



Published in final edited form as:

J Exp Psychol Gen. 1993 September ; 122(3): 316–330.

Lexical Familiarity and Processing Efficiency: Individual Differences in Naming, Lexical Decision, and Semantic Categorization

Mary Jo Lewellen,

Psychology Department, Indiana University Bloomington

Stephen D. Goldinger,

Psychology Department, Arizona State University

David B. Pisoni, and

Psychology Department, Indiana University Bloomington

Beth G. Greene

Center for Reading and Language Studies, Indiana University Bloomington

Abstract

College students were separated into 2 groups (high and low) on the basis of 3 measures: subjective familiarity ratings of words, self-reported language experiences, and a test of vocabulary knowledge. Three experiments were conducted to determine if the groups also differed in visual word naming, lexical decision, and semantic categorization. High Ss were consistently faster than low Ss in naming visually presented words. They were also faster and more accurate in making difficult lexical decisions and in rejecting homophone foils in semantic categorization. Taken together, the results demonstrate that Ss who differ in lexical familiarity also differ in processing efficiency. The relationship between processing efficiency and working memory accounts of individual differences in language processing is also discussed.

The study of individual differences as a content area of psychology has been dominated by intelligence testing. Although intelligence tests help discriminate among individuals for pragmatic purposes, such as predicting academic success, they fail to specify how individuals differ in terms of underlying cognitive processes. Hunt and his colleagues have led the crusade for integrating theories of testing with theories of cognition (Hunt, 1978, 1980, 1983; Hunt, Frost, & Lunneborg, 1973; Hunt, Lunneborg, & Lewis, 1975). The focus in cognitive psychology on the individual as an information processor provides a framework to address individual differences in the perception, encoding, manipulation, storage, and retrieval of information. Our specific interest focused on the following question: Do individuals who differ in familiarity with printed words also differ in word recognition and lexical access?

Related Research on Individual Differences

Hunt and his colleagues (Hunt, 1978, 1980, 1983; Hunt et al., 1973, 1975) have studied differences between high verbal and low verbal individuals. In Hunt's research, individuals were assigned to high and low verbal ability groups by using standardized measures, such as

the Scholastic Aptitude Test (SAT). They were then given several cognitive tasks such as letter matching, lexical decision, semantic categorization, and memory scanning. From the groups' relative performance on these tasks, Hunt and his colleagues concluded that high verbals are faster than low verbals in decoding written symbols (Hunt, Davidson, & Lansman, 1981; Hunt et al., 1973, 1975). As a consequence, high verbals display an advantage in recognizing visual patterns as words or letters.

Whereas Hunt and his colleagues have focused on the cognitive processes that differentiate high and low verbals, other researchers have focused on individual differences in reading ability (Frederiksen, 1981, 1982; Frederiksen, Warren, & Rosebery, 1985a, 1985b; Jackson & McClelland, 1979; Palmer, MacLeod, Hunt, & Davidson, 1985; Perfetti, 1983, 1985; Stanovich, 1980; Stanovich & West, 1989). Like Hunt, these investigators emphasized the importance of decoding skills. For example, Jackson and McClelland (1979) found that high-ability readers were able to retrieve letter and word codes from memory faster than average readers. In a similar vein, Perfetti (1983, 1985) found that high-ability readers were superior to low-ability readers in word decoding skills, and he suggested that their decoding superiority may be due to superior knowledge of word forms and orthographic rules.

Frederiksen and his colleagues (Frederiksen, 1979, 1981, 1982; Frederiksen et al., 1985a, 1985b; see also LaBerge & Samuels, 1974) argued that skilled and unskilled readers differ in several component subprocesses of reading. They found that with training, unskilled readers developed automaticity in their perception of multiletter units within words and improved at encoding orthographic patterns into phonological patterns. Stanovich (1980) has also suggested that good and poor readers differ in word recognition skills. Stanovich and West (1989) argued that orthographic and phonological coding are distinct processing skills, each accounting for individual differences in word recognition. They established a correlational link between individual differences in exposure to print and orthographic processing skills (see also Stanovich & Cunningham, 1992).

In sum, previous research suggests that individual differences in reading reflect differences in the subskills of word recognition. Several findings suggest that experience with words is an important factor in developing efficient word recognition (e.g., Frederiksen, 1979, 1981, 1982; Frederiksen et al., 1985a, 1985b; LaBerge & Samuels, 1974). Thus, individual differences in word recognition may be due, in part, to differential experience with written language.

Visual Word Recognition

Word recognition, as one component process in reading, has received a great deal of empirical and theoretical attention in recent years. Following many current models of word recognition, we assume that a presented word activates a number of orthographically and phonetically similar representations in memory. Word recognition entails the selection of one word from this array of similar competitors (Luce, 1986; Luce, Pisoni, & Goldinger, 1990; Marslen-Wilson, 1984; McClelland & Rumelhart, 1981). Because word recognition involves discrimination among a set of activated candidates, any factor that inhibits or facilitates lexical discrimination is of interest. For example, the size of a subject's lexicon—that is, the number of words an individual knows—should affect lexical discrimination.

Neighborhood density refers to the number of words that are similar to any given word. An estimate of neighborhood density is provided by Coltheart's N-metric (Coltheart, Davelaar, Jonasson, & Besner, 1977), which specifies the number of words that can be derived from a given word by changing one letter. Words with many neighbors belong to dense neighborhoods; words with few neighbors belong to sparse neighborhoods.

The published data on neighborhood effects have shown mixed results. Coltheart et al. (1977) found that neighborhood density affected nonword classification times in lexical decision but not word classification times. However, in visual perceptual identification, Luce (1985) found that words from dense neighborhoods were identified more accurately than words from sparse neighborhoods. In visual lexical decision and naming, Andrews (1989) found a facilitatory effect of high neighborhood density for low frequency words. Andrews (1989) points out that neighborhood effects are not predicted equivalently across theories of word recognition: Search-based theories (e.g., Becker, 1976, 1980; Forster, 1976, 1987; Paap, Newsome, McDonald, & Schvaneveldt, 1982) predict that larger neighborhoods will inhibit word recognition, whereas activation-based theories (e.g., McClelland & Rumelhart, 1981) predict that larger neighborhoods will either inhibit or facilitate word recognition, depending on selected parameter values.

Lexical Familiarity and Individual Differences

The basic approach in previous studies on individual differences has been to differentiate groups by using some accepted criterion, such as an achievement test, and then to administer several cognitive-processing tasks to see if the groups remain distinct. Although this procedure provides valuable insights about individual differences in cognitive processes, the method of separating groups reveals little about the underlying dimensions of individual differences. In the present investigation, we used three quantitative measures of lexical familiarity to select subjects: word familiarity ratings, vocabulary test scores, and a questionnaire designed to measure language experience. Exposure to printed words was assessed by the word familiarity ratings and the questionnaire. Knowledge of word meanings was assessed by the word familiarity ratings and the vocabulary test. For the purposes of this article, *lexical familiarity* refers to those knowledge factors assessed by our selection tasks and converging measures.

In this study, we assessed how subjects who differ in lexical familiarity also differ in word recognition and lexical access. We examined three tasks—naming, lexical decision, and semantic categorization—that require progressively deeper processing of lexical information. Naming requires subjects to match an input to a word in the lexicon; lexical decision requires subjects to match an input to a word in the lexicon and an additional decision stage; semantic categorization requires subjects to match an input to a word in the lexicon, a decision stage, and an additional verification stage (Van Orden, 1987). Comparing high and low subjects across these tasks allowed us to examine the specific dimensions of individual differences in word recognition and lexical access.

We assume that individuals who know more words have larger mental lexicons. Because our high subjects have larger lexicons, we assume they activate more candidates for recognition than do individuals with smaller lexicons. Therefore, neighborhood effects can be used, by analogy, to make predictions regarding group differences. Research on individual differences, coupled with findings on neighborhood density, suggest two hypotheses about individual differences in word recognition. On the one hand, Hunt and his colleagues, Jackson and McClelland, Perfetti, and Stanovich have all argued that high verbals are faster at retrieving lexical information from long-term memory. Also, the interactive-activation model (McClelland & Rumelhart, 1981) and the findings by Luce (1985) and Andrews (1989) imply that a larger lexicon should facilitate lexical processing. On the other hand, Forster's search model, the activation-verification model, and, under some circumstances, McClelland and Rumelhart's interactive-activation model, predict that a larger lexicon should inhibit word recognition. We sought to directly test these predictions.

Selection of Subjects

Initially, 270 subjects were administered five tasks in booklet form. These subjects were introductory psychology students who participated in the study for partial course credit. All were native speakers of English who reported having no history of speech, hearing, or language disorders at the time of testing. Although five measures were included in the screening booklet, subjects were assigned to high and low groups on the basis of their performance on only the first three tasks: the familiarity ratings, the vocabulary test, and the language experience questionnaire.

Seventy-three subjects from the original 270 returned as paid subjects for the experiments reported here. Three subjects were eliminated at random to provide equal cell sizes of 35 high and 35 low subjects. Average scores for each task are presented for the initial group of 270 subjects and for the final group of 70 subjects.

Paper-and-Pencil Familiarity Ratings

Our direct measure of word familiarity was based on earlier research by Nusbaum, Pisoni, and Davis (1984), who collected familiarity ratings for each of the 20,000 words in Webster's Pocket Dictionary. In this study, subjects rated the familiarity of visually presented words on a 7-point scale. A rating of 1 indicated that the subject had never seen the word before; a rating of 7 indicated that the subject was familiar with the word and knew its meaning. Ratings of 2 through 6 indicated intermediate degrees of familiarity. Nusbaum et al. suggested that direct measures of word familiarity provide a better index of exposure to words than do traditional frequency counts (e.g., Kucera & Francis, 1967; Thorndike & Lorge, 1944). Familiarity ratings are highly correlated with behavioral measures of lexical access (Gernsbacher, 1984).

