

# Neighborhood effects in visual word recognition: Facilitatory or inhibitory?

JOAN GAY SNODGRASS and MIRIAM MINTZER  
New York University, New York, New York

In five experiments, in which subjects were to identify a target word as it was gradually clarified, we manipulated the target's frequency of occurrence in the language and its neighborhood size—the number of words that can be constructed from a target word by changing one letter, while preserving letter position. In Experiments 1-4, visual identification performance to screen-fragmented words was measured. In Experiments 1 and 2, we used the ascending method of limits, whereas Experiments 3 and 4 presented a fixed-level fragment. In Experiment 1, there was no relation between overall accuracy and neighborhood size for words between three and six letters in length. However, more errors of commission (guesses) were made for high-neighborhood words and more errors of omission (blanks) were made for low-neighborhood words. Letter errors within guesses occurred at serial positions having many neighbors, and these positions were also likely to contain consonants rather than vowels. In Experiment 2, a small *facilitatory* effect of neighborhood size on both high- and low-frequency words was found. In contrast, in Experiments 3 and 4, using the same set of words, *inhibitory* effects of neighborhood size, but only for low-frequency words, were found. Experiment 5, using a speeded identification task, showed results parallel to those of Experiments 3 and 4. We suggest that whether neighborhood effects are facilitatory or inhibitory depends on whether feedback allows subjects to disconfirm initial hypotheses that the target is a high-frequency neighbor.

For visual word recognition, the size of a word's graphic neighborhood is usually measured with the N-metric, proposed by Coltheart, Davelaar, Jonasson, and Besner (1977). In their definition, the "N" of a letter string is the number of different English words that can be produced by changing just one of the letters in the string to another letter, preserving letter positions.

In recent years, the effect of the size of a word's neighborhood on its visual recognition has attracted attention because the direction of the effect—inhibitory or facilitatory—would appear to sharply constrain models of lexical access. For example, Andrews (1989) has recently argued that her finding that large neighborhoods *facilitate* visual word recognition rules out serial search models of lexical access (Forster, 1976, 1987) and the activation-verification model of Paap, Newsome, McDonald, and Schvaneveldt (1982) in favor of an activation model such as that proposed by McClelland and Rumelhart (1981; Rumelhart & McClelland, 1982), in which top-down activation occurs from the word to the letter level. However, a perusal of the visual word recognition literature reveals that a variety of effects of neighborhood size have been observed, from facilitatory (Andrews, 1989, 1992; Laxon, Coltheart, & Keating, 1988), to zero (Coltheart

et al., 1977), to inhibitory (Grainger, 1990; Grainger, O'Regan, Jacobs, & Segui, 1989; Grainger & Segui, 1990). In the auditory word recognition literature, the consensus seems to be that neighborhood density (a combination of the size of a word's neighborhood and the frequency of its neighbors) is *inhibitory* (Goldinger, Luce, & Pisoni, 1989; Luce, 1986; Luce, Pisoni, & Goldinger, 1990).

This paper was stimulated in part by a previous finding that there was no relationship between word recognition thresholds and neighborhood size for the names of the Snodgrass and Vanderwart (1980) pictures (Snodgrass & Poster, 1992). In the following sections, we first review the empirical literature on neighborhood effects and then consider the implications of these results for models of word recognition.

## Neighborhood Size Effects: Empirical Results

In studies of neighborhood size, a variety of tasks have been used to measure the speed of word recognition, which is usually assumed to have the speed of lexical access as one of its components. These tasks include the lexical decision task, the pronunciation task, the tachistoscopy recognition task, and the reading task. In this review, we focus on the lexical decision task because it is in this task that contradictory results have been reported.

Using lexical decision, Coltheart et al. (1977) found that neighborhood size had no effect on word classification times, but had an *inhibitory* effect on nonword classification times. Andrews (1989) reasoned that neighborhood size might be expected to have a greater effect on low-frequency than on high-frequency words, and that the lack

This research was supported by research grants AFOSR 89-0442 and F49620-92-J-0119 from the Air Force Office of Scientific Research to the first author. The authors wish to thank Meredith Poster for help with Experiment 1. Correspondence concerning this article should be addressed to J. G. Snodgrass, Department of Psychology, New York University, 6 Washington Place, Room 857, New York, NY 10003, (e-mail: gay@xp.psych.nyu.edu).

of an effect for words found by Coltheart et al. might have been due to incomplete control over frequency. Andrews varied both neighborhood size and word frequency orthogonally and found that neighborhood size had a *facilitatory* effect on word classification times, but only for low-frequency words. This effect was replicated in two lexical decision tasks in which the task was either difficult (the nonwords were very similar to the words) or easy (the nonwords were very dissimilar from the words), and in a pronunciation task. Andrews showed that the effects of neighborhood size on pronunciation latencies were not due to the production process itself by showing that neighborhood size effects disappeared in delayed pronunciation (after lexical access had presumably been accomplished). Andrews concluded that her results could not be accommodated by serial search models of lexical access, because these models predict that large neighborhoods should inhibit lexical access and thus slow lexical decision times. An activation model of word recognition in which bottom-up activation from the letter to the word level predominates, such as that of Paap et al. (1982), also predicts inhibitory effects of large neighborhoods because of competition from neighbors receiving bottom-up activation from common letters. According to Andrews's analysis, only a model in which activation from candidate words is fed back to the letter level, thereby increasing the activation of the target word, can predict that large neighborhoods will facilitate word recognition.

Grainger et al. (1989) argued that the important dimension of neighborhood size with respect to serial search models of lexical access was whether a word's neighborhood contained higher frequency words than itself. Using French words with native speakers of French, they compared lexical decision latencies for four classes of words: (1) words with zero neighbors, (2) words with neighbors of only lower frequencies, (3) words with one neighbor of higher frequency, and (4) words with more than one neighbor of higher frequency. Grainger et al. found that neighborhood size had an *inhibitory* effect on lexical decision latencies, but only when the target had one or more neighbors of higher frequency.

Recently, Grainger (1990) found that increasing the number of higher frequency neighbors did not increase the inhibitory effect of neighborhood size beyond the level obtained for words with only one higher frequency neighbor. These results contradict serial search models, which would predict a cumulative effect of the number of higher frequency neighbors because more words need to be verified before the target word can be identified. On the other hand, a simulation based on the McClelland and Rumelhart (1981) interactive-activation model was able to predict these results. But note that this model can predict either facilitatory or inhibitory effects of neighborhood size.

It seems reasonable to assume that low-frequency words have more higher frequency neighbors than high-frequency words do. If this is true, then Andrews's (1989, 1992) results are directly contradictory to those found by Grainger and his collaborators (Grainger, 1990; Grainger

et al., 1989), because Andrews found *facilitatory* effects for low-frequency words with large neighborhoods whereas Grainger et al. found *inhibitory* effects for words with high-frequency neighbors. Grainger (1990) attributed the contradictory results to Andrews's (1989) failure to control for bigram frequency in her set of words. He suggested that the facilitatory effects may not have been due to neighborhood size, but rather merely to the higher bigram frequency found in words with larger neighborhoods. Recently, however, Andrews (1992) replicated her results in a series of experiments in which she controlled for bigram frequency. She replicated the facilitatory effect of neighborhood size on lexical decision latency for low-frequency words, and found only a slight and non-significant inhibitory effect of neighborhood size on lexical decision latency for high-frequency words. Andrews (1992) also manipulated bigram frequency and showed that this frequency had no effect on lexical decision latency. So we are still left with contradictory results from two sets of studies, which cannot be resolved on the basis of stimulus differences.

Recently, Grainger and Segui (1990) introduced a new task, which they found to be more sensitive to the effects of neighborhood size. This task, which they call *progressive demasking*, is similar to Feustel, Shiffrin, and Salasoo's (1983) *continuous threshold latency identification* task. The signal (i.e., word) to noise (i.e., pattern mask) ratio is gradually increased until the subject, by pressing a button, can identify the word. Using this technique, Grainger and Segui (1990) found a strong inhibitory effect of neighborhood size on "masked identification" latencies for low-frequency words. This replicated their lexical decision results, and supported their position that high-frequency neighbors inhibit recognition of low-frequency targets.

For auditory word recognition, the best articulated model of neighborhood effects is the neighborhood activation model (NAM), developed by Luce (1986; see also Goldinger et al., 1989; Luce et al., 1990). This model considers both neighborhood size (the number of neighbors of a word) and neighborhood frequency (the frequencies of the neighbors). The size of a word's neighborhood is determined on the basis of confusion matrices between the word's phonemes and phonemes of potential neighbors. Luce et al. (1990) varied word frequency, neighborhood density, and neighborhood frequency for four tasks—perceptual identification, auditory lexical decision, auditory word naming, and primed perceptual identification—and found the usual advantage of high- over low-frequency words. They also found an inhibitory effect of both neighborhood size and neighborhood frequency; words from sparse and/or low-frequency neighborhoods were recognized more quickly and accurately than words from dense and/or high-frequency neighborhoods. Goldinger et al. (1989) replicated these effects in an auditory priming paradigm. The inhibitory neighborhood effects are accounted for by NAM by the competition from many and more frequent neighbors.

As the preceding review makes clear, the results from visual word recognition studies appear to be more contradictory than the results from auditory word recognition studies. For visual recognition studies that used the lexical decision task, Andrews (1989, 1992) obtained facilitatory effects, Coltheart et al. (1977) obtained no effects, and Grainger and his colleagues (Grainger, 1990; Grainger et al., 1989; Grainger & Segui, 1990) obtained inhibitory effects. For auditory recognition studies under a variety of tasks, inhibitory effects have consistently been obtained (Goldinger et al., 1989; Luce, 1986; Luce et al., 1990). We return to these contradictory findings, and our proposal for reconciling them, in the general discussion.

The experiments reported in this paper used a more direct method of measuring visual recognition—identification of fragmented words presented either with the ascending method of limits, or at a single, moderately fragmented level. Words are fragmented by a procedure we call *screen fragmentation*. In this procedure, the image of a word is fragmented by deleting blocks of pixels. The procedure differs from other word fragmentation techniques in which individual letters are deleted from the word (e.g., Tulving, Schacter, & Stark, 1982), and has the advantage of being appropriate for both short and long words. It differs from other perceptual identification techniques in which a word is presented briefly and then masked (e.g., Jacoby & Dallas, 1981), and has the advantage that each level of the fragmented word is presented for as long as the subject needs to generate a response.

In Experiment 1, we report an extensive analysis of threshold data collected by Snodgrass and Poster (1992) for the names of the Snodgrass and Vanderwart (1980) pictures. In Experiment 2, we orthogonally manipulate frequency and neighborhood size by using the set of four-letter words studied by Andrews (1989). In Experiments 3 and 4, we use the Andrews words with a different psychophysical method, and in Experiment 5 we use those words in a speeded perceptual identification task.

## EXPERIMENT 1

### Method

**Apparatus and Materials.** All subjects were tested individually on an Apple Macintosh Plus microcomputer. Of the 260 original pictures from the Snodgrass and Vanderwart set, 250 served as experimental stimuli to provide the names of the pictures. The set of 250 experimental stimuli were divided into two sets of 125 by assigning odd-numbered stimuli to one set and even-numbered stimuli to the other. Half of the subjects saw Set 1, and the other half saw Set 2. Three pictures selected from the alternative set served as practice stimuli. The experimental stimuli were presented in a random order, which was different for each subject. Words varied in length from 3 to 12 letters, and all two-word phrases were presented as one word.

The words were fragmented in the following manner. First, the word was printed in uppercase Basel typeface in 28-point size, centered within a 246 × 246 pixel square window. No word was longer than 12 letters, and all the words fit within the window. The words were 22 pixels high, and ranged from 49 to 245 pixels long. The fragmentation unit was an 8 × 8 pixel block, so the maximum number of blocks eligible for deletion was 120. The actual number of

blocks used by the fragmentation algorithm was determined by identifying those blocks within the window that contained black pixels. Across the set of 250 words, this number varied from 26 to 118 ( $M = 60$ ).

