

DRC: A Dual Route Cascaded Model of Visual Word Recognition and Reading Aloud

Max Coltheart, Kathleen Rastle, Conrad Perry,
and Robyn Langdon
Macquarie University

Johannes Ziegler
Macquarie University and University of Provence

This article describes the Dual Route Cascaded (DRC) model, a computational model of visual word recognition and reading aloud. The DRC is a computational realization of the dual-route theory of reading, and is the only computational model of reading that can perform the 2 tasks most commonly used to study reading: lexical decision and reading aloud. For both tasks, the authors show that a wide variety of variables that influence human latencies influence the DRC model's latencies in exactly the same way. The DRC model simulates a number of such effects that other computational models of reading do not, but there appear to be no effects that any other current computational model of reading can simulate but that the DRC model cannot. The authors conclude that the DRC model is the most successful of the existing computational models of reading.

The psychology of reading has been revolutionized by the development of computational models¹ of visual word recognition and reading aloud. Computational modeling has many advantages over the alternative way of expressing theories about cognition, which is to represent the theories as “verbal models” (Jacobs & Grainger, 1994). First, an attempt to express any theory about cognition as a computational model immediately reveals many ways in which that theory is incomplete or underspecified, some of which the theorist will not have suspected. A program will not run unless it is fully specified, so a theory cannot yield an executable program unless that theory is also fully specified. Second, once this process has compelled the theorist to fill in the gaps in the theory and an executable program has been created, the adequacy of the theory can be rigorously assessed by simulation. Are all the effects observed in the behavior of people when they are carrying out the cognitive activity in question also seen in the behavior of

the program—for example, in the time the program takes to perform relevant tasks or in the accuracy with which it performs such tasks? Mismatches between human behavior and the behavior of the computational model reveal ways in which the theory from which the model was generated is incorrect as a description of a human information-processing system. Sometimes relatively minor reformulations of specific aspects of the theory can eliminate such mismatches; here modeling has led to a better theory. Sometimes these mismatches are so fundamental that there is no way of making such minor modifications of the theory; here modeling has led to theory refutation.

Of course, even if a theory of cognition is expressible as a computer program that actually runs, and even if all the effects known to characterize human performance in the relevant cognitive domain also characterize the model's performance, it remains possible that some other theory of the same cognitive activity may also be expressible as a computer program that actually runs, and it is also possible that the behavior of this second computational model also captures all the effects known to characterize human performance in the relevant cognitive domain. In other words, even if turning the theory into a computational model has successfully shown the theory to be complete (in the sense that it contains no unspecified processes) and to be sufficient (in the sense that it can offer an account of all the relevant empirical phenomena), that does not guarantee that it is correct.

Nothing ever guarantees, of course, that any theory in any branch of science is correct. But if there is no other theory in the field that has been demonstrated through computational modeling to be both complete and sufficient, resting on laurels is a reasonable thing to do until the emergence of such a competitor—that is,

Max Coltheart, Kathleen Rastle, Conrad Perry, and Robyn Langdon, Macquarie Centre for Cognitive Science, Macquarie University, Sydney, Australia; Johannes Ziegler, Macquarie Centre for Cognitive Science, Macquarie University, Sydney, Australia, and Centre Nationale de la Recherche Scientifique, University of Provence, Marseille, France.

We thank numerous people for valuable assistance, comments, and criticisms: Sally Andrews, Rita Berndt, Derek Besner, Anne Castles, Veronika Coltheart, Martin Davies, Colin Davis, Ken Forster, Carol Fowler, Ram Frost, Jonathan Harrington, Arthur Jacobs, Sachiko Kinoshita, Ken Paap, Karalyn Patterson, David Plaut, Karen Smith-Lock, Marcus Taft, Anna Woollams, and Marco Zorzi. We also thank Alice Coltheart, Robert Mannell, Steve Saunders, and Carl Windhorst for technical assistance.

This work has been funded by Australian Research Council Special Investigator Award A79703207 and Australian Research Council Large Grants AC8932386, AC9231176, AC7970320, and A79700243.

Correspondence concerning this article should be addressed to Max Coltheart, Macquarie Centre for Cognitive Science, Macquarie University, Sydney, NSW 2109, Australia. Electronic mail may be sent to max@maccs.mq.edu.au.

¹ By the term “computational model,” we mean a computer program that is capable of performing the cognitive task of interest and does so by using exactly the same information-processing procedures as are specified in a theory of how people carry out this cognitive activity.

the emergence of a different theory that has also been shown to be both complete and sufficient.

Here the computational modeling approach is yet again valuable, because it facilitates theory adjudication. If the competing theories really are different theories, their computational models will differ too, and analysis of these differences will pinpoint ways in which the models make different predictions about the results of experiments not yet conducted. Carrying out these new experiments may therefore result in adjudication between the two theories, as well as result in new discoveries about the cognitive domain in question.

Two Approaches to the Computational Modeling of Cognition

An immensely popular approach to computational modeling over the past decade has been to develop computational models through a learning algorithm, generally some form of backpropagation (Rumelhart, Durbin, Golden, & Chauvin, 1995; Rumelhart, Hinton, & Williams, 1986). Here, the computational model is typically based on a network with three layers (input units, hidden units, output units) with weights initialized to small random values. A training set of stimulus–response pairs is chosen and submitted to the model for a large number of training epochs. As training proceeds, the learning algorithm progressively adjusts the network's connection weights so that for each stimulus (representation across the input units) the response (representation across the output units that is evoked by that input) more and more closely approximates the correct response. Computational models of reading aloud of English that have been developed using this approach include those of Seidenberg and McClelland (1989), hereafter referred to as the SM model; Plaut, McClelland, Seidenberg, and Patterson (1996), hereafter referred to as the PMSP model; and Zorzi, Houghton, and Butterworth (1998), hereafter referred to as the ZHB model.

Our approach has been a different one. Like McClelland and Rumelhart (1981), Rumelhart and McClelland (1982), Jacobs and Grainger (1992), and Grainger and Jacobs (1996), we (Coltheart, Curtis, Atkins, & Haller, 1993, and the present article) have specified the architecture of the model ourselves, rather than relying on backpropagation to do this. Our computational model is hand-wired rather than learned. We believe that our approach has a number of advantages, even though in the past decade it has been adopted far less often than the learning algorithm approach.

One disadvantage of the learning algorithm approach is that the range of possible model architectures that can be developed through backpropagation is constrained, and the correct model may not be within that range. For example, suppose it were true that the human reading system used a local representation scheme by which individual words were represented as individual processing units. That might not be true, but it certainly could be true; it is clearly logically possible. If it were true, the problem for any approach that sought to develop a computational model of the human reading system by using backpropagation is that this learning algorithm develops distributed, rather than local, representations unless specifically prevented from doing so. Hence if it is the case that representations of words in the human reading system are local, then attempts at developing a correct computational model of that system using backpropagation could not succeed.

Similarly, it is logically possible that the human reading system contains more than two different representational levels. It might contain, say, a letter level, an orthographic word level, a phonological word level, and a phoneme level. This architecture with four explicit representational levels could not arise from the use of backpropagation with a neural network hardwired to have only two explicit representational levels (input units and output units); yet it is certainly logically possible that the human reading system has such an architecture.

Occasionally those modelers who use the backpropagation approach have attempted to deal with this problem by seeking to analyze the internal processing structure of a fully trained network. For example, Plaut et al. (1996) raised the question of whether their one-route model of reading aloud, an orthodox three-layer network trained by backpropagation, was actually a two-route model. Perhaps one part of the trained network was responsible for the ability to use general knowledge about letter–sound relationships, whereas another part of the network was responsible for handling words that violated standard rules about such relationships? They carried out various analyses of the trained network and concluded that the network was not structured in this way.

There is an important distinction here between network architecture and functional architecture. It is easy to find out what the network architecture is, because that is fully decided by the modeler. In the case of the PMSP model, the network architecture consists of three layers, with 105 input units fully connected to 100 hidden units that are, in turn, fully connected to 61 output units (with reciprocal connections from output units to hidden units in the case of the attractor version of the network). In contrast, it is very difficult to discover the functional architecture of a trained network for networks of realistic size—that is, to discover how the trained network has been structured by the learning algorithm so as to be able to perform the task it has learned. The majority of modelers using the learning algorithm approach do not even try to do this; they exhibit no interest in what the functional architecture of the system is, which is why some might see their approach as a kind of New Behaviorism. We are adherents of Old Cognitivism, and so our main interest is in the internal structure—the functional architecture—of human cognitive systems.

There is also a practical difficulty here for the learning algorithm approach. Let us suppose it were actually possible for a network trained to read aloud by backpropagation to develop a functional architecture in which each exception word in the training set had one (or even more than one) hidden unit dedicated to it, and that in addition to this set of dedicated hidden units there were a set of hidden units connected in such a way as to encode general information about letter–sound relationships (permitting the network to read aloud nonwords even though it had never seen these before).

The training set used by Seidenberg and McClelland (1989) and by Plaut et al. (1996) contains about 750 exception words; but the networks they used contained only 200 and 100 hidden units, respectively. These practical facts about the network architecture would prevent the development of this functional architecture by backpropagation, even if that were the correct functional architecture, and even if the backpropagation learning algorithm were capable of developing such a functional architecture when provided with a large enough number of hidden units. The problem cannot be avoided simply by giving the network a large enough

number of hidden units, for two reasons: we have no way of knowing how many is large enough, and even if we tried using, say, 800 hidden units (one per exception word plus 50 to learn general facts about spelling–sound correspondences), this would greatly increase the number of connections in the network and therefore greatly increase the computer time needed to train the network; even with the existing PMSP network, training times pose a practical problem (see Plaut, 1997, Footnote 5, for an example of a desirable simulation that was not attempted because it would have taken about 6 months of CPU time).

Our comments here about learning algorithms are meant only to apply to cases where the functional architecture of the model is determined by the algorithm. A quite different situation is one in which that functional architecture is prespecified by the modeler, and then some form of learning algorithm is used to set the strengths of the connections between the prespecified modules of the architecture. We are in sympathy with this particular way of using learning algorithms and indeed could apply this method to the model we describe in this article.

It is beyond the scope of this article to produce a definitive assessment of the value of the learning algorithm approach to the construction of computational models of cognition, and that has not been our aim in pointing out what we see as some disadvantages of this approach. We have simply wished to explain why we prefer a different approach. Essentially, our view is that the past quarter of a century of empirical and theoretical research on reading has provided us with good reasons for proposing a particular architecture for the human reading system, and our preference is to rely on this body of literature, rather than on backpropagation, for ideas about what this architecture might be. We are thus adopting the same view as Grainger and Jacobs (1998) who wrote, “We therefore argue that in developing algorithmic models of cognitive phenomena, the major source of constraint is currently provided by human behavioral data” (p. 24).

We have relied considerably on an early computational model of visual word recognition, the interactive activation and competition (IAC) model of McClelland and Rumelhart (1981) and Rumelhart and McClelland (1982), a generalization of which functions as a front end of the model we have developed. The architecture of this model was specified by its creators, rather than developed by a learning algorithm. The model has three representational levels: a visual feature level, a letter level, and an orthographic word level. Its sufficiency was evaluated only against data from one particular reading task, the Reicher–Wheeler forced-choice tachistoscopic recognition task (Reicher, 1969; Wheeler, 1970) and the model’s performance in this task was shown to fit quantitatively rather well the performance of human readers. However, Massaro (1989) demonstrated that certain context effects seen in this task were inconsistent with the model. In reply, McClelland (1991) demonstrated that a relatively minor modification to the IAC model (making its computations noisy rather than deterministic) caused the model to show the context effects that human readers also show.

Nested Modeling

The IAC model was the first computational model of reading, and its achievements were impressive. Yet it was abandoned by its creators, one of whom subsequently proposed a completely differ-

ent model of reading (Seidenberg & McClelland, 1989). Future historians of the cognitive psychology of the last century will wonder why this happened, because the IAC model had not been refuted; on the contrary, the attempt at refutation of the model by Massaro (1989) had met a stout defense from McClelland (1991). Commenting on this, Jacobs and Grainger (1994) observed that, “in other sciences it is standard practice that a new model accounts for the crucial effects accounted for by the previous generation of the same or competing models” (p. 1329). We agree wholeheartedly; it is not easy to see how a science can develop cumulatively if previous work is abandoned in what might be seen as an unmotivated way.

Cumulativeness in relation to modeling is referred to as “nested modeling” by Jacobs and Grainger (1994):

The principle of nested modeling (i.e., a new model should be related to or include, at least, its own, direct precursors and be tested against the old data sets that motivated the construction of the old model before testing it against new ones . . . a new model should either include the old one as a special case by providing formal demonstrations of the inclusion, or dismiss with it, after falsification of the core assumptions of the old model. (pp. 1329–1330)

That is the practice that Coltheart et al. (1993) followed and that we follow here; a generalization of the IAC model is part of the computational model we have developed. A major advantage of doing this is that it guarantees that our model will be able to give a good account of anything of which the IAC model originally gave a good account, data from the Reicher–Wheeler task, for example. In contrast, although the SM model included a component whose sole function would seem to have been to allow the model to deal with that task (the feedback connections from hidden units to orthographic input units), this component was not implemented in a way that would allow simulation of the Reicher–Wheeler task, and in any case this component is no longer present in the successor to the SM model, the PMSP model.

A similar attempt at cumulativeness exists in the phonological output side of our model, which can be seen as a (highly simplified) version of certain speech-production models such as that of Dell (1986) and Levelt, Roelofs, and Meyer (1999). Those models have enjoyed some success in explaining certain aspects of normal and aphasic speech production, such as slips of the tongue and types of aphasic speech errors. For our model also to be able to simulate such phenomena, its phonological output side would have to be much more elaborate, but the basic structure of these speech-production models (a phonological lexicon activating a phoneme system) is present, and so could be elaborated.

In the spirit of nested modeling, then, the model we describe in this article has grown directly from previous models of visual word recognition and of spoken word production. In fact, our work even builds on modeling endeavors that are more than a century old.

A Little History

Nineteenth-Century Origins

The idea that the language-processing system includes an orthographic lexicon, as in the IAC model, and a phonological lexicon, as in the speech-production models of Dell (1986) and Levelt et al. (1999), and that reading aloud relies in part on these

two lexicons, is more than a century old. Cognitive neuropsychologists of the 19th century held the view that the language-processing system was highly modular in structure, and they also held the view that an appropriate notation for describing hypothesized architectures of such systems was the box-and-arrow notation. This is clear in, for example, the work of Lichtheim (1885); he proposed an architecture for the phonological components of the language-processing system and described it with the diagram shown in Figure 1.

More than a century later, this modular modeling approach² and the box-and-arrow diagram notation for defining such models, remains popular—and indeed the specific model proposed by Lichtheim (1885) is still popular. The model of language processing proposed by Patterson and Shewell (1987) includes all the modules of Lichtheim’s model (Figure 2 is the part of the Patterson–Shewell model that corresponds to Lichtheim’s model); models of language processing equivalent to the Patterson–Shewell model can be found in other contemporary sources such as Morton and Patterson (1980), Harris and Coltheart (1986), Ellis and Young (1988), and Kay, Lesser, and Coltheart (1992).

Our model is a computational realization of that part of the Morton, Harris–Coltheart, Patterson–Shewell, Ellis–Young, and Kay–Lesser–Coltheart models that is specifically relevant to reading.

Although this approach to modeling cognition, and this particular model of language processing, was popular toward the end of the 19th century and remains popular in the 21st, there was a long period during which both the modular modeling approach and the box-and-arrow notation for expressing theories about cognition had vanished from cognitive psychology—a period from perhaps 1900 to the mid-1950s. This happened for two reasons.

The first reason is that the 19th-century modelers such as Lichtheim and Wernicke were neurologists and so were not content just to build models of the functional architecture of the language-processing system. They also wanted to localize their functional modules in specific brain regions, by carrying out postmortem analyses of the brains of patients who had suffered some form of language impairment caused by stroke or other brain insult. So, for example, they might seek to determine which region of the brain houses a center for auditory word representations by investigating which region of the brain was damaged in patients whose aphasia was interpreted using the model as arising because

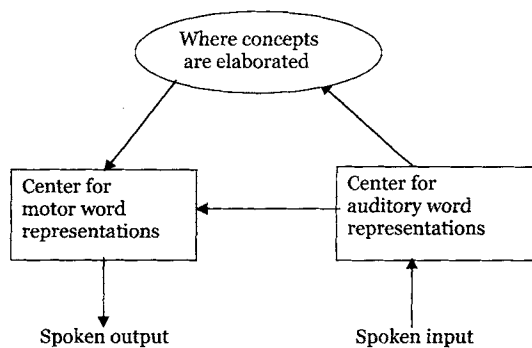


Figure 1. The model of spoken-word processing from Lichtheim (1885) (adapted from Shallice, 1988, Figure 1.1).

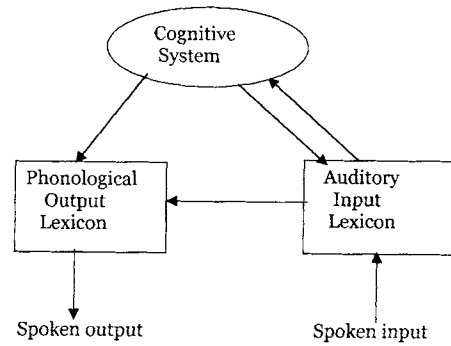


Figure 2. The model of spoken-word processing from Patterson and Shewell (1987).

of specific damage to this particular functional module. It turned out that, in this work, correspondences between the lesions seen in the brains of aphasic patients and the lesions hypothesized to have occurred in the functional architecture of the language-processing systems were so inconsistent that it was easy for neurologists of a noncognitive and nonmodular persuasion—Marie (1906) and Head (1926), for example—to ridicule and discredit the whole modular/box-and-arrow approach. Even as late as 1964 such criticisms were still being made:

The older neurologists, and even some today, thought that the different varieties of aphasia produced by lesions in different situations could be classified in psychological terms . . . but this presupposes first that in the nervous system speech is organized in such a way that anatomical centers correspond to psychological functions, and then that destruction of such a center merely impairs a particular psychological element in speech. This view has largely been abandoned. (Brain, 1964, p. 6)

This “largely abandoned” view returned with a vengeance just one year later (Geschwind, 1965a, 1965b).

It is scarcely surprising that attempts made a century ago to localize the modules of the language-processing system did not succeed, because even today they have not succeeded. When brain imaging is used instead of autopsy for neuroanatomical localization, and late 20th-century models of the functional architecture of the language-processing system are used for functional localization, it has still not been possible to establish decisively any mappings between regions of the brain and modules of the language-processing system (Poeppel, 1996a, 1996b; Van Orden & Paap, 1997).

At the beginning of this century, then, this was the fate of the modular modeling approach as far as neurology and neuropsychology was concerned. With respect to psychology itself, there was an equally devastating turn of events—the onslaught of Behaviorism from Watson (1913) onward. Explaining behavior in terms of functional mental architectures whose elements are “unobservable” is of course anathema to Behaviorism. This was the second reason for the disappearance of the modular modeling approach.

² We use the term “module” to mean a domain-specific cognitive processing system; see Coltheart (1999) for an elaboration and justification of this usage.

It was not until the 1960s and early 1970s that Wernicke-Lichtheim-style cognitive neuropsychology began to rise again, seminal papers being those of Marshall and Newcombe (1966, 1973) and Shallice and Warrington (1970). The renaissance of cognitive psychology itself began a little earlier, in the late 1950s and early 1960s, particularly with Broadbent (1958), and it is significant that one feature of this renaissance was a resumption of the use of box-and-arrow notation, which was used by Broadbent to express his ideas about the functional architecture of the selective-attention system (see, e.g., Figure 7 in Broadbent, 1958).

The Mental Lexicon

The concept of a mental lexicon seems to have been introduced into psycholinguistics in Anne Treisman's (1961) doctoral thesis:

In order to go much further in understanding the nature of the monitoring carried out (during selective attention tasks), it will be necessary to investigate and make some hypotheses about the nature of the word identification system that lies beyond the selective system (pp. 116–117).

The hypotheses she made involved the postulation of a mental lexicon (she used the term "dictionary" rather than "lexicon") with individual lexical entries ("dictionary units" for Treisman) in that lexicon representing individual words: "There is a single channel for recognising words, presumably comprising the matching of signals with some kind of 'dictionary' . . . some of whose units have their thresholds for activation permanently or temporarily lowered" (p. 210). Figure 14 of Treisman's thesis expressed these ideas in box-and-arrow notation, as shown in Figure 3.

And, most interesting, Treisman (1961) discussed the relationship between her theoretical ideas and 19th-century cognitive neuropsychology:

There have . . . been cases reported in the literature of selective losses, which are interesting from the point of view of the suggestions made in this thesis. Wernicke (1874), for example, reported that patients might lose either the "Wortschatz" or "word treasury" which might correspond to the dictionary system, or the "Klangbild" or "sound picture," which might be analyzed before the selective system of attention operated, in the model proposed here. Examples of receptive or perceptual losses are given also by Ziegler (1952), who reports cases of word deafness (sometimes with more general auditory agnosia) although the patients were not deaf to sounds, and by Hemphill and Stengel (1940). These authors report one patient saying "Voice come but no words. I can hear. Sound comes, but words don't separate." Other patients may show marked difficulty in selecting words to name objects. This would also involve the word store or dictionary system, but would mean approaching it from another angle, from the associations with objects rather than spoken sounds. The availability of the word unit is much decreased. Similar difficulties were induced in non-aphasic patients by Penfield's brain stimulation methods. Penfield postulates the existence of two different stores, one for words and one for concepts. . . . Evidence for the existence of the two stores comes from patients who seem to retain concepts, yet are unable to find the appropriate names. They may describe the function, calling a knife "something that cuts," or they may try to find similar words. (pp. 266–267)

Note the various distinct cognitive modules hinted at or specified in just this one paragraph from Treisman's (1961) thesis:

1. A store of concepts (meanings). This is referred to as center

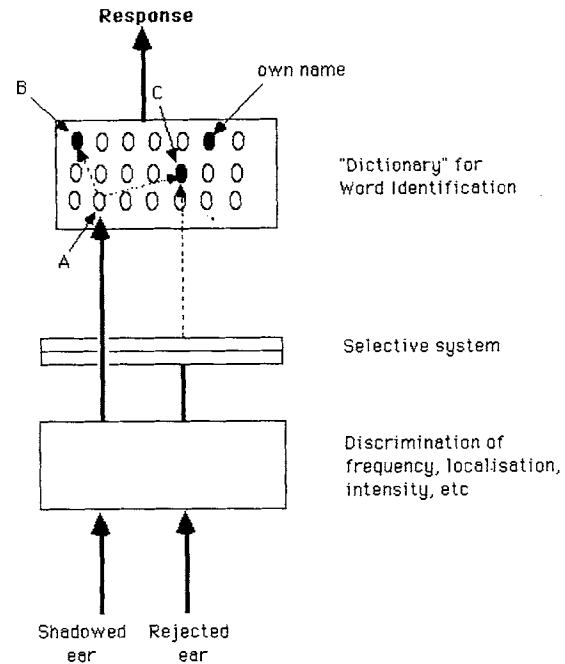


Figure 3. A model of the mental lexicon. The thresholds of Words B and C are lowered by their high transitional probability after Word A. Word C is also activated by the "attenuated" signal from the rejected message and is sometimes heard. From *Attention and Speech*, by A. N. Treisman, 1961, p. 211. Unpublished doctoral thesis. Adapted with permission.

B in the model of Lichtheim (1885; see Figure 1), as the cognitive system in the models of Morton and Patterson (1980) and Patterson and Shewell (1987), as the semantic/cognitive system in the model of Harris and Coltheart (1986) and as the semantic system in the models of Ellis and Young (1988) and of Kay et al. (1992) and in the DRC model (Coltheart et al., 1993).

2. A store of "sound pictures" used to recognize spoken words. This is distinct from the store of auditory representations of environmental sounds, because some patients with word deafness can recognize environmental sounds. This is center A in the model of Lichtheim (1885), the auditory input logogen system in the model of Morton and Patterson (1980), the auditory input lexicon in the models of Patterson and Shewell (1987) and Ellis and Young (1988), the auditory word recognition system in the model of Harris and Coltheart (1986), and the phonological input lexicon in the model of Kay et al. (1992).

3. A store of representations of spoken words that is used for producing spoken words, rather than for recognizing them. This is center M in the model of Lichtheim (1885), the output logogen system in the model of Morton and Patterson (1980), the spoken word production system in the model of Harris and Coltheart (1986), the speech output lexicon in the model of Ellis and Young (1988), and the phonological output lexicon in the models of Patterson and Shewell (1987) and Kay et al. (1992).

4. A store of object representations used to recognize pictures or seen objects. Lichtheim did not discuss vision, but in the same era Lissauer (1890) did, and he drew a distinction between apperceptive visual agnosia (a failure to access a store of object representations from vision) and associative visual agnosia (a failure to

access a store of concepts or meanings from the store of object representations). This store of object representations is the pictogen system in the extension of the logogen model to picture recognition by Seymour (1973) and is the system of object recognition units in the model of Ellis and Young (1988).

The Logogen Model and its Evolution

Morton (1961) also based his theorizing on the concept of a mental lexicon or, rather, on the concept of mental lexicons, because he drew a sharp distinction, not explicitly drawn by Treisman (1961), between a system of knowledge about word meanings (which he referred to as the “cognitive system”) and a system of knowledge about word forms (which he referred to as the “logogen [word generation] system”). Morton’s distinction between cognitive system and logogen system corresponds to the above-mentioned distinction drawn by Wernicke (1874) between the “wortschatz” or “word treasury” (cognitive system) and the “klangbild” or “sound picture” (logogen system). Over the next decade Morton’s logogen model evolved through several increasingly complex forms. We briefly review the model’s evolutionary history, for two reasons.

The first reason is that the model of reading we describe in this article, although it is not a logogen model, evolved from the final form of the logogen model.

The second reason is to show how the evolution of the logogen model was entirely data-driven; complexity was added not for the sake of complexity but as a consequence of hypotheses about the explanations of empirical results that a previous model could not explain. Hence the complex form of the dual-route cascaded (DRC) model, which was inherited from the final version of the logogen model, is motivated by a series of empirical findings. Any simplification of the DRC model—that is, omission of any one of its components or pathways—would leave it unable to offer an explanation of one or more of these findings.

As will be seen from the discussion below of the evolution of the logogen model, the logogen theorist’s response to data that could not be explained by the existing model was to create a new model by adding a new component or new pathway to the existing model. But we do not consider that this is invariably an appropriate response to new data. Specifically, we agree with Stone and Van Orden (1993) that when such data involve evidence that people can strategically vary the use they make of some cognitive processing system, it is inappropriate to propose that this is done by altering the functional architecture of the system; we do not believe that new strategies create new architectures. Instead, we have always (Coltheart, 1978; Coltheart & Rastle, 1994) explained strategy effects as arising through people’s capacities, not to alter the architecture of a processing model, but to vary parameters of the model such as the strength with which one module activates another, or rate at which activation rises, in response to task requirements. This is also how Stone and Van Orden conceptualized strategy effects: they refer to the process as one of “parametric control.”

The initial version of the logogen model (Morton, 1968, 1969, 1970) was described by Morton, once again, in box-and-arrow notation. It is shown in Figure 4(a).

The logogen system is a set of elements called logogens, one for each of the words (or morphemes) in the model’s vocabulary.

Logogens are evidence-collecting devices with thresholds.³ Evidence is collected from visual or auditory input, and when the amount of evidence collected by a word’s logogen exceeds that logogen’s threshold, information about that word in the cognitive system (e.g., its meaning) is accessed, and the word also becomes available as a response in the response buffer. The more frequent a word, the less evidence need be collected from visual or auditory input before threshold is reached, because each logogen has a resting level of activation whose value is proportional to the frequency of occurrence in the language of that logogen’s word. One of the empirical phenomena that Morton (1969) interpreted within the framework of this model was repetition priming, which was attributed to the lowering of thresholds within the logogen system. For example, when a visually presented word is read aloud, its entry in the logogen system will reach threshold, which will result in the threshold being set lower. Lowered thresholds slowly drift back to their original values over time. If the word is presented again before the return to the original threshold value is complete, this lowering of the threshold will facilitate the reading aloud of the word on this second occasion.⁴

However, some data from studies of priming refuted this first version of the logogen model. In this first version of the model, what is important for repetition priming is that the same response is generated on the two occasions. Therefore, if on the first occasion the task is picture naming and on the second occasion the task is reading aloud, repetition priming should still occur, because on both occasions the word’s logogen would be activated. However, Winnick and Daniel (1970) and Clarke and Morton (1983) found that there is no repetition priming between pictures and words except with very short intertrial intervals. This led Morton (1979) to propose a distinction between an input logogen system (responsible for word recognition) and an output logogen system (responsible for word production). If the site of repetition priming were the input logogen system, then picture naming (which does not activate the input logogen system) would not facilitate subsequent recognition of the word that is the picture’s name. Thus a second version of the logogen model (Morton, 1979) was developed in response to evidence from repetition-priming experiments. This second version of the model is shown in Figure 4(b).

This model predicts cross-modal repetition priming within the domain of words, that is, hearing a spoken word should prime subsequent recognition of its printed form, because on both occasions the word’s input logogen will exceed threshold. However, such priming does not occur except with very short intervals between prime and target. From this it follows that, if the site of repetition priming is the input logogen level, spoken word recognition and written word recognition must use different logogen systems, as proposed in the third version of the logogen model (Morton, 1979), shown in Figure 4(c).

³ Thus McClelland and Rumelhart (1981, p. 388) were not really correct in describing their IAC model as a logogen model. The word detectors in the IAC model are not logogens, because they are not thresholded devices.

⁴ Morton (1969) here was not concerned with priming that occurs only at very short intertrial intervals, but priming that lasts for minutes or even days.

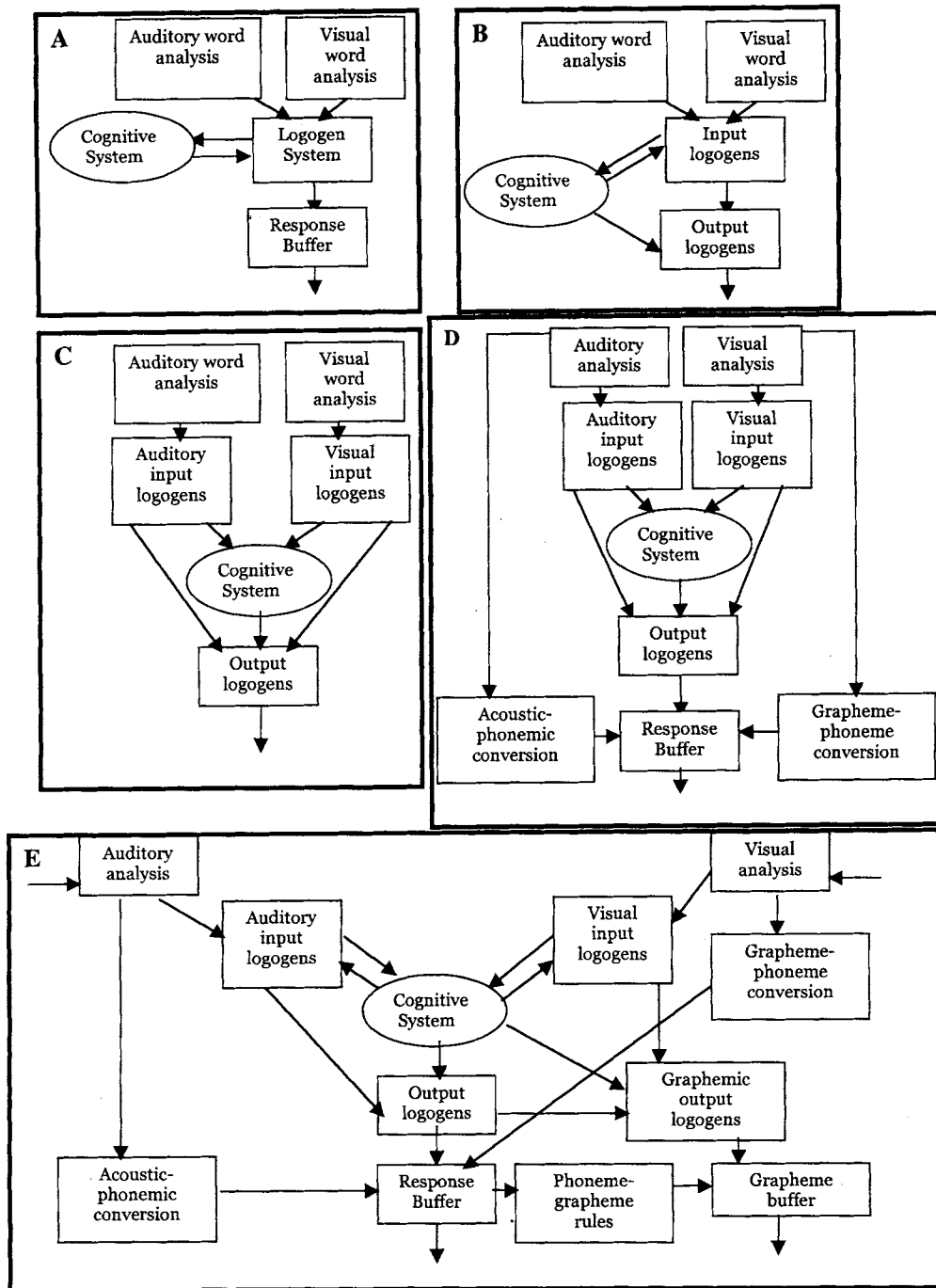


Figure 4. Five versions of the logogen model.

Of course, the idea that spoken words and written words are recognized by distinct word recognition systems is supported by a great deal of cognitive-neuropsychological data, some of it from the 19th century. In the condition known as *word-meaning deafness* (Bramwell, 1897; Howard & Franklin, 1988), printed words can be understood but spoken words cannot, even though hearing is adequate for the task of understanding speech (the patient can

repeat the words that she hears but cannot understand). The reverse holds for pure *alexia* (Coltheart, 1998; Déjerine, 1892), a condition in which spoken words can be recognized but printed words cannot, even though vision is adequate for this task (printed letters can be named).

