Available online at www.sciencedirect.com

SCIENCE DIRECT®

COGNITION

Cognition 88 (2003) 49–78

www.elsevier.com/locate/cognit

Influence of consonantal context on the pronunciation of vowels: A comparison of human readers and computational models

Rebecca Treiman[a,*], Brett Kessler[a], Suzanne Bick[b]

[a]*Washington University, St. Louis, MO, USA*
[b]*Wayne State University, Detroit, MI, USA*

Received 10 April 2002; accepted 6 December 2002

Abstract

In two experiments, we found that college students' pronunciations of vowels in nonwords are influenced both by preceding and following consonants. The predominance of rimes in previous studies of reading does not appear to arise because readers are unable to pick up associations that cross the onset-rime boundary, but rather because English has relatively few such associations. Comparisons between people's vowel pronunciations and those produced by various computational models of reading showed that no model provided a good account of human performance on nonwords for which the vowel shows contextual conditioning. Possible directions for improved models are suggested.

Keywords: Vowel; Onset; Rime; Nonwords; Reading; Computational models

---

1. Introduction

One of the most impressive aspects of human linguistic skill is our ability to deal with novel materials. Encountering Lewis Carroll's (1871/1977) *Jabberwocky* for the first time – "'Twas brillig, and the slithy toves did gyre and gimble in the wabe …" – we not only infer that there was more than one tove that had the characteristic of slithiness but we also assign pronunciations to *slithy* and *toves*. Motivated by the fact that nonword pronunciation is a specific case of our ability to go beyond our linguistic experience, many researchers have examined the knowledge and processes that people use to pronounce nonwords.

---

[*] Corresponding author. Psychology Department, Washington University, Campus Box 1125, One Brookings Drive, St. Louis, MO 63130-4899, USA. Tel.: +1-314-935-5326; fax: +1-314-935-7588.
E-mail address: rtreiman@artsci.wustl.edu (R. Treiman).

0010-0277/03/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0010-0277(03)00003-9

Further impetus for the study of nonword pronunciation is that phonological coding occurs in the reading of real words, even when the words are read silently for meaning (e.g. Van Orden, 1987). Oral reading is particularly important for children, who often encounter words that are new to them. Information about how readers with different levels of experience pronounce nonwords thus provides an important foundation for theories of reading development and for teaching. Interest in oral reading has been heightened by the development of computer models that are designed to simulate humans' reading of words and nonwords, thereby permitting precise tests of particular theories. The resulting debates about model architectures and performance have spurred researchers to collect data that can help distinguish among the models.

In the present study, we examine the pronunciation of nonwords by experienced readers of English. We investigate the effects of the surrounding consonants on college students' pronunciation of vowels in monosyllabic nonwords, and we compare the contextual effects shown by human readers to those shown by computational models. Our focus on vowels is motivated by the fact that the pronunciations of vowel graphemes are quite variable in English. For example, *ea* is pronounced as /i/ in *beat*, /e/ in *steak*, and /ɛ/ in *head*. Quantitative evidence of this variability comes from Kessler and Treiman (2001), who calculated the consistency of different parts of English monosyllabic, monomorphemic words. Consistency has a value of 1 if a letter string is always sounded the same way; otherwise consistency is some number between 0 and 1. When the consonants that surround a vowel are not considered, the average consistency of vowels in the Kessler and Treiman word set is 0.72. The average consistency for *onsets* (initial consonants or consonant clusters) and *codas* (final consonants or consonant clusters) is substantially higher, about 0.98. These findings echo those of Treiman, Mullennix, Bijeljac-Babic, and Richmond-Welty (1995) and Venezky (1970), who also discussed the irregularity of English vowels.

How do readers cope with the variability of vowel pronunciation in English? The analyses reported by Kessler and Treiman (2001) suggest one possibility: Readers could use the context in which a vowel occurs to help narrow the range of likely pronunciations. Supporting the potential value of such a strategy, Kessler and Treiman found that the consistency of vowel pronunciation increases significantly when the coda is considered. As one example, *ea* is more likely to be pronounced as /ɛ/ when it occurs before *d* than when it occurs before many other consonant letters. The majority of English vowel graphemes that have some inconsistency are significantly helped by consideration of the coda letter string. Pooling over all vowel graphemes, knowledge of onset letters does not provide significant help in reading the vowel. Onset effects are reliable in a few cases, however, as when *a* is more likely to be pronounced as /ɑ/ after *w* and *u* than after other letters (e.g. *swamp* vs. *tramp*).

A number of previous results suggest that readers consider the coda when pronouncing vowels in nonwords. Wolf and Robinson (1976) reported that adults were significantly more likely to pronounce *ea* as /ɛ/ before *d* than before consonants such as *m* and *p*. The same kinds of results were obtained for other context-dependent vowels. Glushko (1979) reported similar findings. Additionally, he found that the latency to initiate pronunciation was longer for nonwords such as *wead* than for nonwords such as *weat*. Johnson and Venezky (1976), Ryder and Pearson (1980), and Treiman and Zukowski (1988) also observed coda influences on the pronunciation of vowel graphemes in nonwords. More recently, Andrews and Scarratt (1998) used regression analyses to examine the factors that influence nonword pronunciation. Their results suggest that vowel-coda units play an important role in nonword pronunciation, over and above the roles of individual graphemes. Similar results have been found with real words (e.g. Jared, 1997, 2002). These results fit nicely with the view that the vowel and coda of the spoken English syllable form a linguistic unit, the *rime*. Orthographic units that correspond to rimes appear to play a special role in reading, a role that may derive in part from their importance in the spoken language (Treiman et al., 1995). However, not all researchers agree that orthographic units corresponding to rimes are involved in word pronunciation. Coltheart, Rastle, Perry, Langdon, and Ziegler (2001) argued that the difficulties on words like *drove* that Jared (1997) attributed to the inconsistent pronunciation of the *-ove* unit (cf. *love*, *move*) instead reflect complications in the application of grapheme-to-phoneme rules. The system initially translates *o* to /ɑ/ and only later, when the final *e* is processed, decides on /o/. These difficulties have been called *whammies* (Rastle & Coltheart, 1998).

In the present experiments, we looked again at the effect of codas on the pronunciations of vowels in nonwords. Going beyond many of the previous studies, we selected our stimuli based on quantitative information about the vowel pronunciations that occur in the real words of English. Many previous studies have relied on qualitative information. For example, Glushko (1979) asserted that *wead* has more real-word neighbors with an atypical pronunciation of the vowel than does *weat*, but he did not provide information about the magnitude of the difference. Such data are needed if we wish to compare the strength of context effects in people's reading of nonwords to the strength of the effects in the writing system, as we do here. We used the results of Kessler and Treiman (2001) to select vowel graphemes whose pronunciations are markedly influenced by the coda. We examined college students' pronunciations of these graphemes when they occurred in two contexts – an experimental context and a control context. The control context was defined so that the *typical* pronunciation of the vowel, such as /i/ for *ea*, occurs in almost all real words of English that have that context. In the experimental context, the vowel tends to take on a less typical pronunciation, as when *ea* is pronounced as /ɛ/ before *d*. This pronunciation is called the *critical* pronunciation. We expected that readers would more often pronounce the vowel in the critical manner when it occurred in the experimental context, as in the nonword *clead*, than when it occurred in the control context, as in *cleam*. Such a difference would fit with the idea that orthographic units corresponding to rimes are involved in the pronunciation of nonwords. Because the experimental and control nonwords used the same vowel graphemes and final graphemes of the same length, any obtained differences cannot be attributed to whammies.

In addition to examining the influence of the coda on vowel pronunciation, we looked for a possible influence of the onset. Some previous researchers have suggested that onsets have little or no influence on adults' pronunciations of vowels. For example, Andrews and Scarratt (1998, p. 1067) concluded that "there was little evidence that pronunciation is influenced by CV word neighbors". In another study with nonwords, Treiman and Zukowski (1988) found no statistically reliable effects of onsets on vowel pronunciation, although there were nonsignificant trends. Treiman et al. (1995), based on studies with real words, stated that their results do "not support the suggestion … that readers use C_1V units as well as VC_2 units" (p. 131). Consistent with such statements, some theories of reading grant a special role to rimes that is not given to onset-vowel units (Patterson & Morton, 1985; Zorzi, Houghton, & Butterworth, 1998).

Although the studies cited above did not find statistically significant onset effects on vowel pronunciation, the researchers did acknowledge that such effects might occur in some cases. Indeed, the studies included only a small proportion of stimuli, if any, with onsets that reliably affect the pronunciations of vowels in English words. It is possible that readers are sensitive to the relatively few onset-to-vowel associations that exist in English but that these effects are often swamped by the effects of the more numerous coda-to-vowel associations. Kay (1987) found some evidence for a sensitivity to onset-to-vowel associations in the fact that nonwords such as *wook* (where both *woo*- and -*ook* have a bias toward the /ʊ/ pronunciation) were more likely to receive /ʊ/ pronunciations than nonwords such as *pook* (where -*ook* has a bias toward /ʊ/ but *poo*- does not). However, the study was briefly summarized in Kay's report and details of the stimuli were not presented. Patterson et al. (1996) reported that three patients with acquired surface dyslexia, a disorder characterized by a tendency to regularize words with atypical spelling-to-sound correspondences, were less likely to make regularization errors on words like *swamp* than on standard exception words like *pint*. This difference, Patterson et al. suggested, reflected the patients' sensitivity to patterns involving the onset and the vowel. *Swamp*, in other words, does not have truly exceptional spelling-to-sound correspondences. However, it would be premature to conclude from this result that normal readers show a similar sensitivity to onset-to-vowel associations. Treiman, Kessler, and Bick (2002) found that college students are influenced by onsets as well as codas when spelling vowels in words and nonwords. However, it would be premature to conclude from this result that similar conclusions apply to reading.