The words were fragmented to produce eight levels of fragmentation. The fragmentation was accomplished by randomly deleting blocks containing black pixels, according to the following exponential function:  $P = 0.85^{(8 - \text{level})}$ , where  $P$  is the proportion of critical blocks retained in the image, and level ranges from 1 (most fragmented) to 8 (complete word). So, at Level 1, 32% of the blocks were exposed, whereas at Level 4, 52% of the blocks were exposed. Fragments were added cumulatively from one level to the next, so that an  $n$ th-level stimulus contained all fragments shown at Level  $n - 1$ . Figure 1 shows three examples of words fragmented at the eight levels. The words were fragmented off line, and were stored for subsequent presentation during the experiment proper. Thus, each subject saw identical fragments of a particular word.

**Subjects and Procedure.** Forty-three students enrolled in an introductory psychology course at New York University served as subjects in Experiment 1. They received course credit for their participation in the study. The data of 3 subjects were discarded, because they had performed significantly more poorly than the others, leaving a total of 40 subjects.

When the subjects arrived, they were given a brief description of the experiment and then were asked to sign a consent form. Instructions displayed on the screen of the computer explained that they would be shown fragmented words and that their task was to identify them as soon in the sequence of presentations as possible. They were informed that the words were names of common objects and animals, that the first presentation of the word would be the most fragmented, and that they were to try to identify it. If they could not, they were to press the return key to go on to the more complete version. They were also instructed to type the word in lowercase, even though the word was presented in uppercase.

The word was presented at Level 1 first, and was gradually completed until the subject could correctly type it. The words had to be typed absolutely correctly for the program to score the response as correct. After the experiment, we examined the subjects' guesses prior to their correct responses, and any errors that indicated to the experimenters that the word had been identified were rescored as correct. These included some misspellings, capital letters, punctuation errors, and extra spaces. In particular, misspellings, which when pronounced were judged identical to the pronunciation of the correct spelling, were scored as correct. Nonletter symbols (', [, etc.) were deleted when error responses were analyzed. The subject's threshold for each word was defined as the level of fragmentation at which correct identification occurred.

The task took most of the subjects 45–60 min to complete. When they had completed the experiment, they received a written debriefing statement, and any questions about the experiment were answered.

### Results

For the purposes of this paper, we report threshold data for words of only three to six letters in length, because only six-letter words or shorter varied in neighborhood size. This constituted 68% (171) of the 250 words tested. Results for the entire set of words are presented in Snodgrass and Poster (1992).

Table 1 shows frequency in print (K-F), neighborhood size (N), and percentage of correct identifications (T%) for each of the four word lengths. Frequency-in-print counts were obtained from Kučera and Francis (1967). Frequency ranged from 0 (i.e., the word did not appear in the counts at all) to 591 (for *house*). Neighborhood size was obtained from the Franklin Computer Spelling Ace,

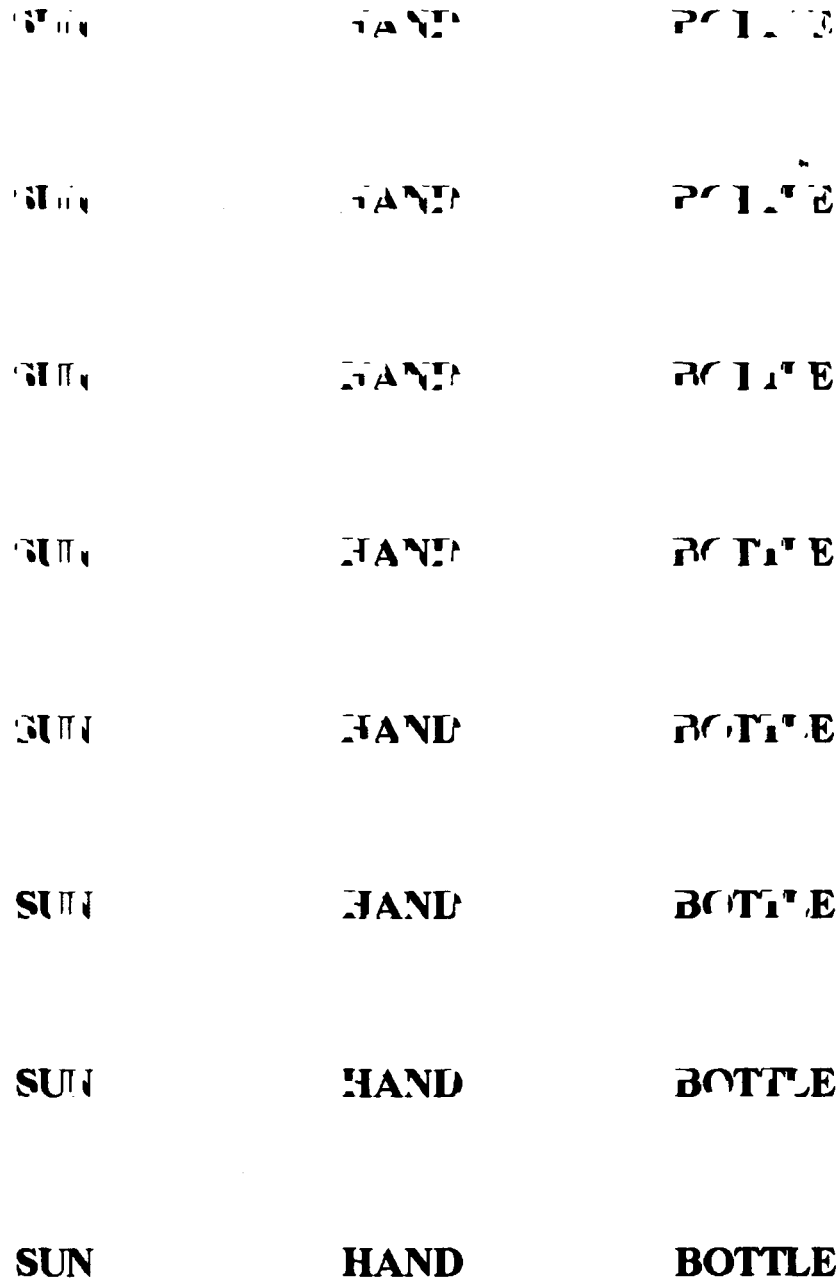


Figure 1. Examples of three fragmented words. Level 1 (the most fragmented stimulus) is shown at the top, and Level 8 (the complete stimulus) is shown at the bottom.

which contains more than 88,000 words in its lexicon (including proper names and abbreviations). The words were entered with the wild-card character “?” in the target position, and all words in the lexicon with letters in that position were counted.

Thresholds were converted to a percent-correct measure by the following transformation:  $T\% = 100 \times (9 - T) / 8$ , where T is the threshold, and varies between 1 (for identification at the most fragmented level) and 8 (for identification at the complete level). The T% measure can be literally interpreted as a percent-correct mea-

sure because it is the percentage of levels of the word that can be identified, assuming that identification at a particular level ensures identification at all higher levels. Therefore, subjects identifying a word at Level 1 would obtain a T% score of 100, indicating that they would have identified that word at all more complete levels.

As shown in Table 1, although both frequency and neighborhood size decreased with word length, identification performance (T%) appears unaffected by word length. Correlation coefficients were computed between T% and each of the three variables—word length, log (fre-

**Table 1**  
 Kučera-Francis (K-F) Frequencies, Neighborhood Size (N),  
 and Percentage of Correct Identifications  
 for the Four Word Lengths in Experiment 1

Word Length (No. of Letters)	K-F (Geometric Means)	N* (M)	T%†	No. of Words in Each Group
3	37.79	14.44	64.6	32
4	20.57	9.41	66.2	51
5	20.43	3.60	63.1	50
6	8.79	2.18	61.7	38

\*Number of words that can be constructed from a target word by changing one letter. †Transformed threshold, computed as  $T\% = 100 \times (9 - T)/8$ , where T is the level at which the word was identified.

quency + 1), and neighborhood size—across the 171 words. None of the three correlations approached significance (all  $ps > .22$ ).

**Predicting patterns of errors.** In order to explore the possible role of neighborhood size in more detail, we undertook an extensive exploration of errors. The subjects could make one of two types of errors—errors of omission (“blanks”), or errors of commission (“guesses”). A blank occurred when a subject failed to type any response onto the keyboard and simply hit the return key for the next most complete fragment. The occurrence of a blank presumably reflects the absence of a sufficiently strong hypothesis about the word’s identity. A guess occurred when a subject typed an incorrect response. Because typing a guess is more effortful than merely hitting the return key to register a blank, we assumed that anytime a subject typed a guess, this represented a legitimate hypothesis about the word’s identity.

On logical grounds, it would seem that words with many neighbors should produce many guesses, and words with few neighbors should produce many blanks. Therefore, total errors (the sum of guesses and blanks) might not show any relationship with neighborhood size, whereas the components of total errors (guesses and blanks) might show strong, but opposite, relationships with neighborhood size.

In order to test the hypothesis that words with many neighbors produced guesses, and words with few neighbors produced blanks, we carried out the following analysis. There was a total of 9,875 errors. Of these, 7,275 (74%) were blanks, and 2,600 (26%) were guesses. Guesses were classified as either those that matched the target word in length, or those that did not. This classification was made after all nonletter symbols were stripped from each guess. Of the 2,600 guesses, 1,755 matched the target word in length.

Table 2 shows the percentage of blanks and guesses for each word length. The number of errors was approximately constant across word length (consistent with the lack of correlation between word length and threshold). There were far more blanks than guesses at each word length (approximately a 3:1 ratio), but both the percentage of guesses and the percentage that matched the target in length decreased with word length.

In the following analyses, all guesses, regardless of whether they matched the target in length, were included. We looked separately at correlations between neighborhood size and both absolute and relative error frequencies for both types of errors. Although the correlation between neighborhood size and total errors was insignificant ( $r = -.11$ ), the correlation between neighborhood size and absolute number of guesses was significant and positive ( $r = +.26$ ), and the correlation between neighborhood size and absolute number of blanks was significant and negative ( $r = -.25$ ). When the absolute numbers of guesses were transformed into proportions by dividing by total errors for each target, the correlation between neighborhood size and guesses was even higher ( $r = +.49$ ); because the proportion of blanks was the inverse of the proportion of guesses, the correlation between neighborhood size and relative blanks was the same value, but opposite in sign ( $r = -.49$ ). These correlations thus support the hypothesis that a large neighborhood produced a particular type of error—a guess, whereas a small neighborhood produced another type of error—a blank.

**Serial position of neighborhood size.** The neighborhood size of a word can be broken down into components by the serial position of the mismatching letter. For example, the word *hand* has a neighborhood size of 9. There are five words that can be formed by changing the first letter (*band, land, rand, sand, and wand*), there is one word that can be formed by changing the second letter (*hind*), there is one word that can be formed by changing the third letter (*hard*), and there are two words that can be formed by changing the fourth letter (*hang* and *hank*). So, the serial position of neighborhood size for this word is U-shaped. As illustrated in this example, most neighbors are formed by replacing consonants by consonants and vowels by vowels. Because there are more consonants than vowels, the serial position of neighborhood size tends to be larger in positions occupied by consonants than in positions occupied by vowels.

When a guess is made, that guess can vary in the degree to which it misses the target. A guess of *form* to the target word *hand* misses the target in all four serial positions, whereas a guess of *sand* misses the target in only the first serial position. So any guess can be analyzed by the serial position of the letters in the guess that mismatch the letters in the target. We will call these mismatches *letter errors* to distinguish them from the more global class of error, word errors, referred to earlier. A word error

**Table 2**  
 Average Number of Errors, Percentage of Blanks, Percentage of  
 Guesses, and Percentage of Guesses Matching  
 the Target in Length in Experiment 1

Word Length (No. of Letters)	No. of Errors	% Blanks	% Guesses	% Match/ Guess
3	55.1	66.2	33.8	75.4
4	55.1	71.3	28.7	74.2
5	59.0	76.6	23.4	60.8
6	61.9	78.5	21.5	56.6

may be a guess or a blank, but a letter error can only occur in a guess that matches the target in length. A guess that matches the target in length can contain from 1 to  $n$  letter errors, where  $n$  is the target word length.

