Journal of Experimental Psychology:
Human Perception and Performance
1998, Vol. 24, No. 4, 1131-1161

0096-1523/98/$3.00

Two Routes or One in Reading Aloud? A Connectionist Dual-Process Model

Marco Zorzi
University of Trieste and University College London

George Houghton and Brian Butterworth
University College London

A connectionist study of word reading is described that emphasizes the computational demands of the spelling–sound mapping in determining the properties of the reading system. It is shown that the phonological assembly process can be implemented by a two-layer network, which easily extracts the regularities in the spelling–sound mapping for English from training data containing many exception words. It is argued that productive knowledge about spelling–sound relationships is more easily acquired and used if it is separated from case-specific knowledge of the pronunciation of known words. It is then shown how the interaction of assembled and retrieved phonologies can account for the combined effects of frequency and regularity–consistency and for the reading performance of dyslexic patients. It is concluded that the organization of the reading system reflects the demands of the task and that the pronunciations of nonwords and exception words are computed by different processes.

Experimental psychologists have long been interested in problems of word recognition and naming, which has given rise to an extensive body of empirical data. The theoretical side of reading research has recently been given substantial impetus by the development of computational models of single-word reading. Such models can provide new and more detailed accounts of existing data and may lead to new predictions. The seminal work in this respect is the Seidenberg and McClelland (1989) parallel distributed processing (PDP) model, which is perhaps most notable, in terms of traditional theoretical discussions of reading, in mounting a challenge to "dual-route theories" of word naming. Such accounts propose that, in alphabetic scripts, written words may be pronounced either by a spelling–sound conversion route (which can also be used on nonwords), or by a "lexical route" in which the whole word is recognized and its pronunciation looked up in a phonological lexicon. The Seidenberg and McClelland (henceforth S&M) model was claimed to be a single-route model with no lexical entries. This challenge to the dual-route theory of reading has been quite influential on the recent literature, and the model's poor performance in nonword reading (Besner, Twilley, McCann, & Seergobin, 1990) does not appear to have had a strong impact on its appeal. Recently, Coltheart, Curtis, Atkins, and Haller (1993) met the S&M challenge by proposing a computational implementation of the dual-route model. In addition, Plaut and colleagues (Plaut & McClelland, 1993; Plaut, McClelland, Seidenberg, & Patterson, 1996) developed a single-route successor to the S&M model that performs well on nonword reading, giving added plausibility to the single-route theory.

Marco Zorzi, Department of Psychology, University of Trieste, Italy, and Department of Psychology, University College London, London, England; George Houghton and Brian Butterworth, Department of Psychology, University College London, London, England.

Marco Zorzi was supported in part by the ERASMUS Exchange Program of the European Community and by a grant from the Ministry of University and Scientific Research, Italy. Brian Butterworth's reading research was supported by Grant G9015838N from the Medical Research Council of the United Kingdom.

Parts of this research were reported at the 12th European Meeting for Cognitive Neuropsychology, Bressanone, Italy, January 1994, and at the September 1995 Congress of the European Society for Cognitive Psychology, Rome, Italy.

We thank Max Coltheart, Remo Job, Karalyn Patterson, David Plaut, Sandy Pollatsek, and Taeko Wydell for helpful discussions and comments on earlier versions of the article.

Correspondence concerning this article should be addressed to Marco Zorzi, Dipartimento di Psicologia, Università di Trieste, Via dell'Università, 7, I-34123 Trieste, Italy. Electronic mail may be sent to zorzi@univ.trieste.it.

In this article, we present a connectionist study that we believe provides new insights into normal and impaired reading. The emphasis in the current work is on the role of the task demands of the spelling–sound mapping in determining the properties of the mechanism implementing that mapping. We show that the regularities in the spelling–sound mapping for English can be quite easily acquired (by using a simple, general, learning rule) from training data containing many idiosyncratic "exception" words, and that the resultant knowledge generalizes well to nonwords. We argue that this knowledge is more easily acquired and used if it is kept separate from lexical knowledge regarding the pronunciation of specific known words. We then show how the interaction of these two sources of knowledge—that is, generative and case-specific—can account for the combined effects of frequency and regularity–consistency and for the reading performance of dyslexic patients.

Empirical Background

English letters and letter patterns largely correspond to speech units (e.g., phonemes) but not in a completely regular way. In many cases, these correspondences are irregular (e.g., the vowels in PINT, HAVE) or arbitrary (e.g., OLO in

COLONEL). Skilled readers can read these words aloud perfectly well, and at the same time will read similar-looking nonwords (e.g., KINT, MAVE, SOLONEL) in a way that reflects, not the individual exception words, but the phonological values the letters standardly represent. Such observations have led to "dual-route" models of reading, which postulate two (nonsemantic) routes from spelling to sound (see, e.g., Baron & Strawson, 1976; Besner & Smith, 1992; Coltheart, 1978, 1985; Morton & Patterson, 1980; Paap & Noel, 1991; Patterson & Morton, 1985). One, usually referred to as the lexical route, is thought to operate by retrieving word pronunciations from an internal lexicon. It is therefore based on a word-specific association mechanism and contains any learned word. The phonological form is thus "addressed" from the visual word form as a whole; therefore, it has been considered a lookup procedure, and its product has been called *addressed* (or *retrieved*) *phonology.* The second route, named the sublexical (or GPC, viz., grapheme-to-phoneme correspondence) route, is described as a letter-to-sound mapping, resulting in the *assembled phonology,* which is based on explicit rules specifying the dominant (e.g., most frequent) relationships between letters and sounds (see Carr & Pollatsek, 1985, and Patterson & Coltheart, 1987, for reviews). The lexical route can handle all known words including exceptional spelling–sound relationships, whereas assembly can be used on nonwords. The addressed route provides the unusual /&/[1] in HAVE, whereas the assembled route produces the /eI/ in MAVE on the basis of words such as GAVE, WAVE, SAVE, SLAVE, and so forth.

If both processes act in parallel on any input string, then exception words will (by definition) lead to disagreements between the two readings. Thus, some models assume a "horse race" between the two outputs (e.g., Paap & Noel, 1991), where the response is determined by the first process to finish. Other models propose that the output of the two processes is pooled continuously until a phonological representation sufficient to drive articulation is generated (e.g., Monsell, Patterson, Graham, Hughes, & Milroy, 1992).

Among dual-route theorists there is a general consensus on the modules involved in word recognition and naming but not on the internal operations of those modules (Carr & Pollatsek, 1985), particularly in regard to the assembly mechanism. Finally, it is widely agreed (but see Van Orden, Pennington, & Stone, 1990) that there is a third important way to retrieve the pronunciation of a known word: the semantic route. The printed form accesses the semantic representation of that word, which in turn activates the corresponding phonology. This procedure has usually been referred to as *reading via meaning.* Most theorists assume that the semantic route contributes little to word naming in skilled readers, because, being indirect, it is slow in delivering a word pronunciation (but see Plaut et al., 1996; Patterson, Graham, & Hodges, 1994); however, Strain, Patterson, and Seidenberg (1995) recently demonstrated some impact of a semantic variable (imageability) in reading low-frequency exception words (i.e., the words that usually yield the longest naming latencies).

The dual-route view was challenged by Glushko (1979), who found that the pronunciation of both words and nonwords is influenced by knowledge of other words with similar spelling patterns. Letter strings that share the word body (vowel and final consonants) with exception words yielded longer naming latencies. This effect was termed the *consistency effect,* because the performance is determined by the consistency of the spelling–sound relationships of the neighborhood of similarly spelled words. Furthermore, Glushko's research participants gave an irregular pronunciation to a significant number of the inconsistent nonwords, that is, nonwords containing substrings with more than one common pronunciation (e.g., OVE may be pronounced as in STOVE, WOVE, GROVE, etc., or as in LOVE, GLOVE, SHOVE). These results have been taken as strong evidence against the dual-route model: If participants use GPC rules, all nonwords should have been given regular pronunciations. The proposal was made instead that nonword pronunciations are lexically based, being somehow synthesized from a "cohort" of similar-looking words activated by the input. The major deficiency of this approach is the absence of any specification of the process by which a coherent pronunciation is generated from the potentially huge number of activated lexical entries. One important conclusion drawn by Glushko (1979) was that the regular–exception classification is, by itself, psychologically useless: The important distinction is whether a word is consistent or inconsistent with the orthographic and phonological structure that it activates.

One main motivation for the separation of lexical and sublexical knowledge, however, comes from neuropsychological studies, which have provided the clearest evidence that the two reading procedures can be selectively impaired by brain damage (see Coltheart, 1985; Denes, Cipolotti, & Zorzi, 1996; Shallice, 1988, for reviews). Reading models have been strongly influenced by data from acquired and developmental reading disorders, and a major claim of dual-route theorists is that the error patterns of dyslexic patients perfectly fit the dual-route architecture (e.g., Besner, in press; Coltheart, 1985; Coltheart et al., 1993; Coltheart, Langdon, & Haller, 1996). Two forms of the dyslexic syndrome are particularly relevant in the present context: surface dyslexia and phonological dyslexia. The main difficulties of phonological dyslexic patients are in nonword reading and are postulated to be due to an impairment of the sublexical route. The purest example of this syndrome is patient W.B., described by Funnell (1983). W.B. showed almost perfect performance in word reading while being unable to read any nonword aloud or to give the sounds of single letters. Surface dyslexic patients read nonwords and regular words well but are impaired on exception words, which they frequently regularize, that is, pronounce according to standard spelling–sound correspondences. Surface dyslexia is assumed to be due to damage to the lexical route.

[1] Pronunciation keys: /i/ in *bean,* /A/ in *bear,* /O/ in *born,* /u/ in *boon,* /3/ in *burn,* /I/ in *pit,* /e/ in *pet,* /&/ in *pat,* /V/ in *bud,* /0/ in *pot,* /U/ in *good,* /eI/ in *bay,* /aI/ in *buy,* /oI/ in *boy,* /@U/ in *no,* /aU/ in *now,* /e@/ in *where,* /I@/ in *here,* /tS/ in *chain,* /dZ/ in *Jane,* /9/ in *think,* /T/ in *thin,* /S/ in *ship.*

The purest examples of this syndrome are patients M.P. (Bub, Cancelliere, & Kertesz, 1985) and K.T. (McCarthy & Warrington, 1986). The interpretation is that the residual reading abilities are performed by means of the assembly procedure, whereas the lexical, lookup procedure is damaged. However, the patient's reading performance typically shows a frequency-by-regularity interaction; that is, the performance in reading exception words gets worse as the word frequency decreases. Such findings have been considered to be evidence that the lexical route can still deliver the correct pronunciation for some exception items.

K.T. is an extreme case in the sense that his reading performance on exception words was severely impaired. He misread even many high-frequency exception words (~ 63%), and most of the errors were regularizations (71%). On the other hand, his regular word and nonword reading was perfect: He scored 100% on Glushko's (1979) nonwords (McCarthy & Warrington, 1986). M.P. is a somewhat different case: Her exception word reading was much less impaired, especially for the high-frequency items. However, this patient is very frequency sensitive, and her reading performance shows a broad variability across different lists, with an error rate ranging from 18% to 59% (cf. Behrmann & Bub, 1992; Bub et al., 1985; Plaut, Behrmann, Patterson, & McClelland, 1993). This variability is attributable to different frequency ranges used in classifying high- and low-frequency words and to different "levels of regularity" of the items in the various lists, that is, how much an item is irregular with respect to its body neighborhood. However, her regular word and nonword reading is very "consistent" and similar to that of K.T.

We should stress that pure cases of surface dyslexia have been found even in the developmental counterpart of the syndrome. In a large study on developmental dyslexia, Castles and Coltheart (1993) identified 10 dyslexic children showing the well-known difficulties with exception words and normal nonword reading. The existence of pure forms of developmental dyslexia has been recently confirmed with a second group study conducted by Manis, Seidenberg, Doi, McBride-Chang, and Petersen (1996).

Reading is not an innate skill, and an interesting source of constraint on models comes from studies of children learning to read. For instance, such data strongly support the idea that the phonological code involves division of the syllable into onset and rime. The onset consists of the consonant cluster preceding the vowel (if any), and rime comprises the vowel and any final consonants (e.g., STRING = STR + ING, MILK = M + ILK). Bradley and Bryant (1983) found a strong connection between children's sensitivity to rhyme and their success in reading: Rhyming skills are an excellent predictor of later reading skill, even 2 years or more before children learn to read. By contrast, there is good evidence that phonemic awareness arises as a consequence of learning to read. Children are not aware of phonemes before they begin learning to read (first or second grade): They have a great deal of difficulty with phonological tasks such as phoneme deletion (e.g., delete /p/ from /plant/) and phoneme tapping (Bruce, 1964; Lieberman, Shankweiler, Fischer, & Carter, 1974; see Goswami & Bryant, 1990, for a review). It is striking that illiterate adults show a similar insensitivity to phonemes (Morais, Cary, Alegria, & Bertelson, 1979), whereas they can manipulate syllables and appreciate rhyme (Morais, Bertelson, Cary, & Alegria, 1986).

We can conclude from this that children may be unable to use GPCs until they begin to be able to read new words. Experimental data support this idea: Children have been found to rely heavily on the larger phonological units of onset and rime when they have to make connections between sounds and letters. In fact, by using larger print-to-sound correspondences, the onset and rime spelling patterns, they are able to draw analogies from words they already know to new words (Goswami, 1986, 1988, 1991; see Goswami & Bryant, 1990, for a review). That the onset–rime representation is an intrinsic feature of the phonological code, applicable both to known words and to the encoding of new words, is independently supported by psycholinguistic studies (Treiman, 1986; Treiman & Danis, 1988; see Treiman, 1989, for a review) and models of speech production (Dell, 1986; Hartley & Houghton, 1996). Most notably, Treiman and collaborators (Treiman & Chafetz, 1987; Treiman, Mullenix, Bijeljac-Babic, & Richmond-Welty, 1995) specifically argued that the orthographic structure of written English reflects the phonological structure of spoken English, that is, that written words have corresponding onset–rime orthographic units. Treiman et al. (1995) found clear evidence that these orthographic units play an important role in both adult's and children's pronunciation of printed words.

A variety of attempts have been made to explain results such as those summarized above. In response to the Glushko (1979) findings, later versions of dual-route theories have assumed that pronunciation rules can involve orthographic units of different sizes. For instance, Patterson and Morton (1985) proposed two distinct levels of orthographic structure: graphemes and word bodies. With regard to single-route models, Shallice and colleagues (Shallice & McCarthy, 1985; Shallice, Warrington, & McCarthy, 1983) proposed a single mechanism operating at different levels of correspondence, from graphemes to whole words. Finally, the latest lexical analogy models (e.g., Humphreys & Evett, 1985) incorporate the assumption that the translation procedure for a nonword is different from that for a known word. At this point, these different verbal theories become difficult to distinguish (Patterson & Coltheart, 1987), partly because of their lack of detail or computational implementation. It is for this reason that the computational modeling of reading performance initiated by Seidenberg and McClelland (1989) has become so important. In the next section, we briefly summarize some recent computational work.

Computational Models

The Seidenberg and McClelland (1989) PDP model of reading (the S&M model) represents the first attempt to overcome the indeterminacy of verbal models by implementing systems capable of performing the spelling–sound mapping.[2] The S&M model is a three-layer, feedforward neural network that attempts to read regular words, exception words, and nonwords in a single route from spelling to sound. The mapping is mediated by a set of "hidden units," and the network is trained using the back-propagation algorithm (Rumelhart, Hinton, & Williams, 1986). The S&M framework can be thought of as an integration of lexical analogy and multiple-level theories (Plaut et al., 1996), because it denies the existence of separate lexical and sublexical procedures.

This model has been criticized by several authors (see, e.g., Besner et al., 1990; Coltheart et al., 1993). For instance, it showed poor nonword reading, and attempts to simulate the surface dyslexic syndrome through damage to the network were unsuccessful. The model has since been superseded by that of Plaut and colleagues (Plaut & McClelland, 1993; Plaut, McClelland, Seidenberg, & Patterson, 1996; hereinafter the PMSP model). This later model is still a single-route model, but it can read monosyllabic nonwords at a level of performance similar to that of humans (also see Seidenberg, Plaut, Petersen, McClelland, & McRae, 1994).

The PMSP model differs from the S&M model both architecturally and representationally. For instance, the PMSP model incorporates recurrent connections, which give rise to attractors (this, however, does not seem to be a key feature for good nonword reading; see Plaut et al., 1996). The S&M model used highly "distributed" representations at input and output, in which no single node stood for anything recognizable as an individual letter or phoneme. The PMSP model, by contrast, segments input and output into onset, vowel, and coda, and the orthographic units consist of graphemes rather than letters; that is, single nodes stand for whole groups of letters (e.g., WH, CH, AY, EA, CK, TCH), which correspond to single phonemes. Compared to the case in the S&M model, these representations are localist; indeed, there are separate nodes for related graphemes (e.g., CH, TCH).[3] The improved nonword (i.e., generalization) performance of the model over that of the S&M model appears to be largely due to this change. With respect to surface dyslexia, the performance of the PMSP attractor network model does not significantly improve on that of the S&M model. For instance, when the model is damaged to the extent needed to mirror the performance of the surface dyslexic K.T. on exception words, its regular word and nonword reading are also substantially affected (see Plaut et al., 1996, Figure 20). Surface dyslexia has been successfully simulated by the PMSP model only within a "two-route" framework, which postulates the active interaction between the (implemented) phonological pathway and the (unimplemented) semantic pathway (see below). Clearly, dyslexic dissociations offer substantial challenges to any single-route account of reading.

In response to these single-route models, Coltheart and colleagues (Coltheart et al., 1993; Coltheart & Rastle, 1994) developed implementations of aspects of dual-route theories. Coltheart et al. (1993) implemented a nonlexical route using explicit rules discovered and stored by a specially constructed "learning algorithm" in a single pass through the training database. The lexical pathway was implemented by Coltheart and Rastle (1994) as an interactive activation model, based on McClelland and Rumelhart's (1981; see also Rumelhart & McClelland, 1982) word recognition model joined to Dell's (1986) spoken word production model. The complete dual-route model is known as the dual-route cascaded (DRC) model. A rather striking feature of this model is that not only are there two routes but there are two distinct sets of computational principles—serial, rule-governed processes in the assembled route and parallel, spreading-activation processes in the lexical route. One would anticipate that interactions between these two routes (as might be required, for instance, to understand Glushko's [1979] type of findings) would prove difficult to model.

In the present article, we present a model of reading that maintains the uniform computational style of the PDP models but avoids the rigid commitment to a single route. We show that the kinds of sublexical spelling–sound regularities that are needed for good nonword performance can be extracted quite easily from exposure to a reasonable corpus of English words (including numerous exception words) by using networks and learning rules of much smaller computational power than those used by Seidenberg, McClelland, and their colleagues. Indeed, we suggest that the acquisition and storage of these regularities will be impeded or corrupted if the network is permitted at the same time to try and deal with exceptions. Next we show how the output of the network storing the sublexical regularities can be integrated with the output of another route (network) capable of handling inputs on a whole-word basis (including, of course, exception words). This interaction between the two routes in computing the final phonological output leads to observed interactions between lexical variables (e.g., word frequency) and sublexical ones (e.g., regularity, consistency). At the same time, the separation of these different kinds of knowledge into different systems enables successful modeling of the surface dyslexic syndrome.

