Journal of Experimental Psychology:
Learning, Memory, and Cognition
2007, Vol. 33, No. 2, 274–296

Copyright 2007 by the American Psychological Association
0278-7393/07/$12.00 DOI: 10.1037/0278-7393.33.2.274

Additive and Interactive Effects on Response Time Distributions in Visual Word Recognition

Melvin J. Yap and David A. Balota
Washington University in St. Louis

Across 3 different word recognition tasks, distributional analyses were used to examine the joint effects of stimulus quality and word frequency on underlying response time distributions. Consistent with the extant literature, stimulus quality and word frequency produced additive effects in lexical decision, not only in the means but also in the shape of the response time distributions, supporting an early normalization process that is separate from processes influenced by word frequency. In contrast, speeded pronunciation and semantic classification produced interactive influences of word frequency and stimulus quality, which is a fundamental prediction from interactive activation models of lexical processing. These findings suggest that stimulus normalization is specific to lexical decision and is driven by the task's emphasis on familiarity-based information.

Keywords: distributional analysis, task-specific effects, stimulus quality, word frequency, visual word recognition

Beginning with Donders (1868/1969), a central goal of understanding human cognition has been to isolate constituent subprocesses through the use of mental chronometry. As there were problems with the insertion procedure Donders advocated, Sternberg (1969a) developed additive-factors logic in which one can provide leverage on the manner in which stages of information processing are organized. Specifically, one can use response time (RT) data from factorial experiments to make inferences about the modules associated with a mental process. For example, Sternberg argued that in an experiment in which two variables are manipulated, *additive* effects of both variables (i.e., main effects for both variables and no interaction) suggest that the variables influence separately modifiable processing stages. In contrast, *interactive* effects are more consistent with the variables influencing at least one stage in common.

In a classic application of additive-factors logic, stimulus quality (intact vs. degraded) and set size (number of items in memory) were manipulated in a memory search task, and these two factors were found to be additive (Sternberg, 1967, 1969b). These additive effects were interpreted as being consistent with a stage model of memory search, in which stimulus quality influences an early encoding stage, and set size influences a subsequent serial comparison stage (see Figure 1). In contrast, factors that interact are assumed to influence a common processing locus. For example, Becker (1979) investigated the effects of word frequency (high vs. low frequency) and semantic context (related vs. unrelated context) on word recognition and reported that word frequency interacts with semantic context. This suggests that word frequency and semantic context influence a common stage.

Following this early classic work, demonstrations of clear additivity have been observed across diverse studies (see Sternberg, 1998, for an extensive review; see also Roberts, 1987; Sanders, 1990), supporting the claim that additive effects reflect discrete stages of processing and that different factors can selectively influence these stages. Given that the notion of discrete stages seems simplistic and architecturally implausible, it is perhaps unsurprising that additive-factor logic has encountered some resistance and skepticism (e.g., Broadbent, 1984, pp. 56–58; Gardner, 1985, pp. 120–124; Luce, 1986, pp. 481–483; Townsend, 1984; Townsend & Ashby, 1983). Although separate stages imply additivity, additivity may not necessarily imply stages.

There are two competing architectures that have been proposed that produce approximate or exact additive effects in mean RT and yet do not assume separate stages: the alternate-pathways model (Roberts & Sternberg, 1993) and the cascade model (Ashby, 1982; McClelland, 1979). In an alternate-pathways model with two pathways, pathway *a* is engaged on a certain proportion of trials, and pathway *b* is engaged on the remaining trials; both pathways are never used during the same trial. In a cascade model with multiple processes, all processes are operating continuously, with the current (partial) output of one process immediately available as an input for the next process. Because the three competing models (i.e., stage, alternate pathways, and cascade) make predictions that are virtually indistinguishable at the level of mean RTs, Roberts

Melvin J. Yap and David A. Balota, Department of Psychology, Washington University in St. Louis.

This research was supported by National Institute of Aging Grant AG03991 and National Science Foundation Grant BCS 0001801 to David A. Balota.

Portions of this research were presented at the 45th (November 2004, Minneapolis, MN) and 46th (November, 2005, Toronto, Ontario, Canada) Annual Meetings of the Psychonomic Society.

We thank Derek Besner, Ken Forster, Dave Plaut, and Sachiko Kinoshita for their insightful reviews, and Jeff Rouder and Roger Ratcliff for assistance with the distributional analyses. Finally, we also extend thanks to Tassy Hayden, Justin Horn, Anuj Shah, and Stephanie To for their help with data collection.

Correspondence concerning this article should be addressed to David A. Balota, Department of Psychology, Washington University, St. Louis, MO, 63130.

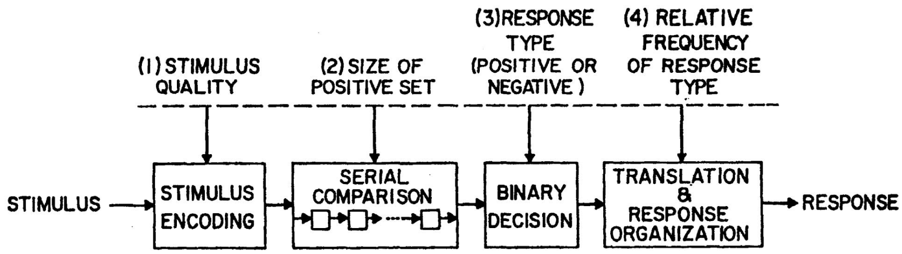

Figure 1. Processing stages in memory search. Reprinted from "The Discovery of Processing Stages: Extensions of Donders' Method," by S. Sternberg, 1969, *Acta Psychologica, 30,* p. 294. Copyright 1969 by North-Holland Publishing. Reproduced with permission from Elsevier.

and Sternberg (1993) argued that the models can be better evaluated at the level of RT variances and RT distributions. Hence, they evaluated the predictions of the models across a wide range of experiments (e.g., detection, identification, and classification). Distributional analyses were carried out with quantile plots of empirical cumulative distribution functions, and additivity was evaluated with the different moments of the RT distribution (mean, variance, skewness, kurtosis). The results of these analyses indicated that the alternate-pathways model failed to successfully simulate any data set. Interestingly, although the cascade model was most successful under parameter settings in which it resembled a stage model, it was rejected after it made incorrect predictions about the relations among means and variances (Roberts & Sternberg, 1993; Sternberg, 1998). Specifically, the distributions produced by an instantiation of the cascade model had twice as much variance as the empirical distributions. In the domain of visual word recognition, the debate between stages and cascading processes has recently been revisited by Borowsky and Besner (2006), who argued that extant empirical findings are accounted more successfully by a stage model of lexical decision performance than a cascaded single-mechanism connectionist model (but see Plaut & Booth, 2006, for a reply).

