

# Spelling-sound effects in reading: Time-course and decision criteria

GLORIA S. WATERS and MARK S. SEIDENBERG  
*McGill University, Montreal, Quebec, Canada*

In these studies, we examined predictions of the time-course model of word recognition (Seidenberg, 1985b; Seidenberg, Waters, Barnes, & Tanenhaus, 1984). The model suggests that effects of irregular spelling or pronunciation should be specific to more slowly recognized words, such as lower frequency items, as shown in previous studies and replicated here. The model also explains why effects of irregular pronunciation are more robust in naming than in lexical decisions: Only the effects in lexical decisions depend on subjects' response criteria. We show that these criteria are affected by the composition of the stimuli in an experiment (i.e., whether irregularly spelled words are present) and by pressure to respond quickly. In contrast to the dual-route model of word recognition, the time-course model accounts for these task differences without assuming that subjects strategically control access to phonology.

Although reading does not logically require the use of phonological information, research dating at least from Huey (1908/1968) has attempted to determine whether this information is utilized nonetheless. Much of this research has centered on whether recognition of visually presented words is "phonologically mediated." Some researchers (e.g., Coltheart, 1978; Gough, 1972; Rubenstein, Lewis, & Rubenstein, 1971) have suggested that recognition is based on phonological information derived on the basis of knowledge of how an alphabetic orthography represents sound (grapheme-phoneme correspondence rules). Other researchers (e.g., Baron, 1973; Bower, 1970; Kolers, 1970) have argued that readers recognize words on a visual basis, without reference to phonology. A reconciliation of these two views is provided by dual-route models of word recognition (e.g., Coltheart, Davelaar, Jonasson, & Besner, 1977; Meyer, Schvaneveldt, & Ruddy, 1974), in which words can be recognized either on the basis of visual information, or through the use of phonology. These two routes are thought to operate in parallel, with a race between the two processes determining whether phonological mediation occurs.

The dual-route model is consistent with the finding that word recognition is not necessarily phonologically mediated. However, it is not necessary to assume the existence of two recognition processes operating in parallel in order to account for this finding. In several papers, we

have developed a time-course model of word recognition (Seidenberg, 1985a, 1985b; Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Waters, Seidenberg, & Bruck, 1984), based on the work of Glushko (1979), Kay and Marcel (1981), McClelland and Rumelhart (1981), and others. Rather than postulating separate orthographic and phonological processes operating in parallel, as in the dual-route model, the time-course model emphasizes a single interactive process with differences in the availability of orthographic and phonological information over time.

In the time-course model, recognition is initiated with the extraction of visual information from the input. As orthographic units are recognized, they activate their phonological representations. Thus, phonological access lags behind the visual analysis. A word is recognized when the activation of its lexical node passes a given threshold. Whether recognition is "direct" or "phonologically mediated" depends upon the time course of the recognition process. If sufficient orthographic information is extracted from the input to permit recognition prior to the access of phonology, direct access results; however, if a word cannot be recognized prior to the activation of phonological information, mediated access results. In this model, prelexical phonological recoding is neither obligatory nor optional. Rather, phonological effects will occur whenever the duration of processing exceeds the latency of phonological code activation. Thus, phonological effects should be found for lower frequency words and for poorer readers, because the recognition process is relatively slow in both cases.

One strategy used to investigate these issues has been to evaluate the reading of regular and exception words using tasks such as pronunciation and lexical decision. Regular words, such as "must," have predictable pronunciations, because all words that are spelled similarly rhyme (i.e., all words that end in "-ust" rhyme with

This research was supported in part by Grants A7924 and A8325 from the Natural Sciences and Engineering Research Council of Canada, and Grant EQ-2074 from the Quebec Ministry of Education. G. S. Waters, who is also at the School of Human Communication Disorders, was supported by a Social Sciences and Humanities Research Council of Canada postdoctoral fellowship. The assistance of Gail Fornarolo and Sharon Greene in running the experiments is gratefully acknowledged. Responsibility for this work was shared equally by the authors. Send reprint requests to either author at the Psychology Department, McGill University, 1205 Docteur Penfield, Montreal, Quebec, Canada H3A 1B1.

“must”). The pronunciations of exception words, such as “have,” are not predictable on the basis of their spelling patterns, because other words that are spelled similarly do not rhyme (i.e., all other words that end in “-ave” are pronounced as in “gave” and “save”). If phonological information enters into the recognition process, it should interfere with the recognition of exception words. If phonological information does not enter into the recognition process, regular and exception words should be read with equal facility, other factors being equal.

Many researchers have found that pronunciation latencies for exception words are longer than those for regular words (Baron & Strawson, 1976; Coltheart, Besner, Jonasson, & Davelaar, 1979; Glushko, 1979; Stanovich & Bauer, 1978; Underwood & Bargh, 1982). We will term this the “exception effect.” Results from the lexical-decision task have been much more variable. Some studies (e.g., Bauer & Stanovich, 1980; Parkin, 1982) yielded an exception effect with this task, and others (e.g., Andrews, 1982; Coltheart et al., 1979) did not. These inconsistent findings have been taken to indicate that readers can strategically control the use of prelexical phonology. According to this view, skilled readers use flexible decoding strategies and rely upon prelexical phonology only when it facilitates performance on a task.

Seidenberg, Waters, Barnes, and Tanenhaus (1984) proposed that several factors account for the inconsistent effects in these experiments. One factor is frequency. The time-course model suggests that effects of irregular pronunciation should be specific to lower frequency, more slowly recognized words. The frequencies of the stimuli in previous experiments varied widely, and frequency effects were not directly investigated. Experiments that included a large proportion of lower frequency words would be expected to produce larger effects. In experiments in which both high- and low-frequency words were used, the exception effect might have been carried by lower frequency items, although this is difficult to determine, because item statistics are rarely provided. In the only other experiment to directly investigate frequency, Andrews (1982) found an exception effect limited to lower frequency items.

Estimates of the magnitude of the exception effect may have been inflated in some studies because of a second factor, the repetition of spelling patterns in different stimuli with different pronunciations (e.g., Glushko, 1979; Treiman, Freyd, & Baron, 1984). Consider a pair such as “gave” [a regular inconsistent word in Glushko’s (1979) taxonomy] and “have” (an exception word). Seidenberg, Waters, Barnes, and Tanenhaus (1984) found that, when exception words such as “have” were preceded in a list of stimuli by matched regular inconsistent words such as “gave,” exceptions yielded longer lexical-decision latencies than did regular words. When exceptions were not preceded by similarly spelled regular inconsistent words, no exception effect obtained. Hence, there are intralist priming effects when spelling patterns are repeated.

A final factor concerns the criteria used in classifying words as exceptions. Seidenberg, Waters, Barnes, and Tanenhaus (1984) noted that, although some researchers limited the exception category to words such as “have” which have common spelling patterns but irregular pronunciations, other researchers also included words that have both irregular pronunciations and uncommon spelling patterns (words such as “ache” and “yacht” that Seidenberg, Waters, Barnes, & Tanenhaus, 1984, label as “strange”). These words have the effect of confounding regularity of spelling with regularity of pronunciation, so that, if these words are different from regular words, the difference could be due to either factor.

Seidenberg, Waters, Barnes, and Tanenhaus (1984) examined lexical-decision and pronunciation latencies for high- and low-frequency regular and exception words. The exception words included only words with common spelling patterns, and spelling patterns were not repeated within the experiment. The results showed a significant 29-msec exception effect for lower frequency words, but no reliable effect for higher frequency words. On the lexical-decision task, there was no exception effect for either higher or lower frequency words. These findings indicate that, for skilled readers, higher frequency words are recognized on a visual basis, with the phonological code accessed postlexically. For low-frequency words, the phonological code is computed prelexically as a part of the recognition process on the pronunciation task, but not on the lexical-decision task.

Seidenberg, Waters, Barnes, and Tanenhaus (1984) also compared subjects’ pronunciation and lexical-decision latencies for high- and low-frequency regular and strange words and found that, in contrast to the results with exception words, low-frequency strange words yielded longer latencies on both tasks than did low-frequency regular words. Furthermore, the magnitude of the strange effect was larger than that of the exception effect. These results suggested that the exception effect in previous lexical-decision experiments was probably due to the inclusion of orthographically strange words. Hence, the effect was due to irregular spelling, not to irregular pronunciation. Together these studies argue that phonological information plays a role only in the recognition of lower frequency words, and only when the task demands overt pronunciation.

These basic findings have been replicated by Seidenberg (1985b) and by Waters, Seidenberg, and Bruck (1984). Seidenberg (1985b) found additional support for the notion that access to phonological information is dependent upon the time course of processing by dividing the college student subjects in his experiment into fast, medium, and slow reader groups on the basis of their overall pronunciation latencies. This analysis showed that the exception effect for lower frequency words was carried by the subjects who were slower decoders. In a similar vein, Backman, Bruck, Hebert, and Seidenberg (1984), Waters, Bruck, and Seidenberg (1985), and Waters, Seidenberg, and Bruck (1984) found that younger subjects and poorer

readers, who typically have longer pronunciation latencies, showed effects of spelling-sound regularity on the pronunciation task for both high- and low-frequency words. Furthermore, Waters, Seidenberg, and Bruck (1984) showed that, in contrast to the findings with adults on the lexical-decision task, young children show effects of spelling-sound irregularity for both high- and low-frequency words, and older children show effects for lower frequency words on this task.

