = 0.32$ ,  $p < .05$ ; for triangle,  $t(30) = 2.59$ ,  $SE = 0.018$ ,  $p < .05$ . In terms of the Regularity  $\times$  Position interaction, only DRC incorrectly predicted a significant interaction,  $F(1, 96) = 34.36$ ,  $MSE = 509$ ,  $p < .001$  (CDP:  $F < 1$ ; triangle:  $F < 1$ ). The item-level correlations with the human data also differed across the models, with only CDP+ explaining a significant amount of the variance (CDP+: 6.29%,  $p < .01$ ; CDP: 1.56%,  $p = .24$ ; triangle: 2.59%,  $p = .10$ ; DRC:  $r = -.066$  [0.44% in the incorrect direction],  $p = .52$ ). Overall, although hard to interpret, these results suggest that the CDP+ was at least as good as the other models in simulating this experiment. For further details, all results are presented in Appendix K.

<sup>5</sup> One problem with trying to interpret the data is that, apart from the confound in the stimuli set mentioned by Zorzi (2000), there are also a few other potential difficulties. The stimuli differ across groups on both orthographic frequency (average frequencies for irregulars at Positions 1, 2, and 3 are 663, 440, and 277 and for controls are 102, 107, and 102; average log frequencies for irregulars are 1.94, 2.29, and 2.08 and for controls are 1.79, 1.83, and 1.77) and phonological frequency (average phonological frequencies for irregulars at Positions 1, 2, and 3 are 2,277, 2,809, and 156 and for controls are 73, 317, and 35; average log phonological frequencies for irregulars are 2.17, 2.12, and 1.48 and for controls are 1.07, 1.19, and 0.99). Thus, some of the regularity effect may have been obscured by frequency confounds. A further difficulty for this study is that Rastle and Coltheart (1999) used randomization tests, rather than more standard  $t$  tests. For the sake of consistency in analyzing and reporting data across different experiments, we recalculated the results of Rastle and Coltheart with pairwise  $t$  tests and did not use covariates.

<sup>6</sup> This analysis was also done on RTs converted into  $z$  scores, as they were the only item results made available to us.

### Summary: Serial Effects

Traditionally, serial effects have been thought to challenge parallel models. Particularly damaging for previous connectionist models was their difficulty in handling length effects in nonword naming. Here we showed that CDP+ successfully simulated length effects in nonword reading and the Length  $\times$  Lexicality interaction. In this respect, CDP+ was largely superior to its connectionist predecessors (the triangle and CDP model) and was comparable with DRC. In addition, CDP+ was able to account for the position-of-irregularity effect; it even accounted for more of the item-specific variance than DRC, whose theoretical commitment to serial processing was at the very origin of this research. We assess later what aspect of CDP+ was responsible for the improved performance.

### Item-Level Database Performance

A particularly hard modeling test concerns the prediction of item-level variance in large-scale databases. Spieler and Balota (1997) argued that a successful model of reading aloud should be able to account for at least as much variance as the three most important factors that affect written word naming. In their analyses, conducted on their database of almost 3,000 monosyllabic words, the three factors were log word frequency, orthographic neighborhood, and orthographic length, which collectively accounted for 21.7% of the variance. Strikingly, none of the computational models came even close to this result. In fact, the proportion of variance accounted for by the three models on the human latencies in Spieler and Balota's database was between 3% and 7% (see Coltheart et al., 2001), which is less than the variance accounted for by the single factor log word frequency.

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. Thus, the critical test was to check how much variance CDP+ would account for both in comparison with its competitors and in comparison with the three most important factors. We computed the percentage of variance accounted for by the different models across the most relevant large-scale databases: (a) Spieler and Balota (1997), (b) Balota and Spieler (1998), (c) Treiman et al. (1995), and (d) Seidenberg and Waters (1989). Because of the

large number of items in all of the databases, we did not use a 3 *SD* cutoff. The results appear in Table 2.

The data showed that CDP+ was far superior to all of its competitors in predicting item-level variance. On average, it accounted for more than three times the variance accounted for by any of the other models. Moreover, CDP+ passed the second and probably even harder modeling test by accounting for as much of the variance as the three most important factors in reading aloud together. We consider this to be a major advancement in the area of computational modeling of reading aloud.

### Other Phenomena in Word Naming

In the spirit of nested modeling, in the following section, we examined the model's performance on a number of relevant phenomena (benchmark effects) reported in Coltheart et al. (2001). Note that we did not attempt to simulate every single experiment reported in Coltheart et al. because some are still controversial and others have been superseded by more important experiments. Furthermore, we discuss and simulate a few additional empirical phenomena that have important theoretical implications.

The parameter set used in the following simulations was identical to that used in all of the simulations reported earlier, except for the dyslexia simulations. We did not simulate any of the lexical decision results because the focus of the model presented here is reading aloud. Note, however, one advantage of the nested modeling approach is the fact that CDP+ is equipped with a lexical route that is identical to that of DRC up to the level of the phonological lexicon, and as such it could perform lexical decision in exactly the same way as DRC.

### Benchmark Effects Reported by Coltheart et al. (2001)

Here, we go through the list of benchmark effects chosen by Coltheart et al. (2001). We discuss how CDP+ accounts for each effect and how relevant each is as a benchmark for computational modeling studies.

*1. Frequency effect.* Reading aloud is faster for high-frequency words than for low-frequency words (e.g., Forster & Chambers, 1973). The model's sensitivity to frequency was shown in the previous section because several of the simulated experiments included a frequency manipulation (see Figures 6 and 10).

Table 2  
Percentage of Variance Accounted for ( $R^2$ ) by the Models, by Word Frequency, and by the Three Most Important Factors (Orthographic Length, Frequency, and Orthographic Neighborhood) on Four Databases

Database	Model data				Factors	
	DRC	CDP	Triangle	CDP+	Frequency	Three factors
Spieler and Balota (1997)	3.69	5.87	3.3	17.28	7.3	21.8
Balota and Spieler (1998)	5.45	6.67	2.9	21.56	12.2	21.8
Treiman et al. (1995)	4.81	6.51	3.3	15.91	4.6	8.2
Seidenberg and Waters (1989)	6.05	2.67	3.0	9.62	1.9	10.1

*Note.* Results for the triangle model are taken from Balota and Spieler (1998). DRC = dual-route cascaded model; CDP = connectionist dual process model; CDP+ = new connectionist dual process model.



2. *Lexicality effect.* Reading aloud is faster for regular words than for nonwords (e.g., McCann & Besner, 1987). Several of our simulations have shown a robust lexicality effect (e.g., the simulations of Weekes, 1997, and Ziegler et al., 2001; see also Figures 9 and 10).

3. *Regularity effect.* Reading aloud is faster for regular words than for irregular words. Previous simulation studies focused on the fact that regularity is reliable for low-frequency words but smaller or absent for high-frequency words (as in the studies of Paap & Noel, 1991; Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Taraban & McClelland, 1987). However, as discussed before, all the earlier studies have methodological shortcomings, although they were certainly state-of-the-art experiments at the time. Today, the best controlled regularity experiments are Jared's (2002) showing that regularity affects both high- and low-frequency words when neighborhood characteristics are controlled for. We note that CDP+ was the only model that successfully simulated all of Jared's experiments. Indeed, when the stimuli of one of the earlier studies (Paap & Noel, 1991) are submitted to CDP+, the model shows a robust Frequency  $\times$  Regularity interaction. A simulation of Paap and Noel's (1991) study is reported in Appendix L.

4. *Position-of-irregularity effect.* The size of the irregularity cost declines as a function of the position at which irregular words have their exceptional grapheme-phoneme correspondence (Rastle & Coltheart, 1999). This effect was successfully simulated in an earlier section (see Appendix J).

5. *Pseudohomophone effect.* Pseudohomophonic nonwords (i.e., nonwords that can be pronounced like words, e.g., *cheet*) are read aloud faster than nonpseudohomophonic nonwords (McCann & Besner, 1987; Seidenberg, Petersen, MacDonald, & Plaut, 1996; Taft & Russell, 1992). A simulation of McCann and Besner's (1987) study showed significantly faster latencies for pseudohomophones compared with nonwords that were not pseudohomophones (137.4 cycles vs. 145.24 cycles),  $t(123) = 2.32$ ,  $SE = 3.39$ ,  $p < .05$ ; 14 errors; 3 outliers. The model also correlated significantly with the individual item results,  $r = .21$ ,  $p < .05$  ( $N = 125$ ).

There is an ongoing debate over whether one should expect a base-word frequency effect for pseudohomophones. According to the recent review of Reynolds and Besner (2005), there is no evidence for a reliable base-word frequency effect when reading aloud pseudohomophones in standard experimental conditions, that is, when pseudohomophones are mixed with nonwords. Instead, they suggested that the effect of base-word frequency was found only in studies in which pseudohomophones of different base-word frequencies were presented in separate blocks. An additional difference they noted that appears to be caused by presenting stimuli in separate blocks is that people also name pseudohomophones more slowly than nonword controls.

To assess whether CDP+ would show a base-word frequency effect, we correlated the base-word frequency of the pseudohomophones in McCann and Besner (1987) with CDP+'s naming latencies. The correlation was weak and only marginally significant ( $r = -.23$ ,  $p = .08$ ,  $N = 60$ ). However, keep in mind that base-word frequency effect is present only in a blocked design, in which pseudohomophones take a longer time to be read than nonwords in pure lists. This might suggest that, in a pure list of pseudohomophones, participants adopt a higher response criterion

to allow more time to retrieve a lexical pronunciation. Accordingly, we ran a simulation with CDP+, in which the phoneme naming activation criterion was increased from .67 to .73. The results showed that under these conditions, the correlation between base-word frequency and the model's latencies became significant ( $r = -.30$ ,  $N = 61$ ,  $p < .05$ ). Conversely, reducing the criterion to .64 to simulate a mixed-list condition reduced the correlation with base-word frequency ( $r = -.099$ ,  $N = 61$ ,  $p = .47$ ) but did not eliminate the pseudohomophone effect (133.08 vs. 139.66 cycles),  $t(124) = 1.90$ ,  $SE = 3.46$ ,  $p = .06$ . Thus, the results suggest that the higher the response criterion used in the model, the greater the size of the base-word frequency effect, and that this effect may not necessarily be significant even when the pseudohomophone effect is.

It is worth noting that our explanation for the occurrence of a base-word frequency effect in pure lists is much more parsimonious than the one proposed by Reynolds and Besner (2005). These authors suggested that DRC could simulate a base-word frequency effect by changing four different parameters of the models. Instead, we suggest that it is sufficient to strategically change a single parameter to simulate the full pattern reported by Reynolds and Besner. This is achieved in the following way: (a) Mixed lists are read aloud with a low phoneme naming activation criterion (.64 in our simulation), as are lists of nonhomophonic nonwords (as shown earlier, in mixed lists there is no base-word frequency effect); (b) pure pseudohomophone lists are read aloud with a higher phoneme naming activation criterion (.73 in our simulation) that results in a base-word frequency effect; and (c) because pseudohomophones are read aloud with a higher phoneme naming activation criterion in pure lists, they are read aloud more slowly compared with nonwords in pure nonword lists. Thus, pseudohomophones in pure lists (.73 criterion) appear to have a disadvantage compared with nonwords in pure lists (.64 criterion), as was indeed found in the human data (pseudohomophones in the .73 criterion condition: 148.92 cycles; nonwords in the .64 criterion condition: 139.66 cycles),  $t(124) = 2.66$ ,  $SE = 3.44$ ,  $p < .01$ .

6. *Orthographic neighborhood (N) effect.* Nonwords with many orthographic neighbors are read aloud faster than nonwords with few or no such neighbors (e.g., Andrews, 1997). Coltheart et al. (2001) used a list of 244 nonwords randomly selected from a database to examine the effect in DRC and obtained a significant correlation of  $-.154$  between neighborhood size and naming latency. On the same set of stimuli, CDP+ produced an almost identical correlation ( $r = -.13$ ,  $p < .05$ ,  $N = 224$ ).

Andrews (1989, 1992) also reported a facilitatory orthographic neighborhood effect for real words, especially when they were of low frequency. DRC failed to simulate this effect (Coltheart et al., 2001, p. 224). To our great surprise, CDP+ also failed to simulate Andrew's *N* effect: for Andrews (1989),  $F < 1$ ; for Andrews (1992),  $F(1, 92) = 2.63$ ,  $MSE = 158$ ,  $p = .11$ . We therefore tried to understand why orthographic neighborhood effects appeared to cause problems for current models of reading aloud.

A review of the recent literature made clear that the empirical situation around orthographic *N* effects for words is actually far from settled. In a recent analysis of naming latencies from a large-scale database containing 2,284 words, Baayen et al. (2006) concluded that the unique variance accounted for by orthographic neighborhood was rather small (0.36%). Although orthographic neighborhood had a facilitatory effect in their analysis, the effect

tended to asymptote quickly (no extra facilitation for words with more than five neighbors). More important, in a methodologically very stringent small-scale study, Mulatti, Reynolds, and Besner (2006) did not find  $N$  effects for words. Indeed, Mulatti et al. argued that previous studies might have found facilitatory  $N$  effects because they confounded orthographic neighborhood with phonological neighborhood. In fact, when Mulatti et al. controlled for phonological neighborhood, they no longer found a reliable effect of orthographic neighbors (for a similar pattern, see also Peereman & Content, 1997), whereas they still found a facilitatory effect of phonological neighbors.

We therefore submitted Mulatti et al.'s (2006) highly controlled stimulus set to CDP+. Note that neither DRC with its standard parameter set nor any version of the triangle models (including the triangle network) was able to capture their data (see Mulatti et al., 2006). However, with the standard parameter set, CDP+ accurately predicted both the absence of an orthographic  $N$  effect (high vs. low orthographic neighbors: 100.3 vs. 98.76 cycles, respectively),  $t < 1$ , as well as the presence of a facilitatory phonological  $N$  effect (high vs. low phonological neighbors: 96.8 vs. 102.0 cycles),  $t(58) = 2.22$ ,  $SE = 2.36$ ,  $p < .05$ .

Given that the empirical situation around orthographic  $N$  effects is not yet settled, we believe that at present, it is not warranted to take neighborhood size effects on words as a major benchmark. In contrast, orthographic neighborhood at a slightly larger grain size (bodies instead of letters) has consistently shown a facilitatory effect in a number of word naming experiments (e.g., Brown, 1987; Forster & Taft, 1994; Ziegler et al., 2001). The latter effect is simulated by CDP+ (see simulation of Ziegler et al., 2001, Figure 10). Given the reliability of the body neighborhood effect, it might be a better benchmark than the effects of orthographic neighborhood.