In our paper-and-pencil familiarity task, 450 words were chosen from Webster's Pocket Dictionary by using the familiarity ratings obtained by Nusbaum et al. Three sets of 150 words were randomly selected from the high, medium, and low familiarity categories. These 450 words were randomized into five different orders for presentation in the booklets. Twenty-five words appeared on each page of the booklet. A 7-point rating scale was placed next to each word.

Average ratings were computed for each subject for the low, medium, and high familiarity words and also for the whole set of 450 words. The average ratings were 6.52 for the high familiarity words ($SD = .39$), 3.88 for the medium familiarity words ($SD = 1.13$), and 2.2 for the low familiarity words ($SD = .84$). The mean rating on all the words was 4.21 ($SD = .72$). The median familiarity rating over all 450 words was used to divide the subjects into high and low groups. To be assigned to the high group, a subject's mean familiarity rating had to be 4.12 or greater. To be assigned to the low group, a subject's mean familiarity had to be lower than 4.12. For the final 70 subjects in the study, the ratings were as follows: The mean ratings for the high, medium, and low familiarity words for the high subjects were 6.79 ($SD = .18$), 5.01 ($SD = .74$), and 3.05 ($SD = .76$), respectively. For the low subjects, the mean ratings were 6.18 ($SD = .47$), 2.82 ($SD = .68$), and 1.52 ($SD = .29$).

Vocabulary Test

The vocabulary test was taken from the Nelson-Denny reading test, which is typically used at the college level to evaluate reading skills (Nelson & Denny, 1960). The vocabulary portion is considered quite difficult. Subjects were given 10 min to respond to 100 statements. Each statement was followed by five vocabulary words. Subjects were instructed to select the word that best expressed the meaning of the opening statement.

The total number of correct responses was tallied for each subject. Scores ranged from 15 to 82 with an average of 43.2 ($SD = 10.7$). The median score of 42 was used as another criterion to separate high and low subjects. To be assigned to the high group, a subject's score had to be 42 or above. To be assigned to the low group, a subject's score had to be lower than 42. For the final 70 subjects, the average score for the high subjects was 53.43 ($SD = 9.03$); the average score for the low subjects was 34.09 ($SD = 5.73$).

Language Experience Questionnaire

A language experience questionnaire was adapted from earlier research in our laboratory to assess subjects' exposure to printed words. Because old familiarity ratings were available for subjects who had completed previous versions of the questionnaire, we constructed a new questionnaire with items that correlated highly (.31 to .66) with the old familiarity ratings. These critical questions were mixed with filler questions about language exposure to make the purpose of the questionnaire less transparent. The critical questions concerned exposure to words, as revealed by reading and writing habits and life experiences.¹ Only these items were scored and used to separate groups.

A 5-point scale was used as the response for each question; higher ratings indicated greater language experience. The scores ranged from 45 to 145 with a mean score of 79.24 ($SD = 16.94$). A median score was used to divide subjects into high and low groups. Subjects scoring 77 or greater were assigned to the high group; subjects scoring less than 77 were assigned to the low group. For the final 70 subjects, the average score for the high subjects was 95.7 ($SD = 15.9$); the average score for the low subjects was 65.8 ($SD = 5.9$).

Thus, subjects were assigned to high and low groups on the basis of their mean familiarity ratings, vocabulary test scores, and responses to the language experience questionnaire. Subjects who scored above the medians on all three measures were assigned to the high group; subjects who scored below the medians on all three measures were assigned to the low group. Subjects who met these criteria were contacted by phone and asked to return as paid subjects for additional experiments.² Correlations among the three selection measures were all significant beyond the $p < .001$ level. They are presented in Table 1.

Converging Measures

Two additional tasks were administered to the original 270 subjects in the paper-and-pencil booklets. The Author Recognition Test and the Magazine Recognition Test were adapted from Stanovich and West (1989). These tests were designed to compensate for subjects' tendency to give socially desirable responses on self-report measures. Stanovich and West characterized both tests as proxy measures of reading activity that assess exposure to print. We used these measures to verify that our groups of subjects truly differed in their visual language experience.

Author Recognition Test—This test consisted of a list of 50 popular authors and the names of 50 individuals who are not authors (i.e., foils). The list of authors included individuals who appear on best-seller lists as well as newspaper and magazine columnists (see Stanovich & West, 1989, for the rationale of author selections). Subjects were advised that they should be certain before they checked off a name because some names in the list

¹Critical questions assessed subjects' reading and writing habits. Specifically, they assessed how often subjects read novels, non-fiction, poetry, and plays and how often they wrote letters, poetry, or in a journal. They also assessed how often subjects used a dictionary or thesaurus and whether they did crossword puzzles in the daily newspaper.

²More than 73 subjects met the criteria for inclusion into the high and low groups. Sixty-three subjects were above the median and 68 were below the median on all three measures. However, not all of the subjects who qualified were willing to return for additional experiments.

were not authors. The number of foils checked was subtracted from the number of correct responses for each subject. The average score for high subjects was 29.89 ($SD = 15.66$); the average score for low subjects was 13.25 ($SD = 6.81$). As expected, high subjects correctly identified more authors than did low subjects, $t(68) = 5.76, p < .001$.

Magazine Recognition Test—This test was similar to the Author Recognition Test, but it contained true and false magazine names rather than author names (see Stanovich & West, 1989, for further details). As in the Author Recognition Test, the difference in responses to targets and foils was calculated. High subjects averaged 54.57 ($SD = 16.08$) and low subjects averaged 42.57 ($SD = 13.47$). This difference was also significant, $t(68) = 3.38, p < .01$. The results for the original 270 subjects and the final 70 subjects replicated the findings reported by Stanovich and West (1989).³

After the final 70 subjects were selected, two additional measures were taken to further validate the separation of high and low subjects. These included a spelling test and a comparison of standardized achievement test scores.

Spelling test—Perfetti (1992) suggests that accurate spelling of irregular words can be used to index the quality of stored orthographic representations. To assess this kind of knowledge, the spelling test used by Stanovich and West (1989) was administered to our final 70 subjects. The test consisted of 34 words, each of which contained an ambiguous segment that departs from a simple phoneme–grapheme mapping (e.g., SERGEANT). The average number of words spelled correctly was 19.91 ($SD = 4.90$) for the high subjects and 17.31 ($SD = 5.43$) for the low subjects, $t(68) = 2.1, p < .05$.

SAT scores—We obtained subjects' SAT or American College Test (ACT) scores. ACT scores were converted to equivalent SAT scores (Langston & Watkins, 1980). High subjects scored an average of 88 points higher than did low subjects on the verbal portion of the SAT (M s of 528 and 440, respectively) and an average of 60 points higher than did low subjects on the mathematics portion of the SAT (M s of 571 and 511, respectively). Both differences were reliable, $t(68) = 4.95, p < .001$, and $t(68) = 3.41, p < .01$.

On-line familiarity task—Because all of our subsequent experiments used on-line measures, we wanted to ensure that the familiarity measures obtained in the paper-and-pencil task would generalize to an on-line task. Subjects were presented with a new list of 450 words, selected according to the same criteria used in the paper-and-pencil familiarity task (150 high, 150 medium, and 150 low familiarity words). None of the words used in the paper-and-pencil task were included in the on-line procedure.

Subjects were tested in groups of 6 or fewer in a room equipped with individual booths. A PDP11/34 computer controlled all experimental procedures. Each word was presented visually in uppercase letters on a cathode-ray tube (CRT). A line of asterisks on the screen indicated the onset of each trial. Each word remained on the screen until the subject responded or until 8 s elapsed. After 8 s, the computer began the next trial. Subjects recorded their familiarity judgments by pressing the appropriate buttons on a seven-button response box interfaced to the computer.

The top panel of Figure 1 displays familiarity ratings and the bottom panel displays response latencies for both groups of subjects and all three categories of words. High subjects gave

³Stanovich and West's mean difference score of 23.8 for the Magazine Recognition Test is comparable with our mean difference scores of 23.7 ($n = 270$) and 24.3 ($n = 70$). Their mean difference score of 9.3 for the Author Recognition Test is comparable with our mean difference scores of 9.8 ($n = 270$) and 10.8 ($n = 70$).

consistently higher ratings than did low subjects for all three sets of words. Familiarity ratings were submitted to a repeated measures analysis of variance (ANOVA) with both words and subjects treated as random, nested factors. A significant main effect of group was observed, $F(1, 75) = 71.65$, $MS_e = 111.21$, $p < .0001$. Post hoc analyses revealed that high subjects rated each subset of words higher than did low subjects. We also found a significant Word Category \times Group Interaction, $F(2, 233) = 25.84$, $MS_e = 23.3$, $p < .0001$. The difference between groups was larger for the medium and low familiarity words than for the high familiarity words. Examination of the frequency distributions of the familiarity ratings showed that all three distributions for the high subjects were shifted above those for the low subjects.

Table 2 shows the correlations between the rating measures for high, medium, and low words, respectively. In all cases, the correlations are high and statistically significant. The high correlations demonstrate that our distinction of high and low groups generalized from the paper-and-pencil task to an on-line task that used another set of words.

Response latencies (see bottom panel of Figure 1) were also recorded from the onset of each word until the subject entered a response. Responses were faster for the high subjects than for the low subjects for all three sets of words. All response latencies 2.5 standard deviations above or below the grand mean were deleted and replaced according to the procedure suggested by Winer (1971). The latencies were then submitted to a repeated-measures ANOVA with subjects and words treated as nested, random factors. A significant main effect of group, $F(1, 69) = 13.23$, $MSe = 72.84$, $p < .001$, and a significant interaction of group and word category were observed, $F(2, 187) = 15.14$, $MS_e = 4.98$, $p < .0001$. Although the high subjects responded significantly faster than did the low subjects to the high and medium familiarity words, $F(1, 72) = 31.17$, $MS_e = 19.25$, $p < .0001$, and $F(1, 72) = 13.08$, $MS_e = 30.91$, $p < .001$, respectively, the difference between groups was not significant for the low familiarity words.