The time-course model explains these developmental and task differences without assuming that subjects strategically control the use of phonology. The developmental differences result because younger children and poorer readers have longer recognition latencies than adults have, so that phonological information has more time to accrue. The task differences result from the different demands of pronunciation and lexical decision. According to the time-course model, access to phonological information is dependent upon the identification of orthographic information and, thus, lags behind the visual analysis. The pronunciation task requires the subject to process each word to the point at which its pronunciation is known. Thus, phonological effects are greater on the pronunciation task than on lexical decision because subjects cannot respond until they have identified the correct pronunciation. The lexical-decision task does not necessarily require processing to this depth, however. Lexical decision is a signal-detection task in which subjects must establish criteria for deciding whether to respond "word" or "nonword" (Seidenberg, Waters, Sanders, & Langer, 1984). These criteria may vary, depending upon factors such as the discriminability of word and nonword stimuli and instructions to the subject. Smaller phonological effects on lexical decisions result because subjects are able to make lexical decisions based on information available prior to the access to phonology. These differences between the tasks result in greater phonological effects on pronunciation than on lexical decision.

This account of lexical-decision performance provides the basis for reconciling the conflicting results of previous studies. Although Andrews (1982) and Seidenberg, Waters, Barnes, and Tanenhaus (1984, Experiment 3) found no exception effect for lower frequency words in lexical decision, Parkin and Underwood (1983) obtained the effect. Our model suggests that effects in lexical decision will depend on the subjects' decision criteria. Effects of irregular pronunciation will be negligible if subjects' response criteria permit fast responses prior to the activation of phonology. If the decision is difficult, there will be more time for phonological information to enter the parallel activation process, producing inhibition for exception words. The conflicting results of the above studies would be reconciled if it were the case that subjects' decision criteria in fact had varied in this way.

One fact about the design of these studies suggests that this account may be correct. The studies which did not obtain an exception effect in lexical decision (Andrews, 1982; Seidenberg, Waters, Barnes, & Tanenhaus, 1984)

were those in which the stimuli included only regular and exception words. Thus, all of the stimuli contained regular spelling patterns that occur frequently in the lexicon. Parkin and Underwood (1983), however, also included words of the type we have termed "strange." Hence, the presence or absence of strange words seems to have affected the relative latencies of regular and exception words. The suggestion, then, is that including orthographically strange words made the lexical decision difficult, slowing the processing of target words, and allowing more time for phonological information to enter the parallel activation process, yielding the exception effect.

The purpose of the present studies was to test this hypothesis directly. Specifically, when the stimuli consist of regular, exception, and strange words, exception words should take longer to read than regular words, although these effects may be limited to lower frequency items. The irregularly spelled strange words should produce longer latencies than regular words, because the difficulty in processing these words results from early visual processing upon which access to phonology is contingent. When the stimuli consist of only exception and regular words, exception words should be comparable to regular words. In addition, we compared performance on both the pronunciation and lexical-decision tasks. Seidenberg, Waters, Barnes, and Tanenhaus (1984) argue that only the lexical-decision task involves variable response criteria; therefore the composition of the stimuli should have less effect on pronunciation.

In Experiment 1, the stimuli consisted of high- and low-frequency regular, exception, and strange words. This was a replication of the Parkin and Underwood (1983) conditions with closer control over frequency and orthographic regularity. In Experiment 2, the stimuli consisted of the regular and exception words from Experiment 1; in Experiment 3, they consisted of the regular and strange words from Experiment 1. If the inclusion of strange words influences performance on the lexical-decision task, the subjects should show an effect of spelling-sound regularity in Experiment 1, but not in Experiment 2. Furthermore, if performance on the pronunciation task is less influenced by task factors, then subjects should show an effect of spelling-sound regularity on the pronunciation task in both Experiments 1 and 2.

## EXPERIMENT 1

### Method

**Subjects.** Fifty-six McGill University undergraduates volunteered or were paid to participate in the experiment. All were native speakers of English and had normal reading skills. The subjects in subsequent experiments were also drawn from this population. Twenty-eight subjects participated in each task.

**Stimuli.** Six classes of words were created by crossing two factors, frequency (high or low) and word type (regular, exception, or strange). There were 12 monosyllabic words in each class (see Appendix A). The words in each frequency group were matched as closely as possible for frequency and length. The regular and exception words were matched for summed bigram frequency

(Mayzner & Tresselt, 1965). Strange words had lower mean bigram frequencies than the other two classes. Strange words also differed from the other two classes in that there were very few or no monosyllabic words in the English language with the same endings (e.g., "-auge" as in "gauge"). Table 1 presents the median word frequencies, based on the Carroll, Davies, and Richman (1971) word count, and the median bigram frequencies, based on the Mayzner and Tresselt (1965) letter count, for the stimuli.

The same word stimuli were used in the pronunciation and lexical-decision tasks. For the latter task, an additional 36 pronounceable nonwords were included. The nonwords were derived from words by changing initial consonants (e.g., "hope" to "bope"). All nonwords had regular orthographic patterns, none were pseudohomophones, and none were derived from words used in the experiment. These stimuli were interspersed randomly among the word trials. The stimuli for each task were divided into two blocks with words from each class randomly ordered within blocks. A block of practice trials preceded each stimulus block.

**Procedure.** The experiment was controlled by an Apple II microcomputer equipped with a real-time clock and video monitor (Amdek Video 300). The clock operated on a time-base of .001 sec. A hardware modification was performed to eliminate the timing error associated with the 60-Hz scanning rate of the CRT (Reed, 1979). For both tasks, each stimulus was presented singly in lower-case letters in the center of the video display. In the pronunciation task, the subject read the word aloud into a microphone connected to a voice key interfaced to the computer. In the lexical-decision task, the subject indicated whether the stimulus was a word or nonword by pressing microswitches interfaced to the computer. Latencies were timed from the onset of the stimulus to the onset of the subject's response. The stimulus remained on the screen until the subject responded. The intertrial interval was 2 sec. The instructions emphasized both speed and accuracy.

## Results

The data for the pronunciation and lexical-decision tasks were analyzed separately. ANOVAs with both subject and item means as units were computed. For each task, analyses were performed on median reaction times for correct responses and on the number of errors.  $\text{MinF}'$  statistics were calculated, and are reported when significant; otherwise, the significant F statistics for the subject and item analyses are reported. The above procedure was used for each of the experiments reported in this paper.

**Pronunciation task.** The mean of subjects' median reaction times and number of errors is presented in Table 2. Data for individual items are presented in Appendix A. The analyses included the factors, word class (regular, exception, or strange) and frequency (high or low). In the analysis of the latency data, there was a significant main effect of word class in the subject analysis [ $F(2,54) = 34.35, p < .001$ ], as well as a significant

**Table 2**  
Mean of Median Reaction Times (in msec) and  
Mean Number of Errors, Experiment 1

Word Class	Word Frequency	
	Low	High
Pronunciation Task		
Regular	518 (0.57)	526 (0.07)
Strange	638 (2.54)	519 (0.00)
Exception	559 (1.21)	509 (0.11)
Lexical Decision Task		
Regular	606 (0.54)	559 (0.21)
Strange	673 (1.82)	572 (0.14)
Exception	632 (1.64)	552 (0.04)

effect of frequency [ $\text{minF}'(1,88) = 10.80, p < .01$ ] and a frequency  $\times$  word class interaction [ $\text{minF}'(2,76) = 3.66, p < .05$ ]. Newman-Keuls tests, using a pooled error term and Satterthwaite degrees of freedom, showed that pronunciation latencies for both low-frequency exception and strange words were longer than those for low-frequency regular words, and that there were no differences across word classes for high-frequency items.

In the analysis of the error data, there were significant main effects of frequency [ $\text{minF}'(1,92) = 18.56, p < .01$ ] and word class [ $\text{minF}'(2,94) = 3.69, p < .05$ ], as well as a frequency  $\times$  word class interaction [ $\text{minF}'(2,98) = 4.22, p < .05$ ]. Post hoc tests showed that subjects made more errors on both low-frequency exception and strange words than on low-frequency regular words, but that the number of errors across word classes did not differ for high-frequency stimuli.

**Lexical decision task.** Results of the lexical-decision task also are presented in Table 2. The item data are given in Appendix A. Analysis of the reaction times revealed a significant effect of frequency [ $\text{minF}'(1,92) = 23.38, p < .01$ ], as well as a significant word class effect [ $F(2,54) = 15.40, p < .01$ ] and a significant frequency  $\times$  word class effect [ $F(2,54) = 3.46, p < .05$ ], in the subject but not the item analysis. Post hoc tests based on subject means indicated that the frequency  $\times$  word class effect resulted from longer reaction times on low-frequency exception and strange words than on low-frequency regular words, but no differences across word classes for high-frequency items.

In the analysis of the error data, there was a significant effect of frequency [ $\text{minF}'(1,86) = 10.51, p < .01$ ], as well as a main effect of word class [ $F(2,54) = 9.51, p < .001$ ], and a frequency  $\times$  word class interaction [ $F(2,54) = 15.94, p < .001$ ] in the subject but not the item analysis. The post hoc tests indicated that the frequency  $\times$  word class interaction was due to more errors on both low-frequency exception and low-frequency strange words than on low-frequency regular words.