*7. Priming of reading aloud.* Priming can be studied using either conscious (unmasked) or unconscious (masked) primes. Masked priming (e.g., Forster & Davis, 1991) is a paradigm in which primes are presented so quickly and are masked such that they are typically not open to conscious awareness. Clearly, the data from masked priming are more relevant for the evaluation of computational models because unmasked priming is subject to all kinds of high-level strategies that are beyond the scope of the models (see, e.g., Borowsky & Besner, 2006; Neely, 1977; Plaut & Booth, 2000). Although there are extremely constraining data from masked priming on the time course of orthographic and phonological activation in French and Hebrew (Ferrand & Grainger, 1992, 1993, 1994; Frost, Ahissar, Gotesman, & Tayeb, 2003; Ziegler, Ferrand, Jacobs, Rey, & Grainger, 2000), such detailed time-course data are not available in English. Therefore, most previous simulation efforts have focused on the onset effect in masked priming (Forster & Davis, 1991).

The onset effect reflects the finding that people are faster in reading aloud a target word when the prime shared the same onset with the target word (an onset prime) compared with when the prime did not share any phonology with the target word (a control prime) and also compared with when the prime rhymed with the target word but did not share the same onset (a rhyme prime). In addition, there was no significant difference between the group where the primes rhymed with the target words and the group where they did not share any phonology (but see Montant & Ziegler, 2001, for an explanation of this null effect).

Coltheart, Woollams, Kinoshita, and Perry (1999) simulated these results with DRC by presenting the prime word for a small number of cycles before the target word was presented to the model. To simulate this effect with CDP+, we used the same strategy as Coltheart et al. and used the same items that were used in the DRC simulation (i.e., a number of items were removed because they were disyllabic and therefore nonextant in the model). We used the same parameters as the normal model, although this was mainly for convenience, and we leave open the possibility that when primes are used, the parameters people use when reading aloud may be different from those when primes are not used. This may be particularly the case with word primes, because readers may need to reduce interference coming from the prime. To simulate the priming effect, we presented prime words to the model for 25 cycles and then presented the target word without any changes to the activation in the model that had built up from the processing of the prime word. The results showed essentially the same pattern as the data. Words preceded by a prime that shared the same onset were read aloud 8.26 cycles ( $SD = 3.21$ ) faster than unrelated controls. Words preceded by a prime that shared the same rime were read aloud only 2.15 cycles ( $SD = 2.24$ ) faster than unrelated controls. A set of  $t$  tests confirmed that CDP+ read aloud target words preceded by an onset prime significantly faster than controls,  $t(18) = 11.32$ ,  $SE = 0.74$ ,  $p < .001$ . Surprisingly, despite the small size of the effect, CDP+ also read aloud words preceded by a rhyme prime significantly faster than the controls,  $t(18) = 4.19$ ,  $SE = 0.52$ ,  $p < .005$ . While the presence of a rhyme priming effect is inconsistent with the original null effect of Forster and Davis (1991), it is worthwhile pointing out that Montant and Ziegler (2001) managed to show a facilitatory rhyme priming effect in masked priming once they neutralized the strongly inhibitory influence of the mismatching onset by replacing the onset with a hash mark (i.e., #ake primed *make* but *fake* did not prime *make*). Thus, CDP+ not only simulates the onset effect but it seems sensitive to the residual effects of rhyme priming.

### Miscellaneous Effects

*1. Whammy effect.* Rastle and Coltheart (1998) reported that five-letter nonwords that contained multiletter graphemes such as *ph*, for which the pronunciation of the first letter is different from the pronunciation of the whole grapheme, were read aloud more slowly than five-letter nonwords made of simple one-letter graphemes. They called this effect the *whammy effect*. CDP+ was presented with these rather difficult nonwords (e.g., *fooph*). On the positive side, CDP+ made only two errors, which is an important achievement given that its predecessor, CDP, produced around 50% errors. On the negative side, however, CDP+ predicted no significant difference between the “whammied” group (complex graphemes) and the “unwhammied” group (152.35 vs. 153.52 cycles, respectively;  $t < 1$ ).

We tried to understand why CDP+ failed to predict a difference between these two groups. The first question one needs to ask is whether the effect is empirically robust. The answer is not very encouraging. First, in Rastle and Coltheart's (1999) own item data, the whammy effect was significant only when a within-item repeated measures ANOVA was used. However, it was no longer significant when a between-items comparison was used,  $F(1,$

46) = 2.82,  $MSE = 3,816$ ,  $p = .10$ . Because one can never match items on all critical variables, it is more conservative when factors that rely on comparisons between different items are analyzed using between-groups comparisons (this has been done throughout the article across all studies).

Second, there are a number of studies that cast doubt on the reliability and meaningfulness of the whammy effect (e.g., Andrews, Woollams, & Bond, 2005; Martensen, Maris, & Dijkstra, 2003). Most important, Lange and Content (2000) showed that grapheme complexity (whammy status) is confounded with grapheme frequency. Whammied items tend to have graphemes of a lower frequency than unwhammied items. When Lange and Content controlled for grapheme frequency, they actually found that nonwords with complex graphemes were named faster than nonwords with simple graphemes (the opposite of the whammy effect).

Nevertheless, one could argue that even if there is ongoing debate about what mechanism explains the whammy effect (competition between rules or grapheme frequency), a model should still be able to account for the data to the extent that the data are empirically robust. We agree. If the robustness of the effect were established in further studies, there would be at least two possibilities to account for the effect in CDP+. First, we could make the graphemic buffer sensitive to grapheme frequency. Because the whammied items in Rastle and Coltheart (1999) had lower grapheme frequencies, the model should be able to simulate the effect. Second, we could implement competition in the graphemic buffer. At present, for any given slot, single letter graphemes are in no competition with multiletter graphemes. By making grapheme parsing competitive, we would have a reasonable chance of picking up the whammy effect.

**2. Strategic effects.** Rastle and Coltheart (1999) examined the effect of regular word and nonword reading in two conditions. In one, the hard condition, first-position irregular words were used as fillers; in the second, the easy condition, third-position irregular words were used as fillers. The results showed that both words and nonwords were read aloud more slowly in the hard condition than in the easy condition. The effect appeared to differ in absolute size for words and nonwords, with words showing about half of the latency cost exhibited by the nonwords (11.7 ms vs. 22.6 ms). Rastle and Coltheart tried to model this finding by slowing down the GPC route, that is, by increasing the amount of time it took to assemble each letter from 17 to 22 cycles. This caused a latency increase that was much stronger with nonwords (23.11 cycles) than with words (0.68 cycles). We carried out a similar manipulation with CDP+. Decreasing the speed of graphemic parsing had an effect very similar to that observed in the human data. For instance, when the delay between the processing of each letter was increased from 15 to 17 cycles, it resulted in a mean latency increase of 7.94 cycles for nonwords and 2.94 cycles for words. Note that DRC, unlike CDP+, seems to markedly overestimate the ratio of nonword to word latency increase induced by the strategic manipulation.

**3. Surface dyslexia.** Patients show a specific impairment of irregular word reading, which is modulated by word frequency (Behrmann & Bub, 1992; McCarthy & Warrington, 1986) but also by the consistency ratio of the words (Patterson & Behrmann, 1997). We have already shown in the previous section that CDP+ (but not DRC) can perfectly capture this pattern of performance.

**4. Phonological dyslexia.** One effect that is particularly challenging is that some (but not all) phonological dyslexic patients read pseudohomophones more accurately than nonpseudohomophonic nonwords (see patient LB; Derouesné & Beauvois, 1985). Moreover, pseudohomophones that are orthographically close to their base word are named more accurately than pseudohomophones that are orthographically far from their base word (e.g., *sead* vs. *phocks*; see Coltheart et al., 2001, for a review). This suggests that when reading nonwords, these phonological dyslexics may try to boost their performance with some sort of lexical support strategy.

Coltheart et al. (2001) modeled the performance of patient LB with DRC by changing two parameters: (a) They reduced the speed at which phonology is assembled, and (b) they increased the phoneme activation criterion to simulate nonspeeded reading aloud. This approach, however, implies that the effect of brain damage is simply a slowing down of phonological assembly. Instead, we have opted for a different approach, according to which neural damage results in lower excitability of the neural structures encoding specific knowledge and in increased susceptibility to noise. Accordingly, we simulated phonological dyslexia with the following manipulation: (a) the sublexical network to phonological output buffer activation was reduced from 0.085 to 0.06 (i.e., we reduced the amount of phonology, rather than the speed of its generation), and (b) the strength of all inhibitory connections in the model was halved. The latter has the effect of increasing 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. Note that the nonlexical network does not produce a clean phonological output representation, but it generates multiple phoneme candidates for any given letter string, particularly for the vowel position (see Zorzi et al., 1998b, for detailed analyses). Thus, inhibition helps in removing competing representations (i.e., noise) that can lead to spurious outputs.

The changes to the parameters led to a set of results that looked very much like those of LB, with the pseudohomophones displaying a strong effect of orthographic similarity (see Figure 11). Orthographically similar pseudohomophones were named much more accurately than orthographically dissimilar ones (72.5% vs. 32.5%) and were named more accurately than two matched nonword control groups (32.5% and 38.5%). In contrast, the model

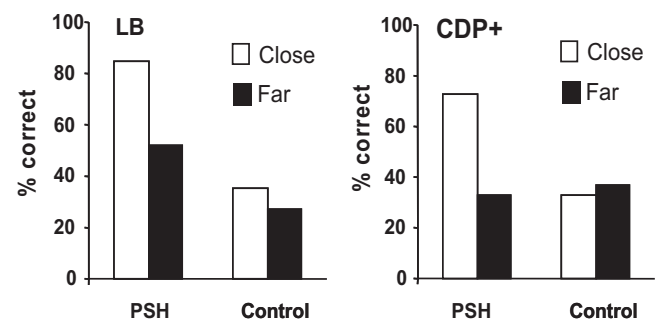


Figure 11. Proportion of correct responses on orthographically close and orthographically distant pseudohomophones and their respective nonword controls for patient LB and CDP+. PSH = pseudohomophone; CDP+ = new connectionist dual process model.

was still fairly accurate in word reading. On the stimuli of Jared (2002), which we used because the original stimuli LB was tested on were not available, the model produced a slightly higher error rate with the dyslexic compared with the normal parameter set (14.03%). Note that the error rate on words was slightly increased in LB too.

### *Reading Aloud Strange Nonwords*

An important challenge for a connectionist account of nonword reading is the ability to read nonwords with spelling patterns that are extremely uncommon or even illegal in real words (such as *jinje* or *rhawnce*). Although the CDP+ model can deal with difficult nonwords like the set of Rastle and Coltheart (1998) with very low error rates (also see the *Credit Assignment and Componential Analyses* section), there might still be a problem with the most odd-looking or illegal nonwords, such as *sklyb* and *ghroumn*.<sup>7</sup> It is at present unclear how people actually read these letter sequences. It is very clear, however, that illegal nonwords are processed in a qualitatively different way from legal nonwords; they do not seem to enter the normal word processing circuit (see Petersen, Fox, Snyder, & Raichle, 1990; Ziegler, Besson, Jacobs, Nazir, & Carr, 1997, for neuroimaging and electrophysiological evidence). Our suggestion is that different strategies are used when nonwords like *ghauxte* or *skryb* are encountered. The most obvious strategy would be an individual grapheme read-out strategy.

Such a strategy can be simulated with CDP+ by processing the graphemes one at a time in the sublexical route. For instance, the nonword *sklyb* can be broken into the graphemes *s-ck-l-y-b* simply by using the normal method with which letter strings are broken down by the model. These can then be presented to the model individually. The problem of finding a grapheme pronunciation at uncommon or illegal positions (e.g., the grapheme *ck* does not occur as an onset in English and therefore it does not produce any activation in an onset position) can be solved by presenting it in the first available position where it does produce activation (e.g., *s\*\*\*\*\**, *\*\*\*\*ck\*\**, *\*\*I\*\*\*\*\**, *\*\*\*\*y\*\*\**, *\*\*\*\*b\*\**). Because the model is quite good at reading single correspondences, this strategy allows the model to adequately read these nonwords (note that a similar strategy might also be used with CDP and the triangle model). Clearly, more empirical data are needed before one should take reading illegal nonwords as a viable benchmark for the evaluation of computational models of reading aloud.

### *Credit Assignment and Componential Analyses*

An important component of our nested modeling approach is the attempt to identify the source of the improved performance of CDP+ compared with its predecessors. Indeed, the new model is more than a simple upgrade of CDP because important changes have been introduced even to the nonlexical part (i.e., the graphemic buffer and serialized input). It is also a major departure from DRC because it dispenses with the hardwired production system of GPC rules and because it also uses a different phonological output system (with onset–vowel–coda organization and no null phonemes).

Later, we present analyses that aim at assigning credit to the major changes. First, we look at the effect of the graphemic buffer

as a determinant of the improved nonword reading performance. Second, we investigate the effect of serializing the nonlexical network by comparing CDP+ with a parallel version of the same model. Third, we examine the contribution of the nonlexical network to the simulations of the consistency effects, using a version of CDP+ where feedback and neighborhood effects are eliminated from the lexical route. Finally, we look at the contribution of isolated lexical and sublexical routes in predicting item latencies in large-scale databases.

### *Role of the Graphemic Buffer in Nonword Reading*

Perhaps the most important failure of Zorzi et al.'s (1998b) CDP model was the high error rates it produced in reading difficult nonwords such as the whammy stimuli of Rastle and Coltheart (1998), where it produced an error rate of about 50%. This data set can therefore be considered the best test for evaluating the effect of replacing the letter-based input of the TLA network (i.e., the sublexical route) with graphosyllabic representations, that is, the graphemic buffer of Houghton and Zorzi's (2003) dual-route model of spelling. We therefore ran just the nonlexical network of CDP+ on the whammy data set. The model produced only one error out of 48 items, reading *glect* as */glest/* (note that the full model made two errors instead of one because it produced one lexicalization error when the lexical route was on). It therefore appears that the nonlexical network of CDP+ is much superior to its predecessor. It should also be noted that the set of graphemes used in the graphemic buffer is the subset selected by Houghton and Zorzi, which is far smaller than the entire set that can be found in English words (for an analysis, see Perry & Ziegler, 2004).

### *Parallel Versus Serial Nonlexical Route*

A second change to the nonlexical route of CDP was the serialization of input. Therefore, we investigated how CDP+ performs when the sublexical network is not serialized but works in parallel like CDP. Perhaps the most uncontroversial benchmark effect for serial processing in reading aloud is the nonword length effect (see earlier discussion). To examine this, we compared the performance of CDP+ on Weekes's (1997) nonwords with that of a parallel version of the model. For this purpose, rather than inputting the graphemes into the sublexical part of the model in the serial way as was described previously, we simply presented all graphemes to the model at the same time. This caused the model to produce four errors. More important, the results showed that there was no hint of a length effect. In fact, the fastest latencies were obtained for the longest words. Moreover, the item correlations were particularly poor. The parallel CDP+ accounted for only 4.92% of the variance, whereas the normal CDP+ accounted for 30.77%. It therefore appears that serial processing in the model is crucial for capturing one of the benchmark effects. The results appear in Figure 12.