An interesting question is whether the frequency of letter errors bears any relationship to neighborhood size. We explored the relationship between the serial position of neighborhood size and the serial position of letter errors for guesses that matched the target in length. For three- and four-letter targets, particularly, there was a striking correspondence. For both three- and four-letter targets, neighborhood size was a U-shaped function of serial position, and so was the function for letter error frequencies; that is, the subjects tended to make letter errors at the ends of words and tended to be correct in the middle. Furthermore, the probability that a word in our corpus contained a consonant at a particular serial position was

also highly correlated with the neighborhood size of that serial position. Pearson correlations ranged from  $+0.76$  (for six-letter words) to  $+0.91$  (for three-letter words). Therefore, the likelihood that a particular serial position will contain neighbors is conditioned by the possibility that many other letters can occupy that position (i.e., by the likelihood that the position is occupied by a consonant). This suggests that serial position of neighborhood size is largely determined by whether a consonant or vowel is likely to occupy a particular serial position. Having a consonant in a particular serial position makes it likely that the neighborhood size will be large. Conversely, having a vowel in a particular serial position makes it likely that the neighborhood size will be small.

Figure 2 shows serial position functions for neighborhood size and letter error probability for three- and four-letter words. Both functions have been normalized so that

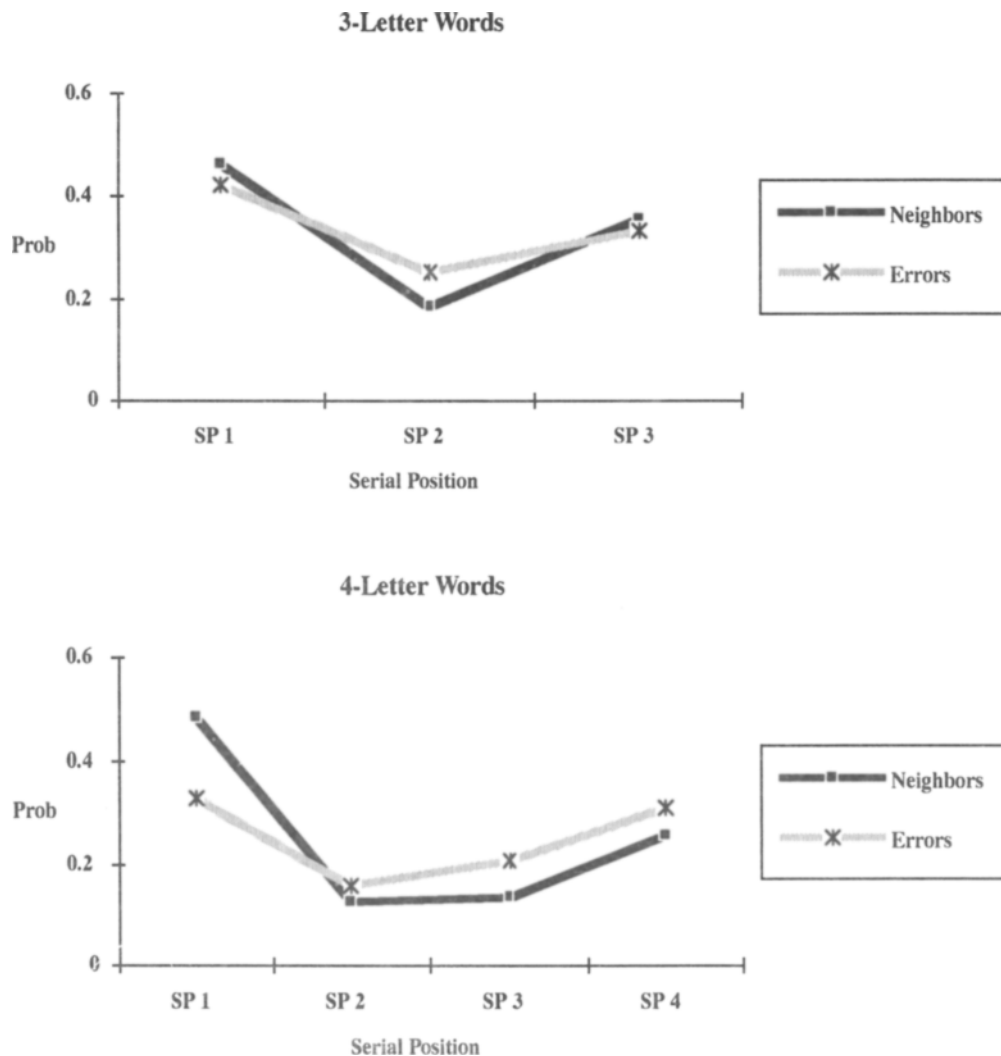


Figure 2. A comparison of neighborhood serial position (SP) functions and error serial position functions for three- and four-letter words for Experiment 1. Both functions have been normalized so that their sum across serial positions equals 1.0.

the sum of values across serial positions equals 1.0, so these represent relative, rather than absolute, functions. Letter error functions generally follow the neighborhood size functions, except that they tend to be flatter; that is, there are fewer errors at the beginnings and endings of words than would be expected. If presentations had been brief, one might attribute the flattening of the U-shaped error functions to lateral masking of the inner letters by the outer letters. However, the fragmented image was available for as long as the subject wished to inspect it. Alternatively, the fragmentation may make outer letters more distinct than inner letters, because their beginnings and endings are better defined. Finally, it is possible that subjects are better at filling in beginning or ending letters than middle letters, because of the organization of the lexicon.

### Discussion

Although no overall effect of neighborhood size on identification accuracy was observed, there were two important effects of neighborhood size on identification performance. First, the subjects were more likely to omit a response when neighborhood size was small, and they were more likely to guess when neighborhood size was large. This makes perfect sense because words with many neighbors, when not identified as themselves, can resemble other words. In contrast, words with few neighbors, when not identified as themselves, may resemble nothing at all.

Second, the serial position of neighborhood size predicted the pattern of letter errors. Neighborhood size showed strong serial position effects, which were generally U-shaped and followed consonant probability functions. Letter error functions were also U-shaped, showing that the subjects were more likely to make errors at serial positions where there were many neighbors (and where there was likely to be a consonant), and to make few errors at serial positions where there were few neighbors (and where there was likely to be a vowel).

In order to more carefully control both frequency and neighborhood size, word frequency and neighborhood size were orthogonally varied in Experiment 2. In addition, only four-letter words were used and the subjects were forced to type four-letter responses. This was expected to increase the guessing rate and thereby to increase the corpus of guess responses.

## EXPERIMENT 2

### Method

**Apparatus and Materials.** All the subjects were tested individually on an Apple Macintosh Plus microcomputer. There were four groups of four-letter words: low frequency–small neighborhood (low/small); low frequency–large neighborhood (low/large); high frequency–small neighborhood (high/small); and high frequency–large neighborhood (high/large). The four-letter words were taken from Andrews (1989) and an additional word was added to each set of 15 to create 16 four-letter words per group. Andrews's frequency counts were based on Carroll, Davies, and Richman (1971).

According to these counts, high-frequency words had a minimum frequency of 85, and low-frequency words had a maximum frequency of 36. According to Kučera-Francis (1967) counts, however, this set of high- and low-frequency words overlapped somewhat; high-frequency words had a minimum frequency of 4, whereas low-frequency words had a maximum frequency of 29.

According to Andrews's measure of neighborhood size, the large-neighborhood words had at least nine neighbors, whereas the small-neighborhood words had no more than five neighbors. We measured neighborhood size with the Language Master™ Desk Accessory from Proximity Technology, which contains more than 80,000 words in its lexicon. Neighborhood size was counted by entering each word with the wild-card character “?” in the desired position, and by counting the number of words returned, excluding proper names and abbreviations. According to this measure, the ranges of small- and large-neighborhood words overlapped somewhat; the large-neighborhood words had a minimum size of eight, and the small-neighborhood words had a maximum size of nine. More statistics on the characteristics of the target words will be presented below. An additional three words (*bird*, *comb*, and *poem*) served as practice stimuli. The experimental stimuli were presented in a random order, which was different for each subject.

Fragmented words were shown in uppercase Basel typeface in 28-point size, centered within a 246 × 246 pixel square window of a computer screen. Each word fit within an area that was 96 pixels long × 32 pixels high, or, when that area was divided into 8 × 8 pixel squares, into a 12 × 4 array of pixel blocks. The words were fragmented in the same manner as those from Experiment 1, except that the exponential rate parameter was set at 0.82 rather than 0.85 to make the words harder to identify. The words were fragmented off line and were stored at eight levels of fragmentation. Thus, each subject saw identical fragments of a particular word.

**Subjects and Procedure.** Sixteen students enrolled in an introductory psychology course at New York University served as subjects in Experiment 2. They received course credit for their participation in the study.

The procedure was identical to that used in Experiment 1, with the following exceptions. The subjects were told that all the words were four letters long and that they would be required to type in a four-letter response to each presentation of the word, even if they had no idea what the word was. The program was modified to continue only when the length of the response typed by the subject was exactly four letters. The subjects were instructed to type the response “pass” if they had no idea what the word was. The “pass” response was intended to correspond to the “blank” response of Experiment 1.

The word was presented at Level 1 first, and was gradually completed until the subject correctly typed the word. The words had to be typed absolutely correctly for the program to accept the response as correct. The subject's threshold for each word was defined as the level of fragmentation at which correct identification occurred.

The task took most of the subjects 30–45 min to complete. When they had completed the experiment, they received a written debriefing statement, and any questions about the experiment were answered.

**Frequency of neighborhood words.** In addition to measuring neighborhood size, we also measured Kučera-Francis frequency of each of the neighbors, as the results of Grainger and his colleagues, reviewed in the introduction, suggest that the number of neighbors of higher frequency than the target may be a more important measure of neighborhood size than the total number of neighbors.

Table 3 shows the means and ranges of target frequency and neighborhood size for the four classes of words. In addition, Table 3 presents four statistics on the frequency of neighbors of the target word. Mean NF is the average of the geometric mean K-F (Kučera-Francis) frequencies of neighbors of each target (excluding the single word having zero neighbors). There was no a priori

Table 3  
 Characteristics of the Words Used in Experiments 2, 3, 4, and 5

Word Class	TargetF	RangeF	MeanN	RangeN	MeanNF	MaxNF	%N > T	# > T
Low Freq, Small N	4.3	(0-23)	3.88	(1-9)	8.44	60.9	87.5	1.69
Low Freq, Large N	5.5	(0-29)	13.44	(8-18)	12.73	524.5	100.0	7.19
High Freq, Small N	113.8	(4-1747)	3.13	(0-7)	19.12	177.3	31.3	0.47
High Freq, Large N	77.8	(5-1013)	13.50	(8-20)	15.96	713.2	87.5	3.69

Note—TargetF is the geometric mean Kučera-Francis (K-F) frequency of target words; MeanN is the mean neighborhood size; MeanNF is the average of the geometric mean K-F frequency of neighbors of the target; MaxNF is the average *maximum* K-F frequency of neighbors of the target; %N > T is the percentage of target words having one or more neighbors of higher frequency; # > T is the average number of neighbors exceeding the target word in frequency.

reason to expect either frequency or neighborhood size to have an effect on this statistic. Surprisingly, however, high-frequency targets have higher frequency neighbors than low-frequency targets, whereas the effects of neighborhood size are inconsistent. The results of a 2×2 between-items analysis of variance (ANOVA) revealed that the main effect of frequency was reliable [ $F(1,59) = 5.30$ ,  $MS_e = 143.57$ ,  $p = .025$ ], but there was no effect of neighborhood size and no interaction. This tendency for high-frequency words to have high-frequency neighbors may be a peculiar property of this set of words, or it may be due to some more basic structural difference between high- and low-frequency words, such as consonant-vowel pattern. If it is a basic property of high-frequency words, here is yet another instance of differences between high- and low-frequency words that are not controlled by various matching strategies (Landauer & Streeter, 1973).

The second statistic is MaxNF, which is the *maximum* K-F frequency of the highest frequency neighbor of each target. On statistical grounds alone, we expect large-N targets to have higher scores than low-N targets, because the expected maximum value of a sample increases with its size (Gumbel, 1958). And, because high-frequency words have higher frequency neighbors than low-frequency words, we expect high-frequency targets to have higher scores than low-frequency targets. Both of these effects are evident in the data shown in Table 3. A 2×2 between-items ANOVA performed on the log frequency of MaxNF showed a significant effect of neighborhood size [ $F(1,59) = 35.62$ ,  $p < .001$ ], and a marginally significant effect of frequency [ $F(1,59) = 3.54$ ,  $p = .06$ ], but no interaction ( $MS_e = 2.03$ , for both).