This model, although capable of recognizing written and spoken words and also capable of producing words, is incapable of doing

anything with nonwords—for example, it can neither repeat nonwords nor read nonwords aloud. Because both tasks can be performed by people, an account is needed of how this nonword processing is done. Morton (1980) therefore embellished the logogen model by adding grapheme–phoneme and acoustic–phonemic routes, to produce a fourth version of the logogen model, shown in Figure 4(d).

Perhaps the distinction here between visual input logogens and auditory input logogens needs a parallel at the output level—that is, there should be two, not one, output logogen systems, one for speaking and another for writing/spelling. This is the fifth and final version of the logogen model (Morton, 1980), shown in Figure 4(e).

The distinction between phonological and orthographic lexicons is in any case strongly suggested by data from cognitive neuropsychology: Some people after suffering brain damage have an impairment of the ability to produce spoken words with relatively intact writing and spelling (Lhermitte & Derouesné, 1974), and other people have an impairment of writing and spelling with relatively intact ability to produce spoken words (Basso, Taborelli, & Vignolo, 1978). This double dissociation would be difficult to explain if the system for producing spoken words were the same as the system for producing written words.

In relation to Figure 4(e), Morton (1980) dryly observed: “The reader who does not like the notation of information processing models will be at a grave disadvantage here” (p. 125). He then went on to add that Figure 4(e) was if anything an oversimplification of what the real system for processing spoken and written language at the single-word level must be like. As our account of evolution of the logogen model has shown, all the complexities of the model shown in Figure 4(e) are motivated. If any box or any arrow were deleted from it, the result would be a system that would fail in at least one language-processing task at which humans succeed.

Beyond Logogens

In the preceding sections we have traced the historical evolution of the model of language processing shown in Figure 5, beginning with models developed by 19th-century cognitive neuropsychologists through theoretical work by Treisman (1961) and Morton (1961), which reintroduced the concept of the mental lexicon, to the development of the logogen model by Morton, which led to the architecture proposed in Figure 4(e).

The final step in this evolution was to *generalize*, by adopting the architecture of the model of Figure 4(e) but avoiding one or more specific theoretical commitments—for example, a commitment to the idea that the system operates according to logogen principles or a commitment to the idea that spelling-to-sound rules used by the nonlexical route for reading aloud are solely grapheme–phoneme correspondence rules.

This step was taken by Harris and Coltheart (1986), Patterson and Shewell (1987), and Ellis and Young (1988), all of whom proposed essentially the same model as the version of the logogen model shown in Figure 4(e), except that their models were no longer necessarily logogen models, even though the architecture of these models derived directly from Morton’s theoretical work. Figure 5 shows the model proposed by these authors.

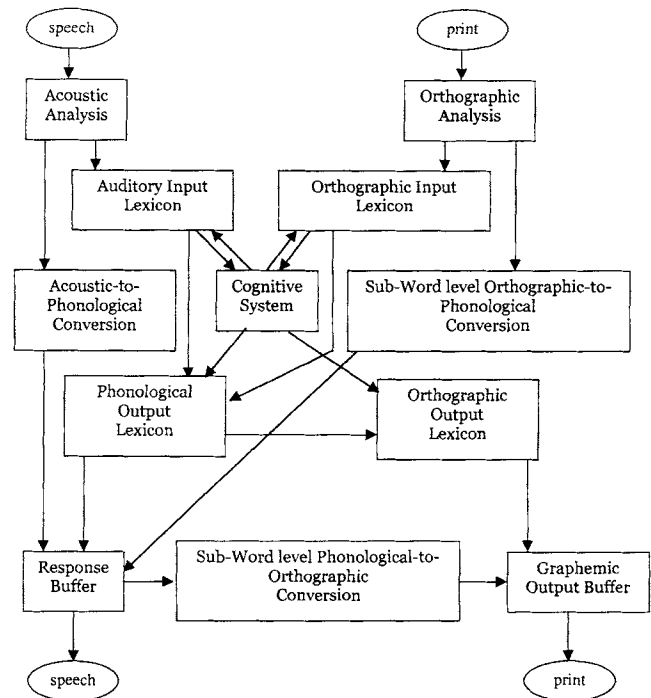


Figure 5. The model of language-processing proposed by Patterson and Shewell (1987).

Logogens are information-gathering devices with thresholds. To claim that the entries in, say, an orthographic input lexicon are logogens is going one step beyond the claim that one module of the language-processing system is an orthographic input lexicon whose elements are lexical entries corresponding to words. That is the sense in which the model of Figure 5 is a generalization of the model of Figure 4(e).

This is made clear by Morton (1982). In discussing the possibility that auditory word recognition may be done in a nonlogogen way, as proposed in the cohort model of Marslen-Wilson and Welch (1978), he noted,

If it turns out that the cohort model wins, then the consequences are clear. The logogen system would have to be modified such that the auditory input categorization system operates on cohort principles and would thus be different from the visual system in this respect. Again, the modularity of the approach prevents the consequences from being more severe than this. (pp. 100–101)

In previous pages we have identified a variety of findings, especially from the cognitive neuropsychology of language, that support this proposed architecture. There are many other findings from this domain of psychology that support the model; for a review of this literature, see Ellis and Young (1988) and Shallice (1988).

Certain theoretical questions were deliberately left unanswered regarding the model shown in Figure 5, so as to preserve its generality. However, as we indicated earlier, if one wants to turn a verbal model (such as the model in Figure 5) into a computational model—and that is our aim—commitments to specific an-

swers to such theoretical questions cannot be avoided. Hence we now consider what these commitments might be.

Thresholded or Cascaded Processing?

It is true by definition for logogen models that the processing going on in any module does not begin to affect subsequent modules at an early point in processing; activation is only passed on to the later modules after a threshold is reached in the earlier module. This is thresholded processing, an alternative to which is cascaded processing (McClelland, 1979). In systems that operate by cascaded processing, there are no thresholds within modules; as soon as there is even the slightest activation in an early module this flows on to later modules.

It is an open question whether the models of Harris and Coltheart (1986), Patterson and Shewell (1987), and Ellis and Young (1988) should be thresholded or cascaded models. Relevant empirical results are needed if this question is to be resolved.

Coltheart et al. (1993) drew attention to some such results in relation specifically to reading. Suppose that in these models processing is thresholded. If so, when a nonword is presented for reading aloud, no entry in the orthographic lexicon will reach threshold, so no information or activation will emerge from that lexicon to affect subsequent phonological processing. That means there will be no lexical influences on nonword reading aloud. However, Glushko (1979), Kay and Marcel (1981), Rosson (1983), and others have reported various ways in which reading a nonword aloud appears to be affected by activity in the lexical route. For example, Glushko (1979) found that nonwords that were inconsistent with real words (e.g., *heaf*; cf. *deaf*) yielded longer naming latencies than nonwords that were not inconsistent with real words (e.g., *hean*). This suggests that at some point during the processing of *heaf*, the phonology of *deaf* must have been activated, a conclusion that is inconsistent with a thresholded version of the model.

Patterson and Morton (1985) proposed a way in which a thresholded model may be reconciled with Glushko's (1979) results: The nonlexical rule system does not operate solely with grapheme-phoneme rules but also has rules relating orthographic bodies such as *eaf* or *ean* to their pronunciations. Because there are two pronunciations for *eaf* (as in *deaf* and as in *leaf*), perhaps the *eaf* rule is weaker or more ambiguous than the *ean* rule, and so the nonlexical route is worse for *heaf* than for *hean* in a purely thresholded system.

However, as Patterson and Morton (1985) acknowledged, this way of accounting for Glushko's (1979) results cannot explain the results of Rosson (1983), who found in a nonword reading task that the pronunciation of the vowel of the nonword *louch* was biased toward the "couch" pronunciation when the preceding item was *sofa* (semantically related to *couch*) and was biased toward the "touch" pronunciation when the preceding item was *feel* (semantically related to *touch*). This result seems to demand the conclusion that, when the input is the nonword *louch*, the orthographic lexical entries for *couch* and *touch* are not only activated themselves but also that this lexical orthographic activation then generates some activation of phonology. That is inconsistent with a thresholded interpretation of the architecture of the model, which is why Coltheart et al. (1993) adopted the view that the system operates in a cascaded fashion (hence the "C" in DRC).

Further and strong evidence for such cascaded processing by the lexical route was subsequently provided by Peereboom and Content (1995) and Job, Peressotti, and Cusinato (1998), because both sets of authors interpreted their results as showing that the orthographic neighbors of a visually presented letter-string nonword are sufficiently activated during reading aloud to influence the computation of the phonology of the to-be-read-aloud nonword.

In a recent and invaluable review of effects of phonology on visual word recognition and reading aloud, Frost (1998) concluded that during reading "phonology is always partly assembled and always partly lexical. It is always activated but not necessarily fully specified" (p. 95). Frost explicitly stated that there is nothing in his review that is inconsistent with the dual-route view: In response to his general conclusion, he says, dual-route theorists could "keep the dual-route view but would reconsider the description of phonological processing within the model" (p. 95). That is what we did by introducing cascaded processing into the dual-route view. As we show below, when the DRC model is reading aloud or performing the lexical decision or Stroop tasks, phonology is always partly assembled and always partly lexical, and always activated but not necessarily fully specified—exactly as prescribed by Frost.

Toward a Computational Model

The model in Figure 5 is meant to offer an account of how all language-processing tasks at the single word and nonword level are performed. This article is specifically concerned, however, with only a subset of such tasks, those involving visual word recognition and reading aloud. Hence the architecture with which the present article is specifically concerned is a subset of the architecture of the model of Figure 5. This architecture for visual word recognition and reading aloud is shown in Figure 6.

Our task is to take this verbal model of visual word recognition and reading aloud, which as we have indicated is well supported by a variety of forms of data, and to turn it into a computational model. We discussed the motivation for doing this in the first part of this article. Initial work along these lines was reported by Coltheart et al. (1993) and Coltheart and Rastle (1994), and Rastle and Coltheart (1998, 1999a, 1999b) have described various results of experiments with human readers and the simulations of these results that have led to the current version of the DRC model.

As we have said, the exercise of deriving a computational model from a verbal model always compels the modeler to make many specific theoretical choices. Our adoption of the IAC model as our starting point committed us to local rather than distributed representations; Grainger and Jacobs (1998) and Page (in press) provide much detailed justification for this choice in relation to cognitive modeling. A second theoretical choice we have already discussed is our use of cascaded rather than thresholded processing. Another such choice made by Coltheart et al. (1993; see also Coltheart, 1978, 1985) was that the spelling-to-sound rules that characterize the nonlexical reading route operate at only one level of phonology, the phoneme. Thus these rules are grapheme-phoneme correspondence (GPC) rules, whereby the term "grapheme" we mean a letter or letter sequence that corresponds to a single phoneme, such as the *i* in *pig*, the *ng* in *ping*, and the *igh* in *high*.

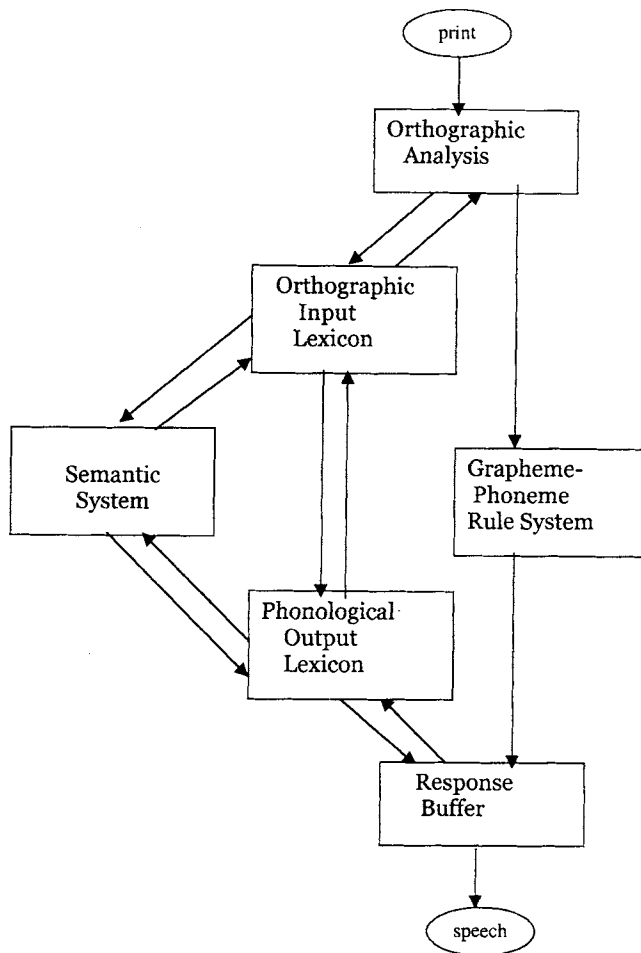


Figure 6. Basic architecture of the dual-route cascaded model of visual word recognition and reading aloud.

McClelland (1993) proposed a set of principles to which computational models should adhere, the GRIN principles. The A in his acronym stands for "adaptive," and refers to models whose structure is learned under some learning algorithm. We have explained why we prefer not to use learning algorithms to develop architectures of computational models. However, we do adhere to the other four principles advocated by McClelland (1993), and so we are guided by the GRIN principles: Activation in our model is graded (rather than all-or-none), it is random (activation update calculations can be noisy rather than deterministic, though in fact we have made little use so far of this feature of the model), it is interactive (activation flows between adjacent modules in both directions),⁵ and it is nonlinear (this is a property of the equations by which unit activations are computed from unit inputs).

DRC: A Dual Route Cascaded Model of Visual Word Recognition and Reading Aloud

As discussed above, a part of the DRC model⁶ is a generalization of the IAC model, and one reason for making this choice was the success enjoyed by the IAC model in accounting for human

data in Reicher–Wheeler experiments. Another reason for the choice was that the IAC is a cascaded model, and, as also discussed above, the results of Glushko (1979), Kay and Marcel (1981), Rosson (1983), and others favor such a model over thresholded models.

The IAC model applied only to four-letter words; its DRC version applies to words from one to eight letters in length, and that is the sense in which the IAC model has been generalized.

General Features of the DRC Model

The overall architecture of the DRC model is illustrated in Figure 7.

The model consists of three routes, the lexical semantic route, the lexical nonsemantic route, and the GPC route. Each of these routes is described individually below. Each route is composed of a number of interacting layers. These layers contain sets of units. The units represent the smallest individual symbolic parts of the model, such as words in the orthographic lexicon or letters in the letter unit layer.

There are two ways in which the units of different layers interact. One is through inhibition, where the activation of a unit makes it more difficult for the activation of other units to rise. The other is through excitation, where the activation of a unit contributes to the activation of other units. Units may also interact on the same level through lateral inhibition. In Figure 7, excitatory links between units are represented by arrows, and inhibitory links between units are represented by circles. Adjacent layers of the model communicate fully in both directions in both excitatory and inhibitory ways. Exceptions are the following:

- Communication between the orthographic lexicon units and phonological lexicon units are only excitatory and only one-to-one, except in relation to homophones and homographs, as discussed below.
- Communication between the feature and letter layers is in one direction only (from features to letters), as in the IAC model.

Layers that have position-specific coding (feature layer, letter layer, phoneme layer) are made up of different subsets of units, one subset for each position in the input string or output string. Here lateral inhibition occurs only within but not between the position-specific subsets of units.

The visual feature units consist of eight different subsets representing the eight possible input positions. The feature sets are based on the 16-feature font of Rumelhart and Siple (1974), where the set consists of individual features that are set to on or off (1 or 0) if they correspond to the properties of the letter that is input. When an input position does not contain a letter at a particular position, that is, contains the null character, all feature units within the subset of feature units for that position are turned off.

⁵ One departure from this principle is that nonlexical activation between the letter and phoneme levels occurs in only one direction in the model we describe here: There is currently no nonlexical activation from phonemes to letters. In the section of this article dealing with spelling we describe how this violation of the GRIN principles is being remedied in current work.

⁶ An executable version of this model is available for download at <http://maccs.mq.edu.au/~max/DRC/>.

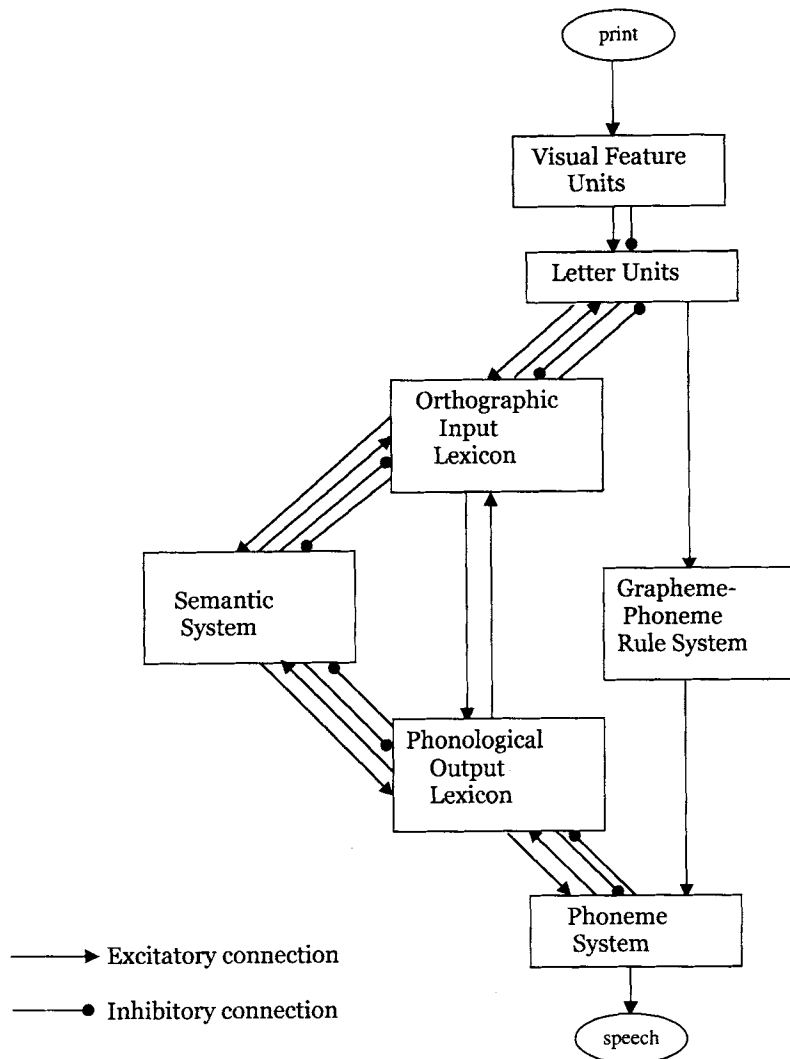


Figure 7. The dual-route cascaded model of visual word recognition and reading aloud.

The letter level also uses eight different subsets, with a structure similar to the feature set. Instead of each subset containing individual features that can be turned on or off, however, each subset contains units for the entire set of letters that can occur, that is, a unit for each of the 26 letters plus one for the blank letter. Lateral inhibition occurs at this level, within but not between each of the eight subsets.

The phoneme units are similar in organization to the letter units, except that each of the eight subsets contains units for the 43 phonemes plus a unit for the blank phoneme.

Units within the orthographic lexicon are not position-specific and each unit inhibits all others. The same is true for units within the phonological lexicon.

The orthographic lexicon contains 7,981 units, one for each of the monosyllabic words in the CELEX database (Baayen, Piepenbrock, & van Rijn, 1993), except for a few rare and foreign words that have been discarded.⁷ For each of these orthographic lexical units there is a corresponding phonological

lexical unit in the phonological lexicon. Heterographic homophones such as SO and SEW have separate units in the orthographic lexicon but a common unit in the phonological lexicon. Homographic heterophones such as LEAD have a single unit in the orthographic lexicon but a separate unit in the phonological lexicon for each of their pronunciations. The phonological lexicon contains 7,131 units (fewer than the orthographic lexicon because there are more heterographic homophones than there are homographic heterophones).

⁷ Plaut et al. (1996) excluded most but not all polymorphic words from the set of monosyllabic words on which they trained their model, and Zorzi et al. (1998) followed suit. Thus for both of these models fewer than 3,000 of the approximately 8,000 monosyllabic words of English can be used in simulations; the remaining 5,000 or so words are effectively nonwords for these models.

Activation Dynamics

Although each layer of the model uses a different set of units, corresponding to the different representation of each layer, each unit individually is governed by the same general activation dynamics. These activation dynamics were originally specified in McClelland and Rumelhart (1981), although, as can be seen below, they have been altered slightly.

There are three factors used when calculating how the activation of a unit changes over a time step. These are the unit's previous activation, how quickly this activation decays, and what the net input of activation into the unit from other units is. This can be represented by the following formula, similar to that of McClelland and Rumelhart (1981):

$$a_i(t + \Delta t) = a_i(t) - \theta_i(a_i(t)) + (\varepsilon_i(t) \times \text{activation rate}) \quad (1)$$

In the formula, a_i represents the unit's current activation, θ_i is a parameter that controls the level of decay, and $\varepsilon_i(t)$ represents the net activation input into the unit after being scaled to a value between 1 and 0. The decay parameter θ_i is set between 0 and 1. It reduces the new activation by a proportion of the old activation. The activation rate is a parameter used to scale the speed at which the activation of the unit changes with respect to the difference between its current activation value and the scaled net input. Small values cause the activation level to change more slowly; high values cause the activation level to change more quickly.

The formula departs from that used for the IAC model in two ways. First, an additional activation rate scaling parameter is available (though in practice we have not made any use of this parameter). Second, in the IAC model a baseline activation level is associated with each unit; for the DRC model, this baseline activation level was moved to the formula modifying the net input (Equation 6).

Unit activations remain between 0 and 1. This is achieved by the following formula:

$$a_i(t) = \begin{cases} 1 & \text{if } a_i(t) > 1 \\ a_i(t) & \text{if } 0 \leq a_i(t) \leq 1 \\ 0 & \text{if } a_i(t) < 0 \end{cases} \quad (2)$$

The amount of net input into the unit at a given time, $n_i(t)$, is a function of the inhibitory and excitatory inputs, as in McClelland and Rumelhart (1981):

$$n_i(t) = \sum_j \alpha_{ij} e_j(t) - \sum_k \gamma_{ik} i_k(t). \quad (3)$$

The first half of the equation specifies the excitatory component of the total net input. In this half, $e_j(t)$ represents the activation produced by an excitatory unit j , and α_{ij} is a weight constant associated with the link from unit i to unit j . This excitatory communication is represented by lines terminated by arrowheads in Figure 7. The second half of the equation specifies the inhibitory component of the net input. In this part i_k represents the activation from an inhibitory unit, and γ_{ik} is a weight constant associated with the link from unit i to unit k . This inhibitory communication is represented by lines terminated by circles in Figure 7. The constant weights α_{ij} and γ_{ik} , associated with the communications between two units, are always the same for all connected units between any two communications between units in two adjacent layers. For

example, the α_{ij} values for all communications between the j units in the orthographic lexicon and the i units in the phonological lexicon were always set to 0.2 in the reported simulations.

Once the net input has been calculated, it is squashed such that unit activations can fall only within the range 0 to 1. The two equations that perform this function (Grossberg, 1978) are

$$\varepsilon_i(t) = n_i(t)(1 - a_i(t)) \quad (4)$$

when the net input, n_i , is above or equal to 0, and

$$\varepsilon_i(t) = n_i(t)(a_i(t)) \quad (5)$$

when the net input, n_i , is below 0.

The Three Routes

Lexical Nonsemantic Route

The lexical nonsemantic route of the DRC model generates the pronunciation of a word through a sequence of processes: The features of the word's letters activate the word's letter units (in parallel across all letter positions), these letters then activate the word's entry in the orthographic lexicon, this word entry in the orthographic lexicon then activates the corresponding word entry in a phonological lexicon, and that word entry in the phonological lexicon then activates the word's phonemes (in parallel across all phoneme position).

As mentioned earlier, the lexical orthographic part of the model (the feature and letter levels and the orthographic lexicon) is a generalization of the IAC model (Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981). The IAC model operates only for words of a fixed length; the DRC model operates with words of any length up to and including eight letters. The model achieves this by adding to each set of 26 letter detectors a blank-letter detector that is activated when there is no letter in that particular position in the letter string. For all eight possible input positions, there is a set of 14 feature-present units (the font used has 14 possible features) and a set of 14 feature-absent units, as in the IAC model—that is, there is a complement-coded representational scheme (Carpenter, Grossberg, Markuzon, Reynolds, & Rosen, 1992). Suppose the model is given a string of, say, five letters as input. For Position 1, the features present in that letter turn on the appropriate feature-present units, and the feature-absent units corresponding to those features not present in the Position 1 letter are also turned on. The same happens with respect to the next four positions. For Positions 6–8, which are not occupied by letters, all 16 feature-present units will be off, and all 16 feature-absent units will be on; that will have the effect of activating the blank-letter detector within the three sets of letter-detectors for those three positions.

Whether a letter causes excitation or inhibition of a unit in the orthographic lexicon is determined as follows: A letter in the N th set of letter units excites all units in the orthographic lexicon for every word that contains that letter in the N th letter position of the word and inhibits all other units in the orthographic lexicon.

Units in the orthographic lexicon are frequency-sensitive; if all other factors are held constant, the activation of high-frequency words rises more quickly than the activation of low-frequency words. To achieve this effect, a constant value is associated with each unit in the lexicon, as was done in the IAC model. This value

is calculated by taking the \log_{10} frequency of the word, dividing by the \log_{10} frequency of the highest frequency word and subtracting 1 to get a standardized \log_{10} frequency value scaled between -1 (corresponding to the lowest frequency word) and 0 (corresponding to the highest frequency word). This value is multiplied by a frequency scaling parameter that takes on a value between 0 and 1 . This parameter controls the dominance of the frequency effect. Whenever activation is calculated for each unit in the orthographic lexicon, this constant is added to the net input, n_i . This means that Equation 3, which was used for calculating the net input into a unit, needs to be modified to incorporate this difference. The net input to unit i in the orthographic lexicon becomes

$$n_i(t) = \sum_j \alpha_{ij} e_j(t) - \sum_k \gamma_{ik} i_k(t) + CFS_i, \quad (6)$$

where CFS_i is a constant frequency scaling variable that equals

$$CFS_i = \left(\left(\frac{\log_{10}(\text{frequency}_i)}{\log_{10}(\text{max frequency in lexicon})} \right) - 1 \right) \times \text{frequency scaling}. \quad (7)$$

Activation of a unit in the orthographic lexicon is transmitted to the corresponding unit in the phonological lexicon. At present, a unit's CFS_i value in the phonological lexicon is set to the same value as its corresponding orthographic frequency. A possibility for future implementations of the DRC model would be to change this to spoken word frequency.

Finally, there is activation of a word's phoneme units by its activated entry in the phonological lexicon. For example, the unit for /bi:t/ activates the phoneme unit for /b/ in the first phoneme set and inhibits all other phoneme units in that set; it activates /i:/ in the second phoneme set and /t/ in the third phoneme set, and inhibits all other phonemes in these sets. As at the letter level, each set of phoneme units also contains a blank phoneme unit; if a word has six phonemes, then it excites the blank phoneme unit in the seventh and eighth phoneme sets and inhibits all the other units in the seventh and eighth phoneme sets.

GPC Route

Derivation and nature of the rules. In earlier work (Coltheart et al., 1993), we developed a rule-discovery algorithm that learned a set of GPC rules from exposure to the database of about 3,000 word spellings and their pronunciations compiled by Seidenberg and McClelland (1989). Our primary aim in developing that algorithm was not to obtain a set of GPC rules, but to show that there was enough information in that database for a procedure to be learned that would then be very accurate at reading nonwords, as was the case with the learned set of GPC rules. From this it follows that the poor nonword reading of the SM model could not be attributed to impoverishment of the database on which that model was trained. Having established that, we did no more work on computational learning of GPC rules, for the reason we gave earlier that unless the learning procedure itself is known to be psychologically real, it may not be able to learn what people learn. We have no reason to believe that the rule-discovery algorithm we used for GPC learning is any closer to the way children learn than is backpropagation (see also Rastle & Coltheart, 1999b).

The GPC route converts a letter string into a phoneme string by using grapheme-phoneme correspondence rules. The route is restricted to using rules where a set of letters map onto a single phoneme, with one exception. The exception is that the letter x is converted into the two-phoneme sequence /ks/. The GPC rules used by the model are those listed in Appendix B of Rastle and Coltheart (1999b).

Rastle and Coltheart (1999b) selected these GPC rules on purely statistical grounds. For any grapheme, the phoneme assigned to it was the phoneme most commonly associated with that grapheme in the set of English monosyllables that contain that grapheme, taking position in the word into account if necessary. As it happens, almost all of the GPC rules selected in this way were ones that the original GPC learning algorithm had yielded. We think of this set of rules as a set of hypotheses about what GPC rules people know.

Some of the GPC rules are context-sensitive—for example, c is translated to /s/ when the following letter is e , i , or y , and is otherwise translated to /k/. Some of the GPC rules are position-sensitive—for example, there are three rules for the grapheme y . In initial position, it is given the phoneme /j/, in medial position it is given the phoneme /ʃ/, in final position it is given the phoneme /ai/. Another category of GPC rules is what we have called “output rules.” Some of these are phonotactic rules; the remainder are morphophonemic rules. Some phonemic sequences generated by the rules are phonotactically illegal; for example, the rules translate the letter-string *enk* as /enk/. This is corrected by an output rule that converts the phoneme /n/ to the phoneme /ŋ/ whenever the following phoneme is /k/. However, purely phonotactic rules are not enough here; they fail to capture a regularity associated with the phoneme /s/ in final position. When this is spelled s , and so is, or could be, the plural morpheme, the phoneme must be converted to /z/ when preceded by any of the phonemes /ɑ: ei ai ə: oi ou iə au eə uə ð ŋ b d g ɪ: l m n u: v/, even though this is not always required by phonotactic considerations (see e.g., Pinker & Prince, 1988, p. 103). For example /ns/ is phonotactically legal, as in *dense*, so /dens/ is a phonotactically legal translation for *dens*, but the morphophonemic rule controlled by the fact that the last letter here is s disallows this phoneme sequence (when it is spelled this way).

Operation of the GPC route. Visual features and corresponding letter units are activated just as with the lexical nonsemantic route because the feature and letter levels are common to the two routes. Then the GPC route operates in the following manner. First, a set number of cycles (10 cycles in the simulations reported in this article) passes before the GPC route begins to operate on the first letter of the input. The set of rules is searched until an appropriate rule is found to convert that letter to a phoneme, and that phoneme's unit in the phoneme system then receives some activation (which adds to the activation it receives from the lexical route). On each subsequent processing cycle, activation is contributed to that phoneme unit in the same way. A constant number of cycles later (17 cycles in the simulations reported here), the next letter becomes available to the GPC route, so that it is now translating the first two letters in the input string. The GPC route uses its rules to translate this two-letter string into a phoneme or phonemes. If this string is a grapheme, such as *ph*, it will be translated into just one phoneme by the rules; if it is not a grapheme, such as *pr*, it will be translated into a sequence of two

phonemes. This process continues, with one letter being added every 17 cycles, until either the letter string is named or the final position in the letter units is reached. Thus the GPC route assembles the letters into phonology serially, letter by letter.

In making the assumption that the nonlexical route operates serially and from left to right, rather than right to left, or in parallel, we followed Forster and Davis (1991). They showed that a briefly presented masked prime facilitated reading aloud of a subsequent target word when mask and prime shared just one letter or phoneme, but only when this was the leftmost letter or phoneme in the prime. Their explanation was that "construction of the nonlexical response to the prime operates in a strict left-to-right manner (see Meyer, 1988), so that by the time the target is presented, the articulation of just the onset of the prime has been planned" (p. 19). Subsequently we have obtained further evidence from other paradigms for this left-to-right operation of the nonlexical route: the positional sensitivity of the regularity effect (Rastle & Coltheart, 1999b), the effect of position of irregularity in filler stimuli on strategy effects in reading aloud (Rastle & Coltheart, 1999b), and the positional sensitivity of the phonological Stroop effect (Coltheart, Woollams, Kinoshita, & Perry, 1999).

Rule selection by the GPC route. An important part of the assembly process is the way the rules are matched to information available to the GPC route. The GPC route works by trying to find a rule whose letters match the leftmost letters available to the route. When it does find a match, the letters that matched are removed from the letter string, and the route tries to find another rule that matches the leftmost letters that remain. Once all of the letters are matched, the phonology generated is subjected to the phonological output rules. On completion of this process, the phonology generated adds activation to the positionally corresponding phonemes in the phoneme unit layer (a layer common to the nonlexical [GPC] and lexical routes) as described above.

The way in which the rules are picked when translating a letter string is not simply by making a random choice from those that could conceivably match. The rules are picked based on the serial order of the rule list. This is done by dividing the rule list into a set of smaller sublists. The order of these sublists is based on the number of letters from which the grapheme is formed and whether the letters are context sensitive. Rules with larger graphemes are tested before rules using smaller graphemes. The specific order of testing is multirules (rules with more than two-letter graphemes), then context-sensitive rules, then two-letter grapheme rules, then the one letter to multiple phonology rule (the letter *x*) and finally single-letter rules.⁸

How the GPC route activates phonemes. The amount each phoneme is activated is a function of two factors. One is a GPC activation parameter, set between 0 and 1. This parameter controls the overall strength of the route. The second is the activation of the letters that the rule matched. In this case, the activation of each assembled phoneme is the average of the activations of the letters from which it was generated. For example, if the system had generated the /θ/ phoneme from the letters *t* and *h*, the amount of activation of this phoneme would be the activation of the letters *t* and *h*, divided by two. If a match occurs using a rule that uses context-sensitive letters, the context-sensitive letters are not entered into the averaging process. The overall activation added to the net input of the phoneme units is then the product of the two factors. It is important to note that this process occurs for both

letters and orthographic word terminators. When orthographic word terminators are encountered, they are assembled into phonemic word terminators. The assembly of these end terminators fundamentally alters the properties of the model because they boost the activation level of the phonemic word terminators at the phoneme level. As is discussed later, these are used to determine when word read-out occurs.