A Two-Layer Network Model of Phonological Assembly

Rationale for the Model—Three Layers Good, Two Layers Bad?

Connectionist models have proved to be more fruitful than traditional approaches in many psychological areas,

[2] The first implemented model of the spelling–sound mapping was actually NETtalk (Sejnowski & Rosenberg, 1987), though it was not considered as a psychological model of the reading process.

[3] Note that grapheme nodes are detectors for letter strings and that some of the PMSP graphemes represent strings as long as or longer than actual words. Given that Plaut et al. (1996) deny the existence of orthographic lexical nodes, which are no more than letter string detectors for whole words, they are in a position of accepting the existence of letter string detectors (and presumably of the means for forming them) but requiring that such detectors never form for whole words. This position may be theoretically incoherent, depending on how Plaut et al. propose that the grapheme detectors come into existence.

providing an account of learning and the emergence of "rule-governed" behavior from a network of simple units and connections (e.g., McClelland & Rumelhart, 1986). The fact that such models can learn is one of their most interesting properties and provides a great opportunity to incorporate constraints from developmental data. In many cases, however, this possibility is ignored (but see Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett, 1997, for an overview of recent connectionist models of cognitive development). Models generated by learning algorithms are taken to implement a fully developed cognitive skill, and developmental data are not considered. This seems an unfortunate omission: In the case of reading aloud, we believe that data regarding children's reading are particularly useful in constraining a network model.

As discussed earlier, the onset–rime structure of syllables is particularly crucial for children's reading: Rhyming skills and the use of onset–rime spelling patterns promote learning to read (Goswami & Bryant, 1990; Treiman et al., 1995). Children's early ability to read new words shows that the acquisition of nonword reading skill does not require long training on words (Goswami, 1986; Treiman, Goswami, & Bruck, 1990). In Treiman et al.'s (1990) study, for instance, even first graders could read correctly about 50% of a set of 48 nonwords (e.g., FESH, TAIN). However, children's insensitivity to phonemes implies that such skill cannot initially depend on GPCs. Indeed, children seem rather to use bigger orthographic units such as onset and rime (Treiman et al., 1990). Hence, one of the aims of the present study is to investigate whether using onset–rime representations in a network model would lead to improved performance on nonword reading, even in the absence of any kind of explicit grapheme–phoneme training (indeed, in the absence of grapheme representations at all). However, along with the developmental evidence, the use of onset–rime representations is independently motivated by the findings in studies on adults (e.g., Jared, McRae, & Seidenberg, 1990; Treiman et al., 1995).

The second important issue, as mentioned in the introduction, involves the network architecture and the learning procedure. The development of algorithms for the training of multilayer neural networks (e.g., Rumelhart et al., 1986) has given modelers a great deal of freedom in designing processing architectures. This freedom encourages a pragmatic approach to modeling, and many theoretically important choices—How many hidden layers should we have? How many hidden units? How are the layers connected? Which learning procedure applies?—are decided either by instinct or trial and error. The problem with this is that the resultant performance of the network may be quite opaque, and no insight is generated into how the problem is being solved (see McCloskey, 1991, for a discussion of these issues). However, this need not be the case, and prior consideration of the task that the model is to perform may provide strong constraints on the network architecture.

Both the S&M and the PMSP models discussed earlier are multilayer, back-propagation-trained networks. Such networks, when provided with a sufficient number of hidden units, can encode arbitrary (nonlinear) input–output mappings. In contrast, two-layered networks have been considered relatively uninteresting because they permit just linearly separable mappings (Minsky & Papert, 1969). However, the mapping from orthography to phonology in English is not arbitrary. There are some inconsistencies, but to a large extent the correspondences between print and sound are systematic. This is even more evident if one looks at spelling–sound consistency at the level of onset–rime orthographic units, rather than at the level of grapheme–phoneme correspondences (Treiman et al., 1995). On the basis of statistical analysis of a large corpus of words, Treiman et al. (1995) conclude that orthographic units VC_2 (vowel plus final consonants, i.e., the orthographic rime) are useful as guides to pronunciation of the vowel; in other words, VC_2 letter clusters have relatively stable pronunciations and are much more consistent than vowels alone (V). In contrast, both routes involved in reading via meaning (i.e., orthography to semantics and semantics to phonology) require the learning of arbitrary mappings (Seidenberg, 1992). These routes have been implemented as multilayer connectionist models by Hinton and Shallice (1991) and Plaut and Shallice (1993). The core of such models is their recurrent architecture, which permits the development of attractors. The PMSP model has exactly the same basic architecture as these other models. Given that *direct* reading (phonological) and *indirect* reading (via semantics) involve substantially different mappings (i.e., systematic vs. arbitrary), we can therefore ask: What is the rationale for having the same network architecture (which implies the same computational power) for two such different tasks? Regular spelling–sound correspondences mean that, in the case of direct reading, the computational load is smaller. Therefore, a simpler mechanism of smaller computational power may be adequate. Such a mechanism might consume fewer computational resources and, crucially, may learn the mapping much more quickly. For these reasons, in the current work we decided to avoid the assumption of the need for hidden units (and associated slow and complex training algorithms) in mapping from orthography to phonology, and to start by looking at the behavior of a two-layer network equipped with psychologically motivated input–output representations.

Method: Description of the Model

The model's architecture is that of a simple two-layer feedforward network. This is a network with input and output layers but no intermediate layers (i.e., without hidden units). The model is trained on a set of 2,774 monosyllables extracted from the *Oxford Psycholinguistic Database* (Quinlan, 1993). This set consists of all the monosyllables with a Kucera–Francis frequency count greater than zero, plus some (missing) monosyllabic words used as word stimuli in various empirical studies. No explicit training was given on any kind of isolated GPC.

Input–output representations. The orthographic representation is strictly position specific, slot based, and equivalent to that used in the McClelland and Rumelhart (1981) model for the letter detector level. For each possible letter position there is an entire set of letter units that are supposed to be activated from a preceding feature detector level. However, the positions are defined with respect to orthographic onset (i.e., letters preceding the vowel letter) and

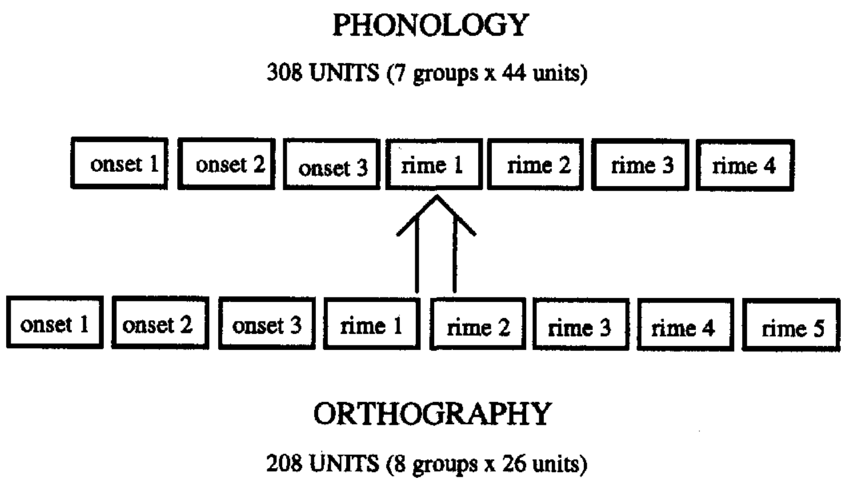

Figure 1. Architecture of the model. The arrow means full connectivity between layers. Each box stand for a group of letters (26) or phonemes (44).

orthographic rime (or word body, i.e., all letters from the vowel onward). The first three positions are for the (orthographic) onset representation, and the onset slots are filled from the first slot onward. The maximum length of the onset in English monosyllables is three letters (e.g., STR in STRING); a word with shorter onset, such as MILK, will have onset Positions 2 and 3 empty (i.e., all the letter units for such positions will be inactive). The orthographic rime, which has a maximum length of five letters (e.g., OUGHT in THOUGHT), is represented on the following five slots. Again, letters are "left justified"; that is, the rime slots are filled from the first rime position (Slot 4) onward (e.g., M_ _I L K_ _; T H_ I N_ _ _). Hence, the input layer consists of eight groups of 26 units, for a total of 208 units.

The input units are fully connected to a set of 308 output units representing the sound of a syllable (see Figure 1). The representation of the phonological output consists of seven groups of 44 units, three for the syllable onset and four for rime. Each unit in a group encodes one of the 44 phonemes of spoken English (20 vowels[4] + 24 consonants). The slots are filled according to the same scheme just described for the orthographic representation (e.g., /T/ _ _ /I/ /n/_ _). This representational scheme is in part motivated by evidence that phonemes in a syllable do not form a linear string: In fact, the vowel has a close bond with the following consonant(s), whereas it is less closely linked to the initial consonant or onset (Treiman et al., 1995). Note, however, that our orthographic and phonological representations do not incorporate explicit onset units or rime units (e.g., a single node representing ILK in MILK; see Norris, 1994); rather, letters and phonemes are "functionally aligned" to these units, so that their relative position in a letter string is invariant with respect to the rime (or onset) units to which they belong. This ensures that the network may detect this invariance during learning and therefore use onset–rime orthographic subpatterns (in addition to the identity of the single letters) as predictors of the word's pronunciation.

Clearly, phonological constraints dictate that most phonemes can occur in only one or two syllabic positions (see Hartley & Houghton, 1996). Orthotactic and phonotactic constraints are absent in our representational scheme, so we trade simplicity at the cost of some redundancy. However, by not prespecifying such structure, the model may learn the orthotactic and phonotactic constraints as a consequence of the exposure to a large set of words. Finally, our representation does not include complex graphemes, as used by Plaut et al. (1993, 1996); that is, there are no nodes standing for whole groups of letters. This ensures that any GPCs the network eventually extracts simply emerge from its attempts to predict onset–rime phonology from onset–rime orthography (see Zorzi, Houghton, & Butterworth, 1998, for simulation studies of the emergence of grapheme–phoneme mappings in the network).

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

$$net_i = \sum_j w_{ij} a_j, \tag{1}$$

where a_j is the activation value of the input unit j, and w_{ij} is the weight of the connections linking the unit j to the output unit i. The activation of the output unit i is determined by Equation 2. This is an S-shaped squashing function (sigmoid) of the net input, bounding phoneme activations in the range [0,1], and with $f(0) = 0$ (that is, no input and no output):

$$O_i = \frac{1}{1 + e^{-(net_i - 1)\tau}}, \tag{2}$$

where τ is a "temperature" parameter determining the slope of the function ($\tau = 3$ for all simulations below). Note that the -1 in the exponent shifts the sigmoid to the right, to give $f(0) = 0$ rather than the standard $f(0) = 0.5$.

Learning rule. The model was trained with the simple gradient descent technique known as the "delta rule" (Widrow & Hoff,

[4] Note that our phonological representation, as a consequence of the characteristics of Quinlan's (1993) database, incorporates 20 vowel phonemes, whereas the representation adopted by Seidenberg and McClelland (1989) for their training corpus had 14 phonemes. For instance, in Quinlan's database, vowel + R is considered as a single vowel phoneme (e.g., /O/ in bORn, /e@/ in squARE, /I@/ in chEER, /3/ in bURn), rather than as /vowel/ + /r/. The Seidenberg and McClelland training corpus was also adopted for the simulations carried out by Plaut et al. (1993, 1996) and by Coltheart et al. (1993).

1960). For any input pattern, the error correction is made by changing the weights according to the difference between the activation of the output units and the desired activation pattern. The desired output is just the correct pronunciation of the orthographic input (nodes that should be on have a target activation of 1, nodes that should be off have a target activation of 0). Formally,

$$\Delta w_{ij} = \epsilon(t_i - o_i)a_j \tag{3}$$

where ϵ is a learning rate (0.05 in the simulations), a_j is the activation of the jth input unit, and t_i and o_i are the teaching input and the actual output of the ith output unit, respectively. Frequency of presentation for the input words during training was not manipulated.

In contrast to back-propagation learning, training with the delta rule does not require that connections are initially randomized to some small value. Accordingly, connection weights are initialized to zero. This can be useful in that only input–output connections that become relevant for solving the task will be strengthened (or, say, actually come into existence). Units have no bias term (an extra weight that represents a threshold for that unit; see Rumelhart et al., 1986). On any trial the delta rule distributes the learning at each point of error over weights from all active input nodes. Early in learning, when all inputs produce errors, many weights are formed that later do not contribute significantly to performance but whose presence complicates analysis of the final weight set. Therefore, at the end of each epoch (one epoch corresponds to a single presentation of all training patterns), a pruning procedure sets to zero all weights below a certain value, which is calculated as a proportion (4% in the present simulations) of the maximum weight value.

Modeling reaction times: The response system. In comparing the behavior of the model to experimental data, we wish to be able to model reaction times (RTs). We take the variance in RTs in phonological reading to (at least partly) reflect differences in the time taken to construct a coherent articulatory plan at the output level. The basic two-layer assembly (TLA) model processes its input in a single pass and hence has a unitary RT. However, depending on the input, this single pass may not generate an unambiguous phonological representation. We thus propose that there is a competitive output process whereby an initial state of activation of the output layer, possibly involving multiple responses at any syllabic position, is resolved over time into a coherent articulatory plan.

For the purpose of modeling RT data, a competitive response system is implemented, which we will refer to as the "phonological decision system" (PDS), following Carr and Pollatsek (1985). The structure of this system is identical to the output layer of the TLA but incorporates features (lateral inhibition and gradual activation decay) that provide a temporal dynamic (see Figure 2). The output produced by the TLA model propagates gradually to the PDS, where the activations change over time until one of the units in each activated phoneme group reaches a response threshold (set to 0.5 for all simulations). In the absence of noise, the winner is the unit that was initially the most active, whereas the activation of the others drops to zero. Variations in the model RTs are due to variations in how long this process takes. The PDS will be used later for modeling the interactions between assembled (sublexical) and retrieved (lexical) phonologies, and the equations that govern the PDS can be found in the same section of the article (Modeling the Interactions Between Assembled and Retrieved Phonologies).

The competitive process takes place among the phoneme units for each position of the phonological output. This requires a certain number of cycles, which is greater if, at the beginning of the competition, two or more alternative phonemes for the same position are active. When within a certain group there is no activity in the PDS (i.e., the activations are under threshold), no response is given for that position (all the units are at 0.0). Note that we avoided the use of a "null phoneme," which is used, for instance, in Coltheart's DRC model (Coltheart & Rastle, 1994). Hence, to ascertain the model's response we just scan the final output pattern to see which units are active. In order to avoid the production of spurious responses that could emerge from low activation values, we use a simple threshold with value $\sigma = 0.2$; however, because any monosyllabic pronunciation must contain a vowel, the threshold value is lowered to $\sigma = 0.05$ for the units corresponding to the vowel phoneme position (i.e., in the first rime position). The interpretation of the system's output is straightforward: It consists of a string of phonemes. The number of cycles needed by the PDS to settle varies for different inputs and may be used as a measure of the naming latency (RT) for the letter string being processed.[5]

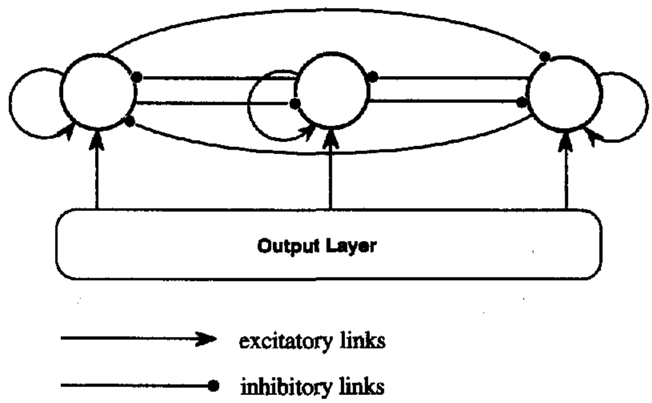

Figure 2. Sketch of the response layer. The 44 units for each phoneme position are connected by inhibitory links. Each unit has an excitatory feedback loop. The figure shows three units only.

The assumption underlying this kind of response system is that response latencies under time pressure are correlated with the amount of noise in the phonological output code. In particular, this noise might consist in more alternative phonemes being active for the same phoneme position. Note that this approach in modeling RTs is not limited to the reading task. Response competition is what accounts in general for the relevant part of the empirical RTs, even across different domains (e.g., in models related to attentional effects; Cohen & Huston, 1994; Houghton & Tipper, 1994; Zorzi & Umiltà, 1995). We should stress that this approach in modeling RTs is quite different from that used by Seidenberg and McClelland (1989). The latency in the S&M model was measured as a phonological error score, which is the sum of the squared differences between the desired output for each unit and its actual activation. In the same vein, the simulations with feedforward networks in Plaut et al. (1996) use an error measure known as "cross entropy." The main problem with such an approach is that we need to know the desired output. But how does one define the "desired" output for a nonword, particularly for a nonword with an inconsistent spelling? Moreover, if the output contains wrong phonemes, the S&M approach would deliver a high error score (and thus a long RT), even if the string of phonemes is relatively

[5] A similar assumption was incorporated in a simulation model of Lacouture (1989), which had a "decoding module" that provided naming latencies.

noise free. By contrast, interpretation of our response system is not dependent on knowledge of the expected output and can give a response latency (along with the response itself) for any orthographic input. Note, however, that the same is true for the attractor network version of the PMSP model, where response latencies are directly measured as the time taken by the network to settle to a stable state.

Results and Discussion

Pronunciation. Since the TLA model is inherently incapable of learning the whole training set, it cannot be trained until errors reach zero. We typically train the model until errors have apparently reached the minimum for the whole training set. After about 10–15 passes through the training set, further improvement in terms of error descent tends to be slow and relatively small. After 12 epochs of training, the model produces the correct pronunciation of about 81% of the words in the training set. Of the words it gets wrong, 75% (15% of the total) result in regularization errors and the rest are "other" errors, most of which (65%) involve the final consonant (i.e., missing consonant or spurious doubling). Following Besner et al. (1990), we tested the model's performance with nonwords using the 52 nonwords of Glushko's (1979) Experiment 2. The network made one true error, BEASH → /bi/, thus showing a performance of 98% correct. A more detailed test of the model was carried out using the stimuli from Glushko's (1979) Experiment 1, which involved both word–nonword and regular–exception dimensions. All but 2 from the 43 regular words of the set were pronounced correctly; the network chose for BATH the pronunciation /b&T/ to rhyme with HATH (in standard British English, the correct pronunciation is /bAT/), and DUNE was pronounced as /djVn/. For the latter case, inspection of the vowel phonemes revealed that the correct vowel /u/ was slightly less active than the wrong /V/ (0.6 vs. 0.7). Among the 43 exception words, 6 were pronounced correctly and 37 were regularized. Inspection of the 6 correctly pronounced words revealed them to be not very "exceptional." Two (MILD, WILD) had the rime ILD pronounced as /aIld/, whereas the regular pronunciation is supposed to be /Ild/. However, at the level of the rime, the most common pronunciation in the training set is just /aIld/. Two other words (DEAD, DREAD) had the rime EAD pronounced as /ed/, whereas the regular pronunciation is supposed to be /id/. Again, at the level of the rime, the most common pronunciation in the training set is /ed/. Finally, the network correctly pronounced the exception words MOST and WERE. Again, the pronunciation of OST as /@Ust/ is the most common in the training set.