Additive-Factors Logic and Stage Models of Visual Word Recognition

A central goal in visual word recognition research is to determine the extent to which visual word recognition involves discrete, sequentially organized modules or simply reflects a highly interactive system that has no independent stages. As one might guess, additive-factors logic has contributed to this endeavor in important ways. For example, the effects of stimulus quality and word frequency are robustly additive in lexical decision performance (Balota & Abrams, 1995; Becker & Killion, 1977; Plourde & Besner, 1997; Stanners, Jastrzembski, & Westbrook, 1975), which is consistent with the idea that these two factors affect different stages in visual word recognition. As discussed earlier, stimulus quality may be influencing an earlier stimulus encoding stage while word frequency may be influencing a later lexical retrieval stage. This simple account is qualified by the finding that both stimulus quality and word frequency also interact with context (i.e., whether a target word, e.g., *cat*, is preceded by a semantically related prime, e.g., *dog*, or unrelated prime, e.g., *moon*). Specifically, context effects (faster latencies for targets primed by a related word, compared with an unrelated word) are stronger for degraded words, compared with intact words, as well as for low-frequency words, compared with high-frequency words (see Borowsky & Besner, 1993; Neely, 1991, for a review). If one uses additive-factors logic, this implies that semantic priming has effects on both encoding (hence the interaction with stimulus quality) and retrieval processes (hence the interaction with word frequency; see Figure 2; see Borowsky & Besner, 1993).

Effects of Stimulus Quality and Word Frequency: Additive or Interactive?

Although the model depicted in Figure 2 is plausible, there is a puzzling disjunction between the foregoing findings and current computational models of visual word recognition. Empirically, additive effects of stimulus quality and word frequency are generally obtained in lexical decision, a pattern that is consistent with multistage models of word recognition (Borowsky & Besner, 1993; Forster, 1976; Paap, Newsome, McDonald, & Schvaneveldt, 1982). However, activation-class models, in which different variables influence a common word detector, are more difficult to reconcile with additivity. Specifically, activation-class word recognition models (Morton, 1969), which assume that stimulus quality and word frequency jointly modulate the activation level of the same word representations, predict interactive effects of the two factors. The most straightforward prediction for such models is that one should observe larger stimulus quality effects for word representations with higher activation thresholds or resting activation levels (i.e., low-frequency words). Therefore, additivity seems prima facie incompatible with influential computational activation-class models such as the interactive activation model (McClelland & Rumelhart, 1981) and, by extension, the compu-

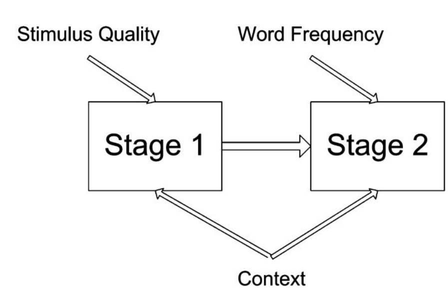

Figure 2. Stimulus quality, word frequency, and context effects in a two-stage model of word recognition.

tational dual route cascaded (DRC) model of word recognition (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), which incorporates the interactive activation model. Indeed, simulations of the computational DRC model confirm our intuitions; stimulus quality and word frequency interacted in pronunciation performance, with larger stimulus quality effects for low-frequency words (Reynolds & Besner, 2004). The connectionist framework of lexical processing (Plaut, McClelland, Seidenberg, & Patterson, 1996), an important alternative to the dual route approach, also appears to predict an interaction between degradation and word frequency. Specifically, D. C. Plaut (personal communication, January 18, 2005), using the attractor network model described in Simulation 3 of Plaut et al.'s (1996) article, demonstrated that word frequency effects in pronunciation were larger for degraded items than for nondegraded items. Therefore, it appears that well-studied computational models of word recognition predict interactive effects of stimulus quality and word frequency, a finding that is markedly inconsistent with the empirical observation of additivity.

Task-Specific Operations in Lexical Decision

One way to resolve the incompatibility between the empirical observations of additive effects of frequency and degradation and the theoretical prediction of an interaction is to posit that additive effects may not faithfully reflect the underlying architecture of the word recognition system but may instead reflect task-specific operations. For example, Balota and Chumbley (1984) compared the size of the frequency effect across lexical decision, pronunciation, and category classification. In the pronunciation task, participants are asked to pronounce visually presented words. In category classification, a category name (e.g., *bird*) is first presented, followed shortly by another word that is either an exemplar (e.g., *robin*) or nonexemplar (e.g., *cow*) of that category; participants decide whether the presented word is an exemplar. Balota and Chumbley found that frequency effects are task modulated, with larger frequency effects in lexical decision than in other tasks (see also Balota & Spieler, 1999). Hence, they argued that in lexical decision, the frequency effect appears to tap both word identification processes and the word–nonword discrimination process that is specific to that task. Just as the frequency effect is exaggerated by the discrimination component of the lexical decision task (LDT), additive effects may also be driven by lexical decision's emphasis on familiarity-based information (Balota & Chumbley, 1984; see also Besner, 1983). Familiarity-based information, in this context, refers to a multidimensional quantity that reflects the orthographic and phonological similarity of a letter string to real words. Specifically, because visual degradation undermines familiarity-based information in the stimuli, degraded stimuli may have to undergo perceptual normalization in an additional early encoding stage. Normalization thus allows familiarity-based information to be recovered and then used to discriminate between familiar words and unfamiliar nonwords. If this line of reasoning is correct, then additive effects should not be observed in lexical processing tasks that do not emphasize familiarity-based information, such as pronunciation or semantic classification.

On the other hand, if the encoding stage implicated by additive effects is indeed a general characteristic of the lexical processing architecture, then it should not be task dependent, and additive effects of stimulus quality and word frequency should be observable across reading tasks. Interestingly, nearly all the studies that have reported additive effects (Becker & Killion, 1977; Plourde & Besner, 1997; Stanners et al., 1975) have used the LDT, and two studies that used the speeded pronunciation task provide at best equivocal support for additive effects. For example, Besner and McCann (1987) found that case alternation (e.g., *dOg*) slowed pronunciation more for low-frequency words than for high-frequency words. Herdman, Chernecki, and Norris (1999) indeed reported additive effects of stimulus intensity and word frequency in pronunciation. Although the interaction between degradation and word frequency was not reliable in Herdman et al.'s data, it is noteworthy that the word frequency effect was 53 ms for degraded words and only 37 ms for nondegraded words. As far as we know, the joint effects of stimulus quality and word frequency have not been investigated in semantic classification.