Lexical Semantic Route

As yet, the semantic part of the model has not been implemented (though see Coltheart et al., 1999, for a small partial implementation). There are a number of possibilities that could be pursued, however. One that seems particularly promising is to use a semantic system similar to that of Dell, Schwartz, Martin, Saffran, and Gagnon (1997), who showed how an interactive activation model of phonology, very similar to the phonological half of the lexical route of the DRC model, could be combined with a system of semantic representations. They then showed that such a model could produce a good fit to a number of behavioral results obtained from aphasic and nonaphasic people. Another approach (Watters & Patel, 1998) has used WordNet (Miller, 1985) as a source for extracting semantic representations and has interfaced a semantic system based on such representations with the DRC model as described in this article.

How the Whole Model Operates

On Cycle 1, the visual feature units are clamped with the features corresponding to the input letter string. This clamping means that on Cycle 2, activation from the feature level will reach the letter level. On Cycle 3 activation will reach the orthographic lexicon and will also be fed back to the letter level, and so on. This process of cascaded activation eventually leads to a build-up of activation in the phonemic layer, and of course to activation feeding back from the phoneme layer to the letter layer. At the same time, as parameters allow, the GPC system will be contributing activation to the phoneme layer. In the present version of the model, the nonlexical route is only feed-forward, but a version of the model in which that route operates bidirectionally is being developed (see the section on Spelling at the end of this article).

How Does a Letter String Get Named?

The model is considered to have determined the pronunciation of a monosyllabic letter string when it has activated (to some criterion of satisfaction) all of the phonemes of that letter string. The processing cycle on which that criterion is achieved is the model's naming latency for that letter string.

The procedure that the DRC model uses to decide whether a pronunciation has been determined is a left-to-right scan across the

⁸ This method of accessing the rules prevents the GPC route from translating a three-letter grapheme such as IGH into three phonemes, because it encounters the GPC rule for IGH before the GPC rules for I, G, and H. There are various other ways in which this problem could be solved, and we are not committed to the method we adopted, which is a mere implementational detail; all that we are committed to is that in the model phonemes are correctly generated from letter strings by the application of GPC rules.

sets of phoneme units. This procedure is carried out at the end of every processing cycle. Each set of phonemes, starting from the first set, is sequentially examined, and the phoneme in the set that is maximally active is identified.

If

(a) a phoneme set is reached in which the phoneme with the highest activation in that set is the phonemic word terminator (blank phoneme),

and if also

(b) the level of activation of the blank phoneme in that set is above the minimum naming threshold (the value of which is 0.43 in the simulations reported here),

then

(c) the model assesses whether the maximally activated phoneme in each of the preceding phoneme sets has an activation above the minimum naming threshold (we used a value of 0.43 here). If this is not so, then pronunciation is considered not yet to be known, and the next cycle of processing begins. If, on the other hand, the maximally activated phoneme in all of the preceding phoneme sets is greater than the minimum naming threshold, then pronunciation of the letter string is considered to be known. The actual pronunciation consists of the maximally activated phoneme in each of the phoneme sets. That will be the DRC model's pronunciation of the letter string, and the model's naming latency will be the number of the cycle at which this event occurred.

Note that this left-to-right scanning process occurs between processing cycles and so does not contribute to DRC's naming latencies. Thus this scanning process does not, for example, cause any effect of string length in DRC's naming latency.

How Simulations are Executed With the DRC Model

Table 1 lists the various parameters of the DRC model. Any of the parameters can be given a value by typing it into the appropriate place in a parameter window that is part of the model's interface. Then the model can be run to perform the naming task or the lexical decision task on a particular set of letter strings under this particular parameter set. Subsequently the model's naming or lexical decision latencies can be analyzed to see whether they correspond to the data from experiments with human readers. Suppose, for example, we were seeking to discover whether in reading words aloud the DRC model shows the interaction between frequency and regularity that is shown in the human data (Paap & Noel, 1991). This would be done by submitting to the DRC model the actual high- and low-frequency regular and irregular words that were used in the human experiment, and carrying out a 2×2 analysis of variance (ANOVA) on the resulting DRC naming latencies. The specific values of the parameters listed in Table 1 are essentially those developed by Rastle and Coltheart (1999b) and used in all of the simulations described later.

Searching Parameter Space

Of course, even if there existed a set of parameters under which DRC would behave in exactly this way, there will be very many sets of parameters under which it will not. If, with the first set of parameters we use, the DRC latencies do not show this effect, where do we go next?

Table 1

Parameters Used in the DRC Simulations of Reading Aloud

Parameter	Value
General Parameters	
Activation rate	.2
Frequency scale	.05
Reading-aloud criterion	.43
Feature Level	
Feature to letter excitation	.005
Feature to letter inhibition	-.15
Noise	.00
Letter Level	
Letter to orthographic excitation	.07
Letter to orthographic inhibition	-.435
Letter to letter inhibition	.00
Noise	.00
Decay	.00
Orthographic Lexicon	
Orthographic to phonological excitation	.20
Orthographic to letter excitation	.30
Orthographic to letter inhibition	.00
Orthographic to orthographic inhibition	-.06
Noise	.00
Decay	.00
Phonological Lexicon	
Phonological to phoneme excitation	.14
Phonological to phoneme inhibition	.00
Phonological to orthographic excitation	.20
Phonological to phonological inhibition	-.07
Noise	.00
Decay	.00
Phoneme Level	
Phoneme to phonological excitation	.04
Phoneme to phonological inhibition	-.16
Phoneme to phoneme inhibition	-.15
Noise	.00
Decay	.00
GPC Route	
GPC to phoneme excitation	.055
Cycles before route begins	10
Cycles before next letter accessed	17

One in-principle approach is the following. There are 31 parameters in the model, so there is a 31-dimensional parameter space. Each point in this space will generate a set of latencies in which the Frequency \times Regularity interaction has a certain value. We could define as a target a simulation in which DRC's high-frequency latencies showed no regularity effect, whereas the low-frequency latencies show a significant regularity advantage. We could then start at some random point in this 31-dimensional parameter space, and use some kind of gradient descent algorithm to trace a path through this space toward a point representing a set of parameters that yield the desired interaction—if such a set exists.

We have not adopted this method, for two reasons. First, the parameter space is so large that it would take an impossibly long time to derive parameters to simulate just one set of human data. Second, we would in any case not be interested in an approach in which each set of human data is simulated with a different set of DRC parameters. Our aim instead (with some exceptions noted later) has been to find just one set of parameters that, unchanged, simulate a wide variety of sets of human data.

Hence we adopted a different method for seeking a parameter set under which the DRC model successfully simulates data from

experiments with human readers. The most delicate issue with the DRC model is to try to set an appropriate balance between the two routes of the model. For very many of the possible parameter sets, the model will read all of the words in its vocabulary perfectly, but will make many errors with nonwords (typically lexicalization errors). Here the lexical route is too strong. For many other possible parameter sets, the model will read nonwords (and regular words) perfectly, but will make many errors with exception words (typically regularization errors). Here the lexical route is too weak.

What is needed is a set of parameters under which DRC reads all exception words and all nonwords correctly. Of course, there is absolutely no guarantee that such a parameter set even exists for the DRC model. That is one of the senses in which the model is falsifiable.

So a sensible course was to use exception words and nonwords when seeking the desired parameter set. One can do better than that. For reasons discussed below (and see Rastle & Coltheart, 1999b), the reading aloud of exception words by the model is harder when these are low in frequency, and hardest of all when the exception words are not only low in frequency but also have an irregular grapheme-to-phoneme correspondence at the initial phoneme. So a word like *chef*—a low-frequency first-position irregular word—is a particularly difficult one for the model.

As for nonwords, there is also a class of these that is particularly hard for the model to read. These are nonwords that have a number of orthographic neighbors all of which differ from the nonword at the last letter position. One example is *starn*, whose neighbors include *start*, *stars*, and *stark*. As discussed above, a nonword excites its neighbors in the orthographic lexicon; these words will excite their phonological forms in the phonological lexicon and, hence, their phonemes at the phoneme level. As also discussed above, the letter *n* of *starn* will not begin to be processed until the $(10 + 4 \times 17) = 78$ th processing cycle. During those 78 processing cycles, the word neighbors *start*, *stars*, and *stark* will build up the activations of the incorrect phonemes /t/, /z/, and /k/ in the fourth phoneme set, and these already-activated phonemes will inhibit the activation of the correct phoneme /n/ when the nonlexical route begins to deliver it to the phoneme level.

To overcome this hostile reception, the nonlexical route will have to be able to deliver strong activation to the phoneme level. But strong nonlexical activation will be very dangerous for the exception word *chef*, because such activation will begin relatively early (on Cycle 10) and so will activate an erroneous phoneme before a great deal of lexical activation of the correct phoneme has been generated.

Hence if we just consider the two items *chef* and *starn*, the problem is clear: to get *starn* right, the nonlexical route must deliver activation above a certain level; to get *chef* right, the nonlexical route must deliver activation below a certain level; and there is no guarantee that a level of nonlexical activation exists that is compatible with both of these requirements. Exactly the same point applies to lexical activation.

Our odyssey through parameter space thus consisted of running exception word–nonword pairs such as this one—indeed, often exactly this one—seeking a point in parameter space at which the DRC model reads both members of the pair correctly. This took a very long time indeed, even though the voyage was by no means a blind one, because analysis of the model's behavior often gave

guidance as to which of its parameters was particularly poorly set and hence should be reset. Table 1 shows the values of the parameters eventually arrived at in this way in the work of Rastle and Coltheart (1998, 1999a, 1999b) and used in the various simulations of reading aloud reported later.

Evaluations of the Model

When all 7,981 words in DRC's orthographic lexicon were presented to the model for reading aloud using this parameter set, 7,898 words were read correctly: the mean naming latency for these items was 79.694 cycles ($SD = 6.244$ cycles; range = 59–116 cycles).

Of the 83 incorrectly read words, 56 were heterophonic homographs such as *BOW*. In every case, DRC's reading was the pronunciation of the more frequent member of the homophone pair. Hence these responses were not actually errors. The remaining 27 word errors were regularization errors for irregular words; 26 of these were irregular in the first position (e.g., *isle*, *heirs*) and 1 (*baths*) was irregular in the second position.⁹

There is an important point here. DRC made these 27 regularization errors because it was run as a simulation of speeded reading aloud: the critical activation level for phonemes was set low (0.43) so as to produce rapid reading aloud. When this value is set at a higher value (any value over 0.70), 26 of these 27 irregular words are read correctly by the model; this is simulation of reading aloud at leisure. Human readers behave the same way; the regularization errors they make in naming-latency experiments are not seen when they are reading these irregular words aloud at their leisure. The DRC model is thus capable of simulating both of these two different modes of human reading aloud, the speeded mode and the at-leisure mode. Other computational models of reading aloud have no natural way of simulating these two modes.

Thus there is only a single word in DRC's 7,981-word vocabulary that the model is incapable of reading aloud at its leisure, given the standard parameters. This is the word *czars*. A combination of factors conspires against the model here. First, this word is irregular in the first phoneme position, and we have explained above why this is relevant. Second, the correct pronunciation /zaz/ has a high frequency phonological neighbor, namely /kaz/, that is, *cars*. Third, this competitor gets support from the fact that the first letter of *czars* translates to /k/, which is the first phoneme of *cars*. The joint effect of these three factors is that *czars* is read as /kaz/ by the model, no matter how high the criterion for pronunciation is set. Perhaps a parameter set that gave a little more weight to the lexical route might permit correct reading of this word, but we were satisfied with a word-reading accuracy of 99.987%.

Word reading by the DRC model is thus almost perfectly accurate. What of nonword reading? We tested this by taking a random sample of 7,000 three- to seven-letter monosyllabic nonwords from the ARC Nonword Database (Rastle, Harrington, Coltheart, & Thomas, 2000) and having the model read these aloud using the standard parameter set. The error rate here was 1.07%,

⁹ BATHS is regular in American English but not in British English or Australian English, where BATS and BATHS have different vowels (short vs. long); the rule-based pronunciation for the grapheme A is as the vowel in BATS, not the vowel in BATHS. Thus, in British English and in Australian English, BATS is regular, and BATHS is irregular.

or 75 errors in 7,000 nonwords. Almost all of these 75 errors (84%) were lexical captures: a nonword was given the pronunciation of an orthographically and/or phonologically similar word. These captures only occurred when the difference between the captive and the capturer was toward the end of the nonword (see the discussion of *starn* above). At least some of the captures were clearly due to the presence of a phonological rather than an orthographic neighbor. For example, the nonword *phliked*, which should have been read (given the set of GPC rules used by DRC) as /flaikt/, rhyming with *spiked*, was read as /flikt/, or *flicked*. Now, *phliked* and *flicked* are not orthographic neighbors; they have only three of their seven letters in common, taking position into account. But they are phonological neighbors, differing only in a single phoneme, and that is the cause of the lexical capture of *phliked*.

It might be possible to make further changes in the parameter set of Table 1 that would reduce the nonword error rate and/or cause the model to read *czars* correctly. However, the number of such errors was so small that we elected to adopt this parameter set for our various simulations. In the rest of this article, these were the parameters we used in all simulations, except where otherwise noted.

This initial evaluation of the model's ability to translate words and nonwords from orthography to phonology having yielded satisfactory results, we went on to carry out various more detailed simulations. These simulations used two tasks, reading aloud and lexical decision, and were intended to deal with the standard effects obtained in studies of human readers as they perform these tasks. Here we follow the recommendation of Jacobs and Grainger (1994) that computational models should be multitask models; in particular, these authors argued that a computational model of reading should be able to perform the lexical decision task, the reading-aloud task, and the perceptual identification task. The DRC model can perform all three of these tasks and is the only computational model of reading that can do so. Other computational models can simulate reading-aloud latencies but not lexical decision latencies (such as the PMSP and ZHB models) or lexical decision latencies but not reading-aloud latencies (such as the model of Grainger & Jacobs, 1996).

Simulation of Reading Aloud

What is There to Simulate in Reading Aloud?

We have identified and listed below certain basic phenomena obtained in experiments on reading aloud using adult skilled readers that we consider as benchmark results that any computational model of reading should be able to simulate if it has claims to adequacy.

We should perhaps say something here about how we chose this particular set of phenomena. Several criteria were relevant. First, we chose every effect that had been simulated by either Plaut et al. (1996) or by Zorzi et al. (1998) so that we would be able to discover whether there were any effects that either of these models could simulate but which the DRC model could not (in the event, there were none). Then we added to this list more recently discovered effects such as the position of irregularity effect (Rastle & Coltheart, 1999b; see also Coltheart & Rastle, 1994) and the homophone and pseudohomophone (PSH) priming effects on reading aloud reported by Rastle and Coltheart (1999a); these effects

are particularly relevant because they arose in experiments specifically attempting to adjudicate between existing computational models of reading. Finally, we added pseudohomophony and *N* (the orthographic neighborhood size of a letter string—i.e., the number of words that differ from that letter string by exactly one letter) to our list of variables to explore, because the effects of these variables on reading aloud have been very widely studied, yet not simulated with other computational models of reading aloud.

We could not consider any effects involving semantics (e.g., the Imageability \times Frequency \times Regularity interaction reported by Strain, Patterson, & Seidenberg, 1995), because the DRC model at present has no semantic system; because it also at present has no system for nonlexical spelling (though see below for how this will be developed), we could not consider any effects on reading to which a nonlexical spelling system is likely to contribute, such as the feedback-consistency effect (also discussed below).

In this way we arrived at the following list of phenomena to simulate:

- (a) Reading aloud is faster for high-frequency words than for low-frequency words (Forster & Chambers, 1973).
- (b) Reading aloud is faster for regular words than for nonwords (McCann & Besner, 1987; Rastle & Coltheart, 1999b).
- (c) Reading aloud is faster for regular words than for irregular words when these are low in frequency; when they are high in frequency, the regularity effect is smaller or absent (Paap, Chen, & Noel, 1987; Paap & Noel, 1991; Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Taraban & McClelland, 1987).
- (d) The size of the regularity advantage declines as a function of the position in irregular words of their exceptional grapheme-phoneme correspondence (Rastle & Coltheart, 1999b).
- (e) Pseudohomophonic nonwords are read aloud faster than nonpseudohomophonic nonwords (McCann & Besner, 1987; Seidenberg, Petersen, MacDonald, & Plaut, 1996; Taft & Russell, 1992).
- (f) Nonwords with many orthographic neighbors are read aloud faster than nonwords with few or no such neighbors; this is also true for words (for a review of numerous studies, see Andrews, 1997).
- (g) Priming of reading aloud: In unmasked priming, where both prime and target are clearly visible, reading aloud is faster when prime and subsequently presented target are phonologically identical, but, at least with long stimulus onset asynchronies (SOAs), this effect interacts with lexicality such that it requires that at least one of the prime and target items is a word (Rastle & Coltheart, 1999a). In masked priming, where the briefly presented prime is not visible to the reader, target naming is faster when target and prime share their initial letter or phoneme than when there is no shared letter or phoneme; this effect does not occur when the shared letter or phoneme is in some position other than the initial position (Forster & Davis, 1991). In repetition priming, a word that has previously been presented in the experiment is read aloud with shorter latency than a word that has not been previously presented.

Our practice in DRC simulations is to use the actual stimuli that were used in the relevant experiments with human readers.

The Effects of Frequency and Lexicality on Reading-Aloud Latencies

Human readers show shorter naming latencies for high-frequency words than for low-frequency words (e.g., Andrews, 1989; Forster & Chambers, 1973). For the 7,898 words in DRC's vocabulary that the model read correctly using the standard parameters, the correlation between DRC naming latency and log printed word frequency was $-.46$ ($p < .0001$); this value is remarkably similar to the correlation of $-.42$ between human word-naming latency and log printed word frequency reported by McCann and Besner (1987).¹⁰

The effect of lexicality on DRC's naming latencies was assessed by using a set of 160 words and a set of 80 nonpseudohomophonic nonwords matched on number of letters, number of neighbors, number of body neighbors, number of body friends, number of body enemies, summed frequency of neighbors, summed frequency of body friends, positional bigram frequency, and positional trigram frequency (Coltheart & Coltheart, 2000). The mean naming latencies were 77.51 cycles (range 68–101) for the words and 156.86 cycles (range 119–234) for the matched nonwords, $t(228) > 50$, $p < .0001$; 10 of the nonwords were misread.

At first sight one might regard these results as showing that the lexicality effect is much bigger for the model than for human readers. In the human data of McCann and Besner (1987), for example, the difference between nonword and word latencies was 173 ms, and the nonword mean was 38% greater than the word mean. This seems much smaller than the 79.35-cycle difference between nonwords and words and the 102% lexicality effect seen in the DRC model's behavior. However, there is a flaw in this reasoning. One cannot reason that if a word is named by humans in 568 ms and by the model in 71 cycles, that 1 ms of human processing corresponds to $568/71 = 8$ cycles of DRC processing. This reasoning fails because there are processes outside the scope of the DRC model (e.g., early visual processing and articulatory execution) that contribute to human naming latencies but not to DRC naming latencies. Hence absolute comparison between the two types of latency cannot be made. This point is discussed further by Rastle and Coltheart (1999a) and by Spieler and Balota (1997), who estimated that 29.9% of the variance of word-reading latencies was due to influence of the initial phoneme of the word on the time at which the voice key is triggered by acoustic energy, a factor outside the scope of the DRC model.

We selected 80 five-letter pseudohomophonic nonwords matched to the nonpseudohomophonic nonwords on number of letters, number of neighbors, number of body neighbors, number of body friends, number of body enemies, summed frequency of neighbors, summed frequency of body friends, positional bigram frequency, and positional trigram frequency. DRC read these PSHs significantly more quickly than the nonpseudohomophonic nonwords (142.60 cycles vs. 156.86 cycles), $t(148) = 5.109$, $p < .0001$.

When the lexical route was disabled (by setting to zero strengths of the excitatory and inhibitory communications between letter level and orthographic lexicon and between phonological lexicon and phoneme level), the PSH advantage disappeared. The mean naming latency for nonpseudohomophonic nonwords increased from 156.86 to 167.99, $F(1, 69) = 62.43$, $p = <.0001$. This indicates that in the DRC model there are lexical influences on nonword reading even for nonpseudohomophonic nonwords and that the net effect of these is beneficial.

However, that does not mean that the lexical route is capable of reading nonwords correctly. We investigated this by disabling the nonlexical route (by setting the parameter GPC activation to zero) and submitting the 80 nonpseudohomophonic nonwords to the model. Here all reading is purely via the lexical route, and 0% of these nonwords were read correctly. All the errors were lexical captures—that is, reading the nonword as an orthographically or phonologically similar real word.

The Effect of Regularity on Reading Aloud

Seidenberg et al. (1984) reported that irregular words yield longer naming latencies than regular words, but only when the words are low in frequency; for high-frequency words there is no significant cost of irregularity. This regularity effect and its interaction with frequency was subsequently confirmed by Taraban and McClelland (1987), Paap, Chen, and Noel (1987), and Paap and Noel (1991).

These results have an obvious explanation in the context of the DRC model. Irregular words suffer a latency cost because there is competition at the phoneme level for the pronunciation of the phoneme that has an exceptional grapheme–phoneme correspondence. So, for example, for the word *comb* in the second-position phoneme set, the lexical route is activating the phoneme /ou/, and the nonlexical route is activating the phoneme /o/. Because within each set of phoneme units each phoneme inhibits all others, the rise of activation of the correct phoneme /ou/ will be slowed by inhibition from the regularization phoneme /o/. Such inhibition will not occur with regular words. And because lexically derived activation of phonemes will rise more rapidly for high-frequency words than for low-frequency words, this nonlexically derived inhibitory effect will be smaller or absent for high-frequency words.

What consequences this inhibition will have for DRC model naming latencies will be parameter dependent. If parameters are set such that lexical activation of the phoneme level is very strong, then the inhibition caused by the nonlexical route will be so weak that there will be no regularity effect, even for low-frequency words. If on the other hand parameters are set such that nonlexical activation of the phoneme level is very strong, then irregular words will be regularized, rather than being correctly pronounced but with a latency cost. It is possible—but there is no guarantee of this—that between these two extremes there is a parameter set that will produce correct responses to irregular words, a cost of irregularity for low-frequency words, and no cost of irregularity for high-frequency words, which is what is seen in human data.

To investigate this, we submitted the high- and low-frequency irregular and regular words¹¹ from Paap and Noel (1991) to the DRC model. The mean DRC latencies for each condition are shown in Table 2.

An ANOVA of the DRC naming latencies revealed significant effects of log word frequency, $F(1, 66) = 60.041$, $p < .0001$; and of regularity, $F(1, 66) = 73.790$, $p < .0001$; and a significant

¹⁰ In calculating this correlation, McCann and Besner (1987) used a transformation of frequency: $40 + 10 \cdot \log(\text{freq} + 1)$. We used the same transformation in the DRC data analysis.

¹¹ A few of these items had to be omitted from this simulation: BURY and WILY are disyllabic, and the DRC model applies only to monosyllables. LURE, LUTE, POUR, BEEN, and DOOR are not irregular words for the DRC model, because they are correctly read by the DRC GPC rules. BOOK and WALL are irregular by these GPC rules. SANS is not in DRC's vocabulary.

Table 2
DRC's Word-Naming Latencies for the Stimuli of Paap and Noel (1991), as a Function of Frequency and Regularity

Frequency level	Regular words	Irregular words
High	72.389	79.944
Low	78.100	88.786

interaction, $F(1, 66) = 4.234, p = .044$, indicating that the regularity effect was significantly larger for low-frequency words than for high-frequency words. All of these are the effects seen with human readers. The only difference between human and DRC results is that the smaller regularity effect with high-frequency words is significant in the DRC latencies but not in the human data; however, in all of the human studies using the standard reading-aloud task, the mean latencies with high-frequency words were longer for irregular words than for regular words.

The Position-of-Irregularity Effect

Irregular words have longer DRC latencies because of conflict at the phoneme level between activation from the lexical route and activation from the nonlexical route. Because the lexical route activates phonemes in parallel, whereas the nonlexical route activates phonemes sequentially from left to right, the magnitude of this conflict will depend on the left-to-right position, in the irregular word, of its irregular grapheme-phoneme correspondence. The earlier the conflicting phoneme is in the word, the less the correct phoneme will have been lexically activated at the point where it begins to receive inhibition from the incorrect phoneme that is activated by the nonlexical route. Rastle and Coltheart (1999b) confirmed this prediction regarding the DRC model by showing that the size of the DRC latency disadvantage incurred by irregular words declines monotonically as a function of the left-to-right position, in the irregular word, of its irregular grapheme-phoneme correspondence. Using the same regular and irregular words as stimuli, they showed that human readers show exactly the same effect. Confirmation of this new prediction provides strong support for the DRC model, and because the result seems most naturally interpreted as evidence that there is a left-to-right process involved in reading aloud, it poses a particular challenge for purely parallel models such as that of Plaut et al. (1996) and Zorzi et al. (1998).¹²

Other Effects of DRC's Serial Component in Reading Aloud

Position-sensitivity of the Stroop effect. When the task is naming the color in which a word is printed, naming latency is affected not only by semantic relationships between word and color (Stroop, 1935) but also by phonological relationships between word and color name (Dalrymple-Alford, 1972). Coltheart et al. (1999) confirmed this by showing that color naming is faster when the word has a phoneme in common with the color name, in the same position, than when it has no phonemes in common with the color name. Thus there is automatic activation of phonology from print in this situation, even though the phonology of the printed word is irrelevant to the task required. Coltheart et al. (1999)

reasoned that if there is a left-to-right component in the system that activates phonology from print, as in the DRC model, a phoneme from the word that is consistent with a phoneme in the color name will facilitate color naming more when it is the first phoneme in the color name and word than when it is the last phoneme in the color name and word. In their experiment, human color-naming latencies showed exactly this position sensitivity.

So did DRC's color-naming latencies. In order to simulate Stroop color naming with the DRC model, Coltheart et al. (1999) added to the DRC model a miniature semantic system containing just three semantic entries, one for RED, one for GREEN, and one for BLUE. This semantic system was linked interactively to the phonological output lexicon of the DRC model. When a word is presented and simultaneously the semantic entry for the desired color name is also activated, this extended model has dual inputs. Hence there will be activation of two entries in the phonological output lexicon, one for the word and one for the color name, and at the phoneme level the phonemes of the color name will be activated from the phonological output lexicon, whereas the phonemes of the word will be activated (in parallel) from the phonological output lexicon and also activated (left to right) from the nonlexical route. This extended DRC model can produce the color name or read the word aloud, just as people can. The instruction to people, "Name the color, don't read the word," was simulated in the DRC model by setting to zero the strength of the communication from the letter level to the orthographic lexicon. Under these conditions, the correct color name was almost always produced by the model. What is more, the model's color-naming latencies were faster when the color name shared a phoneme with the word than when it did not; the size of this effect was larger when the shared phoneme was the first than when it was the last, exactly as shown by the human color-naming latencies.

Strategic effects. In the Stroop study just described, it was proposed that readers can alter the strength of a relevant pathway of communication in their reading systems in response to task instructions. Rastle and Coltheart (1999b) made the same proposal with respect to strategic effects in reading.

Coltheart (1978) reported a small experiment in which readers read aloud a list of items consisting of pronounceable nonwords except for the last item, which was the highly irregular word *wolf*. Numerous readers regularized this word, pronouncing it to rhyme with *golf*; many did not even notice that the last item in the list was a word. Coltheart (1978) suggested an explanation: As more and more nonwords are encountered, the readers turn down the lexical route (because it is never providing a correct response) or turn up the nonlexical route (because it is always providing the correct response), or both. Either or both of these strategic resetttings of the dual-route reading system, if extreme enough, will generate a regularization error to *wolf*.

This kind of strategic effect on reading aloud, however, is not always found. For example, Coltheart and Rastle (1994) measured the size of the regularity effect under two conditions: (a) deemphasizing the nonlexical route (filler items were all exception words) and (b) deemphasizing the lexical route (filler items were all nonwords). Because it is the nonlexical route that generates the

¹² For further discussion as to the correct interpretation of these data, see Zorzi (2000) and the rejoinder by Rastle and Coltheart (2000b).

regularity effect, they expected the regularity effect to be larger in condition (b). But that was not found; the regularity effect did not differ between conditions.

Rastle and Coltheart (1999b) offered an explanation. According to their results, the nonlexical route is maximally problematic for exception words with first-position irregularities; with irregularities at the third or later phoneme positions there is little or no regularity effect. Very many of the exception-word fillers used by Coltheart and Rastle (1994) had late irregularities, so their condition (a) did not really require readers to deemphasize the nonlexical route and so reduce the size of the regularity effect. To really stress the system with exception-word fillers, reasoned Rastle and Coltheart (1999b), all the filler exception words need to be first-position irregular. Third-position irregulars will not be disturbed by the operation of the nonlexical route.

Hence they devised a strategy experiment using a first-position irregular filler condition and a third-position irregular filler condition. Now a strategy effect on human naming latencies was observed: naming of target regular words or nonwords was slowed (as would occur if the nonlexical route were being deemphasized) when first-position irregular fillers were present, compared with target naming when third-position irregular fillers were present. This result was successfully simulated with the DRC model by slowing the activity of the nonlexical route: the parameter controlling how many cycles elapse before the nonlexical route processes the next letter was increased from 17 cycles to 22 cycles.

These results are of consequence for two reasons. First, they provide further evidence for the existence of a left-to-right processing component of the reading system. If no such component existed, why should regular word and nonword reading be slowed by the presence of first-position irregular words in comparison with the presence of third-position irregular words? Second, it is unclear how computational models that only have a single fully implemented reading route, such as the models of Plaut et al. (1996) and Zorzi et al. (1998), could simulate these strategy effects. Furthermore, even if the unimplemented semantic routes of these models were implemented, it is unclear how the models might be able to simulate the influence of position of regularity, because that influence suggests a contribution of serial processing, whereas there is no serial processing of any kind in either of these models.

The Nonlexical Route as a Nonserial Process

In two recent lines of work, it has been argued that the nonlexical route activates phonemes in parallel, rather than serially left to right as occurs in the DRC model. Parallel processing is a feature of the two-cycles model of Berent and Perfetti (1995) and also of the model proposed by Kawamoto, Kello, Jones, and Bame (1998).

The two-cycles model. In the two-cycles model of Berent and Perfetti (1995), there are two processes involved in the assembly of phonology from print, both processes being parallel processes. There is a fast process that is applied in parallel to all of the consonants in the input string, and a slower process that is applied in parallel to all the vowels. This conception of how the assembly process operates is completely at variance with the DRC conception that the nonlexical route operates serially. Evidence for the latter conception has already been described: the effect of position of irregularity on the regularity effect in reading aloud, the effect

of position of irregularity in filler stimuli on strategy effects in reading aloud, and the effect of position of phoneme overlap on Stroop color naming. All of these findings are inconsistent with the two-cycles model.

A direct adjudication between the two conceptions is possible using the Stroop data of Coltheart et al. (1999), who showed that color naming was significantly faster when printed word and color name shared their last phoneme than when there was no phoneme overlap. In two types of stimuli (when the color was red or green) this last phoneme was a consonant; in the third type (when the color was blue) it was a vowel. If the assembly of phonology from print is slower when the phoneme is a vowel than when it is a consonant, then whether the last phoneme overlaps with the color name will have less effect when that phoneme is a vowel than when it is a consonant. No such effect was observed by Coltheart et al. (1999); color did not interact with the size of this effect, which was in fact (nonsignificantly) larger when the color was blue than when it was red or green.

What reasons, then, did Berent and Perfetti (1995) have for proposing such a model? They did so on the basis of a backward masking experiment in which a target word was briefly presented and was followed by a brief backward mask. The mask was either a PSH of the target (rake-RAIK), had the same vowels as the target but one different consonant (rake-RAIB), had the same consonants as the target but one different vowel (rake-RIKK), or was completely different (rake-BLIN). At the shortest target durations, accuracy of target report was higher with the consonant-preserving mask than with the vowel-preserving mask, which Berent and Perfetti took as evidence for the two-cycles model.

However, in the stimuli used by Berent and Perfetti (1995), there is an almost perfect confounding between mask type and position of phoneme at which mask and target differed. In almost every vowel-preserving mask, mask and target differed only at the third phoneme; in almost every consonant-preserving mask, mask and target differed only at the second phoneme. Now, according to the serial theory of assembly of phonology, at the offset of the briefly presented target, the left-to-right assembly process will have dealt only with the first one or two letters in the target. So, although all of the phonemes of the target will have received a little activation through the lexical route, the first one or two phonemes will also have received some additional activation from the nonlexical route. When a backward mask is presented that shares all but one of the phonemes of the target, and therefore boosts their activations, it is plausible to argue that the more weakly activated phonemes would benefit more from this boost (this argument is due to Perry, 1997, p. 73). Because the third phoneme of the target will have been more weakly active at target offset than the second, a mask that contains that third phoneme (the consonant-preserving mask) will be more beneficial than a mask that contains the second phoneme (the vowel-preserving mask).

Thus, because of the confounding between mask type and position of phoneme mismatch in these materials, it is possible to give a plausible account of their results in terms of an assembly process that is a single serial left-to-right process, rather than a pair of parallel processes (i.e., the two-cycles model). Because the serial account is also consistent with various other findings concerning the regularity effect, the Stroop effect, and the onset effect, whereas the two-cycles model is not, it seems clear that we should favor the serial model over the two-cycles model.

The Kawamoto et al. (1998) model. Kawamoto et al. (1998) have proposed that as phonology is being computed from orthography, activations rise in parallel across all phonemes, although the rate of activation varies from phoneme to phoneme because this rate is influenced by the consistency of the grapheme-phoneme relationships. On their model, articulation can be initiated as soon as the initial phoneme reaches a critical activation level, even if this is before subsequent phonemes reach a critical activation level. That claim, however, conflicts with a great deal of evidence on coarticulation effects in speech production; in many speech-production situations, it has been shown that the way in which an initial consonant is articulated is influenced by the nature of the following vowel. In any situation where this happens, the utterer cannot have been producing the initial consonant before knowing the identity of the following vowel.