The network made no errors with the two other lists of nonwords (43 regular nonwords + 43 exception nonwords), although sometimes a pronunciation of the rime which is irregular in terms of GPC rules was chosen. For an initial comparison to human performance, we can consider the number of "regular" pronunciations, that is, those that adhere to Venetzky's (1970) GPC rules. On Glushko's (1979) regular (consistent) nonwords, all of the network's pronunciations were regular, thus reaching 100% correct performance. The model, like humans, produces some irregular pronunciations in reading exception (inconsistent) nonwords, performing 88.4% correct (5 irregularizations). However, as originally found by Glushko, most of these alternative pronunciations are consistent with some other pronunciation of the nonword's body (e.g., GROOK → /grUk/, rhyming with BOOK, not SPOOK) and therefore can be accepted as correct (see Plaut et al., 1996, for a list of accepted pronunciations). This happens with all irregular pronunciations produced by the model (see Table 1), with an overall performance of 100% correct. Note that Glushko (1979), on the basis of the same considerations, found only 4.1% of the responses to be actual errors.

The high rate of irregularizations in reading inconsistent nonwords found by Glushko (1979) in his Experiment 1 has been attributed to intralist priming effects because of the presence of mixed word–nonword stimuli (see Patterson & Coltheart, 1987). In his Experiment 2, which contained only nonwords, Glushko found a lower error rate. Indeed, the model's error rates on the nonword lists are similar to those in Glushko's latter experiment (see Table 2).

Recent studies suggest that human readers exhibit a relative variability in pronouncing nonwords (e.g., Seidenberg et al., 1994); this might be captured in our model (and presumably in the PMSP model) by adding a stochastic component (e.g., noise) that influences the computation of phonology. It is interesting to note that even surface dyslexic patients such as M.P. (Bub et al., 1985) and K.T. (McCarthy & Warrington, 1986) produced some irregularizations in reading inconsistent nonwords. Finally, note that the PMSP attractor network produces more irregular responses than does our model on the same nonword lists (consistent—93% correct, 3 not accepted; inconsistent—62.8% correct, 15 irregular + 1 not accepted; Plaut et al., 1996).

It is important to note that children's ability to generalize to novel items is indeed not the result of a long training. The ability to read new words (and nonwords) emerges soon after some initial reading experience (Goswami, 1986; Treiman et al., 1990). Therefore, a crucial feature for a model of reading would be that of showing a good nonword reading performance in the early stages of training. The development and speed of acquisition of nonword reading skills in the TLA model were specifically tested in one of the simulation studies of Zorzi et al. (1998). It is striking that the model's performance had to be given in terms of training episodes (single exposures to input–output pairs) rather than epochs (exposures to the whole training set), because it was found that significant learning goes on within epochs. The

Table 1
Irregular Pronunciations Chosen by the Model for Glushko's (1979) Exception Pseudowords

Item	Model's response	Regular pronunciation
BILD	/baIld/	/bIld/
COSE	/k@Us/	/k@Uz/
DERE	/de@/	/dI@/
GROOK	/grUk/	/gruk/
PILD	/paIld/	/pIld/

Table 2
Percentages of Regular Pronunciations on Nonwords

Stimuli	Glushko, Experiment 1	Glushko, Experiment 2	Model, Experiment 1	Model, Experiment 2
Consistent nonwords	93.8	94.7	100	97.7
Inconsistent nonwords	78.3	87.7	88.4	100

Note. Lists are from Glushko's (1979) Experiments 1 and 2.

model starts to successfully generalize its learning early in training—for instance, scoring 15% correct after about 700 individual learning trials on different words. The model reaches 90% correct on Glushko's (1979) regular nonwords after about 7,000 learning episodes (i.e., less than 3 epochs) and 100% correct after 22,000–25,000 learning episodes (i.e., fewer than 10 epochs; see Zorzi et al., 1998, Figure 2).

These results show that a two-layer network equipped with psychologically motivated input–output representations can produce nonword reading performance greatly superior to that of the three-layer S&M network and at least equal to that of the three-layer PMSP attractor network. It does this with little training (12 epochs compared with a number ranging from 300 to 3,200 in the simulations of Plaut et al., 1993, 1996) and without the need for grapheme nodes. Plaut et al. (1996) did not explicitly test the model's nonword reading at earlier points of training: However, Figure 23 in Plaut et al. (1996) shows that nonword reading performance in the feedforward version of the model reached about 95% correct after 100 epochs.

Where do regularizations come from? In building our model, we avoided the assumption of the need for hidden units. However, the computational power of a two-layer network is obviously smaller than that of multilayer networks, and the absence of hidden layers leads to a corresponding absence of *intermediate representations* (Rumelhart et al., 1986). The ability to construct such representations means that it is possible for a network with hidden units to identify and correctly respond to individual items in the input set (say, exception words). This is clearly shown by both the S&M and PMSP models, which can correctly respond even to low-frequency exception words. Our model, by contrast, while extracting the statistical regularities of the sample of English words used as the training set, seems to avoid the learning of idiosyncratic cases (exceptions). That is, although the training set contains numerous words containing some degree of spelling–sound irregularity, the model extracted the statistically most reliable spelling–sound relationships and "ignored" the rest. In light of our findings with the two-layer network, this suggests that the intermediate representations formed by the hidden units in mapping from orthography to phonology can be considered, functionally speaking, as implementing a distributed internal lexicon (a similar argument has been made by Besner et al., 1990).

If this interpretation holds, then the lack of hidden units in our model is equivalent to the lack of lexical representations, which would explain why the model produces so many regularization errors in exception word reading, a pattern that resembles the reading performance of surface dyslexic patients. The model's error rate with exception words is 86%, whereas with regular words it is just 4.6%. Table 3 presents a comparison of the model in this respect with patients M.P. (Behrmann & Bub, 1992; Bub et al., 1985) and K.T. (McCarthy & Warrington, 1986). The model's performance shows a closer match to that of K.T., that is, to the severest surface dyslexic pattern. The errors with exception words are typical regularizations (see Table 4). We return to the issue of surface dyslexia in a later section, in relation to the interaction between assembled and lexical processes.

Connections versus GPCs. The use of only two layers in the model has the additional advantage that the connection pattern produced during learning can be readily analyzed. We found that pruning of connections (removal of small weights) dramatically reduced the number of nonzero weights. In some cases this permits the weight pattern to be "read" as context-sensitive and context-free mapping rules operating over letter groups of various sizes, from single letters to the whole orthographic rime. However, these sublexical print–sound correspondences are more sensitive to the local context than are typical GPC rules. For instance, the inhibitory connections from surrounding consonant letters play a crucial role in determining the pronunciation of locally ambiguous vowel letters by inhibiting alternative, but contextually inappropriate, vowel phonemes (see Zorzi et al., 1998, for detailed analyses). In other cases, the contextual information affects the pronunciation of a consonant (e.g., C in the first position + H in the second position produces /tS/; see Figure 3). Finally, for nonambiguous consonants there is a one-to-one mapping with little or no effect of context (initial consonants: B → /b/, D → /d/, L → /l/, M → /m/, etc.). We should stress that these kinds of GPCs developed during learning: They were not prespecified in the system and were not derived from explicit training on GPCs. Furthermore, the model can correctly respond to single letters and to single graphemes (see Zorzi et al., 1998, for the developmental analysis of this capacity). We are not aware of any other connectionist model that shows such capacity without training on GPCs or explicit coding of graphemes. Our model is trained by presenting words: A single letter or a single phoneme is never presented during learning.

Naming latencies. The above results establish that the network can perform the regular and nonword reading tasks

Table 3
Error Rate (%) Comparison of Model With Surface Dyslexic Patients M.P. and K.T.

Stimuli	M.P.	K.T.	Model
Regular words	4–10	3	4.6
Exception words	18–59	63	86
Nonwords	4.5	0	0

Note. The patients' data on regular and exception words are collapsed across frequency.

Table 4
Examples of Regularization Errors Produced by the Model

Item	Target	Model's response
BLOOD	/blVd/	/blud/
BUSH	/bUS/	/bVS/
DEAF	/def/	/dif/
HAVE	/h&v/	/heIv/
LOVE	/lVv/	/l@Uv/
PINT	/paInt/	/pInt/
SAID	/sed/	/seId/
STEAK	/steIk/	/stik/

and as such is a viable candidate for a model of assembled reading. Later we examine the interaction of regularity with lexical frequency by integrating the TLA model with a "lexical" route. This interaction involves variation in naming latency, and the pattern of results produced by the model depends on properties of both routes. However, we have found that the PDS shows systematic variation in naming latency when receiving input from the TLA model in isolation. We therefore feel it is important in understanding the behavior of the full model to first establish the determinants of RT variation that are due to the TLA model.

Naming latency differences are caused by the TLA model when it produces multiple candidates for some phoneme positions. The PDS has to resolve this competition to produce a coherent pronunciation, and it takes longer to do this (for any given position), the greater the amount of response competition. We therefore investigated whether the kinds of words typically used in naming latency experiments lead to significant RT differences when read by the TLA model in isolation and, if so, what factors would correlate with any difference found. Of course, with respect to real words, we cannot sensibly compare the model's latency pattern with human data, because normal participants will not be reading familiar words by an assembly procedure. It is nevertheless important to know what kind of behavior the TLA model produces on such stimuli if we believe there is an influence of assembly on normal word reading. On the other hand, a comparison of model and human data can be made in the case of nonword stimuli.

We examined the latency times for the four word lists from Glushko's (1979) Experiment 1 (see Figure 4). These lists consist of regular and exception words, and regular (i.e., consistent) and exception (i.e., inconsistent) nonwords. The four lists were presented to the model, and RTs were collected, the response criteria for the PDS being as described in the earlier section entitled *Modeling reaction times: The response system.* The model produced one pronunciation (0.6% of the total 172) that we consider implausible (DUNE → /djVn/), that is, one that does not match some other pronunciation of the word's body in the training set. We therefore excluded it from the analysis. The output was analyzed with a two-way analysis of variance (ANOVA) in which the factors were wordness (words vs. nonwords) and regularity (regular vs. exception). The analysis showed a significant RT advantage for regular letter strings: In processing an exception string, the PDS needs more cycles to settle, $F(1, 167) = 8.93$, $p < .005$. The wordness factor and the interaction between the two factors were not significant. Collapsed over regularity, mean latencies for words and nonwords were 4.1 and 4.0 cycles, respectively.

The absence of the wordness effect means that, although the model was repeatedly exposed to words during learning, there is no difference in processing a "known" word or a novel stimulus (nonword); this indicates that the model is truly nonlexical. On the other hand, it is clear that the irregularity of a letter string affects the naming latency. This is clearly a consistency effect, because there is no lexicality effect and the "regularity" effect holds for nonwords. To confirm the consistency effect, we added the two lists of nonwords from Glushko's (1979) Experiment 2, obtaining 69 consistent nonwords and 69 inconsistent nonwords. As

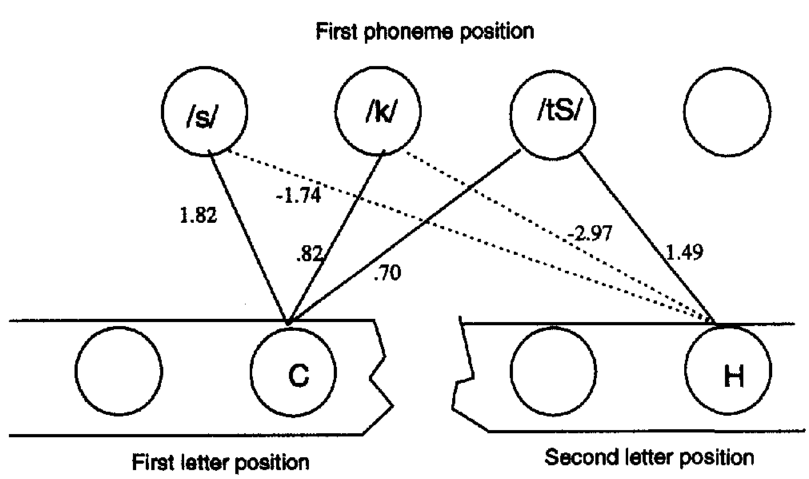

Figure 3. Example of context-sensitive connections. Dashed lines are inhibitory connections, and full lines are excitatory connections. The H in the second position reinforces the pronunciation of C as /tS/ and inhibits the alternative phonemes /s/ and /k/.

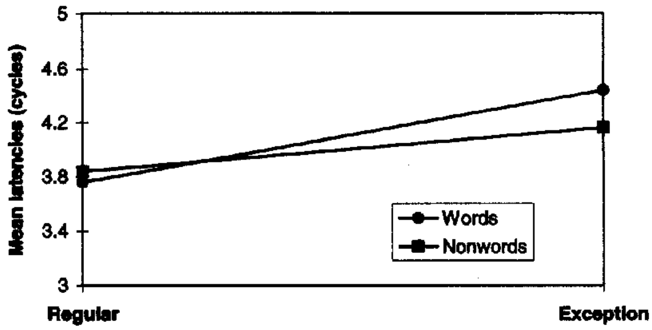

Figure 4. Model's latencies (cycles) on the items of Glushko's (1979) Experiment 1. The graph shows the regularity effect and the lack of a lexicality effect.

mentioned before, one of the consistent nonwords (BEASH) was mispronounced and was therefore excluded from the analysis. Again, the consistency factor reached significance in a one-way ANOVA, $F(1, 135) = 5.82$, $p < .05$. With respect to nonword reading, these results are clearly consistent with those found by Glushko, which showed longer naming latencies for inconsistent nonwords. The Glushko results have always been considered a challenge to dual-route models and as providing strong evidence supporting the lexical analogy model. This is not necessarily the case: Our network model can account for the fact that letter strings that share the rime spelling pattern with exception words can sometimes be produced with the irregular pronunciation and that a longer naming latency occurs.

To summarize these results: For nonwords, we have shown a reliable RT effect, as found in human data, which adds further support for the TLA model as a model of assembled reading; for words, we find the same effect of consistency but no effect of lexicality (it is irrelevant to the model whether or not it has been trained on a particular word). These results are important for understanding the interactions of the TLA model with output from a lexical pathway.

The consistency effect. Following Glushko's (1979) initial findings, consistency effects did not prove to be robust in some experiments (e.g., G. Brown, 1987; Seidenberg, Waters, Barnes, & Tanenhaus, 1984). Recently, Jared et al. (1990) demonstrated that when confounding variables are matched, the consistency effect is "consistent." Using our TLA model, we looked into the consistency effect in more detail by comparing exception words, regular inconsistent words (e.g., COVE), and regular consistent words (e.g., BEAM). We took the stimuli from the Taraban and McClelland (1987) study and used the lists of 24 low-frequency words of each type. We presented these three lists to the model and looked at the latency times. A mean number of cycles was calculated for each list. All regular words were pronounced correctly, whereas, not surprisingly, most of the exception words (71%) were given a regularized pronunciation. The regular inconsistent word COOK was pronounced as /kuk/, which is actually the "regular" pronunciation. A second regular inconsistent word was mispronounced (FOWL → /fOl/), and we therefore excluded its latency from the analysis.

Planned comparisons among the three word types were conducted after a one-way ANOVA was performed to obtain the error term, $F(2, 68) = 7.93$, $MSE = 1.39$, $p < .001$. Regular consistent words (mean latency = 3.33 cycles) were significantly faster than both exception words (mean latency = 4.22 cycles), $F(1, 68) = 15.34$, $p < .001$, and regular inconsistent words (mean latency = 4.66 cycles), $F(1, 68) = 6.60$, $p < .05$. The difference between exception words and regular inconsistent words did not reach significance ($F = 1.7$, $p = .2$).

The longer latencies yielded by inconsistent items (exception and regular inconsistent words) confirm the consistency effect that we found with the nonwords from Glushko's (1979) Experiment 1. Parkin (1983) found that inconsistency leads to slower pronunciation only when it exceeds some critical amount. Nonwords like YINT or BINT that have inconsistent neighborhoods in which one pronunciation clearly predominates seem to suffer no disadvantage (PINT is a unique exception). A disadvantage is suffered by nonwords like POVE that have inconsistent neighborhoods containing several examples of each type of pronunciation (e.g., MOVE, LOVE). This suggests that consistency is not a two-valued variable (see Henderson, 1985). The model shows the same effect: BINT takes 3 cycles to be produced, whereas the mean for inconsistent nonwords is 4.2 cycles. However, POVE takes 4 cycles. We would expect to find the same effect for words given that there is no lexicality effect in the model. Again, PINT (even if regularized as /pInt/) takes 3 cycles. The same happens for regular inconsistent words: WAVE, which has only one inconsistent neighbor (HAVE) but many consistent (CAVE, GAVE) neighbors, is produced in 3 cycles, whereas the mean for regular inconsistent words is 4.22.

Interestingly, Kawamoto and Zemblidge (1992) found that the naming latencies for the regularization errors were faster than the correct irregular responses. In this regard, Van Orden and Goldinger (1994) argued, "[T]his phenomenon would be most likely when irregular words have substructure common to many regular enemy words and have weak friends" (p. 1,282). Note that this is just in the case of PINT, which is produced by the model (in the regularized form) with the same latency as that of the regular enemies (e.g., MINT, LINT). Hence, pronunciation and pronunciation latency in the model seem to depend on a balance between competing neighborhoods. The rime pronunciation is that of the numerically stronger neighborhood, and the latency depends on the size in the training corpus of the winning neighborhood with respect to the losing neighborhood.

The graphs in Figure 5 show how the pronunciation of the same vowel in different words can change as a consequence of having different neighborhoods. The latency time, that is, the number of cycles in the PDS, is related to the activation values of alternative vowel phonemes that are initially active (see Figure 6).

The important point here is that the regular–exception classification is only an approximate description of the word types. It is interesting to note that Laxon, Masterson, and

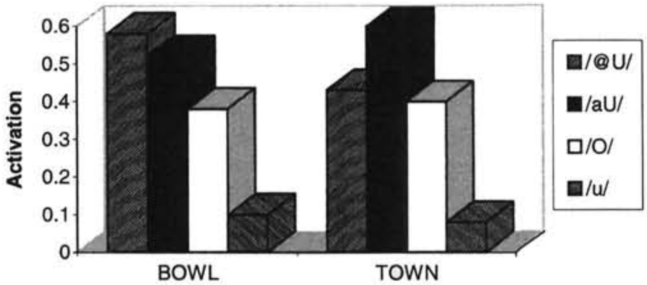

Figure 5. The graph shows the two-layer assembly model's output activation of the phonemes /@U/ and /aU/ for TOWN and BOWL.

