Journal of Experimental Psychology:
Human Perception and Performance
2006, Vol. 32, No. 4, 1040–1062

0096-1523/06/$12.00 DOI: 10.1037/0096-1523.32.4.1040

Is There a Neighborhood Frequency Effect in English? Evidence From Reading and Lexical Decision

Christopher R. Sears and Crystal R. Campbell
University of Calgary

Stephen J. Lupker
University of Western Ontario

What is the effect of a word's higher frequency neighbors on its identification time? According to activation-based models of word identification (J. Grainger & A. M. Jacobs, 1996; J. L. McClelland & D. E. Rumelhart, 1981), words with higher frequency neighbors will be processed more slowly than words without higher frequency neighbors because of the lexical competition mechanism embodied in these models. Although a critical prediction of these models, this inhibitory neighborhood frequency effect has been elusive in studies that have used English stimuli. In the present experiments, the effect of higher frequency neighbors was examined in the lexical decision task and when participants were reading sentences while their eye movements were monitored. Results suggest that higher frequency neighbors have little, if any, effect on the identification of English words. The implications for activation-based models of word identification are discussed.

Keywords: orthographic neighbors, neighborhood frequency effect

Phenomenologically, reading is an effortless process, and skilled readers seldom experience difficulty identifying individual words. However, the ease and speed with which words are identified is misleading. Word identification is the culmination of a sequence of sophisticated information-processing operations, the details of which have been the focus of much empirical attention. An issue of particular interest in recent years has been how the identification of a word is affected by its orthographic neighbors and the implications for our understanding of the word identification process.

The orthographic neighbors of a word are typically defined as the set of different words that can be created by changing one letter of a word while maintaining letter positions (Coltheart, Davelaar, Jonasson, & Besner, 1977). For example, *item, seem, step,* and *stew* are all orthographic neighbors of *stem*. Research on orthographic neighborhood effects is motivated by the fact that, in activation-based models of word identification, the orthographic neighbors of a word play an important role in the lexical selection (i.e., word identification) process (Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981). According to these models, when a word is presented, the lexical representation of the word and the lexical representations of its orthographic neighbors are activated. Selection of the target word then occurs through a process of competitive inhibition, with the lexical units of the word and its neighbors competing against one another by means of mutually inhibitory connections until the target's lexical unit exceeds a threshold level of activation. The existence of orthographic neighborhood effects in word identification tasks is taken as evidence for the lexical competition mechanism embodied in these models.

The written frequency of a word's orthographic neighbors is especially important in these models. For some words, some, or many, of the neighbors are higher in frequency than the word itself. For example, *stem* has a Kučera and Francis (1967) normative frequency of 29, and the normative frequencies of its higher frequency neighbors (*item, seem,* and *step*) are 54, 229, and 131, respectively (*stew* has a normative frequency of 5 and thus is a lower frequency neighbor). According to activation-based models, higher frequency neighbors, because of their higher resting activation levels, can exert more inhibition on the lexical unit of the target word than can lower frequency neighbors. As a result, the lexical unit of a word with higher frequency neighbors will accumulate activation more slowly than the lexical unit of a word without higher frequency neighbors because of this higher degree of interlexical inhibition. Words with higher frequency neighbors will, therefore, reach an activation threshold more slowly than words without higher frequency neighbors. For word identification tasks, the specific prediction is that words with higher frequency neighbors will be responded to more slowly and less accurately than words without higher frequency neighbors, a phenomenon that is usually referred to as an *inhibitory neighborhood frequency effect* (e.g., Grainger, O'Regan, Jacobs, & Segui, 1989).

Using the lexical decision task, a number of investigators have examined the effect of a word's higher frequency neighbors on its identification time (e.g., Carreiras, Perea, & Grainger, 1997; For-

Christopher R. Sears and Crystal R. Campbell, Department of Psychology, University of Calgary, Calgary, Alberta, Canada; Stephen J. Lupker, Department of Psychology, University of Western Ontario, London, Ontario, Canada.

This research was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) grant and an NSERC postgraduate scholarship. We thank Walter van Heuven for providing us with the software used to implement the multiple read-out model. We also thank Suzie Bisson, Tara Wigington, and Tammy Yacyshen for collecting the subjective frequency ratings used in Experiments 1 and 2. We are indebted to Manuel Perea and Kenneth Forster for the invaluable feedback they provided on a previous version of this article.

Correspondence concerning this article should be addressed to Christopher R. Sears, Department of Psychology, University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada, T2N 1N4. E-mail: sears@ucalgary.ca

ster & Shen, 1996; Grainger, 1990; Grainger & Jacobs, 1996; Grainger et al., 1989; Grainger & Segui, 1990; Huntsman & Lima, 1996; Perea & Pollatsek, 1998; Sears, Hino, & Lupker, 1995; Siakaluk, Sears, & Lupker, 2002). Some of these studies show that lexical decision latencies to low-frequency words with higher frequency neighbors are slower than those to low-frequency words without higher frequency neighbors, as activation-based models would predict. In the original report of this effect (Grainger et al., 1989, Experiment 1), neighborhood frequency was manipulated using words with no neighbors, words with some neighbors but none of higher frequency, words with exactly one higher frequency neighbor, and words with many higher frequency neighbors. Target word frequency was equated across these four conditions. Responses to words with at least one higher frequency neighbor were slower than responses to words with no higher frequency neighbors, but there was no cumulative neighborhood frequency effect (i.e., responses to words with many higher frequency neighbors were no slower than responses to words with exactly one higher frequency neighbor). Inhibitory neighborhood frequency effects have also been reported in perceptual identification tasks (e.g., Carreiras et al., 1997; Grainger & Jacobs, 1996; Grainger & Segui, 1990), a semantic categorization task (Carreiras et al., 1997), a naming task (Carreiras et al., 1997), and tasks in which eye movements are monitored (Grainger et al., 1989; Perea & Pollatsek, 1998).

On the whole, the literature does seem to support this key prediction of activation-based models: that words with higher frequency neighbors will be responded to more slowly (and less accurately) than words without higher frequency neighbors. However, as Andrews (1997) noted, almost all of this support has come from studies in languages other than English—namely, French (Grainger & Jacobs, 1996; Grainger et al., 1989; Grainger & Segui, 1990; Jacobs & Grainger, 1992; see Mathey & Zagar, 2000, for an exception), Spanish (Carreiras et al., 1997), and Dutch (Grainger, 1990; van Heuven, Dijkstra, & Grainger, 1998). In studies that have used English stimuli, exactly the opposite results have typically emerged. That is, there have been a number of reports of null and even facilitory effects (Forster & Shen, 1996; Sears et al., 1995; Sears, Lupker, & Hino, 1999; Siakaluk et al., 2002). In Forster and Shen's (1996) lexical decision experiments, for example, responses to words with higher frequency neighbors were generally faster and less error prone than responses to words without higher frequency neighbors (a facilitory neighborhood frequency effect). This was true in their semantic categorization experiments as well, although Forster and Shen ultimately concluded that higher frequency neighbors do not have any effect on word identification because the neighborhood frequency effect was never statistically significant in their item analyses. Alternatively, Huntsman and Lima (1996) observed an inhibitory neighborhood frequency effect in a post hoc analysis of lexical decision latencies and errors: Responses to words with one higher frequency neighbor and to words with two higher frequency neighbors were slower and more error prone than responses to words without higher frequency neighbors. However, in a subsequent study that also used the lexical decision task, Huntsman and Lima (2002) found no evidence for an inhibitory neighborhood frequency effect. Unfortunately, the interpretation of both of these studies is complicated by the fact that they were designed to test for a cumulative effect of higher frequency neighbors (i.e., words with few higher frequency neighbors vs. words with many higher frequency neighbors), not the basic neighborhood frequency effect (words with higher frequency neighbors vs. words without higher frequency neighbors).

The only reasonably clear demonstration of an inhibitory neighborhood frequency effect in English was reported by Perea and Pollatsek (1998). In their Experiment 1, lexical decision latencies to words with higher frequency neighbors were 26 ms slower than latencies to words without higher frequency neighbors (this difference was significant in a subject analysis but not in an item analysis). In a post hoc analysis, Perea and Pollatsek divided their stimuli into low-frequency words (with normative frequencies less than 10) and medium-frequency words (with normative frequencies of 10 or more but less than 58), and in this analysis the inhibitory neighborhood frequency effect was 42 ms for the low-frequency words and 2 ms for the medium-frequency words. Perea and Pollatsek concluded that inhibitory effects of neighborhood frequency could be observed for English words but only when the words are very low in frequency.

Perea and Pollatsek (1998) also examined the reading times for these words by embedding them in sentences and monitoring the eye movements of participants reading the sentences. They reported that first-fixation durations and gaze durations (the sum of all fixations on the target word) to the words with higher frequency neighbors were not any longer than first-fixation durations and gaze durations to the words without higher frequency neighbors. In Perea and Pollatsek's view, the absence of an inhibitory neighborhood frequency effect on these measures indicates that a word's higher frequency neighbors do not affect early stages of lexical processing. Other analyses revealed, however, that there were effects of neighborhood frequency on two of the spillover variables measured, variables that reflect processing that occurs after the reader has left the target word. Specifically, Perea and Pollatsek reported an effect of neighborhood frequency on the probability of regressing back to the target word (13.5% for words with higher frequency neighbors and 6.9% for words without higher frequency neighbors) and on the duration of the first fixation after the target word fixation (261 ms for words with higher frequency neighbors and 249 ms for words without higher frequency neighbors). Perea and Pollatsek concluded that a word's higher frequency neighbors do not have any direct and immediate effect on reading time but do affect later stages of processing (i.e., after a reader has left a word).

Perea and Pollatsek (1998) also conducted a post hoc analysis to determine whether there were any individual differences among participants in terms of their neighborhood frequency effects. They divided their 24 participants into two groups of 12 participants each: those who regressed back to the target word at least 8% of the time, and those who regressed back to the target word less than 8% of the time. For both groups, there was an inhibitory effect of neighborhood frequency on the probability of regressing back to the target word and on the duration of the first fixation after the target word fixation. However, for the participants with few regressions (whom they termed *less impulsive readers*), there was a significant inhibitory neighborhood frequency effect on gaze durations (15 ms), whereas for the participants with more regressions (*more impulsive readers*), the neighborhood frequency effect was facilitory (12 ms) but not statistically significant. This suggested to Perea and Pollatsek that, for readers who make few regressions while reading, the full inhibitory effect of higher frequency neighbors is not delayed until after the target word has been read; for these readers, higher frequency neighbors have a direct and immediate effect on reading times (and thus affect gaze durations). However, as Perea and Pollatsek (1998) pointed out, "any conclusions must be tempered by the fact that the division of participants into groups was made on the basis of data taken from the reading task rather than on the basis of an independent measure of reading ability" (p. 773). Even so, Perea and Pollatsek's results suggest that there may be important individual differences in the neighborhood frequency effect, a possibility that has been overlooked in all other studies.

The Present Research

As many investigators have noted (e.g., Forster & Hector, 2002; Forster & Shen, 1996; Paap & Johansen, 1994; Perea & Rosa, 2000), the neighborhood frequency effect raises critical theoretical issues, the most important being the role of competitive processes in word identification. Unfortunately, the obvious contrast between the English language experiments and those in other languages (with Perea & Pollatsek, 1998, providing the only data that give any support for the existence of an inhibitory neighborhood frequency effect in English) raises questions about whether the neighborhood frequency variable actually does have the impact that the competitive activation models claim it should. The purpose of the present research was to follow up on Perea and Pollatsek's important study as part of a thorough examination of the effect of higher frequency neighbors on the identification of English words. Both lexical decision and eye movement paradigms were used in the present experiments, as were simulations using activation-based models.

Experiment 1A was a replication of Perea and Pollatsek's (1998) lexical decision experiment, using the identical word and nonword stimuli and the identical task instructions. Like Perea and Pollatsek, we instructed participants to stress accuracy when responding ("Participants were instructed to make their responses as rapidly and as accurately as possible; however, we stressed accuracy in order to avoid shallow processing of the stimuli" [p. 770]; "We also stressed to the participants the accuracy of the responses over speed" [p. 769]). These particular instructions seemed important because, in Perea and Pollatsek's experiment, the effect of neighborhood frequency was confined to those words that were quite low in normative frequency (mean normative frequency = 3.4). Because many of these words were probably unfamiliar to participants, it does seem likely that they might be the words most affected by instructions to avoid errors.

With this in mind, we also thought it was important to replicate Perea and Pollatsek's (1998) experiment using lexical decision instructions that ask for both rapid and accurate responding ("respond as quickly and as accurately as possible"). Accordingly, in Experiment 1B, Perea and Pollatsek's stimuli were presented to a new group of participants, who were given lexical decision instructions that did not stress accuracy over speed (hereafter referred to as standard lexical decision instructions). In Experiment 2, the identical experimental design was used with a new and larger set of word and nonword stimuli, carefully chosen to maximize the likelihood of producing an inhibitory neighborhood frequency effect (according to simulations with Grainger & Jacobs's, 1996, multiple read-out model). In Experiment 2A, participants were instructed to give preference to accuracy over speed when responding (i.e., the instructions used by Perea and Pollatsek), and in Experiment 2B, to respond as quickly and as accurately as possible. Taken together, the results of these four experiments will assess the generalizability of Perea and Pollatsek's lexical decision results across both items and task instructions.

In Experiment 3, the eye movements of a new group of participants were recorded while they read sentences that contained the target words used in Experiments 1 and 2. Experiment 3A used the sentences Perea and Pollatsek (1998) created for their target words (the target words used in the present Experiment 1). In Experiment 3B, a new set of sentences was created for the target words used in Experiment 2. The expectation was that both experiments would produce essentially the same results—namely, no effect of neighborhood frequency on first-fixation durations and gaze durations but an inhibitory neighborhood frequency effect on one or more of the spillover variables. Together these experiments should go some distance toward settling the issue of whether inhibitory neighborhood frequency effects occur in English and the implications of those effects (or lack of effects) for activation-based models of word identification.

In all our experiments, we evaluated the possibility that the neighborhood frequency effect is affected by reading ability. This was accomplished by administering the Author Recognition Test (ART; Stanovich & West, 1989). The ART measures reading experience and is known to be correlated with measures of cognitive and reading ability, in particular comprehension, spelling, and vocabulary skills (Stanovich & Cunningham, 1992; Stanovich & West, 1989). The ART provides an assessment of reading ability independent of reading performance. We administered it in all of our experiments to look for interactions between reading ability and neighborhood frequency effects.

The ART seemed particularly appropriate for our purposes because, as reported by Chateau and Jared (2000), another orthographic neighborhood effect—the neighborhood size effect—was shown to be related to reading experience. Chateau and Jared's participants with high ART scores exhibited a smaller facilitory neighborhood size effect than those with low ART scores. (A facilitory neighborhood size effect is defined as faster responding to words with many orthographic neighbors than to words with few orthographic neighbors; for a review, see Andrews, 1997.) Chateau and Jared also reported that the word frequency effect was smaller for participants with high ART scores. Both of these results suggest that the ART would be a good tool to use when looking for interactions between reading ability and neighborhood frequency effects.

Experiment 1

The word and the nonword stimuli used in this experiment were identical to those used by Perea and Pollatsek (1998). In Experiment 1A, participants were instructed to stress accuracy when responding, the instructions Perea and Pollatsek used in their experiment. In Experiment 1B, participants were given standard lexical decision instructions (i.e., "Respond as quickly and as accurately as possible").

Method

Participants

Eighty University of Calgary undergraduate students participated in the experiment in exchange for partial course credit. Forty participated in Experiment 1A and 40 in Experiment 1B. All were native English speakers and reported normal or corrected-to-normal vision. None participated in more than one of these experiments.

Stimuli

The descriptive statistics for the word stimuli are listed in Table 1. There were 92 words presented in the experiment (66 five-letter words and 26 six-letter words). Half of the words had no neighbors substantially higher in frequency than themselves, and half had at least one higher frequency neighbor (the five- and six-letters words were divided equally among these two neighborhood frequency conditions). For the words with higher frequency neighbors, the highest frequency neighbor differed from the stimulus word at one of the middle letter positions (e.g., for the stimulus word *spice*, the highest frequency neighbor is *space*), although for many of the words there were also higher frequency neighbors that differed from the stimulus word in either the initial or the final letter positions (e.g., for the stimulus word *stork,* one of its higher frequency neighbors is *story*). The mean Kučera and Francis (1967) normative frequency per million words of the highest frequency neighbor of each word was 179.4. For the words without higher frequency neighbors, the mean normative frequency of the highest frequency neighbor of each word was slightly lower than the mean target frequency. Perea and Pollatsek (1998) excluded four five-letter words (*lasso, noose, verve,* and *villa*) from all of their analyses because of high error rates (33% or greater); these words were also excluded from all of our analyses. (In Experiment 1A, the error rates for these words were 52%, 31%, 44%, and 34%, respectively, and in Experiment 1B, 57%, 22%, 62%, and 45%, respectively.)