A further question is how crucial the serial processing is for the performance on real words. For this purpose, we compared the parallel model with the serial model on the large-scale database of Spieler and Balota (1997). The results are reported in Table 3. The

<sup>7</sup> We thank Max Coltheart for suggesting that we consider this issue.



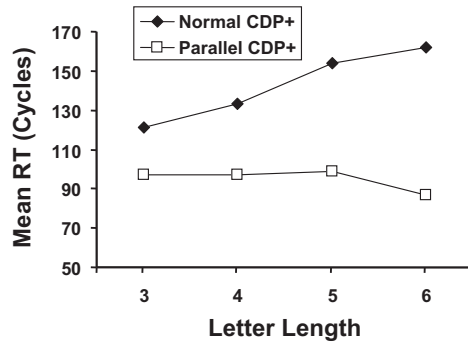


Figure 12. Mean response times (RT) of the new connectionist dual process model (CDP+) and the parallel CDP+ on the nonwords of Weekes's (1997) experiment, which varied as a function of orthographic length.

parallel CDP+ accounted for only 7.72%, whereas the normal CDP+ accounted for 17.27%. Clearly then, serial processing in the model helps not only nonword reading but word reading as well.

### Lexical Route and Consistency Effects

The simulations of Zorzi et al. (1998b) and Zorzi (2000) with CDP showed very clearly that consistency effects arise in the sublexical route, that is, in the TLA network. They cannot arise in the lexical route simply because it was not implemented beyond the provision of frequency-weighted lexical phonological activation. However, as suggested by Coltheart et al. (2001), consistency might also affect naming latencies through neighborhood characteristics of the lexical route. As shown by the simulations with DRC, lexical influences are too weak to account for the majority of the consistency effects reported in the literature. Nonetheless, it is useful to investigate whether the fully implemented lexical route in CDP+ plays any role in its success in simulating Jared's (2002) data.

To examine the contribution of the interactive recurrent processing in the lexical route, we turned CDP+ into a purely feedforward model, and we completely eliminated the activation of orthographic neighbors. To do this, we set all of the feedback parameters in the model to zero. Thus, unlike the normal parameter set, there was no excitation/inhibition from the phonological output buffer to the phonological lexicon; the phonological lexicon did not activate the orthographic lexicon; and there was no excitation/inhibition from the orthographic lexicon to the letter level. The setting of these parameters to zero causes the activation in the

network to build up at the phonological output buffer slightly more slowly than normal; to compensate for this, we increased the phonological lexicon to phonological buffer excitation parameter from .128 to .135. Finally, we increased the strength of inhibition from the letter level to the orthographic lexicon from  $-0.55$  to  $-1.00$ . Increasing this parameter to such a strong level means that no orthographic neighbors of a word are ever activated. Note that these changes mean that the lexical route of the feedforward CDP+ simply produces frequency-weighted lexical phonological activation, much like in the simulations of Zorzi et al. (1998b) with CDP.

The performance of the feedforward CDP+ on the items of Jared's (2002) study was virtually identical to the normal CDP+. All effects that were significant with the normal parameter set were still significant, and all effects that were not significant with the normal parameter set were not significant in the feedforward CDP+. In addition, the correlations with the item data were almost the same (the  $r$  squares were all within .03 of each other). Thus, the feedforward CDP+ allowed us to isolate the source of the consistency effects and to firmly establish that they are entirely produced by the sublexical network.

One additional issue that can be addressed by examining the performance of the feedforward CDP+ is that of the contribution of a fully implemented lexical route to explaining item-level variance on the large-scale databases. The variance accounted for by the feedforward CDP+ on Spieler and Balota's (1997) data set was almost identical to that accounted for by the normal CDP+ (see Table 3). This shows that the improved performance of CDP+ cannot be attributed to the addition of a fully implemented lexical route.

### Contribution of Lexical and Sublexical Routes in Predicting Item Latencies

As another test related to the issue of credit assignment, we examined how the lexical and nonlexical parts of the model contributed in accounting for item-specific variance on the large-scale databases of Spieler and Balota (1997) and Balota and Spieler (1998), which had 2,807 words that were also in the database used by CDP+. To examine the lexical part of the model, we turned off the nonlexical part of the model. The model made 1 error (note that heterographic homophones were considered correct as long as they produced one of the potential pronunciations). To test the sublexical part of the model, we simply ran the model with the lexical route switched off. The model made 268 (9.55%) errors (note that because it uses only a two-layered network, the model cannot reach 100% accuracy on irregular words, unlike the

Table 3

Percentage of Variance Accounted for ( $R^2$ ) by the Different Variants of CDP+, by Frequency, and by the Three Most Important Factors (Orthographic Length, Orthographic Frequency, and Orthographic Neighborhood) on Two Databases

Database	Model data					Factors	
	Normal CDP+	Parallel CDP+	Feedforward CDP+	Lexical CDP+	Nonlexical CDP+	Frequency	Three factors
Spierer and Balota (1997)	17.28	7.72	17.30	9.24	5.83	7.3	21.8
Balota and Spierer (1998)	21.56	12.11	20.69	15.26	4.06	12.2	21.8

Note. CDP+ = new connectionist dual process model.

triangle model). Of the errors, 199 (74.25%) appeared to be legitimate alternative readings (e.g., reading *blood* to rhyme with *food*). The other errors were varied and consisted of responses unlikely to be found with normal readers, including missing phonemes, unlikely spelling–sound translations, cases where the same phoneme was repeated (e.g., /wudd/ for *would*), and combinations of these problems.

As can be seen from Table 3, reading aloud latencies produced by the isolated nonlexical part of the model correlated quite poorly with the human data. More important, however, neither the lexical nor the nonlexical part of the model was able to fulfill the criterion, suggested by Spieler and Balota (1997), that models should be able to perform at least as well on item data as a simple correlation with well-established factors (i.e., orthographic frequency, orthographic neighborhood, and orthographic word length). However, as shown by the simulations with the feedforward CDP+, there is little or no contribution of the lexical phonology beyond the simple effect of frequency. These results therefore show that single parts of the model are inadequate when used in isolation and provide a baseline as to how well the individual parts of the model perform compared with the model as a whole. They also show that a simple feedforward lexical route can account for amounts of variance on the large databases similar to those accounted for by the network used with feedback connections.

### General Discussion

Computational modeling of reading aloud has become one of the most sophisticated and advanced areas of cognitive psychology. One potential criticism, however, is that an incremental and nested modeling approach, according to which a new model should contain the best features of the previous models, is far from being a standard strategy in computational cognitive psychology. As a matter of fact, new models rarely include the old model as a special case, and new models are rarely tested against the data that motivated the development of the earlier model (Jacobs & Grainger, 1994). In the present research, we combined the best features of some of the previous models into a single new CDP+ model of reading aloud. As its name implies, CDP+ belongs to the family of dual-route models like DRC. Unlike DRC, however, it uses a fully connectionist architecture, whereas DRC includes a symbolic system based on production rules (see further discussion later). As required by a nested modeling strategy, the new model was tested both on standard benchmarks as well as on more theoretically and empirically challenging data sets.

Our starting point was to use the basic architecture of CDP developed by Zorzi et al. (1998b). Of particular importance was the sublexical component of their model, the TLA network. The TLA network is sensitive to the statistical distribution of the spelling-to-sound relationships, but it cannot learn whole-word associations; moreover, it has been shown to account for various aspects of reading development (Hutzler et al., 2004; Zorzi et al., 1998a). The distinction between sublexical and lexical processing, a basic tenet of dual-route theories, is therefore supported by a computational account that led to the development of connectionist dual-route (or dual process) models of both reading aloud (Zorzi et al., 1998b) and spelling (Houghton & Zorzi, 2003). Note that even the most recent PDP modeling work (e.g., Harm & Seidenberg,

2004) converged on the importance of a direct mapping between orthographic and phonological units.

The ability of connectionist models to generalize the knowledge of spelling–sound mapping to novel items (i.e., nonwords) has been a major source of controversy since the seminal work of Seidenberg and McClelland (1989). The advantage of DRC over its competitors in nonword reading performance is in large part due to an optimized, handcrafted set of explicit rules that encode grapheme–phoneme correspondences. However, we substantially improved nonword reading performance in the TLA network by replacing the letter input level with the graphemic buffer of Houghton and Zorzi's (2003) model of spelling. This choice is further motivated by the hypothesis that a common graphemic buffer is involved in both reading and spelling (Caramazza et al., 1996; Cotelli et al., 2003; Hanley & Kay, 1998; Hanley & McDonnell, 1997). Furthermore, input into the sublexical network was serialized as a result of graphemic parsing, which was conceived as a process that operates through an attentional window moving from left to right over the letter string (see further discussion later).

The lexical route was not fully implemented in CDP. In most of the simulations of Zorzi et al. (1998b), lexical phonology was simply implemented as a frequency-weighted activation that was pooled online with the sublexical phonology. For the new model, we used the lexical route of DRC up to the phonological lexicon. The use of a localist, interactive activation model of the lexical route is consistent with our nested modeling approach. The advantage of this solution is that the interactive activation model can account for a large number of phenomena related to orthographic processing and lexical access (see Coltheart et al., 2001; Grainger & Jacobs, 1996).

Our results showed that CDP+ was able to simulate not only Coltheart et al.'s (2001) selection of benchmark effects but also all of the critical marker effects that were chosen for their potential to adjudicate between models (i.e., consistency effects, serial effects, etc.). All of the previous models had qualitative shortcomings in some form or other: DRC had difficulties simulating the effects of consistency both on words and nonwords, the triangle model had problems simulating serial effects and the additive effects of regularity and frequency, and finally, CDP had difficulties with simulating length effects. All previous models were also rather limited in their potential to capture item-level variance in large-scale experiments. We now turn to these issues one by one. We then propose an updated list of benchmark effects and discuss which of the components of CDP+ are responsible for its superior performance. We close by indicating some shortcomings and future directions.

### Consistency and Regularity

Simulating graded consistency effects has been a major challenge for model development (see Zevin & Seidenberg, 2006). Connectionist models are well prepared to simulate these effects because in these models the naming of a word is potentially influenced by all other words that the network knows or learns (Treiman et al., 1995). The inability of DRC to account for consistency effects in words (Jared, 2002) and nonwords (Andrews & Scarratt, 1998; Treiman et al., 2003) falsifies its current implementation.

The success of CDP+ in accounting for consistency effects is entirely due to the associative learning network of the sublexical route. Note that the issue of how the sublexical route is conceived has important theoretical implications. DRC is *not* a connectionist model because its GPC route is a production system based on symbolic rules. The combination of a lexical memory with a rule system that allows productivity in language has indeed a long-standing tradition in cognitive science (e.g., Marcus, Brinkmann, Clahsen, Wiese, & Pinker, 1995), and it has faced fierce opposition from the connectionist community (e.g., Seidenberg & MacDonald, 1999). DRC offers a dual-mechanism account of reading aloud that is simply one specific instantiation of the more general class of dual-route models. Thus, falsifying DRC does not falsify dual-route models in general. In fact, CDP+ is a connectionist dual-route model that captures consistency effects better than all of its competitors. Because it is a dual-route model, it can simulate acquired surface and phonological dyslexia in very much the same way as DRC.

There are a number of ways that one might go about upgrading DRC so that it could deal with consistency effects. Perhaps the simplest way would be to suggest that some sort of frequency or consistency weighting applies to the rules that are used. However, having an associative network that learns the statistical distribution of spelling-to-sound correspondences seems to be more elegant and parsimonious than having to find a set of frequency sensitive rules capable of doing a similar job (for a discussion, see Pacton, Perruchet, Fayol, & Cleeremans, 2001). Clearly, consistency effects on words (Jared, 2002) and nonwords (Andrews & Scarratt, 1998) should be in the list of benchmark effects that need to be accounted for by the next generation of reading aloud models (see later in the article for an updated list of benchmark effects).

### *Serial Processing*

Over the past years, there has been a lively debate over whether phonological assembly occurs in a serial or parallel manner (for a review, see Coltheart et al., 2001). The data come from a number of sources: (a) the nonword length effect, (b) the position-of-irregularity effect, and (c) cross-language comparisons.

Length effects are the first class of effects that are often put forward in favor of serial processing in word and nonword reading. The typical result from English is that nonwords produce a robust length effect, whereas real words produce a much smaller length effect (Weekes, 1997; cf. Baayen et al., 2006; Balota et al., 2004). CDP+ perfectly captures this interaction. It also accounts for at least twice as much variance than the other models with words. Models with a serial assembly mechanism explain this effect by suggesting that there is an interaction between phonology generated in parallel by the lexical route and phonology generated serially by the phonological assembly mechanism (Coltheart et al., 2001). Two alternative explanations exist for parallel models. One is to attribute the effect to *dispersion*. The idea is that longer words and nonwords tend to have less frequent and more difficult spelling-sound correspondences. Thus, long nonwords tend to be slower simply because it is more likely that people will encounter a correspondence that is difficult to process. When real words are presented to the network, this effect may be eliminated or reduced because of either a parallel lexical process (according to CDP), or better whole-word orthography to whole-word phonology learning

(according to the triangle model). Simulations show that CDP does indeed produce a serial-like effect in English (Perry & Ziegler, 2002; Zorzi, 1999), whereas the triangle model does not.

Traditionally, the position-of-irregularity effect has been given strong theoretical significance (Rastle & Coltheart, 1999). However, the utility of the position-of-irregularity effect as a marker effect for serial processing has been called into question. Most important, Zorzi (2000) simulated the serial position-of-irregularity effect despite the fact that the letters were presented all at the same time (i.e., in parallel) to CDP. Zorzi attributed the success of a parallel model in simulating an apparently serial effect to a confound in the stimuli used by Rastle and Coltheart (1999). He suggested that if the irregular correspondences in their irregular words were examined in terms of the statistical degree to which they conformed to typical spelling-sound patterns, then early spelling-sound correspondences appeared more atypical and inconsistent than the later ones. Thus, for instance, the way *ch* in *chef* is pronounced is very rare, whereas the way *oo* in *book* is pronounced is not, yet both are irregular according to the definition of Coltheart et al. (1993). Zorzi suggested that his model was sensitive to this kind of grapheme consistency and that this is why CDP showed the same pattern as the human data. Note that CDP+ was also able to simulate the position-of-irregularity effect. It accounted for almost twice as much of the item-specific variance as did DRC. However, given that all models (except the triangle model) can simulate the position-of-irregularity effect, this effect is not the strongest benchmark for the evaluation of models of reading aloud.