In the present study, we asked whether experienced readers' pronunciation of vowels is affected by the onset in those cases in which onset-to-vowel associations exist in English. For example, we compared people's pronunciations of *a* in experimental nonwords such as *wabs* (where critical /ɑ/ pronunciations would be anticipated given the statistics of English) and control nonwords such as *trabs* (where typical /æ/ pronunciations would be expected). As in the case of codas, we compared the strength of the effects in people's pronunciation of nonwords and the strength of the effects in the English writing system itself. If people are sensitive to onset-to-vowel associations when such associations exist in the language, this would suggest that readers use context beyond the rime unit. To the extent that rimes play a special role in reading, this may stem from the fact that associations between vowels and codas are more widespread in English than associations between vowels and onsets, not from people's intrinsically greater sensitivity to within-rime associations.

Our study was also designed to address questions about methods of assessing people's pronunciation of nonwords. Typically, participants are shown nonwords and are asked to pronounce them aloud as if they were real words (e.g. Seidenberg, Plaut, Petersen, McClelland, & McRae, 1994). Some studies measure time to initiate pronunciation as well as the pronunciation itself; other studies do not use RT as a dependent variable. Andrews and Scarratt (1998) used a somewhat different procedure in one of their experiments, asking participants to rate each of their pronunciations after producing it. Participants could indicate that their response was incorrect, that several other equally good pronunciations were possible, or that the pronunciation they had produced was the only plausible alternative. The tasks described so far require participants to be tested individually; they also require a researcher to transcribe the participants' responses. These time-consuming requirements are avoided in a procedure used by Johnson and Venezky (1976) and Ryder and Pearson (1980). Here, participants read nonwords silently and choose, for each nonword, one of several real words that has the same vowel sound as their pronunciation of the nonword. Ryder and Pearson found the same general pattern of results in this group-administered procedure as in the traditional individual testing method. However, they did not directly compare the data from the two methods. We used an individual testing procedure in Experiment 1 and a group testing procedure in Experiment 2 with the same nonwords, and we compared the results from the two methods. If group testing yields results that are indistinguishable from those of individual testing, researchers could collect more data with less time and less effort.

We did not use a voice key to measure participants' latencies to initiate pronunciations because recent results show that such measures can be problematic (Kessler, Treiman, & Mullennix, 2002; Rastle & Davis, 2002). The obtained RTs are affected by the phonetic characteristics of the onset, and different results may be obtained with different types of voice keys. The characteristics of the vowel are also influential. For example, high vowels such as the /i/ of *meal* tend to yield slower RT measurements than nonhigh vowels. If we were to compare RTs for *chead* and *cheal*, we could encounter problems if participants produced different pronunciations for the two vowels, as we expect them to do in some cases. Stimuli that test onset-to-vowel associations, such as *wabs* and *trabs*, have different onsets and may give rise to different vowel pronunciations as well. Here, too, it is difficult to draw strong conclusions from RT data. Kessler et al. (2002) have put forward some possible ways of circumventing these problems, some of which require the availability of data from large numbers of items. In the present study, we took the straightforward approach of examining the phonological forms that people assign to vowels in nonwords and not assessing how long it takes people to produce these forms.

In addition to addressing theoretical and methodological questions about people's pronunciations of nonwords, we wished to examine the performance of computational models of reading and compare their performance with that of humans. The existing models embody different ideas about how people deal with novelty in language. According to dual-route theories of reading, people possess both generative knowledge that allows them to deal with novel items and case-specific knowledge that may be used for known items. The model of Coltheart et al. (2001) takes this view. According to this model, the generative knowledge that helps readers pronounce nonwords is in the form of rules that, for the most part, relate individual graphemes to individual phonemes and that do not consider context. Zorzi et al. (1998) also placed their model in the dual-route category. In their view, the generative knowledge that supports nonword pronunciation is induced by connectionist learning principles and not by explicit rules.

Single-route theories claim that people use similar procedures to pronounce novel items and familiar items. According to connectionist models that take this position, these procedures involve the spread of activation along connections between units, the weights of which change as a function of experience with the correct pronunciations of words. This approach is represented in the connectionist models of Harm and Seidenberg (2002) and Plaut, McClelland, Seidenberg, and Patterson (1996). The model of Norris (1994) may be considered a nonconnectionist single-route theory. As Norris describes it, different levels of spelling-sound correspondence combine to determine the pronunciation of a word or nonword. These levels include the whole word (e.g. *cook*), units that are smaller than a word but larger than a grapheme (e.g. -*ook*), and individual graphemes (e.g. *oo*).

Performance on novel items was an important consideration in the development of current computational models. The first-generation connectionist model of Seidenberg and McClelland (1989) was substantially worse than human college students at pronouncing nonwords (Besner, Twilley, McCann, & Seergobin, 1990), providing a major impetus for the later modeling work of Plaut et al. (1996) and Harm and Seidenberg (2002). Plaut et al. concluded that their models generalized well to novel items, stating for example that "the overall ability of the attractor network [one of the models discussed in the paper] to pronounce nonwords is comparable to that of skilled readers" (p. 86). Harm and Seidenberg likewise concluded that their model generated plausible phonological codes for nonwords. Zorzi et al. (1998) and Coltheart et al. (2001) drew similar conclusions about their dual-route models. For example, Zorzi et al. pointed to their model's "success … in achieving good nonword reading" (p. 1153). Norris (1994) also presented data indicating similar performance on nonwords by humans and by his model. Such findings have led to the belief that all existing models provide an adequate account of the pronunciations that people assign to nonwords. As a result, the issue of generalization ability has not been central in recent years. Debates have instead focused on the models' ability to account for data on time to initiate pronunciation, especially reaction time data from real words (e.g. Coltheart et al., 2001; Spieler & Balota, 1997).

However, the conclusion that current models provide a realistic account of humans' ability to read nonwords may be premature. This is because the models' nonword pronunciations have typically not been evaluated in a stringent manner. Coltheart et al. (2001) counted their model's pronunciations as correct if they fit with the grapheme-to-phoneme rules that were provided to the model. The model was thus scored as correct if it pronounced *glook* as /gluk/. No data were provided to verify that human readers in fact pronounce *glook* in this way. Plaut et al. (1996) counted a model's response as correct if it matched any pronunciation in the training set for the same orthographic rime. For example, both /gluk/ (as in *spook*) and /glʊk/ (as in *book*) were scored as correct for *glook*. Harm and Seidenberg (2003) and Zorzi et al. (1998) used a similar procedure. Based on these findings and those reported by Norris (1994), it appears that none of the models that we consider here produce a large number of "wild" errors such as /glæk/ or /glum/ for *glook*. In this sense, the models generate plausible phonological codes for nonwords. But do the models generate the same phonological codes that people typically do? What if a model produces /gluk/ for *glook* whereas most human readers produce /glʊk/? If model performance is evaluated as described above, one cannot detect a difference of this kind. In the present study, we used more sensitive scoring methods to examine the specific pronunciations produced by various models and how they agree with pronunciations produced by experienced human readers.

Seidenberg et al. (1994) and Andrews and Scarratt (1998) have made some progress in developing ways to assess the nonword pronunciations produced by models and the extent

Table 1
Onset-to-vowel associations tested in Experiments 1 and 2

	Case 1: *a* (followed by consonants other than *r* or velar)	Case 2: *ar*
Preceding context for experimental nonwords	*u* or *w*	*u* or *w*
Preceding context for control nonwords	Any other letter	Any other letter
Critical vowel pronunciation in American English	/ɑ/	/ɔ/
Sample experimental nonword	*squant*	*warge*
Sample control nonword	*spant*	*carge*
Choices for sample pair in Experiment 2	*font*, *rant*	*forge*, *large*
Proportion of English words with critical vowel pronunciation, experimental context[a]	0.81 (0.95)	1.00 (0.96)
Proportion of English words with critical vowel pronunciation, control context[a]	0.01 (0.02)	0.00 (0.00)

[a] The first value is based on the monosyllabic words from Kessler and Treiman (2001); the second value is based on the final syllables of words in the larger sample from the Carnegie Mellon Pronouncing Dictionary (1998), which includes polysyllabic words.

to which they agree with those produced by people. Seidenberg et al. asked 24 Canadian college students to pronounce a large number of nonwords (600; the results for 10 nonwords were subsequently removed due to experimenter errors). The researchers determined the percentage of pronunciations produced by two different models that matched the most common pronunciation produced by people. One model, that of Plaut and McClelland (1993), was an earlier version of the attractor network described by Plaut et al. (1996). Approximately 80% of the pronunciations produced by this model matched the most common pronunciation used by people. The second model, an earlier version of Coltheart et al. (2001), performed at a similar level of accuracy. Andrews and Scarratt (1998) also reported quantitative assessments of an earlier version of the Coltheart et al. model. In the present study, we examine a larger number of models than in previous studies and evaluate all of the models according to the same benchmarks.

2. Experiment 1

Our first experiment was designed to determine whether college students' pronunciations of medial vowels in monosyllabic nonwords are affected by the identity of the preceding and following consonants and to provide data against which to test computational models of reading. We asked whether the same vowel grapheme is pronounced differently when it occurs in certain contexts – those for which Kessler and Treiman (2001) documented a significant consonant-to-vowel association – than when it occurs in other contexts. We selected two cases of onset-to-vowel conditioning and six cases of coda-to-vowel conditioning for study. The number of onset-to-vowel cases was limited by the fact that there are few such associations in English.