As can be seen in Table 1, there were 51 low-frequency words and 37 medium-frequency words according to Perea and Pollatsek's (1998) definitions (words with normative frequencies less than 10 were defined as low-frequency words, and words with normative frequencies of 10 or more but less than 58 were defined as medium frequency words). The low-frequency words had a mean Kučera and Francis (1967) normative frequency of 3.4 (range = 0–9) and the medium-frequency words a mean normative frequency of 25.9 (range = 10–58). Table 1 lists the mean number of neighbors for the words in each of the four conditions; the overall mean neighborhood size was 2.2 (range = 0–7). The mean summed positional bigram frequencies are also listed in Table 1; these data were obtained from the English Lexicon Project database (Balota et al., 2002). Bigram frequencies were slightly higher for words with higher frequency neighbors but not significantly so, $F(1, 84) = 1.70$, $p > .10$, $MSE = 1{,}265{,}854.0$. There were no significant correlations between bigram frequency and response latencies or error rates in Experiment 1A or in Experiment 1B. (This was true for a variety of measures of bigram frequency; e.g., the sum of the bigram count, average bigram count). The lack of any effect of bigram frequency on lexical decision latencies and errors is consistent with the findings of Andrews (1992) and other investigators.

We obtained the subjective frequency of each word to provide an alternative measure of word frequency, given that the Kučera and Francis (1967) norms tend to be somewhat unreliable for low-frequency words (Gernsbacher, 1984; Gordon, 1985). In a separate study, 68 undergraduate students, none of whom participated in any of the present experiments, were asked to estimate how frequently they encountered 446 different words in print using a scale from 1 (*very infrequently*) to 9 (*very frequently*). They were instructed that if they did not think that an item was a word, they should give it a rating of zero. The words were three, four, five, and six letters in length and were listed in a random order on six sheets of paper. The 92 words presented in Experiment 1 were included in this list. The mean subjective frequency ratings for the 88 words used in the analyses of Experiment 1 are listed in Table 1.

The subjective frequency ratings were analyzed using a 2 (word frequency: high, low) × 2 (neighborhood frequency: no higher frequency neighbors, higher frequency neighbors) factorial analysis of variance (ANOVA). Not surprisingly, there was a main effect of word frequency, $F_s(1, 67) = 392.51$, $p < .01$, $MSE = 0.28$, $F_i(1, 84) = 42.58$, $p < .01$, $MSE = 0.81$; the medium-frequency words received higher ratings than the low-frequency words (4.0 vs. 2.7). The main effect of neighborhood frequency was significant in the subject analysis, $F_s(1, 67) = 23.76$, $p < .01$, $MSE = 0.20$, but not in the item analysis, $F_i(1, 84) = 1.80$, $p > .10$, $MSE = 0.81$. As can be seen in Table 1, the words with higher frequency neighbors were judged as less frequently encountered than words without higher frequency neighbors. This was true for the low-frequency words, $F_s(1, 67) = 11.70$, $p < .01$, $MSE = 0.16$, $F_i(1, 49) = 1.13$, $p > .10$, $MSE = 0.62$, and for the medium-frequency words, $F_s(1, 67) = 16.05$, $p < .01$, $MSE = 0.18$, $F_i < 1$, and there was no Word Frequency × Neighborhood Frequency interaction (both $F_s < 1$).[1]

A 2 (word frequency: low, medium) × 2 (neighborhood frequency: no higher frequency neighbors, higher frequency neighbors) factorial design was used for each of the experiments. There were 27 low-frequency words without higher frequency neighbors, 28 low-frequency words with higher frequency neighbors, 18 medium-frequency words without higher frequency neighbors, and 19 medium-frequency words with higher frequency

Table 1
Stimulus Characteristics of the Words Used in Experiment 1

Stimulus characteristic	Neighborhood frequency: No HF neighbors	Neighborhood frequency: HF neighbors
Low-frequency words		
Word frequency	3.8	3.0
Subjective frequency	2.8	2.6
Number of letters	5.3	5.3
Number of neighbors	1.0	2.7
Number of HF neighbors	0.0	1.3
Highest-frequency neighbor	1.9	121.1
Bigram frequency	2,592	3,017
Number of stimuli	25	26
Medium-frequency words		
Word frequency	25.2	26.5
Subjective frequency	4.2	3.9
Number of letters	5.2	5.2
Number of neighbors	1.8	3.4
Number of HF neighbors	0.0	1.3
Highest-frequency neighbor	7.9	269.5
Bigram frequency	2,808	3,016
Number of stimuli	18	19

Note. HF = higher-frequency. Highest-frequency neighbor refers to the mean frequency of the highest-frequency neighbor.

[1] The correlation between the subjective frequency ratings we collected and the familiarity ratings from the MRC Psycholinguistic Database (Coltheart, 1981) was .72 ($N = 62$, $p < .01$). For the 62 of the 88 words (70.4%) included in this database, the words with higher frequency neighbors and the words without higher frequency neighbors had very similar familiarity ratings (493 vs. 502), $t(60) = 0.55$, $p > .10$.

neighbors.[2] There were also 92 nonwords presented in each experiment (66 five-letter nonwords and 26 six-letter nonwords; these were the same nonwords used in Perea & Pollatsek's, 1998, experiment), for a total of 184 trials. Note that three of these items were, in fact, words (*cress, morel,* and *nasal*) but were treated as nonwords in all of our data analyses to maintain equivalence between our analyses and Perea and Pollatsek's. The mean neighborhood size of the nonwords was 2.4 (range = 0–9 neighbors).

Apparatus and Procedure

Stimuli were presented on a color video graphics array monitor driven by a Pentium-class microcomputer. The presentation of stimuli was synchronized with the vertical retrace rate of the monitor (14 ms), and response latencies were measured to the nearest millisecond. At a viewing distance of 50 cm, the stimuli subtended a visual angle of approximately 2.1 degrees.

Each trial was initiated by a 1-s 2000-Hz warning tone, after which a fixation point appeared at the center of the video monitor. As in Perea and Pollatsek's (1998) experiment, the fixation point was presented for 500 ms and was then erased, and 200 ms later a word or a nonword was presented (in lowercase letters). Participants indicated the lexicality of stimuli (word or nonword) by pressing one of two buttons on a response box. The participant's response terminated the stimulus display, and the next trial was initiated after a timed interval of 1.5 s.

In Experiment 1A, participants were given the instructions used by Perea and Pollatsek (1998). In Experiment 1B, participants were instructed to respond as quickly and as accurately as possible but were not instructed to stress accuracy. Each participant completed 24 practice trials before the collection of data. The practice stimuli consisted of 12 words and 12 orthographically legal and pronounceable nonwords. (These practice stimuli were not used in the experiment, and the data from these practice trials were not analyzed.) After the practice trials, the participants were provided with feedback as to the mean latency and accuracy of their responses (percentage of errors), and during the experimental trials this information was presented every 32 trials. The order in which the stimuli were presented in the experiments was randomized separately for each participant.

After completing the lexical decision task, each participant completed the ART (Stanovich & West, 1989). The ART involves the presentation of a list of names that includes both popular authors (e.g., Arthur C. Clark) and individuals who are not popular authors (e.g., Roger Farr). Participants are asked to place a checkmark beside the name of each author they recognize. Guesses are taken into account by subtracting incorrect responses from correct responses when calculating an overall score (the maximum possible score is 45). The ART has been used in previous studies that have examined differences in reading skill among postsecondary students (Chateau & Jared, 2000; Jared, Levy, & Rayner, 1999; Unsworth & Pexman, 2003), and in these studies reliable differences between high- and low-scoring individuals have been observed in a variety of word identification tasks, including the lexical decision task.

Results

For the word data, the response latencies of correct responses and the error rates from each experiment were submitted to a 2 (word frequency: low, medium) × 2 (neighborhood frequency: no higher frequency neighbors, higher frequency neighbors) factorial ANOVA. Both subject and item analyses were performed.

Like Perea and Pollatsek (1998), we treated response latencies less than 300 ms or greater than 1,500 ms as outliers, and these were removed from all analyses. For Experiment 1A, 29 response latencies (0.8% of the data) were removed by this procedure, and for Experiment 1B, 26 response latencies (0.7% of the data) were removed. The mean response latencies of correct responses and the mean error rates in Experiments 1A and 1B are shown in Table 2.

Experiment 1A: Perea and Pollatsek's Lexical Decision Instructions

In the analysis of response latencies, there was a main effect of word frequency, $F_s(1, 39) = 166.24$, $p < .01$, $MSE = 918.03$, $F_i(1, 84) = 18.15$, $p < .01$, $MSE = 4{,}381.81$, as medium-frequency words were responded to 62 ms faster than low-frequency words. The main effect of neighborhood frequency was significant in the subject analysis, $F_s(1, 39) = 16.59$, $p < .01$, $MSE = 624.92$, but not in the item analysis, $F_i(1, 84) = 1.03$, $p > .10$, $MSE = 4{,}381.81$. The Word Frequency × Neighborhood Frequency interaction was also significant in the subject analysis, $F_s(1, 39) = 4.10$, $p = .05$, $MSE = 575.08$, but not in the item analysis ($F_i < 1$). As can be seen in Table 2, responses to low-frequency words with higher frequency neighbors were 24 ms slower than responses to low-frequency words without higher frequency neighbors, $F_s(1, 39) = 17.23$, $p < .01$, $MSE = 656.04$, $F_i(1, 49) = 1.95$, $p > .10$, $MSE = 4{,}753.42$. Responses to medium-frequency words with higher frequency neighbors were 9 ms slower than responses to medium-frequency words without higher frequency neighbors, $F_s(1, 39) = 2.61$, $p > .10$, $MSE = 543.96$, $F_i < 1$. These results essentially replicate those of Perea and Pollatsek (1998), in that there was an inhibitory neighborhood frequency effect for the low-frequency words but not for the medium-frequency words.[3]

In the analysis of error rates, there was a main effect of word frequency, $F_s(1, 39) = 25.54$, $p < .01$, $MSE = 20.44$, $F_i(1, 84) = 3.85$, $p = .05$, $MSE = 72.60$. The main effect of neighborhood frequency was not significant (both Fs < 1). There was an interaction between word frequency and neighborhood frequency in the subject analysis, $F_s(1, 39) = 5.10$, $p < .05$, $MSE = 19.29$, but not in the item analysis ($F_i < 1$). The pattern of error rates was virtually identical to that observed by Perea and Pollatsek (1998). For the low-frequency words, slightly more errors were made to words with higher frequency neighbors (7.0%) than to words without higher frequency neighbors (5.1%), $F_s(1, 39) = 2.40$, $p > .10$, $MSE = 30.87$, $F_i < 1$. For the medium-frequency words,

[2] Word length (five letters or six letters) was not incorporated into our analyses or into Perea and Pollatsek's (1998) analyses. The small number of six-letter words made such comparisons problematic (e.g., there were only five six-letter words of medium frequency with higher frequency neighbors and five six-letter words of medium frequency without higher frequency neighbors), and the fact that the words were not matched on normative frequency or any other variable made such comparisons potentially misleading.

[3] The neighborhood frequency effect and the Word Frequency × Neighborhood Frequency interaction were not statistically significant in the items analysis of Perea and Pollatsek's (1998) experiment either. Given the debate over the use of item analyses in these types of experiments, however, we do not regard these facts as particularly important. That is, as a number of researchers have noted, it is seldom the case in these situations that the selection of items is ever random in any sense of the term; on the contrary, the items used in these types of experiments are typically selected because they satisfied an extensive set of criteria, which is certainly the case here. As such, as a number of researchers (Cohen, 1976; Keppel, 1976; Raaijmakers, Schrijnemakers, & Gremmen, 1999; Smith, 1976; Wike & Church, 1976) have argued, items analyses would clearly be inappropriate in the present situation for a number of reasons, not the least of which is their profound negative bias. Nonetheless, for the interested reader we report the item analyses for all of the experiments.

Table 2
Mean Response Latencies (in Milliseconds) and Error Rates (in %) in Experiment 1A (Perea & Pollatsek's, 1998, Lexical Decision Instructions) and Experiment 1B (Standard Lexical Decision Instructions)

	Neighborhood frequency	
Word frequency	No HF neighbors	HF neighbors
Experiment 1A		
Low-frequency words	628 (5.1)	652 (7.0)
Medium-frequency words	574 (3.0)	583 (1.8)
Experiment 1B		
Low-frequency words	616 (6.9)	621 (6.9)
Medium-frequency words	560 (3.4)	563 (2.1)

Note. HF = higher-frequency. Error rates appear in parentheses. In Experiment 1A, the mean response latency for the nonwords was 717 milliseconds (ms), and the mean error rate was 7.2%. In Experiment 1B, the mean response latency for the nonwords was 697 ms, and the mean error rate was 8.4%.

fewer errors were made to words with higher frequency neighbors (1.8%) than to words without higher frequency neighbors (3.0%), $F_s(1, 39) = 4.13$, $p < .05$, $MSE = 7.13$, $F_i < 1$.

Effect of reader skill: High ART score versus low ART score. To assess the effect of reader skill on the neighborhood frequency effect, a median split of the ART scores ($Mdn = 14$, range = 7–34) was used to create two groups of participants: a low-ART group ($M = 10.6$) and a high-ART group ($M = 20.9$), $F_s(1, 38) = 93.43$, $p < .01$, $MSE = 11.66$. For the responses to words, the participants in the low-ART group were slower than those in the high-ART group (638 ms vs. 580 ms), $F_s(1, 38) = 3.31$, $p = .07$, $MSE = 40{,}287.03$, $F_i(1, 84) = 129.09$, $p < .01$, $MSE = 1{,}060.42$, and the low-ART group made more errors than the high-ART group (5.0% vs. 3.3%), $F_s(1, 38) = 4.69$, $p < .05$, $MSE = 24.00$, $F_i(1, 84) = 9.94$, $p < .01$, $MSE = 12.06$. Similarly, for the nonwords, the participants in the low-ART group were slower than those in the high-ART group (757 ms vs. 677 ms), $F_s(1, 38) = 4.13$, $p < .05$, $MSE = 15{,}196.21$, $F_i(1, 91) = 102.07$, $p < .01$, $MSE = 1{,}959.17$, and were more error prone (8.8% vs. 5.7%), $F_s(1, 38) = 4.36$, $p < .05$, $MSE = 23.25$, $F_i(1, 91) = 9.48$, $p < .01$, $MSE = 49.89$. These results are consistent with those of Chateau and Jared (2000), who also found that participants with low ART scores were slower and more error prone in the lexical decision task than those with high ART scores.

With regard to the neighborhood frequency effect, there were no significant interactions involving ART group and neighborhood frequency in the response latency analyses or in the error analyses (all $ps > .10$). Thus, there was no evidence that the neighborhood frequency effect was affected by reader skill. Similarly, there were no significant Word Frequency × ART Group interactions (all $ps > .10$); the word frequency effect was virtually identical for the two groups of participants (60 ms for the participants with low ART scores and 62 ms for those with high ART scores). This result was unexpected because Chateau and Jared (2000) reported a larger word frequency effect for participants with lower ART scores. We return to this finding later because only in the present experiment was the word frequency effect not modulated by reader skill. Thus, this result may be important for explaining the differences among our lexical decision experiments.

Experiment 1B: Standard Lexical Decision Instructions

As would be expected given the different lexical decision instructions, responses to words and to nonwords were faster and slightly less accurate in Experiment 1B than in Experiment 1A (see Table 2). Responses to words were 19 ms faster in Experiment 1B, $F_s < 1$, $F_i(1, 84) = 28.19$, $p < .01$, $MSE = 494.07$, and responses to nonwords were 20 ms faster, $F_s < 1$, $F_i(1, 91) = 26.61$, $p < .01$, $MSE = 714.70$. For nonwords the difference in accuracy was significant, $F_s(1, 78) = 1.05$, $p > .10$, $MSE = 24.24$, $F_i(1, 91) = 4.53$, $p < .05$, $MSE = 13.22$, but for words it was not, $F_s < 1$, $F_i(1, 84) = 1.31$, $p > .10$, $MSE = 11.62$.