Correlations among tasks—Table 1 also contains correlations for the converging measures for the final group of 70 subjects. Most of these correlations were significant beyond the $p < .05$ level. Taken together, the pattern of correlations suggests that our measures all tap common components of language experience. Most important, the differences between our groups were maintained across all measures.

Experiment 1: Visual Naming

Experiment 1 was conducted to determine if high subjects name printed words faster than do low subjects. Like Andrews (1989) and Luce (1985, 1986), we manipulated the neighborhood density of the stimulus words. Also, because numerous studies have demonstrated that subjects name high frequency words faster than low frequency words (e.g., Andrews, 1989; Becker, 1976; Forster & Chambers, 1973; Frederiksen & Kroll, 1976; Glanzer & Ehrenreich, 1979; James, 1975; Luce, 1985; Rubenstein, Lewis, & Rubenstein, 1971), we varied word frequency.

Method

Subjects—Thirty subjects, 15 from each group, participated in the naming task.⁴ Each subject received \$5.

⁴Experiment 1 was conducted late in the spring semester. Only 30 of our original 70 subjects were willing or able to participate.

Stimuli—Two hundred words were selected from a computerized version of Webster's Pocket Dictionary. Words were three, four, five, six, or seven letters long. Forty words of each length were selected. All words were listed in the Kucera and Francis (1967) corpus and had a rated subjective familiarity of 6.0 or above on Nusbaum et al.'s (1984) 7-point scale. Words of each length were also selected according to orthographic neighborhood density, as determined by Coltheart's N-metric, and were rank-ordered according to frequency. Because of an insufficient frequency range for seven-letter words, they were dropped from the final analysis, leaving a total of 160 words. Mean word frequencies and neighborhood densities, averaged over word length, are shown in Table 3.

Procedure—Subjects were tested individually. Stimulus presentation and data collection were controlled by a PDP 11/34 computer. Each word appeared in uppercase letters on a CRT. Subjects were instructed to name the word into a microphone as quickly and accurately as possible; the microphone was connected to a voice key that registered the onset of the naming response. Response latencies were measured from the onset of the stimulus to the onset of pronunciation. The experimenter monitored naming responses for errors. Stimuli were randomized for each subject.

Results

Because errors occurred on fewer than 1 % of all trials, only latency data are reported. The latency data were adjusted as follows: Responses on errorful trials were not included in the analyses. Latencies 2.5 standard deviations above or below the grand mean were deleted and replaced according to the procedure suggested by Winer (1971). Fewer than 2% of all trials were outliers.

Figure 2 shows mean response latencies for the high and low subjects as a function of word frequency and neighborhood density. Mean latencies for each subject and each word were submitted to separate four-way (Group \times Word Length \times Neighborhood \times Frequency) ANOVAs. Because word length was varied only to prevent habituation, word length effects are not reported here.

The average response latency was 506 ms for high subjects and 602 ms for low subjects, $F_1(1, 28) = 16.02$, $MS_e = 68,891.85$, $p < .001$, and $F_2(1, 288) = 267.43$, $MS_e = 2,761.56$, $p < .001$. (Because the MS_e term is the same throughout the item analysis, it is not repeated for each comparison.) Significant main effects were also observed for neighborhood density, $F_1(1, 28) = 14.24$, $MS_e = 1,321.29$, $p < .001$, and $F_2(1, 288) = 4.29$, $p < .05$, and word frequency, $F_1(1, 28) = 17.57$, $MS_e = 1,269$, $p < .001$, and $F_2(1, 288) = 5.37$, $p < .05$. Mean naming latencies were shorter for words from dense neighborhoods than for words from sparse neighborhoods (548 ms and 560 ms, respectively). High frequency words were named faster than low frequency words (M s of 547 ms and 561 ms, respectively). Neither neighborhood density nor word frequency interacted with group.

Discussion

High subjects were consistently faster than low subjects in naming the stimulus words. We also replicated the neighborhood density effects reported by Andrews (1989) and Luce (1985, 1986); words from dense neighborhoods were named faster than words from sparse neighborhoods. The typical effects of word frequency were also replicated; high frequency words were named faster than low frequency words. High and low subjects showed a uniform difference of about 100 ms, regardless of neighborhood density or word frequency.

To the extent that the naming task is a pure measure, Experiment 1 suggests that high subjects are more efficient than low subjects in word recognition. However, the utility of the

naming task for studying lexical variables has been questioned (Andrews, 1982; Balota & Chumbley, 1984, 1985; Coltheart et al., 1977; Paap et al., 1982). Lexical decision is another task that has been used extensively to study word recognition. However, reservations have also been expressed about this task because it involves discrimination as well as word recognition (Balota, 1990; Balota & Chumbley, 1984). Given the reservations about using either naming or lexical decision alone, several researchers have suggested that the best strategy is to obtain converging evidence from both tasks (Andrews, 1989; Balota & Chumbley, 1984, 1985; Chumbley & Balota, 1984; Monsell, Doyle, & Haggard, 1989). Therefore, we examined lexical decision in Experiment 2.

Experiment 2: Visual Lexical Decision

In the lexical decision task (LDT), subjects discriminate between words and nonword letter strings. Balota and Chumbley (1984) developed a two-stage model that focuses on the decision process in the LDT. In their model, words and nonwords vary along a familiarity/meaningfulness (FM) dimension that is used to make word–nonword decisions. Words and nonwords are assumed to occupy separate distributions along the FM continuum with the word distribution occupying a position of higher FM value than the nonword distribution. Words vary in their discriminability from nonwords as represented by the amount of overlap between the two distributions.

Balota and Chumbley (1984) applied standard signal detection theory to model word–nonword decisions. A subject is assumed to adopt two criteria: If the FM value of a letter string falls above a high criterion or below a low criterion, the subject makes a fast word or nonword response, respectively. If the FM value falls between the low and high criteria, the subject must apply a more extensive analysis of the letter string before a decision can be made. Any words or nonwords that require such extensive analysis yield longer response latencies.

Word frequency is an important component of the FM dimension. High frequency words usually fall above the high criterion, but low frequency words rarely exceed the high criterion and thus require more extensive processing. Another variable that should affect FM values is neighborhood density. Words from dense neighborhoods resemble many other words and should therefore have higher FM values than should words from sparse neighborhoods. These predictions are supported by LDT data (e.g., Andrews, 1989).

Different response times to classify words have also been observed depending on whether pronounceable or unpronounceable nonwords were used as foils (Duchek & Neely, 1984; James, 1975). With unpronounceable nonwords, the discrimination between words and nonwords is easier, allowing the high criterion in Balota and Chumbley's model to be relaxed. Conversely, when pronounceable nonwords are used, the subject adopts a more stringent response criterion.

In Experiment 2, we combined word and nonword manipulations across two blocks of trials. One block of trials consisted of words that varied in density and frequency mixed with pronounceable pseudowords (e.g., GLANT, THRAD). The second block consisted of words that varied in density and frequency mixed with unpronounceable nonwords (e.g., LPREE, YXEDF). This method is assumed to encourage subjects to adjust their criteria for word and nonword responses across blocks of trials. Thus, in Experiment 2, we assessed the effects of frequency, neighborhood density, and nonword type. Our primary concern, however, was with how these manipulations would interact with the group variable. Because we had no evidence to the contrary, we assumed that our subjects did not differ in their familiarity with nonwords. However, because the high subjects were more familiar with words, we assumed that their word distributions would be higher on the FM dimension than would low subjects'

word distributions. Therefore, high subjects should be faster than low subjects in discriminating words from nonwords. Moreover, the group difference should interact with the difficulty of the discrimination: the difference between groups should be larger in the pseudoword block than in the nonword block.

Method

Subjects—All 70 subjects participated in Experiment 2. Each subject received \$5.

Stimuli—Two hundred English words. 100 pronounceable pseudowords (e.g., GLANT, THRAD) and 100 unpronounceable nonwords (e.g., LPREE, YXEDF) were selected as stimuli. The words met the same constraints used in the naming task and varied in neighborhood density and word frequency. As in the naming task, seven-letter words were excluded from the final data analysis. Mean frequencies and neighborhood densities, averaged across word lengths, are shown in Table 4.

Procedure—Subjects were tested in groups of 6 or fewer in a room equipped with individual booths. Stimulus presentation and response collection were controlled by the same apparatus as in Experiment 1, although responses were recorded with a two-button response box. The experiment consisted of two blocks of 200 trials. One block contained words mixed with pronounceable nonwords; the other block contained words mixed with unpronounceable nonwords. The order of blocks and word lists within blocks were counterbalanced across subjects.

Results

Within blocks, all response times above or below 2.5 standard deviations of the grand mean were deleted and replaced according to the procedure suggested by Winer (1971). Fewer than 3% of all responses were outliers.

Two four-way ANOVAs (Group \times Block \times Stimulus Type \times Stimulus Length) were performed on the latency data. The top panel of Figure 3 displays response latencies; the bottom panel displays error rates. The left side of each panel displays results of the pseudoword block; the right side displays results of the nonword block. Over all stimulus conditions, the mean response latency for high subjects was 623 ms, and the mean latency for low subjects was 648 ms. This difference was reliable only in the item analysis, $F_1(1, 68) = 2.36$, $MS_e = 74,661.88$, $p > .12$, and $F_2(1, 608) = 24.36$, $MS_e = 2,285.17$, $p < .001$.