## Discussion

The results from the pronunciation task replicate those of Seidenberg, Waters, Barnes, and Tanenhaus (1984), showing that when subjects were required to pronounce

**Table 1**  
Word and Bigram Frequencies for the Stimuli in Experiment 1

Word Class	Word Frequency	
	Low	High
Regular	21.5 (68.5)	624.5 (79.2)
Strange	17.5 (21.8)	661.0 (26.5)
Exception	29.0 (58.2)	623.5 (71.8)

*Note*—Entries without parentheses are median word frequencies (Carroll et al., 1971). Entries within parentheses are median bigram frequencies (Mayzner & Tresselt, 1965).

words, performance on both low-frequency exception and low-frequency strange words was poorer than that on low-frequency regular words in terms of both latencies and errors. The effects in this experiment are similar in magnitude to those in the Seidenberg et al. study. The latter obtained a 29-msec difference between low-frequency regular and exception words and a 72-msec difference between low-frequency regular and strange words, comparable to the 41-msec and 120-msec differences obtained in this study. In both studies, no differences among word classes were observed with higher frequency stimuli.

The finding of poorer performance on both low-frequency exception and strange words compared to regular words, but of no differences across word classes for high-frequency items, again supports the conclusion that subjects use spelling-sound information when pronouncing low-, but not high-frequency words. Furthermore, the finding of a low-frequency strange effect that was larger in magnitude than the low-frequency exception effect argues that the irregular orthographic patterns in strange words makes them more difficult to process than exception words.<sup>1</sup>

The results from the lexical-decision task are consistent with those of Parkin and Underwood (1983), in which the stimuli also consisted of regular, exception, and strange words. However, they contrast with those of Seidenberg, Waters, Barnes, and Tanenhaus (1984, Experiment 3) and Andrews (1982), in which only regular and exception words were used. Seidenberg et al. found a nonsignificant 3-msec difference between low-frequency exception and regular words, and 2.4% fewer errors on exception words than on regular words. In the present experiment, lower frequency exception words took 26 msec longer than regular words, and produced 9.2% more errors. Lower frequency strange words produced larger effects in both the present experiment and Seidenberg et al.'s Experiment 4 (76 msec in the present study and 40 msec in the Seidenberg et al. experiment). It is important to note that, although the exception effect in the present experiment was similar in magnitude to that in the Parkin and Underwood (1983) study, in the present experiment the effect was only significant when analyzed by subjects, not by items. In contrast, the exception effect with the pronunciation task in the present study is significant when analyzed both by subjects and by items. This pattern reflects systematic differences between the tasks to which we return in the General Discussion.

The exception effect with the lexical-decision task in the present experiment and in Parkin and Underwood (1983), but not in the Andrews (1982) or Seidenberg, Waters, Barnes, and Tanenhaus (1984) experiments, is consistent with the hypothesis that the effect depends on the inclusion of strange words. In order to test this hypothesis, only the regular and exception words were examined in Experiment 2, and only the regular and strange words in Experiment 3. It was expected that removing the strange words would eliminate or greatly reduce the

exception effect in lexical decision but not in pronunciation.

### EXPERIMENTS 2 AND 3

#### Method

**Subjects.** One-hundred-twelve undergraduates volunteered or were paid to participate, 28 for each of the pronunciation and lexical-decision tasks in Experiment 2, and 28 for each task in Experiment 3. Seven of the subjects in Experiment 2 had participated in Experiment 1, one to two months earlier.

**Stimuli.** The stimuli for Experiment 2 were the exception and regular words from Experiment 1. The stimuli for Experiment 3 were the strange and regular words from Experiment 1.

**Procedure.** The procedure was identical to that in the previous study. Subjects either performed the lexical-decision task or pronounced the stimulus word.

#### Results: Experiment 2

**Pronunciation task.** The results of Experiment 2 are summarized in Table 3.<sup>2</sup> Appendix B contains the median reaction time and number of errors for each item. In the ANOVAs based on latency data, there was a main effect of frequency in the subject analysis [ $F(1,27) = 9.17, p < .01$ ], as well as a significant frequency  $\times$  word class interaction [ $\text{min}F'(1,64) = 5.73, p < .05$ ]. Newman-Keuls tests showed that the 31-msec difference between low-frequency regular and exception words was significant, but the difference between the two types of high-frequency words was not.

Analysis of the error data yielded significant main effects of frequency [ $\text{min}F'(1,64) = 5.73, p < .05$ ] and word class [ $F(1,27) = 15.63, p < .01$ ], and a significant frequency  $\times$  word class interaction [ $F(1,27) = 29.15, p < .001$ ] in the subject analysis. Post hoc tests showed that the interaction resulted from more errors on low-frequency exception than on low-frequency regular words.

**Lexical-decision task.** Summary data of the lexical-decision task are presented in Table 3. In the analysis of the latency data, there was a significant main effect of frequency [ $\text{min}F'(1,69) = 16.55, p < .01$ ]. The main effect of word class and the frequency  $\times$  word class interaction were not significant in either the subject or item analyses (all  $p > .25$ ).

Analyses of the error data showed a main effect of frequency [ $\text{min}F'(1,65) = 8.73, p < .01$ ], a main effect of

**Table 3**  
Mean of Median Reaction Times (in msec) and Mean Number of Errors, Experiment 2

Word Class	Word Frequency	
	Low	High
	Pronunciation Task	
Regular	549 (0.36)	552 (0.14)
Exception	580 (1.29)	538 (0.18)
	Lexical Decision Task	
Regular	647 (0.71)	576 (0.14)
Exception	643 (1.64)	563 (0.04)

word class [ $F(1,27) = 8.21, p < .01$ ], and a frequency  $\times$  word class interaction [ $F(1,27) = 12.8, p < .01$ ], in the subject analysis. None of these effects were significant by items. The interaction resulted from more errors on low-frequency exception words than on low-frequency regular words.

### Results: Experiment 3

**Pronunciation task.** The results of Experiment 3 are shown in Table 4. The median reaction time and number of errors for each word in the experiment are presented in Appendix C. In the analysis of the latency data, there was a significant effect of word class in the subject analysis [ $F(1,27) = 20.78, p < .001$ ], as well as a significant effect of frequency [ $\text{min}F'(1,63) = 4.49, p < .05$ ], and a frequency  $\times$  word class interaction [ $\text{min}F'(1,53) = 5.93, p < .05$ ]. The post hoc tests indicated that the interaction was due to longer pronunciation latencies for low-frequency strange words than for low-frequency regular words. In the analysis of the error data, there were significant main effects of frequency [ $\text{min}F'(1,55) = 12.81, p < .01$ ] and word class [ $\text{min}F'(1,60) = 7.45, p < .01$ ], and a frequency  $\times$  word class interaction [ $\text{min}F'(1,60) = 8.83, p < .01$ ]. The interaction resulted from more errors on low-frequency strange words than on low-frequency regular words.

**Lexical-decision task.** In the analysis of the latencies, there was a significant effect of word class  $\times$  subjects [ $F(1,27) = 11.81, p < .01$ ], as well as a significant effect of frequency [ $\text{min}F'(1,65) = 12.24, p < .01$ ], and a frequency  $\times$  word class interaction [ $\text{min}F'(1,56) = 3.85, p < .05$ ]. Post hoc tests indicated that the frequency  $\times$  word class effect was due to longer reaction times on low-frequency strange words than on low-frequency regular words. In the analysis of the error data, there was a significant effect of frequency [ $\text{min}F'(1,57) = 4.02, p < .05$ ], as well as a significant effect of word class [ $F(1,27) = 28.64, p < .001$ ], and a frequency  $\times$  word class interaction [ $F(1,27) = 22.78, p < .001$ ] in the subject analysis. Post hoc analysis of the interaction based on the subject data showed that there were more errors on low-frequency strange words than on low-frequency regular words.

### Discussion

The results of the first three experiments can be summarized simply. First, neither irregular spelling nor

irregular pronunciation influenced the recognition of higher frequency words, replicating the results of Seidenberg, Waters, Barnes, and Tanenhaus (1984). Second, for lower frequency words, irregular spelling (strange words) yielded slower recognition latencies and more errors in both tasks than did regular words, and this effect did not depend on the composition of the stimuli. Third, lower frequency words with irregular pronunciations (exception words) were harder to pronounce than lower frequency words with regular pronunciations (regular words); this effect also did not depend on the composition of the stimuli, and it was significant both by subjects and by items. However, whether lower frequency exception words produced longer lexical-decision latencies than lower frequency regular words depended on the composition of the stimuli. When the stimuli included both irregularly spelled and irregularly pronounced words, an exception effect resulted in both reaction times and errors (Experiment 1), although the effect did not generalize across items. When the stimuli only consisted of regular and exception words, no effect was obtained in reaction time, and a small effect that was only significant in the subject analysis was obtained in the errors (Experiment 2). The effect of the error data should be interpreted with caution: 71% of the errors were due to three words, "wan," "soot," and "caste." This clustering of errors suggests that subjects either did not know the correct pronunciations of these words or were uncertain as to whether they were words.