In the analysis of response latencies of Experiment 1B, the main effect of word frequency was significant, $F_s(1, 39) = 84.45$, $p < .01$, $MSE = 1{,}529.26$, $F_i(1, 84) = 15.56$, $p < .01$, $MSE = 4{,}587.28$. Low-frequency words were responded to 57 ms slower than medium-frequency words. Neither the main effect of neighborhood frequency, $F_s(1, 39) = 1.29$, $p > .10$, $MSE = 589.23$, $F_i < 1$, nor the Word Frequency × Neighborhood Frequency interaction (both $Fs < 1$) were significant. Separate analyses of the low-frequency and medium-frequency words led to the same conclusion, namely, that responses to words with higher frequency neighbors were no slower than responses to words without higher frequency neighbors (all $Fs < 1$).

In the analysis of error rates, there was a main effect of word frequency, $F_s(1, 39) = 37.01$, $p < .01$, $MSE = 18.36$, $F_i(1, 84) = 3.84$, $p = .05$, $MSE = 94.94$, with more errors to low-frequency words than to medium-frequency words (6.9% vs. 2.7%). Neither the main effect of neighborhood frequency, $F_s(1, 39) = 1.90$, $p > .10$, $MSE = 9.55$, $F_i < 1$, nor the Word Frequency × Neighborhood Frequency interaction, $F_s(1, 39) = 1.39$, $p > .10$, $MSE = 13.82$, $F_i < 1$, were significant.

In contrast to the results of Experiment 1A (and those of Perea & Pollatsek, 1998), there was no inhibitory neighborhood frequency effect in this experiment. Consistent with this conclusion was the interaction between experiment and neighborhood frequency in a combined analysis of the response latencies of Experiments 1A and 1B, $F_s(1, 78) = 4.54$, $p < .05$, $MSE = 607.02$, $F_i(1, 84) = 5.24$, $p < .05$, $MSE = 494.07$. This suggests that, at least for these stimuli, the instructions given to participants determine whether or not there is an inhibitory neighborhood frequency effect.

Effect of reader skill: High ART score versus low ART score. To assess the effect of reader skill on the neighborhood frequency effect, a median split of the ART scores ($Mdn = 15$, range = 5–29) was used to create two groups of participants: a low-ART group ($M = 10.4$) and a high-ART group ($M = 20.1$), $F_s(1, 38) = 74.22$, $p < .01$, $MSE = 12.55$. (Note that the mean ART scores of these two groups were virtually identical to the mean ART scores of the two groups created in Experiment 1A.) As was the case in Experiment 1A, for the responses to words, participants in the low-ART group were slower than those in the high-ART group

(624 ms vs. 555 ms), $F_s(1, 38) = 9.14$, $p < .01$, $MSE = 21{,}063.43$, $F_i(1, 84) = 185.43$, $p < .01$, $MSE = 1{,}265.62$, and made more errors (6.4% vs. 3.2%), $F_s(1, 38) = 14.57$, $p < .01$, $MSE = 28.52$, $F_i(1, 84) = 10.12$, $p < .01$, $MSE = 44.02$. Also like Experiment 1A, for the nonwords, the participants in the low-ART group were slower than those in the high-ART group (742 ms vs. 651 ms), $F_s(1, 38) = 6.59$, $p < .05$, $MSE = 12{,}475.63$, $F_i(1, 91) = 171.39$, $p < .01$, $MSE = 1{,}774.89$, and made more errors (10.1% vs. 6.7%), $F_s(1, 38) = 5.24$, $p < .05$, $MSE = 20.94$, $F_i(1, 91) = 15.56$, $p < .01$, $MSE = 32.49$.

Despite these differences between the two groups, again there was no evidence that the neighborhood frequency effect was modulated by reader skill, as there were no significant interactions involving ART group and neighborhood frequency in the response latency analyses or in the error analyses (all *p*s > .10). However, unlike the situation in Experiment 1A, there was evidence that the word frequency effect was modulated by reader skill, as it was in Chateau and Jared's (2000) study. Specifically, in the analysis of response latencies, there was a significant ART Group × Word Frequency interaction, $F_s(1, 38) = 7.13$, $p < .05$, $MSE = 1{,}321.57$, $F_i(1, 84) = 10.72$, $p < .01$, $MSE = 1{,}265.62$, as the word frequency effect was larger for the low-ART group (73 ms) than for the high-ART group (42 ms). The pattern of errors was consistent with the interaction in the response latencies, although for errors the interaction was not statistically significant, $F_s(1, 38) = 3.23$, $p = .08$, $MSE = 17.37$, $F_i(1, 84) = 1.37$, $p > .10$, $MSE = 44.02$. The low-ART participants made more errors to the low-frequency words (9.1%) than to the medium-frequency words (3.7%), as did the high-ART participants (4.7% vs. 1.7%), but the difference was somewhat larger for the low-ART group.

Discussion

In Experiment 1A, participants were given the same instructions that Perea and Pollatsek (1998) used in their lexical decision experiment. As in Perea and Pollatsek's experiment, in Experiment 1A an inhibitory neighborhood frequency effect was observed for the low-frequency words but not for the medium-frequency words. (Note that in most studies these medium-frequency words, with a mean normative frequency of 25.9, would be considered low-frequency words.) In Experiment 1B, participants were given standard lexical decision instructions that did not stress accuracy, and there was no neighborhood frequency effect for either the low-frequency or the medium-frequency words. In neither experiment was there any evidence that the neighborhood frequency effect was modulated by reader ability.

Taken together, these results support the conclusion that, using Perea and Pollatsek's (1998) stimuli, it is possible to produce an inhibitory neighborhood frequency effect in English, but only when participants are instructed to stress accuracy over speed when making their lexical decisions (Experiment 1A). Nonetheless, the contrast between Experiments 1A and 1B could provide the missing piece of the puzzle in trying to understand why inhibitory neighborhood frequency effects have been elusive when using English stimuli. That is, it is possible that the crucial difference between the prior English experiments and the experiments conducted in French, Dutch, and Spanish is not the language of the stimuli but rather how the participants interpret the task instructions. When participants interpret the instructions as requiring them to respond slowly and carefully, making as few errors as possible, a neighborhood frequency effect can emerge. When they try to balance speed and accuracy, the effect disappears. In fact, as is discussed in greater detail later, there is even a mechanism contained within Grainger and Jacobs's (1996) model (the Σ criterion) that, at least in theory, may allow an account of exactly what was changing as a function of task instructions and how those instructions could have produced the patterns in Experiments 1A and 1B.

If this instruction-based hypothesis is correct, it makes an obvious prediction: The neighborhood frequency effect will be observed with any suitable set of English stimuli if the appropriate lexical decision instructions are used. The obvious way to test this prediction is to create a new set of low- and medium-frequency words with and without higher frequency neighbors and to present these stimuli to two new groups of participants, one group receiving standard lexical decisions instructions (as in Experiment 1B) and the other group instructed to stress accuracy when responding (as in Experiment 1A and in Perea & Pollatsek's, 1998, experiment). This was the strategy used in Experiment 2.

Experiment 2

We had two major constraints to keep in mind when selecting a new set of words for use in Experiment 2. The first was to ensure that words of similar Kučera and Francis (1967) normative frequencies were used, because both Perea and Pollatsek's (1998) results and those of Experiment 1A suggest that only for words with very low normative frequencies (< 10/million words) will there be a neighborhood frequency effect. Our second constraint was to ensure that activation-based models (Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981) would predict an inhibitory neighborhood frequency effect for the new set of items. Accordingly, we first selected what we believed would be an appropriate set of words (based on their Kučera & Francis, 1967, normative frequencies) and then conducted simulations with the interactive activation model (McClelland & Rumelhart, 1981) and the multiple read-out model (Grainger & Jacobs, 1996). The details of our stimulus selection and simulations are described next.

Method

Participants

Eighty-one University of Calgary undergraduate students participated in the experiment in exchange for partial course credit. Forty-one participated in Experiment 2A and 40 in Experiment 2B. All were native English speakers with normal or corrected-to-normal vision. None participated in more than one experiment.

Stimuli

The word stimuli presented in the experiment are listed in the Appendix. One hundred twenty words were presented in the experiment (60 four-letter and 60 five-letter words). Ultimately, five words (*bred, farce, gown, mulch,* and *tote*) were excluded from all data analyses because of high error rates. (In Experiment 2A, the error rates for these words were 32%, 32%, 20%, 50%, and 37%, respectively, and in Experiment 2B, 39%, 37%, 18%, 57%, and 42%, respectively.) The descriptive statistics for the remaining stimuli are listed in Table 3.

Table 3
Stimulus Characteristics of the Words Used in Experiment 2

Stimulus Characteristic	Neighborhood frequency	
	No HF neighbors	HF neighbors
Low-frequency words		
Word frequency	4.8	5.0
Subjective frequency	2.7	2.8
Number of letters	4.5	4.5
Number of neighbors	4.5	4.5
Number of HF neighbors	0.0	1.3
Highest-frequency neighbor	4.8	344.3
Bigram frequency	1939	2129
Number of stimuli	28	28
Medium-frequency words		
Word frequency	29.3	29.4
Subjective frequency	4.4	4.5
Number of letters	4.5	4.5
Number of neighbors	4.4	4.4
Number of HF neighbors	0.0	1.2
Highest-frequency neighbor	12.9	270.9
Bigram frequency	1952	2072
Number of stimuli	30	29

Note. HF = higher-frequency. Highest-frequency neighbor refers to the mean frequency of the highest-frequency neighbor.

Neighborhood frequency and word frequency were factorially manipulated while controlling for neighborhood size (range = 1–11 neighbors, $M = 4.5$). Half of these words had no neighbors substantially higher in frequency than themselves, and half had at least one neighbor of much higher frequency. For the majority of the words with higher frequency neighbors, the highest frequency neighbor differed from the stimulus word at the first letter position or at the last letter position (e.g., for the stimulus word *worm,* the highest frequency neighbor is *work*), although for many of the words there were also higher frequency neighbors that differed from the stimulus word at one of the middle-letter positions (e.g., *warm* is also a higher frequency neighbor of the stimulus word *worm*).

Word frequency was manipulated so as to be consistent with Perea and Pollatsek's (1998) definitions of low-frequency and medium-frequency words (i.e., words with normative frequencies less than 10 were defined as low frequency, and words with normative frequencies of 10 or more but less than 58 were defined as medium frequency). Half of the words were of low frequency, and half were of medium frequency. None of these words were used by Perea and Pollatsek. Note that the mean normative frequencies of the low-frequency words (4.8 for words without higher frequency neighbors and 5.0 for words with higher frequency neighbors) were very similar to the mean normative frequencies of Perea and Pollatsek's low-frequency words (3.8 and 3.0, respectively), and the same was true for the medium-frequency words. Also note that, unlike Perea and Pollatsek's word frequency manipulation, our word frequency manipulation was planned, and so we were better able to control neighborhood size (mean number of neighbors) and neighborhood frequency (mean frequency of the highest frequency neighbor) across the four word conditions.

Also listed in Table 3 are the mean summed positional bigram frequencies. Bigram frequencies were slightly higher for the words with higher frequency neighbors, although not significantly so ($F < 1$). There were no significant correlations between bigram frequency and response latencies or error rates in Experiment 2A or Experiment 2B.

After selecting these words, as in Experiment 1, we obtained the subjective frequency of each word to provide an alternative measure of word frequency and to check whether the words selected were of equivalent familiarity to the words used in Experiment 1. In a separate study, 66 undergraduate students, none of whom participated in any of the present experiments, were asked to estimate how frequently they encountered 380 different words in print, using a scale from 1 (*very infrequently*) to 9 (*very frequently*). They were instructed that if they did not think that an item was a word, they should give it a rating of zero. The words were four, five, and six letters in length and were listed in a random order on five sheets of paper. The 120 words presented in Experiment 2 were included in this list. The mean subjective frequency ratings for the 115 words used in the analyses of Experiment 2 are listed in Table 3.

An analysis of these ratings produced a main effect of word frequency, $F_s(1, 65) = 419.80$, $p < .01$, $MSE = 0.45$, $F_i(1, 111) = 67.47$, $p < .01$, $MSE = 1.21$, as the low-frequency words were judged to be less frequently encountered than the medium-frequency words. There was no main effect of neighborhood frequency, $F_s(1, 65) = 3.10$, $p = .08$, $MSE = 0.14$, $F_i < 1$, nor was there a Word Frequency × Neighborhood Frequency interaction (both Fs < 1). There was no effect of neighborhood frequency for the low-frequency words, $F_s(1, 65) = 1.25$, $p > .10$, $MSE = 0.16$, $F_i < 1$, or for the medium-frequency words, $F_s(1, 65) = 1.61$, $p > .10$, $MSE = 0.14$, $F_i < 1$. The mean subjective frequency of the low-frequency words (2.7) was identical to the mean subjective frequency of Perea and Pollatsek's (1998) low-frequency words, as assessed in the present Experiment 1, and the mean subjective frequency of the medium-frequency words (4.4) was very similar to the mean subjective frequency of Perea and Pollatsek's medium-frequency words (4.1). These words were, therefore, of equivalent familiarity to those that Perea and Pollatsek used.[4]

The nonword stimuli were orthographically legal and pronounceable (*gark*) and were matched to the words on length (60 four-letters nonwords and 60 five-letter nonwords). Most of the nonwords had large neighborhoods ($M = 9.0$ neighbors, range = 3–18). This created a difficult word–nonword discrimination, as nonwords with large neighborhoods are orthographically very similar to real words and thus generate a great deal of lexical activity, making it difficult to distinguish them from words on the basis of this activity (Grainger & Jacobs, 1996; Siakaluk et al., 2002).

Simulations

For the simulations with the interactive activation model (McClelland & Rumelhart, 1981), word identification latencies were simulated by noting the number of processing cycles required for a word's lexical unit to reach the activation threshold.[5] For the words used in Experiment 2, the mean number of processing cycles required to reach the activation threshold are shown in Table 4. These data were submitted to a 2 (word frequency: high, low) × 2 (neighborhood frequency: no higher frequency neighbors, higher frequency neighbors) factorial ANOVA. There were main effects of word

[4] The correlation between the subjective frequency ratings we collected and the familiarity ratings from the MRC Psycholinguistic Database (Coltheart, 1981) was .64 ($N = 73$, $p < .01$). Familiarity ratings for 73 of the 115 words were available (63.4%). For these 73 words, the words with higher frequency neighbors and the words without higher frequency neighbors had very similar familiarity ratings (526 vs. 512), $t(71) = 1.17$, $p > .10$.

[5] The parameter values used by Grainger and Jacobs (1996) were also used here, including setting the activation threshold (the M criterion in the multiple read-out model) to 0.67 and setting the letter to word excitation parameter to 0.07 for four-letter words, 0.06 for five-letter words, and 0.055 for six-letter words (see Grainger & Jacobs, 1996, for their rationale). The four-letter and five-letter lexicons used in the simulations consisted of words with Kučera and Francis (1967) frequencies greater than zero. The four-letter lexicon consisted of 1,580 words and the five-letter lexicon, 2,127 words.

Table 4
Mean Number of Processing Cycles for the Word Stimuli Used in Experiments 1 and 2

Word Frequency	Neighborhood frequency: No HF neighbors	Neighborhood frequency: HF neighbors
Experiment 1		
Low-frequency words	16.37 (.33)	18.35 (.42)
Medium-frequency words	16.51 (.38)	17.85 (.44)
Experiment 2		
Low-frequency words	17.39 (.36)	18.33 (.38)
Medium-frequency words	17.14 (.37)	17.79 (.39)

Note. HF = higher-frequency. Mean summed lexical activity appears in parentheses.

frequency, $F(1, 111) = 21.25$, $p < .01$, $MSE = .22$, and neighborhood frequency, $F(1, 111) = 83.89$, $p < .01$, $MSE = .22$, but no interaction, $F(1, 111) = 2.67$, $p > .10$, $MSE = .22$. Thus, for both the low-frequency and medium-frequency words, the model predicts slower responses to words with higher frequency neighbors than to words without higher frequency neighbors (i.e., inhibitory neighborhood frequency effects), $F(1, 54) = 41.35$, $p < .01$, $MSE = .30$, and $F(1, 57) = 44.95$, $p < .01$, $MSE = .14$, respectively.

For completeness, we also conducted simulations with the words from Experiment 1. Table 4 lists the mean number of processing cycles for those words. This analysis revealed a main effect of neighborhood frequency, $F(1, 84) = 46.80$, $p < .01$, $MSE = 1.25$; there was no main effect of word frequency ($F < 1$) and no interaction, $F(1, 84) = 1.72$, $p > .10$, $MSE = 1.25$. That is, the model predicts an inhibitory neighborhood frequency effect for both the low-frequency and the medium-frequency words but no effect of word frequency. Both of these predictions are at odds with the results of Experiment 1, because a neighborhood frequency effect was observed only for the low-frequency words (and only in Experiment 1A) and there was a substantial word frequency effect in Experiment 1A (62 ms) and in Experiment 1B (57 ms).