At this point, it is worth noting the intriguing parallels between lexical decision and memory search. First, both tasks are binary decision tasks that could be viewed as reflecting the presence or absence of a probe stimulus in memory. Second, as described earlier, stimulus quality has additive effects with set size in memory search and with word frequency in lexical decision. In both tasks, these additive effects seem to implicate an early encoding stage in which stimuli are normalized and a later stage in which a comparison or retrieval process is taking place. In memory search, the stimulus encoding stage processes or refines the representation of the degraded stimulus sufficiently so that the serial comparison process works equally efficiently for clear or degraded stimuli (Sternberg, 1969b). In lexical decision, it is plausible that degraded words undergo an analogous normalization procedure prior to lexical retrieval processes, in effect allowing degraded words to be matched to perceptually clear words. Third, Atkinson and Juola (1974) have demonstrated that memory search performance can be modeled by two processes, an initial (fast) familiarity-driven process, and a subsequent (slow) search process. A variation of this model, the two-stage model of lexical decision performance (Balota & Chumbley, 1984; Balota & Spieler, 1999), has also been successful in accounting for various lexical decision phenomena. Finally, Abrams and Balota (1991) found that the word frequency effect in the LDT and the set size effect for target-present trials in the memory scanning task influence not only the onset of the response but also the dynamics of the response after it is initiated. Hence, the similarity between these two binary decision tasks is indeed quite striking, and so it is at least possible that the LDT brings online specific processes that are quite similar to those of the memory scanning task, as opposed to being a reflection of the general lexical processing system.

Objectives of This Study

Given that the additive effects of stimulus quality and word frequency are important for constraining models of word recognition, it is important to establish whether these effects generalize across different measures of reading or whether they are specific to lexical decision. In this article, we have two different, interrelated goals, one theoretical and one methodological. First, we investigate the effects of these two factors across three common word recognition tasks (lexical decision, speeded pronunciation, and semantic classification) using a common stimulus degradation manipulation. To the extent that additive effects are a general lexical phenomenon, additivity should be observed across the tasks.

Second, the data are analyzed both at the level of mean RTs and at the level of distributional characteristics. There is an increasing consensus in the literature that analyzing mean RTs alone can be inadequate and in some cases actually misleading (Andrews & Heathcote, 2001; Balota & Spieler, 1999; Heathcote, Popiel, & Mewhort, 1991; Plourde & Besner, 1997; Yap, Balota, Cortese, & Watson, 2006), because such an analysis does not consider the shape of the RT distribution. In their seminal article, Heathcote et al. (1991) investigated Stroop color-naming performance using both traditional mean and ex-Gaussian analyses. Ex-Gaussian analyses characterize an RT distribution by assuming an explicit function for the shape of the distribution. A convolution of the normal (Gaussian) and exponential distributions, the ex-Gaussian function contains three parameters: μ, the mean of the normal distribution; σ^2, the variance of the normal distribution; and τ, a reflection of the mean and standard deviation of the exponential distribution. Not only does the ex-Gaussian function provide good fits to empirical RT distributions (Luce, 1986), but also the algebraic sum of μ and τ is also approximately equal to the mean of the overall distribution (μ and τ are exactly equal to the mean in the theoretical ex-Gaussian model). This allows differences in means to be partitioned into a component that is associated with distributional shifting (μ) and a component that is associated with distributional skewing (τ; see Table 1 for an example).

Examining mean RTs, Heathcote et al. (1991) observed no difference between the congruent (*RED* displayed in red) and baseline (*XXX* displayed in red) conditions. Although this might suggest that congruency has no effect on color naming, the ex-Gaussian analyses revealed that naming RTs in the congruent condition were facilitated (faster than baseline) in μ but inhibited (slower than baseline) in τ. In this case, congruency shifted the RT distribution leftwards while increasing its skew. These countervailing effects cancelled each other out, spuriously producing null effects in means (see Spieler, Balota, & Faust, 1996, for a replication of this pattern). In the analyses described in this article, more of the information available in an RT distribution is exploited, allowing us to ascertain how a variable modulates the shape, rather than just the mean, of a distribution. In particular, distributional analyses are carried out with ex-Gaussian fitting and a convergent nonparametric technique called *vincentizing* (described in the *Results* section of Experiment 1). As in the case of the Stroop facilitation effect, it is possible that the theoretically important additive effects of word frequency and degradation obtained in means may actually reflect a different pattern once one examines the underlying RT distributions.

Experiment 1

In Experiment 1, we manipulated stimulus quality and word frequency in lexical decision and used distributional analyses to better understand how these two factors influence the underlying RT distributions. On the basis of the literature, we expected additive effects in mean RTs, but we were less certain whether similar patterns of additivity would be present in ex-Gaussian parameters (μ, σ, and τ) and higher order distributional characteristics, such as the second (variance) and third (skewness) moments. The latter point is particularly important when one is using additive-factors logic to make inferences about serially organized, independent stages. As noted, Sternberg (1969b) has pointed out that additivity in means merely supports the existence of successive functional stages. To more rigorously test whether the stages are stochastically independent, additivity must also be demonstrated for the distribution's second and third moments (Andrews & Heathcote, 2001). Interestingly, Plourde and Besner (1997) have in fact manipulated stimulus quality and word frequency in the LDT and also examined the distributional characteristics with an ex-Gaussian analysis. These researchers found evidence of additive effects of these two factors both in means and in the μ and τ parameters. They argued that this provides additional evidence for an early and independent normalization stage.

Plourde and Besner's (1997) study is theoretically informative with respect to models of lexical decision performance as well as more general models of word recognition; however, a few issues remain unresolved. First, although additive effects of stimulus quality and word frequency are typically observed in lexical decision, some researchers (e.g., Norris, 1984; Wilding, 1988) have obtained interactions, with stronger stimulus quality effects for low-frequency words. This empirical inconsistency, coupled with the fact that Plourde and Besner's (1997) study is the only published report exploring these effects with distributional analyses, makes it important to generalize their findings with a different set of stimuli and a different degradation manipulation. Establishing

Table 1
Means of Participants' Lexical Decision Response Time Means, Accuracy, and Ex-Gaussian Parameter Estimates as a Function of Stimulus Quality and Word Frequency (Experiment 1)

Stimulus quality/word frequency	*M*	% of errors	μ	σ	τ
Clear words					
High-frequency words	570	2.4 (2.1)	456	37	114
Low-frequency words	620	8.0 (6.1)	493	47	127
Frequency effect	50	5.6 (4.0)	37	10	13
Mask-degraded words					
High-frequency words	709	7.5 (3.7)	527	40	182
Low-frequency words	765	12.2 (6.6)	561	55	203
Frequency effect	56	4.7 (2.9)	34	15	21
Difference of difference (interaction)	6	−0.9 (−1.1)	−3	5	8

Note. Means are given in milliseconds. Standard deviations are in parentheses.

clear additive effects in Experiment 1 also provides confidence that the degradation manipulation used is not methodologically dissimilar to the manipulations adopted in the other reported studies. Second, in Plourde and Besner's study, stimulus quality was manipulated within participants, which may have encouraged a normalization procedure. Specifically, the random intermixing of clear and degraded stimuli should make stimulus quality across trials unpredictable, and this might engage a strategy in which degraded stimuli were normalized to match perceptually clear stimuli. It is unclear whether this effect generalizes to a between-participants design. In this case, the normalization process may become optional, because degraded and clear stimuli are no longer presented within the same context; that is, participants in the degraded condition are not receiving clear words to which to normalize. (We included both a between-participants, Experiment 1, and a within-participants, Experiment 3, manipulation in the present study.) Finally, as pointed out earlier, we can garner stronger evidence for independent stages if additivity can be demonstrated in the higher moments[1] (i.e., variance and skewness); these analyses are also reported in the present experiment.