The results of these experiments suggest that the exception effect with the lexical-decision task depends on the composition of the stimuli. One problem with this conclusion is that the subjects in Experiments 1 and 2 were not identical. However, 7 subjects participated in both experiments, providing a way to examine this question. The data for these 7 subjects indicated that, although they showed a low-frequency exception effect in Experiment 1 (mean low-frequency regular = 636 msec, mean low-frequency exception = 680 msec), they did not show the effect in Experiment 2 (mean low-frequency regular = 704 msec, mean low-frequency exception = 705 msec). Furthermore, the failure to find an exception effect in terms of reaction times in Experiment 2 cannot be attributed to the fact that 7 of the subjects had participated in Experiment 1; the remaining subjects showed no hint of an effect either.

The data support the hypothesis that the different results found by Parkin and Underwood (1983) and Seidenberg, Waters, Barnes, and Tanenhaus (1984) were due to the fact that Parkin and Underwood included the regular, exception, and strange words in the same experiment, and Seidenberg et al. compared performance on the exception and strange words in different experiments. However, there is one remaining difference between the Parkin and Underwood and the Seidenberg et al. studies. The stimuli in the Seidenberg et al. experiment consisted of short, monosyllabic words, and the stimuli in the Parkin experiment included longer, polysyllabic words. It has often been suggested (e.g., McCusker, Hillinger, & Bias, 1981) that readers are more likely to phonologically recode when

Table 4  
Mean of Median Reaction Times (in msec) and  
Mean Number of Errors, Experiment 3

Word Class	Word Frequency	
	Low	High
Pronunciation Task		
Regular	546 (0.36)	565 (0.14)
Strange	646 (2.82)	543 (0.04)
Lexical Decision Task		
Regular	679 (0.43)	631 (0.25)
Strange	789 (1.64)	621 (0.29)

processing longer, more difficult words. It is possible then, that an exception effect would be found in the Parkin and Underwood study even if the strange words were deleted. We tested this possibility in Experiment 4 by replicating their experiment without the orthographically irregular words.

## EXPERIMENT 4

### Method

**Subjects.** Twenty-eight McGill University undergraduates participated in Experiment 4.

**Stimuli.** The stimuli were those used by Parkin and Underwood (1983), except that the orthographically irregular-phonologically irregular (strange) items were excluded. The stimuli consisted of 24 regular words and 24 exception words. The stimuli for the two word classes had been roughly equated for length, frequency, number of syllables, and part of speech. The median word frequency based on the Carroll et al. (1971) word count was 31.5 for the regular words and 24.0 for the exception words. Note that this places the words in the low-frequency category, according to the criteria used both in the above experiments and in Seidenberg, Waters, Barnes, and Tanenhaus (1984). The median bigram frequency based on the Mayzner and Tresselt (1965) letter count was 33.1 and 40.2 for the regular and exception words respectively. The exception words, then, have slightly lower overall frequencies but higher bigram frequencies than the regular words. Forty-eight nonwords were created in the same manner as in the previous experiments and interspersed randomly among the word stimuli. The stimuli were divided into two blocks, and a set of practice words preceded the first block.

**Procedure.** The procedure was identical to that in the previous studies, but only the lexical-decision task was used.

### Results

The mean of subjects' median reaction times was 636 msec for the regular words and 642 msec for the exception words. This 6-msec difference did not approach significance when analyzed by subjects or by items (both  $F_s > 1$ ). The mean number of errors was 1.32 for regular words and 2.42 for exception words. The difference between these means was significant when analyzed by subjects [ $F(1,27) = 15.27, p < .01$ ], but not when analyzed by items [ $F(1,46) = 1.44, p > .24$ ]. Appendix D presents the item statistics. Inspection of the item means showed that 60% of the errors on exception words were accounted for by three items, "dearth," "plait," and "coup," again reflecting subjects' uncertainty as to the correct pronunciations. The data from this study contrast with those of Parkin and Underwood (1983), who found a 20-msec advantage for regular words over exception words that was significant in both the subject and item analyses. These data argue that the exception effect in the Parkin and Underwood (1983) study was dependent upon the inclusion of strange words in the stimulus set.

### Discussion

The data from the first four studies demonstrate that the composition of the stimulus set influenced whether irregular spelling-sound correspondences affected lexical decisions to lower frequency words. In contrast, performance on the pronunciation task yielded a robust excep-

tion effect which was not influenced by the composition of the stimuli. In addition, effects of orthographic regularity were seen on both tasks regardless of the composition of the stimuli.

We turn now to considering the basis for the variable effects on the lexical-decision task. One possibility is that access to phonological information in the lexical-decision task is a strategic option under the subjects' control (Davelaar, Coltheart, Besner, & Jonasson, 1978). According to this view, when the stimuli contain only regular orthographic patterns, phonological information is not accessed. When at least some of the words are difficult to process visually (i.e., strange words are included), subjects use both visual and phonological information in performing the task. According to the time-course model, however, both orthographic and phonological information automatically become available as a word is being processed; access to phonological information is not under strategic control, but rather depends on how long this information has to accrue. If the visual analysis provides enough information for lexical decisions to be made before the phonological process has been completed, no effect of spelling-sound regularity will be evident. If the decision process is slowed (e.g., by the inclusion of orthographically irregular words), effects of spelling-sound regularity begin to emerge. Thus, the effects of spelling-sound irregularity on lexical-decision performance are variable because the composition of the stimuli determines whether subjects are able to make the decision prior to the activation of phonological information. Phonological effects on pronunciation are larger and less variable across items, because the task demands that words be processed to the point at which their pronunciations are known. Subjects cannot change their response criteria based on the discriminability of words and nonwords (Forster, 1979).

At first glance, the hypothesis derived from the time-course model does not seem to be supported by the above data. The lexical-decision latencies in Experiment 1, in which an effect of spelling-sound regularity was found, are actually shorter than those in Experiment 2, in which there was no regularity effect. However, this comparison may be misleading because it involves different subjects. A better test of the time-course hypothesis can be derived by dividing the subjects in Experiment 1 into subgroups of fast and slow responders on the basis of a median split, using each subject's overall reaction time across all six word classes. Dividing the subjects in this manner, it is found that the exception effect is due to slower subjects. The difference between regular and exception words was 2 msec for fast subjects, and 53 msec for slow subjects. These data suggest that the effect of spelling-sound regularity on the lexical-decision task does depend upon the time course of the recognition process.

The time-course model makes similar predictions concerning the pronunciation task. Subjects who are able to pronounce words very quickly should show smaller exception effects. Seidenberg (1985b) reports this result. The faster subjects in his study recognized lower fre-

quency words about 100 msec faster than slower subjects recognized higher frequency words. Only the slower subjects showed the exception effect for lower frequency words. A further test of the time-course hypothesis could be provided by forcing subjects to make lexical decisions much more quickly than usual. This manipulation should eliminate the exception effect. Subjects' response times can be controlled by using a response-deadline technique (Pachella & Pew, 1968; Pachella, Smith, & Stanovich, 1978; Stanovich & Pachella, 1976; Stanovich, Pachella, & Smith, 1977). A criterion time is defined for each subject prior to each block of trials. The subject attempts to respond prior to the deadline and is given feedback on each trial as to his or her success in doing so. With practice, the subject learns to consistently respond faster than the criterion time. Error rates vary depending upon the stringency of the criterion. In this way it is possible to equate reaction times across conditions and to use error rate as an index of processing difficulty.

Stanovich and Bauer (1978) found that this technique eliminated the exception effect on the lexical-decision task. However, the data from their study are difficult to interpret, because it is uncertain whether subjects could reliably perform the word-nonword task within the time allotted. Response latencies were very fast (mean = 337 msec) and the error rates for nonwords were not reported.

In Experiment 5, we replicated the lexical-decision condition of Experiment 1 using the response-deadline technique. The time-course hypothesis predicts that, if subjects are required to respond more quickly, phonological information will not have time to accrue and an exception effect will not result, in contrast to the findings of Experiment 1. However, there may still be a strange effect that is due to the unusual orthographic patterns of strange words as opposed to their irregular pronunciation.

## EXPERIMENT 5

### Method

**Subjects.** Thirty-five McGill University undergraduates were paid to participate in Experiment 5.

**Stimuli and apparatus.** The stimuli consisted of the high- and low-frequency regular, exception, and strange words used in Experiment 1. The nonwords from Experiment 1 were used along with an additional 36 nonwords. The stimuli were presented on an Apple II computer as in previous experiments.

**Procedure.** Subjects were first run on a block of trials without using the response-deadline procedure. As in Experiment 1, the instructions emphasized speed and accuracy. The purpose was to obtain an estimate of subjects' baseline reaction times. Subjects then were told that they were going to be given several additional blocks of practice trials and that the purpose of the practice trials was to find a deadline that was considerably faster than their baseline reaction time but that they could consistently beat. They were told that the ideal deadline was one which resulted in their making errors on approximately 20% of the trials. Subjects practiced with the response-deadline procedure. On each block of trials, they were informed of the deadline and told to look just below where the word had appeared on the screen for feedback as to the speed of their responses. The feedback was either the word "O.K." or the word

"SLOW" depending upon whether they had emitted the response before or after the deadline. Practice trials were run in blocks of 20, with different stimuli appearing in each block. A different deadline was set for the word and nonword trials, because reaction times on the nonwords were consistently longer than those on the word trials. In the first block of practice trials, the experimenter set the deadlines for the words and nonwords at a rate that was approximately 10% quicker than the subject's baseline reaction time. The experimenter varied the deadline, and practice continued until a deadline was found that the subject could consistently beat while producing errors on approximately 20% of the trials. Subjects required from 4 to 15 blocks of practice trials in order to meet these criteria. Once the appropriate deadline was found, the subject was presented with the two blocks of test trials and instructed to continue responding within the deadline. Subjects were not given feedback as to the speed of their responses during the experimental trials.