The multiple read-out model (Grainger & Jacobs, 1996) is based on the architecture of the interactive activation model and, like the interactive activation model, incorporates competitive inhibition among the lexical units of a word and its neighbors during the lexical selection process. Where the two models differ is that the multiple read-out model incorporates three decision criteria (rather than one) that influence the speed with which lexical decision responses are made. The first is the M criterion, which is sensitive to the activation of single lexical units. According to the model, when the M criterion is exceeded, lexical selection has occurred (i.e., a specific word has been identified). (The M criterion corresponds to the activation threshold in the interactive activation model.) The second is the Σ criterion, which is sensitive to the degree of overall lexical activation (i.e., the total lexical activity generated by the word and its neighbors). If a letter string generates enough lexical activity to exceed the current Σ criterion, then a word response can be made before lexical selection (i.e., before the M criterion being exceeded). The third criterion is the T criterion, which is a temporal deadline used for generating nonword responses. According to the model, when a letter string is presented and either the M criterion or the Σ criterion is reached before the T criterion, then a word response will be made; otherwise, a nonword response will be made.

When responding is based on the M criterion rather than the Σ criterion, the multiple read-out model and the interactive activation model make the same predictions for words; it is only when the Σ criterion comes into play that the models make different predictions. Although the position of the Σ criterion is under strategic control, its placement, and hence its impact, essentially depends on the distributions of lexical activation generated by the words and the nonwords. When the words and the nonwords generate a similar degree of lexical activation, its usefulness would be quite limited, and hence it would be placed at a position where it would, essentially, play no role. In that situation, responding will be driven by the M criterion and, as with the interactive activation model, an inhibitory neighborhood frequency effect would be predicted. Alternatively, if the words tend to produce more lexical activation than the nonwords, then the Σ criterion would be useful for distinguishing words from nonwords. Thus, it will be placed at a position where it will play a major role, reducing the impact of the M criterion. The effect is to make an inhibitory neighborhood frequency effect much less likely to occur. The crucial point, then, is that, for any set of stimuli, the extent to which the words and the nonwords can be distinguished from one another on the basis of lexical activity will largely determine whether Grainger and Jacobs's (1996) model predicts a neighborhood frequency effect.

For the words and the nonwords used in Experiment 2, the distribution of lexical activation values are shown in Figure 1. Note that the words and the nonwords cannot be distinguished from one another on the basis of lexical activation (because of the overlap in the respective distributions), and thus the multiple read-out model, like the interactive activation model, would predict an inhibitory neighborhood frequency effect for this set of words (see Table 4). Figure 2 shows the distribution of lexical activation values for Perea and Pollatsek's (1998) words and nonwords used in Experiment 1. As can be seen in Figure 2, the words tend to generate more lexical activity than the nonwords (the mean lexical activation value for the words is .39 and for the nonwords, .29), and thus it would be possible for participants to use the Σ criterion to distinguish the words from the nonwords. That is, the participants in Experiment 1 could have made use of the Σ criterion when responding, particularly in Experiment 1B, which would have made it difficult to observe strong inhibitory neighborhood frequency effects. For present purposes, however, the crucial point is that this analysis indicates that the stimuli selected for Experiment 2 provide an ideal opportunity, according to the multiple read-out model, to observe an inhibitory neighborhood frequency effect with English stimuli.[6]

[6] We also examined the predictions of parallel distributed processing models (Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989) for the stimuli used in Experiments 1 and 2 (see Sears, Hino, & Lupker, 1999, for a more thorough examination of these models' predictions with regard to the effects of a word's orthographic neighbors). Unlike the multiple read-out model and the interactive-activation model, in these models, there are no lexical units that represent single words. Instead, lexical representations are embodied in the pattern of activation across an interconnected network of units. To relate these patterns of activation to lexical decision and pronunciation latencies, Seidenberg and McClelland (1989) computed orthographic and phonological error scores, which are measures of how close the model's output is to the desired (correct) output. According to the model, lower orthographic error scores should correspond to shorter lexical decision latencies, and lower phonological error scores should correspond to shorter pronunciation latencies. Like the Seidenberg and McClelland model, in the Plaut et al. model (Simulation 4), an error score (cross-entropy error) measures how close the model's output is to the correct pronunciation, with lower cross-entropy errors presumably corresponding to shorter pronunciation latencies.

For the words used in Experiment 1 (Perea & Pollatsek's, 1998, stimuli), the models' lexicon included 45 of the 88 words (51.1%), and an analysis of these error scores revealed that both models predicted an effect of word frequency (with lower error scores for medium-frequency words than for low-frequency words), but neither model predicted an effect of neighbor-

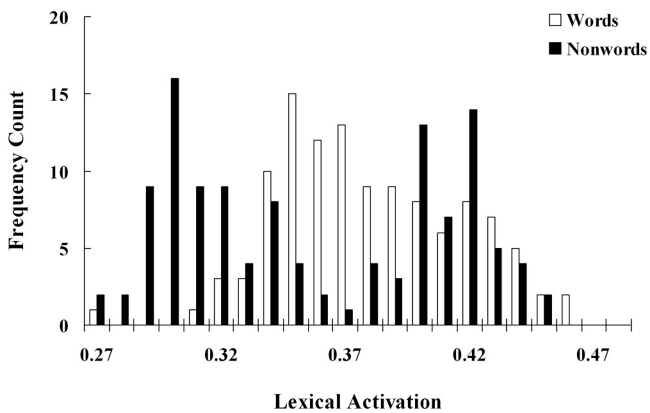

Figure 1. Frequency distribution of lexical activation produced by the stimuli used in Experiment 2.

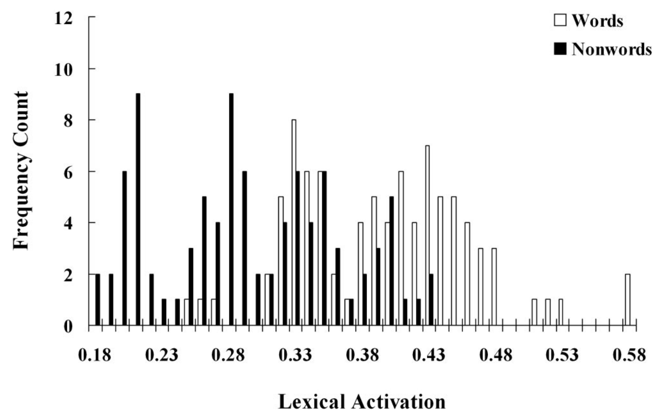

Figure 2. Frequency distribution of lexical activation produced by the stimuli used in Experiment 1.

Apparatus and Procedure

The apparatus and procedure were identical to those used in Experiment 1. In Experiment 2A, participants were instructed to stress accuracy when responding (the lexical decision instructions used by Perea & Pollatsek, 1998). In Experiment 2B, participants were instructed only to respond as quickly and as accurately as possible (standard lexical decision instructions). Each participant completed 24 practice trials before the collection of data. The practice stimuli consisted of 12 words and 12 orthographically legal and pronounceable nonwords. After the practice trials, the participants were provided with feedback as to the mean latency and accuracy of their responses (percentage correct). During the experimental trials this information was presented every 30 trials. The order in which the 240 stimuli were presented in the experiments was randomized separately for each participant.

Although there was no evidence in Experiment 1 that the neighborhood frequency effect was modulated by reader skill, we again administered the ART to each participant after they completed the lexical decision task so that we could look for interactions involving reader skill.

Results

For the word data, the response latencies of correct responses and the error rates from each experiment were submitted to a 2 (word frequency: high, low) × 2 (neighborhood frequency: no higher frequency neighbors, higher frequency neighbors) factorial ANOVA.[7] Both subject and item analyses were performed.

As in Experiment 1, response latencies less than 300 ms or greater than 1,500 ms were considered outliers and were removed from all analyses. For Experiment 2A, a total of 43 observations (0.9% of the data) were removed by this procedure; for Experiment 2B, a total of 15 observations (0.3% of the data) were removed. The mean response latencies of correct responses and the mean error rates in Experiments 2A and 2B are shown in Table 5.

Experiment 2A: Perea and Pollatsek's Lexical Decision Instructions

There was a main effect of word frequency in the analysis of response latencies, $F_s(1, 40) = 212.46$, $p < .01$, $MSE = 1{,}264.13$, $F_i(1, 111) = 71.73$, $p < .01$, $MSE = 2{,}533.86$, as medium-frequency words were responded to 80 ms faster than low-frequency words. Neither the main effect of neighborhood frequency, $F_s(1, 40) = 1.27$, $p > .10$, $MSE = 443.47$, $F_i < 1$, nor the Neighborhood Frequency × Word Frequency interaction, $F_s(1, 40) = 1.70$, $p > .10$, $MSE = 765.27$, $F_i < 1$, were significant. As can be seen in Table 5, there was no hint of an inhibitory neighborhood frequency effect.

In the analysis of error rates, there was a main effect of word frequency, $F_s(1, 40) = 50.54$, $p < .01$, $MSE = 13.28$, $F_i(1, 111) = 26.72$, $p < .01$, $MSE = 17.61$, as the error rates to the low-frequency words (4.8%) were higher than the error rates to the medium-frequency words (0.8%). There was no main effect of neighborhood frequency, $F_s(1, 40) = 1.33$, $p > .10$, $MSE = 8.10$, $F_i < 1$, nor was there a Word Frequency × Neighborhood Frequency interaction, $F_s(1, 40) = 1.45$, $p > .10$, $MSE = 5.12$, $F_i < 1$, were significant. Thus, in contrast to the results of Experiment 1A (and to the results of Perea & Pollatsek, 1998), there was no evidence of an inhibitory neighborhood frequency effect in this experiment.

Effect of reader skill: High ART score versus low ART score. As in Experiment 1, to assess the effect of reader skill on the neighborhood frequency effect, a median split of the ART scores ($Mdn = 14$, range = 5–37) was used to create two groups of participants: a low-ART group ($M = 11.7$) and a high-ART group ($M = 19.7$), $F_s(1, 39) = 35.05$, $p < .01$, $MSE = 18.69$. (Note that

hood frequency (the error scores for words with higher frequency neighbors were no higher than the error scores for words without higher frequency neighbors). For the words used in Experiment 2, the models' lexicon included 97 of the 115 words (84.3%). The Seidenberg and McClelland model predicted an effect of word frequency but not of neighborhood frequency, whereas the Plaut et al. model predicted an effect of word frequency (lower error scores for medium-frequency words) and an inhibitory effect of neighborhood frequency (higher error scores for words with higher frequency neighbors).

[7] Word length (four letters or five letters) was not included in these analyses because this comparison was not planned and, therefore, the four- and five-letter words were not matched on normative frequency, familiarity, neighborhood size, neighborhood frequency, or any other variable.

Table 5
Mean Response Latencies (in Milliseconds) and Error Rates (in %) in Experiment 2A (Perea & Pollatsek's, 1998, Lexical Decision Instructions) and Experiment 2B (Standard Lexical Decision Instructions)

Word frequency	Neighborhood frequency: No HF neighbors	HF neighbors
Experiment 2A		
Low-frequency words	682 (4.8)	673 (4.9)
Medium-frequency words	596 (0.4)	598 (1.3)
Experiment 2B		
Low-frequency words	642 (6.9)	630 (5.8)
Medium-frequency words	559 (0.8)	566 (1.8)

Note. HF = higher-frequency. Error rates appear in parentheses. In Experiment 2A, the mean response latency for the nonwords was 742 milliseconds (ms), and the mean error rate was 5.2%. In Experiment 2B, the mean response latency for the nonwords was 685 ms, and the mean error rate was 4.8%.

the mean ART scores of these two groups were very similar to those of the groups in Experiment 1.)

For the responses to words, unlike Experiment 1, participants in the low-ART group were not significantly slower than those in the high-ART group (both Fs < 1), nor did they make significantly more errors (both ps $> .10$). However, the two groups did differ in their responses to the nonwords (as in Experiment 1): The low-ART participants responded to the nonwords more slowly than the high-ART participants (751 ms vs. 732 ms), $F_s < 1$, $F_i(1, 119) = 24.86$, $p < .01$, $MSE = 1{,}073.22$. The groups did not differ in their nonword error rates (5.0% for the low-ART group and 5.5% for the high-ART group; both Fs < 1).

As was the case in the previous experiments, there was no evidence that the neighborhood frequency effect was modulated by reader skill, as there were no significant interactions involving ART group and neighborhood frequency in the response latency analyses or in the error analyses (all Fs < 1). On the other hand, there was evidence that the word frequency effect was modulated by reader skill, as it was in Chateau and Jared's (2000) experiment and in Experiment 1B. For the low-ART group, the word frequency effect was 90 ms, and for the high-ART group the word frequency effect was 71 ms, producing a marginally significant ART Group × Word Frequency interaction, $F_s(1, 39) = 3.21$, $p = .08$, $MSE = 1{,}197.90$, $F_i(1, 111) = 12.13$, $p < .01$, $MSE = 883.36$.

Experiment 2B: Standard Lexical Decision Instructions

As can be seen in Table 5, the responses to words and to nonwords were faster and slightly less accurate in Experiment 2B than in Experiment 2A, which was expected given the different lexical decision task instructions. Responses to words were 38 ms faster in Experiment 2B, $F_s(1, 79) = 3.94$, $p = .05$, $MSE = 30{,}301.95$, $F_i(1, 111) = 198.04$, $p < .01$, $MSE = 374.55$, and responses to nonwords were 56 ms faster, $F_s(1, 79) = 6.70$, $p < .05$, $MSE = 9{,}787.73$, $F_i(1, 119) = 214.11$, $p < .01$, $MSE = 772.14$. The difference in accuracy was significant for the words, $F_s(1, 79) = 3.63$, $p = .06$, $MSE = 21.25$, $F_i(1, 111) = 7.49$, $p < .01$, $MSE = 7.29$, but not for the nonwords (both Fs < 1).

In the analysis of response latencies of Experiment 2B, there was a main effect of word frequency, $F_s(1, 39) = 257.42$, $p < .01$, $MSE = 843.18$, $F_i(1, 111) = 81.06$, $p < .01$, $MSE = 2{,}042.57$; low-frequency words were responded to 73 ms slower than medium-frequency words. The main effect of neighborhood frequency was not significant (both Fs < 1). The Word Frequency × Neighborhood Frequency interaction was significant in the subject analysis, $F_s(1, 39) = 8.96$, $p < .01$, $MSE = 422.40$, but not in the item analysis, $F_i(1, 111) = 1.15$, $p > .10$, $MSE = 2{,}042.57$. As can be seen in Table 5, the interaction appears to reflect the different neighborhood frequency effects for the low-frequency and medium-frequency words. For the low-frequency words, there appeared to be a small facilitory neighborhood frequency effect: Responses to words with higher frequency neighbors were 12 ms faster than responses to words without higher frequency neighbors. This difference was significant in the subject analysis only, $F_s(1, 39) = 5.48$, $p < .05$, $MSE = 570.37$, $F_i < 1$. Of course, this result is just the opposite of what would be expected given the results of Perea and Pollatsek (1998) and of Experiment 1A. For the medium-frequency words, responses to words with higher frequency neighbors were 7 ms slower than responses to words without higher frequency neighbors. This difference was not significant in the subject analysis, $F_s(1, 39) = 3.22$, $p = .08$, $MSE = 299.68$, or in the item analysis ($F_i < 1$). Thus, the interaction occurred because for the low-frequency words there was a small facilitory neighborhood frequency effect, whereas for the medium-frequency words the effect was in the opposite direction, although it was not statistically significant.

In the analysis of error rates, there was a main effect of word frequency, $F_s(1, 39) = 62.82$, $p < .01$, $MSE = 16.61$, $F_i(1, 111) = 32.32$, $p < .01$, $MSE = 23.18$, with fewer errors to medium-frequency words than to low-frequency words (1.3% vs. 6.3%). The main effect of neighborhood frequency was not significant (both Fs < 1). As can be seen in Table 5, the pattern of error rates mirrored the pattern of response latencies, although the interaction between word frequency and neighborhood frequency was not statistically significant, $F_s(1, 39) = 3.50$, $p = .07$, $MSE = 12.00$, $F_i(1, 111) = 1.30$, $p > .10$, $MSE = 23.18$.