Mean responses (688 ms) in the pseudoword block were reliably slower than mean responses (583 ms) in the nonword block, $F_1(1, 68) = 186.53$, $MS_e = 16,669.30$, $p < .001$, and $F_2(1, 608) = 799.33$, $p < .001$. As expected, the difference between words and pseudowords (627 ms and 749 ms, respectively) was greater than the difference between words and nonwords (570 ms and 595 ms, respectively), $F_1(1, 68) = 232.72$, $MS_e = 2,880.84$, $p < .001$, and $F_2(1, 608) = 180.75$, $p < .001$. Both of these results indicate that discriminations were more difficult in the pseudoword block than in the nonword block.

A significant Group \times Block interaction revealed that high and low subjects differed more in the pseudoword block than in the nonword block, $F_1(1, 68) = 5.57$, $MS_e = 16,669.30$, $p < .05$, and $F_2(1, 608) = 14.09$, $p < .001$. High subjects averaged 579 ms in the nonword block and 667 ms in the pseudoword block; low subjects averaged 586 ms in the nonword block and 710 ms in the pseudoword block.

Two four-way ANOVAs (Group \times Block \times Stimulus Type \times Stimulus Length) were performed on the error data. The bottom panel of Figure 3 reveals that, over all stimulus

conditions, the mean error rates for high and low subjects were 5.4% and 4.0%, respectively, $F_1(1, 68) = 4.83$, $MS_e = 108.14$, $p < .05$, and $F_2(1, 608) = 4.43$, $MS_e = 50.63$, $p < .05$. The difference between the mean errors in the pseudoword block (6.75%) and the nonword block (2.7%) was also reliable, $F_1(1, 68) = 120.69$, $MS_e = 38.63$, $p < .001$, and $F_2(1, 608) = 4.43$, $p < .05$. Moreover, the difference between the words (5.1%) and pseudowords (8.4%) in the pseudoword block was larger than the difference between the words (3.0%) and nonwords (2.3%) in the nonword block, $F_1(1, 68) = 30.25$, $MS_e = 38.02$, $p < .001$, and $F_2(1, 608) = 10.85$, $p < .001$. These results are comparable with the latency data and indicate that discrimination was more difficult in the pseudoword block than in the nonword block.

A significant Block \times Group interaction revealed that high subjects made fewer errors than did low subjects (5.5% and 8.0%, respectively) but only in the pseudoword block, $F_1(1, 68) = 10.53$, $MS_e = 38.63$, $p < .01$, and $F_2(1, 608) = 14.09$, $p < .001$.

Lexical Variables

Because frequency and density were manipulated only in the word stimuli, two five-way ANOVAs (Group \times Block \times Length \times Density \times Frequency) were performed on the word trial response latencies. Figure 4 displays the mean responses for high and low subjects as a function of word frequency and neighborhood density. The top panel displays response latencies; the bottom panel displays error rates. High subjects (593 ms) were reliably faster than low subjects (607 ms) but only in the item analysis, $F_2(1, 256) = 9.77$, $MS_e = 1,715.50$, $p < .01$.

The density and frequency effects observed here replicate previous findings. Responses to words from dense neighborhoods (586 ms) were reliably faster than responses to words from sparse neighborhoods (614 ms), $F_1(1, 68) = 117.97$, $MS_e = 3,659.77$, $p < .001$, and $F_2(1, 256) = 36.46$, $p < .001$. Responses to high frequency words (594 ms) were reliably faster than responses to low frequency words (606 ms), $F_1(1, 68) = 13.52$, $MS_e = 6,604.00$, $p < .001$, and $F_2(1, 256) = 6.22$, $p < .05$. The effects of word frequency were greater in the pseudoword block than in the nonword block, $F_1(1, 68) = 14.52$, $MS_e = 3,155.56$, $p < .001$, and $F_2(1, 256) = 4.37$, $p < .05$. Response latencies for high and low frequency words in the pseudoword block were 617 ms and 638 ms, respectively, and in the nonword block 570 ms and 574 ms, respectively. A similar effect was found for neighborhood density, although the interaction was reliable only in the subject analysis, $F_1(1, 68) = 5.94$, $MS_e = 3,435.83$, $p < .05$, and $F_2(1, 256) = 1.39$, $p < .24$. Response latencies for words from dense and sparse neighborhoods in the pseudoword block were 611 ms and 645 ms, respectively, and in the nonword block 561 ms and 583 ms, respectively. As in the naming experiment, however, neither frequency nor density interacted with group.

Two five-way ANOVAs (Group \times Block \times Length \times Density \times Frequency) were performed on the accuracy data for the word trials. The bottom panel of Figure 4 reveals that the mean error rate for high subjects (3.3%) was reliably lower than the mean error rate for low subjects (4.7%), $F_1(1, 68) = 5.33$, $MS_e = 209.24$, $p < .05$, and $F_2(1, 256) = 4.53$, $MS_e = 33.99$, $p < .05$.

Although the frequency effect was not replicated in the error data, the density effect was replicated, $F_1(1, 68) = 20.51$, $MS_e = 107.25$, $p < .001$, and $F_2(1, 256) = 9.42$, $p < .01$. The mean error rates for words from dense and sparse neighborhoods were 3.1% and 5.0%, respectively. As in the latency data, neither density nor frequency interacted with group.

Discussion

Of primary interest was the comparison of high and low subjects. We varied the difficulty of the task across blocks to determine whether high and low subjects would differ in the

decision component of the LDT. Although high subjects were faster and more accurate than low subjects in the more difficult pseudoword block, they did not differ in the nonword block. This may reflect a floor effect. In any case, the difference between high and low subjects was detectable only when the decision stage required finer discriminations between words and pseudowords. This difference can be interpreted within Balota and Chumbley's (1984) model. Because high subjects are more familiar with words than are low subjects, they are assumed to have less overlap of their word and pseudoword distributions. Consequently, the discrimination between words and pseudowords is less difficult for high subjects than for low subjects.

The failure of the lexical variables to interact with group could indicate that the groups do not actually differ in the processes of word recognition. Alternatively, our range of lexical variables or our range between groups may have been insufficient.

Experiment 3: Semantic Categorization

High subjects performed differently in both naming and lexical decision. Semantic categorization is another task that has been used to study visual word recognition (Hogaboam & Pellegrino, 1978; Hunt et al., 1981). In Experiment 3, we used this procedure to compare the effects of phonology and semantic relatedness on lexical access in our two groups of subjects.

A controversial topic in the literature on visual word recognition concerns the representational codes that readers use to access the mental lexicon. The question is whether lexical access is direct by way of visual representations of words or whether phonological representations must be derived. According to Coltheart's (1978) dual-route theory, independent visual and phonological access routes operate in parallel; effects of phonology will arise only when lexical access cannot be quickly achieved by way of the visual route. Baron and McKillop (1975) found differences in the time course of visual access and, hence, differences in reliance on phonological mediation in college students. Doctor and Coltheart (1980) proposed that familiar words have direct associations between their spelling and their entries in the lexicon. This direct association allows experienced readers to bypass phonology in favor of direct visual access.

According to Van Orden's (1987; Van Orden, Johnston, & Hale, 1988) model of lexical access, however, word identification always includes phonological mediation. Phonological encoding activates a set of lexical candidates, and orthographic information is used in a spelling check to select the correct candidate from the activated set. The verification process on the basis of orthography enables a subject to avoid phonological confusions; the success of verification depends on the subject's knowledge of spelling.

Homophone foils and homophone controls are used in semantic categorization to study the effects of phonology on lexical access. A homophone foil both sounds like and is orthographically similar to a category member (e.g., PAIR is a homophone of PEAR, a member of the category FRUIT). Homophone controls are just orthographically similar to a category member (e.g., PIER). Higher false positive error rates and slower response times have been found for subjects to classify homophones of category exemplars than for subjects to classify homophone controls (Banks, Oka, & Shugarman, 1981; Meyer & Gutschera, 1975; Meyer & Ruddy, 1973; Van Orden, 1987). This evidence has been taken as support for the role of phonological mediation in visual word recognition.

Although the models proposed by Coltheart (1978) and Van Orden (1987) differ in their details, both predict differential effects of phonology on the performance of high and low subjects. By Coltheart's model, we would expect high subjects to rely more on visual access

and low subjects to rely more on phonological access. Consequently, low subjects are expected to experience greater interference from homophones than are high subjects. By Van Orden's model, because we have already established that high subjects are more accurate than low subjects in spelling irregular words, we would also expect high subjects to be better than low subjects at a spelling verification stage. Whether phonology is bypassed or verification is facilitated in high subjects, the empirical prediction is the same: High subjects should be affected less by homophone foils than should low subjects.

Semantic foils and random controls are used in semantic categorization (Banks et al., 1981) to study the effects of semantics on lexical access. Semantic foils are noncategory members that are semantically related to a category (e.g., ZOO, BARN, CAGE for the category ANIMALS). The random controls are noncategory members that have no obvious relationship to a category (e.g., PARK, PLANE, MUD for the category THINGS TO EAT AND DRINK). Banks et al. (1981) found that responses to semantic foils were significantly slower than responses to the random controls, suggesting interference from semantic association. Research by Gernsbacher and Faust (1991; see also Gernsbacher, Varner, & Faust, 1990) addresses possible differences between high and low subjects with respect to semantic interference. They found that less-skilled readers were less able to suppress the inappropriate interpretations of ambiguous words. Following Gernsbacher and Faust, we predicted that high subjects would be more efficient than low subjects in correctly rejecting semantic foils.⁵

Using homophone foils and semantic foils allowed us to manipulate levels of processing in this task. Semantic controls should be easiest to reject because they are not related to the categories in any way; homophone controls should be somewhat more difficult because they are visually similar to a category member; semantic foils should be even more difficult because they are semantically related to the category; homophone foils should be the most difficult because they sound like and are orthographically similar to category members. The primary question, however, was whether we would observe an interaction of group and level of processing.

Method

Subjects—All 70 subjects participated in Experiment 3. Subjects received \$5.