### Results

Although deadlines were chosen such that subjects would make errors on 20% of the word and nonword trials, subjects' accuracy varied. The data of any subject who made errors on fewer than 10% or on more than 30% of either the word trials or the nonword trials were excluded from the analysis. The rationale was that, if subjects were not making errors on at least 10% of the trials, the deadline set for that subject was not stringent enough. On the other hand, if subjects were making errors on more than 30% of the trials, the results probably reflect response biases rather than competent performance of the task. This procedure resulted in the inclusion of 21 subjects in the analysis. These subjects made an average of 22.4% errors on the word trials (range 14-30%) and 24.8% errors on the nonword trials (range 14-30%).

The results for this experiment are shown in Table 5. The item data are given in Appendix E. Analysis of the reaction times showed significant main effects of frequency [ $F(1,20) = 14.03, p < .01$ ] and of word class [ $F(2,40) = 7.76, p < .01$ ] in the subject analysis, but not by items. The interaction was not significant. The word-class effect was found with post hoc tests to be due to faster reaction times on exception words than on regular words and on exception words than on strange words. In the analysis of the error data, there was a significant main effect of frequency [ $\text{min}F'(1,86) = 10.92, p < .01$ ] but no other significant effects.

### Discussion

Although the stimuli in this experiment were identical to those in Experiment 1, the results differed, in that there was no hint of an exception effect. The lack of an excep-

**Table 5**  
Mean of Median Reaction Times (in msec) and  
Mean Number of Errors, Experiment 5

Word Class	Word Frequency	
	Low	High
	Pronunciation Task	
Regular	463.4 (3.09)	442.9 (1.76)
Strange	476.7 (4.14)	454.4 (2.00)
Exception	450.5 (3.19)	426.9 (1.90)

tion effect in this experiment provides strong support for the notion that the effect in the lexical-decision task depends upon the time course of phonological code activation relative to when subjects are able to respond.

While there was no exception effect, there was some suggestion that strange words were still more difficult to process, because the reaction times for strange words were significantly longer than those for exception words, although this difference failed to reach significance. Because the only difference between exception and strange words is the irregularity of the spellings of the latter (i.e., they both have irregular pronunciations), this finding suggests that strange words are difficult to process, not only because of their irregular pronunciations, but also because of their unusual orthographic patterns.

The results of Experiments 1-5 suggest that the effects of phonology on word recognition are quite limited. The effects on the lexical-decision task are variable and depend upon the composition of the stimuli. The effects on the pronunciation task are most consistent but limited to lower frequency words. An important question concerns the extent to which the findings from these tasks can be generalized to normal reading. Although the effects of spelling-sound regularity on the pronunciation task are consistent, some researchers have argued that the processes involved in performing this task may be very different from those involved in silent reading. Coltheart et al. (1977) argued that pronunciation can be accomplished on a nonlexical basis, that is, on the basis of spelling-sound correspondence rules. This process would not require the subject to activate an entry in lexical memory or to retrieve the word's meaning.

There are numerous logical and empirical difficulties with this proposal. First, the fact that subjects can correctly pronounce exception words shows that they are accessing information stored in lexical memory. Even if longer pronunciation latencies for lower frequency exception words were taken as evidence that subjects had attempted to pronounce the words by applying (incorrect) spelling-sound rules (itself not a necessary assumption; Glushko, 1979; Seidenberg, Waters, Barnes, & Tanenhaus, 1984), such words could be correctly pronounced only by checking information stored in lexical memory. It must be remembered that the pronunciation task requires subjects not merely to produce a pronunciation consistent with their knowledge of orthographic-phonological correspondences in English, but rather to produce the correct pronunciation, given their knowledge of the English lexicon. Regular words are not subject to this objection; it is possible in principle to read them correctly nonlexically. However, it cannot be determined whether a word is regular or not until it has been recognized.

There are some data which support the idea that regular words are pronounced lexically. Several studies (e.g., Glushko, 1979) have shown that regular words are pronounced more quickly than regular nonwords. If regular words were pronounced through a nonlexical process, there should be no advantage for these items over reg-

ular nonwords. The finding that naming latencies for regular words are sensitive to frequency also suggests that they are read lexically. Seidenberg, Waters, Barnes, and Tanenhaus (1984) found a 27-msec advantage for high-frequency compared to low-frequency regular words for adult skilled readers, and Waters, Seidenberg, and Bruck (1984) found effects ranging from 26-461 msec for children. However, some studies have failed to find an advantage for high-frequency words. Waters et al. found that there was a trend for skilled adult readers to pronounce low-frequency words more quickly than high-frequency words, and the above studies did not consistently yield an advantage for high-frequency words over low-frequency words.

There are several possible explanations for the inconsistent effects of frequency on pronunciation. Balota and Chumbley (in press) have shown that the frequency effect in pronunciation may be due, at least in part, to differences in ease of pronunciation. Thus, the frequency effects found with regular words in some studies may be due simply to the fact that the high- and low-frequency words were not equated in terms of difficulty of pronunciation. It also is possible that the inconsistent results across studies are due to the fact that the experiments differed in the extent to which the stimuli used in the high- and low-frequency conditions actually differed in terms of frequency. Gernsbacher (1984) has noted the inconsistent effects of frequency on variables such as orthographic regularity, polysemy, and semantic concreteness. She showed that words that are identical in frequency, according to norms such as Carroll et al. (1971), are not equally familiar to subjects. She suggested that subjects' ratings of experiential familiarity may be better indices of word frequency than norms such as Carroll et al.'s.

Experiment 6 was designed to determine whether there is a frequency effect for regular words when the stimuli differ in terms of frequency as measured by both frequency norms and familiarity ratings. The high- and low-frequency stimuli consisted of matched pairs of rhymes (e.g., "must," "lust") as a way to equate them in terms of factors related to pronunciation.

## EXPERIMENT 6

### Method

**Subjects.** Thirty McGill undergraduates were paid to participate in Experiment 6.

**Stimuli.** The stimuli consisted of 52 matched pairs of high- and low-frequency rhymes (e.g., "must," "lust"). All of the words had regular spelling-sound correspondences. The high- and low-frequency lists were matched so that the words in each list began with the same initial phonemes. The mean word frequencies based on the Carroll et al. (1971) word count were 1,059 for the high- and 3.84 for the low-frequency words, respectively.

Familiarity ratings were obtained from 47 McGill undergraduates, none of whom participated in the pronunciation task. The 104 words were randomly arranged and typed on six pages. The words were typed in lowercase letters down the left-hand margin. Opposite each stimulus word was a 7-point numerical scale with its ends labeled

“very unfamiliar” (1) and “very familiar” (7). Subjects were asked to rate the words in terms of their familiarity. The instructions to the subjects encouraged them to use all 7 points of the rating scale. Mean ratings were computed for each item by averaging across all subjects’ responses. The mean rating for the high-frequency words was 6.24 and that for the low-frequency words was 4.02. A *t* test showed that the low-frequency words were rated as significantly less familiar than the high-frequency words [ $t(102) = 23.49, p < .001$ ]. Thus, the high- and low-frequency stimuli differed not only in terms of frequency, as measured by word-frequency norms, but also in terms of subjects’ ratings of familiarity.

**Procedure.** For the pronunciation task, the stimuli were divided into two lists such that the two members of a rhyme pair did not appear in the same list. The mean word frequencies for the high- and low-frequency words in one list were 1,065 and 3.81, while those for the words in the second list were 1,053 and 3.88. Half of the subjects were tested on list 1 and the other half were tested on list 2. The procedure for the pronunciation task was identical to that in the previous studies.

## Results and Discussion

Separate two-way ANOVAs for the reaction time and error data were carried out with the between-subjects factor of list (1 or 2) and the within-subjects factor of frequency (high or low). The mean of subjects’ median reaction times was 481 for high-frequency words and 497 for low-frequency words. In the reaction-time analysis, there was a significant effect of frequency, in both the subject [ $F(1,28) = 23.74, p < .001$ ] and item [ $F(1,100) = 4.25, p < .05$ ] analyses, because lower frequency words were pronounced more slowly than higher frequency words. There was also a significant effect of list in the item analysis [ $F(1,100) = 12.54, p < .001$ ], and a frequency  $\times$  list interaction in the subject analysis [ $F(1,28) = 21.4, p < .001$ ]. No other effects were significant.

The mean number of errors for high-frequency words was .30, and that for low-frequency words was 1.40. Analysis of the error data yielded a significant effect of frequency [ $\text{min}F'(1,117) = 6.95, p < .05$ ], which resulted from more errors on low- than on high-frequency words. No other effects were significant. The reaction time and error data for the individual items are presented in Appendix F.

The finding of a frequency effect in the present study provides support for the view that even regular words are read lexically. A similar result has been found for other languages, such as Serbo-Croatian (Seidenberg & Vidanovic, 1985), Japanese (Besner, Hildebrandt, & McCann, 1984), and Hebrew (Koriat, in press). These data, together with the finding of an advantage for regular words over orthographically legal nonwords in both English (Glushko, 1979) and Serbo-Croatian (Feldman, 1981), provide evidence for the view that words are named lexically.