Kawamoto et al. (1998) acknowledged that coarticulation effects are evident in many speech-production situations, but claimed that their initial-phoneme hypothesis was tenable because coarticulation effects do not occur in speeded word-reading experiments. This claim was directly tested by Rastle, Harrington, Coltheart, & Palethorpe (2000), who studied coarticulatory effects during a speeded word-reading experiment using acoustic, kinematic, and electropalatographic measurements that allowed them to observe horizontal and vertical movement of the lips and the contact of the tongue with the palate during the period between the presentation of the target word and the initiation of the naming response. They found that for words beginning with plosive or with nonplosive consonants, the nature of the second phoneme (always a vowel) did influence the articulation of the initial phoneme—that is, even in this speeded word-reading situation, coarticulatory effects occurred. Thus, they claimed that any model of speeded reading aloud that states that the first phoneme can be uttered before the second phoneme is known—such as the model of Kawamoto et al. (1998)—cannot be correct.

The Effect of Neighborhood Size on Word and Nonword Naming

In the DRC model, one would expect nonword naming to be facilitated by a large orthographic neighborhood (N), as has been shown to be the case in human readers (e.g., Laxon, Masterson, Pool, & Keating, 1992; McCann & Besner, 1987; Peereman & Content, 1995). Cascaded processing in the model allows nonwords to activate orthographically similar words in the orthographic lexicon, and this activation then feeds down to the phonological lexicon and finally to the phoneme system. Because generally the orthographic units that become activated by the nonword stimulus represent words that are neighbors of the nonword, and because these neighbors generally have many phonemes in common with the nonword, phonemic activation generated from the lexical route, paired with correct nonlexical processing, should facilitate nonword naming. That is why DRC's nonword naming is slowed by turning off the lexical route, as we reported above.

In order to determine whether this intuition about how N should affect DRC's nonword naming is correct, we analyzed the naming latencies for the 244 four-letter nonwords that were in the randomly chosen set of 7,000 nonwords described earlier. We did not use the five-, six-, and seven-letter nonwords because most had no neighbors; the mean value of N for these items was well below 1.0.

And we did not use the three-letter nonwords because there were only 23 of them.

The correlation between N and DRC naming latency was negative ($-.154$), and regression analysis showed that it was significant, $F(1, 242) = 5.853, p = .016$. Thus the more neighbors a nonword has the faster it is named by the model, as is the case with human readers.

In the DRC model, naming for word stimuli from large neighborhoods should be facilitated relative to naming for word stimuli from small neighborhoods. The argument is the same as in the case of nonwords: Cascaded processing allows a word stimulus to activate orthographic entries for neighboring words, which in turn activate phonological entries and finally phonemes. However, neighborhood effects for word stimuli should not occur in the model to the same extent as is the case for nonwords. The reason for this is that in the standard set of parameters, the parameter controlling lateral inhibition in the orthographic lexicon is set to a value such that it prohibits much of the facilitatory activation of the orthographic lexicon.

Andrews (1989, Experiment 3), Andrews (1992, Experiment 2), and Sears, Hino, and Lupker (1995, Experiment 3b) all found that, with low-frequency words, high- N words yielded shorter naming latencies than low- N words. The DRC model's naming latencies for the high- N and low- N low-frequency words used in these three studies were obtained. In none of these three simulations was the model's naming latency affected by N . In case this was a problem due to small sample sizes or insufficiently low frequencies, all words in DRC's vocabulary with a CELEX frequency of 1 and between four and seven letters were run through the model, and its naming latencies were obtained. Because N and number of letters are highly correlated, the regression of DRC naming latency on N was calculated separately for each word length (89 four-letter words, 168 five-letter words, 141 six-letter words, and 86 seven-letter words). In none of these four cases did the relationship of N to DRC naming latency even approach significance. Thus with the parameter set of Table 1 there simply is no effect of N on DRC's word-naming latencies; so here there is a major difference between the model and human readers.

It seems likely that the absence of an advantage for high- N words occurs because there is a considerable degree of mutual inhibition between words with the standard DRC parameter set. Lateral inhibition between entries in the orthographic lexicon is set at .06, lateral inhibition between words in the phonological lexicon is set at .07, and inhibition from letters to words is very high (.435); the higher this is, the less a word can excite potentially supportive neighbors. To explore this explanation for the absence of an N effect on word naming, we set both lateral inhibitory parameters to zero and reduced letter-to-word inhibition to .350. With these parameters, there is now a highly significant beneficial effect of N on word naming ($p < .0001$), a significant frequency effect ($p = .045$), and no interaction between these two factors. We note that although both Andrews (1989) and Sears et al. (1995) reported a significant interaction between N and frequency in their word-naming experiments, Andrews (1992) found an N advantage that did not interact with frequency, as occurred in this simulation. Clearly more experimental work and more computational work is needed here; once it is clear what the human results actually are, it will be necessary to study whether the DRC model can yield a beneficial effect of N on word naming through a modification of

the parameter set that does not compromise its successful simulations of the other effects discussed in this article.

We should also point out that in none of the experiments finding an effect of N on human word-naming latencies were nonwords used. The DRC model needs to have a high value for letter-to-word inhibition, such as .435, to prevent lexical capture of nonwords in reading aloud. If only words are presented, this value can be lowered to .350 without causing reading errors. So it would be of particular interest to discover whether human readers continue to show an effect of N on word-reading latencies when words and nonwords in random presentation order make up the materials of the experiment.

The PSH Advantage

Although PSHs are disadvantaged in lexical decision relative to nonpseudohomophonic nonwords (e.g., Besner & Davelaar, 1983; Coltheart & Coltheart, 2000; Coltheart, Davelaar, Jonasson, & Besner, 1977; McQuade, 1981), they are advantaged in naming (e.g., McCann & Besner, 1987; Seidenberg et al., 1996; Taft & Russell, 1992). The conditions under which this advantage appears, and the effects that accompany the advantage, are complex. The effect is modulated by orthographic similarity; PSHs are pronounced more quickly if they are orthographically similar to their base words than if they are orthographically distant (Marmurek & Kwantes, 1996). The PSH advantage is further modulated by base-word frequency, but only when readers are "slow" (Taft & Russell, 1992) or when the stimulus list contains only PSHs (Kwantes & Marmurek, 1995; Marmurek & Kwantes, 1996).

We have carried out a DRC simulation of reading aloud of PSHs and control nonwords with the items used by Taft and Russell (1992). The DRC model operating under the standard set of parameters shows a PSH advantage for both PSHs with high-frequency base words, $t(28) = 7.24$, $p < .01$, and for PSHs with low-frequency base words, $t(28) = 4.12$, $p < .01$. There was an effect of base-word frequency: PSHs with high-frequency base words were named more quickly by the model than were PSHs with low-frequency base words, $t(28) = 2.67$, $p < .05$, as was shown by the slower readers of Taft and Russell (1992).

Taft and Russell (1992) reported that their group of fast readers showed neither effects of PSH base frequency nor an advantage for PSHs at all. We sought to simulate this effect by causing the DRC model to name nonwords more quickly: The interletter interval parameter was decreased from 17 to 8, and the GPC excitation parameter was increased from .055 to .25. Under these conditions, there was no longer any difference between PSHs with high-frequency base words and those with low-frequency base words, $t(28) = .93$. Furthermore, there was no PSH advantage at all, either for PSHs with high-frequency base words, $t(28) = .87$, or for PSHs with low-frequency base words, $t(28) = .22$. Here, the fast DRC model is behaving exactly like the fast readers of Taft and Russell (1992).

The reason why the speed of the nonlexical route modulates the PSH advantage in the DRC model is that there is more time for activation of the base word in the phonological lexicon to rise when the nonlexical route operates slowly and therefore time for this activation to feed forward to facilitate activation at the phoneme level. Because the rise of activation in the phonological lexicon is determined, in part, by word frequency, PSHs with

higher base-word frequencies are facilitated more than are PSHs with lower base-word frequencies.

McCann and Besner (1987) also studied the effect of pseudohomophony on reading aloud, and we obtained the DRC naming latencies for their very well matched sets of PSHs and control nonwords (excluding a few items that are pseudohomophonic in Canadian speech but not in Australian English speech, and their controls, plus a small number of nonwords that the model misread). McCann and Besner found that human readers were faster at naming PSHs than control nonwords; that was so for the DRC model also (pseudohomophone mean latency = 138.4, control nonword mean latency = 152.1, $t(65) = 9.964$, $p < .001$). They then analyzed the regression of N on naming latency, separately for PSHs and control nonwords.

For control nonwords, regressing the linear and quadratic components of the N main effect with the human naming latencies produced $r = .45$, $p < .01$, and with the DRC naming latencies, $r = .52$, $p < .01$. For PSHs, this correlation for human naming latencies was .19 (*ns*), and the correlation for DRC naming latencies was .28 (*ns*). Overall, the fit of model to data here is remarkably close.

Priming of Reading Aloud

A model that adheres to the GRIN principles, such as the DRC model, can approach the simulation of priming in a number of different ways. Priming could occur as a result of short-term changes in between-module communication strengths based on the effects of the prime or could occur as a result of a decrease in the critical activation required for a response after the presentation of a prime. The approach that we have taken to priming is discussed by Rastle and Coltheart (1999a). They suggested that a simulation of priming could be achieved in the DRC model by allowing residual activation from the presentation of the prime to affect the processing of a subsequently presented target. In order to simulate priming in this way in the model, activation within the system is not reset to zero after the offset of the prime stimulus; instead, the target stimulus is presented to a system still partially activated due to the influence of the prime.

If the experiments to be simulated involved a prime-target delay, then in DRC simulations a decay period must be instituted between the presentation of the prime and the presentation of the target, during which activations generated by the prime decay. Thus Rastle and Coltheart (1999a) implemented a proportion decay parameter in the model to decrease the amount of residual activation left behind by the prime. The proportion of decay parameter can be set to one value throughout every unit in the entire system or can be set individually for each module in the lexical route.

Rastle and Coltheart (1999a) implemented a second parameter, the length of decay (prime-target interval), in order to achieve a simulation of priming in the DRC model. Because units in the model are updated on each processing cycle, the influence of frequency or neighborhood might make some units resistant to decay and others more vulnerable to decay. Thus, activation from the prime is not reduced with one massive hit of decay; rather, it is decreased gradually, cycle by cycle.

Orthographic form priming might occur, then, because of the influence of residual activation in the letter or orthographic units;

homophone priming might occur because of the influence of residual activation in the phoneme or phonological units; and all of these effects might contribute to repetition priming.

Given this characterization of priming in the DRC model, we can consider more specifically how the model simulates priming in unmasked and masked conditions.

Simulating unmasked priming. In their phonological priming experiments, Rastle and Coltheart (1999a) measured word and nonword naming latencies in a reading-aloud experiment that varied prime and target lexicality factorially. In the priming conditions of the experiments, prime and target were orthographically different but phonologically identical. They found that homophone priming occurred only when at least one of the prime or the target was orthographically a word; no priming occurred for PSH prime–PSH target pairs or for nonword prime–nonword target pairs. Exactly the same pattern of results occurred with the DRC model when a simulation of unmasked priming, using the same items as were used with human readers, was carried out using that model. Rastle and Coltheart (1999a) further showed that the locus of this priming effect was in the model's local phonological representations and, more subtly, that serial properties of the nonlexical route in the model also played a role in the successful simulation. It is unclear at present how any model in which representations are distributed might be able to accommodate these results. In particular, if phonological representations are distributed as in the model of Plaut et al. (1996), then the items *brayk*, *brake*, and *braik* will have identical representations at the phonological level. Why, then, does the prime *brayk* facilitate *brake* yet have no effect on *braik*?

Because the interstimulus intervals in these experiments were well over 1 s, it might be argued that these priming effects are in some way strategic in nature, but Rastle and Coltheart (1999a) considered this possibility and rejected it for two reasons. First, primes and targets were presented as a continuous stream of items to all of which the readers had to respond, so readers did not know which items were primes and which were targets. Second, Rastle and Coltheart (1999a) followed the recommendations made by Lukatela and Turvey (1994) regarding how to avoid strategy effects in priming experiments, namely, to keep the incidence of related prime–target pairs low. Lukatela and Turvey (1994) argued that “because homophonic similarity was limited to 15% of the stimuli and to only 7.5% of the stimuli if the identity prime conditions are excluded, it would seem that the success of Experiment 5 cannot be attributed to a general strategy” (p. 342). Quoting this reasoning and then adopting it, Rastle and Coltheart (1999a) argued

Likewise, because only an average of 9.3% of the targets used in our experiments were preceded by phonologically matching items, we were also confident that any of the priming effects we observed could not be attributed to a general strategy of using phonological information to anticipate the target. (p. 466)

Simulating masked priming. In masked priming experiments (e.g., Forster & Davis, 1991), a prime is very briefly presented (e.g., for 50 ms) and then replaced by a target that the reader has to read aloud (or perform a lexical decision). This is simulated in DRC by presenting the prime for a relatively low number of processing cycles and then, without resetting the unit activations generated by the prime, presenting a target for reading aloud by the model.

The effect we have studied here is the “onset effect” of Forster and Davis (1991). They found that when prime and target share the same initial phoneme (belly–BREAK), naming latency for the target is facilitated, relative to a no-similarity condition (merry–BREAK) or a rhyming condition (take–BREAK). This onset effect is also present in DRC's word naming latencies when masked form priming is simulated using a prime duration of 15 cycles (Coltheart et al., 1999).¹³ That is, the model names target words faster when they share the first letter or phoneme of the prime word than when they share some other letter or phoneme, or share no letters or phonemes. This onset effect occurs because of the left-to-right operation of the DRC model's nonlexical route; for briefly presented primes, only the first letter of the stimulus has been processed by the nonlexical GPC procedure at the point at which the target word is presented.

We discuss other aspects of masked priming later in this article.

Simulating repetition priming. Visser and Besner (in press) argued that the DRC model predicts that the interaction between regularity and frequency in reading aloud should be reduced by repetition, because repetition of a word facilitates its processing by the lexical route and has no influence on its processing by the nonlexical route, and the more the lexical route contributes to reading aloud the smaller the regularity effect will be. In their human naming latency data using the regular and irregular words from Paap and Noel (1991), they found effects of repetition, frequency, and regularity, and a three-way interaction such that the Frequency \times Regularity interaction was smaller in the repeated than in the nonrepeated condition. They reported a simulation by the DRC model of this experiment, in which repetition priming was simulated as described above. In this simulation, the model latencies showed exactly the same pattern of effects as the human latencies, including the three-way interaction; the Regularity \times Frequency interaction is smaller at second presentation.

Simulation of Lexical Decision

The DRC approach to lexical decision has been developed from work by Coltheart et al. (1977), who discussed the lexical decision task in relation to two broad classes of lexical access model: serial-search models and parallel-activation models. Serial-search models offer a rather natural account of how the lexical decision task is performed. The lexicon is searched serially, in order of frequency. If at any point during this search a match is found between a lexical entry and the stimulus, the decision YES is made. If the search is completed and no such match is detected, the response NO is made. This interpretation is natural because there is a discrete event that justifies a YES decision (the occurrence of a lexical match) and a discrete event that justifies a NO decision (completion of the search with no match having been found). Moreover, some correct predictions follow from the serial-search interpretation of lexical decision: there will be an effect of frequency on YES latencies, and NO latencies will be longer than YES latencies.

Lexical decision is more challenging for parallel-activation models, because there is no discrete event in the processing of such

¹³ This way of simulating masked priming with a computational model was developed by Jacobs and Grainger (1991) and Grainger and Jacobs (1993).

models that signals that the stimulus must be a word, and no discrete event that signals that it must be a nonword. When the stimulus is a word, activation in its lexical entry rises continuously over time toward its asymptotic value. Coltheart et al. (1977) suggested that an activation criterion is set that allows the YES decision to be made; if any entry in the orthographic lexicon attains an activation that exceeds this criterion, then the YES decision is made. They also suggested that the NO decision is made by using a deadline: if a certain processing duration has elapsed and a YES decision has not yet been made, then decide NO. They then reported the results of a lexical decision experiment in which the words and nonwords differed in orthographic neighborhood size (N). Nonwords with high N values yielded significantly longer latencies than nonwords with low N values. This is inconsistent with the serial-search account as formulated above because all NO decisions should take the same time regardless of the nature of the nonword if all NO decisions are made after completion of an unsuccessful serial search of the lexicon. The result is inconsistent with the parallel-activation account as formulated above for a similar reason: the point in time at which the deadline has elapsed will not depend on the nature of the nonword.

Hence Coltheart et al. (1977) considered whether there were any modifications to either of these accounts of lexical decision that could be made that would reconcile them with the data on the effect of N on NO latency. For the serial-search approach, one could suppose that the time made to detect a mismatch between a lexical entry and a nonword stimulus might be longer when the nonword was a neighbor of the lexical entry than when it was not. If so, the exhaustive search would take longer for high- N nonwords than for low- N nonwords, and so there would be a positive relationship between nonword N and NO latency. This account, however, also predicts an effect of N on YES latencies. A word with many neighbors, unless that word's frequency is higher than the frequencies of all of its neighbors, will yield a longer YES latency than a word with no neighbors. In general there should be a positive relationship between word N and YES latency, because the more higher-frequency neighbors a word has, the more mismatch detections will have occurred prior to the correct match. However, Coltheart et al. (1977) reported that N had no effect on YES latency, and they took this as evidence against the serial-search account of lexical decision. (Subsequently, Andrews, 1989, 1992, found that word N did affect YES latencies as long as words were of low frequency, but the effect was opposite to that predicted by the serial-search account—high N led to faster YES latencies. We discuss this further below.)

In contrast, Coltheart et al. (1977) were able to suggest a plausible modification to the parallel-activation approach that could explain the effect of N on NO latencies. They suggested that lexical decision efficiency would be improved if the value of the deadline for NO responses was variable rather than fixed, with the variation being controlled by the total overall activation in the orthographic lexicon early on in processing. When this total is high, the likelihood that the stimulus is a word is high, so it is prudent to have a long deadline to avoid a premature incorrect NO decision. When this total is low, the likelihood that the stimulus is a word is low, so it is relatively safe to set the deadline to some rather low value, enabling the NO response to be made quickly. That would mean that high- N nonwords would yield longer NO latencies than low- N nonwords, whereas N would have no effect

on YES latencies; and that is the pattern of results found by Coltheart et al. (1977). They therefore concluded that their results favored a parallel-activation account of visual lexical access during reading such as the logogen model (Morton, 1969) and were inconsistent with a serial-search account. And, specifically, they argued that in the lexical decision task the YES response is made by using an activation criterion applied to the individual lexical entries in an orthographic lexicon, and the NO response is made when a deadline has elapsed and no YES response has been made. The value of this deadline is computed relatively soon after stimulus onset, from the total activation of the orthographic lexicon at that time; the greater the total, the longer the deadline.

This procedure for making YES and NO decisions in the lexical decision task was implemented in the context of the IAC model by Jacobs and Grainger (1992). They successfully simulated the NO response data from Coltheart et al. (1977); see Figure 7 of Jacobs and Grainger (1992).

Further developments along these lines were recently proposed by Grainger and Jacobs (1996). Their account of how lexical decisions are made adds a third decision criterion to the two proposed by Coltheart et al. (1977) and implemented by Jacobs and Grainger (1992), namely, a "fast guess" procedure for making YES decisions. This procedure works as follows: if early in processing the total activation of the orthographic lexicon is sufficiently high, respond YES even if no single entry has reached the critical activation level. Here, then, there are two different ways to decide YES, and one way to decide NO.

What evidence is there that this third criterion, the fast-guess procedure, ought to be included in an account of the lexical decision task? Grainger and Jacobs (1996) assumed that the critical activation criterion for the YES response is a fixed property of the visual word-recognition system and so cannot vary as a function of properties of the words and nonwords used in lexical decision experiments (i.e., is not strategically variable). Because they did observe effects on YES latencies that were plausibly attributable to strategic responses by readers as a function of the nature of the stimulus materials, they concluded that there was a second method for responding YES, one that was strategically modifiable; a mathematical model of lexical decision involving the three decision procedures offered a good fit to their data.

A second line of evidence in support of the three-criterion account was that it was able to simulate the findings of Andrews (1989, 1992) that with high-frequency words N has no effect on YES latencies, whereas with low-frequency words YES is faster when N is high than when N is low. This interaction between frequency and N in YES responding occurs because for high-frequency words the critical activation criterion is met earlier in time than the fast-guess criterion, and it is only the fast-guess criterion that is sensitive to N .

What is There to Simulate in Lexical Decision?

Our simulation work with the DRC model has been more extensive in relation to the task of reading aloud than in relation to the lexical decision task, but we have identified certain basic phenomena obtained in lexical decision tasks that we consider as benchmark results that any computational model of reading should be able to simulate if it has claims to adequacy, and we report here

simulation studies of all of these phenomena. The phenomena are as follows:

YES latencies are

- (a) Faster for high-frequency words than for low-frequency words (Forster & Chambers, 1973);
- (b) Faster than NO latencies (Forster & Chambers, 1973);
- (c) Unaffected by N when words are of high frequency (Andrews, 1989, 1992);
- (d) Facilitated by N when words are of low frequency (Andrews, 1989, 1992).

NO latencies are

- (a) Longer for high N nonwords than for low N nonwords (Coltheart et al., 1977);
- (b) Slowed by pseudohomophony (Rubenstein, Lewis, & Rubenstein, 1971; Coltheart et al., 1977; Coltheart & Coltheart, 2000);
- (c) Less affected, or not affected at all, by pseudohomophony when the PSH is not orthographically very close to the parent word (e.g., when it is not a neighbor of the parent word; Coltheart & Coltheart, 2000).

As mentioned above, our practice in DRC simulations is where possible to use the actual stimuli that were used in the relevant experiments with human readers, and we have followed that practice here.

As discussed above in relation to simulation of the results from experiments on reading aloud, our goal was to find a single set of parameters under which the model successfully simulates a wide variety of new experimental results without any parameter manipulation. We were not interested in showing that, for each phenomenon to be simulated, there existed a parameter set that gave a successful simulation. We were interested in showing that one single parameter set generated successful simulations of many different data sets illustrating many different empirical phenomena. The only departure from this requirement was that we treated human strategy variation as equivalent to variation of a rationally chosen DRC parameter.

Having found a single parameter set that worked well for simulating reading aloud, then, we went on to explore its use in simulating lexical decision. This was completely unsuccessful, and the reason was obvious. As discussed below, one crucial variable in simulations of lexical decision is neighborhood size, because this influences both YES and NO latencies in human data. Under the parameter set with which we simulated reading aloud, however, the orthographic lexicon of the DRC model was insufficiently sensitive to N , especially when the stimulus was a word; this was clearly because of one particular parameter, letter-to-word inhibition, which is $-.435$ in the parameter set used for simulation results from experiments on reading aloud.

A relatively high value for letter-to-word inhibition is necessary for preventing a nonword from being lexically captured by one of its orthographic neighbors or for preventing a low-frequency word from being captured by a high-frequency neighbor, when the DRC model is performing the reading-aloud task. Such captures would of course cause erroneous responses in this task, but they may not lead to difficulties when the task is lexical decision.

Hence we reduced the value of letter-to-word inhibition to $-.300$. That was the only parameter, from the set that successfully simulates reading aloud, that was changed for the lexical decision simulations. We offer the following justification for treating this

parameter change as a strategic variation adopted specifically for the lexical decision task. Reading aloud requires knowledge of the specific item that has been presented, but lexical decision does not. As we indicate below, accurate lexical decisions can often be made simply on the basis of the activity of the orthographic lexicon, including simply its total activation summed across all entries. Reducing letter-to-word inhibition from the very high value of $-.435$ allows the orthographic lexicon to become much more active early in processing, and hence provides richer data for a lexical decision procedure that uses total activation to make decisions (such as the procedure we describe below).

Here our work is closely related to that of Ziegler, Rey, and Jacobs (1998), who used an IAC-style model to simulate data from a word-recognition experiment that used the progressive demasking or screen-fragmentation technique (Snodgrass & Mintzer, 1993). They found that different readers adopted different response strategies when confronted with this task and were able to model these individual differences in terms of strategic variation of two parameters of the IAC model, word-to-letter feedback and letter-to-word inhibition. It is exactly the latter parameter that we are proposing is strategically reduced when the task is lexical decision, in comparison with reading aloud.

The Effects of Word Frequency and Neighborhood Size on Lexical Decision

The words we have used in these simulations were those used by Andrews in her studies of the $N \times$ Frequency interaction (Andrews, 1989, Experiments 1–4; Andrews, 1992, Sample 1) except that we had to discard four items because they were polysyllabic (and so outside the domain of the DRC model) and some other items so as to achieve matching and equal cell frequencies. This gave us 24 words per cell, with N and frequency orthogonally varied. All words were four letters long.

The high- N and low- N nonwords we used were a newly chosen set, rather than coming from any set previously used with human readers, because we wanted the words and nonwords to be matched on N , and as far as we are aware no study of human lexical decision has done this. So we selected 48 four-letter nonpseudohomophonic nonwords with high N (matched to the N values of the high- N words) and 48 four-letter nonpseudohomophonic nonwords with low N (matched to the N values of the low- N words).

In initial simulations, only the two-decision criteria proposed by Coltheart et al. (1977) were used. Here it was easy to simulate the frequency effect on YES and the harmful effect of N on NO. But we were unable to simulate the Frequency \times N interaction on YES latencies reported by Andrews (1989). When we added the fast-guess criterion for YES proposed by Grainger and Jacobs (1996), however, this became possible.

Hence the lexical decision procedure used in the DRC simulations reported here was that first proposed by Grainger and Jacobs (1996). This procedure operates as follows.

First, we need to define A, S, and D:

A: decide YES if any entry in the orthographic lexicon has reached an activation level of A.

S: decide YES if the sum of the activations of all of the entries in the orthographic lexicon has reached the value S (“fast-guess” mechanism).

D: decide NO if D processing cycles have elapsed and a YES decision has not yet been made. Then initial values for A, S, and D are set. These initial values were $A = 0.69$, $S = 10.00$, and $D = 42$, and they are applied throughout the first 19 cycles of processing.¹⁴ But at Cycle 20, the values of S and D can be updated; whether they are updated or remain at their initial levels depends on the state of the orthographic lexicon at that cycle. The value of A remains constant throughout; it is not subject to updating.

Update S on cycle 20? The total activation criterion for making a fast YES response is reduced on Cycle 20 if there is sufficient activity in the orthographic lexicon at that cycle. Specifically, if the summed activation of all entries in the orthographic lexicon has reached a value of 0.200 by Cycle 20, the value of S is reduced, from its original value of 10.00 to a new value of 1.98. If the summed activation of all entries in the orthographic lexicon has not reached a value of 0.20 by Cycle 20, then S remains at its original value of 10.0. In either case, the fast-guess procedure continues to be applied on subsequent cycles: decide YES if the sum of the activations of all of the entries in the orthographic lexicon has reached the value S.

Update D on cycle 20? The deadline value is lengthened on Cycle 20 if there is sufficient activity in the orthographic lexicon at that cycle. Specifically, if the summed activation of all entries in the orthographic lexicon has reached a value of 0.112 by Cycle 20, the value of D is extended from its original value of 42 cycles to a new value of 48 cycles. If the summed activation of all entries in the orthographic lexicon has not reached a value of 0.112 by Cycle 20, then D remains at its original value of 42 cycles. In either case, the deadline procedure for NO responses continues to be applied on subsequent cycles: decide NO if D processing cycles have elapsed and a YES decision has not yet been made.

It is important to realize that claims about what the procedure is for making lexical decisions are quite distinct from claims about the architecture of any computational model of the reading system. Such a model yields data relevant to lexical decision; then some decision procedure external to the model needs to be applied to those data if the task is lexical decision.

Figure 8 shows the mean YES latencies from the human data (Andrews, 1989, 1992), and the DRC model's mean correct YES and correct NO latencies. The model made two lexical decision errors, both false alarms, both in the high-*N* nonword condition, responding YES to MAVE and RAME. Human data also show higher error rates in this condition than in the low-*N* nonword condition.

Because Andrews (1989, 1992) reported item means for all of these words, we can analyze her data from these items even though they are drawn from two different studies. ANOVAs of the YES latencies showed significant effects of word frequency, human data: $F(1, 92) = 94.305$, $p < .0001$; DRC data: $F(1, 92) = 129.636$, $p < .0001$, and of *N*, human data: $F(1, 92) = 5.015$, $p = .0275$; DRC data: $F(1, 92) = 13.490$, $p = .0004$, and a significant interaction between these factors, human data: $F(1, 92) = 9.161$, $p = .0032$; DRC data: $F(1, 92) = 4.856$, $p = .03$. For both sets of data the *N* effect was significant for low-frequency words, human data: $F(1, 46) = 7.790$, $p = .0076$; DRC data: $F(1, 46) = 13.382$, $p = .007$, but not for high-frequency words. Thus there was perfect correspondence between the effects of word frequency and *N* on

the YES latencies yielded by human readers and on the YES latencies yielded by the DRC model.

Human readers show slower NO responses for high-*N* than for low-*N* nonwords, and as Figure 8 shows this is also true for DRC NO latencies; the *N* effect was highly significant, $F(1, 92) = 75.466$, $p < .0001$.

The effects of lexicality, word frequency, and *N* on human lexical decision times are complex; essentially there is a three-way interaction. High *N* hurts NO and helps YES, but only helps YES when word frequency is low. The DRC model captures this three-way interaction perfectly.

Neighborhood Frequency Effects in Lexical Decision

The three-way interaction referred to in the previous paragraph, complex though it is, might not be sufficiently complex, because there is yet another variable to consider—neighborhood frequency. Andrews (1997, Table 4) reviewed nine studies that investigated whether lexical decision to words is influenced by the word having or not having a single higher-frequency orthographic neighbor, with the actual number of neighbors matched in these two conditions. Many of these studies found that the YES response in lexical decision is slowed by the presence of a higher-frequency neighbor, but in a substantial number of experiments a null effect was obtained, and in two experiments (Experiments 4A and 6 of Sears et al., 1995), the effect obtained was actually facilitatory rather than inhibitory. Even when inhibitory effects were observed, the situation is not straightforward, as Sears et al. (1995) pointed out.

The present results then pose a severe challenge to the existence of a "true" inhibitory neighborhood frequency effect (i.e., that having a higher frequency neighbor per se slows word processing, as Grainger and colleagues have suggested). The reader should also be reminded that even Grainger and colleagues have not consistently obtained the effects. Grainger (1990) observed a trend toward a facilitatory neighborhood frequency effect in a naming task. Grainger et al. (1992) only observed an inhibitory effect for words when the higher frequency neighbor was created by changing a letter in the fourth position and not when the higher frequency neighbor was created by changing the letter in the second position. Finally, as noted, Grainger (1992) reported an inhibitory neighborhood effect only for five-letter words. For four-letter words, an equally large facilitatory neighborhood effect was observed, producing an overall null effect. These results would also appear to call into question the existence of a true inhibitory effect of neighborhood frequency. (p. 879)

We are not suggesting that matters should rest here, because it is clear that if there actually were an inhibitory effect on YES lexical decisions of the presence of a high-frequency neighbor and a beneficial effect on YES lexical decisions of having many rather than few neighbors, this pattern of results might be particularly challenging to simulate with the DRC model. However, it is surely premature to consider this issue at the moment, because we do not know that there actually is an inhibitory effect on YES lexical decisions of the presence of a high-frequency neighbor. Another

¹⁴ As far as S and D are concerned, however, this is equivalent to a procedure according to which consultation of these criteria simply does not begin until cycle 20. That would have the same effect as setting the criteria so high prior to cycle 20 that they are never exceeded.

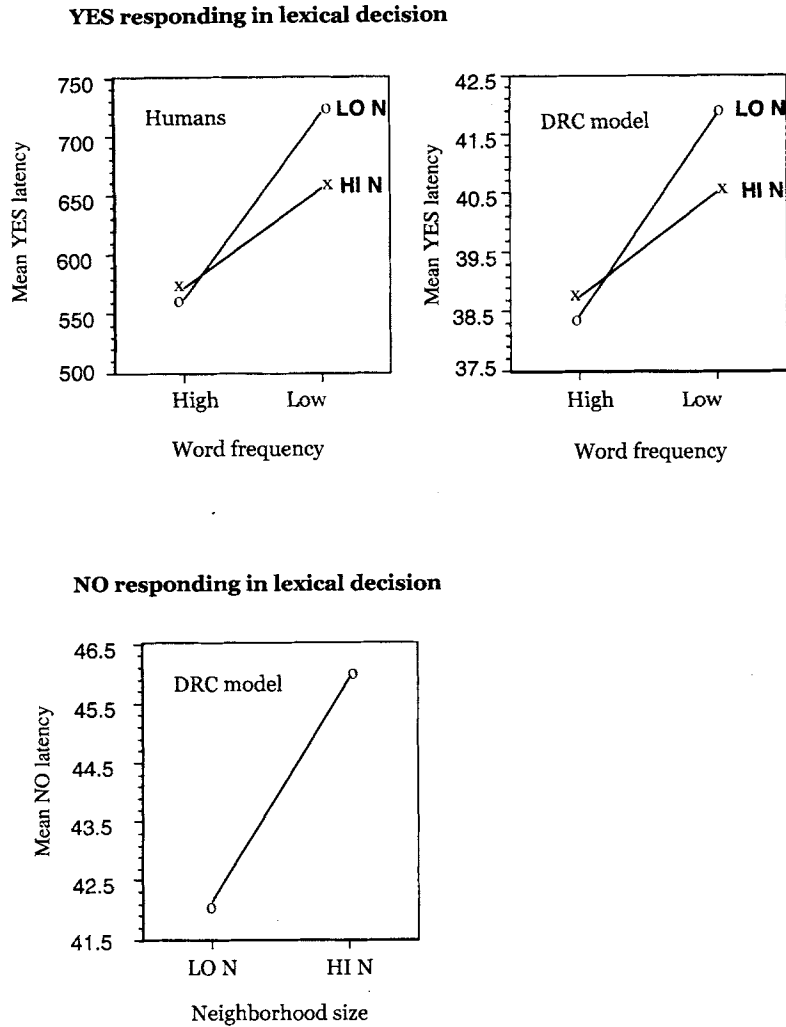


Figure 8. Effects of word frequency and neighborhood size on lexical decision times for human readers and the DRC model.

reason for caution here is that neighborhood frequency effects can be found in the Reicher–Wheeler forced-choice tachistoscopic recognition task (K. Paap, personal communication, April 2000). We have not yet studied this task with the DRC model, but that must be done, because the IAC model, from which the DRC model grew, was extensively applied to this task. When we do this, we will be confronted anew with the neighborhood frequency variable.