Effect of reader skill: High ART score versus low ART score. As before, a low-ART group ($M = 12.5$) and a high-ART group ($M = 20.7$) were created using a median split on ART scores ($Mdn = 16$, range = 5–29), $F_s(1, 38) = 68.07$, $p < .01$, $MSE = 9.75$. (The mean ART scores of these two groups were very similar to those of the corresponding groups in the previous experiments.) For the responses to words, participants in the low-ART group were slower than those in the high-ART group (610 ms vs. 584 ms), $F_s(1, 38) = 1.34$, $p > .10$, $MSE = 19{,}269.96$, $F_i(1, 111) = 56.39$, $p < .01$, $MSE = 607.96$, although they were not more error prone (both Fs < 1). Similarly, for the responses to the nonwords, participants in the low-ART group were slower than those in the high-ART group (703 ms vs. 663 ms), $F_s(1, 38) = 2.65$, $p > .10$, $MSE = 5{,}993.36$, $F_i(1, 119) = 98.55$, $p < .01$, $MSE = 958.36$, but they were not more error prone (5.2% vs. 4.3%; both ps $> .10$). The slower responding of the low-ART group is consistent with the results of the previous experiments and also with the results of Chateau and Jared (2000).

Also consistent with previous experiments (with the exception of Experiment 1A) was the larger word frequency effect for the low-ART group (79 ms for the low-ART group and 67 ms for the high-ART group), although the interaction between word frequency and ART group was not statistically significant (both $ps > .10$). With respect to the neighborhood frequency effect, there were no significant interactions involving ART group and neighborhood frequency in the response latency analyses (all $ps > .10$) or in the error analyses (all $ps > .10$). Consequently, once again, there was no evidence that the neighborhood frequency effect was modulated by reader skill.

Discussion

The purpose of Experiment 2 was to determine whether the lexical decision instructions given to participants determine whether or not there is an inhibitory neighborhood frequency effect, as suggested by the pattern of results from Experiments 1A and 1B. An instruction-based hypothesis predicted that there would be an inhibitory neighborhood frequency effect in Experiment 2A, in which participants were instructed to stress accuracy when responding (as in Experiment 1A and in Perea & Pollatsek's, 1998, experiment), but no inhibitory neighborhood frequency effect in Experiment 2B, in which participants were given standard lexical decision instructions that did not emphasize accuracy (as in Experiment 1B). Essentially, the same pattern of results obtained in Experiment 1 should have been obtained in Experiment 2 using the new and larger set of low- and medium-frequency words created for that experiment. If so, then it would have confirmed that the lexical decision instructions are critical for obtaining an inhibitory neighborhood frequency effect; only when participants respond slowly and carefully, giving preference to accuracy over speed, does the effect appear. This result, in turn, suggests that an inhibitory neighborhood frequency effect is seldom observed in experiments that use English stimuli because the participants in these experiments respond like the participants in Experiment 1B: as quickly and as accurately as possible. That is, the crucial difference between the English experiments and the experiments conducted in French, Dutch, and Spanish would not be the language of the stimuli but rather how the participants interpret the task instructions and how they then respond in the lexical decision task.

Clearly, the results of Experiment 2 did not support this instruction-based hypothesis. The critical result was the absence of an inhibitory neighborhood frequency effect in Experiment 2A, in which the instructions to participants emphasized accuracy. These were the identical instructions used in Experiment 1A, in which an inhibitory neighborhood frequency effect was observed; however, in Experiment 2A a different set of low- and medium-frequency words was used. The absence of an inhibitory neighborhood frequency effect in Experiment 2A is essentially a failure to replicate the effect using the identical instructions but a different set of items. As a result, we can rule out the possibility that the lexical decision instructions alone determine whether or not there is an inhibitory neighborhood frequency effect.

We can also conclude that Perea and Pollatsek's (1998) lexical decision results are not a straightforward demonstration of the inhibitory neighborhood frequency effect predicted by activation-based models. Instead, considered together, the results of Experiments 1 and 2 indicate that the inhibitory effect of neighborhood frequency in Perea and Pollatsek's experiment was due to the particular combination of the lexical decision instructions and the word and nonword stimuli used in that experiment. As a result, the relevance of Perea and Pollatsek's lexical decision results to any evaluation of activation-based models becomes questionable.

What is not clear, of course, is why there is an effect of neighborhood frequency with Perea and Pollatsek's (1998) stimuli when their lexical decision instructions are used (Experiment 1A). Although we cannot single out any one cause, there are likely contributing factors we can identify. First, as noted, simulations with the multiple read-out model (Grainger & Jacobs, 1996) indicated that the model did not predict a word frequency effect for these stimuli, although it did predict a neighborhood frequency effect. This suggests that there is something unusual about this set of words, because the word frequency manipulation is fairly strong (the mean normative frequency of the low-frequency words was 3.4 and of the medium frequency words, 25.9), and because there was a substantial empirical word frequency effect for these stimuli (62 ms in Experiment 1A and 57 ms in Experiment 1B). (In contrast, the word frequency effects observed in Experiment 2 were predicted by the model.)

Second, as our examination of the subjective frequency ratings made clear, many of Perea and Pollatsek's (1998) words are very infrequently encountered in print, even by university students (e.g., on a scale from 1 to 9, *tunic* had a mean subjective frequency of 1.4 and *tassel* a mean subjective frequency of 1.5). For many of these words (e.g., *siege, horde, shawl, mosque, urine*), including those that Perea and Pollatsek ultimately excluded from their analyses (*lasso, noose, verve,* and *villa*), participants may have been unsure of the correct spelling and so, when instructed to stress accuracy, engaged in an additional checking process before responding, slowing their lexical decisions. (This strategy would have also been encouraged by the close resemblance of many of the nonwords to real words; e.g., *vowal, ribban, carpot.*) This checking process may have been more of an issue for the words with higher frequency neighbors because, as noted, the subjective frequency ratings for these words were, on average, lower than the subjective frequency ratings for the words without higher frequency neighbors.[8]

[8] One other difference between the words used in the two experiments is that those used in Experiment 1 had, on average, slightly fewer neighbors than those used in Experiment 2 (2.2 vs. 4.4). Because the number of higher frequency neighbors was controlled, this difference is due to differences in the number of lower frequency neighbors. Although it is not impossible that this difference could matter, it seems quite unlikely that it did. The actual difference in the mean number of neighbors is quite small (i.e., all these words are essentially small neighborhood words). In addition, there seems to be no rationale why even a large difference in the number of lower frequency neighbors would change the probability of observing a neighborhood frequency effect, an effect presumed to be caused by a single higher frequency neighbor. Certainly, there is no mechanism for lower frequency neighbors to have any real effect in the activation-based models being examined. In those models, the impact of neighbors is a direct function of the neighbor's frequency. All of the lower frequency neighbors of the low-frequency words used in both experiments (the ones that supposedly might show a neighborhood frequency effect) would have frequencies very close to zero. In fact, some of these neighbors may not even be known to some of the participants. Thus, their impact on any aspect of processing would have been extremely limited at best.

This proposal—that the stimuli and the task instructions used in Experiment 1A encouraged participants to engage in an additional checking process before responding—could explain another anomalous result of Experiment 1A. In addition to being the only one of our four experiments in which an inhibitory neighborhood frequency effect was observed, Experiment 1A was also the only experiment in which the word frequency effect was not modulated by reader skill. That is, like Chateau and Jared (2000), in Experiments 1B, 2A, and 2B, we found that the word frequency effect was larger for the participants with lower ART scores. (In Experiment 1B, the word frequency effect was 72 ms for the low-ART group and 41 ms for the high-ART group; in Experiment 2A, 90 ms for the low-ART group and 71 ms for the high-ART group; and in Experiment 2B, 79 ms for the low-ART group and 67 ms for the high-ART group.) The exception was Experiment 1A, in which the word frequency effects for the two groups were virtually identical (60 ms for the low-ART group and 62 ms for the high-ART group). Relative to Experiment 1B, in which the identical stimuli were used, the different instructions used in Experiment 1A appear to have decreased the word frequency effect for the low-ART participants (from 72 ms in Experiment 1B to 60 ms in Experiment 1A) and to have increased the word frequency effect for the high-ART participants (from 41 ms in Experiment 1B to 62 ms in Experiment 1A), which eliminated the difference between the two groups. This situation created a three-way Word Frequency $\times$ART Group $\times$ Experiment interaction, $F_s(1, 76) = 4.64$, $p < .05$, $MSE = 1{,}131.50$, $F_i(1, 84) = 8.47$, $p < .01$, $MSE = 1{,}321.22$.

This pattern of results makes sense if the stimuli and the task instructions used in Experiment 1A encouraged participants to engage in a checking process and, in addition, if the engagement of that checking process varied as a function of reader skill. That is, if the low-ART participants were checking both the low- and the medium-frequency words in Experiment 1A but were checking only the low-frequency words in Experiment 1B, then a smaller word frequency effect would be expected in Experiment 1A. Alternatively, if the high-ART participants were checking only the low-frequency words and only in Experiment 1A, then a larger word frequency effect would be expected in Experiment 1A.[9] Experiment 1A was the only one of our four experiments in which the word frequency effect was not modulated by reader skill and also the only experiment in which an inhibitory neighborhood frequency effect was observed. Both of these outcomes appear to be due to the particular combination of lexical decision instructions and the word and nonword stimuli used in that experiment.

Whatever the reasons, the important point is that Perea and Pollatsek's (1998) data are the only data that provide any clear support for the existence of an inhibitory neighborhood frequency effect in English, yet our experiments demonstrate that their lexical decision results do not generalize across items or across task instructions. Consequently, the reality of the inhibitory neighborhood frequency effect in English, at least in the lexical decision task, remains in serious doubt.

Experiment 3

Although there is now an extensive literature on orthographic neighborhood effects, almost all of this research has used tasks that involve making responses to isolated words (e.g., the lexical decision task, perceptual identification tasks). Very few studies have examined the effects of a word's orthographic neighbors in normal silent reading, the exceptions being Perea and Pollatsek (1998) and Pollatsek, Perea, and Binder (1999). In Pollatsek et al.'s experiments, the focus was on the effect of neighborhood size, with neighborhood frequency being controlled. Participants' eye movements were recorded while they read sentences that contained target words with large neighborhoods and target words with small neighborhoods, and the reading times to the words in these two conditions were compared. All the words had at least one neighbor of higher frequency, and the normative frequency of the highest frequency neighbor was equated across the large and the small neighborhood size conditions (the number of higher frequency neighbors was also equated across these two conditions in their Experiment 3). Taken together, Pollatsek et al.'s results were equivocal: There was no clear facilitory neighborhood size effect on reading times, nor was there a clear inhibitory effect (in contrast to the results from lexical decision tasks, including Pollatsek et al.'s Experiment 1, in which a clear facilitory neighborhood size effect was observed). Although Pollatsek et al. did not manipulate neighborhood frequency, from their post hoc regression analyses, they concluded that increasing the number of lower frequency neighbors had a weak facilitory effect on reading times, whereas increasing the number of higher frequency neighbors had an inhibitory effect.

In Perea and Pollatsek's (1998) Experiment 2, the focus was on the effect of neighborhood frequency, with neighborhood size

[9] Our reasoning is as follows: For the low-ART participants, we assume that in both Experiments 1A and 1B these participants used a checking process for the low-frequency words, because many of these words would have been unfamiliar to them (as a result of their lower levels of print exposure; again, recall that many of these words were of very low normative frequency). For the medium-frequency words, we assume that these participants tended not to check these words in Experiment 1B, in which participants were given standard lexical decision instructions, but were inclined to check them in Experiment 1A because of the emphasis on accurate responding. Consistent with these assumptions, the response latencies to low-frequency words were virtually identical in Experiments 1A and 1B (668 ms vs. 661 ms), whereas the response latencies to medium-frequency words were slower in Experiment 1A than in Experiment 1B (608 ms vs. 589 ms). The end result was a smaller word frequency effect in Experiment 1A (60 ms) than in Experiment 1B (72 ms), because, unlike Experiment 1A, in Experiment 1B only the low-frequency words were being checked, increasing the latency difference between the low- and the medium-frequency words.

For the high-ART participants, we assume that in Experiment 1B, in which standard lexical decision instructions were used, these participants did not tend to engage in a checking process, even for the low-frequency words, because these participants were fairly experienced with printed text. However, in Experiment 1A, these participants were inclined to check the low-frequency words before responding, because of the low familiarity of many of these words and the emphasis on accurate responding. Consistent with these assumptions, responses to the low-frequency words were slower in Experiment 1A than in Experiment 1B (611 ms vs. 576 ms), whereas the responses to the medium-frequency words were more similar in the two experiments (549 ms vs. 535 ms). The end result was a larger word frequency effect in Experiment 1A (62 ms) than in Experiment 1B (41 ms), because in Experiment 1A only the low-frequency words were being checked, increasing the latency difference between the low- and the medium-frequency words relative to the situation in Experiment 1B, in which none of the words were being checked.

being controlled (all the words had small neighborhoods). As noted, they concluded that a word's higher frequency neighbors do not have any direct and immediate effect on reading time, as first-fixation durations and gaze durations to words with higher frequency neighbors were no longer than the durations to words without higher frequency neighbors. In contrast, there were effects of neighborhood frequency on a few of the spillover variables measured, which led Perea and Pollatsek to conclude that higher frequency neighbors do affect later stages of processing (i.e., after a reader has left a word).

Given the limited data on the effect of higher frequency neighbors on word identification during normal silent reading, we believed it was important shed more light on this issue by replicating Perea and Pollatsek's (1998) experiment and conducting a new experiment using entirely different stimulus materials. In the present experiments, the eye movements of participants were recorded while they read sentences that contained the words used in Experiments 1 and 2. Experiment 3A used the sentences Perea and Pollatsek (1998) created for their words (the words used in the present Experiment 1 and in Perea and Pollatsek's Experiment 1). In Experiment 3B, a new set of sentences was created for the words used in Experiment 2. We expected that both experiments would produce essentially the same results—namely, no effect of neighborhood frequency on first-fixation durations or on gaze durations but an inhibitory neighborhood frequency effect on one or more of the spillover variables (as Perea and Pollatsek reported). Individual differences in the neighborhood frequency effect were evaluated by performing the same analyses that Perea and Pollatsek used (a median split on the percentage of regressions back to the target word) and by administering the ART to each participant at the end of the experiment.

Method

Participants

Eighty University of Calgary undergraduate students participated in the experiment in exchange for partial course credit. Forty participated in Experiment 3A and 40 in Experiment 3B. All participants were native English speakers with normal or corrected-to-normal vision. None participated in any of the previous experiments or in more than one of the present experiments.

Materials

The sentences used in Experiment 3A were the same as those used by Perea and Pollatsek (1998). The words from Experiment 1 were used in 46 pairs of sentences. The sentences in each pair were identical, except that one contained a target word with no higher frequency neighbors and the other contained a target word with higher frequency neighbors.

For Experiment 3B, a new set of 60 pairs of sentences was created using the words from Experiment 2 (these sentences are listed in the Appendix). The sentences in each pair were identical, except that one contained a target word with no higher frequency neighbors and the other contained a target word with at least one higher frequency neighbor. Half of the sentences contained low-frequency targets and half contained medium-frequency targets. Unlike Perea and Pollatsek's word frequency manipulation, our word frequency manipulation was planned, and so we were able to ensure that for each sentence pair a low-frequency word was paired with a low-frequency word and a medium-frequency word was paired with a medium-frequency word. (In 11 of Perea and Pollatsek's 46 sentences, a low-frequency word was paired with a medium-frequency word.)

We also took steps to reduce any impact that contextual information could have on the processing of the target words. All of the sentences were constructed so that the preceding context was neutral and ambiguous with respect to the target word (e.g., "The pamphlet outlined the *risk* involved in the triathlon"), and thus all of the targets were equally low in predictability (predictable targets are skipped more often than less predictable targets; see Rayner & Pollatsek, 1989, for a review). In addition, the target word was preceded by an adjective only in 3 of the 60 sentences; in the remainder of the sentences, the target word was not modified (e.g., "Justin said that the *lodge* was already booked for the weekend"). Whether the target word was modified or not is an important factor to control because modified words are processed more slowly than unmodified words (e.g., Rayner, Sereno, Morris, Schmauder, & Clifton, 1989, reported that fixation durations for modified nouns were longer than fixation durations for unmodified nouns).

Apparatus and Procedure

Eye movements were recorded by a Sensomotoric Instruments, Inc. (Boston) EyeLink eye-tracking system, which uses infrared video-based tracking technology. The system has a visual resolution of 20 s of arc and a sampling rate of 250 Hz, allowing for a temporal resolution of 4 ms. Participants wore a small lightweight headband equipped with cameras positioned below the eyes that track the position of the pupils as they move while reading. The eye tracker was connected to an IBM 300PL microcomputer and a Sony Multiscan G200 monitor. The computer controlled the visual display and recorded the horizontal and vertical coordinates corresponding to the position of the eye every 4 ms.