One other way to defend parallel models in the light of serial effects is to suggest that the serial effects are simply peripheral to the task of reading aloud and therefore beyond the scope of these models. For example, Seidenberg and Plaut (1998) argued that length effects found in reading aloud may not actually reflect reading aloud per se but rather visual input or articulatory output processes that are currently not captured in the model. An argument against such a possibility is that it is quite difficult to find length effects in picture naming tasks that also require articulation. This difficulty occurs even though the length manipulations in picture naming tend to be more extreme than those in reading aloud (i.e., syllables vs. single letters; e.g., Bachoud-Levi, Dupoux, Cohen, & Mehler, 1998).

Cross-language studies of reading aloud (e.g., Ziegler et al., 2001) also provide particularly good evidence against the hypothesis that length effects are due to peripheral factors, such as articulation. The German-English comparison is particularly meaningful because word length can be manipulated while holding orthography and phonology constant by using cognates (e.g., *land*, *bank*, *sand*, etc.). When identical words and nonwords are tested in both languages, the results show greater length effects in German compared with English (Ziegler et al., 2001). Given that articulatory and orthographic differences are controlled by using identical items across languages, it seems quite clear that length effect differences cannot be due to peripheral factors. Instead, this finding provides very compelling evidence for a serial mechanism or a mechanism that produces serial-like behavior. Moreover, spelling-sound dispersion in German is less than in English (Perry & Ziegler, 2002). Therefore, any account based on this concept must predict a smaller length effect in German than in English, which is the opposite of what was observed (Perry & Ziegler,

2002). Thus, even if some proportion of the length effect might be due to visual input or articulatory output factors, the greater serial effect in German than in English must reflect serial mechanisms beyond input and output processes.

Plaut (1999) offered an account of the length effect based on a simple recurrent network (Elman, 1990) that was trained to generate a sequence of phonemes as output in response to letter strings. The network was also trained to maintain a representation of its current position within the string and to use this signal to refixate a peripheral portion of input when it encountered difficulty in generating a pronunciation. Although this network is very different from the standard triangle model and it gives up the idea of purely parallel processing, one critical issue in this approach is the use of the number of fixations made by the network in pronouncing a word as a measure of naming latency. In particular, there is no empirical evidence that readers typically use more than one fixation when reading monosyllabic words. What seems more plausible to us is the hypothesis that serial processing depends on covert focusing of attention on each sublexical unit (single letter or letter cluster). Note that the operations of Plaut's network on the input could be redescribed in terms of attention shifts rather than fixations, in which case it would be very similar to CDP+.

Serial processing in CDP+ is explicitly linked to graphemic parsing, a process that is likely to involve *left-to-right shifts of spatial attention* over the letter string (also see Facoetti et al., 2006). Left-to-right processing implies spatial coding of the letters; thus, we assume that the letter level, although abstract, is spatially organized according to a word-centered coordinate system (Caramazza & Hillis, 1990; Mapelli et al., 1996).<sup>8</sup> The graphemes inside an attentional window spanning three letters are identified and inserted into the graphemic buffer (where they are kept active throughout the entire naming process). The window is then moved to a new position, graphemes are found, and so on until all the letters have been processed. The role of spatial attention is discussed at some length in a later section. Note, however, that even the other computational models must implicitly assume some attentional operations to achieve the level of representation on which the spelling-sound conversion mechanism operates. The triangle model uses a syllabically structured input based on grapheme units, whereas DRC scans the input string in a letter-by-letter fashion, and at any assembly cycle it performs a full search through the set of GPC rules.

#### *Accounting for Item-Level Variance*

It has been a common practice to evaluate models with regard to their qualitative fit to the data. However, in recent years, modelers started to evaluate and compare models with regard to quantitative fits as well (Coltheart et al., 2001; Spieler & Balota, 1997). This model evaluation and comparison strategy was made possible by the existence of a number of databases that were collected in large-scale experiments.

The results from our study showed that CDP+ accounted for much more variance in the large-scale databases of human naming latencies compared with the other models. Furthermore, CDP+ accounted for as much variance as the three factors mentioned in Spieler and Balota (1997; orthographic length, frequency, orthographic neighborhood). Thus, in this respect, CDP+ is greatly superior to the other models. It is important to note that, given that

the lexical route is essentially the same as that of the DRC, this difference in the amount of the variance accounted for must represent the different way in which sublexical processing is realized and the way in which it influences the final pronunciation.<sup>9</sup>

One general problem with the regression approach is that it somewhat penalizes the models. This is because the parameters that the model uses are always kept the same across all databases and all experiments. However, there is considerable variability due to differences in task or stimulus conditions, such as blocking effects, list effects, context effects, effects of stimulus degradation, and so forth (e.g., Stone & Van Orden, 1993; Visser & Besner, 2001). For example, in one task condition, frequency might play a bigger role than in another task condition, yet the model uses the same frequency scaling value for both. This means that the parameters are not optimized for each data set. This is very different from a large-scale regression analysis (e.g., Balota & Spieler, 1998), in which the regression weights are optimized for a given analysis. In our models, the parameters are chosen as if the entire set of words from all the experiments were put into a regression equation. One way to obtain higher correlations would be to allow parameter variation, as done, for instance, by Dell, Burger, and Svec (1997) in their model of spoken word production. However, given that CDP+ already accounts for a substantial portion of the variance, and given the potential problems associated with allowing parameter changes, we did not follow this strategy.

#### *The New List of Benchmark Effects*

On the basis of these discussions, we are now able to propose a new list of benchmark effects that should help modelers to develop and test the next generation of computational models of reading aloud. Note that we did not include studies that were not significant by items or whose robustness is still strongly debated. The benchmark effects appear in Table 4. All item sets and the corresponding item data can be downloaded at <http://ccnl.psy.unipd.it/CDP.html>.

#### *Credit Assignment*

Our componential modeling approach made it possible to investigate which parts of CDP+ were responsible for its improved performance in comparison with its predecessor and its competitors.

First, to investigate the efficiency of the orthographic buffer, we switched off the lexical route and compared CDP+ against Zorzi et al.'s (1998b) CDP model. Note that both CDP+ and CDP have

<sup>8</sup> Spatial attention can operate on any level of representation that is spatially organized. In the case of number processing, for example, the input level may not be explicitly spatial (e.g., a one-digit number), but spatial attention operates on the abstract semantic representation (i.e., a left-to-right oriented mental number line; see Zorzi, Priftis, & Umiltà, 2002).

<sup>9</sup> Note that the item correlations of DRC on large-scale databases can be improved by changing different parameters of the model. However, such an upgraded DRC would then need to be tested on the old as well as the new benchmark effects simulated in the present article (for an updated list, see Table 4).



Table 4  
New List of Benchmark Effects

Name of effect	Benchmark data set	Description	Triangle	DRC	CDP+
Frequency	Jared (2002, Experiment 2) Weekes (1997)	High-frequency words are faster/more accurate than low-frequency words.	+	+	+
Lexicality	McCann and Besner (1987) Weekes (1997)	Words are faster/more accurate than pseudowords.	+	+	+
Frequency × Regularity	Paap and Noel (1991) Jared (2002, Experiment 2)	Irregular words are slower/less accurate than regular words. Jared (2002) reported no interaction with frequency.	–	+	+
Word consistency	Jared (2002, Experiment 1)	Inconsistent words are slower/less accurate than consistent words. The size of the effect depends on the friend–enemy ratio.	+	–	+
Nonword consistency	Andrews and Scarratt (1998)	Nonword pronunciations show graded consistency effects; that is, people do not always use the most common grapheme–phoneme correspondences.	–	–	+
Length × Lexicality	Weekes (1997) Ziegler et al. (2001)	Nonword naming latencies increase linearly with each additional letter.	–	+	+
Position of irregularity	Rastle and Coltheart (1999)	The size of the regularity effect is bigger for words with first position irregularities (e.g., <i>chef</i> ) than for words with second- or third-position irregularities.	–	+	+
Body neighborhood	Ziegler et al. (2001)	Words with many body neighbors are faster/more accurate than words with few body neighbors.	–	–	+
Masked priming	Forster and Davis (1991)	Words preceded by an onset prime are faster/more accurate than words preceded by unrelated primes.	?	+	+
Pseudohomophone advantage	McCann and Besner (1987) Reynolds and Besner (2005)	Nonwords that sound like real words (e.g., <i>bloo</i> ) are faster/more accurate than orthographic controls.	+	+	+
Surface dyslexia	Patterson and Behrmann (1997)	Patient MP showed specific impairment of irregular word reading, which was modulated by the consistency ratio of the words.	+	–	+
Phonological dyslexia	Derouesné and Beauvois (1985) <sup>a</sup>	Patient LB showed specific impairment of nonword reading, which was reduced when nonwords were orthographically similar pseudohomophones.	+ <sup>b</sup>	+	+
Large-scale databases	Spieler and Balota (1997) Balota and Spieler (1998)	Naming latencies of the model were regressed onto the average naming latency of each item in large-scale databases containing thousands of items.	–	–	+

Note. DRC = dual-route cascaded model; CDP+ = new connectionist dual process model; + = success; – = failure; ? = not sure.

<sup>a</sup> Because the items were in French, Coltheart et al. (2001) created an English list for the simulations. <sup>b</sup> Harm and Seidenberg (1999) simulated not the patient LB but the patient MJ (Howard & Best, 1996).

an identical nonlexical phonological system except for the orthographic buffer. These models were then tested on one of the hardest nonword reading sets, the whammy nonwords (Rastle & Coltheart, 1998). Whereas CDP had an error rate of around 50%, CDP+ had an error rate of only 2.1% (i.e., one error). Clearly, the graphemic buffer tremendously increased the model's accuracy in nonword reading.

Second, we investigated the effect of serializing the sublexical route by comparing a serial version of CDP+ with a parallel version of the same model (parallel CDP+). The results were very clear. The serial version of the model was able to simulate length effects in nonword reading (Weekes, 1997; Ziegler et al., 2001), whereas the parallel version of the model completely lost the ability to capture these effects. We also compared the two versions of CDP+ on large-scale databases of real word reading. The results show that serializing the sublexical route, besides being

crucial for simulating the length effect on nonwords, does significantly improve the correlation with word naming latencies.

Third, we investigated whether the superiority of CDP+ with respect to simulating the consistency effects reported by Jared (2002) could be attributed, at least in part, to the orthographic properties of the inconsistent words (e.g., an orthographic neighborhood confound, see Coltheart et al., 2001). To examine the contribution of the interactive recurrent processing in the lexical route, we turned CDP+ into a purely feedforward model, and we completely eliminated the activation of orthographic neighbors (feedforward CDP+). This allowed us to isolate the source of the consistency effects and to firmly establish that they were entirely produced by the sublexical network. In the simulation with the feedforward CDP+, we also examined to what extent the fully implemented lexical route in CDP+ contributed to explaining item-level variance on the large-scale databases. The feedforward

CDP+ accounted for a proportion of variance almost identical to that accounted for by CDP+. This shows that a fully implemented lexical route does not add much over and above the effect of a frequency-weighted activation of lexical phonology. However, it should be noted that not all phenomena in word naming can be explained by a feedforward model; in particular, our simulation of the pseudohomophone advantage in nonword naming (McCann & Besner, 1987) relies on the existence of feedback connections in the model.

Fourth, we examined how much of the item-level variance in large-scale databases could be accounted for by either lexical or nonlexical parts of the model alone. The results showed that neither the lexical nor the sublexical parts of the model by themselves were able to pass the Spieler and Balota (1997) test, according to which models should be able to perform at least as well on item data as a simple correlation with well-established factors (i.e., orthographic frequency, orthographic neighborhood, and orthographic word length). The sublexical part of the model correlated quite poorly with the data sets, suggesting that, by itself, it does not account for the human performance. The lexical part of the model obtained a stronger correlation, but it did not fare much beyond the simple effect of frequency. These results therefore show that the singular parts of the model are inadequate at explaining the data, even at a superficial level.

#### *On the Role of Spatial Attention in Reading Aloud*

A number of authors have claimed that spatial attention is a necessary condition for word recognition to begin (e.g., Lachter, Forster, & Ruthruff, 2004; McCann, Folk, & Johnston, 1992; Stolz & McCann, 2000), but this assumption has not been incorporated in computational models of reading aloud. Our model is not different in this respect, because it does not contain attention mechanisms operating at the input (i.e., feature) level; that is, processing initiates as soon as a stimulus is visually presented. Our contention, instead, is that a specific form of spatial attention, that is *focused spatial attention*, is involved in the assembly of phonology from print (but not in lexical access) and more specifically in the graphemic parsing process.

It is well known that focused spatial attention enhances visual processing not only in terms of processing speed but also in terms of improved sensitivity (i.e., spatial resolution), reduced interactions with near stimuli (spatial and temporal masking), and elimination of illusory conjunctions (e.g., Braun, 2001; Carrasco & McElree, 2001). Therefore, it is likely to be extremely important for graphemic parsing and segmentation (and in turn for nonword reading). Indeed, some studies suggest that focused visuospatial attention is more important for nonword reading than for word reading. For instance, Sieroff and Posner (1988) used spatial cuing to manipulate focused visual attention during reading. Participants made more errors in reporting the letters from the unattended side of nonwords compared with words (also see Auclair & Sieroff, 2002). Moreover, patients with hemispatial neglect made more errors on the contralesional side of nonwords compared with words (e.g., Sieroff, Pollatsek, & Posner, 1988). Crucially, patients with severe neglect dyslexia showed preserved lexical–semantic access in reading (Ladavas, Shallice, & Zanella, 1997; Ladavas, Umiltà, & Mapelli, 1997), suggesting an interaction between the attentional system and the different reading routes. That is, the

lexical–semantic route is much less affected by neglect than the phonological route because the latter requires a narrower attentional focus to control the sequence of parts of the input string to be admitted to the spelling-to-sound translation process (Ladavas, Shallice, & Zanella, 1997). Most notably, focused spatial attention has been specifically linked to nonword reading performance in developmental dyslexia (Facoetti et al., 2006). Indeed, an RT index of attention orienting accounted for a large proportion of variance in the nonword reading accuracy of dyslexic children, even after partialing out the effects of age, IQ, and phonological skills.

Important support for our proposal is also provided by a recent study of Reynolds and Besner (2006) that used the psychological refractory period paradigm to investigate the attentional demands posed by reading aloud. They found that processing up to the activation of the orthographic lexicon (thus including activation of feature and letter levels) did not require attentional resources. In contrast, phonological assembly was found to be attention demanding. As Reynolds and Besner put it, “Unless there is some as yet unspecified process occurring subsequent to letter identification, but prior to assembled phonology, the observation . . . suggests that assembled phonological recoding uses central attention during reading aloud” (p. 1309). Notably, such an “unspecified process” does indeed exist in CDP+, and it corresponds to an explicit component of the model: the graphemic buffer. Graphemic parsing is attention demanding because it requires shifts of focused spatial attention. In contrast, phonological assembly per se is less likely to be attention demanding because it requires only spread of activation from grapheme nodes to phoneme nodes.