Table 1 provides information about the two onset-to-vowel cases that we examined. In each case, we identified two contrasting environments such that a particular pronunciation

Table 2
Coda-to-vowel associations tested in Experiments 1 and 2

	Case 1: *a*	Case 2: *a*	Case 3: *ea*	Case 4: *i*	Case 5: *o*	Case 6: *oo*
Following context for experimental nonwords	*nge*	*ld* or *lt*	*d*	*nd* or *ld*	*ld* or *lt*	*k*
Following context for control nonwords	*nce*	*nd* or *nt*	*b*, *l*, *m*, *n*, or *p*	*nt* or *lt*	*nd* or *nt*	*m*, *n*, or *p*
Critical vowel pronunciation in American English	/e/	/ɔ/	/ɛ/	/aɪ/	/o/	/ʊ/
Sample experimental nonword	*blange*	*yald*	*clead*	*ild*	*brold*	*blook*
Sample control nonword	*blance*	*yant*	*cleam*	*ilt*	*brond*	*bloon*
Choices for sample pair in Experiment 2	*blade*, *black*	*yawn*, *yap*	*clench*, *cleat*	*aisle*, *id*	*broke*, *bronze*	*pull*, *bloom*
Proportion of English words with critical vowel pronunciation, experimental context[a]	1.00 (0.95)	1.00 (0.80)	0.69 (0.78)	0.89 (0.80)	1.00 (0.99)	0.94 (0.97)
Proportion of English words with critical vowel pronunciation, control context[a]	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.06 (0.04)	0.00 (0.03)	0.00 (0.01)

[a] The first value is based on the monosyllabic words from Kessler and Treiman (2001); the second value is based on the final syllables of words in the larger sample from the Carnegie Mellon Pronouncing Dictionary (1998), which includes polysyllabic words.

of a vowel grapheme, the critical pronunciation, occurs more often in English words in one environment than the other. Consider Case 1, which involves the vowel *a* before consonants other than *r* or velars. In the monosyllabic, monomorphemic words of American English that were analyzed by Kessler and Treiman (2001), the critical pronunciation of this vowel, /ɑ/, occurs 81% of the time when *a* is preceded by *u* or *w*, the experimental context. In contrast, the /ɑ/ pronunciation is found only 1% of the time when *a* is preceded by other consonants, the control context. In this environment, the vowel is generally pronounced with /æ/, its typical pronunciation. The counts of pronunciation frequency in English that we report here and elsewhere in this paper sum across words, each word weighted by the logarithm of 2 plus its frequency of occurrence in Zeno, Ivenz, Millard, and Duvvuri (1995). This decision was based on the finding of Kessler, Treiman and Mullennix (2003) that this counting scheme generally accounts for slightly more variance in word naming data than does a pure type-based scheme that ignores frequency. The results are very similar if counts are based solely on types, however; the average difference in measurements is 1.5 percentage points. We also examined the pronunciation of each vowel in the experimental and control contexts in a larger sample of American English words (Carnegie Mellon Pronouncing Dictionary, 1998), going beyond the monosyllables studied by Kessler and Treiman (2001). In this case, we considered only the first vowel of the words and its environs; the vowel had to be stressed and in an orthographically closed syllable (i.e. *waddle* was included in the counts for Case 1 but not *waded*). As Table 1 shows, the critical pronunciation was substantially more frequent in the experimental context than the control context in the larger word set.

The variant vowel pronunciations that we examine reflect sound changes that occurred after the English spelling system crystallized. Words like *waddle* were once pronounced with the same vowel as *paddle*. The pronunciation of *a* changed under the influence of the preceding *w* but the spelling did not change.

Once the experimental and control contexts were defined, we constructed one set of monosyllabic nonwords in which the target grapheme appeared in the experimental context and a matched set of monosyllabic nonwords in which the grapheme appeared in the control context. Sample items are shown in Table 1. If readers are sensitive to associations between onsets and vowels in English, then pronunciations with the critical vowel phoneme should be more common in the experimental nonwords than the control nonwords.

We used the same logic to examine six cases of coda-to-vowel conditioning. Table 2 provides information about each of these cases. The counts of pronunciation frequency for the larger sample that are shown in the table are based on the last vowels of words stressed on the final syllable. If readers consider the coda in pronouncing the vowel, then the critical vowel pronunciation should be more common for the experimental nonwords than the control nonwords.

2.1. Method

2.1.1. Stimuli

We constructed 10 pairs of experimental and control nonwords for each of our 8 cases. One of the stimuli originally constructed for onset-to-vowel Case 1 was found to contain an error, so the results that are reported for this case are based on 9 pairs. The nonwords in each pair testing onset-to-vowel associations were alike in the vowel and the following letters; they differed only in the initial portion. The nonwords testing coda-to-vowel associations were alike in their onset and vowel letters and differed only in their final portion. Both the pronunciation with the critical vowel and the pronunciation with the typical vowel were phonologically legal, occurring in one or more familiar English words. In addition to the experimental and control stimuli, there were 20 filler nonwords. The fillers included some graphemes that did not appear in the experimental and control stimuli. Their inclusion decreased the repetitiveness of the list. All of the nonwords were orthographically legal in English. Appendix A shows the experimental stimuli, control stimuli, and fillers. Three different sequences were prepared for purposes of presentation. In each order, the experimental items, control items, and fillers were randomly intermixed with the constraint that no more than two consecutive items involved the same case.

2.1.2. Procedure

The participants were tested individually. They were assigned to one of the three sequences according to the order in which they were tested. The participants were told that they would be asked to pronounce a series of "made-up words". They were asked to pretend that these were ordinary, everyday words of English and to pronounce each item the way they thought it would be read if it were a real word. This instruction was designed to counteract any tendency people may have to assume that unfamiliar words are foreign and should therefore be read with atypical patterns. A rest break was given halfway through the list. The participants' pronunciations were tape-recorded. A phonetically trained individual who was not aware of the experimental hypotheses scored the pronunciations and later checked the scoring using the tapes.

2.1.3. Participants

We tested 25 students at Wayne State University who were native speakers of English and who reported no history of speech, hearing, or reading disorders. The students participated in exchange for extra credit in a course. One student pronounced 14 of the 20 fillers in a manner that did not conform to the spelling-sound patterns of English, even given credit for analogies to real words and spelling-sound correspondences that were used in atypical contexts. Data from this participant were not included in the analyses.

2.2. Results and discussion

The pronunciations of experimental and control items were coded as containing the critical pronunciation of the vowel, the typical pronunciation of the vowel, or some other pronunciation. To check reliability, a second individual coded the results for six of the participants using the tapes. The two coders agreed in 95% of the cases. Table 3 shows the proportion of pronunciations that used the critical vowel for the experimental and control stimuli testing onset-to-vowel associations. Table 4 shows the results for the stimuli testing coda-to-vowel associations. Less than 3% of the responses contained a pronunciation of the vowel that was not either the critical or the typical pronunciation.

Table 3
Results for onset-to-vowel associations in Experiments 1 and 2

	Case 1: *a* (followed by consonants other than *r* or velar)	Case 2: *ar*
Experiment 1		
Mean (SD) proportion of critical vowel pronunciations, experimental nonwords	0.64 (0.26)	0.17 (0.29)
Mean (SD) proportion of critical vowel pronunciations, control nonwords	0.06 (0.10)	0.01 (0.03)
P value for difference, one-tailed *t* test by subjects	< 0.001	0.004
P value for difference, one-tailed *t* test by items	< 0.001	< 0.001
Experiment 2		
Mean (SD) proportion of critical vowel pronunciations, experimental nonwords	0.55 (0.22)	0.15 (0.19)
Mean (SD) proportion of critical vowel pronunciations, control nonwords	0.03 (0.06)	0.03 (0.05)
P value for difference, one-tailed *t* test by subjects	< 0.001	0.002
P value for difference, one-tailed *t* test by items	< 0.001	< 0.001

For both onset-to-vowel cases and for all 6 coda-to-vowel cases, participants produced more critical pronunciations of the target vowel for the experimental stimuli than for the control stimuli. In each case, the difference was statistically significant by both subjects and items. Adults' pronunciations of vowels in nonwords are thus affected by both the onset and the coda.

The top two rows of data in Table 5 show the difference in proportion of critical pronunciations between experimental words and control words in the English language and in the participants' data. The results are shown separately for each of the eight cases. Although readers produced more critical vowel pronunciations for the experimental nonwords than the control nonwords, the differences were generally smaller than expected based on the statistics of English. That is, readers' vowel pronunciations are affected by context, but not to the extent that one might expect given the words to which they have been exposed. We further consider both aspects of the results – the existence of contextual effects and the fact that they are smaller than might be anticipated – in the General Discussion (Section 5).

A potential problem in interpreting the results of Experiment 1 is that certain letter sequences were repeated across the experiment. This repetition may have affected the participants' responses. We thus looked at the results for the first presentation of each target vowel grapheme in an experimental nonword and the first presentation of each target vowel in a control nonword. The proportion of pronunciations that used the critical vowel was higher for the experimental stimuli than the control stimuli in both onset-to-vowel cases and all six coda-to-vowel cases. Further evidence that the differences found here do not reflect the repetition of orthographic rimes comes from analyses of the nonword pronunciation data reported by Seidenberg et al. (1994). Among the 590 nonwords for which data were available were 10 that fit one of the present descriptions of an experimental nonword testing coda-to-vowel associations. Another 15 stimuli fit our description of a control nonword testing coda-to-vowel associations. (No stimuli fit the criteria for

Table 4
Results for coda-to-vowel associations in Experiments 1 and 2

	Case 1: *a*	Case 2: *a*	Case 3: *ea*	Case 4: *i*	Case 5: *o*	Case 6: *oo*
Experiment 1						
Mean (SD) proportion of critical vowel pronunciations, experimental nonwords	0.59 (0.34)	0.94 (0.10)	0.13 (0.18)	0.35 (0.25)	0.88 (0.17)	0.70 (0.34)
Mean (SD) proportion of critical vowel pronunciations, control nonwords	0.05 (0.11)	0.08 (0.17)	0.01 (0.03)	0.02 (0.05)	0.05 (0.13)	0.00 (0.00)
P value for difference, one-tailed *t* test by subjects	< 0.001	< 0.001	0.004	< 0.001	< 0.001	< 0.001
P value for difference, one-tailed *t* test by items	< 0.001	< 0.001	0.013	< 0.001	< 0.001	< 0.001
Experiment 2						
Mean (SD) proportion of critical vowel pronunciations, experimental nonwords	0.57 (0.24)	0.80 (0.23)	0.25 (0.28)	0.48 (0.30)	0.87 (0.16)	0.38 (0.32)
Mean (SD) proportion of critical vowel pronunciations, control nonwords	0.17 (0.20)	0.16 (0.19)	0.09 (0.13)	0.07 (0.16)	0.11 (0.13)	0.05 (0.10)
P value for difference, one-tailed *t* test by subjects	< 0.001	< 0.001	0.002	< 0.001	< 0.001	< 0.001
P value for difference, one-tailed *t* test by items	< 0.001	< 0.001	0.001	< 0.001	< 0.001	< 0.001

experimental nonwords testing onset-to-vowel associations.) The proportion of pronunciations that used the critical vowel was 0.45 for the experimental nonwords and 0.06 for the control nonwords. This difference was statistically significant according to a *t* test by items ($t(23) = 4.35$, $P < 0.001$, one-tailed).[1] Thus, contextual effects of the kind found here emerge when the nonwords of interest are interspersed among a large number of other nonwords and when rime spellings are not repeated within an experiment.