The crucial question concerns the *relation* between the frequency of a target and the frequencies of its neighbors; specifically, the degree to which a neighbor will dominate a target in frequency. On the basis of statistical sampling, we expect that the neighbors of low-frequency targets will dominate their targets more often than the neighbors of high-frequency targets will. In the simplest possible case, consider a low- and a high-frequency target word that have a neighbor in common. If the neighbor lies between the two targets in frequency (e.g., *have*, *cave*, *pave*, where *cave* is the neighbor and *have* is the high-frequency target), then *cave* will dominate *pave*, but not *have*. According to Grainger's (1990) hypothesis, this neighbor will interfere with recognition of the low-frequency target, but not with recognition of the high-frequency target.

The likelihood that a target will have a higher frequency neighbor is indexed by two statistics in Table 3. %N > T gives the percentage of target words having one or more neighbors of higher frequency. As can be seen, both frequency and neighborhood size have an effect on this measure. In particular, *all* of the low-frequency large-neighborhood words have at least one neighbor of higher frequency.

The second measure is # > T, which gives the average number of neighbors exceeding the target word in frequency. This statistic shows even more extreme effects of neighborhood size, particularly for low-frequency words. The results of a 2×2 between-items ANOVA on # > T shows a highly significant effect of neighborhood size [ $F(1,59) = 50.58$ ,  $p < .0001$ ], a highly significant effect

of frequency [ $F(1,59) = 14.82$ ,  $p = .0003$ ], and a marginally significant interaction [ $F(1,59) = 3.46$ ,  $MS_e = 5.92$ ,  $p = .07$ ].

Taken together, these results support the idea that targets with large neighborhoods will have more higher frequency neighbors than targets with small neighborhoods. This, of course, is a necessary condition for Grainger's inhibition mechanism to work.

## Results

Table 4 shows the transformed threshold, T% [ $T\% = 100 \times (9 - T) / 8$ ], for the four categories of words. In contrast to Experiment 1, high-frequency words were identified better than low-frequency words: T% for high-frequency words was 51.1, and for low-frequency words was 46.9, a difference that was highly significant when the subjects were used as the units of analysis [ $F(1,15) = 11.14$ ,  $MS_e = 26.21$ ,  $p = .004$ ]. In contrast to Experiment 1, neighborhood size also had a significant effect. Words with large neighborhoods were identified better than words with small neighborhoods: T% for large-neighborhood words was 50.5, and for small neighborhoods was 47.6 [ $F(1,15) = 9.03$ ,  $MS_e = 14.95$ ,  $p = .008$ ]. These two main effects replicate those found by Andrews (1989). However, contrary to her results, there was no interaction. The advantage of large over small neighborhoods for low-frequency words (a difference of 2.4%) was not greater than the advantage of large over small neighborhoods for high-frequency words (a difference of 3.4%) ( $F < 1$ ). In fact, the difference, though not significant, was in the opposite direction.

The significant effects of frequency and neighborhood size were not replicated when items were used as the units of analysis [for frequency,  $F(1,60) = 1.04$ , and for neighborhood size,  $F < 1$ ]. Because of the overlap between both frequency classes and neighborhood sizes found with our measures, we also computed correlations between the T% measure and both frequency and neighborhood size. Neither correlation was significant.

Table 4  
 Percentage of Correct Identifications by Frequency and Neighborhood Size of the Words in Experiment 2

Frequency	Neighborhood Size		M
	Small	Large	
Low	45.7	48.1	46.9
High	49.5	52.9	51.1
M	47.6	50.5	



The lack of an effect for items was largely due to their high variability. Items varied in difficulty from a high of 89% for *game*, a high-frequency, large-neighborhood word, to a low of 20% for *raft*, a high-frequency, small-neighborhood word. Some of the variability in word identification performance can be attributed to the accidents of the fragmentation used. Some fragmentations were simply easier than others, over and above any effect of the word itself. This problem will be addressed in subsequent experiments by using fragmentation series that are randomly generated on each trial.

The main focus of interest was an analysis of error types (blanks and guesses). The subjects were instructed to type "pass" when they could not tell what a word was, so a pass response corresponded to the blank of Experiment 1. We expected that words with small neighborhoods would produce more pass responses than words with large neighborhoods, simply because there are fewer competitors for small-neighborhood words than for large-neighborhood words.

The overall percentage of blanks was much lower in Experiment 2 (20%) than in Experiment 1 (74%). This is presumably because typing the word "pass" was as difficult as typing any other four-letter response. The percentage of passes for high-frequency words was virtually identical to that for low-frequency words (20.1% and 20.0%, respectively) but, as predicted, words with small neighborhoods received more passes (21.6%) than those with large neighborhoods (18.6%). This difference was significant for both subjects [ $t(15) = 2.38$ ] and items [ $t(62) = 1.92$ ] (both one-tailed  $ps < .05$ ). However, the subjects differed widely in their use of the pass response. One subject responded with a total of 300 passes across the 64 words, but 5 subjects made no passes at all.

An examination of the responses made by the subjects who failed to use the pass response revealed that some of their responses might be viewed as stand-ins for the pass response in that they were nonwords that often contained repeating or easy-to-type sequences (e.g., "sqss" or "kdla"). In order to identify responses that were likely stand-ins for pass responses, we classified all guesses by their patterns of consonants (C) and vowels (V), and classified the two patterns CCCC and CCCV as stand-ins for passes.

Adding pass stand-ins added about 10% overall to the percentages, but did not change the pattern of results. Small-neighborhood words continued to produce more passes than large-neighborhood words, but there was no effect of frequency. The difference between large- and small-neighborhood words remained significant for subjects [ $t(15) = 1.90, p < .05$ ], but failed to reach significance for items [ $t(62) = 1.52$ ].

**Serial position of neighborhood size.** In Experiment 2, as in Experiment 1, a comparison of serial position (SP) functions for consonant probability, neighborhood size, and letter error frequencies showed remarkably similar profiles. Table 5 shows these three measures separately for small- and large-neighborhood words. Errors, includ-

Table 5  
Proportion of Consonants, Average Number of Neighbors, and Average Number of Letter Errors as a Function of Serial Position (SP) in Experiment 2

Measure/ Neighborhood Size	SP 1	SP 2	SP 3	SP 4
Consonants (large)	1.00	0.09	0.75	0.69
Consonants (small)	0.84	0.31	0.69	0.78
Neighbors (large)	6.50	1.59	2.59	2.78
Neighbors (small)	0.94	0.63	0.78	1.16
Letter Errors (large)	26.4	17.6	23.6	25.2
Letter Errors (small)	27.5	19.5	25.1	29.8

Note—*Consonants* refers to the probability of a consonant in each serial position in this set of words. *Neighbors* refers to the average number of neighbors at each serial position across the set of words. *Letter Errors* refers to the average number of letter errors in the subjects' guesses at each serial position of the targets.

ing passes (as defined above), were considered to be "true" guesses; that is, responses based on some stimulus information. The frequency of letter errors within guesses was counted by determining which positions within the guess did not match the corresponding letters in the target.

The consonant probability functions are similar to those for four-letter words from Experiment 1; that is, the predominant type of word is CVCC, regardless of whether the word has many or few neighbors. Consistent with this function, there are more neighbors for words in the consonant positions (SP 1, SP 3, and SP 4) than in the vowel positions, and the profiles are the same for large- and small-neighborhood words, even though the average number of neighbors is higher for large-neighborhood words.

Finally, the letter error functions follow the shape of the neighborhood size functions; that is, they show a U-shaped function with a minimum at SP 2, the serial position that is most likely to contain a vowel, and therefore is most likely to have few neighbors. Table 5 shows that the difference in average number of letter errors between large- and small-neighborhood words is quite small compared with the much larger differences between serial positions of letter errors. A 2 (neighborhood size)  $\times$  4 (serial position) mixed ANOVA on letter error frequencies for individual items failed to produce a significant difference between large- and small-neighborhood words ( $F < 1$ ), although it did produce a highly significant effect of serial position [ $F(3,186) = 7.34, MS_e = 145.78, p < .001$ ], and no interaction ( $F < 1$ ). Of course, frequency of letter errors is not perfectly related to frequency of guesses. However, frequency of guesses also did not differ between words with large neighborhoods and words with small neighborhoods (means of 40.7 and 41.0, respectively,  $t < 1.0$ ). The fact that frequency of guesses did not differ between small- and large-neighborhood words means that the effect of neighborhood size on thresholds is due entirely to differences in frequency of passes.

**Further analysis of guesses.** The serial position analysis described above suggests that incorrect guesses tend to share one or more letters with the target in corresponding serial positions. However, this analysis does not tell

us which guesses are neighbors. To be a neighbor, a guess must be a word and must share exactly three letters with the target in corresponding serial positions. In order to determine whether there were more guesses that were neighbors for the high- than the low-neighborhood words, guesses were classified by whether they were neighbors of the target, and by whether that neighbor had a higher frequency than the target.

In order to carry out this analysis, all responses up to and including the correct response were classified into the following four categories: correct responses, passes (as defined previously), guesses that are neighbors of the target (NGuesses), and neighbors that dominate the target in frequency (NFreq > TFreq). The remaining errors fall into a default category, ~NGuesses, which includes all other errors such as words that are not neighbors, neighbors that are not words, and nonwords that are not neighbors. Frequencies were converted to per-item percentages by dividing by the total number of opportunities. This number has a theoretical minimum of 16 (if all 16 subjects identified the target correctly at Level 1), and a theoretical maximum of  $16 \times 8 = 128$  (if all 16 subjects failed to identify the word until Level 8). In fact, total per-item opportunities ranged from 31 to 118.

Table 6 shows these average percentages by word class. Note that because correct responses were computed differently from Table 4, the absolute values differ, but the pattern remains the same.<sup>1</sup> We also show the NFreq > TFreq category conditionalized against the number of opportunities, NGuesses. This was done by dividing the NFreq > TFreq percentage by the NGuess percentage and multiplying by 100 within each word class. However, we did not conduct statistical analyses on this measure because so many of the subject-based and item-based NGuess means were zero, thereby producing many indeterminate values.

For our purposes, the important categories are NGuesses and NFreq > NTarget. As we might expect, more NGuesses were made to large- than to small-neighborhood words [ $F(1,60) = 14.94$ ,  $MS_e = 72.73$ ,  $p < .001$ ]. There was no significant effect of frequency, and no interaction (both  $F_s < 1$ ).

Given that NGuesses do not differ between frequency classes, and that low-frequency targets have more high-

frequency neighbors than high-frequency targets do (see Table 3), we would expect both neighborhood size and target frequency to have effects on NFreq > TFreq. This is exactly the pattern we find in Table 6, for NFreq > TFreq responses [for frequency,  $F(1,60) = 6.41$ ,  $p = .014$ , and for neighborhood size,  $F(1,60) = 10.57$ ,  $MS_e = 45.92$ ,  $p = .002$ , for both]. The interaction was not significant. We conclude from this analysis that, as the serial position analysis suggested, incorrect guesses are more likely to be neighbors of large-neighborhood than small-neighborhood targets, and neighbor guesses are more likely to dominate low-frequency, large-neighborhood targets.

### Discussion

In Experiment 2, unlike Experiment 1, there were significant effects of both frequency and neighborhood size on overall identification accuracy. However, these effects were quite small and did not generalize across items.

These effects replicated those reported by Andrews (1989) in lexical decision, except for the lack of interaction. There was not a greater effect of neighborhood size on low-frequency than on high-frequency words; the effects were equivalent. Because the effects of neighborhood size were facilitatory, they are contrary to those reported by Grainger et al. (1989) for lexical decision and gaze duration, and by Goldinger et al. (1989) for auditory word identification.

As in Experiment 1, much larger effects of neighborhood size were observed when error types were analyzed, and when the serial positions of letter errors were analyzed. The subjects made more passes for small-neighborhood than for large-neighborhood words and serial position functions of letter errors within guesses showed the same U-shaped function evident in the neighborhood serial position function, which in turn followed the consonant probability function.