The PSH Effect in Lexical Decision

Rubenstein, Lewis, and Rubenstein (1971) discovered that NO responses in the visual lexical decision task were slower and had higher error rates when the nonword item was a PSH—a nonword whose pronunciation is that of a word, such as *brane*—than when the nonword item was both orthographically and phonologically a nonword, such as *slint*. This was confirmed by Coltheart et al. (1977).

Coltheart and Rastle (1994) presented a small demonstration that this PSH effect can occur when the DRC model is making lexical decisions, and they explained why the effect occurs with this model. When the model performs the lexical decision task, its NO response is based on a deadline whose magnitude depends on total lexical activation; therefore the PSH effect must arise in the model because the entry in the orthographic lexicon from which the PSH was generated (e.g., the *coat* entry when the nonword is *koat*) is getting activation (through the pathway GPC system to phoneme level to phonological lexicon to orthographic lexicon), and this activation will boost the total level of activation in the orthographic lexicon.

Coltheart and Rastle (1994) further pointed out that, although the orthographic lexical entry for *coat* is receiving excitation through the feedback pathway described above when the stimulus is *koat*, that orthographic lexical entry is also receiving some inhibition from *koat*, according to the DRC model. This inhibition arises because the letter detector for the letter *k* in Position 1

inhibits all entries in the orthographic lexicon for words that do not begin with the letter *k* (e.g., the word *coat*). However, the other three letters in the stimulus *koat* excite, rather than inhibit, the orthographic lexical entry for *coat*, and this will ameliorate the inhibitory contribution from the letter *k*.

Now consider a different and equally valid PSH of *coat*, namely, *kote*. The phonological feedback through the GPC route to the orthographic lexical entry *coat* will be as great for *kote* as it will be for *koat*.¹⁵ But three letters of *kote* inhibit the orthographic lexical entry for *coat*, whereas only one letter of *koat* does so; the other three letters of *koat* actually contribute excitation to the *coat* entry. It follows obviously that the activation of *coat*, and therefore the total activation in the orthographic lexicon, will be larger with *koat* than with *kote*. Therefore the PSH effect in lexical decision by the DRC model should be larger for the PSH *koat* than for the PSH *kote*. For these particular items, Coltheart and Rastle (1994) showed that this was so.

Of course, the nature of this balance between excitation and inhibition will be highly parameter dependent. If letter-to-word inhibition is very high, even a single-letter mismatch between the PSH and the orthographic lexical entry for its parent word will be enough to block the PSH effect on lexical decision; and if letter-to-word excitation is very high, even a single-letter match between the PSH and the orthographic lexical entry for its parent word will be enough to generate the PSH effect on lexical decision. Hence it is highly likely that if parameters are manipulated at will, many different outcomes are possible, including (a) no PSH effect at all and (b) a PSH effect that is independent of the degree of orthographic overlap between the PSH nonword and its lexical parent.

However, Coltheart and Coltheart (2000) measured the PSH effect on lexical decisions by the DRC model without adjusting any parameters at all. They took unchanged the parameter set and the lexical decision procedure from the simulation of neighborhood and frequency effects on lexical decision described above.

Coltheart and Coltheart (2000) selected 40 PSHs that were neighbors of their parent words (i.e., these PSHs differed from their parent words by exactly one letter), 40 control non-PSHs for these, 40 PSHs that were not neighbors of their parent words, and 40 control non-PSHs for these. All items were monosyllabic. The four sets of items were matched on number of letters, number of neighbors, number of body neighbors, number of body friends, number of body enemies, summed frequency of neighbors, summed frequency of body friends, positional bigram frequency, and positional trigram frequency. A set of 160 filler words matched to the nonwords on all of these variables was also chosen, and a lexical decision experiment was carried out with this material. They found that there was significantly worse performance with PSHs than with their controls (more errors and slower RTs) when the PSHs were neighbors of their parent words. With PSHs that were not neighbors of their parent words, pseudohomophony had no effect on speed or accuracy of the NO response.

Then they submitted these 320 items to the DRC model for lexical decision. Analyses of the DRC model's NO latencies indicated no main effects of pseudohomophony or orthographic similarity, but the interaction between these two factors was significant, $F(1, 156) = 4.602, p = .03$. Planned contrasts showed that PSHs did not differ from controls in the orthographically different condition (mean NO latencies = 42.15 vs. 42.30) but did differ significantly in the orthographically similar condition (mean

NO latencies = 42.90 vs. 42.15), $F(1, 77) = 6.65, p = .01$, exactly as was found with human readers.

Coltheart and Coltheart (2000) attributed this effect on the DRC model's NO latencies to feedback to the orthographic lexicon through the following route: letters to GPC rules to phoneme level to phonological lexicon to orthographic lexicon. They verified this interpretation by setting the activation level of the GPC route to zero (which means there is no longer any nonlexical activation of phonemes from letters) and resubmitting the same items for lexical decision. Now, the interaction between pseudohomophony and distance was no longer significant, $F = 0.00$, and there was no longer any effect of pseudohomophony in the neighbor condition, $F < 1$, nor, of course, in the nonneighbor condition, $F < 1$. These results show that in the DRC model the effect of pseudohomophony on lexical decision does indeed arise through the feedback pathway we have described. We propose that this is also the case for human readers.

The Consistency Effect

Consistency Versus Regularity

We have already discussed one stimulus variable based on the relationships of orthography to phonology, namely regularity, and have shown that the DRC model does an excellent job of simulating the results of various studies of human reading involving this variable and its interactions with other variables. A second and different variable based on the relationships of orthography to phonology is consistency.

The Definition of Regularity

A word is regular if its pronunciation is correctly generated by a set of grapheme–phoneme correspondence rules. Hence for some words there is room for debate about whether the word is regular but this will always be a debate about whether a certain GPC rule is appropriate. For any given set of GPC rules, there is no uncertainty about whether any word is regular. We treat as regular all monosyllabic words whose pronunciations are correctly specified by the GPC rules of the DRC model, and as irregular (exceptional) all monosyllabic words whose correct pronunciations are not generated by these rules.

Nonwords, of course, cannot be regular or irregular, because this concept depends on whether dictionary and rule-based pronunciations agree, and nonwords by definition do not have dictionary pronunciations.

¹⁵ For this particular example, this turns out not to be quite true: the example is contaminated by the whammy effect (Rastle & Coltheart, 1998), an effect of which Coltheart and Rastle (1994) were unaware. When a letter string with a multiletter grapheme—VUTH, for example—is being translated left to right, letter by letter by a GPC procedure, when that procedure has gone as far as dealing with the first three letters VUT it will generate the phoneme /t/ in the third position. When it gets to the fourth letter, it will generate the phoneme /θ/ in the third position. Because a different third-position phoneme has already been activated, the correct phoneme /θ/ will suffer some competitive inhibition—it will be “whammed”—and this will cause an increase in naming latency. Such increases are seen not only in the DRC model but also in human readers (Rastle & Coltheart, 1998).

The Definition of Consistency

The concept of consistency comes from Glushko (1979): "I used regular and exception words, but my definitions were not based on grapheme-phoneme rules; I defined a word as an exception if it had a different spelling-to-sound structure than other words with the same vowel and terminal consonants" (p. 684). Thus Glushko's definition of "inconsistent" was as follows: A letter string (word or nonword) is inconsistent if and only if the body of that letter string has at least two pronunciations in the set of words possessing that body. One thing to note is that on this definition, consistency is a categorical (all or none) variable. Another thing to note is that on this definition, "regular" and "consistent" are not synonyms, nor are "irregular" and "inconsistent." Plaut et al. (1996) attempted to conflate regularity and consistency by treating "regular" and "irregular" as two points on a continuum of consistency, but that is indefensible given Glushko's definition. Glushko himself appreciated that according to his definition of consistency there are irregular words that are consistent. His solution to this problem was as follows: "An unfortunate asymmetry here forces me to confess what I did with the 'exception and consistent' words . . . like LAUGH and MAUVE . . . I have chosen to ignore them" (Glushko, 1979, Footnote 3).

Other definitions of consistency exist. For example, various authors have defined consistency in a way that makes it a continuous rather than a categorical variable (e.g., Plaut et al., 1996; Jared, 1997; Rastle & Coltheart, 1999b), relating consistency, for example, to such variables as

- (a) the proportion of words that have the same orthographic body as the word or nonword that also have the majority pronunciation of that body;
- (b) the difference between number of friends (other words that have the word item's body and the same pronunciation of it) and number of enemies (other words that have the word item's body but a different pronunciation of it);
- (c) the summed frequency of friends minus the summed frequency of enemies.

It was originally believed that word naming was faster for consistent than for inconsistent words only when these were of low frequency. However, Jared (1997) argued that this may have been because of a confounding between word frequency and the existence of high-frequency orthographic enemies. When she carried out a study that controlled for the presence of high-frequency enemies, she found effects of both consistency and word frequency on word-naming latencies, but no interaction; high-frequency words benefited just as much from consistency as low-frequency words (Jared, 1997, Experiment 1).

Because many inconsistent words are irregular and many consistent words are regular, one might be tempted to argue that this result is actually a regularity effect, but that cannot be so, because only 30% of Jared's (1997) high-frequency inconsistent words were irregular, and only 30% of her low-frequency inconsistent words were irregular. These percentages are low enough that if the only variable at work here were regularity, no effect of "consistency" could have arisen. Jared's Experiment 4 completely unconfounded regularity from consistency and confirmed that there is a genuine effect of consistency per se: "Experiment 4 has demonstrated that a spelling-sound consistency effect is observed for high-frequency words with lower frequency neighbors even when there are no exception words in the list" (p. 521).

Table 3
The DRC Model's Naming Latencies for the Stimuli of Jared (1997), as a Function of Consistency and Frequency

Word type	High frequency	Low frequency
Consistent	75.63	79.50
Inconsistent	77.37	81.65

The existence of a consistency effect that is independent of the regularity effect would appear to pose a severe problem for the DRC model. The reason is this: consistency is defined in relation to the orthographic body of a word or nonword; the orthographic body is not a representational level in the DRC model, and therefore the model should predict no effect of consistency when it is unconfounded from regularity.

We investigated this by obtaining the DRC model's naming latencies for the words from Jared's (1997) Experiment 1.¹⁶ The mean DRC naming latencies in each condition are shown in Table 3. An ANOVA revealed a significant effect of frequency, $F(1, 74) = 4.027, p < .05$, and consistency, $F(1, 74) = 17.71, p < .0001$, but no interaction between these two variables, $F < 1$. That is exactly the pattern of results shown by Jared's readers. We also analyzed the naming latencies for these items from the three other existing computational models of reading aloud (which are discussed in more detail below), the PMSP model, the ZHB model, and Norris's (1994) model. The latencies of the PMSP and Zorzi models showed main effects of frequency and consistency and no interaction between these variables, as is also shown by the DRC and human latencies. However, the Norris model showed only a frequency effect; the consistency variable yielded $F < 1$. This occurred because all 40 high-frequency words and most of the low-frequency words were named in exactly one cycle by the Norris model, regardless of whether they were consistent or inconsistent.

How could this result with the DRC model have occurred? Why is the model showing an effect of consistency when it does not use the orthographic body as a processing unit? One possibility is that this result is due to the effects of a type of consistency to which the DRC model could be sensitive—neighborhood consistency. Because cascaded processing in the DRC model allows words orthographically similar to the target to be activated in the orthographic lexicon, the possibility exists that activation to the phoneme system from these neighboring items might help or hurt naming, depending on the phonemes that become activated by the neighbors. For example, naming of the highly inconsistent word *bough* will be greatly disadvantaged by activation sent to the phoneme system from its unfriendly neighbors *tough*, *rough*, and *cough*, none of which have any phonemes in common with the to-be-named word. Naming of the highly consistent word *rang*, on the other hand, will benefit from the fact that all 14 of its neighbors share two of its three phonemes.

An item's orthographic neighbors (on which the concept of neighborhood consistency is based) are often highly correlated

¹⁶ We excluded two words here: CLERK, because that is irregular in DRC's Australian English vocabulary but not in the Canadian pronunciation of Jared's subjects, and THOUGH, because inadvertently that word was not in DRC's vocabulary when this work was being done.

with its body neighbors (on which the concept of body consistency is based). In fact, in a recent analysis of Australian English monosyllables, Andrews (1997) found that 46% of the orthographic neighbors of four-, five-, and six-letter words are also body neighbors. This fact about English raises the possibility that the body-consistency effects that have been reported in the literature might actually be neighborhood consistency effects.

If this were the correct explanation of the effect of consistency seen in the DRC simulation of Experiment 1 of Jared (1997), then it would have to be the case that under the parameter set used (the parameters of Table 1), the entries for the neighbors of target words in the orthographic lexicon should be activated while the target word is being processed by the model. However, when we measured such neighbor activation, we found it absent; the high value of the letter-to-word inhibition parameter here caused enough inhibition from one mismatching letter to prevent neighbor activation.

So, as a direct test of the neighborhood consistency hypothesis, we reran the Jared (1997) stimuli through the DRC model but with the nonlexical route turned off (GPC activation set to zero). If the consistency effect is a property of the lexical route, as is the case if it is due to neighborhood consistency, the effect will still be present when the nonlexical route is inoperative. The mean DRC naming latencies in each condition are shown in Table 4. ANOVA revealed a significant effect of frequency, $F(1, 74) = 68.076, p < .0001$, but no effect of consistency, $F(1, 74) = .924, p = .34$, and no interaction between these two variables, $F < 1$.

It is not surprising that the frequency effect is much larger in this simulation than in the previous simulation of Jared's (1997) data, because in this simulation reading is being done solely lexically. But what is surprising is that the consistency effect has disappeared. It follows that in the DRC model the effect is being generated by the nonlexical route. Yet only a small percentage of the inconsistent words are irregular, as we have noted, and so the effect here cannot be due to irregularity. Why, then, are the regular inconsistent words suffering from their nonlexical processing?

There is a second property of words (and nonwords) that generates difficulties for their reading through the nonlexical route, the property of being whammied (Rastle & Coltheart, 1998; and see Footnote 15). Consider the regular word *drove*. As the nonlexical route operates on this word from left to right, there will be a point at which the first four letters have been translated through GPCs, and at that point the phonemes /drov/ will be active. When the nonlexical route gets to the fifth letter, the vowel phoneme now activated will be /ou/. But a different phoneme in the third position, /o/, is already active, and inhibition of /ou/ by /o/ will slow the rise of activation of the correct phoneme /ou/, and so slow the naming of *drove*. In contrast, with the regular word *drift*, no such

conflicts between competing phonemes—no whammy effect—will arise.

We therefore revisited Jared's (1997) stimulus material and classified each word as IW (irregular or whammied; these are the words for which the nonlexical route will generate difficulties) or RU (regular and unwhammied; these are the words for which there will be no difficulties generated by the nonlexical route). For high-frequency words, 55% were IW; for low-frequency words, the percentage was the same. So there is no confounding of frequency with the IW property. However, 75% of the inconsistent words were IW, whereas only 35% of the consistent words were IW, a highly significant difference, $\chi^2(1, N = 80) = 12.94, p < .001$. The same is true for the materials of Experiment 3 of Jared (1997); IW items are significantly more frequent in her inconsistent condition than in her consistent condition. So that is the source of the apparent consistency effect in the DRC simulations of Jared's data: it is a combination of an irregularity effect and a whammy effect, and that is why it vanishes when the nonlexical route is turned off.

We are not, of course, asserting that it follows that the source of the effect in human readers must be a combination of the irregularity effect and the whammy effect. But we have shown that this could be so. More generally, we have shown that the DRC model can simulate the Jared (1997) results and have discovered why; hence we have shown that there is no conflict between her results and the DRC model. Even more generally, for these reasons we believe that the body of experiments showing effects of consistency on reading aloud are compatible with the DRC model despite the fact that this model contains no level of representation specific to orthographic bodies.

“Fast Phonology,” “Strong Phonology,” and the DRC Model

The Horse Race Metaphor

The dual-route theory of reading aloud has sometimes been expressed (see e.g., Norris & Brown, 1985) in terms of a horse race, this metaphor being nicely described as follows by Frost (1998):

The lexical and the prelexical routines can operate independently of each other and . . . the winner of this race is determined by the speed and efficiency of the lexical or assembly process. . . . The “horse race” metaphor that depicts this state of affairs gives these processes a flavour of independence. (p. 86)

None of the present authors has ever described the dual-route theory in this way, however, because the most basic of data refute it. Consider what would happen when regular and irregular words are being read aloud, if the lexical and nonlexical routes were racing toward some winning post. If the lexical route wins, the word will be read correctly, regardless of whether it is regular or irregular, and there will be no difference in latency between the two types of word (because the lexical route is completely unaffected by whether a word is regular or irregular). If, on the other hand, the nonlexical route wins, a regular word would be read aloud correctly, but an irregular word will not: its pronunciation will be according to the nonlexical GPC rules and hence incorrect. It follows that, although irregularity will affect error rate, it will

Table 4
The DRC Model's Naming Latencies for the Stimuli of Jared (1997), as a Function of Consistency and Frequency, With the Model's Nonlexical Route Turned Off

Word type	High frequency	Low frequency
Consistent	73.95	78.30
Inconsistent	73.68	77.60

not affect the latency of correct reading-aloud responses. Yet a very large number of studies have shown that, at least for low-frequency words, latencies of correct responses are longer for irregular words than for regular words.

The relationship between the two routes in dual-route theory is thus not correctly characterized by the horse race metaphor. Nor is it correctly characterized by the term "independence." If two processing routes are independent, that means they share no processing components, and that again is something that no dual-route theorist has proposed. On the contrary, the lexical route and the nonlexical route share two processing components. One is the letter-identification system, because the input to both routes comes from that system. The other is the speech output (phoneme) system, because the output from both routes goes to that system.

As Frost (1998) pointed out, according to the horse race metaphor, "the derived phonology of a specific word in a specific experiment is either assembled or addressed" (p. 88). That is not at all what happens with the DRC model. In that model, the derived phonology for all reading-aloud responses (whether the input be a regular word, an irregular word, or a nonword) is the result of the inputs from both the nonlexical (assembled) and lexical (addressed) routes to the phoneme level. Exactly as Frost (1998) said, "phonology is always partly assembled and always partly lexical" (p. 95).

The horse race metaphor was seductive because it made it easy to understand such statements as "the lexical route is faster than the nonlexical route"—that would just mean that the lexical route wins the race more often than the nonlexical route. But if the horse race metaphor is inappropriate as a characterization of the dual-route theory, how should statements about the relative speeds of the two routes be understood?

Relative Speeds of the Two Routes and the Concept of "Fast Phonology"

The DRC model requires 76 cycles to read aloud the word *strap*. Figure 9 shows separately the nonlexical and lexical activations of the phonemes of this word during this period of 76 cycles.

Consider first the four plots labeled "nonlexical." These show the phoneme activations generated from the nonlexical route. The later a phoneme is in a word, the later is the time at which that phoneme begins to receive activation from the nonlexical route; that is because the nonlexical route operates left to right on the input string, with an interval of 17 cycles between when it begins to operate on letter n and when it begins to operate on letter $(n + 1)$. Activation of the first phoneme, /s/, is zero until Cycle 10 because the nonlexical route does not begin to operate at all until the tenth processing cycle. Nonlexical activation of the second phoneme, /t/, is zero until Cycle 27 because the nonlexical route moves on to treating the $(n + 1)$ th letter 17 cycles after it has begun treating the n th letter. So nonlexical activation of the third phoneme, /r/, is zero until Cycle 44, and nonlexical activation of the fourth phoneme, /a/, is zero until Cycle 61. The fifth phoneme, /p/, is not represented in these plots because the nonlexical route would not begin to activate it until Cycle 78, and the word has been uttered by then (on Cycle 76).

Next consider the plot labeled "lexical." This shows phoneme activation through the lexical route. Because that activation occurs in parallel, all five phonemes in the word receive the same activation from the lexical route on each cycle, so the activation plots for the five phonemes lie on top of each other—that is why there is only one plot for lexical activation.

Now we can consider the statement "the lexical route is faster than the nonlexical route." Is this true or false for the DRC model? It is clearly false as far as the first phoneme is concerned; that

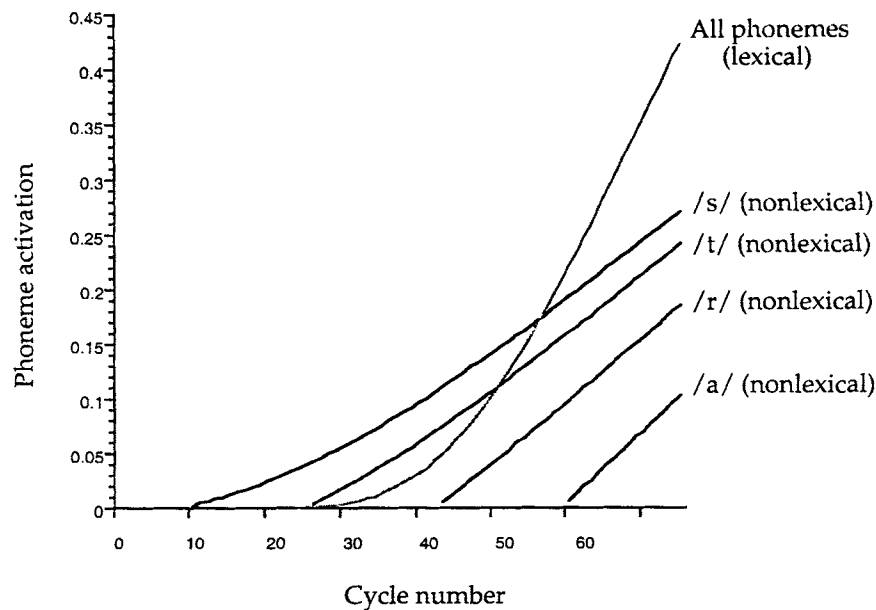


Figure 9. Lexical and nonlexical activation of the phonemes of the word STRAP during its reading by the DRC model.

phoneme begins to receive nonlexical activation on Cycle 10, but receives no lexical activation until much later (Cycle 25). But as far as the third phoneme is concerned, this statement is clearly true; that phoneme begins to receive nonlexical activation on Cycle 44, but is in receipt of lexical activation from Cycle 25 onward. And as for the second phoneme, it begins to receive nonlexical and lexical activation virtually at the same time. Thus the question “which is faster in the DRC model, the lexical route or the nonlexical route?” is ill posed.

Strong Phonology

This term was introduced by Frost (1995), who distinguished between a weak and a strong phonological hypothesis about reading. The weak phonological hypothesis “views the generation of phonological information from print as a process that may involve, in principle, both addressed and assembled phonology” (Frost, 1995), whereas “the strong phonological hypothesis does not make use of the notion of addressed phonology at all because phonology is never entirely addressed, but always computed” (p. 399). Frost discussed this distinction in relation to models of reading aloud only, but it applies more generally—it applies also to models of visual word recognition and to models of reading comprehension. A weak phonological theory of reading posits that reading tasks are not purely orthographic in nature and that all such tasks involve both orthographic and phonological processing. A strong phonological theory of reading posits that in at least some reading tasks processing is solely phonological and not orthographic at all.

All those concerned with building models of reading, at least for alphabetically written languages, agree that the reading system contains a letter (or letter cluster) level, a body of orthographic knowledge (essential for e.g., spelling), and a semantic system. Models differ with respect to how these three levels are linked and, hence, differ in the account they give of how various reading tasks are performed. We consider here three tasks: visual word recognition, (single-word) reading comprehensions, and reading words aloud. The DRC model offers a weak phonological theory of all three of these tasks and is inconsistent with a strong phonological theory of each of them.

Visual Word Recognition

A strong phonological theory for this task denies that there is a direct pathway from the letter level to the body of orthographic knowledge and asserts that communication between these two levels is always phonologically mediated. A weak phonological theory for this task asserts that there is a direct pathway from the letter level to the body of orthographic knowledge, but that there is also indirect phonologically mediated communication between these two levels. The DRC model offers a weak phonological theory for this task. For example, it explains the PSH effect in lexical decision as due to the activation of the orthographic lexicon (the body of orthographic knowledge) through an indirect phonologically mediated pathway (letters to GPC rules to phonemes to phonological lexicon to orthographic lexicon); it has a direct pathway from the letter level to the orthographic lexicon (the pathway denied by a strong phonological theory). We know of no evidence that favors the strong over the weak phonological theory

in relation to this task; if there were such evidence, that would count as evidence against the DRC model.

(Single-Word) Reading Comprehension

A strong phonological theory for this task denies that there is a direct pathway from the body of orthographic knowledge to the semantic system and that communication between these two levels is always phonologically mediated. A weak phonological theory for this task asserts that there is a direct pathway from the body of orthographic knowledge to the semantic system, but that there is also indirect phonologically mediated communication between these two levels. Phonological influences on reading comprehension have been the subject of much work in recent years, typically involving homophone and PSH effects in tasks requiring semantic processing, such as semantic categorization or judgments of semantic relatedness. With semantic categorization, it has been shown that performance on NO trials is influenced by homophony (A type of flower? *rows*) and pseudohomophony (A type of flower? *roze*), and corresponding results are obtained when the task is judgment of semantic relatedness (NO responses are difficult with stimuli such as *tulip rows* or *tulip roze*).

Such results are consistent with both a weak and a strong phonological theory of reading comprehension. The DRC model offers a weak phonological theory of these effects as follows (this analysis of phonological influences on reading comprehension was first proposed by Coltheart, 1980b). When the target item is a homophone of a category exemplar (A type of flower? *rows*) the pathway letters to orthographic lexicon to phonological lexicon to semantic system will generate activation of the semantic representation for that exemplar, which will cause difficulties in deciding on the response NO. When the target item is a PSH of a category exemplar (A type of flower? *roze*) the pathway, letters to GPC rules to phonemes to phonological lexicon to semantic system, will generate activation of the semantic representation for that exemplar, which will cause difficulties in deciding on the response NO. Provided that semantic activation through the direct route, letters to orthographic lexicon to semantic system (activation that will indicate that NO is the correct answer in these cases), begins to rise earlier than semantic activation through the indirect, phonologically mediated routes, the correct decision will usually be made, but the decision will be slowed by the spurious activation arising from the indirect routes. Note that the horse race analogy is again incorrect and misleading here: If there were a race between the direct and indirect routes, a race with a winner, then this kind of interference would not occur. But as mentioned above, the DRC model is not a race model.

Recent work has attempted to decide between the weak and the strong phonological theories of these effects on reading comprehension, and the verdict has gone in favor of the weak phonological theory. For example, consider Luo’s (1996) conclusion:

A strong version of the phonological recoding hypothesis claims that phonological recoding is not only mandatory, but also is in fact the only route to the internal lexicon (e.g., Gough, 1972; Hanson & Fowler, 1987; Lukatela & Turvey, 1991, 1994; Rubenstein et al., 1971; Van Orden, 1987, 1991). . . . Because the evidence collected so far falls short of a strong conclusion, it may be appropriate to take a reconciliatory position. The revised dual-route model shown in Figure 2 represents such an attempt. According to this model, there are

two independent routes to the lexicon (one orthographic and one phonological). (p. 891)

Luo's theory is identical to the weak-phonological DRC theory, except that Luo assumes that the phonological route is on average faster than the orthographic route; but the only basis for this assumption is a horse race conception of the two routes. Without that conception, the Luo/DRC analysis works even if the orthographic route begins to activate semantics earlier than the phonological route.

Folk (1999) also sought to decide between the weak and the strong phonological theories of these effects on reading comprehension, and she reached the same conclusion as Luo (1996):

The present data suggest that phonological codes play an important role in meaning activation during silent reading. However, there was also evidence that orthography interacted with phonology under certain circumstances to influence processing. This is consistent with traditional dual-route models that posit both orthographic and phonological routes to meaning (e.g., Coltheart et al., 1993; Ellis & Young, 1988; Patterson & Morton, 1985) and with numerous studies in the neuropsychological literature. There have been numerous studies of impaired patients who exhibited intact comprehension of written words while demonstrating deficits in their ability to generate the spoken names (e.g., Hanley & McDonnell, 1997; Miceli, Benvegna, Capasso, & Caramazza, 1997; Rapp, Benzing, & Caramazza, 1997; Shelton & Weinrich, 1997). (p. 904)¹⁷

Reading Words Aloud

A strong phonological theory for this task denies that there is a direct pathway from the body of orthographic knowledge to the phonological system and that communication from the letter level to phonology always and only uses sublexical mappings between orthographic and phonological segments. A weak phonological theory for this task asserts that there is a direct pathway from the body of orthographic knowledge to phonological system, but that there is also a system of sublexical mappings between orthographic and phonological segments.

It is difficult to see how the strong phonological theory of reading aloud could be reconciled with phonological dyslexia, the form of acquired dyslexia in which nonword reading is far more impaired than word reading; Frost (1998, p. 93) observed that such a theory "would have to assume that although matching whole-orthographic and whole-phonological units is not a default procedure of the cognitive system, it could in principle be done reasonably well," and that of course would no longer be a strong phonological theory of reading aloud, in the sense in which we are using the term "strong phonological theory." We know of no evidence that favors the strong over the weak phonological theory in relation to reading aloud—that is, no evidence to suggest that there is no direct pathway from the body of orthographic knowledge to the phonological system. If there were such evidence, that would count as evidence against the DRC model.

Frost (1998) has extensively reviewed the literature on phonological effects in reading and has documented an abundance of such effects. None of the effects he discusses, however, requires a strong phonological model of reading (as we have here defined "strong phonological model"); that is, he reports no evidence suggesting that the letter level does not communicate directly with the body of orthographic knowledge, no evidence suggesting that

the body of orthographic knowledge does not communicate directly with the semantic system, and no evidence suggesting that the body of orthographic knowledge does not communicate directly with phonology. Hence the data he reviews are consistent with a weak phonological model of reading (as we have here defined "weak phonological model"), and so are consistent with the DRC model. As is shown in various parts of this article, the DRC model's reading behavior shows numerous phonological influences—PSH effects in lexical decision and in reading aloud, phonological priming of reading aloud, the onset effect in masked priming, the regularity effect in reading aloud words, and strategy effects on reading of regular words.

We are making these claims solely about the reading of English. The facts about reading in other languages may, and in some cases must, be different. The Chinese, Japanese, and Korean writing systems are structurally so different from the English writing system that a model like the DRC model would simply not be applicable: For example, monosyllabic nonwords cannot even be written in the Chinese script or in Japanese kanji, so the distinction between a lexical and nonlexical route for reading aloud cannot even arise. The Hebrew writing system also differs fundamentally from that used for English because text is typically written without vowels. Frost (1995, Figure 2) has proposed an explicit model that suggests how Hebrew readers deal with this problem of missing vowels, a problem with which English readers are not confronted. This is a strong phonological model because it has no connections between whole orthographic word forms and whole phonological word forms. It could well be that a strong phonological theory of visual word recognition is correct for Hebrew even though a weak phonological theory of visual word recognition (such as the DRC model) is correct for English. There are no reasons to expect universals of written language, even if there are universals for spoken language; unlike spoken language, written language is culturally transmitted rather than innate, a recent rather than a long-standing human ability and an ability possessed even today by only a minority of the people in the world.

Even if we just consider languages written alphabetically with vowels present, we might find that one such language, Serbo-Croatian, differs from all other such languages in terms of reading mechanisms, because two alphabets, the Roman and the Cyrillic, are used to write it. This presents unusual problems to those speakers of Serbo-Croatian who are literate in both alphabets, because some letters are common to both alphabets but given different pronunciations (for review, see Lukatela & Turvey, 1998). Differences in the way English and Serbo-Croatian reading is acquired have been found; vowels are easier to learn than consonants in Serbo-Croatian, whereas the reverse is true in English (Ognjenovic, Lukatela, Feldman, & Turvey, 1983). The degree to which conclusions reached about how English is read apply to Serbo-Croatian, and vice versa, is thus an interesting empirical question.

Finally, consider languages written just with the Roman alphabet. We have extended our work with the DRC model to modeling

¹⁷ Note however that Folk argued that her data "suggest that, although both phonological and orthographic codes are involved in word recognition, phonological codes play a central role in meaning activation" (Folk, 1999, p. 904).

the reading of German (Ziegler, Perry, & Coltheart, 2000) and are also working on French and Italian. Although these and other languages written with the Roman alphabet vary in the proportion of irregular words they contain, all do contain some irregular words, and in all it is possible to write unfamiliar words and nonwords. Hence we expect the DRC approach to be appropriate for all such languages. That is not to say that appropriate parameter settings for the model might not be very different for different languages written in the Roman alphabet. For languages where few words are irregular, a highly active nonlexical route would not be as harmful as it would be for English; perhaps that is why Italian is read aloud more rapidly than English (Paulesu et al., 2000).

Comparisons With Other Computational Models of Reading Aloud

As we have mentioned above, there are four computational models of reading aloud of English currently in existence: in addition to the DRC model, there is the Norris model (Norris, 1994), the PMSP model (Plaut et al., 1996), and the ZHB model (Zorzi et al., 1998).¹⁸ All four models can generate model naming latencies for sets of words and nonwords. It is therefore possible to compare these models by selecting sets of items (words or nonwords) for which mean human naming latencies are available, and then studying how closely each model's naming latencies fit the human data. However, investigation of the Norris model's word naming latencies showed that 84% of the 3,130 words used in these analyses were named in exactly one cycle. Hence this model in its present form is not suitable for simulating individual item-naming latencies.