Method

Participants

A total of 79 young adults (mean age = 19.1 years, $SD = 1.09$) participated in this study for course credit. All participants had normal or corrected-to-normal vision and were recruited from the undergraduate student population of Washington University in St. Louis. The participants had an average of 13.1 years of education ($SD = 0.91$) and a mean vocabulary age of 18.4 ($SD = 0.85$) on the Shipley (1940) vocabulary subtest. Using the following procedure, we discarded data from 7 of the 79 participants because of excessively high error rates and/or slow latencies: Each participant's response latencies and error rates were quantified as a vector of four scores (mean RTs for high- and low-frequency words; error rates for high- and low-frequency words), and the Mahalanobis D^2 (Lattin, Carroll, & Green, 2003) was then computed for each participant's vector. The Mahalanobis D^2 reflects a multivariate Z score and indicates how discrepant a vector is from the centroid (multidimensional equivalent of the mean). Participants who had D^2 scores with unusually low probability values (i.e., $p < .05$) were discarded. This approach is advantageous in that it identifies multivariate outliers and does not rely on arbitrary criteria defined with respect to a single variable.[2] In total, there were 37 participants in the clear condition and 35 participants in the degraded condition.

Apparatus

We used an IBM-compatible computer running E-prime software (Schneider, Eschman, & Zuccolotto, 2001) to control stimulus presentation and to collect data. The stimuli were displayed on a 17-in. Super VGA monitor, and participants' responses were made on a computer keyboard.

Stimuli

The stimuli for the LDT consisted of 200 words and 200 length-matched pronounceable nonwords. Using the Lund and Burgess (1996) frequency norms, we designated 100 words as high frequency (mean counts per million = 1,227) and 100 words as low frequency (mean counts per million = 44). We constructed nonwords by changing one to three letters from each of the words. For high-frequency words, the mean length was 4.73 letters ($SD = 0.96$), and on the basis of the values available from the English Lexicon Project (http://elexicon.wustl.edu; see Balota et al., in press), the mean orthographic neighborhood size (Coltheart, Davelaar, Jonasson, & Besner, 1977) was 4.77, and the mean summed bigram frequency was 6,369.86. For low-frequency words, the mean length was 4.78 letters ($SD = 0.85$), the mean orthographic neighborhood size was 4.82, and the mean summed bigram frequency was 6,149.13. *Orthographic neighborhood size* refers to the number of words that can be obtained by changing one letter while preserving the identity and positions of the other letters (e.g., neighbors of *CAT* include *MAT*, *COT*, and *CAN*). *Summed bigram frequency* refers to the sum of frequencies for the successive bigrams in a word, where a bigram is defined as a sequence of two letters (e.g., *DO* and *OG* for *DOG*). There was no significant difference between high- and low-frequency words with respect to length, $t(198) = 0.39$, $p = .70$; orthographic neighborhood size, $t(198) = -0.08$, $p = .94$; and summed bigram frequency, $t(198) = 0.42$, $p = .67$. For the nonwords, the mean orthographic neighborhood size was 3.38, and the mean summed bigram frequency was 5,984.70.

Procedure

Before the experimental trials began, participants completed a computer-administered Shipley vocabulary subtest. Participants were tested individually in sound-attenuated cubicles. They were seated about 60 cm from the computer screen. Participants were told that letter strings would be presented at the center of the screen and that their task was to indicate as quickly and as accurately as possible via a button press on the keyboard whether the letter string was a word or nonword. Participants were presented with 20 practice trials, followed by five experimental blocks of 80 trials, with mandatory breaks occurring between blocks. The order in which stimuli were presented was randomized anew for each participant. The presentation sequence was similar for both clear and visually degraded stimuli. For both conditions, stimuli were presented in 14-point Courier font. For the masked degradation condition, letter strings were rapidly alternated with a randomly generated mask of the same length. For example, the mask *&?#* was presented for 10 ms, followed by *DOG* for 25 ms, and the two repeatedly alternated until the participant responded. The mask was generated from random permutations of the symbols *@#$%&?**, with the proviso that the mask be the same length as the string and that symbols not be repeated within a mask. Each trial consisted of the following order of events: (a) a fixation point (+) at the center of the monitor for 400 ms, (b) a blank screen for

[1] Plourde and Besner (1997) examined additivity in variance but not in skewness.

[2] To determine whether participant and RT screening procedures were influencing the results, we reanalyzed the data from the four experiments, using all participants and less conservative screening criteria (i.e., removing only latencies faster than 200 ms and slower than 3,000 ms). The pattern of results did not change across the four experiments.

200 ms, and (c) a stimulus centered at the fixation point's location. The stimulus remained on the screen until a keyboard response was made. Participants pressed the apostrophe key for words and the *A* key for nonwords. Correct responses were followed by a 1,600-ms delay. If a response was incorrect, a 170-ms tone was presented simultaneously with the onset of a 450-ms presentation of the word *Incorrect* (displayed slightly below the fixation point). In order to keep the response to stimulus interval constant across correct and incorrect trials, the incorrect responses were followed by a 1,150-ms delay.

Design

A 2 × 2 factorial design was used, with stimulus quality (clear vs. mask degradation) manipulated between participants and word frequency (high vs. low) manipulated within participants.

Results

Errors (7.2% across both the degraded and clear conditions) and response latencies faster than 200 ms or slower than 3,000 ms were first excluded, and the overall mean and standard deviation of each participant's word and nonword latencies were then computed on the remaining latencies. Of the remaining latencies, any latencies 2.5 standard deviations above or below each participant's respective mean (across all conditions) were removed. These criteria eliminated a further 2.2% of the lexical decision responses. Analyses of variance (ANOVAs) were then carried out on the mean, accuracy, and ex-Gaussian parameters of the RT data. These data are displayed in Table 1. For means and accuracy, ANOVAs by participants and items were conducted.