Moran (1994) in a study on children's reading found a clear word type effect: Regular consistent words were read more accurately than were regular inconsistent words and exception words. However, there was no difference between the last two. The error pattern demonstrated that an awareness of alternate pronunciations of the rime caused errors in reading. A large proportion of these were regularizations of the exception words, but there were also other irregular versions of the rime for the regular inconsistent items.

Determinants of consistency: Types or tokens? It is clear that what really matters in the model is the number of word types. However, the conclusion of Jared et al. (1990) was that performance on word naming depends entirely on the number of tokens. They argued that it would be difficult for the S&M model to explain residual effects of the number of types, that is, the number of different neighbors, because there are no representations for individual words. However, Kay (1985, 1987; Kay & Bishop, 1987), on the basis of nonword pronunciation experiments, argued that the consistency effect is determined by the number of neighbors and not by their frequency. Given that the Jared et al. (1990) study did not include experiments on nonwords, these contradictory findings might be easily explained if participants were using different routes in the naming task depending on the presence or absence of nonwords. Baluch and Besner (1991) argued that the presence–absence of nonwords in the experimental stimuli induces participants to adopt different reading strategies. In fact, if nonword stimuli are included in the naming experiments, typical lexical effects such as semantic priming and frequency effects are not found (e.g., Baluch & Besner, 1991; Frost, Katz, & Bentin, 1987; Wydell & Humphreys, in press). Such data are considered to be strong evidence that assembled and addressed routines are available in all orthographies (Besner & Smith, 1992). The absence of lexical factors constitutes evidence for the use of the assembled routine. Hence, consistency effects might be somewhat different when nonwords are included in the experiment.

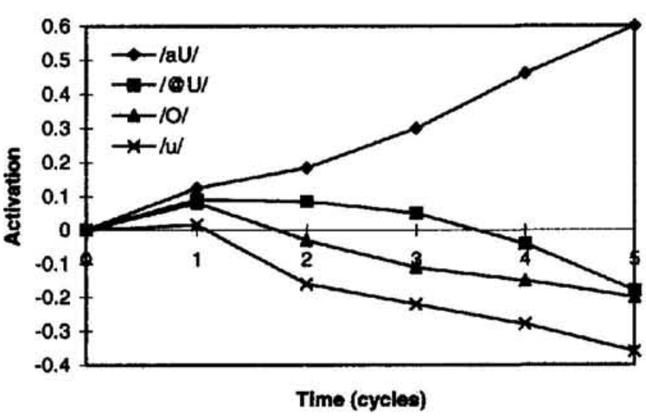

Figure 6. Competition in the phonological decision system between alternative vowel phonemes for pronouncing TOWN.

Units relevant to the definition of "orthographic neighbor." In most of the studies of consistency, the orthographic unit considered to define consistency and to define what counts as a neighbor is the rime. However, it has been suggested that the rime is not the only unit relevant to pronunciation. For instance, Kay (1987) found that the pronunciation of nonwords is influenced by the initial letter: Participants assigned the irregular pronunciation /Uk/ to the rime of WOOK more frequently than they did to POOK. Taraban and McClelland (1987) found other aspects of the word structure to be important, that is, initial letters including the vowel. An alternative measure of neighborhood size is Coltheart's *N* (Coltheart, Davelaar, Jonasson, & Besner, 1977). This refers to the number of words that differ in one letter from the given word.

Given these different perspectives, it would be valuable to analyze what defines an orthographic neighborhood in the model. By considering the pronunciation that the model chose for a given inconsistent word, we tried to discover why that pronunciation was chosen. It turned out that in a few cases the "rime neighborhood" cannot account for the model's pronunciation.

The rime is the most common (and paradigmatic) type of orthographic neighborhood. One example is SHONE (regular inconsistent), which is pronounced by the model as /S@Un/ instead of /S0n/. This case is straightforward: The rime (ONE) is pronounced as /@Un/ nine times and as /0n/ two times in the training corpus. A further example is the pronunciation of the rime ILD as /aIld/ rather than /Ild/; as previously discussed, the former is the most common pronunciation in the training set.

However, in inconsistent nonwords like POOK and ROOK, and in words like BOOK and HOOK, the model pronounces the OOK rime as /uk/ (rhyming with SPOOK), but as /Uk/ in GROOK. We can look at the network's connections to understand why this happens. The phoneme /U/ is inhibited by the letter K in the third rime position, and its activation in the context of OOK is 0.28, whereas the activation of the phoneme /u/ is 0.32. However, the /u/ phoneme receives some small inhibition from the letter G in the first onset position (possibly as a result of exposure to words like GOOD). Hence, the activation of phoneme /u/ in the context of [G] + [OOK] is reduced to 0.18, and the phoneme /U/ wins the competition (see Figure 7). Thus, the phonological interaction between orthographically similar words in the model can be complex, and in some cases the pronunciation of the vowel can be influenced by the onset, as

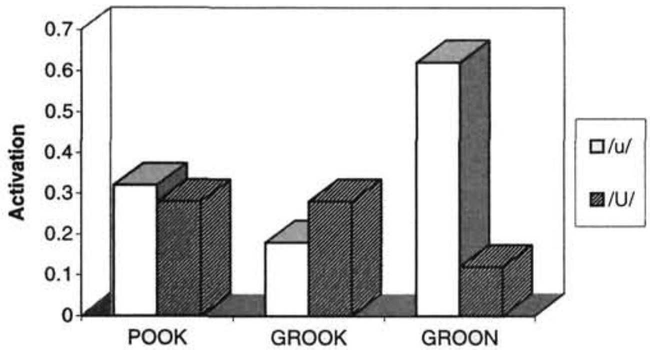

Figure 7. Effect of the onset in changing the vowel pronunciation of the same rime (OOK): the two-layer assembly model's output activation of the phonemes /u/ and /U/ for nonwords GROOK and POOK. The activations for GROON are shown as a control.

originally reported by Kay (1987) and Taraban and McClelland (1987).

Summary of the TLA model

Sublexical spelling-to-sound mappings are easily discovered by a simple two-layer feedforward network, that is, a network without hidden units, which is trained with the delta rule. This rule is formally equivalent to a classical conditioning law (the Rescorla–Wagner rule; Sutton & Barto, 1981) and has been directly applied to human learning by a number of authors (see, e.g., Gluck & Bower, 1988a, 1988b; Shanks, 1991; Siegel & Allan, 1996, for reviews). Although the model is trained on the full set of monosyllabic words, which includes many exception words, the two-layer network can only learn the statistically most reliable spelling–sound relationships. This means that the network is able to extract the statistical regularities of the training corpus but avoids specific learning of the idiosyncratic cases (i.e., the exception words). Note that the TLA model generates the correct pronunciation of about 81% of the words in the training corpus, a figure that is quite similar to the overall performance of the GPC algorithm developed by Coltheart et al. (1993), which scored 78% correct on its training set.

The onset–rime representation exploits the generalization properties and permits a near perfect monosyllabic nonword reading, with a performance that matches that of humans fairly well; the model can also read single letters and graphemes. We claim that the TLA network model can be considered as a computational implementation of phonological assembly, because the model's behavior is precisely what one would expect from the assembly procedure: correct reading of regular words and nonwords but regularization errors on exception words (with a pattern similar to that of a surface dyslexic patient). However, the model shows other interesting features. As we said earlier, the output of the network reflects the relative consistency of a given mapping. It is well known that the major locus of inconsistency in pronouncing English words is the vowel (e.g., EA in HEAD, MEAL, GREAT). The model, along with the most common mapping of the vowel, delivers other alternative, less common mappings, which are activated to a lesser extent. The model's final pronunciation is produced by the phonological decision system (PDS) on the basis of activation competition; this process is sensitive to response competition from alternative mappings, which we postulate to be a causal factor in naming latencies. The more a word is "inconsistent" with respect to the pronunciation of its components, the more the competition and the longer the latency (in processing cycles) that the model needs to produce the final output. This consistency effect is roughly defined in terms of the existence (or absence) of alternative pronunciations of a given word body and is found in the model even for pronounceable nonwords; the effect arises from exposure during learning to orthographically similar words that have different pronunciations. At the same time, there is no sign of lexicality effects in the sense that the "wordness" (i.e., being a word or a nonword) makes no difference to the model.

Most of the sublexical spelling–sound mappings discovered by the network are sensitive to the local context, and their size is variable, but may be single letter to single phoneme in highly regular cases such as initial B, D, L, M, Z, and so forth. The notion of abstract rule proposed by nonlexical route theorists may be intended either in the sense of an (implicit ?) mapping rule or of an (explicit ?) production rule (Patterson & Coltheart, 1987). We would consider the network's connections as mapping rules, whereas the GPC system developed by Coltheart et al. (1993) is more similar to an explicit production rule system. The conceptualization of assembly as a system of "flexible (mapping) rules" is not new. For instance, P. Brown and Besner (1987) suggested that orthographic inputs may be associated with a small number of phonological outputs. The possibility of rules with multiple outputs has also been proposed by Patterson and Morton (1985) in relation to the "body subsystem" contained in their model. Most notably, the idea of multiple outputs is incorporated in the notion of "islands of reliability" in spelling–sound correspondence put forth by Carr and Pollatsek (1985): Correspondences are more reliable for some graphemes than for others. Thus, the system would typically specify a single output for consonants and multiple outputs for vowels. P. Brown and Besner (1987) explicitly suggested "a system of rules which generates small sets of values for consonants and larger sets of values for vowels" (p. 481). This conceptualization of the assembly procedure is quite compatible with our TLA model. On the other hand, it contrasts sharply with the model of phonological assembly proposed by Coltheart et al. (1993), where the algorithm discovers and stores the most common GPCs but eliminates other alternative mappings. In our model, the apparently rule-based behavior and generalization to nonwords is an emergent property of the connectionist approach.

As mentioned before, one prediction is that the model behaves as a purely sublexical system because of a lack of computational power, that is, because the absence of a layer of hidden units leads to a corresponding absence of intermediate, "internal" representations (see Rumelhart et al., 1986). The lack of lexicality effects in the model suggests that the intermediate representations formed by hidden units in mapping from orthography to phonology may implement a kind of distributed internal lexicon. If this hypothesis holds, a three-layer network should process words by using the whole word context to produce the correct phonology: This issue is computationally explored in the next section.

Handling Exception Words: Task Decomposition and the Dual-Process Model

In the previous section we showed the viability of the two-layer network as a model of the assembly process, but clearly, for reading English words generally, some form of mediated mapping is required, at least for exception words. A mediated mapping, of whatever type, must involve an additional set of units lying between input and output. What form should the mediated mapping have? A traditional answer would be that it is simply lexical, as implemented in interactive activation models (e.g., Coltheart & Rastle, 1994; Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981). These models adopt a localist scheme for representing the lexicon, that is, one where each node stands for a single word, and this is clearly a possibility we must consider; this is done later in the article by combining the output of a simulated lexical route with the output of the TLA model. However, an interesting alternative is to simply provide the network with the necessary resources for producing a mediated mapping (viz., a set of hidden units between input and output) without directly stipulating how these resources should be used. The network can be trained with both direct and mediated connections from spelling to sound and can be allowed to construct its own "division of labor" between the two pathways. The important questions are (a) How well does the network generalize (read nonwords), assuming it manages to learn the training set, including exception words? (b) If generalization is adequate, how is the network performing the task? That is, what are the direct connections doing and what are the hidden units doing? (c) How is the division of labor related to the resources available? In particular, what is the effect of having more or fewer hidden units available?

Method: Description of the Model

In this section we investigate how a network provided with direct and mediated connections (i.e., via hidden units) between spelling and sound uses these resources when trained on the set of monosyllables used in the previous simulations. Thus, the input and output layer remain linked by the direct connections, and the hidden layer forms an alternative (fully connected) pathway from input to output (see Figure 8). Note that the only significant architectural differences between this network and those of Plaut, Seidenberg, and colleagues, is that we permit the input and output representations to make direct contact, whereas in all their models, all interactions are mediated by hidden units.

Thus, there are two pathways from input to output: the direct (unmediated) pathway, discussed above, and a second, mediated, pathway via a set of 200 hidden units. The input and output layers are shared by the two pathways, and the orthographic and phonological representations are the same as in the TLA model.

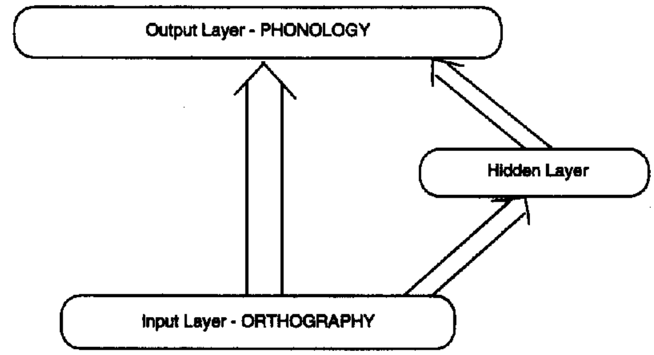

Figure 8. Architecture of the model with the hidden layer pathway. In both the direct pathway and the mediated pathway the layers are fully connected (arrows).

The model is trained with the standard backpropagation learning algorithm (Rumelhart et al., 1986). When an input is presented, activation propagates in parallel along both pathways. At the output, the net input from both pathways is simply summed and passed through a standard sigmoid function to generate the activation level for each output unit. The activation of each output unit is compared with the target activation and used to generate the standard backpropagation error term (Rumelhart et al., 1986). The error signals are then used to change the weights in both the direct and mediated pathways in parallel; that is, learning takes place in both routes at the same time. For the mediated pathway, backpropagation is used to compute hidden unit errors and train the input to hidden unit weights. The direct input–output connections need no backpropagation phase, and hence the training of this pathway essentially reduces to the use of the delta rule, that is, the learning algorithm used for the TLA model. Update of weights takes place after presentation of every input–target pair. The network was trained in this way on the same set of 2,774 monosyllables used for the TLA model for 300 epochs.

Results and Analysis of the Augmented Network

After 300 epochs of training, the model's performance was good for both regular and exception words, scoring 97.3% correct overall. We examined the generalization performance of the model by testing on the same lists of nonwords used for the analysis of the TLA model. We found that nonword reading performance was similar to that of the TLA model (see Table 5). On Glushko's (1979) consistent nonwords, the model produces two irregular (but not accepted) pronunciations (SWEAL → /swel/, WOSH → /wUS/), thus performing 95.3% correct. On the inconsistent nonwords, the model produces 9 irregular responses (see

Table 5
Percentages of Regular Pronunciations on Glushko's (1979) Nonwords

Stimuli	TLA model	Augmented network	Humans (Glushko, 1979)
Consistent nonwords	100	95.3	93.8
Inconsistent nonwords	88.4	79.1	78.3

Note. TLA = two-layer assembly.

Table 6), but all of these are accepted pronunciations (79.1% correct regular, 100% correct). Overall performance on Glushko's nonwords is thus 97.7% correct, leaving only 2.3% as actual errors. As for the TLA model, this level of performance is greatly superior to that of the S&M model and approximately equal to that of the PMSP model. It also matches the performance of human readers fairly well.

The development of intermediate representations in learning the orthography-to-phonology mapping is a recoding process, functionally analogous to the formation of lexical entries. We do not claim that the hidden unit representations strictly correspond to a lexicon, at least not in the sense that cognitive scientists usually use this concept. However, functionally speaking, the internal representations that mediate the spelling–sound mapping can be considered as implementing a distributed lexicon (also see Besner et al., 1990). Interestingly, this finding is consistent with Monsell's (1991) intuition about the different processes of "transcoding" and "identification" that may take place in a network model. According to Monsell, transcoding is the process of "producing an output pattern driven by the net effect of the individual elements of the input pattern.... To learn to identify rather than merely transcode is to learn what elements *co-occur* in input patterns. To learn this, a network needs to be able to form connections among input units.... A possibility is mediated connections via hidden units" (pp. 158–160). Under this interpretation, the hidden pathway would perform the identification part of the reading process (and then, of course, transcoding), whereas the two-layer network would perform transcoding without identification.

A detailed study of the augmented model is beyond the scope of the present article. However, to understand how the model performs the task, we may look at the contribution of the two pathways. To do this, we "lesioned" each pathway in turn: The lesion simply consisted in the complete disconnection of the pathway from the output layer. Use of the direct pathway in isolation produces a large amount of general activation of the phoneme units, at around 0.5.[6] However, particular nodes have activations that stand out against the background noise, and we take these to be the responses selected by the PDS. These responses largely represent "regular" pronunciations of the input string (see below for detailed analysis of a further simulation). The greatest locus of competition for output was in the vowel position, as found with the TLA model trained in isolation.

Table 6
Irregular Pronunciations Chosen by the Model for Glushko's (1979) Inconsistent Nonwords

Item	Model's response	Regular pronunciation
BILD	/baIld/	/bild/
BOST	/bO@st/	/bOst/
GROOK	/grUk/	/gruk/
PILD	/paIld/	/pIld/
PLOVE	/pluv/	/pl@Uv/
POVE	/puv/	/p@Uv/
SHEAD	/Sed/	/Sid/
SOOD	/sUd/	/sud/
WULL	/wUl/	/wVl/

With the direct pathway removed, we found that the mediated pathway produces a very "clean" output pattern, which consists just of the appropriate phonemes for the word being processed. This means that, when used in conjunction with the direct pathway, the hidden units act to inhibit the wrong phoneme candidates activated by the direct connections and to produce (or reinforce) the correct phoneme units. In the case of exception words, the greatest contribution of the mediated pathway is clearly in the major locus of ambiguity, that is, the vowel position, though it can also influence consonant pronunciations. For instance, the word CHROME is pronounced /tSrome/ by the direct pathway, but the hidden unit pathway inhibits the /tSr/ onset and produces the correct /kr/ onset.

In this simulation, we found that when provided with two pathways from input to output, the network still ends up with the direct route behaving like the TLA model. However, given sufficient hidden unit resources, the mediated route can learn to respond individually to most words in the training set (including the regular words) and tends to "ignore" (i.e., inhibit) the output of assembly. In the following simulation, we investigated whether the role of the direct pathway can become more prominent when the mediated pathway has less representational resources, that is, is less able to form distinct representations of the items in the training set.

Role of the mediated pathway: Training with fewer hidden units. To test the idea that the dominance of the mediated route depends on its representational capacity, we trained a new network with the same architecture described above but with a hidden layer of only 50 units. The network was trained the same way as the one in the previous simulation, on a corpus of 461 monosyllabic words (this corpus simply consisted of all words with a Kucera–Francis frequency value greater than 80). This network reached 96% correct performance after 500 epochs, at which point training was stopped.

The testing procedure used on the intact network and after damage to the separate pathways is the same as that used in the previous simulations. As noted above, when the hidden layer is removed, the direct pathway produces background activation of all phonemes at around 0.5 (because of the output activation function). This was disregarded, and the selected phoneme for each position was the one most active above a threshold of 0.95. To analyze the contribution of the network's two pathways, we first look at exception word reading. Table 7 shows the pronunciation of 10 exception words produced by the intact network and by each of the two pathways in isolation.