## GENERAL DISCUSSION

The results of these studies provide strong support for the time-course model. Effects of irregular pronunciation were specific to lower frequency words. Effects in lexical decision depended on subjects’ decision criteria. When

the stimuli contained strange words, an exception effect occurred in terms of both reaction times and errors, although the effect was only significant by subjects. This effect was eliminated by requiring subjects to respond before a deadline. When the stimuli did not contain strange words, there was no difference between regular and exception words in terms of reaction times and only a small difference in errors. This finding was replicated with two sets of stimuli: short monosyllabic words comparable to those used by Andrews (1982) and Seidenberg, Waters, Barnes, and Tanenhaus (1984) and the longer multisyllabic words used by Parkin and Underwood (1983). In contrast to the lexical-decision task, the pronunciation task produced, in both subject and item analyses, stable exception effects which did not depend upon the composition of the stimuli.

We have suggested that differences between the tasks account for these variable results. Lexical decision is a signal-detection task in which subjects set criteria for responding. These criteria vary depending upon the difficulty of the stimuli and the subjects’ own decoding skills. Whether an effect of irregular pronunciation is observed depends on where these criteria are set relative to the timing of the activation of phonological information. Evidence supporting this view is provided by the finding that, in Experiment 1, the exception effect was carried by those subjects who were slow responders, and by the finding in Experiment 5 that, when subjects were required to respond quickly, the exception effect was eliminated. The pronunciation task does not involve such variable response criteria, and is not affected by the composition of the stimuli. In pronunciation, however, a complementary effect can occur: Subjects who are able to pronounce even lower frequency words very rapidly produce no exception effect (Seidenberg, 1985b).

The variable performance on the lexical-decision task suggests that caution should be exercised in drawing conclusions from a particular set of lexical-decision data about the process of reading normal, connected text. Because readers rarely encounter texts consisting of a homogeneous group of words, all having regular orthographic patterns, it might be concluded that the data from studies including a heterogeneous group of words are more representative of natural reading. There are two difficulties with this proposition. First, it is questionable whether stimuli, such as those in Experiment 1, are characteristic of the words in normal texts. What is unique about those stimuli is that they include a high proportion of irregularly spelled words. It is doubtful that normal texts have as many irregularly spelled items, although this question merits further study of the lexical composition of actual texts. The second problem is that, even if the composition of the stimuli in such experiments resembles that of normal text, the lexical-decision task nonetheless differs in important respects from normal reading (Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Waters, Seidenberg, & Bruck, 1984). Only the lexical-decision task involves the discrimination of words and nonwords. This task-spe-

cific element introduces strategic factors which might restrict the relevance of lexical-decision performance to more natural reading tasks.

This discussion raises serious questions about the utility of the lexical-decision task in studying word recognition. The task is often thought to be a useful tool because the word decision can only be made if the target has been recognized as a word. However, as Coltheart (1978) and Coltheart, Davelaar, Jonasson, and Besner (1977) have pointed out, this is surely false. It is possible to create conditions under which the decision can be made based on a superficial analysis of the input. Consider an experiment in which the nonword stimuli are strings of nonalphabetic characters such as \$\$\$\$\$. It is clear that the composition of the nonword stimuli influences the criteria by which lexical decisions are made (Shulman, Hornak, & Sanders, 1978). Furthermore, lexical decisions are also influenced by postlexical processes that result from the use of contextual information or from subject strategies (Seidenberg, Waters, Sanders, & Langer, 1984; Stanovich & West, 1983); these effects are much greater than in pronunciation. The general point is that lexical decision is a signal-detection task, influenced by factors such as the relative discriminability of word and nonword stimuli, and the biases and expectations of the subjects. The behavior of the subject under these conditions is not well understood, and these factors vary between experiments in an uncontrolled fashion. It is not surprising that the results of lexical-decision experiments are often inconsistent.

The pronunciation task is less subject to these variables because it does not involve a word-nonword discrimination, and because of the requirement that the word be pronounced correctly. As such, it might be considered the task of choice. However, the pronunciation task presents two potential problems of its own. First, it requires overt pronunciation; this appears to be a liability in studies of the use of phonological information in reading, because it requires subjects to access phonological information which might not be used in silent reading. This problem is largely irrelevant, however, if we are correct in concluding that phonological code activation is an automatic consequence of decoding, which depends on the time course of the process, not the task. What needs to be determined is the time course of word-recognition processes under natural reading conditions. A second potential problem, noted above, is that pronunciation might be accomplished on a nonlexical basis, that is, without recognizing the word. However, the finding in Experiment 6 of a frequency effect for regular words suggests that pronunciation does entail lexical access.

In reconciling the conflicting results of previous studies, the time-course model also provides a resolution to the debate over whether word recognition is "direct" or "phonologically mediated." The outcome depends on the timing of phonological code activation relative to when subjects make a particular response. In general, only lower frequency, more slowly recognized words are affected by

phonology.<sup>3</sup> However, because of differences between the tasks, these effects are more reliable in pronunciation than in lexical decision. The effects of spelling-sound irregularity on lexical-decision performance vary because subjects establish variable criteria for their decisions. These criteria depend upon the composition of the stimuli, the subject's strategy for performing the task, and the subject's own decoding skill. As a consequence, the amount of time that elapses before the decision is made varies. With more conservative criteria and slower responses, there is more time for phonological information to be activated. Inhibition in lexical decisions to exception words results because of negative phonological feedback in the parallel activation process. Under some conditions, subjects are able to make decisions prior to the activation of phonological information. Whether effects of irregular pronunciation occur in normal reading will depend then on such factors as the characteristics of the words being read, and the speed with which they are decoded.<sup>4</sup>

#### REFERENCES

- ANDREWS, S. (1982). Phonological recoding: Is the regularity effect consistent? *Memory & Cognition*, **10**, 565-575.
- BACKMAN, J., BRUCK, M., HÉBERT, M., & SEIDENBERG, M. S. (1984). Spelling-sound correspondences in reading: Developmental aspects. *Journal of Experimental Child Psychology*, **38**, 114-133.
- BALOTA, D. A., & CHUMBLEY, J. I. (in press). The locus of word frequency effects in the pronunciation task: Lexical access and/or pronunciation? *Journal of Memory & Language*.
- BARON, J. (1973). Phonemic stage not necessary for reading. *Quarterly Journal of Experimental Psychology*, **25**, 241-246.
- BARON, J., & STRAWSON, C. (1976). Use of orthographic and word-specific knowledge in reading of words aloud. *Journal of Experimental Psychology: Human Perception & Performance*, **2**, 386-393.
- BAUER, D. W., & STANOVICH, K. (1980). Lexical access and the spelling-to-sound regularity effect. *Memory & Cognition*, **8**, 424-432.
- BESNER, D., HILDEBRANDT, N., & MCCANN, R. (1984). *Visual and phonological codes in oral reading of Japanese Kana*. Paper presented at the meeting of the Psychonomic Society, Houston, TX.
- BOWER, T. G. (1970). Reading by eye. In H. Levin & J. P. Williams (Eds.), *Basic studies on reading*. New York: Basic Books.
- CARROLL, J., DAVIES, P., & RICHMAN, B. (1971). *The American Heritage Word Frequency Book*. Boston: Houghton Mifflin.
- COLTHEART, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing*. London: Academic Press.
- COLTHEART, M., BESNER, D., JONASSON, J. T., & DAVELAAR, E. (1979). Phonological recoding in the lexical decision task. *Quarterly Journal of Experimental Psychology*, **31**, 489-507.
- COLTHEART, M., DAVELAAR, E., JONASSON, J. T., & BESNER, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI*. Hillsdale, NJ: Erlbaum.
- DAVELAAR, E., COLTHEART, M., BESNER, D., & JONASSON, J. T. (1978). Phonological recoding and lexical access. *Memory & Cognition*, **6**, 391-402.
- FELDMAN, L. B. (1981). *Visual word recognition in Serbo-Croatian is necessarily phonological*. Haskins Laboratory Reports: Status Report on Speech Research SR-66, 167-202.
- FORSTER, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, NJ: Erlbaum.
- GERNSBACHER, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, **113**, 256-281.