In summary, although the empirical data are far too sparse to allow principled assumptions regarding the precise nature of the attentional operations related to the print-to-sound conversion, our proposal of an explicit link between serial processing in the sublexical route and focused spatial attention points to an area where there is still a lot of research to be done.

#### *On the Role of Semantics in Reading Aloud*

Although the computational models discussed in this article may vary to a large degree in terms of their specific architectures and processing assumptions, the emerging consensus view is that the interaction between two different sources of phonological information must be assumed to account for both skilled reading aloud and acquired dyslexia (Zorzi, 2005). After the presentation of a printed word, phonology is retrieved through a lexical–semantic pathway (or network) as well as assembled (or activated) through a spelling–sound mapping process. One major controversy, however, concerns the role of semantics. At the heart of this debate is the interpretation of neuropsychological data from patients with acquired surface dyslexia and/or semantic disorders (for reviews, see Coltheart, 2006; Zorzi, 2005). The conflicting theoretical views are reflected in the way surface dyslexia is accounted for in DRC as opposed to the triangle model (see the *Consistency Effects in Surface Dyslexia* section for a discussion).

Classic dual-route models such as DRC assume that the lexical route can be further divided into two processing routes (note that this distinction means that the dual-route model is in fact a three-route model). That is, phonological word forms can be activated directly from the orthographic lexicon (direct lexical route) or

through the mediation of word meanings (lexical–semantic route). The distinction between a lexical–semantic route and a direct lexical (i.e., nonsemantic) route was first suggested by Schwartz, Saffran, and Marin (1980) in their case study of the acquired dyslexic patient WLP, who could read aloud words (including exception words) that she could not understand. However, following the seminal work of Patterson and Hodges (1992), a series of studies reported a consistent pattern of association between semantic dementia and surface dyslexia (e.g., Funnell, 1996; Graham, Hodges, & Patterson, 1994). The hypothesis that correct exception word reading is dependent on semantic representations (Patterson & Hodges, 1992) was later incorporated into the triangle model, and it formed the basis of Plaut et al.'s (1996) account of surface dyslexia (see also Woollams, Lambon-Ralph, Plaut, & Patterson, in press). According to this model, differences among patients not only reflect a different severity of the lesion, but in particular reflect their different premorbid reading competence, that is, the degree of redistribution of labor between semantic and phonological pathways. However, this hypothesis is challenged by cases of patients showing the corresponding dissociation (i.e., intact reading in the presence of semantic deficits). Indeed, the pattern of lexical nonsemantic reading shown by patient WLP is not an exceptional case, and there are now numerous patient case reports that support the independence of semantic and phonological processing (e.g., Blazely, Coltheart, & Casey, 2005; Cipolotti & Warrington, 1995; Gerhand, 2001; Lambon-Ralph, Ellis, & Franklin, 1995).

One way to reconcile these data with the triangle model is to treat the dissociation between surface dyslexia and semantic dementia as extremes that still fall within the full distribution of cases of semantic impairment (Woollams et al., in press). However, one alternative explanation for the association between semantic dementia and surface dyslexia is that it reflects pathological involvement of functionally and anatomically closely related brain regions (see also Cipolotti & Warrington, 1995). In other words, this specific form of cortical degeneration would lead to semantic impairments but also (and perhaps most often) to the disruption of lexical processing (both orthographic and phonological). This hypothesis would seem to gain support from a recent study that compared the reading performance of patients with different types of dementia (Noble, Glosser, & Grossman, 2000). Despite the presence of a semantic impairment, patients with Alzheimer's disease, frontotemporal dementia, and progressive nonfluent aphasia did not show a pattern of reading difficulty consistent with surface dyslexia; only those with semantic dementia showed the predicted pattern of reading impairment. These data would seem to pose a challenge to the triangle model because any form of semantic impairment in the model should produce a surface dyslexic pattern (unless simulations can prove otherwise).

A related issue is whether semantics contributes to word naming in skilled readers. Strain, Patterson, and Seidenberg (1995) demonstrated that a semantic variable, imageability, can have an impact on naming of isolated words. However, the imageability variable affected only the naming of low-frequency exception words (i.e., the words that usually yield the longest naming latencies). This result would not be problematic for dual-route models: Processing of low-frequency words in a model such as DRC is sufficiently slow to allow semantic effects to emerge from processing in the lexical–semantic route. To complicate the picture,

however, Balota et al. (2004) obtained a significant imageability effect in their large-scale study of naming, whereas the interaction with consistency and frequency reported by Strain et al. did not reach significance. Moreover, Baayen et al. (2006) reanalyzed Balota et al.'s data with regression techniques that are better suited for dealing with collinearity and nonlinearity and found even weaker (i.e., nonsignificant) effects of semantic variables in the naming task.

In summary, it appears that word meaning does not have an important contribution in written word naming. Indeed, it has been argued that reading is fundamentally phonological, because even tasks such as semantic categorization, in which the activation of phonology is, in principle, irrelevant, are strongly affected by the phonological characteristics of the stimuli (see Frost, 1998, for a comprehensive review). In this regard, it should be noted that phonological assembly in CDP+ is fast enough to be consistent with fast phonology theories of reading (see Berent & Perfetti, 1995; Frost, 1998; Rayner, Pollatsek, & Binder, 1998; Van Orden, Pennington, & Stone, 1990). This is a departure from classic dual-route models, where the interaction between lexical and assembled phonology is best characterized as a horse race (Paap & Noel, 1991). Classic dual-route models (e.g., Coltheart, 1978; Meyer, Schvaneveldt, & Ruddy, 1974) give a predominant role to the visual route, because the assembly of phonology is believed to be too slow to affect lexical access.

#### *Limitations of the Model and Future Directions*

The current model is a hybrid of a number of other models, but one aspect that may have detracted from its performance is the lexical route, which is basically the interactive activation model of McClelland and Rumelhart (1981). As discussed earlier, our choice was primarily motivated by the nested modeling strategy. Accordingly, CDP+ incorporates a model of the lexical route that has been used to account for a large amount of empirical data regarding perceptual identification and lexical decision tasks (see Grainger & Jacobs, 1996, for a review), but it may also inherit some of its problems. For example, the interactive activation model has been criticized for its failure to account for letter transposition or body neighborhood effects in lexical decision (e.g., Andrews, 1996; Ziegler & Perry, 1998).

However, there is a good deal of recent work examining ways to model lexical access in a more plausible fashion (e.g., Davis & Bowers, 2004; Houghton & Zorzi, 2003; Shillcock, Ellison, & Monaghan, 2000). If some of these models turn out to offer a better account of the data than the interactive activation model, then there is no reason to think that the lexical route of CDP+ could not be replaced by them. This possibility is facilitated by our finding that the contribution of the lexical route to the overall performance of the model is practically limited to the provision of frequency-weighted lexical phonology (see simulations with the feedforward CDP+). If any of these new lexical routes allow the model to explain more variance than the current model, then it is likely that the strength of the correlation of the model with human data would be much higher than the factors suggested by Spieler and Balota (1997), which would certainly mark a milestone in the modeling of reading aloud.

A current limitation of the model is the absence of learning in the lexical route. The model simulates the lexical route as a localist

interactive-activation network, in which each known word is represented by a dedicated node. It is important to note that such representations can be formed in neural networks with, for instance, competitive learning algorithms (Grossberg, 1980; Kohonen, 1984). In Houghton and Zorzi's (2003) model of spelling, each (localist) word node in the orthographic lexicon had an excitatory feedback loop onto itself, giving it the ability to support its own activation. This is typical of competitive networks, and the strength of this feedback depends on a parameter (the feedback weight) that is the same for all nodes in the network. However, Houghton and Zorzi allowed this feedback weight to vary as a function of word frequency—the more frequent a word was, the stronger the feedback weight. In this way, word frequency was modeled as a dynamic effect rather than as a specific threshold for each word node (as in standard interactive activation models, including the present one). The modulation of a unit's feedback weight could easily be achieved as part of a competitive learning algorithm: If each time a node is activated, its feedback loop is strengthened (thus enabling it to fare better in the competition for activation, which is essential to such algorithms), then more frequent words would be more easily activated (Houghton & Zorzi, 2003). Thus, it should in principle be possible to add learning to the lexical route.

It should also be possible to capture the interaction between lexical and nonlexical parts of the model during learning, as was done by Zorzi et al. (1998b). In one of the simulations, a three-layer feedforward network was trained on a monosyllabic word set with learning taking place in both direct (input–output) and mediated (hidden unit) pathways at the same time. This version of the model was capable of learning the whole training set, including the exception words. The results showed that the direct phonological route, when studied in isolation, still behaved like a spelling-to-sound conversion mechanism and did not acquire lexical properties. In contrast, the hidden unit pathway behaved more like a (distributed) lexical route. When the model was trained with relatively few hidden units (restricting its capacity to represent the training set), the hidden units appeared to dedicate themselves to the exception words by correcting the output of the direct spelling–sound mapping, which, left to itself, would give regularized pronunciations to them. These results provide evidence that a network with lexical and sublexical routes that interact when learning tends to self-modularize so that the regular productive spelling–sound correspondences are learned by the direct phonological route (the TLA network of the current model).

One potential problem with using a distributed lexical route is that it might be difficult to fully account for lexical decision performance (see Borowsky & Besner, 2006, and Plaut & Booth, 2006, for opposing views), and effects such as those reported by Visser and Besner (2001) may also be difficult to capture. However, a distributed model would work almost like a localist one if it had an associative memory with strong attractor dynamics (e.g., Ackley et al., 1985). If a nonword were presented and the input units remained clamped to the nonword, such a network would not settle because the input would not match any stable states (i.e., the learned words). In contrast, if the input units were allowed to change their state shortly after the presentation of the stimulus, the network would settle to the closest attractor (i.e., the closest word neighbor); however, the change of the input units' states could be easily detected. This suggests that there are at least two different

ways to distinguish between words and nonwords. It might be argued that Plaut et al. (1996) used an attractor network for modeling the orthography-to-phonology route and that their network was still able to generalize to nonwords. Plaut et al. pointed out that the good generalization performance was dependent on *componential attractors* developed by the network during learning, that is, attractors for sublexical components rather than for the whole word. However, O'Reilly (2001) has shown that fixed-point recurrent back-propagation develops very weak feedback connections in comparison with other algorithms based on contrastive Hebbian learning. This, in addition to the fact that Plaut et al.'s networks were constrained to settle very rapidly, minimizes the extent to which those networks can be considered interactive; their good generalization performance can thus be attributed to the lack of interactivity rather than to the existence of componential attractors (see O'Reilly, 2001, for further discussion). Further work should investigate how learning in fully interactive networks could be exploited to model lexical access.

One additional issue related to learning concerns the model's orthographic representation in the graphemic buffer. The model is supplied at the outset with a syllabically aligned graphemic representation, which is of great benefit to its learning of the spelling–sound mapping. Although it is known that children have developed syllabic phonology by the time they start to learn to read and write (for a review, see Ziegler & Goswami, 2005), their orthographic representations must develop as a part of this learning (Goswami & Ziegler, 2006). Hence, a plausible developmental model of reading cannot start out with complex grapheme nodes and orthographic syllable structure. Thus, more work is clearly needed to make this a developmentally more plausible model. Ideally, such extensions should be carried out in the spirit of nested modeling and strong inference.

One potential criticism of the model is that it has a large number of free parameters and a complex grapheme buffer. However, it is worth noting that the grapheme buffer is actually less complex than the rule system of the DRC (Coltheart et al., 2001), even though it has a very similar function. Moreover, most of the parameters in the model are related to the lexical route (i.e., the interactive-activation model). One way to greatly reduce the number of parameters in the model would be to use a lexical system, such as that in Zorzi et al. (1998b), in which lexical activation is simply simulated as a frequency function controlled by a single parameter (see also the semantic simulations of Plaut et al., 1996). This would get rid of many parameters but would leave many of the results essentially the same, as clearly shown by the simulation with the feedforward version of the model (feedforward CDP+).

Finally, CDP+ is limited to monosyllabic words, as are the other computational models we have discussed. This reflects the fact that most of the empirical evidence on written word naming comes from research conducted with monosyllabic words. Although most of the words people read are monosyllabic according to a token count, the majority of the words in the lexicon are polysyllabic according to a type count. Some studies have started to explore whether the results on monosyllabic words generalize to polysyllabic words (e.g., Chateau & Jared, 2003; Jared & Seidenberg, 1990). Extending the current models to polysyllabic words requires the consideration of several issues that are not relevant for monosyllables, such as the assignment of stress (e.g., Rastle & Coltheart, 2000), the role of the syllable and the possible ambigu-



ities in segmenting letter strings into larger spelling units. Thus, one important issue for future research is to design appropriate coding schemes for representing the orthography and phonology of polysyllabic words and to assess which orthographic segments become relevant when a simple statistical learning mechanism (such as our phonological assembly network) tries to learn the mapping between spelling and sound.

### Conclusion

The goal of the present research was to design a new model by building on the strengths of the previous models and eliminating their weaknesses. This strategy—*nested modeling*—is commonly used in other areas of science but has rarely been used as a guiding principle in the modeling of cognitive functions (for a notable exception, see the work by Shiffrin and colleagues; Shiffrin, 2003, provided an overview of his 30-year research program on modeling memory).

In the current work, nested modeling was combined with *strong inference testing*. That is, we tested the main alternative models (whether precursors of the new one or not) on critical data sets and on large-scale databases to compare their *descriptive adequacy*, both at the qualitative and quantitative level. Given the fast-spreading use of computational modeling, the approach of testing and comparing competing models is bound to become the standard in cognitive psychology. Empirical studies that aim at adjudicating between competing models need to actually test the competitors (i.e., run the stimuli through the models), because inferring the behavior of a complex (and often nonlinear) model solely from its architectural description and processing assumption is a difficult and potentially misleading enterprise (see Zorzi, 2000). Examples of strong inference testing have been making their way into the area of reading aloud, and these studies have provided very useful insights (e.g., Besner & Roberts, 2003; Jared, 2002; Reynolds & Besner, 2005; Treiman et al., 2003).

Even more important, however, is the use of strong inference testing in the context of model development. Contemporary models of cognition are computationally explicit with regard to their assumptions about architecture, representations, learning, and so forth. These assumptions might be of high theoretical relevance (e.g., the issue of discreteness vs. interactivity in spoken word production; Rapp & Goldrick, 2000), and they might determine the success or failure of a model (e.g., the issue of the format of number semantics in numerical cognition; Zorzi, Stoianov, & Umiltà, 2005). Strong inference testing requires that such alternatives are tested against one another. We believe that nested modeling and strong inference testing are fundamental tools for improving the understanding of the computations underlying human cognition.