Clearly, the direct pathway in isolation produces a regularized pronunciation of the exception words. The mediated pathway in isolation, by contrast, does not produce the entire

[6] This is an artifact due in this simulation to the use of the standard sigmoid activation rule used with backpropagation, which for zero input produces an activation of 0.5. With the activation function used in the TLA, this would result in little or no activation of these phonemes.

Table 7
Pronunciation of Exception Words Produced by the Intact Model and After Lesions (i.e., Disconnection) of the Mediated Pathway and the Direct Pathway

Word	Output of the intact model (correct)	Output after lesion of the mediated pathway	Output after lesion of the direct pathway
BLOOD	blVd	blUd	V
FRONT	frVnt	frOnt	Vn
GREAT	greIt	grett	eIt
LOVE	lVv	l@Uv	V
MOVE	muv	m@Uv	u
OWN	@Un	aUn	@U
TRUTH	truT	trVT	u
TWO	tu	tO	u
WHOM	hum	wOm	hu
WHOSE	huz	wOz	hu

phonological form of the words but only produces the phonemes that represent the exceptional spelling–sound relationship. For instance, the mediated pathway produces the irregular /V/ in LOVE and /u/ in MOVE, in contrast to the regular O → /@U/ produced by the direct pathway. Thus exception words are produced by the combined action of the two pathways.

In the next analysis, we look at the contributions of the two pathways to nonword reading (generalization). Zorzi et al. (1998) show that a good nonword reading performance does not depend on the size of the training corpus, but rather on the existence of a nonword's components (e.g., onsets and rimes) in the training corpus.[7] Therefore, for testing the network's performance in reading nonwords, we selected from Glushko's (1979) 43 regular nonwords those items that share the orthographic rime with a real word in the training corpus, which provided 28 nonwords.

The performance of the intact network is 86% correct, with four errors (see Table 8). One of these errors is a spurious doubling of the final consonant (WEAT → /witt/); another error, WOTE → /r@Ut/, can be considered a lexicalization (from WROTE). Table 8 shows the model's nonword pronunciations after lesion of the mediated pathway and the direct pathway. After disconnection of the mediated pathway, the output of the direct pathway in isolation is similar to that of the intact network. Thus, the ability to read nonwords is well preserved. Interestingly, the lexicalization error has disappeared (WOTE is now pronounced as /w0t/). After disconnection of the direct input–output connections, on the other hand, the output of the mediated pathway consists of isolated phonemes, and not even a single nonword is pronounced correctly. Of the 28 nonword stimuli, there are 2 "null" responses (i.e., no output at all), 18 responses consist of a single phoneme (15 of which are a vowel phoneme), and the remaining 8 consist of two phonemes. Clearly, this performance is far from any acceptable level of nonword reading.

In conclusion, in this simulation the hidden unit pathway may be interpreted as a kind of exception word route, in the sense that the hidden units "identify" exception words as a whole and their output is strictly idiosyncratic. The direct pathway thus handles nonword reading and most of word reading, the hidden units simply correcting the locus of irregularity in exception words. The relative lack of representational power compared with that in the previous simulation leads to reading's being dominated by a sublexical assembly procedure.

Discussion of the Augmented Network

The most interesting finding regarding the augmented network is that it decomposes the reading task into two different procedures, one extracting sublexical spelling–sound relations (the direct pathway) and the other (the mediated pathway) forming word-specific representations (albeit distributed, given the use of backpropagation in training). However, the relative dominance of one procedure over the other depends on the representational power of the hidden unit pathway.

The fact that the two procedures are performed by different computational resources (direct vs. hidden pathways) is a clear example of self-modularization. That is, we do not specify a priori that the direct route should do assembled reading. However, given the amount of sublexical spelling–sound regularity in the training set, this is the most efficient use of this resource in reducing the overall error. In addition, the direct pathway can be relied on to learn these relationships faster than can a mediated pathway. Although these models can be fairly classified as "dual-route" models, it is important to emphasize the point that we do not prespecify what each pathway should do. Thus the possibility arises of explaining why dual-route processing systems might emerge from the interaction of task demands with an initial pattern of connectivity that permits both direct and mediated interactions between two levels of representation.

The capacity for self-organization has been afforded increasing importance in the recent neural network literature. For example, Jacobs, Jordan, and Barto (1991) pre-

[7] The TLA model trained on just 86 words was able to produce an acceptable pronunciation of 90% of the 86 Glushko (1979) nonwords (Zorzi et al., 1998).

Table 8
Nonword Pronunciations Produced by the Intact Network and After Lesions of the Mediated Pathway and the Direct Pathway

Nonword	Output of the intact model	Output after lesion of the mediated pathway	Output after lesion of the direct pathway
DREED	/dred/[a]	/dred/	/aI/
SHEED	/Sid/	/Sid/	/id/
NUST	/nVst/	/nVst/	/V/
POLD	/p@Uld/	/p@Uld/	/@U/
FEAL	/fil/	/fil/	/i/
HEAN	/hin/	/hen/	/n/
BLEAM	/blim/	/blim/	/i/
MUNE	/mun/	/mVn/	/un/
PEET	/pit/	/pit/	/it/
SOAD	/sOd/	/s@Ud/	/d/
STEET	/stit/	/stit/	/it/
WEAT	/witt/[a]	/wet/	/i/
BEED	/bid/	/bid/	—
BELD	/beld/	/beld/	/el/
SUST	/sVst/	/sVst/	/Vt/
WOTE	/r@Ut/[a]	/wOt/	—
BINK	/bI9k/	/bI9k/	/I/
PRAIN	/preIn/	/preIn/	/eI/
BORT	/bOt/	/bOt/	/O/
PLORE	/pl@U/[a]	/pl@U/	/@U/
HOIL	/hoIl/	/hoIl/	/hoI/
LOLE	/l@Ul/	/l@Ul/	/@U/
DOON	/dun/	/daUn/	/u/
GROOL	/grul/	/grUl/	/u/
SPEET	/spit/	/spit/	/it/
DOLD	/d@Uld/	/d@Uld/	/e/
MEAK	/mik/	/mik/	/k/
DORE	/dO/	/dO/	/O/

Note. Dashes indicate no response (i.e., no phonemes were activated).
[a]Nonaccepted pronunciation.

sented a network architecture that tends to modularize when the task consists of two or more subtasks. Applying the paradigm to the visual discrimination of the shape and position of several objects, they found that two such subtasks ("what" vs. "where") were allocated to, and more efficiently performed by, two separate networks. Interestingly, the "where" subtask was linearly separable and could be performed by a two-layer network. Note that it is easy to constrain a single backpropagation network to perform both subtasks (Rueckl, Cave, & Kosslyn, 1989), but this appears not to be the optimal solution. In a similar vein, Schmajuk and DiCarlo (1992) proposed a learning model in which classical conditioning typically occurs by error-driven changes in the weights in a two-layer network representing conditioned stimuli (input units) and conditioned responses (output units). However, it is known that many animals can solve conditioning problems equivalent to the exclusive-or (XOR) problem (so-called "configural learning"), which is not solvable by a two-layer network (Rumelhart et al., 1986). Schmajuk and DiCarlo therefore augmented their basic two-layer network with an additional pathway from input to output via a set of hidden units, trained by a variant of the backpropagation algorithm, which allowed the model to solve configural learning problems by developing appropriate hidden unit representations.

The parallels between these models and our own are clear: The "regular task," that is, regular word reading, can be performed by a two-layer (sublexical) network, whereas the "exception task," that is, exception word reading, requires a hidden unit (lexical) pathway. This task decomposition is a direct response of the network to the quasiregularity of English orthography. Interestingly, recent evidence from a task requiring participants to learn a new visuomotor mapping supports the idea of a modular decomposition strategy during learning (Ghahramani & Wolpert, 1997). The postulation of the existence of the direct pathway, and its role in reading, is what most clearly distinguishes the current model from previous network models of reading, because both the S&M and PMSP models do not include a direct mapping. In these models, all spelling–sound mappings, regular or irregular, depend on the development of intermediate (hidden unit) representations. In the analysis of the attractor network performed by Plaut et al. (1996), it was shown that the good generalization to nonwords achieved by their model was the result of the development of componential attractors, which reflect sublexical correspondences. However, it is interesting to note that the attractors for exception words were "far less componential"; more crucially, Plaut et al. found that the exceptional pronunciation of the vowel depends on the entire orthographic input, and in this respect there is a striking similarity with the role of the mediated pathway in our model when trained with fewer hidden units.

Finally, note that an interesting possibility for obtaining a self-modularization into assembly and lexical pathways might be that of using constructive algorithms, such as cascade-correlation (Fahlman & Labiere, 1990). In this learning procedure, the initial network is two-layered; when the error descent reaches an asymptote, that is, when no further improvement of the performance is achieved, the learning algorithm starts to add a hidden unit pathway, which provides new computational resources.

Modeling the Interactions Between Assembled and Retrieved Phonologies

The simulation study with the TLA network model shows how the assembly procedure can be easily implemented as a connectionist system. However, some form of mediated mapping (that is, one that allows the formation of some kind of whole-word knowledge) is necessary, at least for the correct pronunciation of irregular spelling patterns. As noted before, a mediated mapping, of whatever type, must involve an additional set of (hidden) units lying between input and output. The simulations in the previous section investigated one possibility in which the hidden units are allowed to form distributed representations. An important deficiency, however, is that this model lacks a temporal dynamic at the point of interaction between the assembly and lexical procedures. Furthermore, because word frequency was not manipulated, the simulation of experimental results such as the standard Frequency × Regularity interaction was not possible.

A different, and more traditional solution to the problem of providing case-specific, whole-word knowledge is that of using a localist scheme to realize the mediated mapping. That is, the lexical knowledge is provided by simply having one node for each known word, as implemented in interactive activation models (e.g., Coltheart & Rastle, 1994; Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981; Norris, 1994). The use of a localist scheme for representing lexical knowledge has indeed some merit (see Besner, in press), and the intrinsic limitation of a localist network model of the lexical route (e.g., Coltheart & Rastle, 1994; Grainger & Jacobs, 1996)—that is, the lack of generalization to new stimuli—might be not a problem if it is combined with a second network model that provides generative knowledge (i.e., our TLA model). The combination of the two networks, one providing case-specific knowledge and the other providing generative knowledge, can solve the quasi-regularity problem. The lexical process, operating by recognizing input words, must still send output to the PDS, and the possibility of parallel operation of the two systems clearly exists. Thus, in this section we simulate the interaction of a frequency-sensitive lexical procedure with the TLA model described in the first section.

The input to the PDS from the two processes can be different; for instance, if the word being processed is irregular, the two procedures can reach conflicting decisions about the appropriate pronunciation. According to Carr and Pollatsek (1985), two different kinds of solution are possible. The first possibility, proposed by Coltheart and colleagues (Coltheart, 1978; Coltheart et al., 1977), is that conflicts can arise in principle but are minimal because the visual lexical route is much faster than the phonological GPC route. Thus, for visually familiar words, recognition is always completed before the nonlexical phonological code becomes available; conflicts do occur only when words are somewhat unfamiliar (i.e., low-frequency exception words). Note that this position is mostly maintained in the recent computational implementation of the dual-route model (Colthart et al., 1993, Coltheart & Rastle, 1994), because the GPC route is held to be slow and serial (e.g., one phoneme is produced about every 7 processing cycles of the lexical route). In Coltheart's DRC model, the processing level where lexical and GPC routes converge (and interact) is called the "phoneme system." The second solution proposed by Carr and Pollatsek (1985) is that both procedures work in parallel and that the processing rates of the two systems can be sufficiently similar. This possibility implies that conflict is not only possible but is a rather normal style of processing, in particular if we adopt a model of assembly in which multiple outputs can be delivered in parallel for a single orthographic code. Given the characteristics of the network model of assembly that we have developed in the first section—that is, a process producing parallel activation of phonemes (rather than serial) and involving multiple outputs—we are committed to this second solution. Therefore, the outputs of the two procedures enter the PDS in parallel, where the decision process is based on competitive interactions. We show how this convergence can account for data regarding the interaction of case-specific, lexical variables (word frequency, lexicality) with sublexical variables (spelling–sound consistency).

Method: Description of the Model

The PDS consists of a single layer of phoneme units where the output of the two processes converge into a single phonological code. As previously discussed in relation to the TLA model, the PDS consists of 6 "slots" of phonemes, each comprising 44 phoneme units for a total of 308 units. The units within one slot (i.e., phonemes for the same position) are in a state of competitive interaction by means of mutual (lateral) inhibition.

Nodes receive input from both lexical and assembly routes in parallel. This input builds up gradually over time, to simulate the cascaded, interactive nature of the earlier processes. Hence, the two inputs to the PDS are "ramped" so that they reach their respective maximum levels only after a number of processing cycles. To simulate this process in a simple fashion, we set the maximum level of input for each route to be that produced by the net input rule, and the ramping is realized by multiplying the net input by an exponentially time-varying scaling factor. Formally,

$$Input(t) = net_i * \lambda^{[ramp-t]^+} \tag{4}$$

where $Input(t)$ is the input from either of the processes to the phoneme node u_i at time t, net_i is the asymptotic input to u_i, λ is a scaling factor, *ramp* is the length of time it takes (in discrete processing cycles) for the input to reach its asymptotic value, and $[x]^+ = max(0, x)$. Clearly, it is possible to use two different values of *ramp* for the two processes, to obtain different speeds for the processes in reaching asymptotic output.

The activation level of each phoneme node u_i is governed by the following equation:

$$a_i(t) = net_i^{[ext]}(t) + net_i^{[int]}(t) \tag{5}$$

where $net_i^{[ext]}$ is the external net input to u_i at time t, that is, the input coming from the two processes scaled through the ramping factor, and $net_i^{[int]}$ is the internal net input to u_i at time t, that is, the inhibition coming from other active nodes (however, only positive activations propagate through the lateral connections) plus a percentage of the previous activation state of the node. Formally,

$$net_i^{[int]}(t) = \delta \cdot a_i(t-1) - \sum_j w^- \cdot o_{j\neq i}(t), \tag{6}$$

where a_i is the state of the node at the previous timestep $(t-1)$, δ is a decay parameter, o_j is the output of the unit u_j at time t, and w^- is the weight of the lateral inhibitory links. Finally, the output of the node is determined by an S-shaped squashing function, which bounds activations in the range $[-1, 1]$:

$$o_i = \frac{2}{1 + e^{-\tau a_i}} - 1, \tag{7}$$

where τ is a parameter that determines the slope of the squashing function. Note that this function allows suppressed states of activation (i.e., modeled as negative activations; see Zorzi & Umiltà, 1995), and that for values of 0 the function is the identity, $f(0) = 0$; that is, no input, no output.

The behavior of the system is straightforward: The activation of the nodes builds up gradually over time until a certain response threshold, which we set to 0.5 for all the simulations reported below. The amount of time (in number of processing cycles) that is required to reach the response threshold for a certain input being processed is taken as a measure of the naming latency. The existence of lateral competition between alternative phonemes is what determines the interference (and thus the longer latencies) in the processing of input stimuli that possess some intrinsic ambiguity (e.g., inconsistency of the mapping). This kind of competitive interaction is what accounts in general for the relevant part of empirical RTs (see Houghton & Tipper, 1994; Zorzi & Umiltà, 1995).

Fixed parameters used in the simulations are as follows: $w^- = 0.9$, $\delta = 0.85$, $\tau = 2$, and $\lambda = 0.8$. Although it would seem that the PDS has a number of free parameters, w^- (weight of the lateral inhibitory links), δ (decay), and τ (slope of the activation function) are simply set to values that permit an optimal winner-take-all behavior, that is, one where any nonzero activation value can reach response threshold in a finite number of timesteps and the alternative phoneme candidates are eliminated in the course of the competitive process.

Assembly and lexical processes. We need now to describe in more detail the nature of the output of the two processes, which feed into the PDS just described. The output of the assembly procedure is precisely that of the two-layer network previously described (the TLA model). To summarize, the output of the network is in the [0, 1] range, and more phonemes may be active for the same phoneme position as a result of the inconsistency of the mappings (e.g., both phonemes /i/ and /e/ can be active for a word or a nonword with the body EAD). The output of the TLA model thus constitutes the net input from assembly; as discussed above, this value is "ramped" to reach its asymptotic value over some timesteps (8 cycles in all simulations).

For the lexical procedure, we assume that the output comes from an interactive activation network similar to that of Coltheart and colleagues (Coltheart et al., 1993; Coltheart & Rastle, 1994); in fact, only the output of the phonological output lexicon is simulated: Therefore, any word contained in the (putative) model's lexicon sends excitation to the correct phonemes and inhibition to all the other phonemes. This is consistent with our findings with the three-layer network: The network's analysis suggested that the role of the lexical pathway is that of producing the correct phonemes for a certain word and of inhibiting wrong candidates that may be activated through the sublexical correspondences. Furthermore, it is assumed that the lexical pathway is sensitive to the frequency of the words. The strength of the lexical pathway is given by

$$o_{\text{lex}} = k \cdot f\,(KFfreq), \qquad (8)$$

where k is a constant ($k = 3$ for the excitatory output and $k = 5$ for the inhibitory output) and f is a logarithmic function of the Kucera and Francis (1967) frequency count. This function has a range of [0, 1], with higher values for items of higher frequency. The straightforward result is that the strength of the output of the lexical pathway is modulated by the frequency of the word being processed, although in the case of a nonword no output is produced.[8] Note that the inhibitory input to all wrong phoneme candidates in the PDS is stronger than the excitatory input to the correct ones. This is necessary because, different from the standard dual-route model, the output of the assembly procedure is thought to be fast and parallel.[9] As for the sublexical process, the output of the lexical pathway is "ramped" in order to build up over time, with the asymptotic value reached after 10 processing cycles. This, combined with the frequency manipulation, results in the lexical input to the phoneme nodes in the PDS increasing with a speed that is a monotonic function of word frequency. That is, for a given value of lexical input, the phonemes will receive such input faster (i.e., earlier in time) as the base frequency of the word increases (see Figure 9).

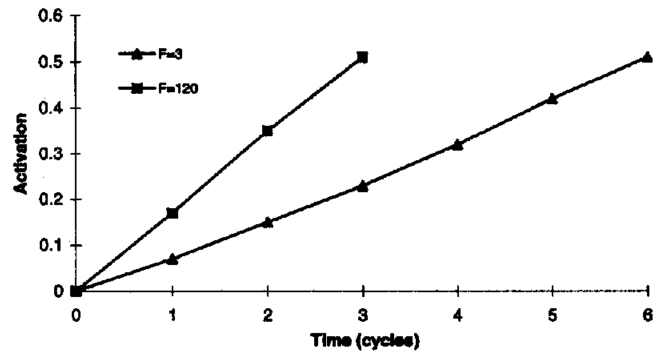

Figure 9. Excitatory lexical input to a correct phoneme in the phonological decision system. The incoming input for two words with frequencies of 3 and 120 is shown over time until the phoneme node reaches response threshold (without input from assembly).