Participants were fitted with the headband when they arrived for the experiment, and then the eye tracking system was calibrated. Viewing was binocular, but eye movements were recorded only from the participant's right eye. The calibration period required approximately 5 min. After the calibration was completed, the procedure used by Perea and Pollatsek (1998) in their Experiment 2 was followed. Participants were told that they would be silently reading sentences, presented one at a time, and that the purpose of the experiment was to determine what people look at while they read. They were asked to read the sentences for normal comprehension. Each sentence was no more than 80 characters in length and was presented on a single line of the computer video monitor. The target word was never the first or the last word of the sentence.[10] Before they read any experimental sentences, participants read 8 practice sentences to familiarize themselves with the procedure. During the experimental trials, participants were asked a series of comprehension questions about what they had read. These questions were asked after every 12 sentences in Experiment 3A and after every 15 sentences in Experiment 3B. Accuracy in answering these comprehension questions was very high ($> 90\%$). After completing the reading task, each participant completed the ART.

Design

In Experiment 3A, each participant read one of two lists. Each list consisted of 23 sentences with target words that had higher frequency

[10] For 35% of the sentences used in Experiment 3B, the target word was the second word in the sentence, and so there was a possibility that participants were processing the first word in the sentence and the target word during the same fixation and then skipping the target word, potentially obscuring any effect of neighborhood frequency (or of word frequency). However, our analyses revealed that these target words were not skipped any more often that target words appearing later in the sentences. More important, there were no differences in our results or our conclusions when these targets were excluded from the data analyses.

neighbors and 23 sentences with target words that had no higher frequency neighbors. The neighborhood frequency of the target word (higher frequency neighbors or no higher frequency neighbors) was counterbalanced across the two lists. That is, if a target word with higher frequency neighbors (e.g., *marsh*) appeared in a sentence in one list, its corresponding target without higher frequency neighbors (e.g., *canal*) appeared in the same sentence in the other list. In Experiment 3B, each participant also read one of two lists. Each list consisted of 30 sentences with target words that had higher frequency neighbors and 30 sentences with target words that had no higher frequency neighbors. The neighborhood frequency of the target words was counterbalanced across the two lists.

A 2 (word frequency: low, medium) × 2 (neighborhood frequency: no higher frequency neighbors, higher frequency neighbors) factorial design was used for each of the experiments. The dependent variables included first-fixation durations (the duration of the first fixation on the target word), gaze durations (the sum of the fixation durations on the target word before the reader left the target word), and the percentage of trials in which the target word was initially skipped. For all these analyses, as in Perea and Pollatsek's (1998) analyses, the target region was defined as the target word, the space that preceded it, and the last two characters of the previous word. For first-fixation durations and gaze durations, trials were included in the analyses only when the reader initially fixated on the target word with a forward saccade (i.e., a trial was not included when the target word was initially skipped).

Also of interest was the processing that occurred after the participant left the target word (spillover effects). The following variables were submitted to separate 2 (word frequency: high, low) × 2 (neighborhood frequency: no higher frequency neighbors, higher frequency neighbors) factorial ANOVAs: the duration of the first fixation after leaving the target word, the probability of making a regression back to the target word, the total time spent on the target word (the sum of all fixation durations on the target word, including regressions), and the total time spent on the target word and the immediate posttarget region. (Like Perea & Pollatsek, 1998, we defined the immediate posttarget region as the two words subsequent to the target word.) Again, trials were included only when the target word was originally fixated. Those trials in which the target word was not fixated on the reader's first pass of the sentence were excluded from all subsequent analyses. For Experiment 3A, 251 trials (14.2% of the data) were excluded, and for Experiment 3B, 431 trials (18.7% of the data) were excluded.

Results

Experiment 3A: Perea and Pollatsek's Stimuli

The data for the first-pass variables are listed in Table 6; the data for the spillover variables are listed in Table 7. Unlike Perea and Pollatsek (1998), we observed a statistically significant effect of word frequency on virtually all of the first-pass and spillover variables. More specifically, a word frequency effect was observed for the percentage of target words skipped during the reader's first pass, $F_s(1, 39) = 42.68$, $p < .01$, $MSE = 117.85$, $F_i(1, 84) = 15.72$, $p < .01$, $MSE = 168.50$, on the first-fixation durations, $F_s(1, 39) = 3.43$, $p = .07$, $MSE = 880.76$, $F_i(1, 84) = 4.57$, $p < .05$, $MSE = 576.73$, on the gaze durations, $F_s(1, 39) = 9.96$, $p < .01$, $MSE = 1{,}509.10$, $F_i(1, 84) = 11.33$, $p < .01$, $MSE = 997.52$, on the total time spent on the target word, $F_s(1, 39) = 13.18$, $p < .01$, $MSE = 2{,}249.67$, $F_i(1, 84) = 7.47$, $p < .01$, $MSE = 2{,}991.23$, on the total time spent on the target word and the posttarget region, $F_s(1, 39) = 12.19$, $p < .01$, $MSE = 15{,}540.13$, $F_i(1, 84) = 6.81$, $p < .05$, $MSE = 21{,}462.20$, and on the percentage of regressions back to the target word, $F_s(1, 39) = 6.66$, $p < .05$, $MSE = 142.63$, $F_i(1, 84) = 4.86$, $p < .05$, $MSE = 119.14$. Only for the first fixation after leaving the target word was the effect of word frequency not statistically significant, $F_s(1, 39) = 3.94$, $p = .06$, $MSE = 1{,}747.10$, $F_i(1, 84) = 2.21$, $p > .10$, $MSE = 1{,}856.26$. Because these were the same stimuli that Perea and Pollatsek used, the fact that we were able to detect an effect of word frequency in our analyses suggests that our analyses had greater statistical sensitivity (probably because we had almost twice as many participants in our experiment). This point is relevant when evaluating any null effects of neighborhood frequency reported later.

Table 6
First-Pass Eye-Movement Measures for Experiment 3A

Word frequency	Neighborhood frequency: No HF neighbors	HF neighbors
First-fixation duration (ms)		
Low-frequency words	229	236
Medium-frequency words	227	222
Gaze duration (ms)		
Low-frequency words	255	260
Medium-frequency words	244	232
Target words skipped (%)		
Low-frequency words	10.4	8.6
Medium-frequency words	23.5	18.0

Note. HF = higher-frequency; ms = milliseconds.

Like Perea and Pollatsek (1998), we did not find an effect of neighborhood frequency on first-fixation durations or on gaze durations (all *F*s < 1), nor were there any interactions between word frequency and neighborhood frequency for these variables (all *p*s > .10). Thus, there was no effect of neighborhood frequency while the participants fixated the target words. However, there was an effect of neighborhood frequency in the analysis of the percentage of target words skipped during the reader's first pass. Specifically, words with higher frequency neighbors were skipped less often than words without higher frequency neighbors (13.3% vs. 16.9%); this effect was significant in the subject analysis, $F_s(1, 39) = 4.92$, $p < .05$, $MSE = 110.96$, but not in the item analysis, $F_i(1, 84) = 1.61$, $p > .10$, $MSE = 168.50$. As can be seen in Table 8, this was true for both the low-frequency and the medium-frequency words. Although the difference was larger for the medium-frequency words, there was no interaction between word frequency and neighborhood frequency, $F_s(1, 39) = 1.32$, $p > .10$, $MSE = 105.53$, $F_i < 1$, although there was an interaction with reader skill, the details of which are reported next. (In Perea and Pollatsek's experiment, medium-frequency words with higher frequency neighbors were skipped less often than medium-frequency words without higher frequency neighbors, but there was no difference for low-frequency words.)

Unlike Perea and Pollatsek (1998), we did not find a significant effect of neighborhood frequency on the percentage of regressions to the target word (both *F*s < 1), the total time readers spent fixating the target word (both *F*s < 1), the duration of the first fixation after the target word fixation (both *F*s < 1), or the total time spent fixating the target word and the posttarget region, $F_s(1, 39) = 1.38$, $p > .10$, $MSE = 13{,}274.60$, $F_i < 1$. In addition, there

Table 7
Spillover Eye-Movement Measures for Experiment 3A

Word frequency	Neighborhood frequency: No HF neighbors	HF neighbors
Total time on target word (ms)		
Low-frequency words	293	304
Medium-frequency words	272	271
First-fixation duration after target word (ms)		
Low-frequency words	253	244
Medium-frequency words	232	239
Total time on target word and the posttarget region (ms)		
Low-frequency words	705	735
Medium-frequency words	644	657
Target words regressed (%)		
Low-frequency words	16.0	16.7
Medium-frequency words	10.3	12.6

Note. HF = higher-frequency; ms = milliseconds.

was no interaction between neighborhood frequency and word frequency for any of these variables (all ps $>$ .10). Consequently, although Perea and Pollatsek concluded that an inhibitory neighborhood frequency effect was most evident on these spillover variables, the results of our analyses lent no support to that conclusion.

Individual differences: Percentage of regressions. We replicated Perea and Pollatsek's (1998) analysis of individual differences in the neighborhood frequency effect by dividing the 40 participants into two equal groups, using a median split on the percentage of trials in which participants regressed back to the target word (*Mdn* = 12.8%, range = 0–38.5%). This procedure created a large difference between the two groups in terms of the mean percentage of regressions (7.8% vs. 20.0%), $F_s(1, 38) = 41.75$, $p < .01$, $MSE = 144.38$, $F_i(1, 84) = 28.34$, $p < .01$, $MSE = 164.71$. Recall that Perea and Pollatsek reported that, for participants who made fewer regressions, the gaze durations to words with higher frequency neighbors were longer than the gaze durations to words without higher frequency neighbors, whereas the opposite was true for participants who made more regressions (i.e., there was a facilitory, but not statistically significant, neighborhood frequency effect). As noted, these results led Perea and Pollatsek to speculate that, for some readers (i.e., those who make relatively few regressions while reading), inhibitory effects of neighborhood frequency may be observed while the target word is fixated.

Unlike Perea and Pollatsek (1998), we did not observe any differences between the two regression groups in the analysis of gaze durations: There were no interactions between regression group and neighborhood frequency (all ps $>$ .10), and for both groups there was no effect of neighborhood frequency. The same was true in the analysis of first-fixation durations. Thus, there was no evidence of an effect of neighborhood frequency for either group while the participants fixated the target words.

There was a minor difference between the two groups in the analysis of the percentage of target words skipped during the reader's first pass. More specifically, there was a marginally significant three-way interaction among regression group, word frequency, and neighborhood frequency, $F_s(1, 38) = 3.80$, $p = .06$, $MSE = 98.46$, $F_i(1, 84) = 3.66$, $p = .06$, $MSE = 86.45$, because there was an effect of neighborhood frequency only for the participants who made fewer regressions and only for the medium-frequency words. For these participants, there was a significant Word Frequency × Neighborhood Frequency interaction in the subject analysis, $F_s(1, 19) = 6.21$, $p < .05$, $MSE = 78.09$, $F_i(1, 84) = 1.83$, $p > .10$, $MSE = 222.33$. For the low-frequency words, there was no effect of neighborhood frequency (both Fs $<$ 1), but for the medium-frequency words, words with higher frequency neighbors were skipped less often than words without higher frequency neighbors (16.2% vs. 26.5%), $F_s(1, 19) = 8.27$, $p < .05$, $MSE = 127.69$, $F_i(1, 35) = 2.20$, $p > .10$, $MSE = 385.27$. For the participants who made more regressions, there was no effect of neighborhood frequency and no interaction between word frequency and neighborhood frequency (all ps $>$ .10), and these participants did not skip the medium-frequency words with higher frequency neighbors any more than the medium-frequency words without higher frequency neighbors (19.8% vs. 20.6%). Recall that in the overall analysis words with higher frequency neighbors were skipped less often than words without higher frequency neighbors, and the difference was larger for the medium-frequency words (see Table 6). Incorporating the regression groups into the same analysis indicates that this effect was confined to the participants who made fewer regressions and to the medium-frequency words.

Our analyses of the spillover variables revealed one other difference between these two groups. In the analysis of the total time spent on the target word, which included regressive fixations, there was a marginally significant three-way interaction among regression group, word frequency, and neighborhood frequency, $F_s(1, 38) = 3.29$, $p = .08$, $MSE = 2{,}487.52$, $F_i(1, 84) = 3.87$, $p = .05$, $MSE = 2{,}239.44$. Separate analyses of the two groups produced an interaction for the participants who made fewer regressions but not for those who made more regressions. Specifically, for the participants who made fewer regressions, there was an interaction between neighborhood frequency and word frequency, $F_s(1, 19) = 4.81$, $p < .05$, $MSE = 1{,}712.37$, $F_i(1, 84) = 4.72$, $p < .05$, $MSE = 2{,}135.64$. For the low-frequency words, words with higher frequency neighbors were examined an average of 28 ms longer than words without higher frequency neighbors, $F_s(1, 19) = 9.82$, $p < .01$, $MSE = 813.54$, $F_i(1, 49) = 4.59$, $p < .05$, $MSE = 2{,}049.08$, and for the medium-frequency words the neighborhood frequency effect was reversed, but the 12-ms difference was not statistically significant, $F_s < 1$, $F_i(1, 35) = 1.07$, $p > .10$, $MSE = 2{,}256.84$. For the participants who made more regressions, there was no effect of neighborhood frequency and no interaction between word frequency and neighborhood frequency (all ps $>$ .10).

Together these analyses produced results that are somewhat consistent with Perea and Pollatsek's (1998) observation that participants who make fewer regressions are more likely to exhibit an inhibitory effect of neighborhood frequency. Specifically, for these participants, the medium-frequency words with higher frequency neighbors were skipped less often than the medium-frequency words without higher frequency neighbors, and the low-frequency words with higher frequency neighbors were examined an average of 28 ms longer than the low-frequency words without higher frequency neighbors. However, note that neither of these specific results was reported by Perea and Pollatsek. More important, unlike in Perea and Pollatsek's experiment, in our experiment there was no indication that these participants had longer gaze durations (or longer first-fixation durations) to words with higher frequency neighbors. Our results provide no support for the hypothesis that, for some readers, higher frequency neighbors have a direct and immediate effect on reading times.

Individual differences: High ART score versus low ART score. We used a median split on the ART scores (Mdn = 15, range = 6–32) to create a low-ART group (N = 19, M = 10.8) and a high-ART group (N = 21, M = 20.9), $F_s(1, 38) = 54.65$, $p < .01$, MSE = 18.48. The low-ART group did not regress to the target words more often than the high-ART group, $F_s(1, 38) = 1.62$, $p > .10$, MSE = 290.65, $F_i(1, 84) = 4.32$, $p < .05$, MSE = 126.07, and actually made slightly fewer regressions overall (12.1% vs. 15.5%). Relatedly, the correlation between the ART scores and the mean percentage of regressions was essentially zero (r = .05). These results indicate that the ART and the regression data are measuring different individual differences associated with reading.

There were several reasons to believe that the ART was tapping individual differences in reader ability. For one, participants with lower ART scores were slower to read the target words than participants with higher ART scores, as they had longer first-fixation durations (234 ms vs. 224 ms), $F_s(1, 38) = 1.13$, $p > .10$, MSE = 3,854.57, $F_i(1, 84) = 14.36$, $p < .01$, MSE = 784.67, and longer gaze durations (257 ms vs. 239 ms), $F_s(1, 38) = 2.04$, $p > .10$, MSE = 6,584.50, $F_i(1, 84) = 17.47$, $p < .01$, MSE = 1,469.11. (These differences were especially pronounced for the low-frequency words.) These results are consistent with those of other studies that have examined the eye movement latencies of readers of higher and lower ability (e.g., Jared et al., 1999). They are also consistent with the fact that lexical decision responses of low-ART participants were generally slower than those of high-ART participants in Experiments 1 and 2. Second, participants with lower ART scores skipped the target words much less often than those with higher ART scores (9.8% vs. 20.0%), $F_s(1, 38)$ = 13.88, $p < .01$, MSE = 297.99, $F_i(1, 84) = 35.36$, $p < .01$, MSE = 158.26, consistent with the notion that they were less skilled readers. In addition, word frequency effects were larger for participants with lower ART scores than for those with higher ART scores, as evidenced by statistically significant interactions between ART group and word frequency for gaze durations, $F_s(1, 38) = 7.85$, $p < .01$, MSE = 1,283.59, $F_i(1, 84) = 3.92$, $p = .05$, MSE = 1,469.11, and for the duration of the first fixation after leaving the target word, $F_s(1, 38) = 10.46$, $p < .01$, MSE = 1,405.91, $F_i(1, 84) = 5.35$, $p < .05$, MSE = 2,094.67. These results are consistent with those of Experiments 1B, 2A, and 2B, in which participants with lower ART scores also had larger word frequency effects than those with higher ART scores.