3. Experiment 2

Experiment 1, like most other studies in which participants' pronunciations of nonwords are used to gain insight into the reading process (e.g. Andrews & Scarratt, 1998; Seiden-

---

[1] The data were made available to us in a form that did not allow for by-subject tests.

Table 5

Difference between proportion of critical pronunciations of vowels in experimental nonwords and control nonwords in English monosyllabic words (results for all words in parentheses), human data of Experiment 1, and models

	Case							
	CV1	CV2	VC1	VC2	VC3	VC4	VC5	VC6
English words	0.80 (0.93)	1.00 (0.96)	1.00 (0.95)	1.00 (0.79)	0.69 (0.78)	0.83 (0.76)	1.00 (0.96)	0.94 (0.96)
Human data	0.58	0.16	0.55	0.86	0.12	0.33	0.83	0.70
Coltheart et al., 2001	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Zorzi et al., 1998	0.00	0.60	0.90	0.20	0.00	1.00	1.00	0.20
Plaut et al., 1996, Simulation 2	0.33	0.50	1.00	0.30	0.60	0.50	0.90	0.80
Plaut et al., 1996, Simulation 3	0.89	0.80	0.90	0.20	0.40	0.70	0.80	0.70
Plaut et al., 1996, Simulation 4	0.44	0.00	0.60	0.10	0.20	0.50	1.00	0.60
Plaut & McClelland, 1993	0.56	0.50	1.00	0.70	0.00	0.90	1.00	0.70
Powell et al., 2001	0.56	0.60	0.70	0.30	0.10	0.70	1.00	0.80
Harm & Seidenberg, 2003	0.56	0.70	0.90	0.70	0.40	1.00	1.00	0.80
Norris, 1994, WS parameters	0.00	0.10	1.00	0.90	0.30	0.50	0.90	1.00
Norris, 1994, TM parameters	0.00	0.20	1.00	0.90	0.30	0.40	0.80	1.00

berg et al., 1994), used a labor-intensive procedure in which participants are tested individually and their responses scored by an experimenter. In Experiment 2, we asked whether the same pattern of results would emerge when participants are tested in groups and when they score their own responses. We asked participants to pronounce each nonword silently and then choose which of two real words was most similar to their pronunciation of the nonword. For example, participants were asked whether they pronounced *squant* more similarly to *font* or to *rant*. This procedure is similar to that of Johnson and Venezky (1976) and Ryder and Pearson (1980). However, we used two real-word choices instead of four, as in those previous studies, and we gave participants the opportunity to respond that their pronunciation was similar to neither of the real words that were presented. Unlike the other studies, we included filler items to assess whether participants were paying attention to and understood the task. Experiment 2 used the same critical nonwords as Experiment 1. If the results mirror those of Experiment 1, this would suggest that a group testing procedure can yield reliable results in studies of nonword pronunciation.

3.1. Method

3.1.1. Stimuli

The experimental and control nonwords were the same as in Experiment 1. For each pair of experimental and control nonwords, two real words were selected. One real word had the critical pronunciation of the vowel and the other had the typical pronunciation. In the onset-to-vowel cases, both of the real-word choices shared the nonword's coda. For these items, the *VC choice* items, the participants' task involved choosing the real word that rhymed with their pronunciation of the nonword. For most of the coda-to-vowel stimuli, both of the choices shared the nonword's onset. For these *CV choice* items, the participant judged which real word began like the nonword. For most of the stimuli for coda-to-vowel Case 6, no real-word pairs could be found that fit the above descriptions. In the resulting *V choice* items, the participant judged which real word shared its vowel with their pronunciation of the nonword. The real-word choices are shown in Appendix A. In most instances, the vowel grapheme in the real word with the typical vowel pronunciation was the same as the vowel grapheme in the nonword but different from the vowel grapheme in the real word with the critical vowel pronunciation. For example, the vowel grapheme in *squant* matches the one in *rant* (which has the typical /æ/ pronunciation of *a*) but differs from the one in *font* (which has the critical /ɑ/ pronunciation in American English). Thus, any tendency to choose a response based on orthographic similarity to the nonword would lead to a low rate of critical vowel choices for both experimental and control nonwords.

An additional 20 filler items were included. Most of these were taken from Experiment 1. Two real-word choices were selected for each filler nonword. The incorrect choice generally corresponded to the most popular illegal pronunciation of the grapheme in Experiment 1, and sometimes involved a consonant rather than a vowel. For example, because *feg* was sometimes pronounced as /fɛdʒ/ in Experiment 1, *leg* and *ledge* were chosen as the options in Experiment 2. An additional six real words were used for practice. The filler items and the practice items, along with the choices for each, are listed in

Appendix A. The nonwords were presented one at a time on an overhead projector in one of the orders that was used for Experiment 1. The practice items were presented before the nonwords.

3.1.2. Procedure

The participants were tested in groups. They were told that a series of “made-up words” would be presented. The participants were asked to pretend that these were ordinary, everyday words of English and to pronounce each item to themselves the way they thought it would be pronounced if it were a real word. After the participants had been given time to pronounce each item, the experimenter read aloud the two real-word choices in a randomly chosen order. The participants were asked to circle the letter A on their answer sheet if the nonword, as they had pronounced it, was similar to the first real word. They were asked to circle B if the nonword was similar to the second real word and to circle “neither” if it was not similar to either real word. For the practice items, the participants selected their response and the experimenter then explained which response was correct. No such feedback was given on the test items. There was a rest break halfway through the experiment.

3.1.3. Participants

Thirty-two students from the same population as in Experiment 1 were tested. None of them had participated in Experiment 1. The students’ responses to the fillers were examined to determine whether they understood and were paying attention to the task and whether they could make the required judgments. Nine of the students produced 13 or fewer correct responses on the 20 filler items (counting “neither” responses as incorrect). The data from these participants were not included in the reported analyses. For the remaining 23 students, the number of correct responses on the filler items averaged 16.8 of 20 (84% correct).

3.2. Results and discussion

Table 3 shows the proportion of responses with the critical vowel pronunciation for the stimuli testing onset-to-vowel associations. Table 4 shows the results for the stimuli testing coda-to-vowel associations. Participants produced more critical pronunciations of the target vowel for the experimental stimuli than for the control stimuli. The difference was statistically significant both by subjects and items in each case, as the tables show. As in Experiment 1, the differences were smaller than might be anticipated on the basis of the statistics of English (shown in Tables 1 and 2). We also looked at the results for the first presentation of each target grapheme in an experimental nonword and the first presentation of each target grapheme in a control nonword. In each case, there were more pronunciations with the critical vowel for the experimental nonwords than the control nonwords.

By comparing the results of Experiments 1 and 2, we can address questions about methods of assessing nonword pronunciation. There was no significant difference between the proportion of critical vowel pronunciations in the individual test of Experiment 1 and the group test of Experiment 2 ($P = 0.64$ according to a t test across all items). However, the proportion of “neither” responses – those with neither the critical vowel pronunciation nor the typical vowel pronunciation – was 0.06 across all items in the group test as compared to 0.03 in the individual test, a significant difference ($P < 0.001$, two-tailed). This difference reflects, in part, the nature of the choices that were offered in Experiment 2. When the choice was based on a single shared vowel, a V choice, participants made substantially more "neither" responses than when the choice was based on more than one shared phoneme, a VC or CV choice. Supporting this conclusion, an analysis of variance across all experimental and control items using the factor of choice type (VC, CV, or V) showed a significant effect ($F(2, 155) = 58.48$, $P < 0.001$). A post hoc test revealed significantly more "neither" responses for those relatively few items with a V choice (0.24) than for those items with a VC (0.04) or CV (0.02) choice. This result suggests that participants found it harder to make judgments involving a single vowel than judgments involving more than one phoneme. The difficulty may have been exacerbated by the fact that choices involving a single vowel did not occur very often in the experiment. The observed differences do not appear to reflect the specific phonemes in the nonwords for which the choice hinged on a single vowel. Items for which the choice was based on a single vowel in Experiment 2 did not yield an elevated number of unusual vowel pronunciations in Experiment 1. The nature of the choices for the filler items also appeared to contribute to some participants' relatively poor performance on these items. For a number of the fillers, the correct response shared just one phoneme with the expected pronunciation of the item.

Methodologically, the results of Experiment 2 suggest that group testing with self-scoring yields less reliable results than individual testing. In a group setting, with little outside pressure to attend to the task, some participants appear to respond randomly. In addition, some participants find it difficult to make the required similarity judgments. The presentation of the two choices may call attention to the alternative pronunciation in a way that would not normally happen. Despite these problems, it is notable that Experiment 2 yielded the same general pattern of results as Experiment 1. A group-administered nonword pronunciation task can yield credible data if efforts are made to make the choices as easy as possible and ensure that participants try to perform well.

4. Performance of computational models

We turn now to a comparison between the human results and those of the computational models that currently exist for reading aloud in English. The most prominent model in the dual-route category is that of Coltheart et al. (2001). The lexical route of this model provides case-specific knowledge. Generative knowledge is provided by a route that converts letter strings into phoneme strings by means of rules at the grapheme–phoneme level. Some of these rules are sensitive to context, as when *c* is translated to /s/ before some letters but to /k/ before others, but most rules do not take context into account. The rules are specified by the researchers rather than discovered by the model. The publicly available implementation of this model has a lexicon of 7981 monosyllabic words, both single-morpheme words and words of more than one morpheme. It uses Australian English pronunciations.