Note that the time for the output of the lexical process to reach asymptotic value is slightly longer than that for the output of the assembly process. This is contrary to the standard assumption of the dual-route model, where the assembly route is thought to be far slower than the lexical route. However, the idea that assembled phonology can be activated faster than retrieved phonology has gained empirical support (Berent & Perfetti, 1995; Colombo, Cubelli, Zorzi, & Caporali, 1996). Furthermore, given that both processes operate in parallel, it is conceivable that a direct pathway (assembly process) might be faster than a mediated pathway (lexical process); that is, in a mediated pathway there is at least one additional layer of nodes that must be traversed.

No other assumptions are needed for the simulation of this lexical pathway. Note that this kind of lexical pathway could be simulated in a variety of ways. In particular, its behavior is precisely what would be expected from an interactive activation model, such as that implemented by Coltheart and Rastle (1994) in their DRC model. In this case, the activation coming from the lexical procedure would be sent by a single node representing the phonological form of the word in the phonological output lexicon. Alternatively, if distributed representations were preferred, the lexical procedure might be implemented by a recurrent network (with hidden units) that develops strong lexical attractors (so that the cycles needed by the network to settle to a stable output would depend on the frequency of the input word). However, an important advantage of a localist, interactive activation model of the lexical route is that it can be readily used to simulate the lexical decision task. Although we are not concerned with lexical decision in the present work, it is worth noting that Grainger and Jacobs (1996)

[8] Note that this way of implementing a frequency-weighted lexical pathway is very similar to Plaut et al.'s (1996) simulation of the influence of a "semantic" pathway (see General Discussion).

[9] Excitatory and inhibitory input are set to values that allow the model to produce the correct pronunciation of low-frequency exceptions (i.e., the correct phonemes can win the competition with alternative candidates).

have successfully simulated all major data on lexical decision with a semi-stochastic version of the interactive activation model; however, modeling of word recognition has proved difficult within a distributed model such as the S&M model (see Besner et al., 1990; Fera & Besner, 1992).

The architecture of the complete model, depicted in Figure 10, is therefore formed by three distinct connectionist modules. One module provides an assembled phonology (our TLA model), another provides a retrieved phonology (Coltheart's interactive activation network, or a similar implementation), and the two modules feed into a shared output module (our PDS). As discussed above, both retrieved and assembled phonologies enter the PDS in parallel, and the two processes (lexical and sublexical) have similar processing rates. It should be clear at this point that there is no real "horse race" between the routes, at least not in the classic sense of the standard dual-route model, but rather something similar to the "pooling process" proposed by Monsell et al. (1992). However, we have assumed that, at least in normal reading by skilled readers, the output of the lexical pathway is stronger than that of the assembly procedure, because lexical reading is generally reliable and can be unambiguously used on any known word. This interpretation of the dominance of one route over another accords with the interpretation of automaticity developed by Cohen, Dunbar, and McClelland (1990) and Cohen, Servan-Schreiber, and McClelland (1992) for connectionist models such as that of the Stroop task.

Results and Discussion

The Taraban and McClelland (1987) study. In this simulation we replicate with the model the study of Taraban and McClelland (1987), which involved three different types of word stimuli (regular consistent, regular inconsistent, and exception) and two frequency bands (high and low frequency). All words are produced by the model with the correct pronunciation; Figure 11 shows the mean latencies (in processing cycles) for the different lists of stimuli.

The RTs produced by the model were submitted to a two-way ANOVA with the factors of word type (regular consistent, regular inconsistent, or exception) and frequency (low or high). Overall, high-frequency words are faster than low-frequency words, $F(1, 138) = 51.56, p < .001$. There is a significant main effect of type, $F(2, 138) = 48.53, p < .001$, and a significant interaction of type with frequency, $F(2, 138) = 10.88, p < .001$. Post hoc comparisons (with Bonferroni correction) revealed that, among low-frequency words, exception words are slower than both regular inconsistent words, $F(1, 46) = 18.75, p < .001$, and regular consistent words, $F(1, 46) = 42.18, p < .001$, whereas regular inconsistent words are slower than regular consistent words, $F(1, 46) = 12.39, p < .001$. For high-frequency words, only the difference between regular consistent words and exception words reaches significance, $F(1, 46) = 9.31, p < .005$. Within each word type, the effect of frequency is significant for both regular inconsistent words, $F(1, 46) = 12.39, p < .001$, and exception words, $F(1, 46) = 67.46, p < .001$, but not for regular consistent words ($F < 1$). Therefore, the model replicates the basic finding of the interaction between frequency and consistency, with the consistency factor affecting the RTs more on the low-frequency items.

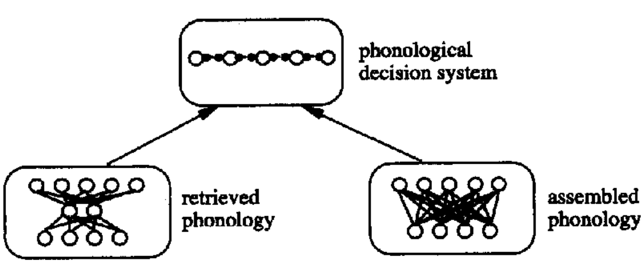

Figure 10. Lexical and sublexical procedures in reading aloud, and their interaction in the phonological decision system, where the final phonological code is computed for articulation.

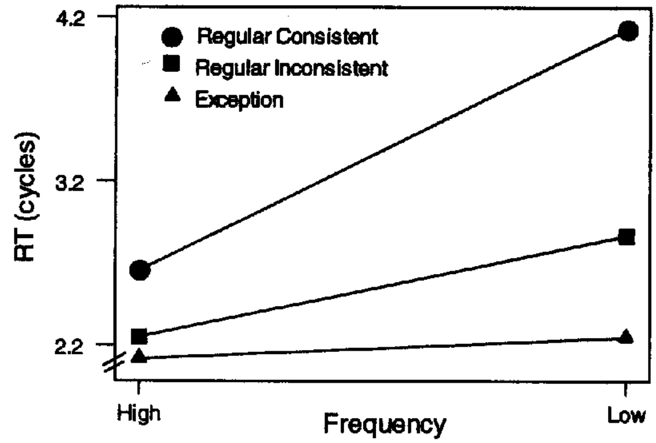

Figure 11. Naming latency (RT = reaction time) of the model on various lists of stimuli from Taraban and McClelland (1987).

The frequency-by-regularity interaction clearly arises from the interaction between the two processes: The output of the lexical pathway for the low-frequency words is weak and the buildup of a phonological code can either suffer (if the word is an exception) or enjoy (if the word is regular) the contribution of assembly. Note, however, that the consistency effect for regular words is entirely due to the characteristics of the assembly procedure: The disadvantage of regular inconsistent words compared with regular consistent words is the result of the activation of multiple phoneme candidates produced by the TLA model. The presence of multiple phonemes activated for the same phoneme position causes competition in the PDS that takes time to be resolved.

How do regular inconsistent words differ from exceptions? Clearly, in both cases there is competition caused by the activation of multiple vowel phonemes from the assembly procedure. However, in the case of regular inconsistent words, the correct vowel phoneme is usually the most active one, whereas in the case of exceptions, the correct phoneme is weakly activated (or even not active—e.g., the irregular /V/ in BLOOD) and there is a stronger alternative candidate (i.e., that resulting from the regular, most common spelling–sound relationship). This is clearly demonstrated by the behavior of the TLA model in isolation (analyzed in the first section of the article): Regular inconsistent words are pronounced correctly (i.e., the stronger vowel phoneme is the correct vowel pronunciation), whereas exception words are regularized (i.e., the stronger vowel phoneme is an alternative, more "regular," pronunciation). Therefore, for a regular inconsistent word, the lexical route and the assembly route reinforce each other, and thus the correct phoneme can inhibit the (wrong) alternative candidates via the lateral inhibitory connections (also, the lexical route sends inhibition to the wrong candidates). For an exception word, however, the phoneme with higher activation is a wrong candidate that can be "killed" only by the lexical route, and the activation of the correct phoneme relies mostly on the retrieved phonology.

In summary, for regular inconsistent words, the stronger vowel phoneme within the candidates produced by the assembly process is the *same* vowel phoneme activated by the lexical process; for exception words, the stronger vowel phoneme within the candidates produced by the assembly process is *different* from the vowel phoneme activated by the lexical process. This simple analysis reveals that the regularity effect and the consistency effects have different natures. Note that both effects show up in the model at the output level (the PDS): The two effects can be easily confounded because they arise in the PDS for the same reason—that is, owing to the competition between alternative phoneme candidates, which results in longer processing times. Nonetheless, in spite of this apparent similarity, the two effects have different origins.

Lexicality effect. In addition to the standard frequency-by-regularity interaction, the model ought to show another basic phenomenon—that is, the lexicality effect. Low-frequency regular words are read faster than nonwords (P. Brown, Lupker, & Colombo, 1994; Glushko, 1979; McCann & Besner, 1987). For dual-route theorists, the interpretation of this effect is in terms of sources of phonological information: Phonological codes for known words come from two sources; that is, both retrieved and assembled phonologies are delivered by the two routes, respectively, whereas phonological codes for nonwords are entirely based on the output of the assembly process. If the two routes are in agreement (as they should be for regular words), then they reinforce each other, and the response is quicker. Turning to single-route models, we note that the lexicality effect has not been discussed by Plaut et al. (1996). Within their framework, the source of this effect could be either within the orthography-to-phonology network, or because real words enjoy a contribution from semantics, or both.

To investigate the presence of a lexicality effect, we compared the naming latencies produced by the model on the 24 low-frequency regular words (from Taraban & McClelland, 1987) with those obtained on a list of 24 nonwords. The list of nonwords (from Plaut et al., 1996, Appendix 1) was generated from the Taraban and McClelland (1987) stimuli by altering the word onsets. All words and nonwords were given correct pronunciation; a one-way ANOVA on the RTs produced by the model showed a significant effect of lexicality, $F(1, 46) = 54.04$, $p < .001$, with low-frequency regular words ($M = 2.25$ cycles) yielding faster RTs than nonwords ($M = 4.37$ cycles). Thus, the latencies for nonwords are generally longer than those for known words, even longer than low-frequency words collapsed over word types (3.08 cycles using a composite mean; $F(1, 46) = 41.57$, $p < .001$. The lexicality effect—that is, the RT advantage of words over nonwords—is due to the fact that even the lowest frequency regular words enjoy some contribution from the lexical pathway, whereas for a nonword, only a single source of phonological information feeds into the PDS.

Besides the lexicality effect, it is worth noting that we have already demonstrated in the first section of this article that the model shows a reliable effect of consistency for the nonwords, with consistent nonwords yielding faster RTs than inconsistent nonwords.

Further Issues: Modeling Surface Dyslexia

We showed earlier that the reading performance of the TLA network model is similar to that of a severe surface dyslexic patient. However, surface dyslexic patients such as K.T. and M.P. show a frequency-by-regularity interaction; that is, their performance on exception words gets worse as the frequency of the words decreases. However, they do not show such an effect for regular words. How can the model accommodate this finding?

We have presented computational evidence that a network model acquires lexical properties by means of intermediate representations, that is, when a hidden layer is present. Both the S&M and PMSP models read via a hidden unit pathway only and clearly show frequency effects. By contrast, as discussed earlier, our network model of assembly is not sensitive to word frequency, because any kind of word level is absent. However, the lexical route is sensitive to the word's base frequency, so a lesion to it would result in a greater impairment of the lower frequency items (but regardless of word type). Therefore, what does a dual-process model predict about the behavior of the whole system? Again, we must look at the sources of phonological information that are available to the system. If the assembly procedure is not impaired, regular word and nonword reading would be unaffected. Nonetheless, the impaired lexical route would still be able to deliver the phonology of many high-frequency words; however, the lexical output for the lower frequency words would be too weak (or too slow) to impede the production of a final phonological code that is mostly based on the assembly code: If the word is an exception, the result is a regularization error.

We can test this hypothesis by lesioning the lexical route of the model so that the production of lexical (retrieved) phonology would be affected. A simple way to do this consists in slowing the rate of the activation flow for the retrieved phonology, which is controlled by the ramp parameter mentioned earlier. Therefore, by increasing the value of the ramp parameter of the lexical route, the activation feeding into the PDS at a given timestep will be weaker than that in the normal operation mode. Note, however, that we do not manipulate the absolute output strength of the lexical route, but rather the amount of time that it takes to reach asymptotic output activation. An alternative way would be to decrease the actual strength of the lexical output by decreasing the value of the excitatory or inhibitory connections (or both) that send activation to the PDS. In either case, however, the lesion does not affect the assembled phonology in any way.

To evaluate the effect of a lesion to the lexical route, we tested the model on Taraban and McClelland's (1987) regular and exception words. We can compare the model's performance to that of the surface dyslexic patient K. T. (McCarthy & Warrington, 1986) by finding the value for the ramp parameter at which the performance of the model on low-frequency exception words provides the best match to that of the patient. We then take the model's performance on the other word lists with that same parameter value and compare this with the patient's data. It turns out to be easy to find a suitable parameter value, which we finally set to *ramp* = 18. The comparison of the model's output with the data from K.T. is shown in Figure 12. It is clear that the model shows the same interaction between frequency and regularity, with the performance on low-frequency exception words being much worse than that on high-frequency exception words. Regular word (and of course nonword) reading, however, is clearly unaffected.

In our model even a severe lesion of the lexical route cannot affect the reading performance on regular words and nonwords. K.T.'s performance is indeed the result of a severe impairment, as shown by the close match to the performance of the isolated TLA model and by the dramatic slowing–weakening of lexical phonology that was necessary to produce a close match to the patient's data (the speed of activation flow was halved).

Thus, the different degrees in the severity of the syndrome may be modeled by varying the lesion severity to the lexical route: This would give rise to different patterns of impairment in exception word reading (and thus different rates of regularization errors). This is not the case for single-route, hidden-unit models, however, as demonstrated by the repeated failure to model surface dyslexia within a single pathway from spelling to sound. As we see it, the problem with both the S&M and PMSP models is that the lexical and sublexical levels are highly interwoven and are required to use the same computational resources (a single, hidden, mediated pathway). The S&M model appears to operate on a largely lexical basis across the board, in the sense that each learned word generates a unique hidden-unit response in the network. Thus, contrary to the data, the lesioned S&M model was also impaired on regular word reading and showed a further decline in its already poor nonword reading (Patterson, Seidenberg, & McClelland, 1989). Similar problems also afflict the PMSP orthography-to-phonology network: When the model is damaged to the extent required to generate a performance on exception word reading comparable to that of K.T., the model's performance falls dramatically on all dimensions. It gets worse for both high-frequency and low-frequency regular words (50–55% correct) and also for nonwords (49%). At the same time, the regularization rate drops to a low level (18%; Plaut et al., 1996). The nonword reading performance of the damaged model is particularly disappointing given that the reported performance is obtained on the relatively simple Glushko (1979) nonwords, on which K.T. attained 100% correct. Overall, the behavior of the damaged PMSP network is not very different from that of the damaged S&M model and may be ascribed to lack of a pathway dedicated to the regular mapping.

To accommodate this problem, Plaut et al. (1996) developed a dual-route explanation of surface dyslexia, which is based on the interaction of the orthography-to-phonology pathway with the semantic pathway (the latter was acknowledged, but not implemented, in the broader S&M framework). This nontraditional account of surface dyslexia is based on the hypothesis of Patterson and Hodges (1992) that the integrity of phonological representations is dependent on an interaction between meaning and phonology. In one simulation of Plaut et al., the orthography-to-phonology network is trained with an additional external input to the phoneme units that would model the contribution of a (putative) semantic pathway; furthermore, the magnitude of this external input increases in the course of training to simulate an increased competence of the putative semantic pathway. Plaut et al. showed that in this version of the model the orthography-to-phonology network is relieved from mastering low-frequency exception words because the correct output for these words is provided by the semantic pathway. They argued that as the semantic pathway's competence improves, the phonological pathway becomes specialized for regular words; this results in a redistribution of labor between the semantic and the phonological pathways. Therefore, when a lesion completely eliminates the contribution of the semantic pathway, the severity of the dyslexic pattern exhibited by the orthography-to-phonology network is a function of the redistribution of labor between the semantic and the phonological pathways that occurred prior to the lesion.

Therefore, differences among patients reflect not only different severities of lesions, but also and in particular their different premorbid reading competences, that is, how much their phonological pathways "deteriorated" before the brain damage. Needless to say, this new interpretation deserves a careful neuropsychological investigation. In support of the hypothesis that correct exception word reading is dependent on semantic representations, Patterson and Hodges (1992)

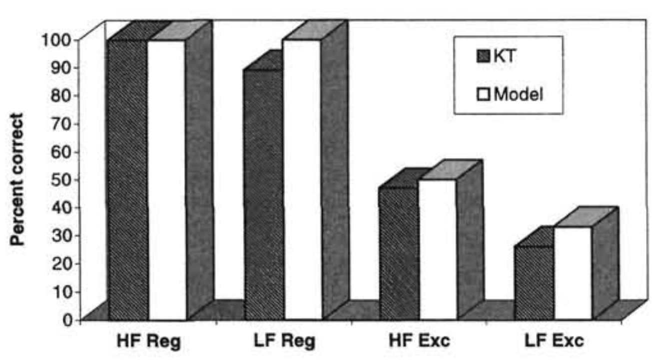

Figure 12. Correct performance of patient K.T. (McCarthy & Warrington, 1986) and lesioned model on word lists from Taraban and McClelland (1987). HF = high frequency; LF = low frequency; Reg = regular words; Exc = exception words.

reported the reading performance of six patients suffering from semantic dementia: All patients presented with a surface dyslexia. However, a patient reported on by Schwartz, Saffran, and Marin (1980) was able to read exception words that she could not understand. In addition, the strongest evidence in support of the independence of semantic and phonological processing comes from a recent study by Cipolotti and Warrington (1995). Their patient D.R.N., though presenting with a semantic dementia, had a strikingly well-preserved ability to read both regular and exception words that he failed to comprehend. Three other patients with the same characteristics were later described (Lambon-Ralph, Ellis, & Franklin, 1995; Cipolotti & Warrington, 1996). This pattern, known as lexical nonsemantic reading, is accounted for in the dual-route model by the existence of the lexical nonsemantic route, which can still be intact in the case of damage to the semantic system (see Coltheart et al., 1996, for further discussion). However, Plaut (1997) argued that the different effects of a semantic deficit can be accommodated within an individual differences account, in which, for a given individual (i.e., network), the reliance on the semantic pathway for reading exception words depends on a combination of two parameters of the model (i.e., semantic strength and weight decay).