Our results directly contradict the results and hypothesis of Grainger et al. (1989). Our low-frequency, large-neighborhood words do have more neighbors that dominate them in frequency than our high-frequency, large-neighborhood words do. Therefore, we would expect, on the basis of the Grainger hypothesis, that these higher frequency neighbors of low-frequency words would inhibit

Table 6  
Percentage of Responses per Item in Experiment 2 in Each Response Category

Word Class	Response Category				
	Correct Responses*	Passes*	NGuesses*	NFreq > TFreq	NFreq > TFreq/ NGuess
Low Freq, Small N	19.9	29.5	5.0	4.2	84.0
Low Freq, Large N	21.8	27.8	13.5	10.4	77.0
High Freq, Small N	21.3	30.2	5.7	0.6	10.5
High Freq, Large N	23.9	25.5	13.6	5.4	39.7

Note—Passes include CCCCs and CCCVs. NGuesses are neighbors of the target word that are also words. NFreq > TFreq is the percentage of total responses having higher frequencies than the target, and NFreq > TFreq/NGuess is the percentage of neighbor responses having higher frequencies than the target (based upon word class means). \*The sum of these categories equals 100% minus the percentage in the default category, ~NGuess.

their recognition. Instead, we get a small facilitation effect of neighborhood size.

However, more guesses were neighbors of the target for large- than for small-neighborhood words. We assume that when subjects guess the neighbor of a target word and are told that they are wrong, they eliminate this neighbor from further consideration. Therefore, guessing neighbors early in the ascending sequence of presentations could act to remove their inhibitory effect. Thus, one way to account for the lack of an inhibitory effect of neighborhood size is to assume that the subjects were able, through guessing, to eliminate more competitors for high-neighborhood than for low-neighborhood words. This explanation does not, of course, account for the small facilitatory effect of neighborhood size that was found in this experiment.

In the next two experiments, we introduce a single-trial procedure, in which the subjects were given only one opportunity to identify the stimulus. This procedure rules out the possibility of eliminating high-frequency-neighbor competitors of the target by the judicious use of guessing. If these early guesses prevented the subjects from showing inhibitory effects of large neighborhoods for low-frequency words in Experiment 2, the next two experiments should show the inhibition.

In addition, we compared two methods of presenting the words. In the fixed condition, the subjects were shown a single level of a moderately fragmented stimulus. In the ascending condition, the subjects were shown the same moderately fragmented stimulus, preceded by brief presentations of three less complete versions of the word, presented in an ascending series. In previous research (Snodgrass & Hirshman, 1991), we found that subjects did more poorly in identifying an ascending than a fixed presentation, and we attributed this to the activation of competitors of the target stimulus. The reason for introducing this manipulation here was to determine whether words with large neighborhoods would show relatively greater interference in the ascending versus fixed conditions, compared with words with small neighborhoods.

Finally, in order to eliminate extraneous sources of item variability caused by using identical fragmentation sequences, the words were fragmented on-line with a different random selection of pixel blocks on each trial. This random method of presentation is more comparable to on-line masking procedures in its introduction of variability at the point of stimulus presentation.

### EXPERIMENT 3

#### Method

**Apparatus and Materials.** All the subjects were tested individually on an Apple Macintosh Plus microcomputer. We used the same set of 64 experimental and 3 practice words that were used in Experiment 2.

To accomplish on-line fragmentation, the set of 48 pixel blocks was permuted and successive subsets of pixel blocks were selected for display according to the following exponential function:  $P = 0.82^{(8 - \text{level})}$ , where level was varied from 1 to 4 for the ascending

condition, and was fixed at 4 for the fixed condition. So at Level 1, 25% or 12 blocks were displayed, and at Level 4, 45% or 22 blocks were displayed. On each trial, the sequence of blocks was randomly permuted, so that each trial typically displayed a different subset of fragments. On ascending trials, the fragments were shown cumulatively; that is, the same subset of fragments shown at Level 1 were included in the fragments shown at Level 2, and so on.<sup>2</sup>

**Subjects and Procedure.** Thirty-one students enrolled in an introductory psychology course at New York University received course credit for their participation. The data for 1 subject were lost due to a technical malfunction, leaving a total of 30 subjects.

When the subjects arrived, they were given a brief description of the experiment, and then were asked to sign a consent form. Instructions displayed on a screen of the computer informed them that they would be shown four-letter words that would be fragmented, and that their task was to identify them. They were also informed that some of the words would be slowly completed on the screen, but others would immediately be shown in their most complete form.

Half of the words were shown in the ascending condition, and the other half were shown in the fixed condition. Two counterbalancing series were constructed so that each word would be seen in each condition an equal number of times. The subjects were instructed to type the word onto the computer keyboard in lowercase, even though the word was presented in uppercase, and were required to type a four-letter response, even if they had no idea what the word was. They were instructed to type the response "pass" if they had no idea what the word was. Words had to be typed absolutely correctly for the program to accept the response as correct. Because only one opportunity for a correct response occurred, the dependent variable was proportion correct.

In the ascending condition, fragments at Levels 1-3 were each presented for 1 sec. In both the ascending and fixed conditions, the fragment at Level 4 was presented until the subject responded. If the subjects typed a response that was not four letters long, they were instructed to try again. They were given feedback about whether their response was correct or not, and were told the correct name of the word. The feedback was displayed for 1 sec and was followed immediately by the next word.

The task took most of the subjects 30-45 min to complete. When they had completed the experiment, they received a written debriefing statement, and any questions about the experiment were answered.

#### Results and Discussion

The two presentation conditions, ascending and fixed, produced statistically equivalent performance (30% for ascending, and 33% for fixed), and presentation condition did not interact with any other variable, so the two conditions were combined. The lack of an effect for presentation condition is contrary to the results of Snodgrass and Hirshman (1991), who found that the fixed condition produced performance superior to that of the ascending condition. The lack of an effect here is most likely

Table 7  
Percentage of Correct Identifications by Frequency and Neighborhood Size of the Words in Experiment 3

Frequency	Neighborhood Size		<i>M</i>
	Small	Large	
Low	35.2	20.4	27.8
High	35.4	35.6	35.5
<i>M</i>	35.3	28.0	

Note—Data have been combined across ascending and fixed presentations.

due to the low level of performance. Subsequent investigations of the perceptual interference effect have revealed that it only occurs for the conditions used here when identification performance is well above 50% (Luo & Snodgrass, 1993).

Table 7 shows the percentage of correct identifications for each of the four classes of words combined across the ascending and fixed conditions. In Experiment 3, unlike Experiment 2, neighborhood size had an *inhibitory* rather than a *facilitatory* effect on word identification. Furthermore, the expected interaction between frequency and neighborhood size was obtained; that is, the inhibitory effect of neighborhood size only occurred for low-frequency words, just as Grainger's (1990) inhibition hypothesis predicts.

A 2 (presentation condition)  $\times$  2 (neighborhood size)  $\times$  2 (frequency) within-subjects ANOVA was carried out on the percentage of correct identifications. There was no main effect of presentation condition [ $F(1,29) = 1.69$ ,  $MS_e = 259.61$ ,  $p = .20$ ], a significant main effect of neighborhood size [ $F(1,29) = 12.19$ ,  $MS_e = 261.76$ ,  $p = .002$ ], and a significant main effect of frequency [ $F(1,29) = 14.48$ ,  $MS_e = 246.14$ ,  $p < .001$ ]. Most important, there was a significant interaction between neighborhood size and frequency [ $F(1,29) = 7.70$ ,  $MS_e = 438.60$ ,  $p = .010$ ]. Simple effects tests confirmed that the source of that interaction was a significant difference between large and small neighborhoods for low-frequency words, and no effect of neighborhood size for high-frequency words. No other interactions were significant.

The most striking finding in the present experiment is that large neighborhoods inhibit identification performance, but only for low-frequency words. The finding of inhibition rather than facilitation is contrary to the results reported by Andrews (1989, 1992), and to the results reported here in Experiment 2. It is, however, consistent with the results reported by Grainger and his colleagues (1989, 1990), and suggests that Andrews's results were not, contrary to the suggestions of Grainger (1990), due to the fact that bigram frequency was higher in words with large neighborhoods than those with small neighborhoods. As mentioned previously, Andrews (1992) showed that bigram frequency had no effect on lexical decision latency.

Before considering possible explanations for the effect, including those based on analyses of the errors from Experiment 3, we first present the next experiment. In Experiment 4, the design of Experiment 3 was replicated, with the following two changes. First, in order to boost performance somewhat, we increased the completion rate from .82 to .85. Second, in order to obtain more data on each trial, we continued the ascending and fixed series of presentations beyond Level 4 to whatever level was required for correct identification.

## EXPERIMENT 4

### Method

**Apparatus and Materials.** The same apparatus and materials that were used in Experiment 3 were used in Experiment 4, with the ex-

ception that the rate-of-completion parameter was changed from .82 to .85, so the exponential function used was  $P = 0.85^{(8-\text{level})}$ . This meant that at Level 1, 32% or 15 blocks were displayed, and at Level 4, 52% or 25 blocks were displayed. As before, fragmentation was accomplished on line, and the series of blocks was permuted randomly each time, so a unique fragmentation series was presented on each trial. And as before, on ascending trials the fragments were shown cumulatively.

**Subjects and Procedure.** Twenty-two students enrolled in an introductory psychology course at New York University served as subjects. They received course credit for their participation in the study. The data for 1 subject were lost due to a technical malfunction, leaving a total of 21 subjects. They were divided approximately equally between the two counterbalancings.

The procedure was identical to that used in Experiment 3, with the following additional step. After the word was presented at Level 4, it was gradually completed (up to Level 8, if necessary) until the subject correctly typed it. All prerecognition guesses were recorded. This change in procedure permitted us to analyze two dependent variables: identification accuracy at Level 4 (as in Experiment 3), and the subject's threshold, defined as the level of fragmentation at which correct identification occurred (as in Experiment 2).

The task took most of the subjects 30–45 min to complete. When they had completed the experiment, they received a written debriefing statement, and any questions about the experiment were answered.

### Results and Discussion

Two measures of performance were analyzed in this experiment. The first, identical to the measure used in Experiment 3, was the percentage of correct identifications at Level 4. The second was the percentage of identifications based on thresholds, in which the threshold was converted by the formula  $T\% = 100 \times (9 - T) / 8$ , where  $T$  is the threshold (the level at which the subject first identified the stimulus). This  $T\%$  measure is comparable to the  $T\%$  measure used in Experiment 2, except that here the minimum threshold is 4, so the maximum value of  $T\%$  is 62.5%, rather than 100%.

The overall levels of performance in the two presentation conditions were virtually identical, as was the pattern of results across neighborhood and frequency conditions. Accordingly, the data were collapsed across the two presentation conditions. Table 8 shows the percentage of correct identifications for the four types of words and two dependent variables. As can be seen, neighborhood size again had an *inhibitory* effect on identification performance, but only for low-frequency words. This same pattern was evident in both dependent measures.

The results of two 2 (neighborhood size)  $\times$  2 (frequency) within-subjects ANOVAs on the two dependent variables revealed approximately the same effects. For percent correct at Level 4, there was a significant main effect of neighborhood size [ $F(1,20) = 8.12$ ,  $MS_e = 105.93$ ,  $p < .001$ ], no main effect of frequency ( $F < 1$ ), and no interaction between neighborhood size and frequency [ $F(1,20) = 1.96$ ,  $p = .18$ ]. However, because the interaction had been significant in Experiment 3, we carried out simple effects tests, which showed that neighborhood size was significant for low-frequency stimuli [ $F(1,20) = 9.75$ ,  $MS_e = 110.31$ ,  $p = .005$ ], but not for high-frequency stimuli.

**Table 8**  
Percentage of Correct Identifications by Frequency and Neighborhood Size of the Words in Experiment 4

Frequency	Neighborhood Size		M
	Small	Large	
Percentage Correct at Level 4			
Low	55.4	45.2	50.3
High	53.3	50.6	51.9
M	54.3	47.9	
Transformed Threshold (T%)			
Low	52.1	49.1	50.6
High	52.1	52.3	52.2
M	52.1	50.7	

Note—Data have been combined across ascending and fixed presentations. Transformed threshold was computed as  $T\% = 100 \times (9 - T)/8$ , where T is the level at which the word was identified.