We also considered the model of phonological assembly of Zorzi et al. (1998), a feedforward connectionist network with an input layer (orthography) and an output layer (phonology) and no hidden units. This model, when supplemented by a pathway that provides case-specific knowledge, may be considered a dual-route model. The input and output representations of the Zorzi et al. model are organized in terms of onset slots and rime slots. For example, the *t* of *tip* fills the first onset slot and the *i* and *p* fill the first and second rime slots. In *stamp*, the *s* fills the first onset slot and the *t* the second, with *a*, *m*, and *p* filling the first, second, and third rime slots, respectively. This coding scheme is meant to help the model use patterns within the rime to predict pronunciation. This model was trained on 2774 monosyllables with British pronunciations.

Turning to single-route models of a purely connectionist variety, we examined the models of orthography-to-phonology conversion discussed by Plaut et al. (1996). These models have three layers: an input layer of grapheme units, a layer of hidden units, and an output layer of phoneme units. A grapheme in the onset is represented differently from the same grapheme in the coda; the same holds true for phonemes. This coding system, together with several additional assumptions, allows the model to effectively represent the relative order of letters and phonemes. We considered Simulation 2 of Plaut et al., a feedforward network that was trained on 2998 monosyllabic words with American English pronunciations, mostly monomorphemic. In Simulation 2, the words were presented to the model at frequencies proportionate to their raw frequencies from Kučera and Francis (1967). We also considered Simulation 3 of Plaut et al., an attractor network that was trained on the same words, which were again presented proportionately to their raw Kučera and Francis frequencies. In an attractor network, the output representations have "cleanup" units that tend to drive degraded patterns of activity toward more stable, familiar patterns. Simulations 2 and 3 yielded slightly better agreement with our human data than did Simulation 1 of Plaut et al., a feedforward network in which words were presented in line with the logarithm of their Kučera and Francis frequencies, and an alternate version of Simulation 2 that used a square root compression of frequency. Also, in analyses reported by Plaut et al., Simulation 2 with raw frequencies and Simulation 3 generally gave better fits to human data than the other two versions mentioned above. Finally, we examined Simulation 4 of Plaut et al., a feedforward network in which the correct phoneme units receive external activation in a way that is meant to simulate the contribution of semantics in word reading. We followed the procedure of Plaut et al. in using results for each simulation that were based on a single trained model. It is possible that somewhat different responses would occur on different runs as a function of the initial weights on the connections or the order in which words are presented in training, but this remains to be investigated.

We also examined two models that are similar to those of Plaut et al. (1996) but that were trained on isolated grapheme-phoneme correspondences as well as on whole words. The first of these was the model of Plaut and McClelland (1993), which was otherwise very similar to Simulation 3 of Plaut et al. The second was developed by Powell, Plaut, and Funnell (2001) in an attempt to simulate data on children's acquisition of reading.

The connectionist model of Harm and Seidenberg (2002) was also tested. This model includes mappings from orthography to semantics and mappings between semantics and phonology as well as mappings from orthography to phonology. There are direct connections between orthographic and phonological representations as well as connections mediated by hidden units. The model uses attractor structures for both the phonological and semantic representations. This model's phonological representations, unlike those of the other models considered here, incorporate phonetic features. The model was trained on 6103 monosyllabic words with American English pronunciations, including words with more than one morpheme as well as single-morpheme words. The first phase of training involved phonology, semantics, and the relationships between them. This was meant to simulate the fact that children are familiar with the phonological and semantic forms of many words before learning to read them. The connections from orthography to the other units were trained in a second phase. Words were presented in training proportionate to the square root of their frequency in the Wall Street Journal Corpus (Marcus, Santorini, & Marcinkiewicz, 1994).

The final model we examined is that of Norris (1994). According to this model, generation of the pronunciation of a unit (onset, vowel, or coda) considers spelling–sound mappings that involve the unit itself, the unit in the context of adjacent unit(s), and whole words. For example, the vowel of *book* would be read by considering the pronunciation of *oo* in all words with *-oo-*, *boo-*, *-ook*, and *book*. We examined the performance of the Norris model using two sets of parameters. The first parameter set provided the best fit for the data of Waters and Seidenberg (1985) in analyses reported by Norris; the second set provided the best fit for data collected by Taraban and McClelland (1987). These will be referred to as the WS parameters and the TM parameters, respectively. Norris used 2900 mostly monomorphemic words for his simulations. The sublexical spelling-to-sound correspondences that the model induced were based on the logarithm of the number of words in the set that embodied each correspondence. The words used by Norris overlapped in large part with those used by Plaut et al. (1996), but some of the pronunciations were changed so as to represent British English.

Our experimental and control nonwords were submitted for pronunciation to each of the models. We compared the pronunciations produced by the models to those produced by the human participants in Experiment 1, the results of which appear more reliable than those of Experiment 2. A few issues arise when comparing the results for models that are based on British or Australian English to pronunciations produced by speakers of American English. The standard accents of the US, Britain, and Australia distinguish the vowels in the experimental and control contexts in all eight cases, and we expect that the proportions of critical vowels in experimental and control contexts in American English words (Tables 1 and 2) will be similar in the other dialects. However, the sets of words that have the relevant critical and typical vowels and the identities of the critical and typical vowels sometimes vary across accents. In onset-to-vowel Case 1, for example, the critical vowel in British English is /ɒ/, as in *hot*, which differs from the vowel in *harm*. In American English the critical vowel is /ɑ/, which appears in both *hot* and *harm*. We classified each model's pronunciation of each vowel as the critical vowel for the dialect assumed by the model, the typical vowel for the dialect, or some other vowel. One method of assessing agreement between each model's performance and the humans' performance, which we will call the *binary scoring method*, began by determining how human readers most commonly pronounced the vowel of each nonword – with the critical pronunciation for our participants' dialect, the typical pronunciation, or some other pronunciation. A model's pronunciation was scored as correct if its vowel pronunciation fell into the same category as the humans' preferred pronunciation, and otherwise as incorrect. (The humans' preferred pronunciation of the vowel was either the critical pronunciation or the typical pronunciation for all of the nonwords.) Our second method of assessing agreement was to determine, for each nonword, the proportion of cases in which humans' pronunciation of the vowel fell into the same category as the model's pronunciation. We then calculated the proportion of matching responses relative to all responses for all nonwords as a whole and for various types of items. For "other" responses, a match was scored if a model's vowel exactly matched the vowel that was used by a human reader.[2] This second method will be called the *proportional method*.

In interpreting the results, bear in mind that the models differ in the size of the training set and the identity of the words in the training set, as well as in the pronunciations of some of the words and the amount of training that was given. These factors could influence the rate of agreement above and beyond the architectural features of the models and the nature of the models' representations. Also, bear in mind that our stimuli are not a representative sample of nonwords. The experimental nonwords are inherently more complex than the nonwords without contextual conditioning, and we subject the models to a stringent test by assessing their performance on such nonwords. However, it is important to do this in order to determine whether the models pronounce nonwords in the way that people do.

Table 6 shows the results using the two methods of assessing human-model agreement. The figures are generally somewhat higher for the binary method than for the proportional method. This may occur because some human responses reflect attentional or visual lapses or are otherwise idiosyncratic (e.g. *quead* read as /kjudi/, *squean* read as /kwin/). Arguably, no model of spelling-to-sound translation should account for such responses, and indeed the binary method effectively ignores them. Despite the differences between the two methods, the patterns of results are similar.

The general conclusion from Table 6 is that none of the models do a very good job of accounting for the averaged human data. Most of the models do well on the control nonwords. Even the model that performs worst on the control nonwords, one version of Norris (1994), scores 86% correct according to the binary criterion. However, none of the models do well on the experimental nonwords. The model that performs best on the experimental nonwords, one version of the Norris model, gets only 68% correct by the binary criterion. The model that performs worst on the experimental nonwords, that of Coltheart et al. (2001), scores only 38% correct. These levels of performance are not impressive.

To get a better understanding of the models' performances, we may examine the differences that they show between experimental and control nonwords in the proportion of critical vowel pronunciations and compare the difference scores to those of humans. These informal analyses give us a feel for the models' strengths and weaknesses. Table 5 shows the difference scores for each model for each of the eight cases we examined. Also shown are the differences for human subjects and the differences in the language itself. The Coltheart et al. (2001) model produces the same vowel pronunciations for the experimental items and the control items in all cases. The only exception is that the model

---

[2] In a handful of cases in which a model assumed British English pronunciations and it was not meaningful to expect an exact match with American English pronunciations, the results for "other" responses were based on the total number of such responses. The scoring decision about these cases had a negligible effect on the outcomes.