- GLUSHKO, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception & Performance*, *5*, 674-691.
- GOUGH, P. B. (1972). One second of reading. In J. F. Kavanagh & I. T. Mattingly (Eds.), *Language by ear and by eye*. Cambridge, MA: MIT Press.
- HUEY, E. B. *The psychology and pedagogy of reading*. Cambridge, MA: MIT Press, 1968 (original work published 1908).
- KAY, J., & MARCEL, A. J. (1981). One process, not two, in reading aloud: Lexical analogies do the work of nonlexical rules. *Quarterly Journal of Experimental Psychology*, *33*, 397-414.
- KOLERS, P. A. (1970). Three stages of reading. In H. Levin & J. P. Williams (Eds.), *Basic studies on reading*. New York: Basic Books.
- KORIAT, A. (1985). Lexical access for low- and high-frequency words in Hebrew. *Memory & Cognition*, *13*, 37-44.
- MAYZNER, M. S., & TRESSELT, M. E. (1965). Tables of single-letter and digram frequency counts for various word length and letter position combinations. *Psychonomic Science Monograph Supplements*, *1*, 13-32.
- MCCLELLAND, J., & RUMELHART, D. (1981). An interactive-activation model of context effects in letter perception, Part 1: An account of basic findings. *Psychological Review*, *88*, 375-407.
- MCCUSKER, L. X., HILLINGER, M. L., & BIAS, R. G. (1981). Phonological recoding and reading. *Psychological Bulletin*, *89*, 217-245.
- MEYER, D. E., SCHVANEVELDT, R. W., & RUDDY, M. G. (1974). Functions of graphemic and phonemic codes in visual word recognition. *Memory & Cognition*, *2*, 309-321.
- PACHELLA, R. G., & PEW, R. W. (1968). Speed-accuracy tradeoff in reaction time: Effect of discrete criterion times. *Journal of Experimental Psychology*, *76*, 19-24.
- PACHELLA, R. G., SMITH, J. E. K., & STANOVICH, K. E. (1978). Qualitative error analysis and speeded classification. In J. Castellan & F. Restle (Eds.), *Cognitive theory* (Vol. 3), Hillsdale, NJ: Erlbaum.
- PARKIN, A. J. (1982). Phonological recoding in lexical decision: Effects of spelling-to-sound regularity depend upon how regularity is defined. *Memory & Cognition*, *10*, 43-53.
- PARKIN, A. J., & UNDERWOOD, G. (1983). Orthographic versus phonological irregularity in lexical decision. *Memory & Cognition*, *11*, 351-355.
- REED, A. V. (1979). Microcomputer display timing: Problems and solutions. *Behavior Research Methods & Instrumentation*, *11*, 572-576.
- RUBENSTEIN, H., LEWIS, S. S., & RUBENSTEIN, M. A. (1971). Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning & Verbal Behavior*, *10*, 645-657.
- SEIDENBERG, M. S. (1985a). The time-course of information activation utilization in visual word recognition. In D. Besner, T. G. Waller, & G. E. Mackinnon (Eds.), *Reading research: Advances in theory and practice* (Vol. 5). New York: Academic Press.
- SEIDENBERG, M. S. (1985b). The time-course of phonological code activation in two writing systems. *Cognition*, *19*, 1-30.
- SEIDENBERG, M. S., & VIDANOVIC, S. (1985, November). *Word recognition in Serbo-Croatian and English: Do they differ?* Paper presented at the meeting of the Psychonomic Society, Boston, MA.
- SEIDENBERG, M. S., WATERS, G. S., BARNES, M., & TANENHAUS, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning & Verbal Behavior*, *23*, 383-404.
- SEIDENBERG, M. S., WATERS, G. S., SANDERS, M., & LANGER, P. (1984). Pre- and postlexical loci of contextual effects on word recognition. *Memory & Cognition*, *12*, 315-328.
- SHULMAN, H. G., HORNAK, R., & SANDERS, E. (1978). The effects of graphemic, phonemic, and semantic relationships on access to lexical structure. *Memory & Cognition*, *6*, 115-123.
- STANOVICH, K. E., & BAUER, D. (1978). Experiments on the spelling-to-sound regularity effect in word recognition. *Memory & Cognition*, *6*, 410-415.
- STANOVICH, K. E., & PACHELLA, R. G. (1976). The effect of stimulus probability on the speed and accuracy of naming alphanumeric stimuli. *Bulletin of the Psychonomic Society*, *8*, 281-284.
- STANOVICH, K. E., PACHELLA, R. G., & SMITH, J. E. K. (1977). An analysis of confusion errors in naming letters under speed stress. *Perception & Psychophysics*, *21*, 545-552.
- STANOVICH, K. E., & WEST, R. F. (1983). On priming by a sentence context. *Journal of Experimental Psychology: General*, *112*, 1-36.
- TREIMAN, R., FREYD, J. J., & BARON, J. (1984). Phonological recoding and use of spelling-sound rules in reading sentences. *Journal of Verbal Learning & Verbal Behavior*, *22*, 682-700.
- UNDERWOOD, G., & BARGH, K. (1982). Word shape, orthographic regularity and contextual interactions in a reading task. *Cognition*, *12*, 197-209.
- WATERS, G. S., BRUCK, M., & SEIDENBERG, M. S. (1985). Do children use similar processes to read and spell words? *Journal of Experimental Child Psychology*, *39*, 511-530.
- WATERS, G. S., SEIDENBERG, M. S., & BRUCK, M. (1984). Children's and adults' use of spelling-sound information in three reading tasks. *Memory & Cognition*, *12*, 293-305.

## NOTES

1. In the present studies, as in our previous experiments (e.g., Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Waters, Seidenberg, & Bruck, 1984), there is a trend for a small advantage for high-frequency strange words compared to regular. One explanation for this finding is that when a word is high in frequency and orthographically irregular, it is easier to eliminate from the cohort of alternatives. That is, the orthographically irregular word is easier to discriminate from its competitors. When the orthographically irregular word is low in frequency, there is more negative feedback from word-level candidates, making recognition harder.

2. Part of the results from this task were also reported by Seidenberg (1985b, Experiment 2).

3. Strictly speaking, the relative speed of word recognition is relevant only insofar as it is related to the amount of phonological feedback in the parallel activation process. It is possible to slow recognition latencies without affecting the amount of phonological feedback; for example, degrading stimuli or presenting them backward slows overall latencies without altering the basic frequency by regularity interaction (unpublished data from our laboratory). Altering the stimuli in this way slows the recognition of features and letters; however, it does not affect the rest of the parallel activation process.

4. Our current account of the time-course model is quite general and does not, for example, specify the kinds of phonological units that are activated and exactly how they feed into the parallel activation process. A more specific version of the model will have to await empirical evidence concerning issues such as what the relevant units actually are. Nonetheless, the model accounts for a wide range of phenomena having to do with task, frequency, word class, and skill differences. The central point of the model is that orthographic information and phonological information become available at different latencies. This basic property of the model will be preserved whether the relevant phonological units turn out to be phonemes, phoneme-clusters, or entire words.

**Appendix A**  
**Median Reaction Times (in msec) and Number of Errors by Items, Experiment 1**

	Regular		Exception		Strange			
	Pron.	LD	Pron.	LD	Pron.	LD		
Low Frequency								
mode	521 (0)	664 (1)	deaf	453 (0)	570 (0)	gauge	541 (11)	790 (3)
dock	474 (0)	553 (2)	worm	455 (3)	554 (1)	sword	619 (8)	615 (0)
pest	479 (0)	576 (0)	phase	548 (1)	563 (0)	seize	569 (3)	605 (2)
hike	517 (0)	557 (1)	plaid	650 (9)	718 (4)	heir	680 (13)	711 (2)
math	502 (1)	602 (1)	tomb	586 (2)	627 (0)	sieve	629 (10)	689 (8)
greed	559 (0)	655 (0)	soot	632 (4)	837 (11)	chute	669 (5)	689 (4)
chore	724 (13)	614 (1)	wand	491 (0)	640 (4)	choir	601 (2)	590 (0)
grill	489 (0)	573 (0)	sew	619 (4)	585 (2)	brooch	569 (6)	665 (9)
bakes	563 (2)	840 (5)	wan	561 (0)	777 (19)	corps	584 (9)	624 (3)
fern	577 (0)	690 (1)	caste	680 (6)	671 (5)	tsar	744 (3)	881 (19)
tile	504 (0)	546 (0)	steak	582 (2)	538 (0)	aisle	618 (1)	640 (0)
rust	458 (0)	554 (3)	gross	531 (3)	782 (0)	weird	463 (0)	595 (1)
High Frequency								
still	566 (0)	517 (0)	give	489 (0)	562 (0)	earth	441 (0)	511 (0)
feel	574 (0)	536 (1)	says	567 (2)	552 (0)	piece	483 (0)	571 (0)
thin	535 (0)	559 (0)	break	494 (0)	521 (9)	sign	562 (0)	580 (0)
corn	522 (0)	540 (0)	touch	486 (0)	505 (0)	view	560 (0)	581 (0)
nine	498 (0)	565 (3)	lose	491 (0)	573 (0)	knife	518 (0)	541 (0)
race	529 (0)	575 (1)	choose	555 (0)	575 (0)	eye	474 (0)	538 (0)
least	496 (1)	567 (0)	watch	455 (0)	547 (0)	friend	547 (0)	529 (0)
face	534 (0)	555 (0)	heard	513 (0)	543 (0)	once	506 (0)	657 (1)
wake	482 (0)	570 (0)	doll	531 (1)	520 (0)	ghost	585 (0)	594 (0)
these	546 (0)	615 (1)	some	568 (0)	542 (1)	two	488 (0)	670 (3)
beach	534 (1)	553 (0)	wool	508 (0)	557 (0)	climb	526 (0)	577 (0)
shell	559 (0)	542 (0)	wash	441 (0)	529 (0)	tongue	493 (0)	570 (0)

Note – Pron. =Pronunciation; LD=Lexical Decision.