With regard to the neighborhood frequency effect, our analyses revealed only one difference between the low-ART and high-ART participants. In the analysis of the total time spent on the target word and the posttarget region (a spillover variable), there was a significant interaction between ART group and neighborhood frequency in the subject analysis, $F_s(1, 38) = 4.61$, $p < .05$, MSE = 12,150.52, $F_i < 1$. For the low-ART participants, there was no effect of neighborhood frequency for the low-frequency words or for the medium-frequency words (all Fs < 1). For the high-ART participants, there was an effect of neighborhood frequency in the subject analysis, $F_s(1, 20) = 5.78$, $p < .05$, MSE = 11,811.15, $F_i(1, 84) = 1.09$, $p > .10$, MSE = 30655.37; low-frequency words with higher frequency neighbors were examined 55 ms longer than low-frequency words without higher frequency neighbors, $F_s(1, 20) = 5.47$, $p < .05$, MSE = 5,714.16, $F_i(1, 49) = 1.52$, $p > .10$, MSE = 29,499.55. Similarly, medium-frequency words with higher frequency neighbors were examined 60 ms longer than medium-frequency words without higher frequency neighbors, but this difference was not significant, $F_s(1, 20) = 2.70$, $p > .10$, MSE = 13,732.76, $F_i < 1$.

Experiment 3B: New Stimuli

The data for the first pass variables are listed in Table 8; the data for the spillover variables are listed in Table 9. The effect of word frequency was significant for gaze durations, $F_s(1, 39) = 8.21$, $p < .01$, MSE = 532.90, $F_i(1, 111) = 2.61$, $p > .10$, MSE = 1,458.26, and for the total time spent on the target word, $F_s(1, 39) = 14.50$, $p < .01$, MSE = 999.72, $F_i(1, 111) = 3.09$, $p = .08$, MSE = 3,342.46, but not for any of the other first-pass or spillover variables (all ps > .10).

With respect to neighborhood frequency, the results were very straightforward. For none of the first-pass or the spillover variables was there an inhibitory (or a facilitory) neighborhood frequency effect (all ps > .10), nor were there any statistically significant interactions between word frequency and neighborhood frequency (all ps > .10).

Individual differences: Percentage of regressions. As noted, Perea and Pollatsek (1998) reported that participants who made fewer regressions had longer gaze durations to words with higher frequency neighbors than to words without higher frequency neighbors, whereas the opposite was true for participants who made more regressions (i.e., there was a slight facilitory neighborhood frequency effect). As in Experiment 3A, we replicated Perea and Pollatsek's analysis by dividing our 40 participants into two equal groups, using a median split on the percentage of trials in

Table 8
First-Pass Eye-Movement Measures for Experiment 3B

	Neighborhood frequency	
Word frequency	No HF neighbors	HF neighbors
First-fixation duration (ms)		
Low-frequency words	225	217
Medium-frequency words	220	217
Gaze duration (ms)		
Low-frequency words	252	260
Medium-frequency words	245	245
Target words skipped (%)		
Low-frequency words	21.1	18.9
Medium-frequency words	18.6	17.1

Note. HF = higher-frequency; ms = milliseconds.

Table 9
Spillover Eye-Movement Measures for Experiment 3B

Word Frequency	Neighborhood frequency: No HF neighbors	HF neighbors
Total time on target word (ms)		
Low-frequency words	288	298
Medium-frequency words	276	272
First-fixation duration after target word (ms)		
Low-frequency words	226	223
Medium-frequency words	221	220
Total time on target word and the posttarget region (ms)		
Low-frequency words	637	629
Medium-frequency words	630	606
Target words regressed (%)		
Low-frequency words	12.4	16.1
Medium-frequency words	13.1	12.0

Note. HF = higher-frequency; ms = milliseconds.

which participants regressed back to the target word (*Mdn* = 12.9%, range = 0–32.5%). The mean percentage of regressions for the group with the lower percentage of regressions was 7.5%, and for the group with the higher percentage of regressions, 19.3%, $F_s(1, 38) = 66.47$, $p < .01$, $MSE = 83.18$, $F_i(1, 111) = 36.48$, $p < .01$, $MSE = 215.50$.

There was some evidence of an interaction between regression group and neighborhood frequency in the subject analysis of gaze durations, $F_s(1, 38) = 3.92$, $p = .06$, $MSE = 619.89$, $F_i(1, 111) = 1.73$, $p > .10$, $MSE = 1{,}715.30$. However, whereas Perea and Pollatsek found that the neighborhood frequency effect was inhibitory for participants who made fewer regressions, in our analysis there was no effect of neighborhood frequency for these participants for the low-frequency or the medium-frequency words (all *p*s > .10). Instead, in our analyses, if anything, the participants who made more regressions tended to exhibit an inhibitory neighborhood frequency effect, just the opposite of what Perea and Pollatsek found. For these participants, the effect of neighborhood frequency was marginally significant in the subject analysis, $F_s(1, 19) = 4.03$, $p = .06$, $MSE = 685.09$, $F_i(1, 111) = 1.16$, $p > .10$, $MSE = 2{,}355.23$. The gaze durations to words with higher frequency neighbors were an average of 12 ms longer than the gaze durations to words without higher frequency neighbors. There was no interaction between word frequency and neighborhood frequency (both *F*s < 1).

In contrast, for the participants who made fewer regressions, the effect of higher frequency neighbors was usually facilitory, the opposite of what Perea and Pollatsek (1998) observed. In the analysis of the first-fixation durations, there was some evidence of an interaction between regression group and neighborhood frequency, $F_s(1, 38) = 3.09$, $p = .08$, $MSE = 447.23$, $F_i < 1$. For the participants who made fewer regressions, first-fixation durations to words with higher frequency neighbors were 11 ms shorter than first-fixation durations to words without higher frequency neighbors, $F_s(1, 19) = 6.61$, $p < .05$, $MSE = 398.21$, $F_i(1, 111) = 2.06$, $p > .10$, $MSE = 1{,}110.75$. Similarly, in the analysis of the total time spent on the target word and the immediate posttarget region, there was some evidence of an interaction between regression group and neighborhood frequency, $F_s(1, 38) = 1.74$, $p > .10$, $MSE = 5{,}053.51$, $F_i(1, 111) = 4.71$, $p < .05$, $MSE = 8{,}276.11$, as the participants who made fewer regressions spent an average of 30 ms less time on the words with higher frequency neighbors than on the words without higher frequency neighbors, $F_s(1, 19) = 9.40$, $p < .01$, $MSE = 1{,}973.88$, $F_i(1, 111) = 3.05$, $p = .08$, $MSE = 13{,}899.91$, whereas for the participants who made more regressions there was no neighborhood frequency effect (both *F*s < 1). Statistically, most of these effects are marginal and unimpressive, and so we have little confidence in their replicability. All they may demonstrate is the absence of any consistent effect of neighborhood frequency.

Individual differences: High ART score versus low ART score. A median split on the ART scores (*Mdn* = 14, range = 3–26) created a low-ART group ($M = 9.1$) and a high-ART group ($M = 18.5$), $F_s(1, 38) = 68.77$, $p < .01$, $MSE = 12.77$. Like the situation in Experiment 3A, the high-ART group made more regressions than the low-ART group (16.1% vs. 10.9%), and in this experiment the difference was statistically significant, $F_s(1, 38) = 5.43$, $p < .05$, $MSE = 200.08$, $F_i(1, 111) = 10.37$, $p < .01$, $MSE = 170.25$. The correlation between the ART scores and the mean percentage of regressions was again essentially zero ($r = .04$), indicating that the ART and the regression data are measuring different individual differences associated with reading. Also like Experiment 3A, the participants with lower ART scores skipped the target words less often than those with higher ART scores, although in this experiment the difference was significant only for the low-frequency words (17.4% vs. 22.9%), producing an interaction between ART group and word frequency, $F_s(1, 38) = 4.78$, $p < .05$, $MSE = 98.80$, $F_i(1, 111) = 3.13$, $p = .08$, $MSE = 125.46$.

As was the case in Experiment 3A, there were a number of indications that the ART was tapping individual differences in reading ability. First, participants with lower ART scores were slower to read the target words than those with higher ART scores, as they had longer first-fixation durations (224 ms vs. 215 ms), $F_s(1, 38) = 1.11$, $p > .10$, $MSE = 3{,}388.53$, $F_i(1, 111) = 13.68$, $p < .01$, $MSE = 687.50$, and longer gaze durations (264 ms vs. 236 ms), $F_s(1, 38) = 4.74$, $p < .05$, $MSE = 6{,}398.58$, $F_i(1, 111) = 39.22$, $p < .01$, $MSE = 1{,}283.52$. The total time spent on the target word was also slightly longer for the participants with lower ART scores (292 ms for the low-ART group and 275 ms for the high-ART group), $F_s(1, 38) = 1.42$, $p > .10$, $MSE = 9019.89$, $F_i(1, 111) = 13.21$, $p < .01$, $MSE = 1{,}842.80$, as was the duration of the first fixation after leaving the target word (226 ms for the low-ART group and 219 ms for the high-ART group), $F_s < 1$, $F_i(1, 111) = 7.15$, $p < .01$, $MSE = 788.59$. The two groups also differed in the total time spent on the target word and the immediate posttarget region, with slightly longer latencies for the low-ART group (650 ms for the low-ART group vs. 598 ms for the high-ART group), $F_s(1, 38) = 2.47$, $p > .10$, $MSE = 44{,}510.00$, $F_i(1, 111) = 25.66$, $p < .01$, $MSE = 6{,}840.77$. (As was the case in Experiment 3A, differences between the groups were larger for the low-frequency words.)

Apart from these differences in reading times, the two groups differed from one another in one minor respect. In the analysis of the first-fixation durations after leaving the target word (a spillover effect), the interaction between ART group and neighborhood frequency was significant, $F_s(1, 38) = 7.48$, $p < .01$, $MSE = 433.99$, $F_i(1, 111) = 5.58$, $p < .05$, $MSE = 788.59$. For the low-ART group, when the target word had higher frequency neighbors, the first fixation after the leaving the target was an average of 11 ms shorter than when the target word had no higher frequency neighbors (a facilitory neighborhood frequency effect), $F_s(1, 20) = 4.74$, $p < .05$, $MSE = 490.13$, $F_i(1, 111) = 3.19$, $p = .08$, $MSE = 1{,}025.49$. This was not true for the high-ART group, as there was no effect of neighborhood frequency, $F_s(1, 18) = 2.89$, $p > .10$, $MSE = 371.62$, $F_i(1, 111) = 1.56$, $p > .10$, $MSE = 857.81$.

Discussion

The most important finding of Experiments 3A and 3B was the absence of an inhibitory neighborhood frequency effect on first-fixation durations or on gaze durations. Further, in contrast to Perea and Pollatsek's (1998) results, there was also little or no evidence for an effect of neighborhood frequency on any of the spillover variables measured in Experiments 3A and 3B, despite the expectation that the effect would be most evident on these variables. Also in contrast to Perea and Pollatsek's results, incorporating individual differences (percentage of regressions and reader skill) into the analyses did not provide any evidence that some readers had longer gaze durations (or longer first-fixation durations) to words with higher frequency neighbors. Incorporating reader skill into the analyses of the spillover variables in Experiment 3A did produce a few results consistent with Perea and Pollatsek's report that some readers were more likely to show inhibitory neighborhood frequency effects on these variables. None of these results, however, were replicated when using the new sentences created for Experiment 3B.

General Discussion

A key prediction of activation-based models of word identification is that words with higher frequency neighbors will be processed more slowly (and less accurately) than words without higher frequency neighbors. This is a consequence of the lexical competition mechanism embodied in these models: Words with higher frequency competitors (higher frequency neighbors) experience more interlexical inhibition than words without higher frequency competitors and are, therefore, slower to accumulate activation and to reach an activation threshold. The existence of an inhibitory neighborhood frequency effect in word identification tasks is thus a critical test of the models' assumption that competitive activation is central to lexical selection.

Most of the support for the inhibitory neighborhood frequency effect has come from studies in languages other than English (French, Spanish, and Dutch); the studies that have used English stimuli typically report null or facilitory neighborhood frequency effects (e.g., Forster & Shen, 1996; Sears et al., 1995; Siakaluk et al., 2002), the notable exception being Perea and Pollatsek's (1998) experiments. Perea and Pollatsek concluded that an inhibitory effect of neighborhood frequency could be observed for English words in a lexical decision task but only when the words are very low in normative frequency. On the basis of a second experiment, they also concluded that a word's higher frequency neighbors do not have any direct and immediate effect on reading time (as assessed by first-fixation durations and gaze durations) but do affect later stages of processing (i.e., after a reader has left a word). However, on the basis of their analyses of individual differences, Perea and Pollatsek left open the possibility that, for some readers (those who make few regressions while reading), higher frequency neighbors may directly affect reading time.

The purpose of the present research was to follow up on Perea and Pollatsek's (1998) assessment of the effect of higher frequency neighbors on the identification of English words. Our four lexical decision experiments were designed to assess the generalizability of Perea and Pollatsek's lexical decision results across both items and task instructions. In Experiment 1, we used Perea and Pollatsek's word and nonword stimuli and varied the lexical decision instructions given to participants. In Experiment 1A, participants were instructed to stress accuracy when responding, the same instructions Perea and Pollatsek used, whereas in Experiment 1B participants were given lexical decision instructions that did not stress accuracy (standard lexical decision instructions). The results of Experiment 1A were essentially the same as Perea and Pollatsek's results: An inhibitory neighborhood frequency effect was observed for the low-frequency words but not for the medium-frequency words. However, in Experiment 1B, there was no neighborhood frequency effect for the low-frequency words or for the medium-frequency words.

This outcome implied that the lexical decision instructions given to participants may be the critical determinant of whether or not there is an inhibitory neighborhood frequency effect (i.e., only when participants respond slowly and carefully, giving preference to accuracy over speed, does the effect appear). To assess this possibility, in Experiment 2, we used the identical experimental design with a new and larger set of word and nonword stimuli. Given the pattern of results from Experiments 1A and 1B, the expectation was for an inhibitory neighborhood frequency effect in Experiment 2A, in which participants were instructed to stress accuracy over speed (as in Experiment 1A and in Perea & Pollatsek's, 1998, experiment), but no inhibitory neighborhood frequency effect in Experiment 2B, in which participants were given standard lexical decision instructions (as in Experiment 1B). There was, however, no evidence of an inhibitory neighborhood frequency effect in either experiment. Consequently, considered together, our lexical decision experiments indicate that the inhibitory effect of neighborhood frequency in Perea and Pollatsek's experiment was due to the particular combination of lexical decision instructions and word and nonword stimuli used in that experiment. The fact that the effect did not generalize across items or across task instructions has obvious implications for its relevance to the predictions of activation-based models.

In Experiment 3 we recorded the eye movements of participants while they read sentences that contained the words used in Experiments 1 and 2. Experiment 3A used the sentences Perea and Pollatsek (1998) created for their words (the words used in the present Experiment 1 and in Perea and Pollatsek's Experiment 1), making Experiment 3A an attempted replication of the only experiment designed specifically to examine the effects of higher frequency neighbors on reading times. Like Perea and Pollatsek, we did not find an effect of neighborhood frequency on first-fixation durations or on gaze durations, implying that there is no effect of neighborhood frequency while the participants fixate the target words. Unlike Perea and Pollatsek, we did not find an overall effect of neighborhood frequency on any of the spillover variables measured either. In Experiment 3B, a new set of sentences was created for the words used in Experiment 2, and in that experiment there was no effect of neighborhood frequency on any of the first-pass or spillover variables. Our conclusion is that a word's higher frequency neighbors have no direct effect on reading times and have little, if any, effect on postidentification processing either.

Finally, in all of our experiments, we examined the impact of reader ability on the neighborhood frequency effect. In none of our lexical decision experiments was there any evidence that the neighborhood frequency effect was modulated by reader ability, as measured by the ART, in contrast to the clear interactions between reader ability and the word frequency effect. Although there were some indications that better readers were more prone to an inhibitory neighborhood frequency effect in Experiment 3A, particularly on the spillover variables, our confidence in these results is not high given the number of analyses performed and the fact that similar findings were not obtained in Experiment 3B. Nevertheless, our results do not rule out the possibility that reader ability could be a source of individual differences in the neighborhood frequency effect. This issue may deserve further attention.