Table 6
Agreement between vowels produced by models and vowels produced by humans in Experiment 1 using the binary method (results using proportional method are shown in parentheses)

	All experimental and control nonwords	Experimental nonwords			Control nonwords		
		All	CV	VC	All	CV	VC
Coltheart et al., 2001	0.69 (0.66)	0.38 (0.38)	0.53 (0.56)	0.33 (0.33)	1.00 (0.94)	1.00 (0.95)	1.00 (0.94)
Zorzi et al., 1998	0.70 (0.67)	0.43 (0.42)	0.00 (0.11)	0.57 (0.52)	0.96 (0.92)	0.95 (0.91)	0.97 (0.92)
Plaut et al., 1996, Simulation 2	0.76 (0.70)	0.56 (0.50)	0.32 (0.35)	0.63 (0.55)	0.96 (0.90)	1.00 (0.95)	0.95 (0.89)
Plaut et al., 1996, Simulation 3	0.75 (0.69)	0.57 (0.48)	0.47 (0.39)	0.60 (0.51)	0.94 (0.90)	0.95 (0.90)	0.93 (0.90)
Plaut et al., 1996, Simulation 4	0.79 (0.73)	0.63 (0.56)	0.63 (0.56)	0.64 (0.55)	0.95 (0.91)	1.00 (0.95)	0.93 (0.90)
Plaut & McClelland, 1993	0.78 (0.72)	0.63 (0.56)	0.32 (0.31)	0.73 (0.64)	0.92 (0.87)	0.89 (0.84)	0.93 (0.88)
Powell et al., 2001	0.78 (0.72)	0.61 (0.53)	0.32 (0.34)	0.70 (0.60)	0.95 (0.90)	0.95 (0.90)	0.95 (0.91)
Harm & Seidenberg, 2003	0.74 (0.70)	0.54 (0.52)	0.32 (0.38)	0.62 (0.56)	0.94 (0.88)	1.00 (0.95)	0.92 (0.86)
Norris, 1994, WS parameters	0.76 (0.69)	0.66 (0.57)	0.11 (0.19)	0.83 (0.68)	0.86 (0.82)	0.68 (0.66)	0.92 (0.87)
Norris, 1994, TM parameters	0.79 (0.71)	0.68 (0.59)	0.16 (0.25)	0.85 (0.69)	0.89 (0.83)	0.74 (0.69)	0.93 (0.88)

pronounces *wadge* with the critical vowel; this is because Coltheart et al. list *wadge* as a real word, and their model therefore gets the correct pronunciation by lexical lookup. *Wadge* is not a word for our US participants, and so the results on this item are not included in the data for the Coltheart et al. model shown in Table 5. On the average, human readers produced more critical vowel pronunciations for the experimental items than the control items in all eight cases. Indeed, no human reader of the 47 whose data were included across the two experiments failed to show at least some contextual effects. Because the model of Coltheart et al. did not show any contextual effects, it does not account for the human results.

The other models reproduce the human data at a broad level in that all show more critical pronunciations for the experimental nonwords as a group than for the control nonwords. However, the models do not reproduce the human data well at a detailed level in that the sizes of the differences do not closely mirror the average difference scores shown by people. The connectionist models overestimate the human differences in more cases than they underestimate them. The Norris (1994) model generally shows larger differences than humans for coda-to-vowel associations but not for onset-to-vowel associations.

The poor performance of the Coltheart et al. (2001) model on experimental nonwords fits with other reports that the model does not account for consistency effects in the oral reading of real words (e.g. Jared, 2002). The generative route of the Coltheart et al. model includes rules that operate at the level of single graphemes and single phonemes; in most cases, including those examined here, the rules do not take context into account. Vowels in nonwords are thus given their typical pronunciations by the nonlexical route. The influence of the lexical route on nonword reading is not large enough to alter these pronunciations. Nor is it large enough to alter the response times. The number of cycles taken by the model on the present experimental nonwords is not significantly greater than the number of cycles taken on the control nonwords (153 vs. 151; $P = 0.28$). The Coltheart et al. model could in principle be modified by adding rules that operate on larger units than graphemes and phonemes or by incorporating additional context-sensitive rules. Another possible modification would be to increase the role of the lexical route in nonword reading. Additional work would be needed to determine whether and how such changes impact the model's ability to account for other findings. The model did perform well on our control nonwords, and other aspects of the model's performance appear to correspond well to human performance (e.g. Coltheart et al., 2001; Kessler et al., 2003).

The model of Zorzi et al. (1998) was partly motivated by a desire to account for coda-to-vowel effects in reading. This led the researchers to adopt a rime-based coding system. This coding system, accordingly, appears to hinder the model's ability to detect and generalize onset-to-vowel effects. For example the *u* in *quad* and the *u* in *squad* are in different slots, preventing the model from making the generalization that they influence the pronunciation of the following *a* in the same way. In his current work, Zorzi (personal communication, December 6, 2001) has abandoned rime-based coding.

The connectionist models that we examined, although they outperformed the Coltheart et al. (2001) and Zorzi et al. (1998) models on our experimental nonwords, were often more influenced by the vowel's context than were human readers. Connectionist models are designed to pick up those patterns in the input that are most useful for predicting the response. These patterns will involve vowels and following consonants with such words as *mild* and *child*. However, as also suggested by Kay (1987), people may sometimes operate at the grapheme-phoneme level even when that level is not the most predictive of pronunciation. One might hypothesize that training the models on grapheme-phoneme correspondences as well as on whole words, as children are taught in many schools, would improve their ability to account for human data. The model of Plaut and McClelland (1993), which was trained in this way, performed somewhat but not substantially better on the experimental nonwords than the most similar model that did not receive such training (Simulation 3 of Plaut et al., 1996). Nor did the model of Powell et al. (2001), which was also trained on isolated grapheme–phoneme correspondences, perform substantially better than the other connectionist models.

The Norris (1994) model did better than the other models on the experimental nonwords, although worse on the control nonwords. Some of this model's errors on control nonwords, such as /gææ/ for *garx* and /ldɪɪ/ for *ilt*, do not match anything that our readers said, nor were they even consistent with English phonological constraints. Some of the other models contain mechanisms to rule out such responses, including the output filters of Coltheart et al. (2001), the phonological representations of Plaut et al. (1996) that disallow certain illegal sequences, and the phonological cleanup units in the attractor networks of Plaut et al. and Harm and Seidenberg (2002). The inclusion of some such mechanism would likely improve the performance of the Norris model. Another problem with the Norris model for items with consonant clusters is that the smallest unit used by the model is the entire cluster. The above-mentioned error on *garx* appears to arise because the model was not exposed to words with final *rx* and cannot induce a pronunciation for this sequence from experience with words such as *far* and *lax*.

5. General discussion

Our work was motivated by the fact that the links between spellings and sounds in English, when viewed as context-insensitive correspondences between single graphemes and single phonemes, are not very regular. Vowel graphemes, in particular, have multiple pronunciations. How do readers cope with this variability? Analyses of the English writing system (e.g. Kessler & Treiman, 2001) suggest that one possible method would be to consider the vowel's context. By determining whether there are consonants that precede and/or follow the vowel and, if so, their identity, one can improve the likelihood of selecting the correct pronunciation for the vowel. We asked whether experienced readers consider these factors and also whether computational models of reading do so.

The results with human readers may be discussed with respect to two main questions. First, do people take account of the neighboring consonants when pronouncing vowels in nonwords? Second, are people as sensitive to consonant-to-vowel associations as one would expect given the strength of the associations in the real words that they know?

The answer to the first question is clearly "yes". People's pronunciations of vowels are influenced by both onset and coda. Many previous studies have found coda-to-vowel effects in English (e.g. Andrews & Scarratt, 1998; Glushko, 1979; Johnson & Venezky, 1976; Ryder & Pearson, 1980; Treiman et al., 1995; Treiman & Zukowski, 1988; Wolf &

Robinson, 1976), although not all investigators find the data convincing (Coltheart et al., 2001). The present findings strengthen the evidence for coda-to-vowel effects by showing that they are not necessarily whammy effects in disguise, as Coltheart et al. (2001) suggested. The observed onset-to-vowel effects are more surprising. As mentioned earlier, several previous studies have suggested that onsets have little or no influence on adults' pronunciation of vowels (e.g. Andrews & Scarratt, 1998; Treiman et al., 1995; Treiman & Zukowski, 1988). We found, to the contrary, that onsets play a large role in some cases. That these cases are limited – there are few incontrovertible cases beyond those involving onset /w/ – could lead some to question their importance. However, we see the findings as valuable because they show that people pick up those onset-to-vowel associations that exist in the English writing system. The apparent lack of onset-to-vowel influences in the previous studies could be explained in at least two ways. One possible explanation is that people represent spoken syllables as containing separate onset and rime units and cannot learn spelling–sound contingencies that cross the boundary between these constituents. A second possibility is that people are sensitive to onset-to-vowel associations when they exist but that, because these effects are so localized, they are difficult to see in some studies. Our findings support the second explanation. People consider the preceding consonant in pronouncing the vowel in those (relatively rare) cases in which the statistics of the English language make it helpful to do so. Readers can take account of units beyond the rime when pronouncing the vowel. The same appears to be true of spellers (Treiman et al., 2002).

Our findings do not allow us to make definitive statements about the level at which the context effects arise. In the experimental contexts, the pronunciation with the critical vowel results in phoneme strings that are more common than the pronunciation with the typical vowel in all eight cases examined here. For instance, the sequence /ɛd/ (with the critical vowel for coda-to-vowel Case 3) is more common than the sequence /id/ (with the typical vowel). This pattern appeared in the monosyllabic words of the Kessler and Treiman (2001) set and the larger sample of words in the Carnegie Mellon Pronouncing Dictionary (1998); these analyses ignored spellings and examined only pronunciations. In the control contexts, in contrast, pronunciations with the critical vowel tend to be less common than pronunciations with the typical vowel. For example, /ɛp/ is less common than /ip/. It is possible that some readers will select, from multiple legal spelling-to-sound correspondences, the one that yields more typical phoneme strings. This could explain the greater tendency to pronounce *ea* as /ɛ/ before *d* than before *p*. This explanation – that context effects arise at the level of pronunciation rather than at the level of spelling-to-sound translation – is difficult to test with the present stimuli, where effects at the two levels almost always go in the same direction. The most natural mechanism for a pronunciation frequency effect in oral reading would involve feedback, such that the output (pronunciation) feeds back to influence the processing of the input. Thus, it is a potential problem for this view that feedback effects have not been convincingly demonstrated (see Kessler et al., 2003; Peereman, Content, & Bonin, 1998; Stone, Vanhoy, & Van Orden, 1997).

The observed contextual effects, regardless of the level at which they occur, mean that care needs to be taken in the selection of stimuli for experiments. In a number of studies, investigators have attempted to select real words that deviate from the spelling-to-sound patterns of English, or exception words. The vowel pronunciation in a word such as *plaid* is truly irregular. However, the vowel pronunciation in *wand* is not irregular when context is considered, as Venezky (1970) pointed out some time ago and as Kessler and Treiman (2001) confirmed statistically. To determine how often investigators have considered words like *wand* to be exceptional, we searched PsycInfo for journal articles published between 1991 and October 2001 that presented original data on the reading of irregular English words by English-speaking adults and that contained a list of the words used. Of the 17 such articles that we could locate, 9 included one or more words that followed one of the patterns examined here. For example, *wand*, *swarm*, *dread*, *bind*, and *brook* were considered exceptional by some researchers even though the vowel's pronunciation is affected by the preceding and/or following consonant in a systematic way. In most cases, only a few of the stimuli in each study were problematic. In two experiments, however, 19% of the exception words fit one of the patterns examined here (Burt & Humphreys, 1993; Strain & Herdman, 1999).[3] Mixing more predictable words like *wand* with less predictable words like *plaid* may cause researchers to underestimate the effects of unpredictable spelling-to-sound relationships; additional problems may arise if the numbers of such stimuli are not matched across conditions.