### References

- Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science*, 9, 147–169.
- Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 802–814.
- Andrews, S. (1992). Neighbourhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 234–254.
- Andrews, S. (1996). Lexical retrieval and selection processes: Effects of transposed-letter confusability. *Journal of Memory and Language*, 35, 775–800.
- Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin and Review*, 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, 1052–1086.
- Andrews, S., Woollams, A., & Bond, R. (2005). Spelling–sound typicality only affects words with digraphs: Further qualifications to the generality of the regularity effect on word naming. *Journal of Memory and Language*, 53, 567–593.
- Auclair, L., & Sieroff, E. (2002). Attentional cueing effect in the identification of words and pseudowords of different length. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 55(A), 445–463.
- Baayen, R. H., Feldman, L. B., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, 55, 290–313.
- Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database (CD-ROM)*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- Bachoud-Levi, A. C., Dupoux, E., Cohen, L., & Mehler, J. (1998). Where is the length effect? A cross-linguistic study of speech production. *Journal of Memory and Language*, 39, 331–346.
- Balota, D. A., Cortese, M. J., Hutchison, K. A., Neely, J. H., Nelson, D., Simpson, G. B., & Treiman, R. (2002). *The English Lexicon Project: A Web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords*. Available from <http://lexicon.wustl.edu/>
- Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, 133, 283–316.
- Balota, D. A., & Spieler, D. H. (1998). The utility of item-level analysis in model evaluation: A reply to Seidenberg and Plaut. *Psychological Science*, 9, 238–240.
- Behrmann, M., & Bub, D. (1992). Surface dyslexia and dysgraphia: Dual routes, single lexicon. *Cognitive Neuropsychology*, 9, 209–251.
- Berent, I., & Perfetti, C. A. (1995). A rose is a reez: The two-cycles model of phonology assembly in reading English. *Psychological Review*, 102, 146–184.
- Besner, D. (1999). Basic processes in reading: Multiple routines in localist and connectionist models. In R. M. Klein & P. McMullen (Eds.), *Converging methods for understanding reading and dyslexia* (pp. 413–458). Cambridge, MA: MIT Press.
- Besner, D., & Roberts, M. A. (2003). Reading nonwords aloud: Results requiring change in the dual route cascaded model. *Psychonomic Bulletin and Review*, 20, 398–404.
- Besner, D., Twilley, L., McCann, R. S., & Seergobin, K. (1990). On the association between connectionism and data: Are a few words necessary? *Psychological Review*, 97, 432–446.
- Blazely, A. M., Coltheart, M., & Casey, B. J. (2005). Semantic impairment with and without surface dyslexia: Implications for models of reading. *Cognitive Neuropsychology*, 22, 695–717.
- Borowsky, R., & Besner, D. (2006). Parallel distributed processing and lexical-semantic effects in visual word recognition: Are a few stages necessary? *Psychological Review*, 113, 181–195.
- Braun, J. (2001). Visual attention: Light enters the jungle. *Current Biology*, 12, 599–601.
- Brown, G. D. A. (1987). Resolving inconsistency: A computational model of word naming. *Journal of Memory and Language*, 26, 1–23.
- Bub, D., Cancelliere, A., & Kertesz, A. (1985). Whole-word and analytic translation of spelling to sound in a non-semantic reader. In K. E.

- Patterson, J. C. Marshall, & M. Coltheart (Eds.), *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (pp. 15–34). London: Erlbaum.
- Caramazza, A., Capasso, R., & Miceli, G. (1996). The role of the graphemic buffer in reading. *Cognitive Neuropsychology*, *13*, 673–698.
- Caramazza, A., & Hillis, A. E. (1990). Spatial representation of words in the brain implied by studies of a unilateral neglect patient. *Nature*, *346*, 267–269.
- Caramazza, A., & Miceli, G. (1990). The structure of graphemic representations. *Cognition*, *37*, 243–297.
- Caramazza, A., Miceli, G., Villa, G., & Romani, C. (1987). The role of the graphemic buffer in spelling: Evidence from a case of acquired dysgraphia. *Cognition*, *26*, 59–85.
- Carrasco, M., & McElree, B. (2001). Covert attention accelerates the rate of visual information processing. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 5363–5367.
- Chateau, D., & Jared, D. (2003). Spelling–sound consistency effects in disyllabic word naming. *Journal of Memory and Language*, *48*, 255–280.
- Cipolotti, L., & Warrington, E. K. (1995). Semantic memory and reading abilities: A case report. *Journal of the International Neuropsychological Society*, *1*, 104–110.
- Clark-Carter, D. (2004). *Quantitative psychological research. A student's handbook*. London: Psychology Press.
- Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 151–216). London: Academic Press.
- Coltheart, M. (2006). Acquired dyslexias and the computational modelling of reading. *Cognitive Neuropsychology*, *23*, 96–109.
- Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, *100*, 589–608.
- 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*, 1197–1211.
- 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*, 204–256.
- Coltheart, M., Woollams, A., Kinoshita, S., & Perry, C. (1999). A position-sensitive Stroop effect: Further evidence for a left-to-right component in print-to-speech. *Psychonomic Bulletin and Review*, *6*, 456–463.
- Cortese, M. J. (1998). Revisiting serial position effects in reading. *Journal of Memory and Language*, *39*, 652–665.
- Cotelli, M., Abutalebi, J., Zorzi, M., & Cappa, S. F. (2003). Vowels in the buffer: A case study of acquired dysgraphia with selective vowel substitutions. *Cognitive Neuropsychology*, *20*, 99–114.
- Crick, F. H. C. (1989). The recent excitement about neural networks. *Nature*, *337*, 129–132.
- Cubelli, R. (1991). A selective deficit for writing vowels in acquired dysgraphia. *Nature*, *353*, 209–210.
- Davis, C. J., & Bowers, J. S. (2004). What do letter migration errors reveal about letter position coding in visual word recognition? *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 923–941.
- Dell, G. S. (1986). A spreading activation theory of retrieval in language production. *Psychological Review*, *93*, 283–321.
- Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, *104*, 123–147.
- Denes, F., Cipolotti, L., & Zorzi, M. (1999). Acquired dyslexias and dysgraphias. In G. Denes & L. Pizzamiglio (Eds.), *Handbook of clinical and experimental neuropsychology* (pp. 289–317). Hove, England: Psychology Press.
- Derouesn e, J., & Beauvois, M. F. (1985). The “phonemic” stage in the non-lexical reading process: Evidence from a case of phonological alexia. In K. E. Patterson, J. C. Marshall, & M. Coltheart (Eds.), *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (pp. 399–458). London: Erlbaum.
- Ellis, A. (1988). Normal writing processes and peripheral acquired dysgraphias. *Language and Cognitive Processes*, *3*, 99–127.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.
- Facoetti, A., Zorzi, M., Cestnick, L., Lorusso, M. L., Molteni, M., Paganoni, P., et al. (2006). The relationship between visuo-spatial attention and nonword reading in developmental dyslexia. *Cognitive Neuropsychology*, *23*, 841–855.
- Ferrand, L., & Grainger, J. (1992). Phonology and orthography in visual word recognition: Evidence from masked nonword priming. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *45(A)*, 353–372.
- Ferrand, L., & Grainger, J. (1993). The time course of orthographic and phonological code activation in the early phases of visual word recognition. *Bulletin of the Psychonomic Society*, *31*, 119–122.
- Ferrand, L., & Grainger, J. (1994). Effects of orthography are independent of phonology in masked form priming. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *47(A)*, 365–382.
- Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*, 627–635.
- Forster, K. I., & Davis, C. (1991). The density constraint of form-priming in the naming task: Interference effects from a masked prime. *Journal of Memory and Language*, *30*, 1–25.
- Forster, K. I., & Taft, M. (1994). Bodies, antibodies, and neighborhood-density effects in masked form priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 844–863.
- Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin*, *123*, 71–99.
- Frost, R., Ahissar, M., Gotesman, R., & Tayeb, S. (2003). Are phonological effects fragile? The effect of luminance and exposure duration on form priming and phonological priming. *Journal of Memory and Language*, *48*, 346–378.
- Funnell, E. (1996). Response biases in oral reading: An account of the co-occurrence of surface dyslexia and semantic dementia. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *49(A)*, 417–446.
- Gerhand, S. (2001). Routes to reading: A report of a non-semantic reader with equivalent performance on regular and exception words. *Neuropsychologia*, *39*, 1473–1484.
- Gluck, M. A., & Bower, G. H. (1988a). Evaluating an adaptive network model of human learning. *Journal of Memory and Language*, *27*, 166–195.
- Gluck, M. A., & Bower, G. H. (1988b). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, *117*, 227–247.
- Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 674–691.
- Goswami, U., & Ziegler, J. C. (2006). A developmental perspective on the neural code for written words. *Trends in Cognitive Sciences*, *10*, 142–143.
- Graham, K. S., Hodges, J. R., & Patterson, K. (1994). The relationship between comprehension and oral reading in progressive fluent aphasia. *Neuropsychologia*, *32*, 299–316.
- Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, *103*, 518–565.
- Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, *87*, 1–51.

- Hanley, J. R., & Kay, J. (1998). Does the graphemic buffer play a role in reading? *Cognitive Neuropsychology*, *15*, 313–318.
- Hanley, J. R., & McDonnell, V. (1997). Are reading and spelling phonologically mediated? Evidence from a patient with a speech production impairment. *Cognitive Neuropsychology*, *14*, 3–33.
- Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, *106*, 491–528.
- Harm, W. M., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*, 662–720.
- Houghton, G., & Zorzi, M. (1998). A model of the sound–spelling mapping in English and its role in word and nonword spelling. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the twentieth annual conference of the Cognitive Science Society* (pp. 490–495). Mahwah, NJ: Erlbaum.
- Houghton, G., & Zorzi, M. (2003). Normal and impaired spelling in a connectionist dual-route architecture. *Cognitive Neuropsychology*, *20*, 115–162.
- Howard, D., & Best, W. (1996). Developmental phonological dyslexia: Real world reading can be completely normal. *Cognitive Neuropsychology*, *13*, 887–934.
- Huey, E. B. (1908). *The psychology and pedagogy of reading*. Cambridge, MA: MIT Press. (Original work published 1908)
- Hutzler, F., Ziegler, J. C., Perry, C., Wimmer, H., & Zorzi, M. (2004). Do current connectionist learning models account for reading development in different languages? *Cognition*, *91*, 273–296.
- Jacobs, A. M., & Grainger, J. (1994). Models of visual word recognition: Sampling the state of the art. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1311–1334.
- Jared, D. (1997). Spelling–sound consistency affects the naming of high-frequency words. *Journal of Memory and Language*, *36*, 505–529.
- Jared, D. (2002). Spelling–sound consistency and regularity effects in word naming. *Journal of Memory and Language*, *46*, 723–750.
- Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language*, *29*, 687–715.
- Jared, D., & Seidenberg, M. S. (1990). Naming multisyllabic words. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 92–105.
- Jónsdóttir, M. K., Shallice, T., & Wise, R. (1996). Phonological mediation and the graphemic buffer disorder in spelling: Cross-language differences? *Cognition*, *59*, 169–197.
- Kohonen, T. (1984). *Self-organization and associative memory*. New York: Springer.
- Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after Broadbent (1958): Still no identification without attention. *Psychological Review*, *111*, 880–913.
- Ladavas, E., Shallice, T., & Zanella, M. T. (1997). Preserved semantic access in neglect dyslexia. *Neuropsychologia*, *35*, 257–270.
- Ladavas, E., Umiltà, C., & Mapelli, D. (1997). Lexical and semantic processing in the absence of word reading: Evidence from neglect dyslexia. *Neuropsychologia*, *35*, 1075–1085.
- Lambon-Ralph, M., Ellis, A. W., & Franklin, S. (1995). Semantic loss without surface dyslexia. *Neurocase*, *1*, 363–369.
- Lange, M., & Content, A. (2000, November 16–19). *Grapheme complexity and length effects in visual word recognition*. Poster presented at the 41st meeting of the Psychonomic Society, New Orleans, LA.
- Lesch, M. F., & Pollatsek, A. (1993). Automatic access of semantic information by phonological codes in visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 285–294.
- Mapelli, D., Umiltà, C., Nicoletti, R., Fanini, A., & Capezzani, L. (1996). Prelexical spatial representation. *Cognitive Neuropsychology*, *13*, 229–255.
- Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R., & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology*, *29*, 189–256.
- Martensen, H., Maris, E., Dijkstra, T. (2003). Phonological ambiguity and context sensitivity: On sublexical clustering in visual word recognition. *Journal of Memory and Language*, *49*, 375–395.
- McCann, R., & 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*, 14–24.
- McCann, R. S., Folk, C. L., & Johnston, J. C. (1992). The role of spatial attention in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 1015–1029.
- McCarthy, R. A., & Warrington, E. K. (1986). Phonological reading: Phenomena and paradoxes. *Cortex*, *22*, 359–380.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: 1. An account of the basic findings. *Psychological Review*, *88*, 357–407.
- Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1974). Functions of graphemic and phonemic codes in visual-word recognition. *Memory & Cognition*, *2*, 309–321.
- Montant, M., & Ziegler, J. C. (2001). Can orthographic rimes facilitate naming? *Psychonomic Bulletin and Review*, *8*, 351–356.
- Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*, 165–178.
- 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*, 799–810.
- Murre, J. M. J., Phaf, R. H., & Wolters, G. (1992). CALM: Categorizing and learning module. *Neural Networks*, *5*, 55–82.
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*, 226–254.
- Noble, K., Glosser, G., & Grossman, M. (2000). Oral reading in dementia. *Brain & Language*, *74*, 48–69.
- Norris, D. (1994). A quantitative multiple-levels model of reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1212–1232.
- O'Reilly, R. C. (1998). Six principles for biologically based computational models of cortical cognition. *Trends in Cognitive Science*, *2*, 455–462.
- O'Reilly, R. C. (2001). Generalization in interactive networks: The benefits of inhibitory competition and Hebbian learning. *Neural Computation*, *13*, 1199–1242.
- Paap, K. R., & Noel, R. W. (1991). Dual route models of print to sound: Still a good horse race. *Psychological Research*, *53*, 13–24.
- Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning out of the lab: The case of orthographic regularities. *Journal of Experimental Psychology: General*, *130*, 401–426.
- Patterson, K. (1990). Alexia and neural nets. *Japanese Journal of Neuropsychology*, *6*, 90–99.
- Patterson, K., & Behrmann, M. (1997). Frequency and consistency effects in a pure surface dyslexic patient. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 1217–1231.
- Patterson, K. E., & Hodges, J. R. (1992). Deterioration of word meaning: Implications for reading. *Neuropsychologia*, *12*, 1025–1040.
- Peereman, R., & Content, A. (1997). Orthographic and phonological neighborhoods in naming: Not all neighbors are equally influential in orthographic space. *Journal of Memory and Language*, *37*, 382–410.
- Perry, C., & Ziegler, J. C. (2002). Cross-language computational investi-