Our results appear to pose a rather serious problem for activation-based models like the interactive activation model (McClelland & Rumelhart, 1981) and the multiple-read out model (Grainger & Jacobs, 1996). As can be seen in Table 3, simulations with the interactive activation model indicate that in all of our experiments the model predicts slower latencies to words with higher frequency neighbors than to words without higher frequency neighbors (for both the low-frequency and the medium-frequency words). When the words and the nonwords cannot be distinguished from one another on the basis of lexical activation, as was the case with the stimuli used in Experiments 2A, 2B, and 3B (see Figure 1), the multiple read-out model also predicts an inhibitory neighborhood frequency effect. Again, this is a consequence of the lexical competition mechanism embodied in these models: Words with higher frequency neighbors experience more interlexical inhibition than words without higher frequency neighbors and are, therefore, slower to accumulate activation and to reach an activation threshold. The problem for the models is that in only one of our six experiments was there a clear inhibitory neighborhood frequency effect (Experiment 1A) and, contrary to the models' predictions, only for the low-frequency words.

These observations, together with those made by Siakaluk et al. (2002), suggest that these models overestimate the role of inhibition in the orthographic processing of English words. This is also made clear by comparing the models' predictions for the neighborhood frequency effect with their predictions for the word frequency effect: The models predict that the neighborhood frequency effect will be larger than the word frequency effect for both sets of stimuli used in our experiments. As can be seen in Table 3, for the stimuli used in Experiments 1A, 1B, and 3A (Perea & Pollatsek's, 1998, stimuli), the predicted word frequency effect (in terms of cycles of processing) is 0.18 cycles, and the predicted neighborhood frequency effect is 1.66 cycles. For the stimuli used in Experiments 2A, 2B, and 3B, the predicted word frequency effect is 0.40 cycles, and the predicted neighborhood frequency effect is 0.80 cycles. Given the substantial word frequency effects obtained (averaging 68 ms in the four lexical decision experiments), the predicted neighborhood frequency effects should have been readily observed. Yet, even in Experiment 1A, in which there was evidence of an inhibitory neighborhood frequency effect, there was no suggestion that the neighborhood frequency effect was even as large as the word frequency effect. Thus, at the very least, as currently instantiated, the lexical competition mechanism embodied in these models exaggerates the impact that lexical competition will have on the identification latencies for English words.

Is there an inhibitory neighborhood frequency effect in English, as there appears to be in French, Dutch, and Spanish? To date, the balance of the evidence suggests there is not. However, that does not necessarily mean that inhibitory processes play no role in the visual identification of English words. Certainly, lexical competition is more difficult to discern in English relative to other languages, possibly because the inhibitory process is simply more dominant in these other languages. English words are, on average, shorter than words in French, Dutch, and Spanish, with less variability in length (e.g., Carlson, Elenius, Granström, & Hunnicutt, 1985; Wimmer, Köhler, Grotjahn, & Altmann, 1994). Because shorter words have more neighbors as a result of the smaller number of orthographically permissible letter combinations (Andrews, 1997; Frauenfelder, Baayen, Hellwig, & Shrender, 1993), English words tend to have more neighbors than the words in these other languages. Thus, on average, English words will also have more higher frequency neighbors. In fact, most English words three to five letters in length do have higher frequency neighbors (Andrews, 1997; Siakaluk et al., 2002). This neighborhood structure for English words (i.e., larger neighborhoods and many higher frequency neighbors) may necessitate a lexical processor with weaker inhibitory connections than those in other languages. Otherwise, it might be extremely difficult for low-frequency words to accumulate enough activation to reach their identification thresholds.

Inhibitory processing in English may be more readily detectable in other experimental paradigms, such as masked priming using word neighbor primes, which may permit a more fine-grained examination of lexical inhibition. (In fact, Davis & Lupker, in press, and Nakayama, Sears, & Lupker, 2005, have reported masked inhibitory priming using English stimuli, similar to the inhibitory priming reported by Segui & Grainger, 1990, using French stimuli.) Further, it is possible that inhibitory processing may be more readily detectable in readers of certain skill levels, as it is known that the activation of orthographic (and phonological) representations is affected by reader skill (e.g., Chateau & Jared, 2000; Unsworth & Pexman, 2003). In any event, it is becoming clear that there are important language differences in the role that inhibition plays in orthographic processing, and these differences and their origins deserve further study.

References

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 234–254.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin and Review, 4,* 439–461.

Balota, D. A., Cortese, M. I., Hutchison, K. A., Loftis, B., Neely, J. H., Nelson, D., et al. (2002). *The English Lexicon Project: A Web-based repository of descriptive and behavioural measures for 40,481 English words and nonwords.* Retrieved January 11, 2005, from http://elexicon.wustl.edu/default.asp

Carlson, R., Elenius, K., Granström, B., & Hunnicutt, S. (1985). Phonetic and orthographic properties of the basic vocabulary of five European languages. *Quarterly Progress and Status Report, 1,* 63–94.

Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of orthographic neighborhood in visual word recognition: Cross-task comparisons. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 857–871.

Chateau, D., & Jared, D. (2000). Exposure to print and word recognition processes. *Memory & Cognition, 28,* 143–153.

Cohen, J. (1976). Random means random. *Journal of Verbal Learning and Verbal Behavior, 15*, 261–262.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33A,* 497–505. Retrieved January 11, 2005, from http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale, NJ: Erlbaum.

Davis, C. J., & Lupker, S. J. (in press). Masked inhibitory priming in English: Evidence for lexical inhibition. *Journal of Experimental Psychology: Human Perception and Performance*.

Forster, K. I. (1976). Assessing the mental lexicon. In W. Marslen-Wilson (Ed.), *Lexical representation and processing* (pp. 75–107). Cambridge, MA: MIT Press.

Forster, K. I., & Hector, J. (2002). Cascaded versus noncascaded models of lexical and semantic processing: The *turple* effect. *Memory & Cognition, 30,* 1106–1117.

Forster, K. I., & Shen, D. (1996). No enemies in the neighborhood: Absence of inhibitory neighborhood frequency effects in lexical decision and semantic categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 696–713.

Frauenfelder, U. H., Baayen, R. H., Hellwig, F. M., & Shrender, R. (1993). Neighborhood density and frequency across languages and modalities. *Journal of Memory and Language, 32,* 781–805.

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.

Gordon, B. (1985). Subjective frequency and the lexical decision latency function: Implications for mechanisms of lexical access. *Journal of Memory and Language, 24,* 631–645.

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language, 29,* 228–244.

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

Grainger, J., O'Regan, J. K., Jacobs, A. M., & Segui, J. (1989). On the role of competing word units in visual word recognition: The neighborhood frequency effect. *Perception and Psychophysics, 45,* 189–195.

Grainger, J., & Segui, J. (1990). Neighborhood frequency effects in visual word recognition: A comparison of lexical decision and masked identification latencies. *Perception and Psychophysics, 47,* 191–198.

Huntsman, L. A., & Lima, S. D. (1996). Orthographic neighborhood structure and lexical access. *Journal of Psycholinguistic Research, 25,* 417–429.

Huntsman, L. A., & Lima, S. D. (2002). Orthographic neighbors and visual word recognition. *Journal of Psycholinguistic Research, 31,* 289–306.

Jacobs, A. R., & Grainger, J. (1992). Testing a semistochastic variant of the interactive activation model in different word recognition experiments. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 1174–1188.

Jared, D., Levy, B., & Rayner, K. (1999). The role of phonology in the activation of word meanings during reading: Evidence from proofreading and eye movements. *Journal of Experimental Psychology: General, 128,* 219–264.

Keppel, G. (1976). Words as random variables. *Journal of Verbal Learning and Verbal Behavior, 15,* 263–265.

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

Mathey, S., & Zagar, D. (1996). Rôle du voisinage orthographique lors de la reconnaissance visuelle des mots de 4, 6, et 8 lettres [The role of orthographic neighborhood in the visual recognition of 4-, 6-, and 8-letter words]. *Revue de Neuropsychologie, 6,* 205–217.

Mathey, S., & Zagar, D. (2000). The role of neighborhood distribution in visual word recognition: Words with single and twin neighbors. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 184–205.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88,* 375–407.

Nakayama, M., Sears, C. R., & Lupker, S. J. (2005, November). *Masked priming with orthographic neighbors: An interaction between relative prime–target frequency and neighborhood size.* Poster presented at the 46th Annual Meeting of the Psychonomic Society, Toronto.

Paap, K. R., & Johansen, L. S. (1994). The case of the vanishing frequency effect: A retest of the verification model. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1129–1157.

Perea, M., & Pollatsek, A. (1998). The effects of neighborhood frequency in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 767–779.

Perea, M., & Rosa, E. (2000). The effects of orthographic neighborhood in reading and laboratory word identification tasks: A review. *Psicológica, 21,* 327–340.

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

Pollatsek, A., Perea, M., & Binder, K. (1999). The effects of "neighborhood size" in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1142–1158.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with "the language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language, 41*, 416–429.

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin, 85,* 618–660.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 371–422.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading.* Englewood Cliffs, NJ: Prentice Hall.

Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton, C. (1989). Eye movements and on-line language comprehension processing. *Language and Cognitive Processes, 4,* SI 21–49.

Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 876–900.

Sears, C. R., Hino, Y., & Lupker, S. J. (1999). Orthographic neighborhood effects in parallel distributed processing models. *Canadian Journal of Experimental Psychology, 53,* 220–229.

Sears, C. R., Lupker, S. J., & Hino, Y. (1999). Orthographic neighborhood effects in perceptual identification and semantic categorization tasks: A test of the multiple read-out model. *Perception and Psychophysics, 61,* 1537–1554.

Segui, J., & Grainger, J. (1990). Priming word recognition with orthographic neighbors: Effects of relative prime-target frequency. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 65–76.

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

Siakaluk, P. D., Sears, C. R., & Lupker, S. J. (2002). Orthographic neighborhood effects in lexical decision: The effects of nonword orthographic neighborhood size. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 661–681.

Smith, J. E. K. (1976). The assuming-will-make-it-so fallacy. *Journal of Verbal Learning and Verbal Behavior, 15*, 262–263.

Stanovich, K. E., & Cunningham, A. E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory & Cognition, 20*, 51–68.

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly, 24,* 402–433.

Unsworth, S. J., & Pexman, P. M. (2003). The impact of reader skill on phonological processing in visual word recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 56*(A), 63–81.

van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language, 39,* 458–483.

Wike, E. L., & Church, J. D. (1976). Comments on Clark's "the language as fixed-effect fallacy." *Journal of Verbal Learning and Verbal Behavior, 15*, 249–255.

Wimmer, G., Köhler, R., Grotjahn, R., & Altmann, G. (1994). Towards a theory of word length distribution. *Journal of Quantitative Linguistics, 1,* 98–106.

Appendix

Items Used in Experiment 2

Low-Frequency Words Without Higher Frequency Neighbors

acre, brag, bred, buff, clog, cuff, emit, glee, hazy, raft, romp, slab, snag, snug, soda, alley, bloat, bulky, comic, cramp, crane, croak, mulch, petty, scoop, scrap, slang, super, tribe, wreck

Low-Frequency Words With Higher Frequency Neighbors

dorm, gene, glue, harp, herb, hike, moth, pour, reel, roam, robe, surf, tote, worm, yelp, barge, basin, dense, farce, focal, heave, leash, scent, sneak, snort, spear, stool, thief, valve, woven

Medium-Frequency Words Without Higher Frequency Neighbors

acid, bomb, bowl, copy, dirt, drug, gift, glad, jump, push, soap, span, suit, swim, tube, blame, cloud, crawl, curve, delay, guilt, lodge, loose, merge, moist, relax, shift, steel, straw, trust

Medium-Frequency With Higher Frequency Neighbors

calf, foam, fuel, gown, hero, horn, kiss, knee, noon, pair, risk, shoe, tool, wool, yard, beard, blond, bloom, chill, flood, prime, prize, skill, smart, smell, spend, stall, storm, suite, worse

Nonwords

aton, bace, balt, beld, bire, boad, bope, bort, bown, brab, chep, cing, clar, dace, dast, doss, dute, faie, fand, fank, farg, fing, forp, fost, fure, gark, gost, gulm, helt, jame, lape, leck, mant, marl, nime, noot, nush, pait, pean, plem, puel, rean, ribe, rist, sark, skib, slan, sork, sund, surl, swid, tean, tond, tord, tunk, vade, vire, wend, whot, wull, barch, begen, blace, blate, blick, brack, crake, creat, crill, datch, eatch, fatch, glake, gough, grase, greep, gress, gried, grome, herry, hetch, hower, jaked, litch, maken, marth, mater, mired, morth, naste, natch, nevel, parth, pelch, plake, porse, pribe, rable, reace, seave, shafe, shart, shaze, smill, snock, sount, spide, stabe, stape, starm, steck, stort, swart, sweft, touth, trave, trown, vired, whare, whone

Sentences Used in Experiment 3B

Low-Frequency Words

Patricia said that the [hike, romp] would benefit us all.
David learned to [surf, raft] while on vacation last summer.
That was a [sneak, petty] attack and he should be ashamed.
Samantha started to [pour, buff] the wax on to the hood of the car.
Mary said that the [harp, slab] was too heavy for her to lift alone.
They thought the [snort, croak] came from behind that tree.
She forgot her [tote, soda] on the kitchen counter at home.
She returned the [worm, mulch] to the compost heap.
My dad said that the [spear, tribe] came from an area of Brazil.
The guide said that the [barge, wreck] was at the bottom of the ocean.
The [stool, comic] was in the center of the stage.
She made a [focal, super] point in her speech on health care.
He started to [yelp, cramp] as he neared the finish line.
Martin said that the [valve, clog] was not allowing the water to drain.
Justin said that the [dorm, alley] was not a comfortable place to sleep.
The [basin, acre] was located next to the Pacific Ocean.
The [moth, snag] made a large hole in the camper curtains.
The fashion editor said that the [robe, cuff] made the garment look cheap.
The [woven, bulky] jacket was her favorite on rainy days.
The [dense, hazy] air made it difficult to finish the race.
They [roam, bred] the ponies at a large ranch in the foothills.
The [farce, slang] that was added to the script made the play more enjoyable.
He left the [reel, scoop] in the bottom of the boat.
He started to [brag, glue] but then he realized his mistake.
The [gene, crane] took years of engineering to modify.
The ground started to [heave, emit] molten lava on to the highway.
The [herb, snug] garden had just enough room for everything she wanted to grow.
The [leash, scrap] was made out of iron and aluminum.
She found the [thief, glee] from her first case difficult to restrain.
The [bloat, scent] of the carcass made Hank feel nauseous.

Medium-Frequency Words

Sara said that the [gown, suit] in the store window looked expensive.
Tommy liked to play in the [yard, dirt] on hot summer days.
Mrs. Mackie made John another [pair, copy] in case he lost his.
Matthew said that a [horn, bomb] went off a couple blocks away.
Donna said that the [foam, tube] was for packaging a parcel.
My professor said that [tool, drug] use is common among animals.
The pamphlet outlined the [risk, swim] involved in the triathlon.
The mayor said that Nicholas was a [hero, gift] to the community.
The fireman said there was [fuel, acid] spilled all over the road ahead.
The boys like to [kiss, push] the girls during recess.

(*Appendix continues*)

Justin said that the [suite, lodge] was already booked for the weekend.
The [knee, jump] had to be repaired before her next competition.
The [storm, cloud] rolled in from the north end of the lake.
The [calf, straw] grew to be four feet tall in just one season.
The salesman said the [wool, steel] his company makes lasts a lifetime.
The [shoe, bowl] was made in England during the Victorian Period.
His [beard, soap] lathered up nicely, making shaving easier.
The [noon, shift] meeting was rescheduled for later in the day.
Her [skill, span] for remembering numbers greatly improved with practice.
She looked [worse, glad] after the eight hour operation was over.
Andrew wanted to [spend, relax] all day reading his favorite book.
Nancy said that the [blond, moist] cake from the bakery was delicious.
The lawyer wanted to [prime, trust] her client before the trial.
The [smell, delay] was almost unbearable.
He knew the [prize, blame] was eventually going to him.
The [bloom, curve] along the garden path was beautiful and graceful.
A numbing [chill, guilt] seemed to grip her entire body.
The cat was not [smart, loose] enough to wiggle free from inside the pipe.
Just as he began to [stall, merge] he was hit by another car.
The [flood, crawl] of cars on the freeway made David take another route.

Note. The target words are in brackets; the word with higher frequency neighbors is listed first, the word without higher frequency neighbors is listed second.

Received February 11, 2005
Revision received December 14, 2005
Accepted December 15, 2005 ■