Our second question about human readers concerned the degree to which they use consonantal context in generating pronunciations for vowels in nonwords. The results suggest that human readers, on average, are less sensitive to both coda-to-vowel and onset-to-vowel associations than one might expect given the spelling-to-sound patterns in real words. Although previous studies have not made comprehensive quantitative comparisons, there are signs of this phenomenon in earlier reading data (Patterson & Morton, 1985; Wolf & Robinson, 1976), as well as in spelling data (Treiman et al., 2002). Why do people not use consonantal context as much as one might expect? For example, if *-ead* corresponds to /ɛd/ in over two-thirds of the words that a reader knows, why don't readers pronounce *chead* as /tʃɛd/ at a similar rate? One possible explanation is that people's representations of the contingencies in the system – the extent to which a following *d* conditions the /ɛ/ pronunciation of *ea*, for example – are biased or inaccurate. However, results reviewed by Shanks (1995) indicate that associative judgments are generally accurate and unbiased when people have had extensive exposure to a relationship, as with the experienced readers studied here.

There are several potential explanations for readers' apparent underuse of context that will need to be tested in future studies. One possibility is that readers are accustomed to getting by without context in many cases. If the vowel of a real word is assigned its most common context-free pronunciation, the resulting pronunciation, although incorrect, will often be similar to the correct one. For example, *book* will in the first instance be pronounced wrongly, as /buk/, but the correct /bʊk/ may subsequently be retrieved because it is a close phonological neighbor to that nonword. As a result, people may sometimes use the shortcut of ignoring context. Another possibility stems from the observation that use of context sometimes requires people to consider the identity of letters across a fairly long span. Compare *quand*, where context should lead to /ɑ/ pronunciations of the vowel, and

---

[3] The patterns examined in the present study do not exhaust the context-dependent patterns in English. As a result, these values may underestimate the true size of the problem.

quank, where /æ/ is expected. In this case, selection of the appropriate vowel pronunciation requires the consideration of *u*, before the vowel, through to *d*/*k*, three letters later. Readers may prefer not to use such a range of letters, especially since the eye may not take in all of the relevant letters during the same fixation. Yet another possible explanation for the apparent underuse of context is that participants avoid pronunciations for nonwords that are highly similar to pronunciations for existing words. There may be a pull toward /bluk/, in pronouncing *blook*, because this pronunciation better distinguishes the nonword from real words such as *book* and *look*. Finally, and we think quite likely, conflicting pronunciation patterns may contribute to people's apparent underuse of context in pronouncing vowels. For instance, *chead* shares *-ead* with *dead* and *head*, promoting use of /ɛ/, but it also shares units with *cheap* and *meat*, promoting /i/. The rate of /ɛ/ pronunciations, although thus higher than for *cheal*, is not as high as the rate of /ɛ/ pronunciations among real words with *-ead*. This phenomenon could be captured by a model like that of Norris (1994).

This brings us to a consideration of the existing models of oral reading and the extent to which they reproduce the averaged human data. As mentioned earlier, Besner et al. (1990) argued that the first-generation connectionist model of Seidenberg and McClelland (1989) was much poorer than people at pronouncing nonwords. The later connectionist models of Plaut et al. (1996) and Harm and Seidenberg (2003) have, their developers argued, overcome this limitation. Even critics of the earlier models have tended to agree with this assessment (Besner, 1999). However, researchers have generally considered a model's pronunciation of a nonword to be correct if it is consistent with any pronunciation in the training set. For example, either /gruk/ or /grʊk/ would be scored as correct for *grook*. Both pronunciations are justifiable, and certainly a model that yielded /gruk/ or /grʊk/ would be preferred to one that yielded /glæk/. However, a model that yielded /grʊk/ – the pronunciation that the majority of our readers produced – would be more realistic than a model that yielded /gruk/. None of the current models, according to our analyses, do a good job of accounting for the pronunciations that human readers produce on our experimental nonwords. All of the models except for that of Coltheart et al. (2001) capture the overall pattern of performance shown by human readers, in that they produce more critical pronunciations for experimental nonwords as a group than for control nonwords. However, none of the models do a very good job at the level of individual items. This conclusion is similar to the one that Spieler and Balota (1997) drew based on RT data from real words.

Coltheart et al. (2001) listed seven phenomena related to oral reading in adults that any computational model of reading should be able to simulate. For example, any model should account for the fact that reading aloud is faster for high-frequency words than for low-frequency words and for nonwords that sound like real words than for nonwords that do not. We suggest that an eighth benchmark, at least, should be added to the list. Any model of reading should account for the specific pronunciations that people produce and for the fact that their choices for vowels are influenced by both the onset and the coda. We think that this benchmark is as important, if not more important, than the others. The seven phenomena listed by Coltheart et al. all involve differences among stimuli in time to initiate pronunciation. However, RT measures can be problematic from the point of view of models and from the point of view of people. On the model side, there is no direct analogue of naming latency in many models, and assumptions must be made to link the behavior of the model to performance on naming tasks. On the human side, it can be difficult to interpret RT data collected by voice keys (Kessler et al., 2002; Rastle & Davis, 2002). Further complicating interpretation, it is not clear whether people compute the full pronunciation of an item before they start saying it or whether they begin talking when just some of the phonemes have been activated (Kawamoto, Kello, Jones, & Bame, 1998; Rastle, Harrington, Coltheart, & Palethorpe, 2000). The pronunciations that people and models produce can be interpreted more straightforwardly. We acknowledge that the scoring of people's pronunciations is not free of problems: Certain vowel pairs, especially those like /u/ and /ʊ/, can be somewhat difficult to distinguish. However, reliability of scoring was adequate in the present study and could likely be improved further in future studies. Some early researchers did collect fairly detailed information about people's choices of pronunciations for nonwords (e.g. Johnson & Venezky, 1976; Wolf & Robinson, 1976). Regrettably, information about the specific pronunciations that people produce for nonwords (and words) was less likely to be reported as voice keys became more available and as RT became the primary dependent measure for most researchers. This led to the state of affairs summarized by Andrews and Scarratt (1998, pp. 1054–1055): "[T]here is surprisingly little systematic evidence about the pronunciations that people do assign to nonwords." We recommend that researchers collect more such evidence, while also seeking a better understanding of RT measures and of how to analyze them.

If we accept that the ability to explain the pronunciations that people assign to nonwords is an important benchmark for computational models of reading, the conclusion is discouraging. None of the current models, our data show, do a good job of accounting for the contextual effects on vowel pronunciation that are shown by human readers. We have suggested some possible ways of modifying existing models that could be explored. Also worthy of further exploration are nonconnectionist single-route models. In such a model, a unit in a word or nonword would be pronounced by considering the mappings between spellings and sounds that exist in the words that are known to the reader. Relevant mappings involve the unit on its own, the unit in the context of others, and whole words. Information about the frequencies of the spelling-sound mappings in various contexts would combine in some way to yield a pronunciation for the item. The knowledge of sublexical correspondences could be precompiled. Alternatively, it could be computed when needed from appropriately structured lexical representations, yielding a type of analogy model. These ideas are similar to those of Norris (1994). However, as mentioned previously, the Norris model lacks a mechanism to rule out unlikely or impossible pronunciations. Such a mechanism is needed for speech production, where errors rarely yield illegal strings of phonemes, and should arguably be included in models of reading. The Norris model also lacks a way to operate on individual consonants in onset and coda clusters, as discussed earlier. Finally, the Norris model cannot learn patterns in which vowel pronunciation is affected by both onset and coda. Such patterns do exist in English. For example, only by considering both onset and coda can one predict that *wand* is pronounced with /ɑ/ (the onset is *w* and the coda does not contain a velar or *r*), whereas *wag*, *wax*, and *quack* are pronounced with /æ/ (the onset is *w* but the coda is velar). We suspect that people can consider the vowel in the context of both onset and coda, and if so models should too. A model with appropriate weightings for various types of context could potentially explain both aspects of the present results – the effects of preceding and following context on people's pronunciations of vowels and the fact that these effects are smaller than might be expected. Of course, these ideas remain to be tested.

The present results pose a challenge to those who seek to model the reading process. Whether investigators build new models of reading or modify existing ones, we hope that they can rise to this challenge. If so, our understanding of the reading process will be improved.

Acknowledgements

This research was supported by NSF Grants SBR-9807736 and BCS-0130763. Some of these data were presented at the 2002 meeting of the Psychonomic Society. We thank Max Coltheart, Mike Harm, Jason Zevin, Dennis Norris, Dave Plaut, Mark Seidenberg, and Marco Zorzi for making their data and modeling results available to us and for helpful discussions. Thanks to Mark Seidenberg for comments on a draft of the manuscript and to Christine Abadir for her help with the literature search.