Stimuli—A total of 120 stimulus words were selected to cover three categories: THINGS TO EAT AND DRINK, ANIMALS, and PARTS OF THE BODY. Sixty words were members of target categories, with 20 words per category. Sixty words were noncategory members, with 20 words per category. Within each group of noncategory members, 5 were homophones foils, 5 were homophone controls, 5 were semantic foils, and 5 were semantic controls. All the words were three to five letters in length and all were highly familiar (rated 6.17 or higher). A complete list of stimulus words is provided in the Appendix.

Procedure—Stimulus presentation and data collection were controlled by the same apparatus used in Experiment 2. On each trial, subjects were shown the name of a category followed 1 s later by a stimulus word. Subjects pressed the *yes* or *no* button on a response box to indicate if the stimulus word was a member of the category. Response latencies and errors were recorded.

⁵Although Gernsbacher and Faust (1991) studied subjects who differed in reading comprehension, we can still use their results to make predictions for our subjects. On the basis of their SAT verbal scores, we assumed that our high subjects were better readers and comprehenders than were our low subjects.

Results

The top panel of Figure 5 shows response latencies for high and low subjects on all stimuli; the bottom panel shows error rates. A one-way ANOVA (group) on the mean response latencies for category words revealed that high and low subjects (718 ms and 750 ms, respectively) differed reliably but only in the item analysis, $F_2(1, 118) = 5.65$, $MS_e = 6,852.55$, $p < .05$. Of greater interest were the response latencies on the foils. A 2×4 (Group \times Foil) ANOVA was conducted on the latency data for the foils. Over all foils, high subjects (841 ms) were reliably faster than low subjects (899 ms), $F_1(1, 68) = 4.29$, $MS_e = 54,927.7$, $p < .05$, and $F_2(1, 112) = 26.21$, $MS_e = 3,538.49$, $p < .001$.

The relative difficulty of foil rejection was as anticipated. For all subjects, homophone rejection (936 ms) was more difficult than semantic foil rejection (878 ms), which was more difficult than either the homophone control or semantic control rejection (828 ms and 840 ms, respectively). The difference between the homophone foils and semantic foils was reliable, $F_1(1, 68) = 23.91$, $MS_e = 4,984.47$, $p < .001$, and $F_2(1, 56) = 13.67$, $p < .001$, as was the difference between the semantic foils and the semantic controls, $F_1(1, 68) = 18.87$, $MS_e = 2,714.97$, $p < .001$, and $F_2(1, 56) = 5.92$, $p < .05$. The homophone control and semantic control conditions did not differ reliably.

The greatest difference between high and low subjects occurred on the homophone foils. High subjects (895 ms) were reliably faster than low subjects (977 ms) in correctly rejecting homophones, $F_1(1, 68) = 6.77$, $MS_e = 17,451.07$, $p < .05$, and $F_2(1, 28) = 7.45$, $MS_e = 4,773.68$, $p < .05$. High and low subjects differed reliably (800 ms and 856 ms, respectively) in their response latencies for homophone controls, $F_1(1, 68) = 3.94$, $MS_e = 13,885.84$, $p = .051$, and $F_2(1, 28) = 7.27$, $MS_e = 3,392.78$, $p < .05$. High and low subjects also differed reliably (808 ms and 872 ms, respectively) in their response latencies for semantic controls, $F_1(1, 68) = 4.51$, $MS_e = 15,790.81$, $p < .05$, and $F_2(1, 28) = 11.16$, $MS_e = 2,590.01$, $p < .01$. They did not differ reliably in responses to the semantic foils.

The bottom panel of Figure 5 shows error rates for all stimuli. A one-way ANOVA (group) revealed that errors on the category words for high and low subjects did not differ reliably. A 2×4 (Group \times Foil) ANOVA was conducted on the error data for the foils. Over all foils, high subjects made fewer errors than did low subjects (5.5% and 8.1%, respectively), although only the item analysis approached significance, $F_1(1, 68) = 4.35$, $MS_e = 114.49$, $p < .05$, and $F_2(1, 112) = 3.32$, $MS_e = 63.51$, $p = .08$.

The error data demonstrate that homophone foils were more difficult to reject than semantic foils, both of which were more difficult to reject than either type of control foils. For all subjects, the mean error of 18.1% on homophones was reliably greater than the mean error of 5.2% on semantic foils, $F_1(1, 68) = 51.85$, $MS_e = 114.92$, $p < .001$, and $F_2(1, 56) = 20.9$, $p < .001$, and the mean error of 5.2% on semantic foils was reliably greater than the mean error of 1.8% on the semantic controls, $F_1(1,68) = 15.38$, $MS_e = 25.29$, $p < .001$, and $F_2(1,56) = 12.83$, $p < .001$. No differences were found between the homophone controls and the semantic controls.

The greatest difference between the high and low subjects was on the homophone foils (mean error rates of 13.0% and 23.2%, respectively). This difference was reliable in the subject analysis, $t_1(68) = 3.06$, $p < .01$, and approached reliability in the item analysis, $t_2(28) = 1.88$, $p = .07$. High and low subjects did not differ in errors on any of the other types of foils.

Discussion

High subjects were faster than low subjects across all foils, replicating Hunt et al.'s (1981) results. Our finding that all subjects were slowest and least accurate on homophone foils is also consistent with previous research (Banks et al., 1981; Meyer & Gutschera, 1975; Meyer & Ruddy, 1973; Van Orden, 1987; Van Orden et al., 1988) and indicates that both groups used phonological encoding in this task. Moreover, our finding that both groups were slower on the semantic foils than on the semantic controls replicates Banks et al.'s (1981) results.

As shown by the upward slope of the bars for foils in both panels of Figure 5, both types of control foils were easier to reject than were semantic foils, which were easier to reject than homophone foils. In the controls, although high subjects were reliably faster than low subjects, they did not differ in accuracy. In the semantic foil condition, there was a strong trend for high subjects to respond faster than low subjects, but there were no reliable differences in either latency or accuracy. In the homophone foil condition, however, high subjects were reliably faster and more accurate than were low subjects. Performance in rejecting homophones indexes the use of phonological, orthographic, and semantic knowledge. Our lexical familiarity measure included both experience with word forms and their meanings. The difference between groups suggests that high subjects were less dependent on phonological encoding and, hence, were able to use their orthographic representations in working memory more effectively to make a correct rejection. High subjects may be more efficient at spelling verification simply because their orthographic representations in working memory may be less noisy with respect to phonological interference.

General Discussion

The present investigation was conducted to learn more about the relationship of lexical familiarity and word recognition. Subjects who differed in lexical familiarity also differed in their performance on three tasks. In naming, the difference between groups could have been due to faster word recognition, conversion of orthography to phonology, or physical response execution. The lexical decision results suggest that the difference between groups may be localized in the decision stage rather than the word recognition stage of the task (Balota & Chumbley, 1984). The semantic categorization results suggest that high subjects experienced less interference than did low subjects by applying their knowledge of orthography to phonological representations. Taken together, the results suggest that high subjects are more efficient than low subjects across a variety of lexical processing skills. Beyond access to lexical representations, these skills include converting print to sound and sound to print and comparing lexical representations with semantic memory.

We predicted that group differences would interact with word frequency and neighborhood density. We expected high subjects' greater familiarity with written language to enhance the effects of word frequency, because their frequency of exposure to any given word is assumed to be higher. However, we did not find any interaction of group with word frequency.

Neighborhood density effects were also used to predict a difference between our groups. We assumed that high subjects activate more lexical candidates in response to a stimulus. More activated candidates could either increase or decrease the effects of neighborhood density for high subjects relative to low subjects. However, we did not find any interaction of group with neighborhood density. Thus, lacking any evidence to the contrary, we can only suppose that word frequency and neighborhood density effects are not modulated by individual differences in lexical familiarity.

Previous research on individual differences can be divided into two main approaches: On the one hand, Hunt and his colleagues have focused on identifying cognitive processes that relate to measures of intelligence. On the other hand, researchers such as Daneman and Carpenter (1980, 1983), Frederiksen (1979, 1981, 1982), Just and Carpenter (1992), King and Just (1991), Perfetti (1983, 1988), and Stanovich (1980) have focused on identifying the cognitive processes that differentiate good readers from poor readers. We have taken a third approach: Subjects in the present investigation were assigned to different groups on the basis of lexical familiarity. We then assessed the relationship between lexical familiarity and specific lexical processing skills with experimental tasks that were similar to those used by Hunt, Perfetti, Frederiksen, and Stanovich. We now ask the following question: How do individuals who differ in their lexical familiarity differ in word recognition and lexical access?

In the reading research that has focused on individual differences, two general accounts have been proposed. The first, represented by the work of Perfetti, Frederiksen, and Stanovich, assumes that reading ability is determined by the operating efficiency of its component processes. The second, represented by the work of Just and Carpenter (1992; see also Daneman & Carpenter, 1980, 1983; King & Just, 1991) addresses the role of working memory capacity in reading comprehension. Although overall verbal efficiency may vary across experimental tasks due to differing processing and storage demands (e.g., word naming vs. sentence comprehension), all such functions presumably share a limited-capacity working memory system. As such, working memory has been suggested as a possible locus of individual differences in verbal ability.

Perfetti's verbal efficiency theory assumes that reading consists of both automatic and attention-demanding processes. Because the latter processes are constrained by the limited capacity of working memory, developing automaticity in lower subskills leaves more resources available for higher skills, such as comprehension. Perfetti (1985) suggested that processes of lexical access are the most likely to develop automaticity. To the extent that lexical access becomes automatic, propositions in working memory can be efficiently constructed and combined.

Perfetti (1983) found that high-ability readers were better than low-ability readers at tasks that draw on working memory capacity. Hunt et al. (1975) showed that high verbals search short-term memory faster in a Sternberg (1969) scanning task, have better memory for order information, are better at integrating letter patterns to form words, and are better in verification and computational tasks than are low verbals. Thus, high verbals are apparently more efficient in memory operations and therefore show greater verbal fluency than do low verbals.