Stimulus Quality and Word Frequency

Mean response latencies and accuracy. The ANOVA on mean response latencies yielded significant main effects of stimulus quality, $F_p(1, 70) = 30.65$, $p < .001$, $MSE = 23{,}693.58$, $\eta^2 = .31$; $F_i(1, 198) = 1{,}266.26$, $p < .001$, $MSE = 1{,}721.56$, $\eta^2 = .87$, and word frequency, $F_p(1, 70) = 201.77$, $p < .001$, $MSE = 493.94$, $\eta^2 = .74$; $F_i(1, 198) = 63.81$, $p < .001$, $MSE = 5{,}339.21$, $\eta^2 = .24$. The Stimulus Quality × Word Frequency interaction did not approach significance (F_p and $F_i < 1$). Turning to accuracy, the main effects of stimulus quality, $F_p(1, 70) = 22.96$, $p < .001$, $MSE = 0.0034$, $\eta^2 = .25$; $F_i(1, 198) = 44.75$, $p < .001$, $MSE = 0.0033$, $\eta^2 = .18$, and word frequency, $F_p(1, 70) = 61.92$, $p < .001$, $MSE = 0.0015$, $\eta^2 = .47$; $F_i(1, 198) = 26.35$, $p < .001$, $MSE = 0.013$, $\eta^2 = .12$, were again significant. The interaction between stimulus quality and word frequency was not significant by participants ($p = .46$) but did reach significance by items, $F_i(1, 198) = 4.98$, $p = .027$, $MSE = 0.0033$, $\eta^2 = .03$. The simple main effect of frequency (high-frequency words more accurate than low-frequency words) was slightly larger in the clear condition ($d = .94$) than in the degraded condition ($d = .46$). Hence, if anything, this pattern runs counter to the expected greater word frequency effect in the degraded compared with the clear condition. More crucially, the multiple potential sources of lexical decision error (see Balota & Spieler, 1999), and the fact that the interaction was not reliable by participants, makes it difficult to interpret this interaction in a principled manner.

Ex-Gaussian analyses. Ex-Gaussian parameters (μ, σ, τ) were obtained for each participant by use of continuous maximum-likelihood estimation in the R statistics software (R Development Core Team, 2004). Continuous maximum-likelihood estimation provides efficient and unbiased parameter estimates (Van Zandt, 2000) using all the available raw data (see Heathcote, Brown, & Mewhort, 2002, for an alternative approach). Through use of Nelder and Mead's (1965) simplex algorithm, negative log-likelihood functions were minimized in the R statistics package (cf., Speckman & Rouder, 2004), with all fits successfully converging within 500 iterations.

Turning to the ex-Gaussian parameters, for μ, the main effects of stimulus quality, $F(1, 70) = 31.79$, $p < .001$, $MSE = 5{,}576.03$, $\eta^2 = .31$, and word frequency, $F(1, 70) = 144.82$, $p < .001$, $MSE = 313.53$, $\eta^2 = .67$, were significant. The Stimulus Quality × Word Frequency interaction was not reliable ($F < 1$). Turning to σ, the main effects of stimulus quality, $F(1, 70) = 4.24$, $p = .043$, $MSE = 330.50$, $\eta^2 = .057$, and word frequency, $F(1, 70) = 34.32$, $p < .001$, $MSE = 165.43$, $\eta^2 = .33$, were significant. The Stimulus Quality × Word Frequency interaction was not reliable ($p = .27$). Turning to τ, the main effects of stimulus quality, $F(1, 70) = 17.36$, $p < .001$, $MSE = 10{,}623.69$, $\eta^2 = .20$, and word frequency, $F(1, 70) = 16.26$, $p < .001$, $MSE = 625.02$, $\eta^2 = .19$, were significant. The Stimulus Quality × Word Frequency interaction was not reliable ($F < 1$). Hence, the ex-Gaussian analysis is very clear; all parameters produced main effects, but none of the parameters produced interactions.

Vincentile analyses. A converging procedure for investigating the effects of variables on response latencies is to plot the mean vincentiles for the data. Vincentizing is used to average RT distributions across a number of participants (Andrews & Heathcote, 2001; Rouder & Speckman, 2004; Vincent, 1912) to produce the RT distribution for a typical participant. This approach does not depend on prior distributional assumptions and examines the raw data directly. To carry out vincentizing, one first computes a predefined number of vincentiles for each participant, where a vincentile is defined as the mean of observations between neighboring percentiles. For example, to obtain 10 vincentiles, the RT data for a participant are first sorted (from fastest to slowest responses), and the first 10% of the data is then averaged, followed by the second 10%, and so on. Individual vincentiles are then averaged across participants. Plots of mean vincentiles are useful for investigating how different variables influence different regions of the RT distribution and provide a complementary perspective to ex-Gaussian analysis. For example, μ effects are reflected in additive changes in the vincentiles along the *y*-axis, and τ effects are reflected in the slowest (rightmost) vincentiles.

The mean vincentiles for the raw data are plotted in the top two thirds of Figure 3, and in the bottom third, one can see more clearly the word frequency effect for the clear and degraded conditions across the vincentiles. As shown in the bottom third of the figure, the frequency effect increases across vincentiles for both clear and degraded conditions, with only a slight increase in the degraded condition, which ultimately decreases at the last vincentile. As described below, this slight nonsignificant increase in the frequency effect for the degraded condition is dramatically different from conditions in which there is clear evidence of interactive effects.

Higher moments (variance and skewness). As advocated by Roberts and Sternberg (1993), we also computed estimates of variance and skewness[3] for each condition (see Table 2) as a more rigorous test of evidence for stochastically independent stages. The results from the ANOVA on the variance estimate indicated that there was a main effect of stimulus quality, $F(1, 70) = 11.60$, $p = .001$, $MSE = 2{,}296{,}966{,}670$, $\eta^2 = .14$, and the main effect of word frequency approached conventional levels of significance, $F(1, 70) = 3.01$, $p = .087$, $MSE = 90{,}030{,}849$, $\eta^2 = .041$. The Stimulus Quality × Word Frequency interaction again did not approach significance ($F < 1$). Turning to skewness, the main effect of stimulus quality was significant, $F(1, 70) = 8.50$, $p = .005$, $MSE = 1.44 \times 10^{15}$, $\eta^2 = .11$, and the main effect of word frequency approached significance, $F(1, 70) = 3.90$, $p = .052$, $MSE = 1.09 \times 10^{14}$, $\eta^2 = .053$. The interaction was not significant ($F < 1$). Hence, the higher moments analysis (see Table 2) converged with the earlier results, suggesting that stimulus quality and word frequency are indeed additive across means, ex-Gaussian parameters, and the higher order moments. In addition to the above analyses, we also investigated the Stimulus Quality × Lexicality interaction (see mean response latencies in Table 3).