- gation of the length effect in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 990–1001.
- Perry, C., & Ziegler, J. C. (2004). Beyond the two-strategy model of skilled spelling: Effects of consistency, grain size, and orthographic redundancy. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 57(A), 325–356.
- Petersen, S. E., Fox, P. T., Snyder, A. Z., & Raichle, M. E. (1990, August 31). Activation of extrastriate and frontal cortical areas by visual words and word-like stimuli. *Science*, 249, 1041–1044.
- Platt, J. R. (1964, October 16). Strong inference testing. *Science*, 146, 347–353.
- Plaut, D. C. (1999). A connectionist approach to word reading and acquired dyslexia: Extension to sequential processing. *Cognitive Science*, 23, 543–568.
- 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, 786–823.
- Plaut, D. C., & Booth, J. R. (2006). More modeling but still no stages: Reply to Borowsky and Besner. *Psychological Review*, 113, 196–200.
- 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, 56–115.
- Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, 107, 460–499.
- Rastle, K., & Coltheart, M. (1998). Whammies and double whammies: The effect of length on nonword reading. *Psychonomic Bulletin and Review*, 5, 277–282.
- Rastle, K., & Coltheart, M. (1999). Serial and strategic effects in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 482–503.
- Rastle, K., & Coltheart, M. (2000). Lexical and nonlexical print-to-sound translation of disyllabic words and nonwords. *Journal of Memory and Language*, 42, 342–364.
- Rayner, K., Pollatsek, A., & Binder, K. S. (1998). Phonological codes and eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 476–497.
- Rey, A., & Schiller, N. O. (2005). Graphemic complexity and multiple print-to-sound associations in visual word recognition. *Memory & Cognition*, 33, 76–85.
- Rey, A., Ziegler, J. C., & Jacobs, A. M. (2000). Graphemes are perceptual reading units. *Cognition*, 75, B1–B12.
- Reynolds, M., & 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, 13–31.
- Reynolds, M., & Besner, D. (2005). Basic processes in reading: A critical review of pseudohomophone effects in reading aloud and a new computational account. *Psychonomic Bulletin and Review*, 12, 622–646.
- Reynolds, M., & Besner, D. (2006). Reading aloud is not automatic: Processing capacity is required to generate a phonological code from print. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 1303–1323.
- 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 and Review*, 10, 405–414.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representation by error propagation. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Volume 1. Foundations* (pp. 318–362). Cambridge, MA: MIT 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, 60–94.
- Schwartz, M. F., Saffran, E. M., & Marin, O. S. M. (1980). Fractionating the reading process in dementia: Evidence for word-specific print-to-sound associations. In M. Coltheart, K. E. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 259–269). London: Routledge & Kegan Paul.
- Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, 23, 569–588.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523–568.
- Seidenberg, M. S., Petersen, A., MacDonald, M. C., & Plaut, D. C. (1996). Pseudohomophone effects and models of word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 48–72.
- Seidenberg, M. S., & Plaut, D. C. (1998). Evaluating word-reading models at the item level: Matching the grain of theory and data. *Psychological Science*, 9, 234–237.
- 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, 1177–1196.
- Seidenberg, M. S., & Waters, G. S. (1989). Word recognition and naming: A mega study [Abstract]. *Bulletin of the Psychonomic Society*, 27, 489.
- Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior*, 23, 383–404.
- Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge, England: Cambridge University Press.
- Shanks, D. R. (1991). Categorization by a connectionist network. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 433–443.
- Shiffrin, R. M. (2003). Modeling memory and perception. *Cognitive Science*, 27, 341–378.
- Shillcock, R., Ellison, T. M., & Monaghan, P. (2000). Eye-fixation behavior, lexical storage, and visual word recognition in a split-processing model. *Psychological Review*, 107, 824–851.
- Siegel, S., & Allan, L. G. (1996). The widespread influence of the Rescorla-Wagner model. *Psychonomic Bulletin and Review*, 3, 314–321.
- Sieroff, E., Pollatsek, A., & Posner, M. (1988). Recognition of visual letter strings following injury to the posterior visual spatial attention system. *Cognitive Neuropsychology*, 5, 427–449.
- Sieroff, E., & Posner, M. (1988). Cueing spatial attention during processing of words and letters strings in normals. *Cognitive Neuropsychology*, 5, 451–472.
- Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science*, 8, 411–416.
- Stolz, J. A., & McCann, R. S. (2000). Visual word recognition: Reattending to the role of spatial attention. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 1320–1331.
- Stone, G. O., & Van Orden, G. C. (1993). Strategic control of processing in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 744–774.
- Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1140–1154.
- Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, 88, 135–170.
- Taft, M. (1979). Lexical access via an orthographic code: The basic



- orthographic syllabic structure (BOSS). *Journal of Verbal Learning and Verbal Behavior*, 18, 21–39.
- Taft, M., & Russell, B. (1992). Pseudohomophone naming and the word frequency effect. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 45(A), 51–71.
- Taraban, R., & McClelland, J. L. (1987). Consistency effects in word recognition. *Journal of Memory and Language*, 26, 608–631.
- Treiman, R., Kessler, B., & Bick, S. (2003). Influence of consonantal context on the pronunciation of vowels: A comparison of human readers and computational models. *Cognition*, 88, 49–78.
- Treiman, R., Mullenix, 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, 107–136.
- Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition*, 15, 181–198.
- Van Orden, G. C., Pennington, B. F., & Stone, G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review*, 97, 488–522.
- Visser, T. A. W., & Besner, D. (2001). On the dominance of whole-word knowledge in reading aloud. *Psychonomic Bulletin and Review*, 8, 560–567.
- Weekes, B. S. (1997). Differential effects of number of letters on words and nonword naming latency. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 50(A), 439–456.
- Wickelgren, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review*, 76, 1–15.
- Widrow, G., & Hoff, M. E. (1960). Adaptive switching circuits. In *Institute of Radio Engineers, Western Electronic Show and Convention record, Part 4* (pp. 96–104). New York: Institute of Radio Engineers.
- Woollams, A. M., Lambon-Ralph, M. A., Plaut, D. C., & Patterson, K. (in press). *SD-squared*: On the association between semantic dementia and surface dyslexia. *Psychological Review*.
- Zevin, J. D., & Seidenberg, M. S. (2006). Simulating consistency effects and individual differences in nonword naming. *Journal of Memory and Language*, 54, 145–160.
- Ziegler, J. C., Besson, M., Jacobs, A. M., Nazir, T. A., & Carr, T. H. (1997). Word, pseudoword, and nonword processing: A multitask comparison using event-related brain potentials. *Journal of Cognitive Neuroscience*, 9, 758–775.
- Ziegler, J. C., Ferrand, L., Jacobs, A. M., Rey, A., & Grainger, J. (2000). Visual and phonological codes in letter and word recognition: Evidence from incremental priming. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 53(A), 671–692.
- Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, 131, 3–29.
- Ziegler, J. C., & Perry, C. (1998). No more problems in Coltheart's neighborhood: Resolving neighborhood conflicts in the lexical decision task. *Cognition*, 68, B53–B62.
- Ziegler, J. C., Perry, C., & Coltheart, M. (2000). The DRC model of visual word recognition and reading aloud: An extension to German. *European Journal of Cognitive Psychology*, 12, 413–430.
- Ziegler, J. C., Perry, C., & Coltheart, M. (2003). Speed of lexical and nonlexical processing in French: The case of the regularity effect. *Psychonomic Bulletin and Review*, 10, 947–953.
- Ziegler, J. C., Perry, C., Jacobs, A. M., & Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science*, 12, 379–384.
- Zorzi, M. (1999). *The connectionist dual-process model: Development, skilled performance, and breakdowns of processing in oral reading*. Unpublished doctoral dissertation, University of Trieste, Trieste, Italy.
- Zorzi, M. (2000). Serial processing in reading aloud: No challenge for a parallel model. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 847–856.
- Zorzi, M. (2005). Computational models of reading. In G. Houghton (Ed.), *Connectionist models in cognitive psychology* (pp. 403–444). London: Psychology Press.
- Zorzi, M., Houghton, G., & Butterworth, B. (1998a). The development of spelling–sound relationships in a model of phonological reading. *Language & Cognitive Processes*, 13, 337–371.
- Zorzi, M., Houghton, G., & Butterworth, B. (1998b). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1131–1161.
- Zorzi, M., Priftis, K., & Umiltà, C. (2002). Neglect disrupts the mental number line. *Nature*, 417, 138–139.
- Zorzi, M., Stoianov, I., & Umiltà, C. (2005). Computational modeling of numerical cognition. In J. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 67–84). New York: Psychology Press.

(Appendixes follow)

## Appendix A

## Complex Graphemes Used in the CDP+ Sublexical Network

The complex graphemes are identical to those implemented in the connectionist model of spelling of Houghton and Zorzi (2003).

The onset consonants were as follows: *ch, gh, gn, kn, ph, qu, sh, th, wh, and wr*.

The vowels were as follows: *air, ai, ar, au, aw, ay, ear, eau, eir, eer, ea, ee, ei, er, eu, ew, ey, ier, ieu, iew, ie, ir, oar, oor, our, oa, oe, oi, oo, ou, or, ow, oy, uar, ua, ue, ui, ur, uy, ye, and yr*.

The coda consonants were as follows: *ght, tch, que, ch, ck, dd, dg, ff, gh, gn, ll, mb, ng, ph, sh, ss, th, tt, and zz*.

## Appendix B

## Parameters Used in the Model

Parameter type	Parameter value
Lexical route	
Features	
Feature-to-letter excitation	0.005
Feature-to-letter inhibition	-0.150
Letters	
Letter-to-letter inhibition	0
Letter-to-orthography excitation	0.075
Letter-to-orthography inhibition	-0.550
Orthographic lexicon	
Orthography-to-orthography inhibition	-0.06
Orthography-to-phonology excitation	1.40
Orthography-to-letter excitation	0.30
Phonological lexicon	
Phonology-to-phonology inhibition	-0.160
Phonology-to-phoneme excitation	0.128
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
Overall parameters	
Overall activation rate	0.2
Lexicon frequency scaling	$> 0.4 \times \log$ (word frequency)
Phoneme naming activation criterion	0.67
Cycle-to-cycle stopping criterion	0.0023
Maximum number of cycles a word is run for before being timed out and considered an outlier	250
Parameters used in the sublexical network	
Network to phonological output buffer activation	0.085
Number of cycles taken for each letter to be processed	15
Level of activation that a letter must be over before grapheme identification begins	.21
Temperature ( $\tau$ ) in the assembly network	3
Learning rate ( $\epsilon$ ) in the assembly network	0.05

## Appendix C

## Activation and Learning Equations Used With the Sublexical Network

The sublexical spelling-to-sound network is identical to the two-layer assembly network of Zorzi et al. (1998b), except that instead of letter units, we have grapheme units. These include the complex (i.e., multiletter) graphemes listed in Appendix A in addition to all single letters.

## Activation Function

For any given input pattern, the input units are clamped to a value of 1.0 or 0.0, according to the presence or absence of the grapheme they encode; the net input to each output unit is simply

$$net_i = \sum_j w_{ij} a_j,$$

where  $a_j$  is the activation value of the input unit  $j$ , and  $w_{ij}$  is the weight of the connections linking the unit  $j$  to the output unit  $i$ . The activation of the output unit  $i$  is determined by an S shaped squashing function (sigmoid) of the net input, bounding phoneme activations in the range [0,1] and with  $f(0) = 0$  (i.e., no input and no output):

$$O_i = \frac{1}{1 + e^{-(net_i - 1)\tau}},$$

where  $\tau$  is a temperature parameter determining the slope of the function ( $\tau = 3$  for all simulations). Note that the  $-1$  in the exponent shifts the sigmoid to the right, such that  $f(0)$  is very close to 0, rather than the standard  $f(0) = 0.5$ . As in Zorzi et al. (1998b), in the simulations reported here, values less than 0.05 are set to 0, so no input really does mean no output.

## Learning Rule

The model was trained with the simple gradient descent technique known as the delta rule (Widrow & Hoff, 1960). For any input pattern, the error correction is made by changing the weights according to the difference between the activation of the output units and desired activation pattern. The desired output is just the correct pronunciation of the orthographic input (nodes that should be on have a target activation of 1, nodes that should be off have a target activation of 0). Formally,

$$\Delta w_{ij} = \epsilon (t_i - o_i) a_j,$$

where  $\epsilon$  is a learning rate (0.05 in the simulations),  $a_j$  is the activation of the  $j^{\text{th}}$  input unit and  $t_i$  and  $o_i$  are the teaching input and the actual output of the  $i^{\text{th}}$  output unit, respectively (for further details, see Zorzi et al., 1998b, pp. 1136–1137).

Appendix D

Grapheme–Phoneme Correspondences Used in Pretraining

Orthography	Phonology	Orthography	Phonology	Orthography	Phonology
---u-e--	---u----	---ew----	---u----	----f--	----f--
---o-e--	---5----	---ue----	---u----	----f--	----f--
---i-e--	---2----	gn-----	n-----	----g--	----g--
---a-e--	---1----	ph-----	f-----	----g--	----g--
---y----	---2----	---eu----	---u----	----k--	----k--
---eigh--	---1----	b-----	b-----	----k--	----k--
---augh--	---9----	d-----	d-----	----k--	----k--
---tsch--	---J----	f-----	f-----	---I---	---I---
---tch--	---J----	g-----	g-----	---I---	---I---
---ck--	---k----	h-----	h-----	---m---	---m---
---ee----	---i----	j-----	j-----	---m---	---m---
---ea----	---i----	k-----	k-----	---n---	---n---
---sh--	---s----	I-----	I-----	---n---	---n---
---sh--	---S----	-I-----	-I-----	---p---	---p---
sh-----	S-----	-I-----	-I-----	---p---	---p---
---ai----	---1----	m-----	m-----	---r---	---r---
---oa----	---5----	n-----	n-----	---r---	---r---
---ng--	---N----	-n-----	-n-----	---s---	---s---
---oo----	---u----	-n-----	-n-----	---s---	---s---
---ou----	---6----	p-----	p-----	---s---	---s---
---ow----	---6----	-p-----	-p-----	---t---	---t---
---ay----	---1----	r-----	r-----	---t---	---t---
---th--	---T----	-r-----	-r-----	---t---	---t---
---th--	---T----	-r-----	-r-----	---t---	---t---
th-----	T-----	s-----	s-----	---z---	---z---
---oi----	---4----	t-----	t-----	---z---	---z---
---au----	---9----	-t-----	-t-----	---tt---	---t---
ch-----	J-----	v-----	v-----	---nn---	---n---
---ch--	---J----	w-----	w-----	---ss---	---s---
---ch--	---J----	-w-----	-w-----	---II---	---I---
---ie----	---2----	z-----	z-----	---rr---	---r---
wh-----	w-----	---a----	---{----	---ff---	---f---
wr-----	r-----	---e----	---E----	---ph---	---f---
---oe----	---5----	---i----	---I----	---ph---	---f---
---oy----	---4----	---o----	---Q----	---pp---	---p---
---ui----	---u----	---u----	---		
kn-----	n-----	---b---	---b---		
---ei----	---1----	---b---	---b---		
---ey----	---1----	---d---	---d---		
---uy----	---2----	---d---	---d---		

Note. Notation is taken from the CELEX database (Baayen et al., 1993). The hyphens represent empty slots in the orthographic or phonological representation.