With regard to the issue of a link between semantic dementia and surface dyslexia, the strong (but not necessary) correlation between the two syndromes might be attributed to the anatomical proximity of the brain areas involved in semantic and phonological processing (Cipolotti & Warrington, 1995). In our simulations with the three-layer network, the lexical mappings depend on the same order of computational resources required for semantic processing. It is the presence of a layer of hidden units that allows the system to establish arbitrary (rather than linear) mappings. Considering that semantic dementia is affecting just those kind of arbitrary, higher order computations, it is plausible that the lexical pathway (based on the mediated mappings) would be in most (but not necessarily all) cases affected in the same way as semantics. Note that this explanation is based on computational, rather than anatomical, considerations. The idea that the higher orders of computation are more sensitive to brain damage is not new. Shallice and colleagues (Shallice et al., 1983; Shallice & McCarthy, 1985) attempted to explain surface dyslexia within their multiple-level model of reading by proposing that the higher levels (e.g., the lexical level) are more susceptible to neurological damage. This is surely most likely to occur in the degenerative forms of damage, therefore explaining the correlation of semantic dementia and surface dyslexia. However, it remains difficult to explain the correlation across items in comprehension and naming of exception words that has been observed in some patients (Graham, Hodges, & Patterson, 1994). In this regard, a recent proposal of Funnell (1996) is that in surface dyslexic patients, a lexical response is preferred over a sublexical, regularized response only when vestiges of word meaning remain. Patient E.P., who presented with semantic dementia (and surface dyslexia), showed knowledge of both lexically derived and sublexically derived pronunciation in reading exception words (e.g., presented with GLOVE she responded "/gl@Uv/ or /glVv/?"; see Funnell, 1996). When all meaning of a given word was lost, the patient selected a sublexical response, even when she had also produced the correct lexical response. Clearly, further work is necessary to simulate more details of the surface dyslexic syndrome.

General Discussion

Reading English words aloud, as a mapping from orthography to phonology, might be described as a quasi-regular task in the sense that these input–output mappings are in most cases systematic (regular words) but in many cases somewhat arbitrary (exception words). The TLA model clearly shows that a simple two-layer network, equipped with psychologically motivated input–output representations, is able to extract the statistical regularities of a corpus of English monosyllabic words. The TLA model implements a sublexical assembly procedure in which the phonology of any letter string (word or nonword) is computed according to the most common spelling–sound relationships. Words with irregular pronunciations are regularized in a way that corresponds to the behavior of surface dyslexic patients. Nonword reading performance of the TLA model is equal to that of the PMSP model (Plaut et al., 1996), but the TLA model reads nonwords with very little training and without the need for grapheme units in the orthographic representation. Following the success of our TLA model in achieving good nonword reading, some work was done on using a similar architecture for nonword spelling (Glasspool, Houghton, & Shallice, 1995; Shallice, Glasspool, & Houghton, 1995).

The results obtained with the TLA model lead us to reject Coltheart et al.'s (1993) claim that the nonlexical route must be based on a system of linguistic rules. Even though, as previously discussed, the network connections in our model can be considered as implementing (implicit) mapping rules, this interpretation is still in contrast with Coltheart et al.'s conceptualization of the GPC route, which is simulated as a system of (explicit) production rules. In addition, the production of assembled phonology is parallel in our TLA model, whereas it is strictly serial in the GPC route. Finally, the output of the TLA model is sensitive to the relative consistency of the spelling–sound mappings: This results in the activation of multiple phoneme candidates for the same phoneme position, at the vowel position in particular. This pattern closely resembles other proposals about the nature of the assembly procedure, particularly the ideas of "flexible mapping rules" (P. Brown & Besner, 1987) and "islands of reliability" (Carr & Pollatsek, 1985). Strong arguments in favor of a fast and parallel assembly process can also be found in recent work by Berent and Perfetti (1995).

With the second, augmented, model, we presented computational evidence that learning to read demands task decomposition: During learning, the network tends to self-modularize by allocating different resources to perform the two complementary subtasks of regular and exception word reading. As in the first model, the regular task is performed by the direct, unmediated route from spelling to sound. The exception task requires the formation of higher order intermediate representations, which the provision of hidden units permits. If the computation is forced to be completely mediated by the hidden layer pathway, then the model acquires lexical properties, as we believe the S&M model also demonstrates. Of course, Seidenberg and McClelland (1990) strongly reaffirmed their belief that their model contains no lexicon. However, from our investigations, we propose that its behavior is in fact best understood by seeing it as implementing a distributed lexicon, as has been argued by Besner et al. (1990) and Monsell (1991). According to Monsell (1991),

> [A]lthough Seidenberg & McClelland emphasize with polemical fervour that their network does not, after learning, contain any representation of words, this is a touch too strong in that *the network must in some sense be "identifying" whole words in order to transcode successfully spelling patterns with very exceptional spelling-sound correspondences* [italics added]. (p. 162)

In our final section we proposed a connectionist dual-process architecture of the reading system in which lexical, case-specific knowledge about learned words and sublexical, generative knowledge about spelling–sound relationships interact for the formation of a final phonological code to drive articulation. The TLA model is the implementation of the assembly component, whereas the lexical component can be conceptualized as an interactive activation network such as that implemented by Coltheart and Rastle (1994), or alternatively, by any network that develops mediated, internal representations for the known words. As previously noted, however, the additional advantage of a (localist) interactive activation model of the lexical route is that visual word recognition (e.g., perceptual identification and the lexical decision task) can be readily simulated (Grainger & Jacobs, 1996). Clearly, with regard to word recognition and retrieval of lexical phonology, any advantage of distributed over local representations has yet to be demonstrated. The point of interaction between the two procedures is realized in our model in the PDS, where the on-line competitive interaction between the output of the two processes results in the observed interactions of frequency and regularity–consistency and in the lexicality effect. Finally, we showed how a lesion to the lexical component of the dual-process model can easily provide a close match to the dramatic surface dyslexic pattern of patient K. T. (McCarthy & Warrington, 1986). Such a pattern has proven difficult (or perhaps impossible) to model within a single-route connectionist system (see Patterson et al., 1989; Plaut et al., 1996) and has led Plaut et al. (1996) to adopt a "dual-route" perspective (i.e., one that assumes active interactions between phonological and semantic routes) and to advocate nontraditional explanations (e.g., redistribution of labor and individual differences).

Computational Models: Similarities and Differences

Modeling reading aloud is, we believe, a special case in the growing field of computational cognitive modeling. The S&M model is indeed one of the first and most successful computational models of cognitive functions: This is testified to by the huge number of citations and by the major debate in the reading literature generated by the model. The theoretical side of reading research has been given substantial impetus by the further development of different (and alternative) computational models of single-word reading. Although in the past, the success of a computational model could be judged by simple "proof of simulation," this has now quite changed. A good model should account for data coming from experimental studies on normal individuals (both in terms of RTs and error rates), from studies on patients with neuropsychological disorders, and ultimately from developmental studies. However, the strengths of a model must be judged not only in terms of the number of facts it can account for but also in terms of the economy and relative "transparency" of the model: If the model's internal operations are difficult to interpret, little insight can be gained from it (see McCloskey, 1991, for discussion). A further important point regards the novel or differential predictions models may provide. Here we outline the major similarities and differences between the models to see if different predictions can be derived from them.

With respect to the PMSP model, it is worth pointing out that the way in which we simulated the contribution of retrieved phonology (i.e., the lexical route) is similar to the way in which Plaut et al. (1996) simulated the contribution of a "semantic" pathway. In both cases, what it really amounts to is a further (frequency-weighted) input to phonology. As argued by Bernstein and Carr (1996), there is nothing compellingly "semantic" in Plaut et al.'s simulation: Rather, these properties can be ascribed to a lexical nonsemantic route. One limitation of the simulation of Plaut et al. is that no assumptions are being made about the time course of the semantically driven activation of phonology; that is, the input from the two routes to the phonological output begins simultaneously, and this seems rather implausible. Given that the phonological and the semantic pathways are thought to operate in parallel, the phonological pathway should be faster than the semantic pathway, because the latter contains two additional layers of nodes that must be traversed (i.e., the semantic layer and a further layer of hidden units). In our model, however, even the assumption of faster assembly (contrary to the standard dual-route model) produces the standard empirical effects, such as the interaction between frequency and regularity.

If we turn to the DRC model of Coltheart and colleagues (Coltheart et al., 1993; Coltheart & Rastle, 1994), the main differences between it and our model concern the nature of the assembly procedure and the on-line interactions between the two routes. The GPC route is held to be slow and serial, and therefore interaction (e.g., conflict) between the two routes is relegated "to the status of a minor nuisance that mainly plagues laboratory tasks" (Carr & Pollatsek, 1985, p. 9). In this respect, it can be noted that all models have converged to a similar assumption; that is, the interaction between two different sources of phonological information is necessary to explain both behavioral and neuropsychological data. In the DRC model this interaction is still best characterized as a "horse race," which resembles the assumption of the original dual-route "horse race" model (Coltheart, 1978; Paap & Noel, 1991). In the PMSP model and in the present model, however, the interaction is a sort of pooling process, as proposed in alternative versions of the dual-route model (the "pooling" model of Monsell et al., 1992, and the "summation" model of Hillis & Caramazza, 1991). More important, the GPC route is not sensitive to consistency and delivers a single phoneme for each position. Our model, in contrast, suggests that processing in the assembly route (of monosyllabic words) can be fast and parallel: This is highly consistent with recent proposals (and experimental evidence) about the nature of the assembly process (see, e.g., Berent & Perfetti, 1995). In both the DRC model and ours, word frequency has no role in the assembly procedure. However, consistency effects in our model arise within the assembly component for both words and nonwords, whereas in Coltheart et al.'s (1993) model they are supposed to arise from the interaction between lexical and assembly routes. This idea is somewhat odd—at least in the case of consistency effects for nonwords, because they should depend on subthreshold activations in the lexical route in processing nonword stimuli—and has not been simulated to our knowledge. In the S&M framework, the role of the implemented spelling-to-sound network (i.e., the PMSP model) is indeed that of an assembly procedure, which can be used on any string of letters (although it can produce the correct pronunciation of any known word, including exceptions). Processing in the PMSP model is, like in ours, parallel. However, the PMSP model is also sensitive to word frequency, and consistency effects are tightly linked with frequency effects.

Dissociations or changes in the relative contributions between the two routes have been sought both in neuropsychological and neuroimaging studies (see Petersen & Fiez, 1993, and Posner & Carr, 1992, for reviews) and in experimental studies (using manipulations that create priming or interference effects, e.g., Monsell et al., 1992; Paap & Noel, 1991). For instance, a striking dissociation between frequency and regularity effects has been reported by Balota and Ferraro (1993) across groups of participants ranging from young adults to older adults and to patients with senile dementia of the Alzheimer's type. This finding is clearly difficult to reconcile with the PMSP model, in which the two variables are truly interdependent (in addition, note that consistency and regularity are confounded in the PMSP model). This is not a problem for either the DRC model or for ours, because the two effects lie in different components of the models. On the experimental side, a way to dissociate the two routes involves the creation of interference that is thought to debilitate one route more than the other: this is achieved with a concurrent task. For instance, a concurrent task involving phonological short-term memory (e.g., a digit memory load) seems to interfere with assembly, and in some experiments has resulted in a release-from-competition effect, that is, a relative improvement in naming low-frequency exception words (e.g., Bernstein & Carr, 1996; Herdman & Beckett, 1996; Paap & Noel, 1991). However, note that the existence of the release-from-competition effect is highly controversial, and some researchers have argued that it applies only to a subset of readers (Bernstein & Carr, 1996) or that it is not a real phenomenon altogether (see Bernstein, DeShon, & Carr, 1998; Pexman & Lupker, 1995, 1998). The opposite pattern—that is, an increased competition coming from assembly for low-frequency exceptions—was reported by Herdman and Beckett (1996) with a visual short-term memory load, which they argued interferes with the lexical route.

The important question at this point regards a possible dissociation between consistency and regularity. In the PMSP model, for instance, regularity and consistency are confounded, so that the two effects cannot be dissociated (moreover, Plaut et al. [1996] denied that regularity is different from consistency). In Coltheart et al.'s (1993; Coltheart & Rastle, 1994) DRC model, consistency and regularity are different effects, but both arise from the interaction between the two routes. In our dual-process model, by contrast, consistency effects arise within the assembly process, whereas regularity effects arise from the interaction between the two processes. Therefore, different predictions can be derived from the models: Our model can predict a clear dissociation between consistency and regularity.

On the Role of Strategic Control

As previously discussed, recent studies demonstrate that participants' performance in oral reading is sensitive to the context in which the stimuli occur. For instance, the presence of a word frequency effect is dependent on the presence or absence of nonword stimuli (Baluch & Besner, 1991; Frost et al., 1987; Wydell & Humphreys, in press). These results have been taken as strong evidence that participants can strategically rely on the outputs of the assembled (i.e., sublexical) or the addressed (i.e., lexical) routines, depending on the experimental material (Besner & Smith, 1992). For instance, the presence of nonwords encourages participants to use the assembly process, the result of which is that lexicality effects disappear. The important critique of Besner and his colleagues (Besner, in press; Besner & Smith, 1992; Besner et al., 1990) is that connectionist models such as those of Seidenberg and McClelland (1989) and Van Orden et al. (1990) are discomfited by these data, because they cannot have a route that is insensitive to word frequency. The same argument appears to be true for the Plaut et al. (1996) model (see Besner, in press, for further discussion). However, the critique does not hold for our dual-process model because it incorporates a distinct nonlexical process (the TLA model) that is not sensitive to word frequency. This is clearly demonstrated by the fact that consistent nonwords (i.e., words with a frequency of zero) yield the same naming latency as consistent words that have been presented hundreds of times to the TLA model during training. In the same vein, Monsell et al. (1992) found that participants deemphasize the assembly procedure when they are not prepared for nonwords. For instance, in a pure block of exception words participants name exception words faster and make fewer regularization errors. This finding suggests that strategic control is based on an inhibitory, rather than an excitatory, device, which is responsible for the selection of the contextually appropriate translation process by deemphasizing the other process. It is very difficult to see how this kind of flexible, strategic, modulation of processing could be implemented in a single-route model, because it appears to be impossible to isolate, functionally or structurally, those components of single-route models responsible for different aspects of the task. If this is so, how can one subprocess be selected over another? Our dual-process model, in contrast, can be readily used to investigate the effects of strategic modulation by changing the speed of processing in the two routes (e.g., by changing the ramp parameter).

More recently, Lupker, Brown, and Colombo (1997) provided new evidence on strategic effects and suggested that these depend on a "criterion-setting effect" on the time to start articulation, rather than on route selection or deemphasis. In this account, the effect of blocking a list of homogeneous stimuli results in the setting of the most appropriate time criterion for that list, which depends on the relative difficulty of the stimuli (e.g., exceptions vs. regular words) or, in other words, on the strength of the mappings. In a mixed list, however, the criterion is set at a position that is intermediate to the positions used for the two kinds of stimuli, so that RTs tend to homogenize (i.e., fast stimuli slow down and slow stimuli speed up). In our dual-process model, the criterion to start articulation is provided by the response threshold in the PDS. The time to reach response threshold, however, is a function of the strength of the incoming phonological codes. Therefore, the phenomena described by Lupker et al. (1997) can be explored with the dual-process model by manipulating the response threshold in the PDS by setting the threshold value at different (and most appropriate) levels in the case of blocked, homogeneous lists.

Developmental Dyslexia

Castles and Coltheart (1993), in a large study on dyslexic children, provided the clearest evidence that surface and phonological dyslexia have a developmental counterpart (also see Manis et al., 1996). They found a group of children who were selectively impaired at exception word reading and a second group of children selectively impaired at nonword reading. Thus, a computational model of reading should give some insights into the failure to acquire fully competent reading skills. Our study demonstrates that the regular task requires less computational power. This fact is interesting from a developmental perspective. As previously discussed, children are able to read new words soon after starting reading. They start to use a whole-word strategy only in a successive reading stage. This stage change might involve the development of more powerful computational resources. Interestingly, Hynd and Semrud-Clikeman (1989) reported neuroimaging evidence that some forms of dyslexia are related to a failure in allocating neural resources to parts of the reading task. This suggests that the core of the surface dyslexia syndrome might be explained in terms of insufficient (neural) computational resources for the exception task. These resources might be simply unavailable in the developmental form due to "retarded maturation"; to brain damage in the acquired form. One might therefore conclude that the computational resources required for exception word reading are more powerful than those required for regular mapping. This need for a higher order of computational power (i.e., the hidden unit representations) is just what one of our simulations shows. Indeed, the surface dyslexic performance of the pure TLA model may be analogous to the developmental form of the syndrome because it results from training without hidden-units resources.

An interesting possibility for modeling the stage change in normal reading might be to use constructive algorithms, such as cascade-correlation (Fahlman & Labiere, 1990). This learning procedure starts with a two-layer network and allocates new computational resources (i.e., hidden units) when no further improvement of the performance is achieved. Shultz (1991) suggested that cascade-correlation is particularly suitable for modeling stage changes in child development.

Developmental phonological dyslexia, on the other hand, might reflect the fact that the development of a sublexical process requires earlier basic segmentation skills, that is, the ability to subdivide a word into some basic components. Participant R.E. (Campbell & Butterworth, 1985), for instance, had associated deficits of phonemic processing and phonemic awareness: He was unable to make the initial segmentation of a string of letters (word or nonword). This contention is supported by comparing the performance of the existing computational models. All the models capable of good nonword reading assume some segmentation of the orthographic code intermediate between the single-letter and whole-word levels. In the case of our model, the word components are onset and rime; in Plaut et al.'s (1996) model they are onset, vowel, and coda (in addition to grapheme units); in Coltheart et al.'s (1993) rule system they are the whole set of graphemes. The Seidenberg and McClelland (1989) model, on the other hand, does not employ such a segmentation and shows poor nonword reading (Besner et al., 1990). Because this behavior is not a consequence of damage to the network, it can be considered a form of developmental phonological dyslexia. We discuss these issues further, with simulations of developmental data, in Zorzi et al. (1998). In particular, we show that if the orthographic and phonological representations cannot make direct contact (i.e., learning is forced to occur through the hidden-units pathway), the rapid acquisition of the ability to generalize to nonwords is severely impaired (Zorzi et al., 1998).

Implications for Other Languages

The degree of regularity in the computation from print to sound has an immediate impact on the possible role of the direct input–output mapping (i.e., the TLA model) in performing the task and also on the necessity of a lexically mediated mapping. If the direct (assembly) pathway and the mediated (lexical) pathway were the basic starting architecture of all phonological mechanisms, the role of the mediated pathway would depend on the orthographic transparency of the language that the model is trained on. Hence, an opaque orthography such as Japanese Kanji (see, e.g., Wydell, Butterworth, & Patterson, 1995) would mostly rely on the mediated pathway (although sublexical, character-level phonological information might still be necessary; see Wydell, Butterworth, Shibahara, & Zorzi, 1998, for discussion), whereas in completely shallow orthographies such as Japanese Kana, the direct pathway might be allocated all the work. In other cases of shallow orthographies, the mediated pathway would be necessary at least for resolving minor inconsistencies, such as the stress assignment in Italian. The issue of the effect of orthographic transparency on the development of the reading mechanism is worthy of more detailed investigation.