Table 2
Means of Participants' Lexical Decision Response Time Variance and Skewness as a Function of Stimulus Quality and Word Frequency (Experiment 1)

Word frequency/stimulus quality	Variance	Skewness
Clear words		
High-frequency words	18,395	5.46E+06
Low-frequency words	20,870	7.11E+06
Frequency effect	2,475	1.65E+06
Mask-degraded words		
High-frequency words	45,335	2.21E+07
Low-frequency words	48,351	2.73E+07
Frequency effect	3,016	5.23E+06

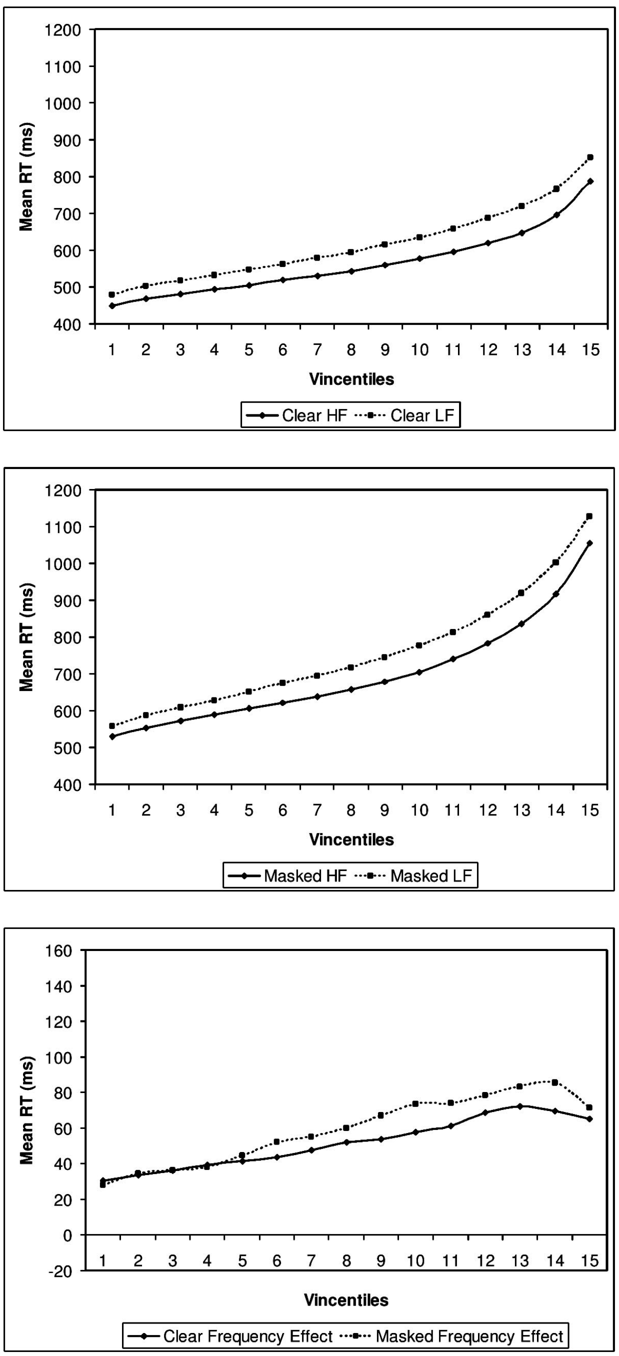

Figure 3. Experiment 1 vincentile means of participants' lexical decision response times as a function of stimulus quality and word frequency. RT = response time; HF = high frequency; LF = low frequency.

Stimulus Quality and Lexicality

Mean response latencies and accuracy. For mean response latencies, the main effects of stimulus quality, $F_p(1, 70) = 36.80$, $p < .001$, $MSE = 33{,}820.29$, $\eta^2 = .35$; $F_i(1, 398) = 3{,}218.05$, $p < .001$, $MSE = 2{,}224.54$, $\eta^2 = .89$, and lexicality, $F_p(1, 70) = 182.53$, $p < .001$, $MSE = 2{,}262.53$, $\eta^2 = .72$; $F_i(1, 398) = 326.95$, $p < .001$, $MSE = 6{,}807.93$, $\eta^2 = .45$, were significant. In contrast to the additive effects of frequency and degradation, the Stimulus Quality × Lexicality interaction was highly reliable, $F_p(1, 70) = 29.93$, $p < .001$, $MSE = 2{,}262.53$, $\eta^2 = .30$; $F_i(1, 398) = 155.18$, $p < .001$, $MSE = 2{,}224.54$, $\eta^2 = .28$; the lexicality effect (word RT < nonword RT) was larger when stimulus quality was degraded. Turning to accuracy, the main effect of stimulus quality was significant, $F_p(1, 70) = 28.15$, $p < .001$, $MSE = 0.0027$, $\eta^2 = .29$; $F_i(1, 398) = 128.05$, $p < .001$, $MSE = 0.0035$, $\eta^2 = .24$. The main effect of lexicality was not significant by participants ($p = .30$) or by items ($F_i < 1$). The Stimulus Quality × Lexicality interaction was not significant by participants ($F_p < 1$) but was significant by items, $F_i(1, 398) = 4.25$, $p = .040$, $MSE = 0.0035$, $\eta^2 = .011$.

Ex-Gaussian analyses. Turning to the ex-Gaussian parameters, for μ, the main effects of stimulus quality, $F(1, 70) = 49.20$, $p < .001$, $MSE = 6{,}900.51$, $\eta^2 = .41$, and lexicality, $F(1, 70) = 273.72$, $p < .001$, $MSE = 930.08$, $\eta^2 = .80$, were significant. The

[3] The third cumulant was estimated by $2\tau^3$ (Andrews & Heathcote, 2001).

Table 3
Means of Participants' Lexical Decision Response Time Means, Accuracy, and Ex-Gaussian Parameter Estimates as a Function of Stimulus Quality and Lexicality (Experiment 1)

Stimulus quality/lexicality	*M*	% of errors	μ	σ	τ
Clear					
Words	594	5.2 (3.7)	468	42	126
Nonwords	657	4.7 (3.7)	526	49	132
Lexicality effect	63	−0.5 (0.0)	58	7	6
Mask degraded					
Words	736	9.9 (4.5)	539	47	198
Nonwords	887	9.2 (5.3)	649	71	237
Lexicality effect	151	−0.7 (0.8)	110	24	39
Difference of difference (interaction)	88	−0.2 (0.8)	52	17	33

Note. Means are given in milliseconds. Standard deviations are in parentheses.

Stimulus Quality × Lexicality interaction was also significant, $F(1, 70) = 27.29$, $p < .001$, $MSE = 930.08$, $\eta^2 = .28$; the lexicality effect was larger when stimulus quality was degraded. Turning to σ, the main effects of stimulus quality, $F(1, 70) = 16.23$, $p < .001$, $MSE = 403.23$, $\eta^2 = .19$, and lexicality, $F(1, 74) = 49.53$, $p < .001$, $MSE = 170.80$, $\eta^2 = .41$, were significant. The Stimulus Quality × Lexicality interaction was again significant, $F(1, 70) = 13.84$, $p < .001$, $MSE = 170.80$, $\eta^2 = .17$; the lexicality effect was larger when stimulus quality was degraded. Turning to τ, the main effects of stimulus quality, $F(1, 70) = 17.23$, $p < .001$, $MSE = 16{,}482.71$, $\eta^2 = .20$, and lexicality, $F(1,70) = 12.41$, $p = .001$, $MSE = 1{,}537.27$, $\eta^2 = .15$, were significant. The Stimulus Quality × Lexicality interaction was also significant, $F(1, 70) = 6.63$, $p = .012$, $MSE = 1{,}537.27$, $\eta^2 = .087$; the lexicality effect was significant only in the degraded condition.