The same ANOVA on T% also showed a significant effect of neighborhood size [ $F(1,20) = 7.16, MS_e = 5.87, p = .015$ ], a marginally significant effect of frequency [ $F(1,20) = 3.53, MS_e = 15.96, p = .075$ ], and a significant interaction [ $F(1,20) = 4.35, MS_e = 12.93, p = .05$ ]. Simple effects tests on T% also showed that neighborhood size was significant for low-frequency words [ $F(1,20) = 15.12, MS_e = 6.47, p = .001$ ], but not for high-frequency words ( $F < 1$ ).

Accordingly, two experiments with somewhat different levels of performance (32% in Experiment 3 compared with 51% at Level 4 in Experiment 4) both show an inhibitory effect of neighborhood size on word identification, but only for low-frequency words.

This pattern of results differs from our results in Experiment 2 and from Andrews's (1989, 1992) results, but is consistent with those found by Grainger and his col-

leagues (Grainger, 1990; Grainger et al., 1989). The only difference between Experiment 2 and Experiments 3 and 4 was the fact that the subjects had no opportunity to guess until Level 4 in Experiments 3 and 4. We speculated earlier that this might prevent them from eliminating high-frequency competitors of low-frequency words early in the sequence. If this were true, we would expect to observe more high-frequency-neighbor guesses for low-frequency, high-neighborhood words in Experiments 3 and 4 than we had in Experiment 2. To test this conjecture, exactly the same analysis of incorrect responses that was used in Experiment 2 was carried out for Experiments 3 and 4.

**Analysis of types of errors.** Responses were divided into four classes: correct responses, passes (including stand-ins for passes, CCCCs, and CCCVs), guesses that are neighbors of the target (NGuesses), and neighbors that dominate the target in frequency (NFreq > TFreq). The remaining errors fall in the default category, ~NGuesses, which includes words that are not neighbors, neighbors that are not words, and nonwords that are not neighbors. The NFreq > TFreq measure is presented both as a percentage of total trials and as a percentage of NGuesses. As for the previous analysis, the latter measure was computed on the averages for each word class, because of the large number of individual subject or item means with zero NGuesses. For comparison, a similar analysis of Experiment 2, only for responses made at Level 4, is also shown. Table 9 presents these results.

**Pass responses.** In Experiment 3, as in Experiments 1 and 2, small-neighborhood words received more passes than large-neighborhood words. A 2 (neighborhood size)  $\times$  2 (frequency) ANOVA was conducted on the mean percentage of passes for both subjects and items. As expected,

**Table 9**  
Mean Percentages of Responses per Item in Experiments 3 and 4 Falling Into Each Response Category as a Function of Word Class, Compared With the Same Measures at Level 4 for Experiment 2

Word Class	Response Category				
	Correct Responses*	Passes*	NGuesses*	NFreq > TFreq	NFreq > TFreq/NGuess
Experiment 3					
Low Freq, Small N	35.2	18.3	12.5	8.8	70.0
Low Freq, Large N	20.4	16.3	32.3	24.4	75.5
High Freq, Small N	35.4	20.6	5.6	1.5	25.9
High Freq, Large N	35.6	11.7	24.2	11.0	45.7
Experiment 4 at Level 4					
Low Freq, Small N	55.4	7.7	11.6	7.7	66.7
Low Freq, Large N	45.2	8.3	22.9	16.1	70.1
High Freq, Small N	53.3	8.3	9.5	2.1	21.9
High Freq, Large N	50.6	4.2	27.1	12.8	47.3
Experiment 2 at Level 4					
Low Freq, Small N	25.0	29.3	6.6	5.5	82.4
Low Freq, Large N	31.3	23.4	16.4	9.4	57.1
High Freq, Small N	29.7	28.5	3.1	0.4	12.5
High Freq, Large N	38.7	19.1	9.4	3.9	41.7

Note—Passes include CCCCs and CCCVs. NGuesses are neighbors of the target word that are also words. NFreq > TFreq is the percentage of total responses having higher frequencies than the target, and NFreq > TFreq/NGuess is the percentage of neighbor responses having higher frequencies than the target (based on averages by word class). \*The sum of these categories equals 100% minus the default category, ~NGuess, within each word class.

words with small neighborhoods received more passes (19.5%) than words with large neighborhoods (14.0%). This difference was significant both by the subject analysis [ $F(1,29) = 18.43$ ,  $MS_e = 49.62$ ,  $p < .001$ ] and by the items analysis [ $F(1,60) = 5.24$ ,  $MS_e = 51.70$ ,  $p = .03$ ]. Neither the main effect of frequency nor the interaction were significant in either analysis.

In Experiment 4, the percentage of passes out of total wrong answers was computed at Level 4 in order to make the measure comparable to that used in Experiment 3. Overall, the mean percentage of passes at Level 4 was only 7.1%, which was much lower than in Experiment 3 (16.6%). This reduction in passes is probably due to the higher performance in Experiment 4 (although a similar decrease in incidence of NGuesses was not observed). In Experiment 4, neither neighborhood size nor frequency were significant in either the subject-based or the item-based ANOVAs on passes.

**NGuess and NFreq > TFreq responses.** Our major interest is whether the inhibition observed for low-frequency, large-neighborhood words can be attributed to a greater frequency of guesses that are neighbors of the target (NGuess responses), particularly those that surpass the target in frequency (NFreq > TFreq responses).

In both experiments, the percentage of NGuesses is higher for large- than for small-neighborhood words, and more NGuesses surpass the target in frequency for large- than for small-neighborhood words. Our claim here is that the low-frequency, large-neighborhood words are being inhibited by competition from high-frequency neighbors, which are still viable candidates because they have not been eliminated from contention. This claim is supported by the fact that, even though Experiments 3 and 4 differed in level of performance, both showed the greatest percentages of NFreq > TFreq responses for the low-frequency, large-neighborhood words.

In the following analyses, all ANOVAs are based on subjects, unless indicated otherwise. For Experiment 3, both neighborhood size and frequency had significant effects on percentage of NGuesses [for neighborhood size,  $F(1,29) = 156.24$ ,  $MS_e = 70.54$ ,  $p < .001$ ; for frequency,  $F(1,29) = 23.51$ ,  $MS_e = 71.79$ ,  $p < .001$ ]. The interaction was not significant.

For Experiment 4, only neighborhood size had a significant effect on percentage of NGuesses [ $F(1,20) = 68.43$ ,  $MS_e = 63.94$ ,  $p < .001$ ]. There was no effect of frequency and no interaction.

For Experiment 3, both neighborhood size and frequency had significant effects on percentage of NFreq > TFreq responses, and the interaction was significant [for neighborhood size,  $F(1,29) = 74.90$ ,  $MS_e = 63.63$ ,  $p < .001$ ; for frequency,  $F(1,29) = 61.40$ ,  $MS_e = 51.96$ ,  $p < .001$ ; for the interaction,  $F(1,29) = 4.30$ ,  $MS_e = 63.63$ ,  $p = .05$ ]. As shown in Table 9, the effect of neighborhood size on NFreq > TFreq responses is larger for low-frequency than for high-frequency words. The interaction between neighborhood size and frequency was exactly as

predicted, with low-frequency, large-neighborhood words receiving a particularly large percentage of responses.

In Experiment 4, the NFreq > TFreq measure was significant for both neighborhood size and frequency, but not for their interaction [for neighborhood size,  $F(1,20) = 40.24$ ,  $MS_e = 47.34$ ,  $p < .001$ ; for frequency,  $F(1,20) = 11.11$ ,  $MS_e = 37.67$ ,  $p = .003$ ; for the interaction,  $F < 1$ ]. The lack of an interaction for Experiment 4 is not consistent with our predictions; the fact that low-frequency, high-neighborhood words had the highest percentage of NFreq > TFreq responses stems from the fact that both frequency and neighborhood size were significant in the analysis.

**Comparison of error patterns in Experiment 3 with those in Experiment 2.** For comparison, the distribution of responses at a comparable level in Experiment 2 is also shown at the bottom of Table 9. Only responses that occurred at Level 4 were analyzed. Correct responses include all trials in which the item was identified prior to Level 4. Because the level of accuracy was approximately the same for Experiments 3 and 2 at Level 4, we concentrate on this comparison.

Figure 3 shows the difference in percentage of responses between Experiments 3 and 2 for the four response categories by word class. The low-frequency, large-neighborhood condition has been highlighted, because it shows the biggest difference in percentage of correct responses. For correct responses, Experiment 3 differs from Experiment 2 in showing lower performance for low-frequency, large-neighborhood words and higher performance for low-frequency, small-neighborhood words. Experiment 3 produced a smaller percentage of passes than Experiment 2, but this effect is seen uniformly across word classes. We assume that fewer passes occurred in Experiment 3 than in Experiment 2, because the subjects in Experiment 3 knew they had only one opportunity to identify the target, whereas those in Experiment 2 knew they would eventually identify the item. But the fact that this difference in passes occurs evenly across word class suggests that passes are not responsible for the different pattern of correct responses in the two experiments.

In contrast, Experiment 3 differs from Experiment 2 in both the NGuess and NFreq > TFreq categories, particularly for low-frequency, large-neighborhood words. Similar effects, but of smaller magnitude, are shown for high-frequency, large-neighborhood words. We would argue that these differences are attributable to subjects' having eliminated likely guesses earlier in the sequence, resulting both in better performance in Experiment 2 for the large-neighborhood words (particularly those of low frequency), and a higher incidence of NGuess and NFreq > TFreq responses in Experiment 3.

The results of the last two experiments suggest that high-frequency neighbors are fundamentally inhibitory and interfere with perceptual accuracy. However, much of the contradictory data on neighborhood size uses lexical decision latency, rather than error rates, as the dependent

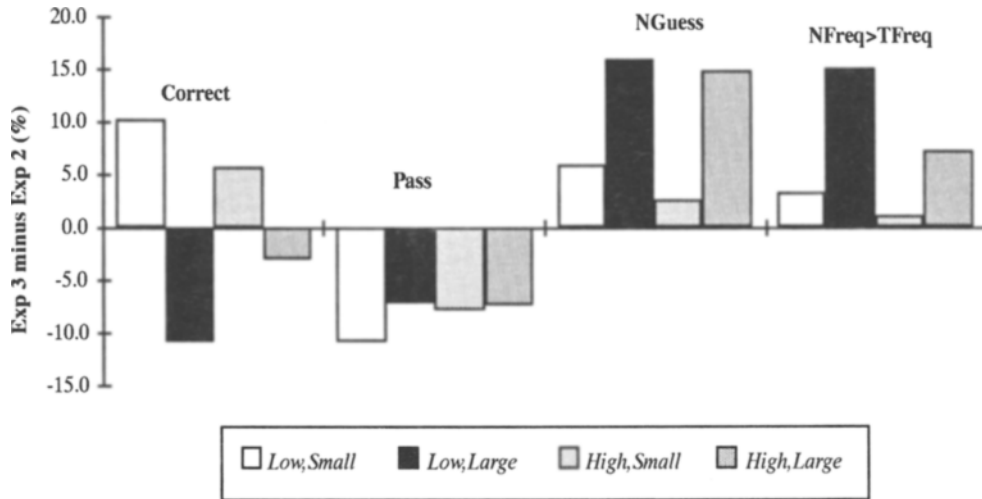


Figure 3. Differences in response patterns between Experiments 3 and 2 measured at Level 4 for each of the four response categories for each word class. Bars below the 0% line mean that Experiment 2 exceeded Experiment 3 on the measure, and those above the 0% line mean that Experiment 3 exceeded Experiment 2.

variable. Accordingly, in the next experiment, we used a latency measure in a speeded identification task. This task was similar to Feustel et al.'s (1983) continuous threshold latency identification task, and Grainger and Segui's (1990) progressive demasking task. In the speeded identification task, the subjects were presented with an ascending series of fragmented images, each presented for a fixed duration. They pressed a key as soon as they recognized the word, and the total duration of presentation (a latency measure) was the dependent variable.

## EXPERIMENT 5

### Method

**Apparatus and Materials.** Exactly the same apparatus and materials that were used in Experiment 4 were also used in Experiment 5. The rate of completion was kept at .85. As before, fragmentation was accomplished on line, and the series of blocks was permuted randomly each time, so a unique fragmentation series was presented on each trial. However, in contrast to Experiments 2 and 3, the series was presented rapidly and the subjects' task was to stop the series as soon as they could identify it.