As discussed by Spieler and Balota (1997), there are two ways of assessing the goodness-of-fit of model to data, which we will refer to as the *factorial* and *regression* methods. To use the factorial method, an ANOVA of the model latencies and of the human latencies is carried out. Then a model is assessed by determining whether all the effects significant with the human latency data are also significant with the model latency data, and whether all the effects that are nonsignificant with the human data are also nonsignificant with the model data. To use the regression method (Spieler & Balota, 1997), the correlation across items between the human latencies and the model latencies is computed and squared; the squared correlation is a measure of the proportion of variance in the human latencies for which the model "can account." We have applied both methods (or just the regression method where the data did not come from a factorial design) to six databases of human naming latencies and three databases of computational model naming latencies.¹⁹ The human databases are as follows:

- (a) The Whammy database. This database of human nonword naming latencies is taken from Rastle and Coltheart (1998) and consists of mean naming latencies for 48 five-letter monosyllabic and pronounceable nonwords, 24 with five phonemes and 24 with three phonemes.
- (b) The Length database. This database of human word and nonword naming latencies was taken from Weekes (1997) and consists of mean naming latencies for 300 monosyllabic items—100 high-frequency words, 100 low-frequency words, and 100 pronounceable nonwords. Within each of these three
- sets of 100 items, there are 25 items at each of the lengths 3–6 letters.
- (c) The Position database. This database of human word naming latencies is taken from Rastle and Coltheart (1999b) and consists of mean naming latencies for 88 monosyllabic irregular words with irregular grapheme-to-phoneme correspondences at either the first, second, or third phoneme position, and 88 matched monosyllabic regular words.
- (d) The SB database (Spieler & Balota, 1997). This consists of human naming latencies for 2,870 monosyllabic words from the training corpora of the PMSP model and its predecessor, the SM model.
- (e) The WS (Wayne State) database (Treiman, Mullenix, Bijeljac-Babic, & Richmond-Welty, 1995). This consists of human naming latencies for 1,327 monosyllabic words with the phonemic structure consonant–vowel–consonant (CVC).
- (f) The SW database (Seidenberg, Waters, Barnes, & Tanenhaus, 1984). This consists of human naming latencies for 1,153 monosyllabic words with the phonemic structure CVC.

We report first the results using the factorial method with the first three databases, and then the results using the regression method with all six databases. In all cases, incorrect responses by the models, or responses to words that were not in a model's training set (this does not apply to the DRC model) were treated as missing data.

The Factorial Approach

The Whammy Database

Rastle and Coltheart (1998) found that human naming latencies for 48 five-letter nonwords that had either three or five phonemes were greater for the three-phoneme items than for the five-phoneme items and explained this effect in terms of a phenomenon they termed the "whammy effect," described earlier in this article.

One complication to which we should immediately draw attention is that these 48 nonword stimuli are particularly difficult for the PMSP and ZHB models because a number of the items end in consonant + final s. Because most plural monosyllables of English were excluded from the training set used for those models, the models had little experience of the pattern consonant + final s as they learned to read. An additional difficulty for the ZHB model with these items is that a number of them have orthographic bodies that do not occur in any item in the training set (e.g., *elst*), and body representations are relevant to the way the ZHB model learns. These are the reasons for the PMSP and ZHB models achieving only 75% and 48% accuracy, respectively, in reading these 48 nonwords. The DRC model read 47% of the nonwords correctly.

These accuracy data complicate the analysis. The two sets of nonwords were pairwise matched on *N*, initial phoneme, number of letters, number of body friends, number of body enemies, and the summed frequencies of body friends and of body enemies, so

¹⁸ A computational model of reading aloud in French also exists; see Ans, Carbonnel, and Valdois (1998).

¹⁹ We are grateful to David Balota, David Plaut, Daniel Spieler, Rebecca Treiman, Brendan Weekes, and Marco Zorzi for making human and modeling data available to us.

a within-item analysis of the whammy effect is appropriate. But if one member of a matched pair is wrongly read by a model, its matched twin should also be excluded from the analysis; if there are many wrongly read items, the number of item pairs included in the analysis becomes very small. It is not illegitimate to apply a between-item analysis to data of this kind (although it is likely to have lower power than a within-item analysis), so we report both types of analysis.

Mean latency data for the within-item and between-item analyses are shown in Table 5.

Human data. The human data yielded a whammy effect in the within-item analysis such that the whammied items (three phonemes) produced longer naming latencies than the unwhammied items (five phonemes), $F(1, 23) = 6.20, p = .02$; the between-item analysis yielded $F(1, 46) = 2.82, p = .10$. Given the greater power of the within-item analysis, we take this as evidence that human readers are susceptible to the whammy effect.

DRC data. Both the within-item analysis, $F(1, 23) = 8.54, p = .008$, and the between-item analysis, $F(1, 46) = 10.03, p = .003$, produced a significant whammy effect such that whammied items were read more slowly by the model than unwhammied items. Thus, the DRC model is, like human readers, susceptible to the whammy effect.

PMSP data. Five whammied items and seven unwhammied items were read wrongly by the model; these items were removed from the between-item analysis, and they and their matched controls were removed from the within-item analysis. The within-item analysis yielded, $F(1, 13) = 4.34, p = .058$, and the between-item analysis yielded, $F(1, 34) = 2.88, p = .099$. Hence, the whammy effect was not significant, but there was a strong trend toward significance, particularly in the more powerful within-item analysis.

ZHB data. Thirteen whammied items and 12 unwhammied items were read wrongly by the model. These items were excluded from the between-item analysis, and they and their matched controls were excluded from the within-item analysis. The within-item analysis yielded, $F(1, 5) = 2.5, p = .175$, and the between-item analysis yielded, $F(1, 21) = 3.22, p = .087$. Note that the whammy effect, nearly significant in the between-item analysis, is in the opposite direction to that shown by the human readers. Hence, this model does not show the whammy effect seen in the human data.

Table 5

Mean Naming Latencies for Whammied and Unwhammied Nonwords, for Human Participants and Three Computational Models

Source of data	Unwhammied	Whammied	Total N (MAX = 48)
Humans	567.6	585.5	48
DRC	157.4	167.4	47
PMSP	1.784	1.886	36
ZHB	3.917	3.364	23

Note. DRC = dual-route cascaded model; PMSP = Plaut, McClelland, Seidenberg, and Patterson (1996) model; ZHB = Zorzi, Houghton, and Butterworth (1998) model.

The Length Database

Weekes (1997) used a multiple regression approach so as to determine whether length affected naming latency when confounding variables, particularly N , were controlled. He found an interaction between length and lexicality: Length had no effect on naming latency for words, but was significantly related to naming latency for nonwords. We have analyzed these data using an analysis of covariance (ANCOVA) in which length and lexicality were treated as between-items factors, and N was treated as a covariate.

Human data. Naming latencies for two items were removed from the analysis of the human data because they are not monosyllabic (*trying* and *flying*). The ANCOVA revealed a significant interaction between length and lexicality, $F(3, 289) = 15.60, p < .001$, as there was a clear effect of length for nonwords but not for words. Figure 10 shows the form of this interaction. The main effects of length, $F(3, 289) = 14.95, p < .001$, and lexicality, $F(1, 289) = 189.73, p < .001$, were also significant.

DRC data. Four items, in addition to the two disyllabic words, were removed from the analysis. One item, *blest*, is not in DRC's vocabulary; the other three items were nonwords read incorrectly by the model (*fretch* read as *french*, *grite* read as *grit*, and *sush* read as *such*).

The ANCOVA revealed a significant interaction between length and lexicality, $F(3, 285) = 109.54, p < .001$. The main effects of length, $F(3, 285) = 61.16, p < .001$, and lexicality, $F(1, 285) = 3518.13, p < .001$, were also significant. Like the human data, there was an effect of length for nonwords but not for words. The form of the interaction, shown in Figure 10, is as in the human data.

PMSP data. In the ANCOVA, the main effects of length, $F(3, 254) = 3.46, p < .05$, and lexicality, $F(1, 254) = 19.34, p < .001$, were significant. There was a near significant interaction between length and lexicality, $F(3, 254) = 2.43, p = .066$. However, Figure 10 shows that the form of this interaction was quite different from that seen in the human data.

ZHB data. The ANCOVA yielded a nonsignificant interaction between length and lexicality, $F(3, 253) = 1.64, p = .18$. As shown in Figure 10, the model latencies show no suggestion of a larger length effect for words than for nonwords. The model did not produce an overall main effect of length, $F(3, 253) = .35, ns$, but it did produce a main effect of lexicality, $F(1, 253) = 103.37, p < .001$.

The Position Database

These data were analyzed using an ANCOVA in which regularity was treated as a within-items factor, and position was treated as a between-items factor. The covariates were N and consistency (see Rastle & Coltheart, 1999b, for details on how consistency was measured). Results are shown in Figure 11.

Human data. The ANCOVA yielded a significant Position \times Regularity interaction, $F(2, 83) = 11.12, p < .0001$, as the cost of irregularity declined monotonically across position of irregularity. The effect of regularity at position 3 was analyzed separately with N as a covariate; it was not significant, nor were the main effects of position and regularity.

DRC data. As with the human data, the ANCOVA yielded a significant interaction between position of irregularity and regu-

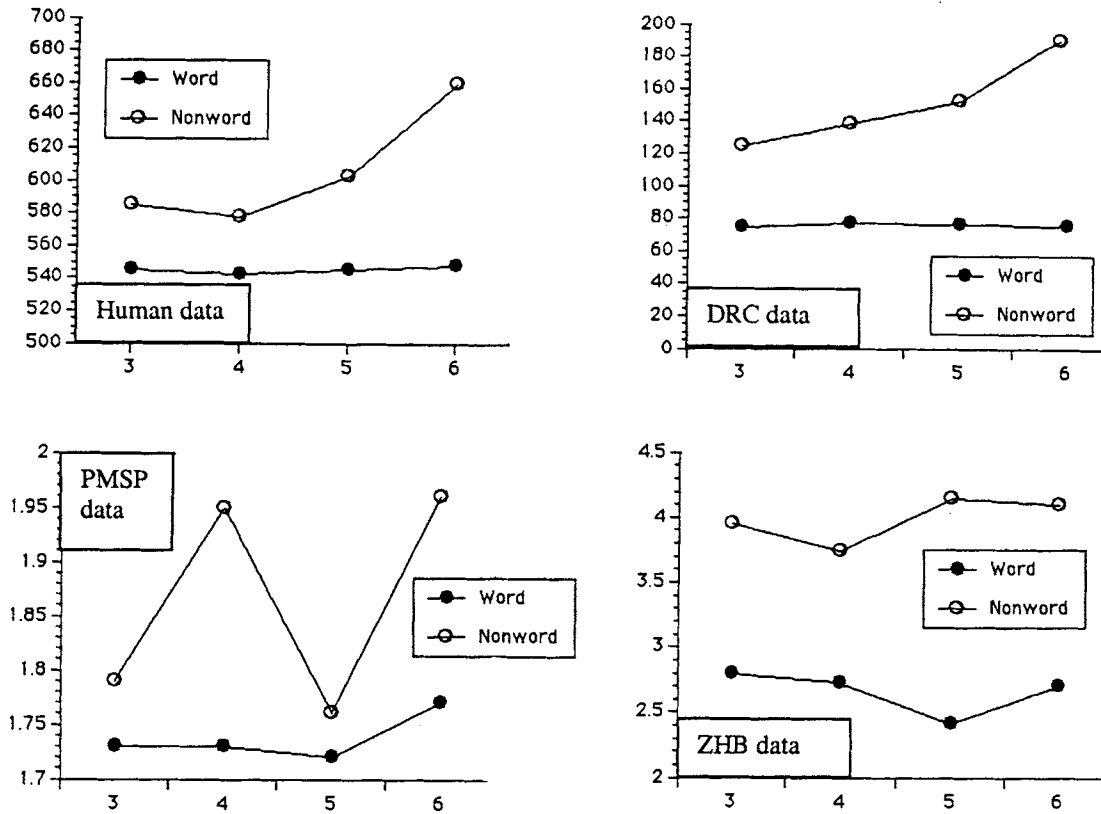


Figure 10. Effects of length and lexicality on naming latencies from human readers and the three computational models.

larity, $F(2, 80) = 33.78, p < .0001$, with cost of irregularity declining monotonically across position of irregularity to a non-significant effect, $F(1, 27) = 3.43, p = .075$. The main effects of position, $F(2, 80) = 47.20, p < .0001$, and regularity, $F(1, 80) = 43.67, p < .0001$, were both significant.

PMSP data. The ANCOVA did not yield a significant Position \times Regularity interaction, $F(2, 41) = .08, ns$. Neither the main effect of position, $F(2, 41) = .14, ns$, nor the main effect of regularity, $F(1, 41) = .13, ns$, was significant. Thus the PMSP model did not produce the same pattern of data as was produced by human participants—the cost of irregularity did not decline with position of irregularity.

ZHB data. As with the human data, the ZHB model yielded a significant interaction between regularity and position of irregularity, $F(2, 39) = 3.32, p = .047$, with the cost of irregularity declining monotonically over position of irregularity. The effect of regularity at position 3 was analyzed separately with N as a covariate; this effect was significant; $F(1, 4) = 6.44, p = .024$. Whereas the main effect of regularity was significant, $F(1, 39) = 6.80, p = .013$, the main effect of position of irregularity was not, $F(2, 39) = 2.18, ns$.

Summary of Results Obtained Using the Factorial Approach

With the whammy database, there was a whammy effect on the human latencies and on the DRC model latencies, but not on the

PMSP model latencies (though the effect was close to significance) nor on the ZHB model latencies.

With the length database, in the human data length and lexicality interacted such that there was a length effect on naming latencies for nonwords but no length effect for words. This was true also for the DRC latencies, but not for the latencies from the other two models.

With the position database, in the human data the cost of irregularity declined as the position of irregularity increased. This was also true of the DRC and ZHB model latencies, but not of the PMSP model latencies. For words with position 3 irregularities, there was no significant regularity effect in the human or DRC model latencies, but there was in the ZHB model latencies.

For all three databases, then, the DRC model provided a closer fit to human performance than either of the other two models. Indeed, in all cases every effect that was significant in the human latencies was also significant in the DRC latencies, and every effect that was insignificant in the human latencies was also insignificant in the DRC latencies.

The Regression Approach

All three models made a few errors when reading words aloud, so we eliminated from consideration any word that any of the models misread. We also eliminated from consideration any word that was not in both the PMSP model's training set and the ZHB model's training set as well as being in the DRC model's vocab-

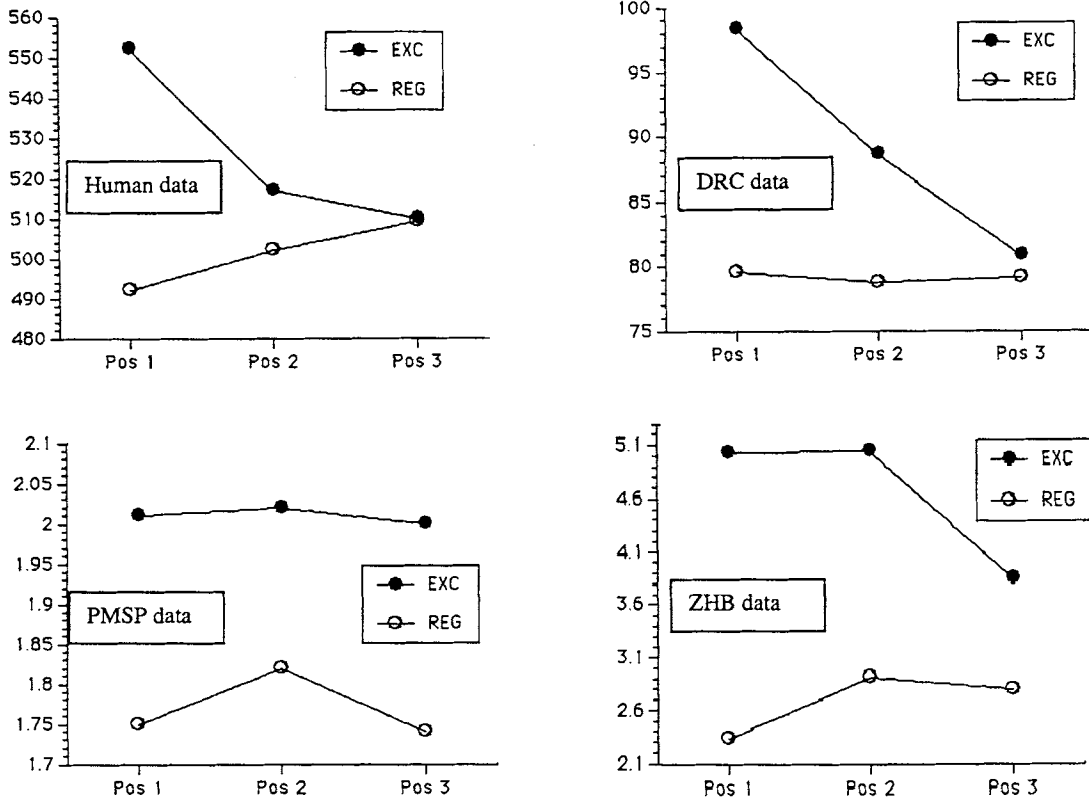


Figure 11. Effects of irregularity and position of irregularity on naming latencies from human readers and the three computational models.

ulary. This yielded a set of 2,516 words for which we had the latency of correct reading by all three models. Model naming latencies used in the regression analyses were drawn from this set.

The Whammy Database

Latencies for the items that each model read correctly were regressed against human naming latencies, and we calculated the percentage of variance in the human data accounted for by each model. Both the DRC model and the PMSP model accounted for a significant amount of the variance in the human data. The PMSP model accounted for 12.97% ($p = .03$) of the variance ($N = 36/48$) and the DRC model accounted for 12.44% ($p < .01$) of the variance ($N = 48/48$). The ZHB model did not account for a significant amount of the variance in the human data.

The Length Database

Latencies for correctly named items were regressed against human naming latencies for each model. All models accounted for a statistically significant percentage ($p < .05$) of the variance in human naming latency, but these percentages varied widely among the three models. The PMSP model accounted for 2.8% of the variance in human naming latency, the ZHB model accounted for 11.6%, and the DRC model accounted for 47.97%.

Nonwords played a large role in the high percentage of variance accounted for by the DRC model. Table 6 shows the percentage of

variance in human naming latency accounted for by each model when words and nonwords are analyzed separately.

All three of the models accounted for a significant percentage of the variance in the word data ($p < .05$), with the DRC model accounting for the largest percentage; for all three models, these percentages are small. Only the DRC model accounts for a significant percentage of the variance in nonword naming latency.

The Position Database

For the 176 items, model naming latency was regressed on human naming latency for all three models. The DRC model accounted for 13.08% ($p < .01$) of the variance in human naming

Table 6
Percentages of Variance of Word-Naming Latency and Nonword-Naming Latency Accounted for by the Three Models

Model	Words	Nonwords
PMSP	2.5%	0.1% (ns)
ZHB	3.1%	0.03% (ns)
DRC	4.5%	39.4%

Note. PMSP = Plaut, McClelland, Seidenberg, and Patterson (1996) model; ZHB = Zorzi, Houghton, and Butterworth (1998) model; DRC = dual-route cascaded model.

latency, the PMSP model accounted for 1.1% (nonsignificant), and the ZHB model accounted for 11.7% ($p < .01$).

Because exception words and regular words were pairwise matched, we regressed the model cost of irregularity (exception word latency minus matched regular control latency) on human cost of irregularity. When model costs are regressed on human costs, the DRC model accounts for 24.36% ($p < .01$) of the variance in human cost of irregularity, the PMSP model accounts for 0.7% (nonsignificant) of the variance, and the ZHB model accounts for 14.02% ($p < .01$).

The SB Database

All three models accounted for a significant percentage ($p < .05$) of the variance of the human naming latencies, although these percentages were all small: ZHB = 7.73%, DRC = 3.49%, PMSP = 2.54%.

The WS Database

Once again, all three models accounted for a significant percentage ($p < .05$) of the variance of the human naming latencies, although these percentages were all small: DRC = 4.89%, ZHB = 4.70%, PMSP = 1.67%.

The SW Database

Again, all three models accounted for a significant percentage ($p < .05$) of the variance of the human naming latencies, although these percentages were all small: DRC = 6.37%, ZHB = 6.33%, PMSP = 2.54%.

Summary of Results Obtained Using the Regression Approach

In all of these regression analyses involving word stimuli, the PMSP model accounted for considerably less of the variance of human word naming latency than did the two dual-route models (DRC and ZHB), whereas the latter two models performed comparably. Only the DRC model was able to account for any significant proportion of the variance of human nonword naming latencies in both nonword datasets; the PMSP model accounted for a significant proportion in one of these, and the ZHB did not account for a significant amount of variance in either.

It is not surprising that the ZHB and DRC models behave similarly on word reading, because, as Zorzi et al. (1998) noted, "For the lexical procedure, we assume that the output comes from an interactive activation network similar to that of Coltheart and colleagues (1993, 1994)" (p. 1149). Thus the phonological lexicon and its parallel activation of the phoneme level in the ZHB model are essentially the same as in the DRC model. It should be noted, however, that the lexical route in the ZHB model, like the semantic route in the PMSP model (but unlike the lexical route in the DRC model), is not fully implemented. There is no input from orthography to the phonological lexicon in the ZHB model. For the purposes of running simulations of word reading, the nonlexical route gets input from letters, but the response of the lexical route is mimicked by directly turning on the word's representation in the phonological lexicon of the model, not by providing the model with orthographic input.

Where the ZHB and DRC models do differ is in how the nonlexical route operates—for example, one is parallel and the other is serial. Here, the models perform quite differently. The DRC model accounts for significant proportions of the variances of nonword naming latencies, whereas the ZHB model does not.

The proportion of variance in nonword naming latencies accounted for by the DRC model was substantial, yet the proportions of variance in word reading latency accounted for by the ZHB and DRC models was not very high. Why? One reason—though clearly not the only one—is that there are factors affecting naming latency that lie outside the scope of these models, factors to do with early visual and late articulatory processes. However, such factors would affect nonword naming latencies too, and because the DRC model accounts for much less of word reading variance than nonword reading variance, these factors can only be a small part of the story. Nor can unreliability of word naming latencies be capping the amount of systematic variance that is there to be accounted for, because Spieler and Balota (1997) have shown that a regression using the three stimulus dimensions of length, log frequency, and neighborhood size accounted for 21.7% of the variance of word naming latencies in their human data set. Hence, even though the DRC model does rather a good job of explaining how the nonlexical route for reading works, there must be aspects of the lexical route for reading that are inadequately characterized by the DRC model. That provides a focus for future work with the model. For example, systematic study of the words for which the DRC model makes a good prediction of human naming latency versus the words for which it does not might identify word properties to which human readers are much more sensitive than the model, and this would provide guidance as to how to improve the model.

Be that as it may, the outcome of this work using the factorial and regression approaches for comparative evaluation of the three computational models of reading aloud is clear: The DRC model provides a much better account of human reading aloud than do the other two models.

Modeling Reading Impairments

Acquired Dyslexia

Any computational model of skilled reading should also aspire to be able to simulate various distinct patterns of acquired dyslexia. Cognitive neuropsychologists interpret such patterns as arising from various distinct patterns of preservations and impairments of the subcomponents of the skilled reading system (see, e.g., Coltheart, 1981, 1982, 1985). Computational cognitive neuropsychologists thus attempt to reproduce patterns of acquired dyslexia by lesioning specific components of their computational models and studying how closely the resultant acquired dyslexia in the computational model corresponds to the acquired dyslexia in the patient being simulated.

Plaut and Shallice (1993) described simulations of one acquired dyslexia, deep dyslexia (Coltheart, Patterson, & Marshall, 1980) but the computational model they were lesioning was not a model of normal reading (e.g., it had no procedure for reading nonwords), and so this work is not relevant here. The two forms of acquired dyslexia on which relevant computational modeling work has been done are phonological dyslexia and surface dyslexia.

Phonological dyslexia (see, e.g., Coltheart, 1996) is a condition in which after brain damage a previously skilled reader has a selective difficulty in reading nonwords aloud; for example, the patient WB (Funnell, 1983) was completely unable to read even the simplest CVC nonwords aloud after his stroke, whereas reading accuracy for words, even long low-frequency polymorphemic abstract words such as *preliminary* or *satirical*, was close to normal.

Surface dyslexia (Patterson, Marshall, & Coltheart, 1985) is a condition in which after brain damage a previously skilled reader has a selective difficulty in reading irregular words aloud; errors with such words are generally regularization errors. Nonword reading is relatively preserved, perhaps even normal in some patients (e.g., KT, see McCarthy & Warrington, 1986; or MP, see Behrmann & Bub, 1992).

To simulate extreme versions of these two acquired dyslexias is trivial with the DRC model. Turning down the activation parameter of the nonlexical route to zero will produce extreme phonological dyslexia; the model will have normal accuracy at word reading but will score zero percent on nonword reading.²⁰ Turning down an appropriate parameter of the lexical route to zero—for example, letter-to-word excitation—will produce extreme surface dyslexia. The model will have normal accuracy at nonword reading and regular word reading but will produce a regularization error with every irregular word.

Using a preliminary version of the DRC model (Coltheart et al., 1993), Coltheart, Langdon, and Haller (1996) attempted more challenging simulations of these two acquired dyslexias.

Surface Dyslexia and the DRC Model

With respect to the surface dyslexics MP and KT, Coltheart et al. (1996) sought to lesion the DRC model so that its nonword reading accuracy and regular word reading accuracy would remain normal, but accuracy of irregular word reading would be impaired but not abolished, with a greater impairment for low-frequency than for high-frequency irregular words, and with the errors for irregular words being regularization errors. This pattern of results was chosen because both MP and KT showed exactly this pattern, with KT being more severely impaired than MP. Coltheart et al. (1996) began with a DRC model that was normal at reading irregular words, regular words, and nonwords. They altered just one parameter of that model, reducing letter-to-word excitation to 25% of its normal value. The model's reading accuracy for irregular words, regular words, and nonwords now fitted the quantitative data from MP very well; reducing letter-to-word excitation still further, to 21% of its normal value, gave a model whose reading accuracy for irregular words, regular words, and nonwords now fitted the quantitative data from the more severe patient, KT, very well.

Here we take a slightly different approach to the simulation of surface dyslexia. First, because surface dyslexic patients are typically tested under unspeeded conditions, we used in our simulations of acquired dyslexia a value of .70 for the minimum phoneme activation criterion; that value simulates normal unspeeded reading aloud, whereas a value of .43 simulates speeded reading aloud. Second, we chose a different lesion site in the DRC model, the orthographic lexicon itself, rather than the excitatory connections to it from the letter level (which was the site chosen by Coltheart

et al., 1996, in their preliminary simulation). We did this because MP was not only surface dyslexic but also surface dysgraphic. In the version of the DRC model that will be capable not only of reading but also of spelling (see below for discussion of this extension of the model), a lesion of the orthographic lexicon should produce not only surface dyslexia but also surface dysgraphia.

Thus we are suggesting that both MP's acquired dyslexia and her acquired dysgraphia occur because entries in her orthographic lexicon have become less accessible, that is, more difficult to activate. Entries in the DRC model's orthographic lexicon can be made less excitable by increasing the Frequency Scale parameter. As in the IAC model, each word in the orthographic lexicon of the DRC model has a frequency constant associated with it, which varies in value from zero (for the most common word in the lexicon) to -1.0 (for the least common word), as we described earlier. Whenever the activation level of a unit in the orthographic lexicon is updated, adding this frequency constant is part of the updating (see Equation 7). The sensitivity of the model to frequency can be varied through the Frequency Scale parameter, by which the frequency constant is multiplied at each updating. The larger this Frequency Scale parameter is, the smaller will be the amount of activation added to the unit's activation level when the unit is updated, that is, the less responsive the unit will be to input. Hence we simulated a decrease in the responsiveness of units in the orthographic lexicon by increasing the value of the Frequency Scale parameter from 0.105 (its value in the standard parameter set) to 0.25.

The data from MP we have simulated are from Behrmann and Bub (1992, Appendix 1). These are data from a study of MP in which she was given regular and irregular words of varying frequencies to read aloud, plus the nonwords from Glushko (1979). Her performance with these items, and the performance of the DRC model with its lesioned orthographic lexicon, are shown in Figure 12. The correspondence between MP's data and the data from the lesioned model is extremely close.

More severe cases of surface dyslexia, such as KT (McCarthy & Warrington, 1986) can be simulated by further increasing the value of the Frequency Scale parameter, which has the effect of making the orthographic lexicon even less responsive to input. This is illustrated in Figure 13, which shows that the larger the value of this parameter the worse performance is with irregular words. At all values of this parameter tested, the model made no errors in reading regular words.

Phonological Dyslexia and the DRC Model

With respect to phonological dyslexia, Coltheart et al. (1996) successfully simulated an effect seen in some but not all phonological dyslexics (e.g., patient LB of Derouesné & Beauvois, 1985, and the adult developmental phonological dyslexic studied by Howard & Best, 1996), namely, that nonwords that are PSHs are read with greater accuracy than nonpseudohomophonic nonwords and that this PSH advantage is greater when the PSH is only one

²⁰ Here nonwords would be read as orthographically similar real words. That is the most common error type seen in the nonword reading of people with phonological dyslexia.

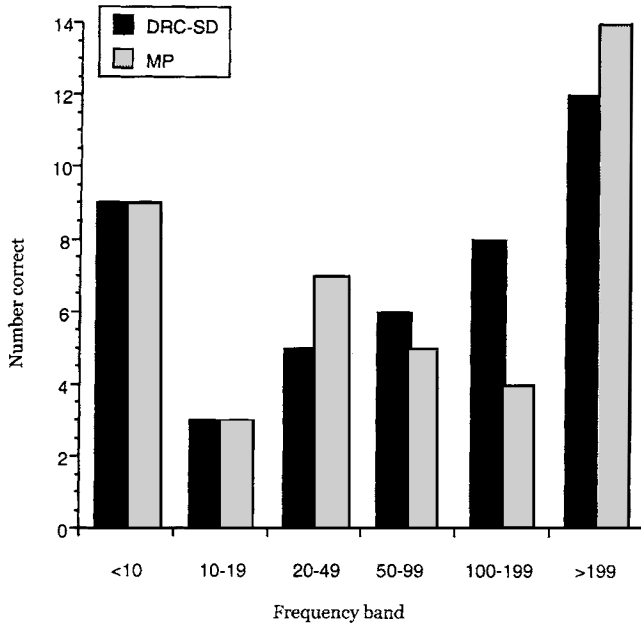


Figure 12. Simulation of the reading of the surface dyslexic patient MP.

letter different from its parent word than when the difference is greater. Exactly this pattern of results was seen in the nonword reading of the preliminary version of the DRC model when its nonlexical route was lesioned. This route moves from left to right across a letter string at a rate controlled by a parameter that specifies how many cycles elapse before the next letter in the string is processed by the nonlexical route. The successful simu-

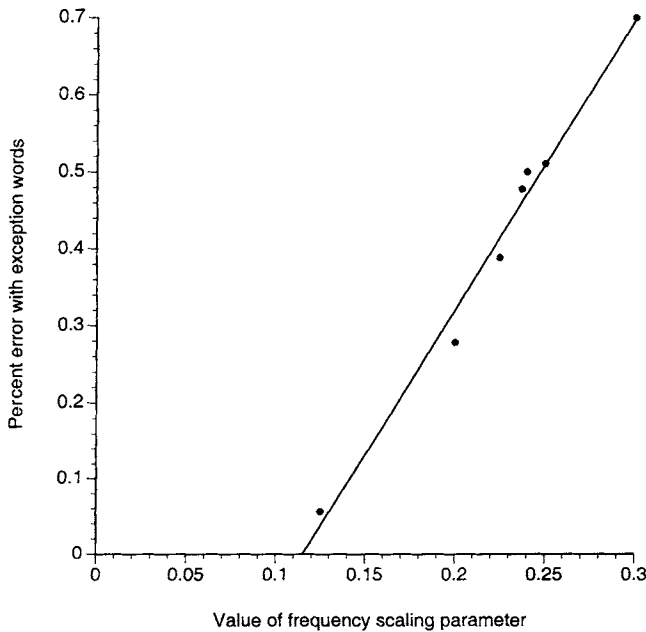


Figure 13. Simulation of surface dyslexia at varying levels of severity by a DRC model with an orthographic lexical lesion.

lation of phonological dyslexia was achieved by increasing this parameter. We did exactly this with the final version of the DRC model, increasing the parameter from its standard value of 17 cycles to a value of 27 cycles, and also, as with the simulation of surface dyslexia, using a phoneme activation criterion of .70 to simulate unspeeded reading aloud. Figure 14 shows data from the acquired phonological dyslexic LB (Derouesné & Beauvois, 1985; the adult developmental phonological dyslexic Melanie-Jane, Howard & Best, 1996, showed a very similar pattern) and the nonlexically lesioned DRC model. In all three cases, PSHs yield a higher level of accuracy than nonpseudohomophonic nonwords, and the effect is larger when the PSH is an orthographic neighbor of its parent word than when it is not. The PSH advantage with the DRC model latencies is significant for the condition in which PSHs were neighbors of their parent words, $\chi^2(1, N = 80) = 3.8$, $p = .05$, but not significant for the condition in which PSHs were not neighbors of their parent words; this was also true for Howard and Best's case.

Not all phonological dyslexics exhibit this PSH advantage. In the original study by Beauvois and Derouesné (1979), two phonological dyslexic patients showed this effect, and two did not. More recently, Berndt, Haendiges, Mitchum, and Wayland (1996) have reported a study of 11 phonological dyslexic patients in which the PSH advantage is shown by 6 patients and not by the other 5. Coltheart (1985) and Derouesné and Beauvois (1985) have argued that, because the nonword reading route has a number of distinct processing stages, there will be a number of different ways in which nonword reading can be impaired, and hence a number of different forms of phonological dyslexia. We argue that whether a phonological dyslexic will show a PSH advantage will depend on the nature of the damage to the nonlexical reading route. In our simulation, this damage takes the form of abnormally slow operation of an otherwise intact nonlexical route. This route therefore produces correct but weak excitation at the phoneme level. For PSHs, this abnormally weak excitation is boosted by top-down interactive activation from the entry in the phonological lexicon of the PSH's parent word; that is the source of the PSH advantage in our simulation of phonological dyslexia.