(Appendixes continue)

## Appendix E

Mean Human Latencies (in Milliseconds) and Model Reaction Times for the Experiments Reported in Jared (2002)

Data set	Human results from Jared (2002)		Model							
			DRC		CDP		Triangle		CDP+	
	Ex/I	Cont	Ex/I	Cont	Ex/I	Cont	Ex/I	Cont	Ex/I	Cont
Experiment 1										
Ex (F < E)	584	548	83.7	77.9	4.06	3.05	0.210	0.017	112.8	102.4
Ex (F > E)	543	536	84.1	78.1	4.06	3.11	0.062	0.013	105.1	100.5
RI (F < E)	572	544	78.8	78.1	3.85	3.11	0.110	0.011	108.6	101.3
RI (F > E)	562	555	76.8	77.4	3.17	3.20	0.054	0.019	101.4	100.6
Fit ( $r^2$ )			1.21		15.11		7.26		23.58	
Fit dif $Z^a$			3.65		1.02		2.19			
Model errors			2		6		1		2	
Outliers			1		3		6		1	
Experiment 2										
HF (F < E)	540	517	80.6	73.6	3.50	2.50	0.230	0.016	92.6	84.5
HF (F > E)	533	521	79.6	73.6	3.30	2.42	0.220	0.022	89.4	87.4
LF (F < E)	610	566	83.7	77.9	4.06	3.05	0.210	0.017	112.8	102.4
LF (F > E)	562	553	84.1	78.1	4.06	3.11	0.062	0.013	105.1	100.5
Fit ( $r^2$ )			8.72		14.74		0.51		40.02	
Fit dif $Z$			3.82		2.93		5.82			
Model errors			2		6		1		1	
Outliers			0		1		5		0	
Experiment 3										
HF Ex	537	518	80.6	73.6	3.50	2.50	0.23	0.016	92.60	84.5
HF RI	537	525	74.1	73.6	2.90	2.39	0.10	0.021	87.25	85.8
LF Ex	596	573	83.7	77.9	4.06	3.05	0.21	0.017	112.8	102.4
LF RI	593	564	78.8	78.1	3.85	3.11	0.11	0.011	108.6	101.3
Fit ( $r^2$ )			17.80		31.25		6.72		52.38	
Fit dif $Z$			3.84		2.46		5.62			
Model errors			0		3		0		1	
Outliers			1		2		4		1	
Experiment 4										
HF Ex	530	509	80.6	73.6	3.50	2.50	0.23	0.016	92.60	84.5
HF RI	528	517	74.1	73.6	2.90	2.39	0.10	0.021	87.25	85.8
LF Ex	573	542	83.7	77.9	4.06	3.05	0.21	0.017	112.80	102.4
LF RI	566	540	78.8	78.1	3.85	3.11	0.11	0.011	108.60	101.3
Fit ( $r^2$ )			18.50		30.38		5.40		46.28	
Fit dif $Z$			3.22		1.82		5.12			
Model errors			0		3		0		1	
Outliers			1		2		4		1	

*Note.* DRC = dual-route cascaded model; CDP = connectionist dual process model; CDP+ = new connectionist dual process model; Ex = exception; I = inconsistent; Cont = control; RI = regular inconsistent; F = friends; E = enemies; HF = high frequency; LF = low frequency.

<sup>a</sup> Fit dif  $Z$  =  $Z$  score differences.  $Z$  score differences were calculated to examine the difference in correlation strengths between CDP+ and the other models on each data set. This was done with the following formula (e.g., Clark-Carter, 2004, p. 310),

$$z = \frac{r'_1 - r'_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}$$

where  $r'_1$  is the Fischer's transformation—that is,  $r'_v = 0.5 \times \log_e\left(\frac{1+r}{1-r}\right)$ —of one correlation coefficient;  $r'_2$  is the Fischer's transformation of the other; and  $n$  is the number of items in the group. All  $r^2$  values reported are multiplied by 100 (as they are in all of the other appendixes), therefore reflecting the percentage of variance explained by the models.



Appendix F

Mean Human and Model Results (Percentage of Regular Responses) on Items Reported in Experiment 1 and Experiment 2 of Andrews and Scarratt (1998)

Variable	Human results	Model results			
		DRC	CDP	Triangle	CDP+
Experiment 1					
CV con/VC con	92.2	92.2	100.0	100.0	100.0
CV con/VC inc	86.8	100.0	86.8	84.2	88.1
CV inc/VC con	94.0	98.1	95.8	96.1	100.0
CV inc/VC inc	86.9	97.8	86.3	85.1	93.2
No regular analogy	32.3	67.8	21.4	66.7	25.0
Fit ( <i>RMSE</i> ) <sup>a</sup>		17.72	6.05	15.87	6.19
Model errors		3	23	11	9
Outliers		5	4	5	5
Experiment 2					
Consistent	92.5	100.0	97.2	100.0	97.3
Inconsistent	87.4	97.4	83.8	86.8	82.1
No regular analogy (many bodies)	19.2	62.5	4.5	47.8	10.5
No regular analogy (few bodies)	41.2	63.6	55.6	47.8	50.0
Fit ( <i>r</i> <sup>2</sup> )		33.68	61.61	22.24	44.72
Fit dif Z		1.13	-1.87	2.23	
Model errors		2	12	8	6
Outliers		2	3	3	3

Note. DRC = dual-route cascaded model; CDP = connectionist dual process model; CDP+ = new connectionist dual process model; CV = consonant-vowel; VC = vowel-consonant; con = consistent; inc = inconsistent; *RMSE* = root-mean-square error; Fit dif Z = Z score based on the difference in correlation strengths between the CDP+ and other models (see Appendix E for details).

<sup>a</sup>No item data were available for Experiment 1. Therefore, fits were calculated using *RMSE* values computed from the means. *RMSE* scores were calculated with the following formula:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N e_i^2}$$

where *e* is the error (i.e., observed score minus the actual score), and *N* is the number of groups. Note that the smaller the *RMSE* value the better the fit. Fits for Experiment 2 were created by coding whether a word was produced as regular by the model or not as 0 or 1 and by correlating those numbers with the probability that participants gave a regular response.

Appendix G

Mean Human and Model Results (Percentage of Correct Answers)

Variable	Human results	Model results			
		DRC	CDP	Triangle	CDP+
Degree of consistency					
Low (F < E)	38.0	58.5	37.5	29.0	41.7
Medium (F > E)	57.0	70.0	56.7	63.0	56.7

Variable	Human results	Model results			
		DRC	CDP	Triangle	CDP+
High ( <i>wa</i> words)	75.0	50.0	83.3	83.0	66.7
Fit ( <i>RMSE</i> )		20.1	4.8	7.8	5.2
Control 1-Word	100.0	95.8	91.7	96.0	91.7
Control 2-Word	83.0	96.7	80.0	83.0	86.7
Control 3-Word	83.0	100.0	83.3	100.0	100
Fit ( <i>RMSE</i> )		12.8	5.1	10.1	11.1

Note. Human results are for patient MP (Patterson & Behrmann, 1997). Means for the CDP and triangle model are taken from Zorzi (1999) and Patterson and Behrmann (1997), respectively. Fits were calculated with an *RMSE* measure using the mean scores (see Appendix F for details). DRC = dual-route cascaded model; CDP = connectionist dual process model; CDP+ = new connectionist dual process model; F = friends; E = enemies; *RMSE* = root-mean-square error.

Appendix H

Human Data (in Milliseconds) and Simulation Results on the Items Reported in Weekes's (1997) Study, Which Manipulated Word Length, Lexicality, and Frequency

Length	Human results	Model data			
		DRC	CDP	Triangle	CDP+
Words					
LF 3	535	72.2	3.08	0.023	94.2
LF 4	549	77.4	3.28	0.029	100.9
LF 5	552	78.1	3.42	0.018	106.1
LF 6	566	77.8	3.72	0.041	110.6
HF 3	535	69.4	2.21	0.012	77.8
HF 4	532	74.3	2.64	0.024	87.0
HF 5	546	75.0	2.71	0.031	91.1
HF 6	542	73.9	2.70	0.020	98.3
Fit ( <i>r</i> <sup>2</sup> )		4.0	0.48	0.15 <sup>a</sup>	8.54
Fit dif Z		0.7	1.91	1.20	
Model errors		0	0	0	0
Outside 3 SD		4	4	6	2
Nonwords					
3	576	121.2	4.25	0.34	120.0
4	577	138.4	4.29	0.16	131.1
5	606	152.4	4.81	0.21	150.5
6	666	186.7	4.76	0.21	160.9
Fit ( <i>r</i> <sup>2</sup> )		40.3	1.11	2.90	30.80
Fit dif Z		-0.8	4.56	4.46	
Model errors		1	9	1	8
Outside 3 SD		0	2	3	3

Note. DRC = dual-route cascaded model; CDP = connectionist dual process model; CDP+ = new connectionist dual process model; LF = low frequency; HF = high frequency; Fit dif Z = Z score based on the difference in correlation strengths between the CDP+ and other models (see Appendix E).

<sup>a</sup>The correlation was negative for this fit value.

(Appendixes continue)

## Appendix I

Human Data (in Milliseconds) and Model Results on the Items Reported in Ziegler et al.'s (2001) Experiment as a Function of Word Length, Lexicality, and Body Neighborhood

Length	Human results		Model data							
	LBN	HBN	DRC		CDP		Triangle		CDP+	
			LBN	HBN	LBN	HBN	LBN	HBN	LBN	HBN
Words										
3	526	506	70.3	70.4	2.4	2.8	0.039	0.020	82.9	83.8
4	527	506	74.4	74.1	3.0	2.7	0.017	0.023	91.6	86.7
5	530	512	74.5	74.0	3.0	2.4	0.054	0.019	93.4	90.7
6	534	525	75.0	74.0	2.9	2.4	0.046	0.057	96.6	96.1
Fit ( $r^2$ )			9.78		0.74		11.03		6.44	
Fit dif Z			-0.40		1.07		-0.52			
Errors			0		0		0		0	
Outliers			2		0		3		0	
Nonwords										
3	577	568	122.0	121.5	4.44	4.1	0.35	0.160	125.2	118.2
4	617	589	136.8	133.9	4.50	4.0	0.34	0.028	132.1	127.1
5	625	623	154.8	148.9	4.56	4.5	0.15	0.190	156.8	150.1
6	662	627	173.5	188.4	5.44	4.4	0.33	0.042	170.3	164.4
Fit ( $r^2$ )			22.41		15.41		(0.03)		23.04	
Fit dif Z			0.06		0.64		3.30			
Errors			0		6		0		3	
Outliers			0		1		2		1	

*Note.* Parentheses indicate that the correlation between the model and the data is negative. DRC = dual-route cascaded model; CDP = connectionist dual process model; CDP+ = new connectionist dual process model; LBN = low body neighborhood; HBN = high body neighborhood; Fit dif Z = Z score based on the difference in correlation strengths between the CDP+ and other models (see Appendix E).

## Appendix J

Human Data (in Milliseconds) and Simulation Results (in Cycles) for Items Used in the Position-of-Irregularity Experiment by Rastle and Coltheart (1999)

Position of irregularity	Human results		Model results							
	Irreg	Reg	DRC		CDP		Triangle		CDP+	
			Irreg	Reg	Irreg	Reg	Irreg	Reg	Irreg	Reg
Position 1	556	494	97.5	78.4	4.24	3.25	0.23	0.015	114.8	105.6
Position 2	510	497	87.7	78.2	3.97	3.27	0.17	0.031	109.6	103.1
Position 3	512	512	79.7	78.2	4.07	3.64	0.13	0.024	109.9	107.1
Fit ( $r^2$ )			12.81		11.76		5.94		21.30	
Fit dif Z			1.15		1.30		2.31			
Model errors			3		9		2		0	
Outliers			1		2		3		0	

*Note.* DRC = dual-route cascaded model; CDP = connectionist dual process model; CDP+ = new connectionist dual process model; Irreg = irregular; Reg = regular; Fit dif Z = Z score based on the difference in correlation strengths between the CDP+ and other models (see Appendix E for details).

## Appendix K

## Human Data (in ms) and Simulation Results (in Cycles) for Items Used in the Position-of-Irregularity Experiment by Roberts et al. (2003)

Position of irregularity	Model results									
	Human results		DRC		CDP		Triangle		CDP+	
	Irreg	Reg	Irreg	Reg	Irreg	Reg	Irreg	Reg	Irreg	Reg
Position 2	553	526	90.22	78.66	4.39	3.26	0.088	0.030	118.0	105.8
Position 3	555	521	79.72	77.56	4.25	3.50	0.082	0.033	113.9	106.8
Fit ( $r^2$ )			0.43 <sup>a</sup>		1.57		2.95		6.29	
Fit dif Z			2.24		0.88		0.56			
Model errors			2		13		1		3	
Outliers			2		0		0		1	

*Note.* DRC = dual-route cascaded model; CDP = connectionist dual process model; CDP+ = new connectionist dual process model; Irreg = irregular; Reg = regular; Fit dif Z = Z score based on the difference in correlation strengths between the CDP+ and other models (see Appendix E for details).  
<sup>a</sup> DRC  $r^2$  is from a negative correlation.

## Appendix L

Frequency  $\times$  Regularity Interaction (Paap & Noel, 1991)

Paap and Noel (1991) performed a classic study examining the Frequency  $\times$  Regularity interaction. The results showed a significant effect of regularity, but only with low-frequency words. CDP+ correctly predicted Paap and Noel's data (unlike the DRC, where there was a significant effect with high-frequency words), with a main effect of frequency,  $F(1, 72) = 129.24$ ,  $MSE = 10,436$ ,  $p < .001$ ; a main effect of regularity,  $F(1, 72) = 19.87$ ,  $MSE = 1,604$ ,  $p < .001$ ; and an interaction between them,  $F(1, 72) = 11.02$ ,  $MSE = 890$ ,  $p < .005$ . The model produced one error and one outlier. Two  $t$  tests examining the high- and low-frequency groups showed that only the low-frequency words produced a significant regularity effect: for low frequency, 113.65 versus 97.58 cycles,  $t(34) = 5.25$ ,  $SE = 1.99$ ,  $p < .001$ ; for high frequency, 83.30 versus 80.95 cycles,  $t < 1$ .

Received July 7, 2005

Revision received November 3, 2006

Accepted November 10, 2006 ■