**Subjects and Procedure.** Twenty-three students enrolled in an introductory psychology course at New York University served as subjects. They received course credit for their participation in the study.

When the subjects arrived, they were given a brief description of the experiment, and then were asked to sign a consent form. Instructions displayed on the screen of the computer informed them that they would be shown common four-letter words that would be fragmented, and that their task was to identify them. They were informed that each word would be presented as a series of fragmented images that would be quickly completed. The subjects were instructed to hit the space key on the keyboard as soon as they could identify the word. As soon as a subject hit the space key, the word was erased, and the subject was prompted to type in the word. They were urged to hit the space bar as soon as they could, but not before they could identify the word.

In order to motivate them to respond quickly and accurately, the subjects were awarded points according to how quickly in the series they hit the space key. They were given 8 points for stopping

the series at Level 1, 7 points for stopping it at Level 2, down to 1 point for stopping it at Level 8, but only if their subsequent response was correct. They did not receive any points if they did not correctly identify the word. After each word, the subjects were told the number of points they had received. The subject with the highest total number of points was awarded a prize of \$25.

The subjects were instructed to type the word in lowercase, even though the word was presented in uppercase, and were also required to type a four-letter response, even if they had no idea what the word was. The words had to be typed absolutely correctly for the program to accept the response as correct.

Before the experimental trials began, the subjects were permitted to practice on the word *bird* for as many trials as they wished. During this practice trial, they were given feedback about whether they were correct or incorrect, and were told the number of points they had earned when they were correct.

The speeded identification test consisted of 66 trials. The first 2 were practice trials and were not scored; the remaining 64 presented the words in the four classes of frequency and neighborhood size. Prior to each speeded trial, the message "get ready" appeared in the center of the screen, and was erased after 0.5 sec. Each level of fragmented image was then presented for 167 msec so that the entire series of eight fragmented images required 1,336 msec for presentation. As soon as the subject hit the space key, the level of fragmentation being shown at that instant was recorded, and the image was erased. The subject was then instructed to type the word.

Two dependent measures were obtained on each trial—the fragmentation level at which the clock was stopped (stop level) and whether the word was identified correctly (naming accuracy). The feedback was displayed for 1 sec, followed by a 2-sec intertrial interval.

The task took most of the subjects 15–20 min to complete. At the end of the experiment, the subjects were shown the total number of points they had won, and were thanked and given a written debriefing statement.

### Results

Table 10 shows stop level (our measure of identification speed) for correct responses only, and the percentage of naming errors for each category of word. Identification was slower for large-neighborhood, low-frequency

**Table 10**  
**Stop Level for Correct Responses and Naming Accuracy by**  
**Frequency and Neighborhood Size of the Words in Experiment 5**

Measure	Neighborhood Size/Frequency			
	Large/ High	Large/ Low	Small/ High	Small/ Low
Stop Level	6.45	7.16	6.68	6.89
Naming Errors (%)	14.9	18.8	9.5	17.1

words than for the other classes, which replicates the pattern of accuracy data from Experiments 3 and 4. In contrast to the accuracy data, however, identification was faster for large-neighborhood, high-frequency words. The pattern of naming errors shows similar, but not identical, effects.

A 2 (neighborhood size)  $\times$  2 (frequency) within-subjects ANOVA on stop level showed no main effects of neighborhood ( $F < 1$ ), a main effect of frequency [ $F(1,22) = 45.63$ ,  $MS_e = .105$ ,  $p < .001$ ], and a significant interaction [ $F(1,22) = 20.98$ ,  $MS_e = .068$ ,  $p < .001$ ]. Simple effects showed that neighborhood effects were significant at each frequency level: For low-frequency words, high-neighborhood words were identified slower than low-neighborhood words [ $F(1,22) = 11.73$ ,  $MS_e = .072$ ,  $p = .002$ ], but for high-frequency words, high-neighborhood words were identified *faster* than low-neighborhood words [ $F(1,22) = 7.13$ ,  $MS_e = .085$ ,  $p = .014$ ]. A comparable 2 $\times$ 2 between-items ANOVA on stop level showed essentially the same effects. Again, there was no main effect of neighborhood ( $F < 1$ ), a main effect of frequency [ $F(1,60) = 24.38$ ,  $p < .001$ ], and a significant interaction [ $F(1,60) = 6.73$ ,  $MS_e = .144$ ,  $p = .012$ ; for both]. However, for the item analysis, simple effects showed that neighborhood effects were significant only for low-frequency words [ $F(1,60) = 4.05$ ,  $MS_e = .144$ ,  $p = .05$ ]. Taken together, then, these results show that large-neighborhood, low-frequency words were identified more slowly than the other categories of words, and large-neighborhood, high-frequency words were identified more quickly (but only in the subject analysis). So, we replicated the inhibitory effects of neighborhood size for low-frequency words, but there was some evidence that size was facilitatory for high-frequency words.

A somewhat different pattern of results was obtained for naming errors. A 2 (neighborhood size)  $\times$  2 (frequency) within-subjects ANOVA on percentage of naming errors showed a significant main effect of neighborhood size, with small-neighborhood words identified more accurately than large-neighborhood words [ $F(1,22) = 4.53$ ,  $MS_e = 63.30$ ,  $p = .05$ ], a main effect of frequency [ $F(1,22) = 9.62$ ,  $MS_e = 77.81$ ,  $p = .005$ ], and no interaction ( $F < 1$ ). However, a comparable 2 $\times$ 2 between-items ANOVA on naming errors showed no main effect of neighborhood size ( $F < 1$ ), a main effect of frequency [ $F(1,60) = 6.87$ ,  $MS_e = 75.85$ ,  $p = .01$ ], and no interaction ( $F < 1$ ). Taken together, these results show that

high-frequency words were identified more accurately than low-frequency words after their series was stopped, but no other effects were significant in both analyses.

We hoped through instructions and payoffs to keep naming errors at a low and uniform rate across the four word categories, so that only speed need be considered. However, because only the frequency variable was significant for naming errors, and both measures showed an advantage for high- over low-frequency words, we do not need to qualify our conclusion that large-neighborhood, low-frequency words are inhibited in the speed of their identification.

## GENERAL DISCUSSION

As we noted in the introduction, there are major discrepancies in the literature regarding the effect of neighborhood size on word recognition. Although a sizable amount of evidence has accumulated to show that neighbors, particularly high-frequency neighbors, inhibit recognition of a target, other evidence exists showing that neighbors facilitate target recognition, particularly for low-frequency words. Using the same set of words that Andrews (1989) used, we found facilitatory effects of neighborhood size in Experiment 2. In contrast, we found inhibitory effects of neighborhood size in Experiments 3, 4, and 5. These inhibitory effects only occurred for low-frequency words, and were observed for both an accuracy measure in Experiments 3 and 4 and for a speed measure in Experiment 5.

The major question, then, is what differences among the four experiments led to facilitation in one situation and inhibition in the others? Recall that in Experiment 2, we used the ascending method of limits, beginning with Level 1, and forced the subjects to make some response on each presentation. In contrast, in Experiments 3 and 4, the subjects only responded at Level 4. Therefore, we believe that the operative difference among the experiments was the opportunity, in Experiment 2, to eliminate likely competitors of the target as sources of interference. These likely competitors tended to be high-frequency neighbors of the target, particularly for low-frequency, large-neighborhood words. In Experiments 3 and 4, in which the subjects did not guess until Level 4, these high-frequency neighbors were still viable responses to the low-frequency, high-neighborhood targets. Therefore, the subjects in Experiments 3 and 4 guessed high-frequency neighbors, because they had no opportunity to eliminate them in previous responses.

This hypothesis is supported by several of the statistics we have already considered. First, high-neighborhood words typically lead to more guesses than low-neighborhood words. This is a necessary condition for the mechanism to work. Second, these guesses are more likely to be neighbors of the target and, for low-frequency words, are more likely to be of higher frequency than the target.



### Lexical Decision Results

In the preceding section, we were able to account for both facilitatory and inhibitory effects of neighborhood size by differential opportunities for guessing. Is it also possible to account for contradictory effects of neighborhood size in lexical decision? We think so, but for entirely different reasons.

Lexical decision may be a particularly inappropriate task to study neighborhood effects in word identification. Balota and Chumbley (1984) have criticized the lexical decision task as a tool for investigating word frequency effects in lexical access, and their arguments could also apply to its use in studying neighborhood effects. Balota and Chumbley point out that the lexical decision task is really a discrimination task between words and nonwords. They propose a hybrid model for lexical decision, inspired in part by the Atkinson and Juola (1974) model of memory search, in which extreme values along a global familiarity/meaningfulness continuum produce fast, but error-prone, word and nonword decisions, whereas intermediate values of familiarity/meaningfulness produce slower, but accurate, decisions based on lexical search. In this model, subjects set two criteria for word/nonword decisions. If the familiarity/meaningfulness of a target falls above the upper criterion, the subject makes a fast "word" response, and if the familiarity/meaningfulness of a target falls below the lower criterion, the subject makes a fast "nonword" response. If the target value falls between the two criterion, the subject carries out a more time-consuming analysis, such as a lexical search. Balota and Chumbley argue that, because the slower search process is more likely for low-frequency than for high-frequency words, the difference between low- and high-frequency words in lexical decision latencies may be exaggerated by differences in the decision process, rather than by differences in speed of lexical access.

Balota and Chumbley's analysis depends primarily on the assumption that lexical search takes longer than criterion-based decisions. Thus, their model can predict that the advantage of high- over low-frequency words will vary, but never that low-frequency words will be faster than high-frequency words. We have the somewhat more difficult job of showing that a model like this could predict a reversal of reaction times (RTs) between small-neighborhood and large-neighborhood words. We use a neglected feature of the original Atkinson and Juola (1974) model—that criterion-based decisions are faster for targets whose values on the familiarity/meaningfulness continuum fall farther from the criterion.<sup>3</sup> This feature of the model predicts that criterion-based decisions to large-neighborhood words will be faster than criterion-based decisions to small-neighborhood words, because large-neighborhood words are more familiar/meaningful than small-neighborhood words. So, in a case where subjects make predominantly criterion-based decisions (by moving their two response criteria close together), the model predicts that large-neighborhood words will have faster correct RTs than small-neighborhood words.

However, we assume that when subjects engage in lexical search, large-neighborhood words will take longer to locate than small-neighborhood words, particularly when the neighbor is of higher frequency. This follows either from a frequency ordered serial search model or from a parallel search model, in which high-frequency neighbors compete with the target for limited resources. Therefore, when subjects make predominantly search-based decisions (by moving their two response criteria far apart), large-neighborhood words will have *slower* correct RTs than small-neighborhood words, because words with many neighbors will take longer to locate than words with few neighbors.

According to this hybrid model for neighborhood size effects, the subject's placement of the two criteria dividing the search from nonsearch spaces for words and nonwords can determine whether neighborhood size will facilitate or inhibit lexical decision times. Neighborhood size could have a facilitatory effect if the criterion for word decisions was set low on the familiarity/meaningfulness dimension (so that most decisions were based on fast global judgments), and could have an inhibitory effect if the criterion for word decisions was set high on the familiarity/meaningfulness dimension (so that most decisions were based on lexical search).

Because criterion-based decisions will produce higher error rates than search-based decisions, this analysis predicts that studies in which facilitation from neighborhood size was observed should have higher error rates than studies in which inhibition was observed. Lowering the upper (word) criterion will increase errors on nonwords, and raising the lower (nonword) criterion will increase errors on words, particularly errors on low-frequency, low-neighborhood words. It seems reasonable to assume that subjects wishing to be fast at the expense of accuracy will move both criteria together, simultaneously lowering the word criterion while raising the nonword criterion. Similarly, subjects wishing to be accurate at the expense of speed will likely raise the word criterion and lower the nonword criterion. Thus, error rates are expected to be roughly symmetric.