Perfetti's account of individual differences is also consistent with Just and Carpenter's (1992) conception of working memory as the locus of individual differences in verbal processing. To comprehend written material, an individual must execute encoding processes, store information as it is obtained, and integrate that information into coherent propositional structures. Daneman and Carpenter (1980, 1983) found that poor readers devote more of their total capacity to low-level encoding processes and leave less capacity for storing and integrating information. Indeed, Just and Carpenter (1992) recently presented a simulation model of reading that demonstrates the central role of working memory in text comprehension. The accounts proposed by Perfetti and by Just and Carpenter clearly complement each other: Differences in processing efficiency could lead to differences in working memory capacity, just as differences in working memory capacity could lead to differences in processing efficiency.

The results of the present investigation are also consistent with working memory accounts of individual differences. Individuals who differ in lexical familiarity also differ in naming, lexical decision, and semantic categorization. As Just and Carpenter (1992) have argued, all of these processes clearly involve aspects of working memory. Differences in working memory capacity could account for all of the present data, perhaps more parsimoniously, than could any other accounts. Indeed, a comparison of our high and low subjects that used an auditory serial recall task revealed differences that are consistent with changes in the efficiency of working memory (see Lewellen, Goldinger, Pisoni, & Greene, 1991). High and low subjects listened to lists of spoken words and attempted to recall all the words in their correct serial order. Although our subjects were not separated into groups on the basis of any memory measures, high subjects recalled more words from early list positions than did low subjects. This difference could be due to maintenance rehearsal, transfer of information to long-term memory, or retrieval of information. All of these processes clearly involve working memory and lend support for Just and Carpenter's theory.

The present investigation focused on individual differences in lexical familiarity and processing efficiency. We have tried to relate measures of lexical familiarity to measures of word recognition and lexical access. The overall pattern of results is consistent with recent accounts of individual differences in working memory capacity and the efficiency of lexical processing. High subjects excel at levels of lexical processing beyond early pattern recognition. The challenge for future research is to develop new experimental tasks and stimulus manipulations that can further pinpoint the locus of the differences in storage and processing operations.

Acknowledgments

Preparation of this article was supported in part by National Institutes of Health Research Grant DC-00111-16 and in part by National Institute of Deafness and Communication Disorders Training Grant DC-00012-13 to Indiana University Bloomington.

We thank Paul Luce, David Balota, and Rebecca Treiman for their constructive suggestions. We also thank Harold Lindman for his suggestions on the data analysis and Heidi Banholzer, Denise Beike, Ellen Garber, and Susan Rivera for their assistance in data collection.

References

- Andrews S. Phonological recoding: Is the regularity effect consistent? *Memory & Cognition*. 1982; 10:565-575.
- Andrews S. Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1989; 15:802-814.
- Balota, DA. The role of meaning in word recognition. In: Balota, DA.; Flores d'Arcais, GB.; Rayner, K., editors. *Comprehension processes in reading*. Hillsdale, NJ: Erlbaum; 1990. p. 9-32.
- Balota DA, Chumbley JI. 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 and Performance*. 1984; 10:340-357. [PubMed: 6242411]
- Balota DA, Chumbley JI. The locus of word frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language*. 1985; 24:89-106.
- Banks, WP.; Oka, E.; Shugarman, S. Recoding of printed words to internal speech: Does recoding come before lexical access?. In: Tzeng, OJL.; Singer, H., editors. *Perception of print: Reading research in experimental psychology*. Hillsdale, NJ: Erlbaum; 1981. p. 137-170.
- Baron J, McKillop BJ. Individual differences in speed of phonemic analysis, visual analysis, and reading. *Acta Psychologica*. 1975; 39:91-96.
- Becker CA. Allocation of attention during visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*. 1976; 2:556-566. [PubMed: 1011005]

- Becker CA. Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory & Cognition*. 1980; 8:493–512.
- Chumbley JI, Balota DA. A word's meaning affects the decision in lexical decision. *Memory & Cognition*. 1984; 12:590–606.
- Coltheart, M. Lexical access in simple reading tasks. In: Underwood, G., editor. *Strategies of information processing*. San Diego, CA: Academic Press; 1978. p. 151-216.
- Coltheart, M.; Davelaar, E.; Jonasson, JT.; Besner, D. Access to the internal lexicon. In: Dornic, S., editor. *Attention and Performance VI*. Hillsdale, NJ: Erlbaum; 1977. p. 535-555.
- Daneman M, Carpenter PA. Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*. 1980; 19:450–466.
- Daneman M, Carpenter PA. Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1983; 9:561–584.
- Doctor EA, Coltheart M. Children's use of phonological encoding when reading for meaning. *Memory & Cognition*. 1980; 8:195–209.
- Duchek JM, Neely JH. A dissociative word-frequency \times levels-of-processing interaction in episodic recognition and lexical decision tasks. *Memory & Cognition*. 1984; 2:148–162.
- Forster, KI. Accessing the mental lexicon. In: Wales, RJ.; Walker, E., editors. *New approaches to language mechanisms*. Amsterdam: North-Holland; 1976. p. 257-287.
- Forster, KI. Form priming with masked primes: The best match hypothesis. In: Coltheart, M., editor. *Attention and Performance XII*. Hillsdale, NJ: Erlbaum; 1987. p. 201-219.
- Forster KI, Chambers S. Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*. 1973; 12:627–635.
- Frederiksen, JR. Component skills in reading: Measurement of individual differences through chronometric analysis. In: Snow, RR.; Federico, PA.; Montague, WE., editors. *Aptitude, learning, and instruction: Cognitive process analysis*. Hillsdale, NJ: Erlbaum; 1979. p. 105-138.
- Frederiksen, JR. Sources of process interactions in reading. In: Lesgold, AM.; Perfetti, CA., editors. *Interactive processes in reading*. Hillsdale, NJ: Erlbaum; 1981. p. 361-386.
- Frederiksen, JR. A componential theory of reading skills and their interactions. In: Sternberg, RJ., editor. *Advances in the psychology of human intelligence*. Vol. 1. Hillsdale, NJ: Erlbaum; 1982. p. 125-180.
- Frederiksen JR, Kroll JF. Spelling and sound: Approaches to the internal lexicon. *Journal of Experimental Psychology: Human Perception and Performance*. 1976; 2:361–379.
- Frederiksen JR, Warren BM, Rosebery AS. A componential approach to training reading skills: Part 1. Perceptual units training. *Cognition and Instruction*. 1985a; 2:91–130.
- Frederiksen JR, Warren BM, Rosebery AS. A componential approach to training reading skills: Part 2. Decoding and use of context. *Cognition and Instruction*. 1985b; 2:271–338.
- Gernsbacher M. Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*. 1984; 113:256–281. [PubMed: 6242753]
- Gernsbacher M, Faust ME. The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1991; 17:245–262.
- Gernsbacher M, Varner KR, Faust ME. Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1990; 16:430–445.
- Glanzer M, Ehrenreich SL. Structure and search of the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*. 1979; 18:381–398.
- Hogaboam TW, Pellegrino JW. Hunting for individual differences in cognitive processes: Verbal ability and semantic processing of pictures and words. *Memory & Cognition*. 1978; 6:189–193.
- Hunt E. Mechanics of verbal ability. *Psychological Review*. 1978; 85:109–130.
- Hunt E. Intelligence as an information-processing concept. *British Journal of Psychology*. 1980; 71:449–474. [PubMed: 7437671]
- Hunt E. On the nature of intelligence. *Science*. 1983; 219:141–146. [PubMed: 6849125]

- Hunt E, Davidson J, Lansman M. Individual differences in long-term memory access. *Memory & Cognition*. 1981; 9:599–608.
- Hunt, E.; Frost, N.; Lunneborg, C. Individual differences in cognition: A new approach to intelligence. In: Bower, G., editor. *Advances in learning and motivation*. Vol. 7. San Diego, CA: Academic Press; 1973. p. 87-122.
- Hunt E, Lunneborg C, Lewis J. What does it mean to be high verbal? *Cognitive Psychology*. 1975; 7:194–227.
- Jackson MD, McClelland JL. Processing determinants of reading speed. *Journal of Experimental Psychology: General*. 1979; 108:151–181. [PubMed: 528903]
- James CT. The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance*. 1975; 104:130–136.
- Just MA, Carpenter PA. A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*. 1992; 99:122–149. [PubMed: 1546114]
- King J, Just MA. Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*. 1991; 30:580–602.
- Kucera, H.; Francis, W. *Computational analysis of present-day American English*. Providence, RI: Brown University Press; 1967.
- LaBerge D, Samuels SJ. Toward a theory of automatic information processing in reading. *Cognitive Psychology*. 1974; 6:293–323.
- Langston, IW.; Watkins, TB. *SAT—ACT Equivalents (Research Memorandum 80–5)*. Champaign: University of Illinois, University Office of School and College Relations; 1980.
- Lewellen, MJ.; Goldinger, SD.; Pisoni, DB.; Greene, B. *Word familiarity and lexical fluency: Individual differences in serial recall of spoken words (Research on Speech Perception Progress Rep. No. 17)*. Bloomington: Indiana University, Psychology Department, Speech Research Laboratory; 1991.
- Luce, PA. *Similarity neighborhoods and word frequency effects in visual word identification: Sources of facilitation and inhibition (Research on Speech Perception Progress Rep. No. 11)*. Bloomington: Indiana University, Psychology Department, Speech Research Laboratory; 1985.
- Luce, PA. *Neighborhoods of words in the mental lexicon (Research on Speech Perception Tech. Rep. No. 6)*. Bloomington: Indiana University, Psychology Department, Speech Research Laboratory; 1986.
- Luce, PA.; Pisoni, DB.; Goldinger, SD. Similarity neighborhoods of spoken words. In: Altmann, G., editor. *Cognitive representations of speech*. Cambridge, MA: MIT Press; 1990. p. 122-147.
- Marslen-Wilson, WD. Function and process in spoken word-recognition. In: Bouma, H.; Bouwhuis, D., editors. *Attention and performance X*. Hillsdale, NJ: Erlbaum; 1984. p. 125-150.
- McClelland JL, Rumelhart DE. An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review*. 1981; 88:375–407.
- Meyer, DE.; Gutschera, KD. *Orthographic versus phonemic processing of printed words; Paper presented at the meeting of the Psychonomic Society; St. Louis, MO. 1975 Nov.*
- Meyer, DE.; Ruddy, MG. *Lexical-memory retrieval based on graphemic and phonemic representation of printed words; Paper presented at the meeting of the Psychonomic Society; St. Louis, MO. 1973 Nov.*
- Monsell S, Doyle MC, Haggard PN. Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*. 1989; 118:43–71. [PubMed: 2522506]
- Nelson, MJ.; Denny, EC. *The Nelson-Denny Reading Test*. Boston: Houghton Mifflin; 1960.
- Nusbaum, HC.; Pisoni, DB.; Davis, CK. *Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words (Research on Speech Perception Progress Rep. No. 10)*. Bloomington: Indiana University, Psychology Department, Speech Research Laboratory; 1984.
- Paap KR, Newsome SL, McDonald JE, Schvaneveldt RW. An activation-verification model for letter and word recognition: The word-superiority effect. *Psychological Review*. 1982; 89:573–594. [PubMed: 7178333]
- Palmer J, MacLeod CM, Hunt E, Davidson JE. Information processing correlates of reading. *Journal of Memory and Language*. 1985; 24:59–88.