If instead the grapheme-phoneme correspondences themselves are damaged in the DRC model, the nonlexical route will produce incorrect rather than weak phoneme activations. In this case, a

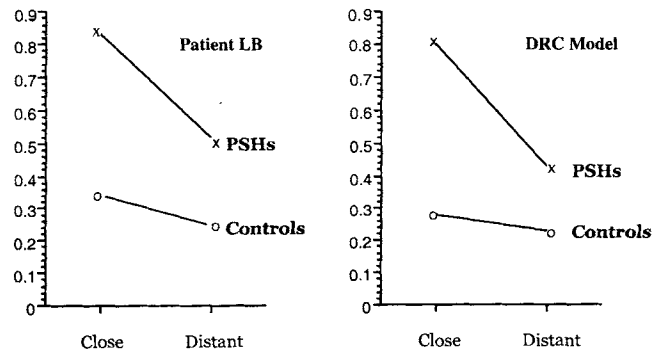


Figure 14. Simulation of phonological dyslexia by a DRC model with a lesion of the nonlexical route: Effects of pseudohomophony and orthographic similarity to base word.

PSH will not be able to excite the representation of its parent word in the phonological lexicon, and so here there will be impaired nonword reading but no PSH advantage.

A third form of acquired dyslexia, pure alexia or letter-by-letter reading (Coltheart, 1998) is perhaps worth mentioning here. In this disorder words are not read aloud by rapid immediate recognition; instead, reading is slow and arduous, often taking the form of naming the individual letters of a word from left to right. If all the letters are correctly named, the patient will then usually be able to say the whole word. Writing and spelling are typically intact. Behrmann, Plaut, and Nelson (1998) have discussed how pure alexia might be explained within the context of the IAC model (and hence, a fortiori, within the context of the DRC model) in terms of abnormally weak activation at the level of the letter units, although there have been as yet no actual simulation studies to investigate whether pure alexia can be successfully simulated in this way.

Future work will be required to determine how successful simulations of pure alexia by the DRC model will be; it is already clear that successful simulations of phonological dyslexia and surface dyslexia by the model have been achieved.

Acquired Dyslexia and the PMSP Model

Both surface and phonological dyslexia have been considered in the context of the PMSP model. One way to approach such simulation is to lesion the trained orthography-to-phonology attractor network (which, unlesioned, reads both exception words and nonwords very well) in order to try to simulate surface dyslexia. Plaut et al. (1996, pp. 92–94) explored this possibility in various ways. They were able to simulate the accuracy data from high- and low-frequency regular and irregular words from the surface dyslexic MP (Bub, Cancelliere, & Kertesz, 1985), but not the data from the more severe surface dyslexic KT (McCarthy & Warrington, 1986).

Hence Plaut et al. (1996) suggested a different approach. The verbal (i.e., unimplemented) form of their model includes two pathways from orthography to phonology, the direct orthography → phonology (OP) pathway (which they did implement) and an indirect pathway through semantics, the orthography → semantics → phonology (OSP) pathway (which has not been implemented). In relation to the use of these two pathways for reading aloud, they proposed a division-of-labor hypothesis. Presuming that the OSP pathway would only be able to read words (because nonwords have no semantic representations), Plaut et al. argued that

if the semantic pathway contributes significantly to the pronunciation of words, then the phonological pathway need not master all of the words by itself. Rather, it will tend to learn best those words high in frequency, consistency, or both; low-frequency exception words may never be learned completely . . . readers of equivalent skill may differ in their division of labor between the two pathways. In fact, if the semantic pathway continues to improve with additional reading experience, the phonological pathway would become increasingly specialized for consistent spelling-sound mappings at the expense of higher frequency exception words. (p. 91)

Given this division-of-labor hypothesis, explanations of acquired dyslexia can be approached by the PMSP model. Plaut et al. (1996) described this as follows:

A new network was trained in the context of an approximation to the contribution of semantics. Including a full implementation of the semantic pathway is, of course, beyond the scope of the present work. Rather, we will characterize this pathway solely in terms of its influence on the phoneme units within the phonological pathway . . . Over the course of training, the magnitude . . . of the input to phoneme units from the (putative) semantic pathway for a given word was set to be (a function of word frequency). (pp. 95–96)

Surface Dyslexia and the PMSP Model

The division-of-labor hypothesis led to an interpretation of surface dyslexia as follows: “At any point, brain damage that impaired or eliminated the semantic pathway would lay bare the latent inadequacies of the phonological pathway. . . . Surface dyslexia . . . seems to involve reading primarily via the phonological pathway because of an impairment of the semantic route” (Plaut et al., 1996, p. 92). In the trained division-of-labor network, the strength of the contribution of semantics to reading is a function of training epoch. Early in training, the phonological pathway is relatively good at reading irregular words—indeed, at around 400 epochs it was essentially perfect with high-frequency irregular words—but as training proceeds further and further, the phonological pathway becomes worse and worse with irregular words. In other words, this pathway becomes more and more surface dyslexic. After 400 epochs of training, the network’s performance when it was semantically lesioned was similar to the performance of the less severe surface dyslexic MP. After 2,000 epochs of training, its performance when semantically lesioned was similar to that of the more severe surface dyslexic KT.

This account obviously predicts that all patients with a severe semantic impairment should be surface dyslexic. However, this is not the case, as Plaut et al. (1996) acknowledged. Both WLP (Schwartz, Saffran, & Marin, 1979) and DRN (Cipolotti & Warrington, 1995) exhibited severe semantic impairments yet very good reading of irregular words. The account of this offered by Plaut et al. (1996) was as follows: “By our account, these observations suggest that, in these individuals, the phonological pathway had developed a relatively high degree of competence without assistance from semantics; but this post hoc interpretation clearly requires some future, independent source of evidence” (p. 99).

Here, Plaut and colleagues (1996) are drawing attention to the following problem: In order to reconcile the data from WLP and DRN with the PMSP model’s account of surface dyslexia, they need to suppose that individual differences in the competence of the phonological pathway of intact readers are so large that in some cases that pathway was premorbidly almost perfect at reading irregular words (WLP, DRN), whereas in other cases was poor, for example, capable of reading only about 70% of low-frequency irregular words (MP), and in other cases extremely poor, for example, capable of reading only about 25% of low-frequency irregular words (KT). That makes their theory unfalsifiable without some independent source of evidence about what the phonological reading pathway might have been like premorbidly in these patients.

Such evidence is not completely unattainable, however. According to the division-of-labor hypothesis, the more reading experience a person has, the greater will be the contribution of the semantic pathway to reading. One might therefore expect that people with a great deal of reading experience would have pho-

nological pathways that are rather poor at irregular word reading, and people who do little reading would have much more competent phonological pathways. So, for a fixed degree of semantic impairment, the more premorbid reading experience a person has had, the more severe their surface dyslexia should be. DRN and MP are an appropriate comparison here. MP was educated to secondary level, and her occupation was homemaker. DRN was a biological scientist with a tertiary education. It would be difficult not to conclude that DRN had much more experience with reading than MP. Because both had profound semantic impairments, it follows that DRN should have been much more surface dyslexic than MP, given the PMSP model's analysis of surface dyslexia. But the reverse was the case: MP had a clear surface dyslexia, whereas DRN was not surface dyslexic at all.

Plaut (1997) has approached this issue in relation to the patients DRN, discussed above, and DC (Lambon Ralph, Ellis, & Franklin, 1995), both of whom had semantic impairments but no surface dyslexia. DC left school at age 14 so is very unlikely to have had as much reading experience as the biological scientist DRN. Because neither was surface dyslexic, both must have premorbidly possessed a nonsemantic pathway for reading that was similarly good at reading low-frequency exception words, despite the presumably large difference in their premorbid reading experience. Simulations of surface dyslexia by removing the semantic input from a trained feed-forward network showed that relatively good reading of low-frequency exception words was present if either of two things had been true when the network was being trained:

- (a) Weight decay in the network had a low value, and the strength of the semantic input during training was high, or
- (b) Weight decay in the network was relatively high, and the strength of the semantic input during training was low.

Plaut (1997) aligned DRN with situation (a) and DC with situation (b), but this seems quite arbitrary: he gave no justification for the argument that someone who reads a lot would have a network with lower levels of weight decay than someone who does not, and no justification for the argument that someone who reads a lot would have a network with a higher semantic contribution to the phonological route than someone who does not. Why isn't the reverse just as plausible?

Phonological Dyslexia and the PMSP Model

Plaut et al. (1996) offered the following interpretation of phonological dyslexia in terms of the PMSP model: "Phonological dyslexia has a natural interpretation within the SM89 framework in terms of selective damage to the phonological pathway (or perhaps within phonology itself; see Patterson & Marcel, 1992) so that reading is accomplished primarily (perhaps even exclusively in some patients) by the semantic pathway" (p. 92). Our interpretation is the same: if the route on which nonword reading depends is damaged at any point after the letter level and before the phoneme level, words will still be read well, but nonwords will be read badly. The letter level and the phoneme level are needed for word reading, so damage at either of those levels would affect word reading as well as nonword reading, but even here nonwords might be more affected than words, because interactive support from orthographic lexicon to letter level, or from phonological lexicon to phoneme level, would help overcome the difficulties for words without assisting nonwords.

There is a difference between the two models, however, because in the DRC model the route on which phonological dyslexics are presumed to be relying for word reading has been implemented, whereas it has not been implemented in the PMSP model. Hence the DRC model (but not the PMSP model) can be used for quantitative simulation of data from phonological dyslexic patients, and such simulations have been successful, as we have reported above.

Our conclusion, then, concerning the simulation of acquired dyslexia is that the PMSP model has not successfully computationally simulated anything about phonological dyslexia, whereas the DRC model has.

As for surface dyslexia, the DRC model has succeeded in a genuine simulation, whereas the PMSP model cannot do so, because one needed component of such a simulation, the semantic route of the model, has not been implemented. To the extent to which this has been approached, the attempt to explain semantic impairment without surface dyslexia in patients with very different levels of premorbid literacy has led the PMSP model into some rather strange territory. Note that the association and dissociation between surface dyslexia and semantic impairment is handled very simply by the DRC model. Damage just to the semantic system will produce just a semantic impairment without surface dyslexia. Additional damage that affects the reading system—for example, damage to the orthographic lexicon, or to its links to the phonological lexicon—will produce semantic impairment accompanied by surface dyslexia.

Deep Dyslexia

The computational cognitive neuropsychology of reading investigates whether it is possible to lesion a computational model of the normal reading system in such a way that the model now makes reading errors that correspond to the kinds of errors made by patients with specific forms of acquired dyslexia. We have discussed in some detail successes achieved in lesioning the DRC model in a way that make its reading resemble the reading of people with surface dyslexia (selective impairment of the lexical route) and in a way that make its reading resemble the reading of people with phonological dyslexia (selective impairment of the nonlexical route). We have also mentioned work on a third form of acquired dyslexia, "pure alexia" (also known as letter-by-letter reading; see Coltheart, 1998, for a review of this form of acquired dyslexia); Behrmann et al. (1998) have discussed how pure alexia might be explained within the context of the IAC model (and hence, a fortiori, within the context of the DRC model) in terms of abnormally weak activation at the letter-identification level.

Might this approach also be applied to a fourth kind of acquired dyslexia, deep dyslexia (Coltheart, Patterson, & Marshall, 1980)? The characteristic symptoms of deep dyslexia are semantic errors in reading aloud (*ill* read as "sick"; *bush* read as "tree"), visual errors (*sword* read as "words"), and morphological errors (*fleeing* read as "flee"), particular difficulties in reading abstract words and function words, and a complete inability to read nonwords. Coltheart (1980a) contrasted two possible ways to explain this pattern of symptoms. One (Morton & Patterson, 1980) was to seek to show that some specific pattern of impairments of the components of the normal reading system would generate all of the error types evident in the reading of deep dyslexia. The other (Coltheart,

1980a; Saffran, Bogyo, Schwartz, & Marin, 1980) was to argue that deep dyslexics are reading, not with a damaged version of the normal reading system, but with a completely different reading system, located in the right hemisphere.

Subsequent research has provided very strong evidence for this right-hemisphere interpretation of deep dyslexia. Patterson, Vargha-Khadem, and Polkey (1989) studied the reading of a teenage girl who had been acquiring reading at a normal rate until she developed a brain disease that necessitated removal of her left cerebral hemisphere when she was 13 years of age. Her reading at age 15 showed all the symptoms that deep dyslexic patients show. As Patterson et al. (1989) remarked "the empirical conclusion is clear: adult deep dyslexics, who may be reading with the right hemisphere, and N.I., who must be reading with the right hemisphere, are strikingly similar" (p. 56). Michel, Henaff, & Intrilligator (1996) reported the case of AC, a 23-year-old college student who suffered a lesion of the posterior part of the corpus callosum, impairing communication between the two hemispheres. His reading was tested by briefly presenting words and nonwords to the left or right visual hemifields. With right-hemifield (left-hemisphere) presentation, no abnormalities in his reading were detected. With left-hemifield (right-hemisphere) presentation, his reading showed the symptoms of deep dyslexia. Weekes, Coltheart, & Gordon (1997) reported data from a brain-imaging study of a deep dyslexic patient, a surface dyslexic patient, and two normal controls that show that in a visual word recognition task there was more activation of the right hemisphere than the left for the deep dyslexic patient but not for the other three. Finally, Coltheart (2000) has argued that the data from another brain-imaging study of deep dyslexia (Price et al., 1998) also supports the right-hemisphere interpretation of deep dyslexic reading.

In the work on surface dyslexia, phonological dyslexia, and pure alexia discussed above, the idea has been that someone with acquired dyslexia is reading using a damaged form of the normal reading system (this system being located in the left hemisphere)—that is, the normal system but with one or more of its boxes or arrows damaged or deleted. Within cognitive neuropsychology, this is known as the assumption of subtractivity (Saffran, 1982; Caramazza, 1984), the assumption that brain damage does not add to, but only subtracts from, functional architectures of cognitive systems. That assumption needs to hold for a patient with acquired dyslexia if studies of that patient are to tell us anything about the normal reading system. It appears to hold for surface dyslexics, phonological dyslexics, and pure alexics, but not for deep dyslexics. If so, it is fruitless to seek to interpret deep dyslexia in relation to a model of the normal reading system, and thus fruitless to seek to simulate it by lesioning the DRC model, even though this is a fruitful enterprise in relation to the other three forms of acquired dyslexia. So the explanation of any symptom of deep dyslexia is outside the scope of the DRC model, even such intriguing symptoms as the occurrence of a PSH effect in lexical decision in deep dyslexic readers who cannot succeed in reading aloud any nonwords at all (Buchanan, Hildebrandt, & MacKinnon, 1994).²¹

Learning to Read and Developmental Dyslexia

Because the DRC model is not developed through any kind of learning algorithm, for reasons we explained earlier, it has nothing

to say about the actual process of learning to read. Nevertheless, it does offer a perspective on learning to read and developmental dyslexia. A child in the process of learning to read is a child in the process of acquiring the reading system whose architecture is shown in Figure 7. An impairment in learning to read could be an impairment in acquiring any one component of this architecture.

For example, some children might be acquiring the components of the lexical route at a normal rate, but be having difficulty with one or more components of the nonlexical route. Such children would have a selective difficulty in reading nonwords aloud. This is developmental phonological dyslexia, and it has been documented in various single case studies and group studies (Broom & Doctor, 1995; Campbell & Butterworth, 1985; Castles & Coltheart, 1993; Cestnick & Coltheart, 1999; Manis, Seidenberg, Doi, McBride-Chang, & Petersen, 1996; Snowling & Hulme, 1989; Snowling, Stackhouse, & Rack, 1986; Temple, 1984, 1997; Temple & Marshall, 1983).

Other children might be acquiring the components of the nonlexical route at a normal rate, but be having difficulty with one or more of the components of the lexical route. Such children would have a selective difficulty in reading irregular words aloud. This is developmental surface dyslexia, and it too has been documented in various single case studies and group studies (Broom & Doctor, 1995; Castles & Coltheart, 1993, 1996; Cestnick & Coltheart, 1999; Coltheart, Byng, Masterson, Prior, & Riddoch, 1983; Goulandris & Snowling, 1991; Hanley, Hastie, & Kay, 1992; Holmes, 1973, 1978; Job, Sartori, Masterson, & Coltheart, 1984; Manis et al., 1996; Temple, 1984, 1997).

The occurrence of these two broad subtypes of developmental dyslexia provides support for the view that the dual-route theory of reading of which the DRC model is a computational realization is useful as a framework for accounting not only for skilled reading but also for learning to read and developmental dyslexia.

Further support for this view is provided by quantitative studies of the prediction of accuracy of reading regular words in samples of young normal readers and dyslexic and brain-damaged children (Coltheart, Dufty, & Bates, 2000). In this work various data sets in which children read regular words, irregular words, and nonwords were analyzed. According to the dual-route framework, although there are three variables being measured here, there are only two underlying processes controlling the reading-aloud performance, the lexical reading procedure and the nonlexical reading procedure. It should therefore be possible to predict scores on any one of the reading tasks from scores on the other two. For example, percent correct on reading irregular words is a direct estimate of the competence of the lexical route, and percent correct on reading nonwords is a direct estimate of the competence of the nonlexical route. Because according to dual-route theory a regular word will be read correctly if either route succeeds, the proportion correct for regular words should be predictable from the proportions correct on irregular words and nonwords by the following equation:

$$P(\text{reg}) = P(\text{irreg}) + (1 - P(\text{irreg})) \times P(\text{nwd}) \quad (8)$$

²¹ Note that these authors explicitly comment that "The Right Hemisphere hypothesis can, however, accommodate the present findings since one need only claim that the right hemisphere is capable of implicit translation of orthography to phonology but is incapable of explicitly accessing that information" (Buchanan et al., 1994, p. 174).

Coltheart et al. (2000) applied this equation to eight data sets in which reading accuracy for regular and irregular words and non-words was measured in children aged 7 to 15. Some of these samples were normal readers, some were dyslexic, and one was a sample of children who had suffered strokes. In all of these eight samples, the correlations between predicted and obtained accuracy of regular word reading were uniformly and remarkably high, ranging from .825 to .980. Figure 15 shows a plot of predicted versus obtained accuracy of regular word reading for the entire set of 1,488 children. The correlation between predicted and observed regular word reading accuracy here is $+0.921$.

So although the dual-route framework is not a model of learning to read, it works well as a means of characterizing what a normal or dyslexic child has so far learned at any particular point in time during the course of learning to read. If that is entirely so, then the correlation between predicted and obtained regular word reading accuracy should be just as high for the youngest children in this sample as for the oldest. This turned out to be the case: the correlation was $+0.84$ for the 7-year-olds and $+0.78$ for the 15-year-olds.

This suggests that even very young or very poor readers do not have reading systems with architectures radically different from that shown in Figure 7. Instead, their reading systems are scaled-down versions of the dual-route system that the skilled adult reader uses.

Is the DRC Model Falsifiable?

Occasionally we have heard doubts expressed about the falsifiability of the DRC model. One way in which these doubts might be expressed is to argue that a model with such a large number of parameters can explain everything, and to explain everything is to explain nothing. We have four points to make here.

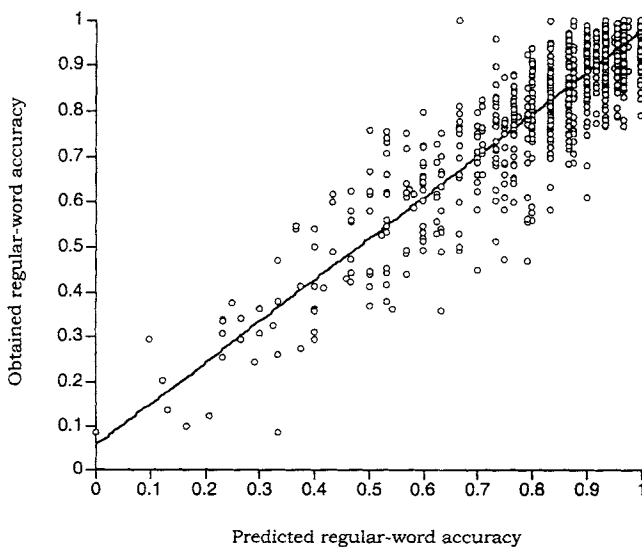


Figure 15. Prediction from dual-route theory of regular word reading accuracy from accuracy of reading irregular words and nonwords, in a sample of 1,488 children aged 8–14 years (some normal readers, some developmentally dyslexic, some with brain damage).

- (a) It is true that a mathematical model with 31 parameters can fit perfectly an extremely complex data set—any empirical function with up to 31 points in it. But that point is not relevant to the DRC model. The parameters of mathematical models are generally uninterpreted, whereas each parameter of the DRC model has a specific meaning. In mathematical modeling, the modeler is free to choose how many parameters to use to fit the data, but that is not so for DRC modeling. The number of parameters is dictated by the architecture of the model; given the architecture, the model cannot have fewer than, or more than, 31 parameters.
- (b) Fitting the DRC model to data is not an exercise in parameter estimation. As we have emphasized at various points in this article, what we did is to find a set of parameters that fitted one set of human data on reading aloud, and then showed that numerous other data sets were also well fitted by the model without changing any parameter. Here, for all but the initial data set, the model does not have 31 free parameters; on the contrary, it has no free parameters. Then we made a single—and rationally justified—change to this parameter set (reducing the strength of letter-to-word inhibition) and showed that human data from lexical decision tasks were well-simulated by the DRC model when it was performing the visual lexical decision task.
- (c) In the claim, “the model can explain everything,” what is meant by “everything”? If it means “everything we currently know about reading aloud and visual lexical decision,” then this can scarcely be taken as a criticism of the model. If instead it means “all logically possible outcomes of experiments on reading aloud and visual lexical decision,” then the claim is false. There are many possible outcomes of such experiments that would conflict with predictions from the model. Some of these outcomes could be dealt with by minor or major modifications of the model. For example, suppose Rastle and Coltheart (1999b) had found that the size of the regularity advantage in reading aloud was independent of the position in an irregular word of its irregular grapheme–phoneme correspondence. That refutes the idea that the nonlexical GPC route operates left to right, but it does not refute the idea that there is a nonlexical route that uses GPC rules.

Here, one could seek to build a new version of the DRC model in which the nonlexical route operates in parallel across the letters of the input string. Other logically possible outcomes of experiments would, however, refute the DRC model entirely. For example, there is no parameter set, and no version of the model, that could yield faster reading aloud of irregular words than regular words, or faster reading aloud of nonwords than regular words. Another outcome that would refute the model would be to find that when body consistency, as defined by Glushko (1979), is taken into account, regularity has no effect on reading aloud. Such an experiment would use regular and irregular words matched for body consistency (and of course other relevant factors). One way to do that is to use words that have unique bodies; all such words have an equal degree of body consistency. Thus if body consistency and not regularity were the critical factor here, irregular words like *yacht* and *waltz* would be no slower to read aloud than regular words like *blitz* or *gauze*. That result would refute any model

that proposes that there exists a reading route that uses GPC rules.

- (d) As we discussed earlier, a word like *chef* will be regularized by the DRC model if the GPC route is too active, whereas a nonword like *starm* will be lexicalized by the model if the GPC route is not active enough. If the DRC model were unfalsifiable, we would have known in advance that there would be a value for the level of activity for the GPC route that is low enough to avoid regularization errors for first-position irregular words like *chef* yet also high enough to avoid lexicalization errors for nonwords with last-position neighbors, such as *starm*. We did not know this; on the contrary, we went through a very large number of parameter sets that erred with *chef* or else erred with *starm* until we eventually located a region in parameter space for which both items were read correctly.

These are some of our reasons for rejecting the claim that the DRC model is unfalsifiable.

The history of the logogen model that we recounted earlier in this article is relevant to this issue of falsification. As we explained above, the original version of the logogen model, shown in Figure 4(a), predicted that picture naming should prime word reading at relatively long intervals, and this turned out not to be the case. So a new model was developed that was consistent with these priming data: that is the model shown in Figure 4(b). Was this a falsification of the logogen model? No, the correct way of describing what happened was that a logogen model was falsified, the model shown in Figure 4(a). The Figure 4(b) model is also a logogen model. The term "logogen model" refers to a class of models, to which all of the models in Figure 4 belong. Is this whole class of models falsifiable, or only specific examples from the class? Because the class has certain defining properties, such that processing is thresholded rather than cascaded, the class itself is falsifiable: evidence that processing is cascaded rather than thresholded would falsify the whole class. We think that this is in fact the case. As explained earlier, consistency effects on nonword reading aloud appear to be mediated through the lexical route, but nonwords could not get to phonology through the lexical route if processing in the visual word recognition system were thresholded, because that system would yield no output if the input were a nonword.

One can view the DRC model as a class of models, a falsifiable class, in just the same way. Any exemplar from this class can be false just because its parameter set is wrong, a change in the parameters might be all that is needed to reconcile it with all available data. Or the exemplar might be rather more seriously wrong at an architectural level; for example, data might be collected that show that orthographic word bodies influence nonword reading independently of graphemes. A new version of the DRC model in which the nonlexical route used nonlexical orthography-to-phonology rules that operated at the body level as well as rules operating at the grapheme level might succeed in simulating the new data. If so, the new data would falsify one exemplar of, but not the whole class of, DRC models. But the whole class itself is falsifiable, just like the whole class of logogen models, as we have shown in point (c) above.

The Issue of Model Complexity

Suppose that the DRC model and one of the other computational models discussed here turned out to have exactly equal explanatory

capacity—all the effects that the DRC model could simulate could be simulated by the other model, and vice versa. Because here the data do not discriminate between the two models, how might we choose which model to favor? Jacobs and Grainger (1994) proposed the criterion of model simplicity, described how this can be evaluated, and suggested that when two models have equal observational adequacy, what we should do is choose the simpler one.

That is certainly a defensible methodology, but we take a different methodological stance, namely, that when two models have equal observational adequacy, what we should do is collect more data with the aim of discriminating between the two theories. Unless the two models are only notational variants of each other, it will be possible to devise experiments about whose outcomes the two models make different predictions. When these experiments are done, it might be the simpler model that makes the correct predictions, or it might be the more complex. Our preference is to use the ability to make correct predictions, rather than the criterion of simplicity, as the basis for preferring one model over another.

Future Developments

Spelling to Dictation and Auditory Lexical Decision

Reading aloud by the DRC model begins with activation of a letter string's representation at the visual feature level and ends with activation of that letter string's representation at the phoneme level. However, because the model is an interactive one, activation of a letter string's representation at any level of the model will eventually create activation of its representation at all other appropriate levels. Thus turning on the phonemes representing the correct pronunciation of any word or nonword will lead to the activation of the letter units representing the spelling of that word or nonword, and that amounts to the task of spelling to dictation.

This is not quite so with the version of the model we have described so far, however. In this model, there is feedback from the phoneme level to the letter level on the lexical route, but not on the nonlexical route. So the model could spell words to dictation, but not nonwords. Currently we are adding a phoneme-grapheme correspondence (PGC) rule system to the model; in this version of the model, there is both lexical and nonlexical activation from phoneme level to letter level. This will allow the model to spell regular words, irregular words, and nonwords.

Just adding a PGC system, however, is insufficient. If one attempts to simulate spelling to dictation by turning on the appropriate phoneme units and awaiting activation of corresponding letter units, what can happen is that activations of the phoneme units representing the stimulus decline to zero before adequate activation of the letter units occurs. This is because there is interactive activation between phonological lexicon and phoneme level. This interactive activation can sometimes switch phoneme units off. That is why the communication from visual feature level to letter level in the IAC model, and hence in the DRC model, is one-way. There is no feedback from letters to the feature level, so nothing can turn feature units off; once they are turned on, they are effectively clamped on. If spelling to dictation is to be possible by the model, there has to be a similar arrangement at the phoneme level. Hence we have implemented a phonetic feature level that has excitatory and inhibitory links to the phoneme level, but receives

no input from that level. The spelling-to-dictation task is carried out by initially activating the sets of phonetic feature units corresponding to the phoneme string to be spelled. These units will remain on, and spelling will have been achieved when units at the letter level have reached an adequate level of activation.

This will allow us to explore simulation of acquired dysgraphias such as surface dysgraphia (Weekes & Coltheart, 1996), a selective difficulty in spelling irregular words with relative preservation of nonword spelling, and phonological dysgraphia (Shallice, 1981), a selective difficulty in spelling nonwords to dictation with relative preservation of word spelling. We will also be able to explore various results that implicate feedback interactions from phonemic to orthographic levels. One of these is the orthographic effect on auditory rhyme judgments reported by Seidenberg and Tanenhaus (1979). A second is the feedback inconsistency effect in lexical decision and naming (Ziegler, Montant, & Jacobs, 1997). A third is the effect reported by Lesch & Pollatsek (1998). The task in their experiment was to judge whether two printed words were semantically related; they found that readers were slower to respond NO to items like *pillow bead* than to items like *pillow bend*. The point here is that although *bead* is not a homophone of "bed," nor is it pronounced "bed" by the nonlexical route of a dual-route model, there are words such as *head* or *dead* in which the segment *ead* is pronounced /ed/. A DRC model with both lexical and nonlexical feedback from phoneme to letter level should be capable of simulating this effect, as follows. The word *bead* will excite a number of neighbors such as *head* and *dead* in the orthographic lexicon, hence activating their entries in the phonological lexicon and hence activating the phoneme /e/ in the second position and the phoneme /d/ in the third; the phonemes /b/ in the first position and /d/ in the third will have been activated through the orthographic and phonological lexical entries for the actual stimulus *bead*. Because all of the phonemes for the word "bed" will be active at the phoneme level, the excitatory links from phonemes to phonological lexicon will activate the phonological lexical entry for "bed" and hence its semantics. Further access to the semantics of "bed" will be provided by activation of the letters *e* in the second position and *d* in the third position through the PGC rule route applied to the phonemes /e/ in the second position and /d/ in the third, and these letter activations will contribute activation to the *bed* entry in the orthographic lexicon. That explanation is well supported by the finding of Lesch and Pollatsek (1998, p. 578) that the more orthographic neighbors of the false homophone (here *bead*) there are that support the pronunciation of the semantically related item (here *bed*), the longer the NO latency is.

The DRC model extended in this way will also be able to perform the auditory lexical decision task, applying to the activation of the phonological lexicon the decision criteria described earlier in connection with the model's performance of the visual lexical decision task.

Polysyllabic Words

All of the current computational models of reading English, including the DRC model, are restricted to the processing of monosyllabic stimuli.²² The problems encountered when items of more than one syllable are considered are numerous; when such items are considered, one is forced to develop hypotheses about how a nonlexical procedure for translating from orthography to

phonology could accomplish the placement of stress and the reduction of vowels. These problems have been underrepresented in the psycholinguistic literature and virtually ignored in modeling reading.

If a dual-route model is to be applied to polysyllabic word reading, then it must be shown how the nonlexical route deals with the difficulties encountered when these words are considered. What are the nonlexical rules for assigning stress and reducing vowels? If such a set of rules can be identified, then it seems clear that some words would follow those rules and other words would not, and thus a stress regularity effect may emerge in reading aloud.

The approach adopted by Rastle and Coltheart (2000a) to developing a set of rules for the nonlexical route was based on the work of Garde (1968) and the more recent work of Fudge (1984), who took the view that certain orthographic patterns may function as morphemes, and these morphemes may influence the placement of stress. Rastle and Coltheart (2000a) thus designed an algorithm that searches a letter string for the presence of a possible affix and computes phonology, assigns stress, and reduces vowels according to a series of rules. Because this algorithm is meant to function as a nonlexical route, it can be applied to nonword reading.

Rastle and Coltheart (2000a) found that the algorithm that they developed predicted the assignment of stress to disyllabic nonwords well when tested against a group of human readers. Moreover, when they classified words as irregularly or regularly stressed on the basis of the algorithm, they found that those items classified as irregularly stressed were named more slowly than those items classified as regularly stressed, particularly if they were of low frequency.

Potential Problems for the DRC Model

Masked Phonological Priming Effects

With various masking paradigms, evidence has been reported that phonological properties of very briefly presented stimuli can affect performance in reading aloud and lexical decision tasks. Perfetti and Bell (1991) and Perfetti, Bell, and Delaney (1988) investigated report of a target presented very briefly (for, say, between 35 and 55 ms) and followed by a brief backward mask (a letter string) and then a pattern mask. They found that target report was more accurate when the mask that immediately followed the target was phonologically identical to it (blue BLOO XXXX) in comparison with an orthographically related control condition (blue BLAR XXXX). Using a different kind of masking design, Lukatela, Frost, and Turvey (1998) reported that when a prime was briefly presented and then backward-masked by a subsequent target word, lexical decision to the target word was faster when the prime was a PSH of the target (KLIP clip) than when it was a matched control (PLIP clip).

Both types of study thus suggest that a phonological representation of a letter string (the target in one case, the prime in the other) is available very early because these letter strings were presented so briefly. We have already discussed how and why

²² See Ans, Carbonnel, and Valdois (1998) for a computational model of reading French that considers the issue of reading polysyllabic words.

early activation of phonological information occurs with the DRC model, so evidence of such activation is not intrinsically problematic for the DRC model. However, whether the model could actually simulate these effects needs to be investigated, and this cannot be done at the moment because the model lacks an adequate account of how masking occurs—that is, an account of how one briefly presented letter string is influenced by, and influences, a subsequent briefly presented letter string. The only account of masking offered by the model involves residual activation from the first string affecting the processing of the second string—that is how the onset effect was explained—but there is certainly more to masking than just this mechanism.