Vincentile analyses. The mean vincentiles for these data are plotted in the top two thirds of Figure 4, and the mean lexicality effects for the clear and degraded conditions as a function of vincentile are displayed in the bottom third. As clearly shown in the bottom third, the lexicality effect is substantially larger for the degraded condition compared with the clear condition, and this effect increases quite dramatically across vincentiles. Comparing the bottom thirds of Figures 3 and 4 clearly shows the difference between additive and interactive effects of variables.

It should also be noted here that because there is already evidence from the analyses of the means of an interaction between stimulus quality and lexicality, there is no need to report the analyses of the higher order moments.

Summary

Stimulus quality and lexicality showed clear interactive effects in means and ex-Gaussian parameters, with larger lexicality effects in the degraded condition across all the parameters (see Table 3). This result is consistent with the literature (see Borowsky & Besner, 1993; Stanners et al., 1975) and furthermore reveals that the lexicality effect in the clear condition was mediated primarily by μ, whereas in the degraded condition, it was mediated by a mixture of μ and τ. "Nonword" lexical decision RTs were shifted and skewed when visually degraded. The vincentile plot also confirms that the lexicality effect for degraded items is larger than for the clear items, and this trend increases across the entire RT distribution.

Discussion

Experiment 1 replicates and extends Plourde and Besner's (1997) results, demonstrating that the additive effects of stimulus quality and word frequency are robust, even when stimulus quality is manipulated between participants. The general pattern of additivity across means, ex-Gaussian parameters, and the higher order moments suggests that when words are processed in lexical decision, stimulus quality and word frequency are indeed influencing separate and independent stages. In particular, the pattern of additivity in the higher order moments is more consistent with a stage model than a cascade model (see Roberts & Sternberg, 1993). The most straightforward interpretation of these results is that during lexical decision, each degraded word undergoes a normalization process before engaging the processes tied to the decision process in this task. This of course is quite similar to the account of the additive effects of degradation and set size in memory scanning. Although the data do not directly address what happens during normalization, they suggest that models of lexical decision performance (e.g., Balota & Spieler, 1999; Coltheart et al., 2001; Grainger & Jacobs, 1996; Plaut, 1997; Ratcliff, Gomez, & McKoon, 2004; Seidenberg & McClelland, 1989) may need to incorporate a perceptual normalization procedure under degraded conditions that precedes the normal lexical decision process.

Interestingly, although such a normalization process might also suggest equivalent effects of degradation for words and nonwords, this does not appear to be the case since there is a reliable significant Stimulus Quality × Lexicality interaction (replicated in Borowsky & Besner, 1993, and Stanners et al., 1975). Because words and nonwords undergo a common normalization stage prior to lexical decision, it is unclear why the stimulus quality effect should be larger for nonwords (see Table 3). One possible account of this pattern is that participants are simply more conservative when they have to reject a degraded letter string as a nonword. Such a conservative nonword response bias may reflect a relatively late decision-level influence. In particular, because all degraded letter strings, whether words or nonwords, initially look unfamiliar, participants may become particularly cautious before making a nonword response. Because half of these degraded letter strings

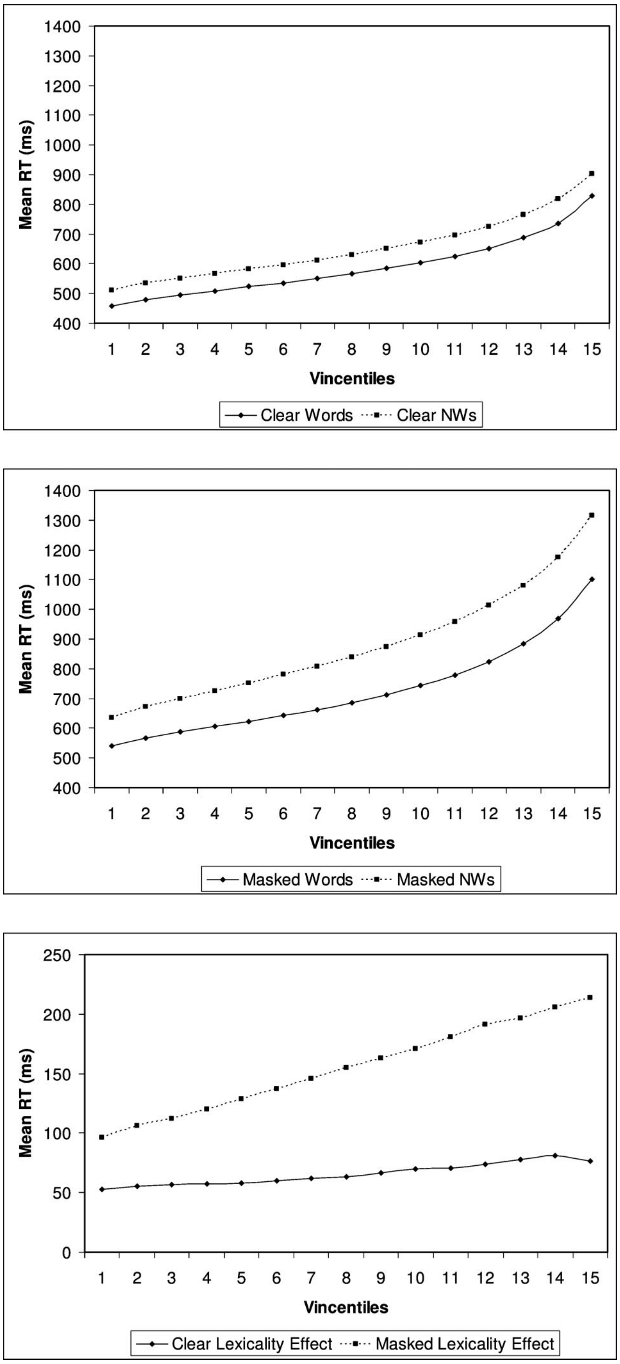

Figure 4. Experiment 1 vincentile means of participants' lexical decision response times as a function of stimulus quality and lexicality. RT = response time; NWs = nonwords.

form words and are eventually resolved because of the normalization process, participants may allocate more time for processing the strange-looking letter strings to ensure that they are not actually words—that is, that they are indeed nonwords.

Mechanistically, it is relatively straightforward to incorporate this conservative bias within both the activation and search metaphors of lexical access. For example, in activation-class models, "nonword" responses are produced when lexical activity does not reach threshold after some temporal deadline (Coltheart et al., 2001). Degraded items might lengthen this nonword deadline, which will of course exaggerate lexicality effects for degraded items. Alternatively, in search models of lexical access (e.g., Murray & Forster, 2004), orthographic bins are searched exhaustively, with "word" responses produced when the target word is located, and "nonword" responses produced when the search is unsuccessful. Presenting degraded nonwords may lead to multiple, perseverative searches of the bin before a search is terminated, leading to larger lexicality effects in the degraded condition (K. I. Forster, personal communication, January 12, 2004). In either case, this perspective suggests that participants become more conservative about responding "nonword" when degraded letter strings are presented. Of course, these accounts still demand empirical validation. However, the robust Stimulus Quality × Lexicality interaction, coupled with the robust additive effects of stimulus quality and word frequency in lexical decision performance, appear problematic for a simple model that attributes the effects of word frequency and lexicality to a common mechanism in this task.