Conclusions

The internal operations of the module involved in the ability to deal with novel stimuli, that is, to read new words and nonwords, have been the subject of a considerable theoretical debate (e.g., Carr & Pollatsek, 1985). More recently, the development of computational models of reading had the effect of enhancing the contrast between theories postulating one versus two routes. On the one hand, Seidenberg, McClelland, and colleagues (Seidenberg & McClelland, 1989; Seidenberg et al., 1994; Plaut et al., 1996) argued against the notion of specific and distinct sublexical and lexical processes. On the other hand, Coltheart et al. (1993; Coltheart & Rastle, 1994) claimed that the sublexical process is implemented by a specific mechanism operating on the basis of a set of rules, while they maintained a neural-like approach to the simulation of the lexical route.

With the TLA model, we have presented computational evidence that a sublexical assembly process can be implemented in a connectionist network. Therefore, in complete agreement with Seidenberg et al. (1989, 1994; Plaut et al., 1996), we reject Coltheart et al.'s (1993) claim of the need for a system of explicit GPC rules to explain our ability to deal with nonwords. However, we also demonstrated that a network-based sublexical process can exist even in the absence of a lexical counterpart, in contrast to the assumptions of any single-route theory (or model). In addition, we have presented a "dual-process" connectionist model in which the interactions between different sources of phonological information—assembled phonology and retrieved phonology—can account for the standard experimental effects in the oral reading of single words. The model provides insight into a number of important theoretical issues: the architecture of the phonological mechanism, the effects of consistency, the impaired reading performance of acquired and developmental dyslexia patients, and the role of strategic control. In addition, we believe that the relative transparency of the model sheds light on how previous network models of reading have accomplished the task and highlights the origins of their strengths and weaknesses.

Much work remains to be done with the model. For instance, the various dyslexic syndromes need more detailed simulation, and the role of strategic control in modulating the contributions of each process warrants investigation. Nonetheless, our findings reaffirm the basic assumption of any two-process theory—that is, that the pronunciations of nonwords and exception words are computed by different processes.

References

Balota, D. A., & Ferraro, R. F. (1993). A dissociation of frequency and regularity effects in pronunciation performance across young adults, older adults, and individuals with senile dementia of the Alzheimer type. *Journal of Memory and Language, 32,* 573–592.

Baluch, B., & Besner, D. (1991). Visual word recognition: Evidence for strategic control of lexical and nonlexical routines in oral reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 644–652.

Baron, J., & Strawson, C. (1976). Use of orthographic and word-specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 386–392.

Behrmann, M., & Bub, D. (1992). Surface dylexia and dysgraphia: Dual routes, single lexicon. *Cognitive Neuropsychology, 9,* 209–251.

Berent, I., & Perfetti, C. A. (1995). A rose is a reez: The two-cycles model of phonology assembly in reading English. *Psychological Review, 102,* 146–184.

Bernstein, S. E., & Carr, T. H. (1996). Dual-route theories of pronouncing printed words: What can be learned from concurrent task performance? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 86–116.

Bernstein, S. E., DeShon, R. P., & Carr, T. H. (1998). Concurrent task demands and individual differences in reading: How to discriminate artifacts from real McCoys. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 822–844.

Besner, D. (in press). Basic processes in reading: Multiple routines in localist and connectionist models. In R. M. Klein & P. A. McMullen (Eds.), *Converging method for understanding reading and dyslexia.* Cambridge, MA: MIT Press.

Besner, D., & Smith, M. C. (1992). Basic processes in reading: Is the orthographic depth hypothesis sinking? In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 45–66). Amsterdam: Elsevier.

Besner, D., Twilley, L., McCann, R. S., & Seergobin, K. (1990). On the connection between connectionism and data: Are a few words necessary? *Psychological Review, 97,* 432–446.

Bradley, L., & Bryant, P. E. (1983). Categorising sounds and learning to read: A causal connection. *Nature, 301,* 419–521.

Brown, G. D. A. (1987). Resolving inconsistencies: A computational model of word naming. *Journal of Memory and Language, 26,* 1–23.

Brown, P., & Besner, D. (1987). The assembly of phonology in oral reading: A new model. In M. Coltheart (Ed.), *Attention and performance XII* (pp. 471–489). Hillsdale, NJ: Erlbaum.

Brown, P., Lupker, S. J., & Colombo, L. (1994). Interacting sources of information in word naming: A study of individual differences. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 537–554.

Bruce, D. J. (1964). The analysis of word sounds. *British Journal of Educational Psychology, 34,* 158–170.

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.

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 Section A: Human Experimental Psychology, 37A,* 435–475.

Carr, T. H., & Pollatsek, A. (1985). Recognizing printed words: A look at current models. In D. Besner, T. G. Waller, & G. E. MacKinnon (Eds.), *Reading research: Advances in theory and practice* (Vol. 5, pp. 1–82). San Diego, CA: Academic Press.

Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition, 47,* 148–180.

Cipolotti, L., & Warrington, E. K. (1995). Semantic memory and reading abilities: A case report. *Journal of the International Neuropsychological Society, 1,* 104–110.

Cipolotti, L., & Warrington, E. K. (1996, January). *Deterioration of semantic knowledge in surface dyslexia.* Paper presented at the 14th European Meeting for Cognitive Neuropsychology, Bressanone, Italy.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97,* 523–547.

Cohen, J. D., & Huston, T. A. (1994). Progress in the use of interactive models for understanding attention and performance. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV* (pp. 453–476). Cambridge, MA: MIT Press.

Cohen, J. D., Servan-Schreiber, D., & McClelland, J. L. (1992). A parallel distributed processing approach to automaticity. *American Journal of Psychology, 105,* 239–269.

Colombo, L., Cubelli, R., Zorzi, M., & Caporali, A. (1996, November). *Consonant and vowel information in the phonological assembly of Italian words.* Paper presented at the 33rd Annual Meeting of the Psychonomic Society, Chicago, IL.

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. (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., 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., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and Performance IV* (pp. 535–555). Hillsdale, NJ: Erlbaum.

Coltheart, M., Langdon, R., & Haller, M. (1996). Computational cognitive neuropsychology and acquired dyslexia. In B. Dodd, L. Worrall, & R. Campbell (Eds.), *Models of language: Illuminations from impairment* (pp. 9–36). London: Whurr.

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.

Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

Denes, F., Cipolotti, L., & Zorzi, M. (1996). Dislessie e disgrafie acquisite [Acquired dyslexia and dysgraphia]. In L. Pizzamiglio & G. Denes (Eds.), *Manuale di neuropsicologia* (pp. 386–422). Bologna, Italy: Zanichelli.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1997). *Rethinking innateness: A connectionist perspective on development.* Cambridge, MA: MIT Press.

Fahlman, S. E., & Labiere, C. (1990). The cascade-correlation learning architecture (Tech. Rep. No. CMU-CS-90-100). Pittsburgh, PA: Carnegie-Mellon University, School of Computer Science.

Fera, P., & Besner, D. (1992). The process of lexical decision: More words about a parallel distributed processing model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 749–764.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographic depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 104–115.

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

Funnell, E. (1996). Response biases in oral reading: An account of the co-occurrence of surface dyslexia and semantic dementia. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 49A,* 417–446.

Ghahramani, Z., & Wolpert, D. M. (1997). Modular decomposition in visuomotor learning. *Nature, 386,* 392–395.

Glasspool, D. W., Houghton, G., & Shallice, T. (1995). Interactions between knowledge sources in a dual-route connectionist model of spelling. In L. S. Smith & P. J. B. Hancock (Eds.), *Neural computation and psychology* (pp. 209–226). London: Springer-Verlag.

Gluck, M. A., & Bower, G. H. (1988a). Evaluating an adaptive network model of human learning. *Journal of Memory and Language, 27,* 166–195.

Gluck, M. A., & Bower, G. H. (1988b). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117,* 227–247.

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

Goswami, U. (1986). Children's use of analogy in learning to read: A developmental study. *Journal of Experimental Child Psychology, 42,* 73–83.

Goswami, U. (1988). Orthographic analogies and reading development. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 40A,* 239–268.

Goswami, U. (1991). Learning about spelling sequences in reading: The role of onsets and rimes. *Child Development, 62,* 1110–1123.

Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read.* London: Erlbaum.

Graham, K. S., Hodges, J. R., & Patterson, K. (1994). The relationship between comprehension and oral reading in progressive fluent aphasia. *Neuropsychologia, 32,* 299–316.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103,* 518–565.

Hartley, T., & Houghton, G. (1996). A linguistically-constrained model of short-term memory for nonwords. *Journal of Memory and Language, 35,* 1–31.

Henderson, L. (1985). Issues in the modelling of pronunciation assembly in normal reading. In K. E. Patterson, J. C. Marshall, & M. Coltheart (Eds.), *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (pp. 453–508). London: Erlbaum.

Herdman, C. M., & Beckett, B. L. (1996). Code-specific processes in word naming: Evidence supporting a dual-route model of word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1149–1165.

Hillis, A. E., & Caramazza, A. C. (1991). Mechanisms for accessing lexical representations for output: Evidence from a category-specific semantic deficit. *Brain and Language, 40,* 106–144.

Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review, 98,* 74–95.

Houghton, G., & Tipper, S. P. (1994). A model of inhibitory mechanisms in selective attention. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 53–112). New York: Academic Press.

Humphreys, G. W., & Evett, L. J. (1985). Are there independent lexical and nonlexical routes in word processing? An evaluation of the dual-route theory of reading. *Behavioral and Brain Sciences, 8,* 689–740.

Hynd, G. W., & Semrud-Clikeman, M. (1989). Dyslexia and brain morphology. *Psychological Bulletin, 106,* 447–482.

Jacobs, R. A., Jordan, M. I., & Barto, A. G. (1991). Task decomposition through competition in a modular connectionist architecture: The what and where vision tasks. *Cognitive Science, 15,* 219–250.

Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language, 29,* 687–715.

Kawamoto, A. H., & Zemblidge, J. (1992). Pronunciation of homographs. *Journal of Memory and Language, 31,* 349–374.

Kay, J. (1985). Mechanisms of oral reading: A critical appraisal of cognitive models. In A. W. Ellis (Ed.), *Progress in the psychology of language* (Vol. 2, pp. 73–105). London: Erlbaum.

Kay, J. (1987). Phonological codes in reading: Assignment of sub-word phonology. In A. Allport, D. G. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production* (pp. 181–197). London: Academic Press.

Kay, J., & Bishop, D. (1987). Anatomical differences between nose, palm, and foot, or, the body in question: Further dissection of the processes of sub-lexical spelling-sound translation. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 449–463). Hillsdale, NJ: Erlbaum.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Lacouture, Y. (1989). From mean squared error to reaction time: A connectionist model of word recognition. In D. S. Touretzky, G. E. Hinton, T. J. Sejnowsky (Eds.), *Proceedings of the 1988 Connectionist Models Summer School* (pp. 371–378). San Mateo, CA: Morgan Kaufmann.

Lambon-Ralph, M., Ellis, A. W., & Franklin, S. (1995). Semantic loss without surface dyslexia. *Neurocase, 1,* 363–369.

Laxon, V., Masterson, J., & Moran, R. (1994). Are children's representations of words distributed? Effects of orthographic neighbourhood size, consistency and regularity of naming. *Language and Cognitive Processes, 9,* 1–27.

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

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

Manis, F. R., Seidenberg, M. S., Doi, L. M., McBride-Chang, C., & Petersen, A. (1996). On the bases of two subtypes of developmental dyslexia. *Cognition, 58,* 157–195.

McCann, R. S., & Besner, D. (1987). Reading pseudohomophones: Implications for models of pronunciation and the locus of the word-frequency effect in word 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., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88,* 375–407.

McClelland, J. L., & Rumelhart, D. E. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 2: Psychological and biological models.* Cambridge, MA: MIT Press.

McCloskey, M. (1991). Networks and theories: The place of connectionism in cognitive science. *Psychological Science, 2*(6), 387–395.

Minsky, M., & Papert, S. (1969). *Perceptrons.* Cambridge, MA: MIT Press.

Monsell, S. (1991). The nature and locus of word frequency effects in reading. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 148–197). Hillsdale, NJ: Erlbaum.

Monsell, S., Patterson, K. E., Graham, A., Hughes, C. H., & Milroy, R. (1992). Lexical and sublexical translation of spelling to sound: Strategic anticipation of lexical status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 452–467.

Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition, 24,* 45–64.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phonemes arise spontaneously? *Cognition, 7,* 323–331.

Morton, J., & Patterson, K. E. (1980). A new attempt at an 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 & Kegan Paul.

Norris, D. (1994). A quantitative, multiple levels model of reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1212–1232.

Paap, K. R., & Noel, R. W. (1991). Dual route models of print to sound: Still a good horse race. *Psychological Research, 53,* 13–24.

Parkin, A. J. (1983, July). *Two processes not one in reading aloud.* Paper presented at the meeting of the Experimental Psychology Society, Oxford, England.

Patterson, K. E., & Coltheart, V. (1987). Phonological processes in reading: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 421–447). Hillsdale, NJ: Erlbaum.

Patterson, K. E., Graham, H., & Hodges, J. R. (1994). Reading in Alzheimer's type dementia: A preserved ability? *Neuropsychologia, 8,* 395–412.

Patterson, K. E., & Hodges, J. R. (1992). Deterioration of word meaning: Implications for reading. *Neuropsychologia, 12,* 1025–1040.

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: Neuropsychological and cognitive studies of phonological reading* (pp. 335–359). London: Erlbaum.

Patterson, K. E., Seidenberg, M. S., & McClelland, J. L. (1989). Connections and disconnections: Acquired dyslexia in a computational model of reading processes. In R. G. M. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neurobiology* (pp. 131–181). London: Oxford University Press.

Petersen, S. E., & Fiez, J. A. (1993). The processing of single words studied with positron emission topography. *Annual Review of Neuroscience, 16,* 509–530.

Pexman, P. M., & Lupker, S. J. (1995). Effects of memory load in a word naming task: Five failures to replicate. *Memory & Cognition, 23,* 581–595.

Pexman, P. M., & Lupker, S. J. (1998). Word naming and memory load: Still searching for an individual differences explanation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 803–821.

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes, 12,* 767–808.

Plaut, D. C., Behrmann, M., Patterson, K. E., & McClelland, J. L. (1993, November). *Impaired oral reading in surface dyslexia: Detailed comparison of a patient and a connectionist model.* Paper presented at the 34th Annual Meeting of the Psychonomic Society, Washington, DC.

Plaut, D. C., & McClelland, J. L. (1993). Generalization with componential attractors: Word and nonword reading in an attractor network. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (pp. 824–829). Hillsdale, NJ: Erlbaum.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103,* 56–115.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10,* 377–500.

Posner, M. I., & Carr, T. C. (1992). Lexical access and the brain: Anatomical constraints on cognitive models of word recognition. *American Journal of Psychology, 105,* 1–26.

Quinlan, P. T. (1993). *The Oxford psycholinguistic database.* London: Oxford University Press.

Rueckl, J. G., Cave, K. R., & Kosslyn, S. M. (1989). Why are "what" and "where" processed by separate cortical visual systems? A computational investigation. *Journal of Cognitive Neuroscience, 1,* 171–186.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations* (pp. 318–362). Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effects and some tests and extensions of the model. *Psychological Review, 89,* 60–94.

Schmajuk, N. A., & DiCarlo, J. J. (1992). Stimulus configuration, classical conditioning, and hippocampal functioning. *Psychological Review, 98,* 268–305.

Schwartz, M. F., Saffran, E. M., & Marin, O. S. M. (1980). Fractionating the reading process in dementia: Evidence for word-specific print-to-sound associations. In M. Coltheart, K. E. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 253–269). London: Routledge & Kegan Paul.

Seidenberg, M. S. (1992). Beyond orthographic depth in reading: Equitable division of labor. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 85–118). Amsterdam: Elsevier.

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

Seidenberg, M. S., & McClelland, J. L. (1990). More words but still no lexicon: Reply to Besner et al. (1990). *Psychological Review, 97,* 447–452.

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

Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behaviour, 23,* 383–404.

Sejnowski, T. J., & Rosenberg, C. R. (1987). Parallel networks that learn to pronounce English text. *Complex Systems, 1,* 145–168.

Shallice, T. (1988). *From neuropsychology to mental structure.* Cambridge, England: Cambridge University Press.

Shallice, T., Glasspool, D., & Houghton, G. (1995). Can neuropsychological evidence inform connectionist modelling? Analyses from spelling. *Language and Cognitive Processes, 10* (3/4), 195–225.

Shallice, T., & McCarthy, R. (1985). Phonological reading: From patterns of impairment to possible procedures. In K. E. Patterson, J. C. Marshall, & M. Coltheart (Eds.), *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (pp. 361–397). London: Erlbaum.

Shallice, T., Warrington, E. K., & McCarthy, R. (1983). Reading without semantics. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 35A,* 111–138.

Shanks, D. R. (1991). Categorization by a connectionist network. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 433–443.

Shultz, T. R. (1991). Simulating stages of human cognitive development with connectionist models. In *Machine learning: Proceedings of the Eighth International Workshop* (pp. 105–109). San Mateo, CA: Morgan Kaufmann.

Siegel, S., & Allan, L. G. (1996). The widespread influence of the Rescorla-Wagner model. *Psychonomic Bulletin & Review, 3,* 314–321.

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1140–1154.

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review, 88,* 135–170.

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

Treiman, R. (1986). The division between onsets and rimes in English syllables. *Journal of Memory and Language, 25,* 476–491.

Treiman, R. (1989). The internal structure of the syllable. In G. Carlson & M. Tanenhaus (Eds.), *Linguistic structure in language processing* (pp. 27–52). Dordrecht, The Netherlands: Kluwer.

Treiman, R., & Chafetz, J. (1987). Are there onset- and rime-like units in printed words? In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 281–298). Hillsdale, NJ: Erlbaum.

Treiman, R., & Danis, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 145–152.

Treiman, R., Goswami, U., & Bruck, M. (1990). Not all nonwords are alike: Implications for reading development and theory. *Memory & Cognition, 18,* 559–567.

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., & Goldinger, S. D. (1994). Interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1269–1291.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review, 97,* 488–522.

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

Widrow, G., & Hoff, M. E. (1960). Adaptive switching circuits. *Institute of Radio Engineers, Western Electronic Show and Convention Record, Part 4* (pp. 96–104).

Wydell, T. N., Butterworth, B., & Patterson, K. E. (1995). The inconsistency of consistency effects in reading: The case of Japanese Kanji. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 491–514.

Wydell, T. N., Butterworth, B., Shibahara, N., & Zorzi, M. (1998). *The irregularity of regularity effects in reading: The case of Japanese Kanji.* Manuscript submitted for publication.

Wydell, T. N., & Humphreys, G. W. (in press). Kana processing: Effects of lexicality and word frequency in naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Zorzi, M., Houghton, G., & Butterworth, B. (1998). The development of spelling-sound relationships in a model of phonological reading. *Language and Cognitive Processes, 13,* 337–372.

Zorzi, M., & Umiltà, C. (1995). A computational model of the Simon effect. *Psychological Research, 58,* 193–205.

Received June 25, 1996
Revision received May 16, 1997
Accepted June 18, 1997 ■