Even if the DRC model were extended by adding to it a computational model of backward and forward masking, however, there currently exist some difficulties concerning exactly what the effects are that would need to be simulated. This is so for both of the masking paradigms discussed above. For example, both Brysbaert and Praet (1992) and Verstaen, Humphreys, Olson, and d'Ydewalle (1995) found that the effect of phonological identity of a backward mask on report of a target—the phenomenon investigated by Perfetti and colleagues (1988)—only occurs when a very large percentage of the trials involve phonological identity of target to mask, or when the experimental situation encourages in other ways the use of phonological information. Without such encouragement, the phonological priming effect does not occur. Hence this effect cannot be due to there always being rapid automatic phonological encoding of briefly presented letter strings; and indeed Verstaen et al. (1995) expressed serious doubts about the relevance of this paradigm to normal reading processes: “Our study shows that a particular reading strategy may be adopted in the backward masking paradigm, favouring phonemic effects. Since the orthographically determined process of word recognition is highly vulnerable to the effects of backward masking, subjects are encouraged to use a phonological procedure that is relatively invulnerable to the effects of pattern masking. Therefore it is not justified to claim that phonological information is always used in word identification on the basis of phonemic effects in backward masking” (pp. 351–352).

With respect to the klip CLIP effect referred to above, Lukatela, Frost, and Turvey (1998, Footnote 2) observed “For unknown reasons, the dim lighting proves to be crucial. Pilot work failed to find any differences, at very brief prime durations, among full and partial phonological primes under the conditions of high illumination in our research room which had previously served as an office. It was only by reducing the room illumination to that provided by a single desk lamp at floor level that we could obtain reliable priming differences.” However, Lukatela, Frost, and Turvey (1999) reported that masked identity priming was greater for consistent words such as *bent* than for inconsistent words such as *bowl*, an effect they ascribed to phonological processing of the brief prime—yet this experiment was done “in a well-lit room” (Lukatela et al., 1999, p. 778). It seems essential to discover what the unknown reasons are for the need to have very dim lighting to obtain the klip CLIP effect before one would attempt to simulate this result. The situation is made even more complicated by a finding reported by Lukatela and Turvey (2000), namely, that masked priming by a PSH did not occur when PSH primes differed from targets by their vowel spellings (e.g., NAIM–name) but did when they differed in consonant spellings (e.g., KLIP–clip)—

indeed, in their Experiments 2 and 3 the vowel PSHs actually slowed target processing.

Hence the implementation of a computational account of masking effects in the DRC model would need to be accompanied by considerable further empirical work on masked phonological priming to establish exactly what the effects are that need to be simulated.

Orthographic Bodies and Consistency

We have argued that there is not yet any convincing evidence that the orthographic body is a level of representation in the human reading system, but many authors (e.g., Taft, 1991) do consider that this is so, and convincing evidence may well emerge. If it does, then the DRC model in its present form will have been refuted. Modification of the model, however, might not be difficult. One obvious modification would be to relax the requirement that the nonlexical rules operate only with graphemes and phonemes and to add to the rule set a set of body-to-rime rules, and perhaps also a set of head-to-onset rules. Such a model would still be a dual-route model in that it would still possess lexical and nonlexical procedures for reading aloud.

Lexical Decision

The procedure by which the DRC model makes lexical decisions is still extremely crude, particularly with respect to NO decisions, the latencies of which can at present have only two possible values: the default value, or the extended value adopted when overall orthographic activation is high early in processing. This coarseness may well prevent the model from being capable of simulating some subtle effects on lexical decision NO responses. A more continuous procedure, whereby the deadline for NO responding is adjusted afresh on each cycle rather than just on one occasion, may turn out to be needed.

Amount of Variance of Word Naming Latencies Accounted for by the DRC Model

As shown earlier, this is disappointingly low (especially when compared with the amount of variance of nonword naming latencies accounted for), although statistically significant and no worse than it is with the other two computational models we have discussed. Because the problem is specific to words, it must presumably be caused by an inadequacy of some part of the model that is specific to words (so not caused, e.g., by the choice of representational scheme for letters or for phonemes). Perhaps the simple system by which the letter level is connected to the orthographic lexicon is at fault; one possibility worth exploring is the insertion of an intermediate level of representation, perhaps orthographic heads and bodies as advocated by Taft (1991). Computationally, this is perfectly feasible and easily evaluated, but does it substantially increase the amount of variance of word naming latencies accounted for?

Polysyllabic Words

As described above, Rastle and Coltheart (2000a) have done some preliminary work relevant to extending the DRC model beyond the monosyllable, but much more work needs to be done

in this domain, and many potential problems are evident. For example, how might we reconcile the nonlexical stress-assignment algorithm proposed by Rastle and Coltheart (2000a)—which makes use, in part, of the ends of words—with a nonlexical system that operates serially? The representation of lexical stress in the model may also cause difficulty: in particular, should stress be stored in each phonological lexical entry, or should segmental and suprasegmental information be stored separately?

Conclusions

The model of visual word recognition and reading aloud we have described in this article is the most recent development in a modeling endeavor that began more than a century ago, with the 19th-century cognitive neuropsychologists who believed that the language-processing system was best modeled in terms of a multicomponent modular cognitive information-processing system that included not only a semantic system but also lexicons, modules containing word-specific local representations. This approach languished in the first half of the 20th century, but was revived by Morton and Treisman in the 1960s. Morton went on to develop a detailed version of his logogen model that dealt with visual word recognition, auditory word recognition, spelling, and spoken word production. A generalization of this model—one that was not tied specifically to the concept of the logogen—was offered by various authors in the 1980s (e.g., Ellis & Young, 1988; Harris & Coltheart, 1986; Patterson & Shewell, 1987).

The first step toward making this model computational was taken by McClelland and Rumelhart (1981) because their IAC model was a computational realization of the part of the model that dealt with visual word recognition; subsequently Jacobs and Grainger (1992) and Grainger and Jacobs (1996) developed the IAC model of visual word recognition further, as did Coltheart et al. (1993), whose DRC model envisaged an IAC-style model for handling visual word recognition plus a phonological lexicon and a nonlexical GPC route to handle reading aloud. The DRC model described by Rastle and Coltheart (1999a, 1999b) and in this article is the complete version of that model.

To evaluate DRC's worth as a model of visual word recognition and reading aloud, we selected a variety of basic effects observed in studies of lexical decision and reading aloud, and investigated the ability of the model to simulate these effects. The effects that were successfully simulated are as follows:

Reading Aloud

1. Frequency effect
2. Lexicality effect
3. Regularity effect
4. Interaction of regularity with frequency
5. Interaction of regularity with position of irregularity
6. Consistency effect
7. Pseudohomophony effect
8. Base word frequency effect on pseudohomophone reading
9. Absence of *N* effect on pseudohomophone reading
10. Presence of *N* effect on nonword reading
11. Whammy effect
12. Strategy effects
13. Homophone and pseudohomophone priming
14. Repetition priming
15. Onset effect in masked form priming

16. Triple interaction between regularity, frequency, and repetition
17. Length effect
18. Interaction between lexicality and letter length

Lexical Decision

1. Word frequency effect
2. Pseudohomophone effect
3. Interaction between pseudohomophone effect and orthographic similarity
4. *N* effect on NO responding
5. Interaction between *N* and frequency on YES responding

Acquired Dyslexias

1. Quantitative simulation of exception word, regular word, and nonword reading in surface dyslexia, and of the effect of frequency on surface dyslexic reading of exception words.
2. Quantitative simulation of the interaction between pseudohomophony and orthographic similarity in phonological dyslexia.

Other Effects

1. Stroop effect and its interaction with position of overlap
2. Prediction of regular-word reading accuracy from accuracy exception-word and nonword reading in normal, dyslexic, and brain-damaged children

We have already mentioned that no other computational model of reading can simulate both the reading-aloud task and the lexical decision task; the DRC model does well with both kinds of simulation. Moreover, even if we confine our attention just to the reading-aloud task, the set of phenomena that the DRC model can simulate is much larger than the set that any other current computational model of reading aloud can simulate; and, to the best of our knowledge, there is no effect seen in reading aloud that any of these other models can simulate but that the DRC model cannot. We suggest, therefore, that the DRC model is the most successful of current computational models of reading.

References

- 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). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 234–254.
- Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin and Review*, *4*, 439–461.
- Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic word reading. *Psychological Review*, *105*, 678–723.
- Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database* (CD-ROM). Linguistic Data Consortium, University of Pennsylvania.
- Basso, A., Taborelli, A., & Vignolo, L. A. (1978). Dissociated disorders of speaking and writing in aphasia. *Journal of Neurology, Neurosurgery and Psychiatry*, *41*, 556–563.
- Beauvois, M. F., & Derouesné, J. (1979). Phonological alexia: Three dissociations. *Journal of Neurology, Neurosurgery, and Psychiatry*, *42*, 1115–1124.
- Behrmann, M., & Bub, D. (1992). Surface dyslexia and dysgraphia: Dual routes, single lexicon. *Cognitive Neuropsychology*, *9*, 209–251.
- Behrmann, M., Plaut, D. C., & Nelson, J. (1998). A literature review and

- new data supporting an interactive account of letter-by-letter reading. *Cognitive Neuropsychology*, *15*, 7–51.
- 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.
- Berndt, R. S., Haendiges, A. N., Mitchum, C. C., & Wayland, S. C. (1996). An investigation of nonlexical reading impairments. *Cognitive Neuropsychology*, *13*, 763–801.
- Besner, D., & Davelaar, E. (1983). Suedohomophony effects in visual word recognition: Evidence for phonological processing. *Canadian Journal of Psychology*, *37*, 300–305.
- Brain, L. (1964). Statement of the problem. In A. V. S. De Reuck & M. O'Connor (Eds.), *Ciba Foundation Symposium on Disorders of Language* (pp. 5–20). London: Churchill.
- Bramwell, B. (1897). Illustrative cases of aphasia. *The Lancet*, *i*, 1256–1259. (Reprinted in *Cognitive Neuropsychology*, *1984*, *1*, 249–258)
- Broadbent, D. E. (1958). *Perception and communication*. Oxford, England: Oxford University Press.
- Broom, Y. M., & Doctor, E. A. (1995). Developmental phonological dyslexia: A case study of the efficacy of a remediation programme. *Cognitive Neuropsychology*, *12*, 725–766.
- Brybaert, M., & Praet, C. (1992). Reading isolated words: No evidence for automatic incorporation of the phonetic code. *Psychological Research*, *54*, 91–102.
- 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.
- Buchanan, L., Hildebrandt, N., & MacKinnon, G. E. (1994). Phonological processing of nonwords by a deep dyslexic patient: A rowze is implicitly a rose. *Journal of Neurolinguistics*, *8*, 163–181.
- Campbell, R., & Butterworth, B. (1985). Phonological dyslexia and dysgraphia in a highly literate subject: A developmental case with associated deficits of phonemic processing and awareness. *Quarterly Journal of Experimental Psychology*, *37A*, 435–475.
- Caramazza, A. (1984). The logic of neuropsychological research and the problem of patient classification in aphasia. *Brain and Language*, *21*, 9–20.
- Carpenter, G. A., Grossberg, S., Markuzon, N., Reynolds, J. H., & Rosen, D. B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks*, *3*, 698–713.
- Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition*, *47*, 149–180.
- Castles, A., & Coltheart, M. (1996). Cognitive correlates of developmental surface dyslexia: A single case study. *Cognitive Neuropsychology*, *13*, 25–50.
- Cestnick, L., & Coltheart, M. (1999). The relationship between language-processing and visual-processing deficits in developmental dyslexia. *Cognition*, *71*, 231–255.
- Cipolotti, L., & Warrington, E. K. (1995). Semantic memory and reading abilities: A case report. *Journal of the International Neuropsychological Society*, *1*, 104–110.
- Clarke, R., & Morton, J. (1983). Cross modality facilitation in tachistoscopic word recognition. *Quarterly Journal of Experimental Psychology*, *35A*, 79–96.
- 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. (1980a). Deep dyslexia: A right hemisphere hypothesis. In M. Coltheart, K. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 326–380). London: Routledge and Kegan Paul.
- Coltheart, M. (1980b). Reading phonological recoding, and deep dyslexia. In M. Coltheart, K. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 197–226). London: Routledge and Kegan Paul.
- Coltheart, M. (1981). Disorders of reading and their implications for models of normal reading. *Visible Language*, *15*, 245–286.
- Coltheart, M. (1982). The psycholinguistic analysis of acquired dyslexias. *Philosophical Transactions of the Royal Society of London*, *B298*, 151–164.
- Coltheart, M. (1985). Cognitive neuropsychology and the study of reading. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and performance XI* (pp. 3–37). Hillsdale, NJ: Erlbaum.
- Coltheart, M. (Ed.). (1996). *Phonological dyslexia*. Hove, England: Erlbaum.
- Coltheart, M. (Ed.). (1998). *Pure alexia (letter-by-letter reading)*. Hove, England: Psychology Press.
- Coltheart, M. (1999). Modularity and cognition. *Trends in Cognitive Sciences*, *3*, 115–120.
- Coltheart, M. (2000). Deep dyslexia is right-hemisphere reading. *Brain and Language*, *71*, 299–309.
- Coltheart, M., Byng, S., Masterson, J., Prior, M., & Riddoch, M. J. (1983). Surface dyslexia. *Quarterly Journal of Experimental Psychology*, *35A*, 469–495.
- Coltheart, M., & Coltheart, V. (2000). *Interaction of pseudohomophony and orthographic similarity in visual lexical decision*. Manuscript in preparation.
- 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., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale, NJ: Erlbaum.
- Coltheart, M., Dufty, D., & Bates, T. (2000). *The view from above: Learning to read, developmental dyslexia, and the dual route model of reading*. Manuscript in preparation.
- Coltheart, M., Langdon, R., & Haller, M. (1996). Computational cognitive neuropsychology. In B. Dodd, R. Campbell, & L. Worrall (Eds.), *Evaluating theories of language: Evidence from disordered communication* (pp. 9–36). London: Whurr Publishers.
- Coltheart, M., Patterson, K., & Marshall, J. C. (1980). *Deep dyslexia*. London: Routledge and Kegan Paul.
- 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., 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.
- Dalrymple-Alford, E. C. (1972). Sound similarity and color word interference in the Stroop task. *Psychonomic Science*, *28*, 209–210.
- Déjerine, J. (1892). Contribution à l'étude anatomo-pathologique et clinique des différentes variétés de cécité verbale. *Compte Rendu Hebdomadaire des Séances et Mémoires de la Société de Biologie*, *4*, 61–90.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321.
- Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*, 801–838.
- Derouesné, 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. W., & Young, A. (1988). *Human cognitive neuropsychology*. London: Erlbaum.
- Folk, J. R. (1999). Phonological codes are used to access the lexicon during

- silent reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 892–906.
- 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 on form-priming in the naming task: Interference effects from a masked prime. *Journal of Memory and Language*, 30, 1–25.
- Frost, R. (1995). Phonological computation and missing vowels: Mapping lexical involvement in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 398–408.
- Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin*, 123, 71–99.
- Fudge, E. C. (1984). *English word stress*. London: Allen and Unwin.
- Funnell, E. (1983). Phonological processes in reading: New evidence from acquired dyslexia. *British Journal of Psychology*, 74, 159–180.
- Garde, P. (1968). *L'Accent*. Paris: Presses Universitaires de France.
- Geschwind, N. (1965a). Disconnection syndromes in animals and man: I. *Brain*, 88, 237–294.
- Geschwind, N. (1965b). Disconnection syndromes in animals and man: II. *Brain*, 88, 585–644.
- 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.
- Goulandris, N. K., & Snowling, M. (1991). Visual memory deficits: A plausible cause of developmental dyslexia? Evidence from a single case study. *Cognitive Neuropsychology*, 8, 127–154.
- Grainger, J., & Jacobs, A. M. (1993). Masked partial-word priming in visual word recognition: Effects of positional letter frequency. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 951–964.
- Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, 103, 518–565.
- Grainger, J., & Jacobs, A. M. (1998). On localist connectionism and psychological science. In J. Grainger & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition*. (pp. 1–38). Mahwah, NJ: Erlbaum.
- Grossberg, S. (1978). A theory of visual coding, memory, and development. In E. L. J. Leeuwenberg & H. F. J. M. Buffart (Eds.), *Formal theories of visual perception* (pp. 7–26). New York: Wiley.
- Hanley, J. R., Hastie, K., & Kay, J. (1992). Developmental surface dyslexia and dysgraphia: An orthographic processing impairment. *Quarterly Journal of Experimental Psychology*, 44A, 285–319.
- Harris, M., & Coltheart, M. (1986). *Language processing in children and adults: An introduction*. London: Routledge and Kegan Paul.
- Head, H. (1926). *Aphasia and kindred disorders of speech*. Cambridge, England: Cambridge University Press.
- Holmes, J. M. (1973). *Dyslexia: A neurolinguistic study of traumatic and developmental disorders of reading*. Unpublished doctoral dissertation, University of Edinburgh.
- Holmes, J. M. (1978). "Regression" and reading breakdown. In A. Caramazza & E. B. Zurif (Eds.), *Language acquisition and language breakdown* (pp. 87–98). Baltimore: Johns Hopkins University Press.
- Howard, D., & Best, W. (1996). Developmental phonological dyslexia: Real word reading can be completely normal. *Cognitive Neuropsychology*, 13, 887–934.
- Howard, D., & Franklin, S. (1988). *Missing the meaning? A cognitive neuropsychological study of the processing of words by an aphasic patient*. Cambridge, MA: MIT Press.
- Jacobs, A. M., & Grainger, J. (1991). Automatic letter priming in an alphabetic decision task. *Perception and Psychophysics*, 49, 43–52.
- Jacobs, A. M., & Grainger, J. (1992). Testing a semistochastic variant of the interactive activation model in different word recognition experiments. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 1174–1188.
- 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.
- Job, R., Peressotti, F., & Cusinato, A. (1998). Lexical effects in naming pseudowords in shallow orthographies: Further empirical data. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 622–630.
- Job, R., Sartori, G., Masterson, J., & Coltheart, M. (1984). Developmental surface dyslexia in Italian. In R. N. Malatesha & H. A. Whitaker (Eds.), *Dyslexia: A global issue* (pp. 133–142). The Hague, Netherlands: Martinus Nijhoff.
- Kawamoto, A. H., Kello, C. T., Jones, R., & Bame, K. (1998). Initial phoneme versus whole-word criterion to initiate pronunciation: Evidence based on response latency and initial phoneme duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 862–885.
- Kay, J., Lesser, R., & Coltheart, M. (1992). *PALPA: Psycholinguistic assessments of language processing in aphasia*. Hove, England: Erlbaum.
- Kay, J., & Marcel, A. (1981). One process, not two, in reading aloud: Lexical analogies do the work of non-lexical rules. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 33A, 397–413.
- Kwantes, P., & Marmurek, H. H. C. (1995, November). *Orthographic and lexical frequency effects in naming pseudohomophones*. Paper presented at the annual meeting of the Psychonomic Society, Los Angeles.
- Lambon Ralph, M. A., Ellis, A. W., & Franklin, S. (1995). Semantic loss without surface dyslexia. *Neurocase*, 1, 363–370.
- Laxon, V. J., Masterson, J., Pool, M., & Keating, C. (1992). Nonword naming: Further exploration of the pseudohomophone effect in terms of orthographic neighborhood size, graphemic changes, spelling-sound consistency, and reader accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 730–748.
- Lesch, M. F., & Pollatsek, A. (1998). Evidence for the use of assembled phonology in accessing the meaning of printed words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 573–592.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–75.
- Lhermitte, F., & Derouesné, J. (1974). Paraphasies et jargonaphasies dans le langage oral avec conservation du langage écrit. *Revue Neurologique*, 130, 21–38.
- Lichtheim, L. (1885). On aphasia. *Brain*, 7, 433–484.
- Lissauer, H. (1890). Ein Fall von Seelenblindheit nebst einem Beitrag zur Theorie derselben. *Archiv für Psychiatrie und Nervenkrankheiten*, 21, 222–270. [English translation published in *Cognitive Neuropsychology*, 5, 157–192]
- Lukatela, G., Frost, S. J., & Turvey, M. T. (1998). Phonological priming by masked nonword primes in the lexical decision task. *Journal of Memory and Language*, 39, 666–683.
- Lukatela, G., Frost, S. J., & Turvey, M. T. (1999). Identity priming in English is compromised by phonological ambiguity. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 775–790.
- Lukatela, G., & Turvey, M. T. (1994). Visual lexical access is initially phonological: 2. Evidence from phonological priming by homophones and pseudohomophones. *Journal of Experimental Psychology: General*, 123, 331–353.
- Lukatela, G., & Turvey, M. T. (1998). Reading in two alphabets. *American Psychologist*, 53, 1057–1072.
- Lukatela, G., & Turvey, M. T. (2000). An evaluation of the two-cycles

- model of phonology assembly. *Journal of Memory and Language*, 42, 183–207.
- Luo, C. R. (1996). How is word meaning accessed in reading? Evidence from the phonologically mediated interference effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 883–895.
- Manis, F. R., Seidenberg, M. S., Doi, L. M., McBride-Chang, C., & Petersen, A. (1996). On the bases of two subtypes of development dyslexia. *Cognition*, 58, 157–195.
- Marie, P. (1906). Révision de la question de l'aphasie: Le troisième convolution frontale gauche ne joue aucun rôle spéciale dans la fonction du langage. In M. F. Cole & M. Cole (Eds.) (1971). *Pierre Marie's Papers on Speech Disorders*. New York: Hafner.
- Marmurek, H. H. C., & Kwantes, P. J. (1996). Reading words and words: Phonology and lexical access. *Quarterly Journal of Experimental Psychology*, 49A, 696–714.
- Marshall, J. C., & Newcombe, F. (1966). Syntactic and semantic errors in paralexia. *Neuropsychologia*, 4, 169–176.
- Marshall, J. C., & Newcombe, F. (1973). Patterns of paralexia: A psycholinguistic approach. *Journal of Psycholinguistic Research*, 2, 175–199.
- Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10, 29–63.
- Massaro, D. W. (1989). Testing between the TRACE model and the fuzzy logical model of speech perception. *Cognitive Psychology*, 21, 398–421.
- 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.
- McCarthy, R., & Warrington, E. K. (1986). Phonological reading: Phenomena and paradoxes. *Cortex*, 22, 359–380.
- McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, 86, 287–330.
- McClelland, J. L. (1991). Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology*, 23, 1–44.
- McClelland, J. L. (1993). Toward a theory of information processing in graded, random, and interactive networks. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance 14: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 655–688). Cambridge, MA: MIT Press.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, 88, 375–407.
- McQuade, D. V. (1981). Variable reliance on phonological information in visual word recognition. *Language and Speech*, 24, 99–109.
- Michel, F., Henaff, M. A., & Intrilligator, J. (1996). Two different readers in the same brain after a posterior callosal lesion. *NeuroReport*, 7, 786–788.
- Miller, G. A. (1985). WordNet: A lexical database for English. *Communications of the ACM*, 38, 39–41.
- Morton, J. (1961). *Reading, context and the perception of words*. Unpublished PhD thesis, University of Reading, Reading, England.
- Morton, J. (1968). Considerations of grammar and computation in language behavior. In J. C. Catford (Ed.), *Studies in language and language behavior* (pp. 499–545). C.R.L.L.B. Progress Report No. VI, University of Michigan.
- Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, 76, 165–178.
- Morton, J. (1970). A functional model for memory. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press.
- Morton, J. (1979). Facilitation in word recognition: Experiments causing change in the logogen model. In P. A. Kollers, M. E. Wroldstad, & H. Bouma (Eds.), *Processing of visual language: I*. (pp. 259–268). New York: Plenum Press.
- Morton, J. (1980). The logogen model and orthographic structure. In U. Frith (Ed.), *Cognitive processes in spelling*. London: Academic Press.
- Morton, J. (1982). Disintegrating the lexicon: An information processing approach. In J. Mehler, E. C. T. Walker, & M. Garrett (Eds.), *Perspectives on mental representation* (pp. 89–109). Hillsdale, NJ: Erlbaum.
- Morton, J., & Patterson, K. E. (1980). A new attempt at interpretation, or, an attempt at a new interpretation. In M. Coltheart, K. E. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 91–118). London: Routledge and Kegan Paul.
- Norris, D. (1994). A quantitative multiple-levels model of reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 1212–1232.
- Norris, D., & Brown, G. (1985). Race models and analogy theories: A dead heart? Reply to Seidenberg. *Cognition*, 20, 155–168.
- Ognjenovic, V., Lukatela, G., Feldman, L. B., & Turvey, M. T. (1983). Misreadings by beginning readers of Serbo-Croatian. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 35A, 97–109.
- Paap, K. R., Chen, L., & Noel, R. W. (1987, November). *Word recognition: Is the sky falling on top-down processing?* Paper presented at the annual meeting of the Psychonomic Society, Seattle, WA.
- Paap, K. R., & Noel, R. W. (1991). Dual route models of print to sound: Still a good horse race. *Psychological Research*, 53, 13–24.
- Page, M. (in press). Connectionist modelling in psychology: A localist manifesto. *Behavioral and Brain Sciences*.
- Patterson, K., & Marcel, A. (1992). PHONOLOGICAL alexia or phonological ALEXIA? In J. Alegria & D. Holender (Eds.), *Analytic approaches to human cognition* (pp. 259–274). Amsterdam, Netherlands: North-Holland.
- Patterson, K. E., Marshall, J. C., & Coltheart, M. (1985). *Surface dyslexia: Cognitive and neuropsychological studies of phonological reading*. Hove, England: Erlbaum.
- Patterson, K. E., & Morton, J. (1985). From orthography to phonology: An attempt at an old interpretation. In K. E. Patterson, J. C. Marshall, & M. Coltheart (Eds.), *Surface dyslexia: Cognitive and neuropsychological studies of phonological reading* (pp. 335–360). Hove, England: Erlbaum.
- Patterson, K., & Shewell, C. (1987). Speak and spell: Dissociations and word-class effects. In M. Coltheart, G. Sartori, & R. Job (Eds.), *The cognitive neuropsychology of language* (pp. 273–294). London: Erlbaum.
- Patterson, K., Vargha-Khadem, F., & Polkey, C. (1989). Reading with one hemisphere. *Brain*, 112, 39–63.
- Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S. F., Cotelli, M., Cossu, G., Corte, F., Lorusso, M., Pesenti, S., Galagher, A., Perani, D., Price, C., Frith, C. D., & Frith, U. (2000). A cultural effect on brain function. *Nature Neuroscience*, 3, 91–96.
- Peereman, R., & Content, A. (1995). Neighborhood size effect in naming: Lexical activation or sublexical correspondence? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 409–421.
- Perfetti, C. A., & Bell, L. C. (1991). Phonemic activation during the first 40 ms of word identification: Evidence from backward masking and priming. *Journal of Memory and Language*, 30, 473–485.
- Perfetti, C. A., Bell, L. C., & Delancy, S. M. (1988). Automatic (prelexical) phonetic activation in silent word reading: Evidence from backward masking. *Journal of Memory and Language*, 27, 59–70.
- Perry, C. (1997). *A computational and experimental investigation of visual word recognition and naming*. Unpublished doctoral thesis, Swinburne University of Technology.
- Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28, 73–193.
- Plaut, D. C. (1997). Structure and function in the lexical system: Insights

- from distributed models of word reading and lexical decision. *Language and Cognitive Processes*, 12, 765–805.
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. E. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103, 56–115.
- Plaut, D. C., & Shallice, T. (1993). Deep dyslexia. A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, 10, 377–500.
- Poeppel, D. (1996a). A critical review of PET studies of phonological processing. *Brain and Language*, 55, 317–351.
- Poeppel, D. (1996b). Some remaining questions about studying phonological processing with PET: Response to Demonet, Fiez, Paulesu, Petersen, and Zatorre (1996). *Brain and Language*, 55, 380–385.
- Price, C. J., Howard, D., Patterson, K., Warburton, E. A., Friston, K., & Frackowiak, R. S. J. (1998). A functional neuroimaging description of two deep dyslexic patients. *Journal of Cognitive Neuroscience*, 10, 303–315.
- Rastle, K., & Coltheart, M. (1998). Whammy and double whammy: The effect of length on nonword reading. *Psychonomic Bulletin and Review*, 5, 277–282.
- Rastle, K., & Coltheart, M. (1999a). Lexical and nonlexical phonological priming. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 461–481.
- Rastle, K., & Coltheart, M. (1999b). Serial and strategic effects in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 482–503.
- Rastle, K., & Coltheart, M. (2000a, in press). Lexical and nonlexical print-to-sound translation of disyllabic words and nonwords. *Journal of Memory and Language*, 42, 342–364.
- Rastle, K., & Coltheart, M. (2000b). Serial processing in reading aloud: A reply to Zorzi. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 1232–1235.
- Rastle, K., Harrington, J., Coltheart, M., & Palethorpe, S. (2000). Reading aloud begins when the computation of phonology finishes. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 1232–1235.
- Rastle, K., Harrington, J., Coltheart, M., & Thomas, B. (2000). *The ARC Nonword Database*. Manuscript submitted for publication.
- Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 81, 275–280.
- Rosson, M. B. (1983). From SOFA to LOUCH: Lexical contributions to pseudoword pronunciation. *Memory and Cognition*, 11, 152–160.
- Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. (1971). Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior*, 10, 645–657.
- Rumelhart, D. E., Durbin, R., Golden, R., & Chauvin, Y. (1995). Back-propagation: The basic theory. In Y. Chauvin & D. E. Rumelhart (Eds.), *Backpropagation: Theory, architectures, and applications. Developments in connectionist theory* (pp. 1–34). Hillsdale, NJ: Erlbaum.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533–536.
- 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.
- Rumelhart, D. E., & Siple, P. (1974). Process of recognizing tachistoscopically presented words. *Psychological Review*, 81, 99–118.
- Saffran, E. M. (1982). Neuropsychological approaches to the study of language. *British Journal of Psychology*, 73, 317–337.
- Saffran, E., Bogyo, L. C., Schwartz, M. F., & Marin, O. S. M. (1980). Does deep dyslexia reflect right-hemisphere reading? In M. Coltheart, K. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 381–406). London: Routledge and Kegan Paul.
- Schwartz, M. F., Saffran, M., & Marin, O. S. M. (1979). Dissociations of language function in dementia: A case study. *Brain and Language*, 7, 277–306.
- Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 876–900.
- 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, 48–62.
- Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 546–554.
- 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.
- Seymour, P. H. (1973). A model for reading, naming and comparison. *British Journal of Psychology*, 64, 35–49.
- Shallice, T. (1981). Phonological agraphia and the lexical route in writing. *Brain*, 104, 413–429.
- Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge, England: Cambridge University Press.
- Shallice, T., & Warrington, E. K. (1970). Independent functioning of the verbal memory stores: A neuropsychological study. *Quarterly Journal of Experimental Psychology*, 22, 261–273.
- Snodgrass, J. G., & Mintzer, M. (1993). Neighborhood effects in visual word recognition: Facilitatory or inhibitory? *Memory and Cognition*, 21, 247–266.
- Snowling, M., & Hulme, C. (1989). A longitudinal case study of developmental phonological dyslexia. *Cognitive Neuropsychology*, 6, 379–401.
- Snowling, M., Stackhouse, J., & Rack, J. (1986). Phonological dyslexia and dysgraphia: A developmental analysis. *Cognitive Neuropsychology*, 3, 309–339.
- Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science*, 8, 411–416.
- 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. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1140–1154.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662.
- Taft, M. (1991). *Reading and the mental lexicon*. Hove, England: Erlbaum.
- Taft, M., & Russell, B. (1992). Pseudohomophone naming and the word frequency effect. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 45A, 51–71.
- Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word recognition. *Journal of Memory and Language*, 26, 608–631.
- Temple, C. M. (1984). Surface dyslexia in a child with epilepsy. *Neuropsychologia*, 22, 569–576.
- Temple, C. M. (1997). *Developmental cognitive neuropsychology*. Hove, England: Psychology Press.
- Temple, C. M., & Marshall, J. C. (1983). A case study of developmental phonological dyslexia. *British Journal of Psychology*, 74, 517–533.
- Treisman, A. M. (1961). *Attention and speech*. Unpublished doctoral thesis, Oxford University.

- 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., & Paap, K. R. (1997). Functional neuroimages fail to discover pieces of mind in the parts of the brain. *Philosophy of Science*, *64* (4 Suppl S), S85–S94.
- Verstaen, A., Humphreys, G. W., Olson, A., & d'Ydewalle, G. (1995). Are phonemic effects in backward masking evidence for automatic prelexical phonemic activation in visual word recognition? *Journal of Memory and Language*, *34*, 335–356.
- Visser, T., & Besner, D. (in press). On the dominance of the addressed routine in reading aloud. *Psychonomic Bulletin and Review*.
- Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, *20*, 158–177.
- Watters, P. A., & Patel, M. (1998). Modelling lexical-semantic processes using WordNet. *Glott International*, *3*, 23–24.
- Weekes, B. (1997). Differential effects of number of letters on word and nonword naming latency. *The Quarterly Journal of Experimental Psychology*, *50A*, 439–456.
- Weekes, B., & Coltheart, M. (1996). Surface dyslexia and surface dysgraphia: Treatment studies and their theoretical implications. *Cognitive Neuropsychology*, *13*, 277–315.
- Weekes, B., Coltheart, M., & Gordon, E. (1997). Deep dyslexia and right-hemisphere reading: A regional cerebral blood flow study. *Aphasiology*, *11*, 1139–1158.
- Wernicke, C. (1874). *Der aphasische symptomkomplex* Breslau, Germany: Cohn and Weigart. (Translated in *Boston Studies in Philosophy of Science*, *4*, 34–97)
- Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology*, *1*, 59–85.
- Winnick, W. A., & Daniel, S. A. (1970). Two kinds of response priming in tachistoscopic recognition. *Journal of Experimental Psychology*, *84*, 74–81.
- Ziegler, J. C., Montant, M., & Jacobs, A. M. (1997). The feedback consistency effect in lexical decision and naming. *Journal of Memory and Language*, *37*, 533–554.
- Ziegler, J., 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., Rey, A., & Jacobs, A. M. (1998). Simulating individual word identification thresholds and errors in the fragmentation task. *Memory and Cognition*, *26*, 490–501.
- 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., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1131–1161.

Received February 4, 1999

Revision received June 27, 2000

Accepted July 5, 2000 ■