Appendix A

A.1. Items testing onset-to-vowel associations

Case 1, Experimental: *squant*, *quab*, *wabs*, *twamp*, *wadge*, *squamp*, *quatch*, *quap*, *guat*

Case 1, Control: *spant*, *clab*, *trabs*, *glamp*, *tadge*, *namp*, *flatch*, *blap*, *trat*

Choices for Experiment 2: *font/rant*, *sob/dab*, *cobs/cabs*, *pomp/camp*, *lodge/badge*, *pomp/camp*, *botch/batch*, *fop/sap*, *hot/hat*

Case 2, Experimental: *warge*, *wark*, *warse*, *warx*, *quarb*, *quarge*, *quarm*, *quarn*, *swarb*, *swark*

Case 2, Control: *carge*, *tark*, *sharse*, *garx*, *darb*, *garge*, *narm*, *starn*, *tarb*, *vark*

Choices for Experiment 2: *forge/large*, *pork/park*, *horse/parse*, *corks/parks*, *orb/barb*, *forge/large*, *storm/farm*, *horn/barn*, *orb/barb*, *pork/park*

A.2. Items testing coda-to-vowel associations

Case 1, Experimental: *blange*, *brange*, *crange*, *drange*, *shange*, *quange*, *sange*, *spange*, *slange*, *snange*

Case 1, Control: *blance*, *brance*, *crance*, *drance*, *shance*, *quance*, *sance*, *spance*, *slance*, *snance*

Choices for Experiment 2: *blade/black*, *brave/brat*, *crave/cram*, *drape/drab*, *shape/shack*, *quake/quack*, *sane/sand*, *space/spam*, *slate/slam*, *snake/snap*

Case 2, Experimental: *yald*, *dald*, *frald*, *fralt*, *talt*, *nald*, *nalt*, *pralt*, *shald*, *tald*

Case 2, Control: *yand*, *dant*, *frand*, *frant*, *tant*, *nand*, *nant*, *prant*, *shand*, *tand*

Choices for Experiment 2: *yawn/yap*, *dawn/dab*, *fraud/frat*, *fraud/frat*, *taut/tab*, *gnaw/nap*, *gnaw/nap*, *prawn/practice*, *shawl/shack*, *taut/tab*

Case 3, Experimental: *clead*, *chead*, *swead*, *glead*, *pread*, *quead*, *splead*, *squead*, *stread*, *yead*

Case 3, Control: *cleam*, *cheal*, *swean*, *gleap*, *preal*, *queam*, *spleab*, *squean*, *streal*, *yeab*

Choices for Experiment 2: *clench/cleat*, *check/cheat*, *swell/sweet*, *glen/glee*, *press/preen*, *quell/queen*, *splendid/spleen*, *squelch/squeeze*, *stress/street*, *yes/yield*

Case 4, Experimental: *ild*, *brild*, *chind*, *crind*, *drind*, *smind*, *shrind*, *slind*, *snild*, *swild*

Case 4, Control: *ilt*, *brilt*, *chint*, *crint*, *drint*, *smint*, *shrint*, *slint*, *snilt*, *swilt*

Choices for Experiment 2: *aisle/id*, *bright/brick*, *chide/chip*, *crime/crib*, *drive/drip*, *glide/glib*, *shrine/shrill*, *slight/slip*, *snide/snip*, *swipe/swim*

Case 5, Experimental: *brold*, *chold*, *crold*, *golt*, *jold*, *nolt*, *polt*, *prold*, *rolt*, *solt*

Case 5, Control: *brond*, *chond*, *crond*, *gont*, *jond*, *nont*, *pont*, *prond*, *ront*, *sont*

Choices for Experiment 2: *broke/bronze*, *choke/chop*, *crone/crop*, *ghost/god*, *joke/john*, *node/nod*, *poke/pod*, *probe/prod*, *robe/rob*, *soak/sock*

Case 6, Experimental: *blook*, *grook*, *clook*, *drook*, *glook*, *prook*, *pook*, *plook*, *slook*, *trook*

Case 6, Control: *bloon*, *groon*, *cloom*, *droon*, *gloon*, *proom*, *poom*, *ploon*, *sloom*, *troon*

Choices for Experiment 2: *pull/bloom*, *pull/grew*, *pull/clue*, *pull/drew*, *pull/gloom*, *pull/moon*, *pull/pool*, *pull/plume*, *pull/slew*, *pull/true*

A.3. Fillers for Experiment 1

bluth, *bripe*, *feg*, *gletch*, *yud*, *korf*, *mobe*, *poin*, *splem*, *reet*, *shig*, *sabe*, *sneff*, *telp*, *troke*, *vay*, *zung*, *glish*, *thruff*, *sploich*

A.4. Fillers (correct choice/incorrect choice) for Experiment 2

blut (*mutt/moot*), *bripe* (*ripe/rip*), *feg* (*leg/ledge*), *gletch* (*etch/each*) *yud* (*mud/mood*), *korf* (*born/burn*), *fope* (*soap/soapy*), *poin* (*coin/con*), *splem* (*peck/peek*), *reet* (*meet/met*), *shig* (*pit/Pete*), *sabe* (*lake/lack*), *sneff* (*pet/pit*), *telp* (*desk/disk*), *troke* (*cope/cup*), *vay* (*sake/sack*), *zung* (*lung/lunge*), *glish* (*lick/leak*), *thruff* (*thread*, *tread*), *sploich* (*Roy/rah*)

A.5. Practice items (correct choice/incorrect choice) for Experiment 2

cat (*bat/but*), *dog* (*front/frog*), *snake* (*rate/rut*), *and* (*cap/cup*), *shook* (*book/beak*)

References

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

Besner, D. (1999). Basic processes in reading: Multiple routines in localist and connectionist models. In R. M. Klein & P. A. McMullen (Eds), *Converging methods for understanding reading and dyslexia* (pp. 413–458). Cambridge, MA: MIT Press.

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

Burt, J. S., & Humphreys, M. S. (1993). Delayed priming of the pronunciation of inconsistent words and pseudowords. *Journal of Memory and Language*, *32*, 743–765.

Carnegie Mellon pronouncing dictionary [Data file]. (1998). Retrieved from Carnegie Mellon University, Speech at CMU Web site: ftp://ftp.cs.cmu.edu/afs/cs.smu.edu/data/anonftp/project/fgdata/dict/cmulex.0.6

Carroll, L. (1977). *Through the looking-glass, and what Alice found there*. New York: St. Martin's Press. (Original work published 1871).

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

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

Harm, M. W., & Seidenberg, M. S. (2002). *Division of labor in a multicomponent connectionist model of reading: I. Computing the meanings of words*, submitted for publication.

Jared, D. (1997). Spelling-sound consistency affects the naming of high-frequency words. *Journal of Memory and Language*, *36*, 505–529.

Jared, D. (2002). Spelling-sound consistency and regularity effects in word naming. *Journal of Memory and Language*, *46*, 723–750.

Johnson, D. D., & Venezky, R. L. (1976). Models for predicting how adults pronounce vowel digraph spellings in unfamiliar words. *Visible Language*, *10*, 257–268.

Kawamoto, A., Kello, C. T., Jones, R., & Bame, K. (1998). Initial phoneme versus whole-word criterion to initiate pronunciation: Evidence based on response latency and initial phoneme duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 862–885.

Kay, J. (1987). Phonological codes in reading: assignment of sub-word phonology. In A. Allport, D. G. MacKay, E. Prinz, & E. Scheerer (Eds.), *Language perception and production: Relations between listening, speaking, reading, and writing* (pp. 182–196). London: Academic Press.

Kessler, B., & Treiman, R. (2001). Relationships between sounds and letters in English monosyllables. *Journal of Memory and Language*, *44*, 592–617.

Kessler, B., Treiman, R., & Mullennix, J. (2002). Phonetic biases in voice key response time measurements. *Journal of Memory and Language*, *47*, 145–171.

Kessler, B., Treiman, R., & Mullennix, J. (2003). *A new look at the factors that affect oral word reading*, submitted for publication.

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

Marcus, M., Santorini, B., & Marcinkiewicz, M. A. (1994). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, *19*, 313–330.

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

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., Plaut, D. C., McClelland, J. L., Seidenberg, M. S., Behrmann, M., & Hodges, J. R. (1996). Connections and disconnections: A connectionist account of surface dyslexia. In J. A. Reggia, E. Ruppin & R. S. Berndt (Eds.), *Neural modeling of brain and cognitive disorders* (pp. 177–199). Singapore: World Scientific.

Peereman, R., Content, A., & Bonin, P. (1998). Is perception a two-way street? The absence of feedback consistency effects in visual word recognition. *Journal of Memory and Language*, *39*, 151–174.

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

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

Powell, D., Plaut, D. C., & Funnell, E. (2001). A developmental evaluation of the Plaut, McClelland, Seidenberg, & Patterson (1996) connectionist model of single word reading [Abstract]. Meeting of the British Psychological Society, Developmental and Education Sections, Worcester, UK.

Rastle, K., & Coltheart, M. (1998). Whammy and double whammy: The effect of length of nonword reading. *Psychonomic Bulletin and Review*, *5*, 277–282.

Rastle, K., & Davis, M. H. (2002). On the complexities of measuring naming. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 307–314.

Rastle, K., Harrington, J., Coltheart, M., & Palethorpe, S. (2000). Reading aloud begins when the computation of phonology finishes. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1232–1235.

Ryder, R. J., & Pearson, P. D. (1980). Influence of type-token frequencies and final consonants on adults' internalization of vowel digraphs. *Journal of Educational Psychology*, *72*, 618–624.

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

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

Shanks, D. R. (1995). *The psychology of associative learning*. Cambridge, UK: Cambridge University Press.

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science*, *8*, 411–416.

Stone, G. O., Vanhoy, M. D., & Van Orden, G. C. (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language*, *36*, 337–359.

Strain, E., & Herdman, C. M. (1999). Imageability effects in word naming: An individual differences analysis. *Canadian Journal of Experimental Psychology*, *53*, 347–359.

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

Treiman, R., Kessler, B., & Bick, S. (2002). Content sensitivity in the spelling of English vowels. *Journal of Memory and Language*, *47*, 448–468.

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

Treiman, R., & Zukowski, A. (1988). Units in reading and spelling. *Journal of Memory and Language*, *27*, 466–477.

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory and Cognition*, *15*, 181–198.

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

Waters, G. S., & Seidenberg, M. S. (1985). Spelling-sound effects in reading: Time course and decision criteria. *Memory and Cognition*, *13*, 557–572.

Wolf, C. G., & Robinson, D. O. (1976). Use of spelling-to-sound rules in reading. *Perceptual and Motor Skills*, *43*, 1135–1146.

Zeno, S. M., Ivenz, S. H., Millard, R. T., & Duvvuri, R. (1995). *Educator's word frequency guide*, Brewster, NY: Touchstone Applied Science Associates.

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