Andrews's (1989) subjects, who showed facilitation from large neighborhoods, showed a particularly high error rate for low-frequency words (16.3% in Experiment 1 and 9.4% in Experiment 2), whereas the subjects of Grainger et al. (1989), who showed inhibition from large neighborhoods, had a much lower error rate (5.5%) for words with high-frequency neighbors (the comparable class of stimuli). Andrews's subjects also showed a high nonword error rate (8.0% in Experiment 1, unreported in Experiment 2), whereas the nonword error rates for Grainger et al.'s subjects are also unreported.

The change in error rates from Experiments 1 to 2 in Andrews's data is also consistent with the model. In Experiment 1, nonwords that were similar to the target words were used, whereas nonwords that were dissimilar to the target words were used in Experiment 2. If subjects continued to use a primarily criterion-based strategy, this

would produce the same amount of facilitation on the RTs (as found), but a lowered error rate on the words because of less overlap between the word/nonword distributions. This model also accounts for the inhibitory effect of neighborhood size on nonword lexical decisions, found by Coltheart et al. (1977), because nonwords with few word neighbors will be low on the familiarity/meaningfulness continuum and thus can be quickly rejected as words, whereas nonwords with many word neighbors will be higher on the familiarity/meaningfulness continuum and thus will have slower criterion-based RTs and also may require a lexical search, which will produce even longer RTs.

### Pronunciation Results

As discussed in the introduction, Andrews (1989, 1992) showed that large neighborhoods facilitated pronunciation. Grainger (1990) also found a slight facilitatory effect of neighborhood size on pronunciation latencies for words with one or more higher frequency neighbors. Yet, if large neighborhoods slow access to the lexicon, and if pronunciation depends upon lexical access, then large-neighborhood words, particularly those of low frequency, should show slower naming times than small-neighborhood words. As in the lexical decision task, it is possible to account for facilitatory effects of neighborhood size on pronunciation by assuming that subjects do not always access the lexicon.

Grainger (1990) discusses two models that specify *non-lexical* mechanisms unique to the pronunciation task, which could accommodate facilitatory effects of neighborhood size on pronunciation latency. According to analogy models of word naming, pronunciation is thought to be accomplished by synthesizing the phonological contributions of all activated lexical representations (Glushko, 1979; Taraban & McClelland, 1987). The neighbors, which share part of the target's phonology, therefore provide support for the activation of the target's representation. The facilitatory effect is stronger for a low-frequency word than a high-frequency word, because the initial activation level of the low-frequency word is lower, forcing the synthesis to rely more on the representations of the neighbors. Models that incorporate the analogy mechanism with the spelling-to-sound correspondences of dual-route theory are also able to account for facilitatory effects of neighborhood size.

Yet another model that could accommodate facilitatory effects is an extension to the Paap et al. (1982) activation-verification model (Paap, McDonald, Schvaneveldt, & Noel, 1987). This model assumes that there are both lexical and nonlexical pathways to pronunciation, that pronunciation of low-frequency words is often accomplished by spelling-to-sound translation rules (the nonlexical pathway), and that, in consequence, the difference between high- and low-frequency words is often *underestimated* by the pronunciation task. These authors argue, in contrast to Balota and Chumbley (1984), that lexical decision is the better task to investigate frequency effects

because the pronunciation task may underestimate frequency effects. It seems reasonable to assume that high-neighborhood words share more pronunciation patterns with their neighbors than low-frequency words, and thus that their pronunciation is more likely to be accomplished by the nonlexical route. This could produce a reversal of the neighborhood-size effect, thereby underestimating (indeed reversing) the advantage that small-neighborhood words have in the lexical access pathway to pronunciation.

### Conclusions

It is a peculiar comment on the history of research in word recognition that, although the process was initially studied by use of the perceptual identification task, typically using tachistoscopic recognition thresholds as the dependent variable (e.g., Howes & Solomon, 1951), today the most popular tasks for measuring word recognition are lexical decision and pronunciation. Yet, as the preceding discussion makes clear, questions have been raised about both the lexical decision and pronunciation tasks as viable ways of measuring lexical access. We believe that the two tasks used in this paper—perceptual identification and speeded identification—can provide important converging evidence about the roles of target frequency, neighborhood size, and neighbor frequency on the process of word recognition. Although we admit that perceptual identification is not free of response bias effects, the identification technique has the advantage of revealing something about the nature of errors on both the perceptual and response sides. And, as detailed below, we believe that there are important reasons for wanting to study both speed and accuracy as types of dependent measures.

Investigators often assume that speed and accuracy of responses index the same underlying process and are sensitive to the same manipulations. Elsewhere, we have called this assumption *Cattell equivalence* (Snodgrass, 1991). Cattell equivalence proposes that speed of response to a suprathreshold stimulus will show the same functional relationship to some manipulation of an independent variable that accuracy of response to a threshold stimulus does. The speeded identification task, in which the perceptual features of a stimulus are slowly revealed until identification is achieved, would appear to yield a measure of speed that is more comparable to the accuracy measure obtained in classical threshold studies than either the lexical decision or pronunciation task.

As mentioned previously, Grainger and Segui (1990) found that their progressive demasking procedure yielded latency data that was much more sensitive to both frequency and neighborhood effects than lexical decision latencies were. In the present studies, the results from speeded identification obtained in Experiment 5 parallel those from perceptual identification obtained in Experiments 3 and 4.

This paper would argue for a return to the more "direct" method of measuring word recognition of perceptual identification, by using either an accuracy or a

speed measure, or, preferably, both. It is a particularly instructive method when subjects' guesses are recorded, because one can then determine what they thought the word was (if anything) when they did not correctly identify the target. Such erroneous guesses can be a rich source of hypothesis testing for the many models of word recognition extant today in the literature.

## REFERENCES

- ANDREWS, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 802-814.
- ANDREWS, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 234-254.
- ATKINSON, R. C., & JUOLA, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 1, pp. 243-293). San Francisco: W. H. Freeman.
- BALOTA, D. A., & CHUMBLEY, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 340-357.
- CARROLL, J. B., DAVIES, P., & RICHMAN, B. (1971). *The American Heritage word frequency book*. New York: American Heritage.
- 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.
- FEUSTEL, T. C., SHIFFRIN, R. M., & SALASOO, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. *Journal of Experimental Psychology: General*, **112**, 309-346.
- FORSTER, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches in language mechanisms* (pp. 257-287). Amsterdam: North-Holland.
- FORSTER, K. I. (1987). Form priming with masked primes: The best match hypothesis. In M. Coltheart (Ed.), *Attention and performance XII* (pp. 201-219). Hillsdale, NJ: Erlbaum.
- GLUSHKO, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception & Performance*, **5**, 674-691.
- GOLDINGER, S. D., LUCE, P. A., & PISONI, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory & Language*, **38**, 501-518.
- GRAINGER, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory & Language*, **29**, 228-244.
- 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 & 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 & Psychophysics*, **47**, 191-198.
- GUMBEL, E. J. (1958). *Statistics of extremes*. New York: Columbia University Press.
- HOWES, D. H., & SOLOMON, R. L. (1951). Visual duration threshold as a function of word probability. *Journal of Experimental Psychology*, **41**, 401-410.
- JACOBY, L. L., & DALLAS, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, **110**, 306-340.
- KUČERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- LANDAUER, T. K., & STREETER, L. A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning & Verbal Behavior*, **12**, 119-131.
- LAXON, V. J., COLTHEART, V., & KEATING, C. (1988). Children find friendly words friendly too: Words with many orthographic neighbors are easier to read and spell. *British Journal of Educational Psychology*, **58**, 103-119.
- LUCE, P. A. (1986). *Neighborhoods of words in the mental lexicon* (Research on speech perception Tech. Rep. No. 6). Bloomington: Indiana University, Speech Research Laboratory, Psychology Department.
- LUCE, P. A., PISONI, D. B., & GOLDINGER, S. D. (1990). Similarity neighborhoods of spoken words. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 122-147). Cambridge: MIT Press.
- LUO, C. R., & SNODGRASS, J. G. (1993). *What factors affect perceptual interference?* Manuscript submitted for publication.
- MCCLELLAND, J. L., & RUMELHART, D. E. (1981). An interactive activation model of context effects in letter perception: Pt. 1. An account of basic findings. *Psychological Review*, **88**, 375-405.
- PAAP, K. R., McDONALD, J. E., SCHVANEVELDT, R. W., & NOEL, R. W. (1987). Frequency and pronounceability in visually presented naming and lexical decision tasks. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 223-243). Hillsdale, NJ: Erlbaum.
- PAAP, K. R., NEWSOME, S. L., McDONALD, J. E., & SCHVANEVELDT, R. W. (1982). An activation-verification model for letter and word recognition: The word superiority effect. *Psychological Review*, **89**, 573-594.
- RUMELHART, D. E., & MCCLELLAND, J. L. (1982). An interactive activation model of context effects in letter perception: Pt. 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, **89**, 60-94.
- SNODGRASS, J. G. (1991, November). *Perceptual fluency and recognition memory—What's the connection?* Paper presented at the meeting of the Psychonomic Society, San Francisco.
- SNODGRASS, J. G., & HIRSHMAN, E. (1991). Theoretical explorations of the Bruner-Potter (1964) interference effect. *Journal of Memory & Language*, **30**, 273-293.
- SNODGRASS, J. G., & POSTER, M. (1992). Visual-word recognition thresholds for screen-fragmented names of the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments, & Computers*, **24**, 1-15.
- SNODGRASS, J. G., & VANDERWART, M. (1980). A standardized set of 260 pictures: Norms for naming agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory*, **6**, 174-215.
- TARABAN, R., & MCCLELLAND, J. L. (1987). Conspiracy effects in word pronunciation. *Journal of Memory & Language*, **26**, 608-631.
- TULVING, E., SCHACTER, D., & STARK, H. (1982). Priming effects in word-fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **8**, 336-342.

## NOTES

1. The T% measure is computed as  $T\% = 100 \times (9 - T) / 8$ , where T is the threshold. This was the measure used in Table 4 to describe levels of correct performance. Note that because T% is a linear transformation of threshold, analyses based on T% measures would be identical to those based on raw T values. The percentage correct measure in Table 6, in contrast, is computed for a single trial as  $100 \times I / T$  (i.e., the number of correct responses on that trial, which equals I, divided by the number of opportunities, which equals the threshold, T). These two measures are not linearly related to one another. In fact, for items, the correlations between the two measures are very high ( $r = +.95$ ). For evaluating overall correct performance, we believe the T% (or T) measure is the appropriate one. However, when describing the distribution of responses across possible opportunities, as was done in Table 6, the only appropriate measure for percentage correct is the one we used.
2. There are at least two senses in which a fragmentation series can be random. In our intended sense, the pixel blocks for each word were randomly permuted at the beginning of each trial and then the appropriate proportion was selected for display. So, for example, although 52% of

pixel blocks were always shown at Level 4, the particular subset that was selected varied randomly from subject to subject. A second sense is one in which each subset of pixel blocks at each level is randomly selected from the entire population of pixel blocks. This was never done in any of the experiments reported here. Instead, in the ascending method of presentation, pixel blocks were always added cumulatively, so that the blocks shown at Level 1 were included in those shown at Level 2, and so on.

3. Although Atkinson and Juola (1974) included a parameter in their original model for the rate at which RT decreased with distance from the criterion, in a simulation of their memory data they found that the model fit as well when the parameter was omitted, so that all responses that exceeded the criteria were assumed to have the same (fast) RT.

(Manuscript received September 30, 1991;  
revision accepted for publication August 14, 1992.)

**23rd Annual Meeting of the Society for Computers in Psychology  
Washington, D.C.  
November 4, 1993**

**Call for Papers**

The 23rd Annual Meeting of the Society for Computers in Psychology will be held at the Shoreham Hotel in Washington, DC, on November 4, 1993, the day before the annual meeting of the Psychonomic Society. The meeting will include presentations, workshops, tutorials, and demonstrations. The application of computer-based solutions to all areas of psychology will be featured, including research, education, clinical practice, and industrial applications. The proceedings will be published in *Behavior Research Methods, Instruments, & Computers*.

The deadline for submissions is June 25, 1993.

For further information, contact Nancy Duncan, Department of Psychology, Hampton University, Hampton, VA 23668 <71043,1340@compuserv.com> or Ellen Rosen, Department of Psychology, College of William and Mary, Williamsburg, VA 23187 <Internet:EFROSE.MAIL.WM.EDU>.