- Perfetti, CA. Individual differences in verbal processes. In: Dillon, R.; Schmeck, RR., editors. *Individual differences in cognition*. San Diego, CA: Academic Press; 1983. p. 65-104.
- Perfetti, CA. *Reading ability*. New York: Oxford University Press; 1985.
- Perfetti, CA. Verbal efficiency in reading ability. In: Besner, D.; Waller, TG.; MacKinnon, GE., editors. *Reading research: Advances in theory and practice*. Vol. 6. San Diego, CA: Academic Press; 1988. p. 109-143.
- Perfetti, CA. The representation problem in reading acquisition. In: Gough, PB., editor. *Reading acquisition*. Hillsdale, NJ: Erlbaum; 1992. p. 145-174.
- Rubenstein H, Lewis SS, Rubenstein MA. Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning & Verbal Behavior*. 1971; 10:645-657.
- Stanovich KE. Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*. 1980; 16:32-71.
- Stanovich KE, Cunningham AE. Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory & Cognition*. 1992; 20:51-68.
- Stanovich KE, West RF. Exposure to print and orthographic processing. *Reading Research Quarterly*. 1989; 24:402-433.
- Sternberg S. Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*. 1969; 57:421-457. [PubMed: 5360276]
- Thorndike, E.; Lorge, I. *The teacher's word book of 30,000 words*. New York: Columbia University, Teacher's College Press; 1944.
- Van Orden GC. A row is a rose: Spelling, sound, and reading. *Memory & Cognition*. 1987; 15:181-198.
- Van Orden GC, Johnston JC, Hale BL. Word identification in reading proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1988; 3:371-386.
- Winer, BJ. *Statistical principles in experimental design*. New York: McGraw-Hill; 1971.

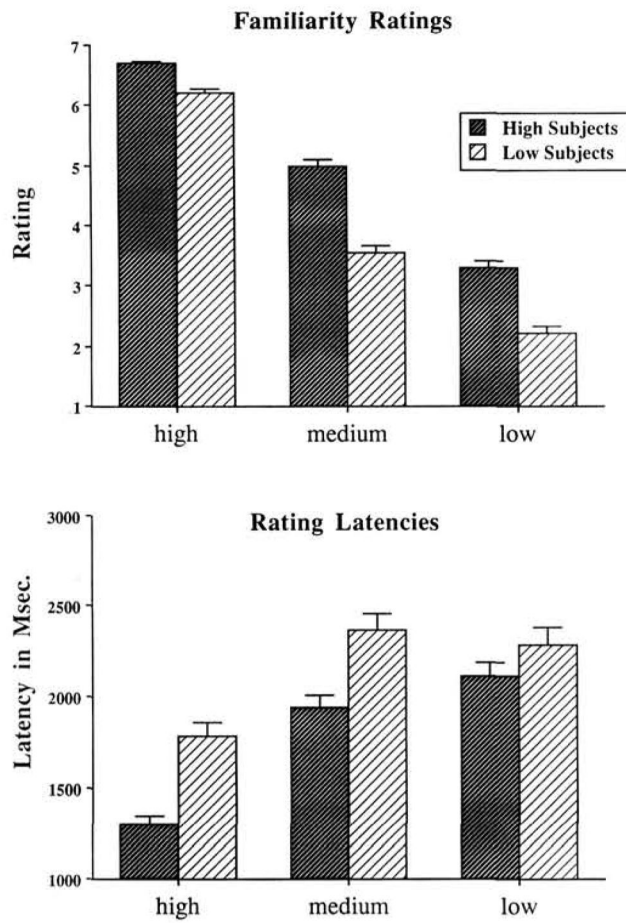


Figure 1. Familiarity ratings and latencies for the on-line familiarity task. (The top panel displays familiarity ratings; the bottom panel displays response latencies.)

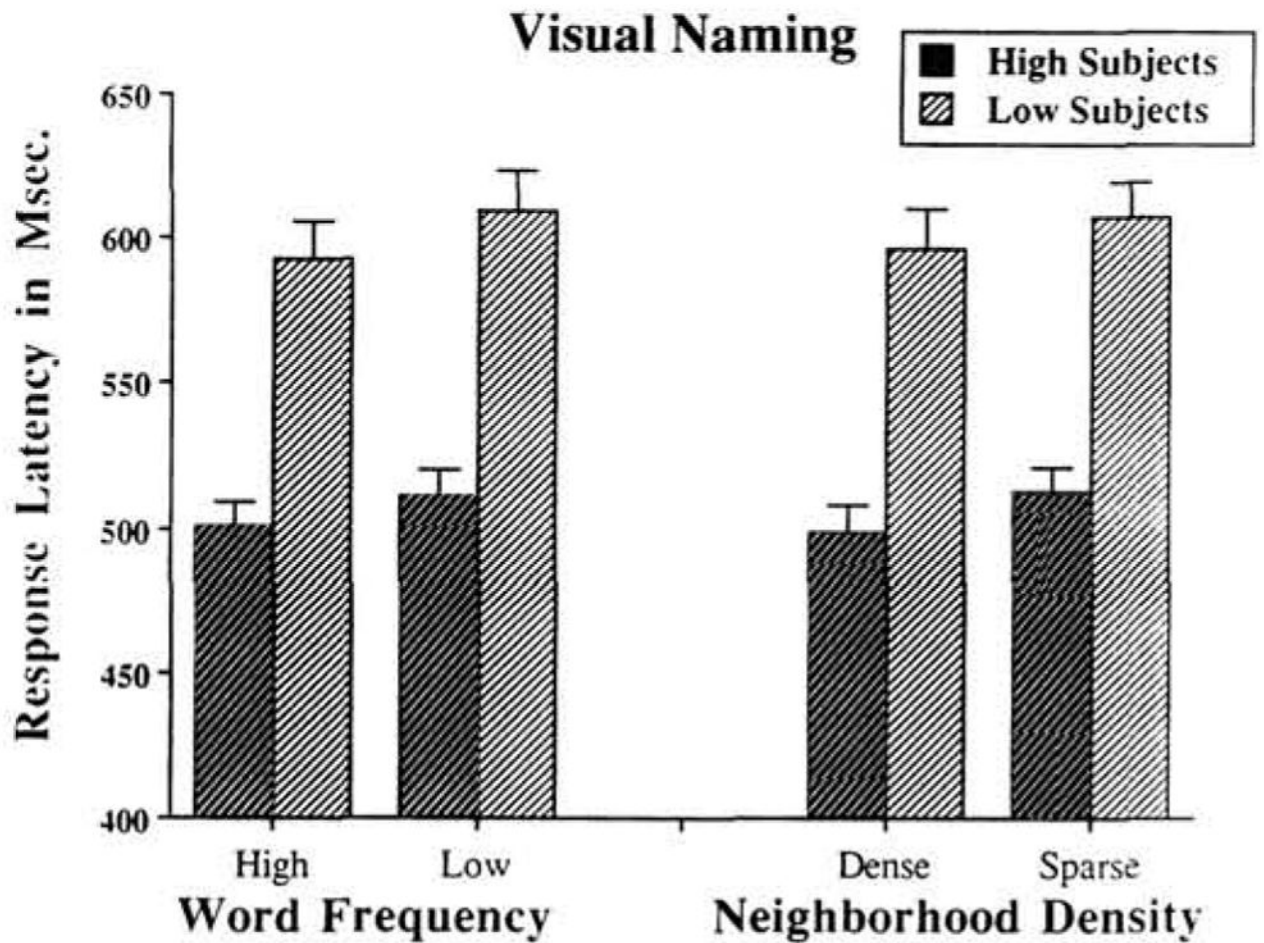


Figure 2.

Visual naming response latencies for high and low subjects averaged across word length.

(The left side of the graph displays high and low subjects' response latencies as a function of word frequency; the right side of the graph displays high and low subjects' response latencies as a function of neighborhood density.)