**Appendix B**  
**Median Reaction Times (in msec) and Number of Errors by Items, Experiment 2**

	Regular		Exception		
	Pron.	LD	Pron.	LD	
Low Frequency					
mode	522 (1)	679 (3)	deaf	598 (0)	783 (3)
dock	493 (0)	720 (3)	worm	503 (0)	557 (0)
pest	543 (0)	704 (3)	phase	563 (0)	579 (0)
hike	517 (0)	609 (1)	plaid	480 (11)	673 (2)
math	506 (3)	548 (0)	tomb	607 (3)	576 (0)
greed	580 (0)	612 (1)	soot	696 (7)	762 (11)
chore	570 (3)	580 (0)	wand	505 (0)	687 (5)
grill	526 (0)	582 (1)	sew	608 (3)	594 (2)
bakes	568 (2)	752 (5)	wan	521 (3)	910 (16)
fern	605 (0)	756 (3)	caste	603 (7)	661 (6)
tile	532 (1)	614 (0)	steak	606 (0)	551 (0)
rust	481 (0)	586 (0)	gross	567 (2)	627 (1)
High Frequency					
still	586 (0)	581 (1)	give	544 (0)	582 (0)
feel	541 (0)	553 (0)	says	553 (2)	572 (1)
thin	573 (0)	577 (0)	break	481 (1)	525 (0)
corn	505 (0)	582 (0)	touch	485 (0)	559 (0)
nine	523 (3)	629 (2)	lose	536 (1)	534 (0)
race	488 (0)	548 (0)	choose	566 (0)	509 (0)
least	531 (1)	505 (0)	watch	515 (0)	531 (0)
face	541 (0)	530 (0)	heard	532 (1)	571 (0)
wake	512 (0)	598 (0)	doll	488 (0)	574 (0)

## Appendix B (continued)

	Regular		Exception	
	Pron.	LD	Pron.	LD
these	589 (0)	643 (0)	some	573 (0) 503 (0)
beach	534 (0)	539 (0)	wool	503 (0) 525 (0)
shell	602 (0)	508 (1)	wash	498 (0) 543 (0)

Note – Pron. = Pronunciation; LD = Lexical Decision.

**Appendix C**  
**Median Reaction Times (in msec) and Number of Errors by Items, Experiment 3**

	Regular		Strange	
	Pron.	LD	Pron.	LD
Low Frequency				
mode	546 (1)	750 (0)	gauge	628 (10) 846 (2)
dock	516 (0)	624 (1)	sword	582 (4) 606 (0)
pest	486 (1)	640 (1)	seize	602 (1) 646 (1)
hike	527 (0)	628 (0)	heir	714 (13) 725 (2)
math	515 (0)	623 (0)	sieve	623 (16) 743 (4)
greed	523 (0)	652 (1)	chute	672 (4) 910 (3)
chore	749 (6)	696 (1)	choir	615 (2) 666 (0)
grill	551 (0)	723 (1)	brooch	685 (10) 953 (12)
bakes	739 (2)	769 (4)	corps	581 (16) 770 (2)
fern	610 (0)	777 (2)	tsar	850 (0) 1171 (17)
tile	552 (0)	635 (0)	aisle	611 (3) 739 (2)
rust	515 (0)	632 (1)	weird	514 (0) 611 (1)
High Frequency				
still	688 (0)	585 (0)	earth	504 (0) 594 (0)
feel	568 (0)	624 (0)	piece	495 (0) 607 (0)
thin	553 (0)	660 (1)	sign	561 (0) 757 (4)
corn	516 (0)	601 (0)	view	566 (0) 633 (1)
nine	519 (2)	702 (6)	knife	507 (0) 595 (0)
race	510 (0)	584 (0)	eye	534 (0) 527 (0)
least	567 (1)	609 (0)	friend	595 (0) 605 (0)
face	558 (0)	590 (0)	once	515 (1) 603 (1)
wake	565 (1)	733 (0)	ghost	562 (0) 606 (0)
these	700 (0)	689 (0)	two	552 (0) 614 (1)
beach	557 (0)	589 (0)	climb	553 (0) 612 (0)
shell	573 (0)	556 (0)	tongue	545 (0) 613 (1)

Note – Pron. = Pronunciation; LD = Lexical Decision.

**Appendix D**  
**Median Reaction Times (in msec) and Number of Errors by Item, Experiment 4**

Regular		Exception	
token	612 (0)	massage	617 (0)
spark	584 (0)	dearth	936 (12)
grill	622 (1)	loser	654 (0)
vocal	556 (1)	draught	675 (1)
motive	603 (0)	sloth	751 (7)
plug	603 (1)	bough	828 (7)
insist	627 (1)	plait	832 (18)
feature	570 (0)	coup	773 (11)
cross	667 (1)	vase	680 (0)
market	563 (0)	demon	673 (0)
divine	601 (2)	caste	697 (4)
melon	621 (0)	police	524 (0)
cask	779 (6)	gross	668 (2)
thug	793 (4)	chamber	586 (0)
shack	626 (0)	steady	567 (0)
flute	598 (0)	indict	683 (5)
lathe	831 (6)	bowl	590 (0)

**Appendix D (continued)**

Regular			Exception		
screech	732	(1)	prayer	594	(0)
click	557	(0)	lever	621	(0)
sable	754	(4)	cough	668	(1)
throng	817	(7)	steak	579	(0)
treason	775	(1)	fever	581	(0)
mutton	674	(1)	ballet	668	(0)
slipper	547	(0)	premise	647	(0)

**Appendix E  
Median Reaction Times (in msec) and Number of Errors by Items, Experiment 5**

Regular			Exception			Strange		
Low Frequency								
mode	449	(6)	deaf	411	(3)	gauge	395	(6)
dock	486	(4)	worm	479	(8)	sword	453	(2)
pest	503	(4)	phase	398	(2)	seize	464	(4)
hike	506	(6)	plaid	501	(10)	heir	425	(12)
math	445	(2)	tomb	452	(1)	sieve	508	(10)
greed	449	(3)	soot	391	(9)	chute	470	(9)
chore	412	(5)	wand	469	(7)	choir	484	(0)
grill	494	(5)	sew	499	(6)	brooch	442	(7)
bakes	588	(10)	wan	471	(10)	corps	524	(8)
fern	395	(8)	caste	485	(0)	tsar	505	(20)
tile	421	(9)	steak	435	(1)	aisle	496	(7)
rust	426	(3)	gross	429	(10)	weird	524	(2)
High Frequency								
still	437	(3)	give	455	(4)	earth	401	(1)
feel	442	(3)	says	491	(5)	piece	448	(2)
thin	448	(2)	break	478	(2)	sign	482	(7)
corn	477	(4)	touch	434	(1)	view	429	(6)
nine	498	(7)	lose	437	(3)	knife	471	(4)
race	450	(4)	choose	398	(2)	eye	453	(3)
least	504	(2)	watch	456	(3)	friend	403	(0)
face	449	(2)	heard	437	(2)	once	486	(2)
wake	426	(3)	doll	395	(5)	ghost	497	(1)
these	426	(4)	some	439	(6)	two	502	(7)
beach	409	(3)	wool	460	(6)	climb	461	(4)
shell	403	(1)	wash	396	(1)	tongue	470	(5)

**Appendix F  
Median Reaction Times (in msec) and Number of Errors by Items, Experiment 6**

High Frequency			Low Frequency		
can	476	(0)	ban	474	(0)
that	474	(0)	mat	475	(0)
will	430	(0)	sill	598	(2)
more	418	(0)	fore	546	(0)
made	410	(0)	wade	462	(0)
did	485	(0)	hid	473	(1)
like	427	(0)	hike	507	(0)
how	492	(0)	sow	569	(5)
since	505	(0)	mince	505	(2)
went	444	(2)	lent	482	(0)
last	453	(0)	mast	446	(0)
came	504	(1)	tame	495	(0)
job	467	(0)	sob	506	(0)
house	474	(0)	louse	513	(1)
gave	525	(1)	shave	566	(0)
feet	482	(0)	beet	498	(4)

## Appendix F (continued)

High Frequency		Low Frequency	
saw	522 (0)	claw	547 (0)
least	461 (1)	beast	544 (1)
take	467 (0)	rake	475 (1)
need	435 (0)	weed	477 (0)
best	447 (1)	pest	503 (2)
paid	455 (1)	raid	471 (0)
run	438 (0)	nun	478 (3)
terms	476 (0)	germs	604 (0)
reach	435 (0)	peach	526 (1)
shape	510 (0)	nape	452 (1)
must	459 (0)	lust	445 (0)
had	490 (0)	lad	509 (1)
but	475 (0)	rut	457 (1)
not	451 (1)	jot	503 (0)
him	482 (0)	whim	467 (0)
time	492 (0)	chime	556 (0)
back	525 (0)	tack	496 (0)
men	482 (0)	den	511 (0)
still	567 (0)	gill	528 (0)
long	481 (0)	tong	501 (0)
room	456 (0)	loom	476 (0)
line	456 (0)	thine	528 (1)
part	496 (0)	cart	519 (0)
place	550 (0)	lace	471 (6)
god	581 (1)	plod	506 (1)
feel	551 (0)	heel	472 (0)
side	553 (0)	tide	469 (1)
large	487 (0)	barge	491 (0)
told	526 (0)	fold	489 (0)
white	503 (0)	mite	436 (0)
blue	473 (0)	hue	525 (2)
club	542 (0)	stub	521 (0)
chance	538 (0)	lance	481 (1)
stage	566 (0)	cage	514 (0)
bring	491 (0)	sting	573 (3)
land	459 (0)	bland	441 (1)

(Manuscript received September 17, 1984;  
revision accepted for publication June 10, 1985.)