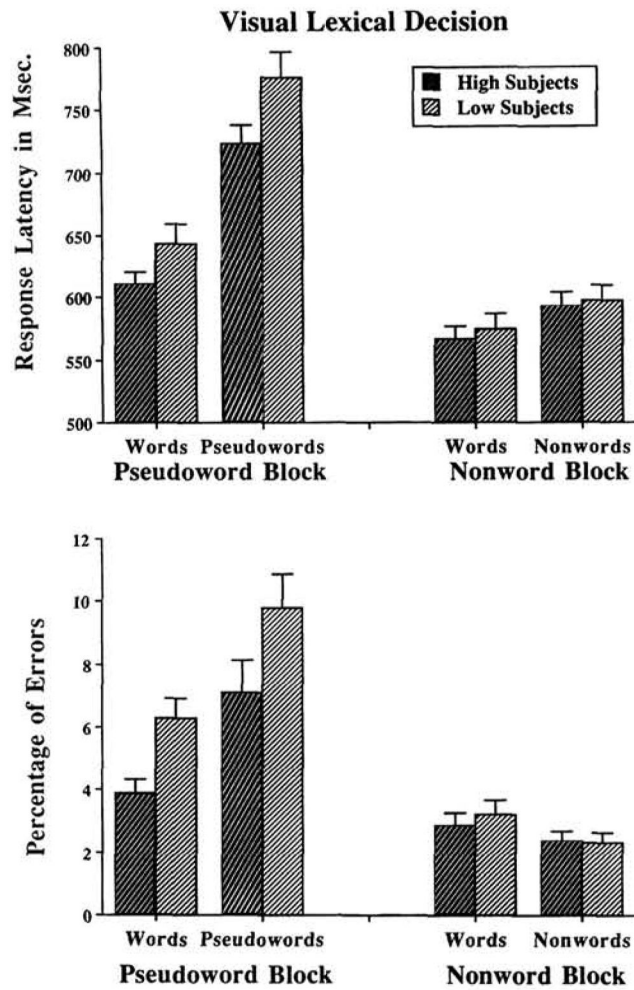


Figure 3. Response latencies and error percentages in lexical decision. (The top panel displays response latencies; the bottom panel displays error percentages. The left side of each graph displays means for the pseudoword block; the right side of each graph displays means for the nonword block.)

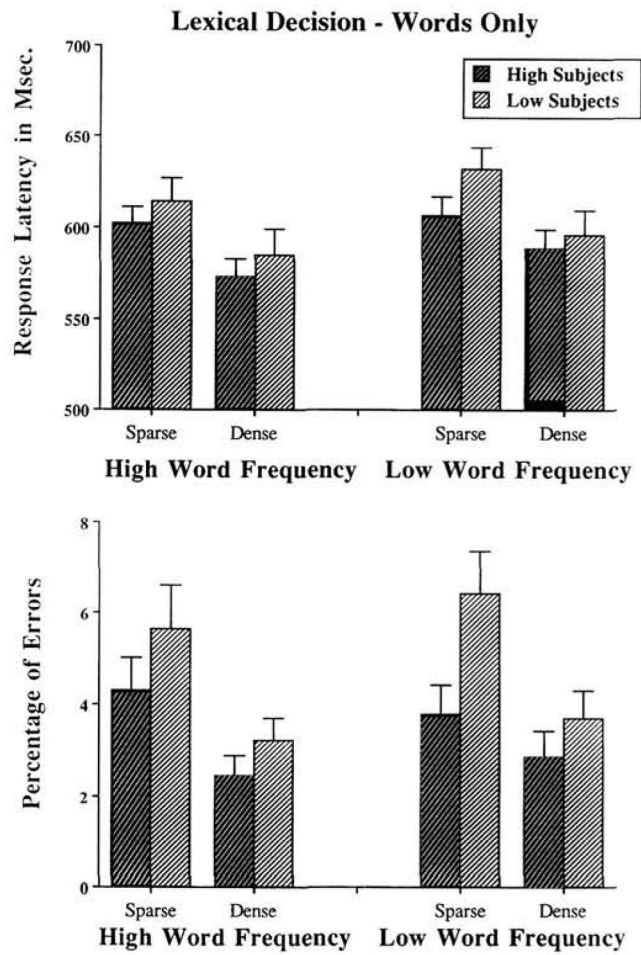


Figure 4. Response latencies and error percentages in lexical decision for the words only. (The top panel displays response latencies; the bottom panel displays error percentages. The left side of each graph displays means for high frequency words; the right side of each graph displays means for low frequency words.)

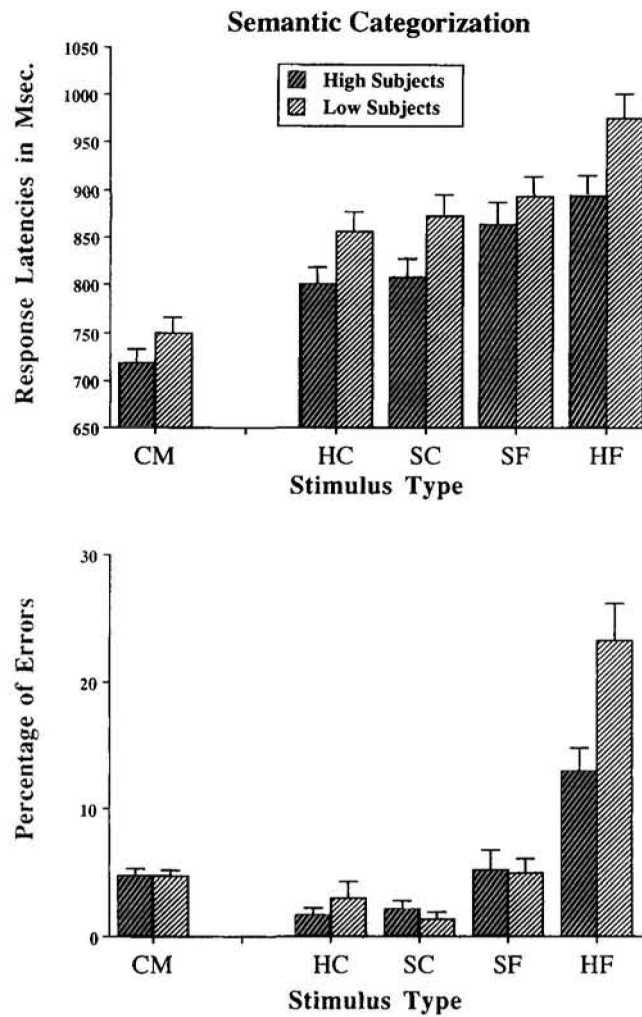


Figure 5. Response latencies and error percentages in semantic categorization for all stimulus conditions. (The top panel displays response latencies; the bottom panel displays error percentages. CM = category members; HC = homophone controls; SC = semantic controls; SF = semantic foils; HF = homophone foils.)

Table 1
Correlations Between Selection and Converging Measures for the Final 70 Subjects

Measure	1	2	3	4	5	6	7	8	9
1. Paper-and-pencil familiarity ratings	—	.72***	.68***	.46***	.25**	.16	.50***	.37***	.83***
2. Nelson-Denny Vocabulary Test		—	.61***	.70***	.53***	.43***	.76***	.52***	.62***
3. Language experience questionnaire			—	.45***	.23*	.18	.38***	.18	.68***
4. Author Recognition Test				—	.62***	.36**	.53***	.25**	.34**
5. Magazine Recognition Test					—	.13	.46***	.12	.19*
6. Spelling test						—	.44***	.32**	.11
7. SAT Verbal							—	.57***	.45***
8. SAT Math								—	.42***
9. On-line familiarity ratings									—

Note. SAT = Scholastic Aptitude Test.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

Table 2

Correlations Between Paper-and-Pencil and On-Line Measures of Word Familiarity for the Final 70 Subjects

Paper-and-pencil measure	On-line measure		
	High	Medium	Low
High	.79	.76	.66
Medium	.77	.85	.77
Low	.64	.72	.66

Note. All correlations are significant beyond the $p < .001$ level.

Table 3

Mean Word Frequencies and Neighborhood Densities for Naming Stimuli

Word frequency	<i>M</i>	Neighborhood density	<i>M</i>
Low	4.50	Sparse	3.50
Low	8.87	Dense	10.67
High	440.50	Sparse	3.52
High	926.37	Dense	10.77

Note. Seven-letter words were excluded.

Table 4

Mean Word Frequencies and Neighborhood Densities for Lexical Decision Task Stimuli

Word frequency	<i>M</i>	Neighborhood density	<i>M</i>
Low	3.47	Sparse	2.97
Low	12.97	Dense	10.60
High	2,017.20	Sparse	3.07
High	361.77	Dense	11.75

Note. Seven-letter words were excluded.

Appendix

Stimuli Used in the Semantic Categorization Task

Member	Homophone	Homophone control	Semantic	Semantic control
Category: Things to Eat and Drink				
peach	pair	pier	stove	stone
honey	meet	moat	fork	park
milk	stake	stack	plate	plane
bean	beat	belt	cup	mud
grape	tee	team	table	sable
veal				
lime				
lemon				
coke				
beet				
candy				
juice				
water				
roll				
bread				
cake				
ham				
bacon				
pork				
pizza				
Category: Animals				
goat	dear	deed	zoo	tee
cat	dough	doubt	farm	harm
dog	hair	harp	leash	reach
tiger	wail	wave	barn	horn
sheep	aunt	act	cage	rage
snake				
hog				
Pig				
cow				
rat				
mouse				
lynx				
fox				
trout				
lion				
bear				
robin				

Member	Homophone	Homophone control	Semantic	Semantic control
gull				
bee				
zebra				
Category: Parts of the Body				
wrist	hare	harm	sock	dock
bone	feat	fees	shoe	shop
thigh	waste	waits	shirt	shore
cheek	tow	toy	hat	mat
brain	heal	heed	ring	burn
leg				
arm				
head				
nose				
eye				
knee				
foot				
ankle				
mouth				
ear				
heart				
chin				
toe				
thumb				
hand				