Experiment 2

On the basis of the additive effects from Experiment 1 and Plourde and Besner's (1997) study, there does seem to be compelling evidence for an early normalization stage that is insensitive to word frequency. However, as argued in the introduction, it is unclear whether this normalization stage reflects the cognitive architecture of the word recognition system or is specific to the task requirements of lexical decision. Again, because of the similarity with the additive effects observed in memory scanning (a nonlexical task), one might be concerned that task-specific operations are producing these effects. If normalization is a general property of lexical processing, additive effects of stimulus quality and word frequency should be observable in other word recognition tasks. Hence, the goal of Experiment 2 was to essentially replicate the design of Experiment 1 with the dependent measure being speeded pronunciation performance.

Method

Participants

A total of 88 young adults (mean age = 19.8 years, *SD* = 1.31) participated in this study for course credit. All participants had normal or corrected-to-normal vision and were recruited from the undergraduate student population of Washington University in St. Louis. The participants had an average of 13.3 years of education (*SD* = 1.14) and a mean vocabulary age of 18.6 (*SD* = 0.88) on the Shipley vocabulary subtest. Data from 10 of the 88 participants were discarded because of excessively high error rates or slow latencies, using the same multivariate outlier procedure described in Experiment 1. This left 39 participants in each of the two conditions.

Apparatus and Stimuli

An IBM-compatible computer was used to control stimulus presentation and to collect responses. The stimuli were displayed on a 17-in. Super VGA monitor, and participants' pronunciation responses were detected by an Audio-Technica microphone (Audio-Technica, Stow, OH) connected to a PST serial response box (Psychology Software Tools, Inc., Pittsburgh, PA) with an integrated voice key. The stimuli for the pronunciation task consisted of the 200 words used in Experiment 1.

Procedure

Participants were told that words would be presented in the center of the screen, and their task was to read aloud each word as quickly and as accurately as possible. This was followed by 20 practice trials and four experimental blocks of 50 trials, with mandatory breaks occurring between blocks. The order in which stimuli were presented was randomized anew for each participant. The presentation sequence was similar for both clear and visually degraded stimuli. Each trial consisted of the following order of events: (a) a fixation point (+) at the center of the monitor for 500 ms, (b) a blank screen for 750 ms, and (c) the stimulus at the fixation point's location. The stimulus word remained on the screen until a pronunciation response was detected. Participants then coded their responses by pressing the left mouse button for a correct response and the right mouse button for an incorrect response. Responses were followed by a 2,000-ms delay.

Design

A 2 × 2 factorial design was used: Stimulus quality was manipulated between participants, and word frequency was manipulated within participants. The presentation format for the mask-degradation condition was identical to the manipulation used for Experiment 1.

Results

Errors (1.6% across the two stimulus quality conditions) and extreme response latencies (faster than 200 ms or slower than 3,000 ms) were first excluded from the analyses, and the overall mean and standard deviation of each participant's word latencies were computed on the remaining latencies. Response latencies 2.5 standard deviations above or below each participant's respective mean latency were removed. These criteria eliminated a further 2.4% of the pronunciation responses. ANOVAs were then carried out on the mean, accuracy, and ex-Gaussian parameters of the RT data (see Table 4).

Mean response latencies and accuracy. For means, the main effects of stimulus quality, $F_p(1, 76) = 13.91$, $p < .001$, $MSE = 23{,}622.18$, $\eta^2 = .16$; $F_i(1, 198) = 908.01$, $p < .001$, $MSE = 943.79$, $\eta^2 = .82$, and word frequency, $F_p(1, 76) = 68.53$, $p < .001$, $MSE = 440.37$, $\eta^2 = .47$; $F_i(1, 198) = 23.10$, $p < .001$, $MSE = 3{,}589.42$, $\eta^2 = .10$, were significant. More important, the Stimulus Quality × Word Frequency interaction was significant,[4] $F_p(1, 76) = 9.19$, $p = .003$, $MSE = 440.37$, $\eta^2 = .11$; $F_i(1, 198) = 11.40$, $p = .001$, $MSE = 943.79$, $\eta^2 = .054$, with a larger stimulus quality effect for low-frequency words than for high-frequency words. Turning to accuracy, the main effects of stimulus quality, $F_p(1, 76) = 3.43$, $p = .07$, $MSE = 0.00063$, $\eta^2 = .043$; $F_i(1, 198) = 10.42$, $p = .001$, $MSE = 0.0013$, $\eta^2 = .050$, and word frequency, $F_p(1, 76) = 7.63$, $p = .007$, $MSE = 0.00018$, $\eta^2 = .091$; $F_i(1, 198) = 10.18$, $p = .002$, $MSE = 0.0028$, $\eta^2 = .049$, were significant. The interaction between stimulus quality and word frequency was not reliable by participants or items (Fs < 1).

Ex-Gaussian analyses. Turning to the ex-Gaussian parameters, for μ, only the main effect of word frequency was significant, $F(1, 76) = 13.51$, $p < .001$, $MSE = 392.16$, $\eta^2 = .15$. Neither the main effect of stimulus quality ($p = .13$) nor the interaction ($F < 1$) were significant. Likewise, for σ, only the main effect of word frequency was significant, $F(1, 76) = 5.62$, $p = .020$, $MSE = 180.30$, $\eta^2 = .069$. The main effect of stimulus quality and the interaction were not significant. For τ, the main effects of stimulus quality, $F(1, 76) = 18.01$, $p < .001$, $MSE = 9{,}248.72$, $\eta^2 = .19$, and word frequency, $F(1, 76) = 12.20$, $p = .001$, $MSE = 835.24$, $\eta^2 = .14$, were significant. The Stimulus Quality × Word Frequency interaction approached significance ($p = .07$), reflecting a larger stimulus quality effect for low-frequency words than for high-frequency words.

Vincentile analyses. The mean vincentiles for these data are plotted in the top two thirds of Figure 5, and the frequency effect for the clear and degraded conditions are displayed in the bottom third of the figure. As most clearly shown in the bottom third, there was a larger frequency effect in the degraded condition, primarily at the slower vincentiles, which converges with the results from the ex-Gaussian analyses. Again, comparing Figure 5 and Figure 2, one can see a strong difference in the joint effects of stimulus quality and word frequency in pronunciation and lexical decision performance.

Discussion

In Experiment 1, using the LDT, we observed robust additive effects of stimulus quality and word frequency. In Experiment 2, through use of speeded pronunciation, the same stimuli and stimulus quality manipulation produced a clear interaction between stimulus quality and word frequency in mean RTs, with larger stimulus quality effects for low-frequency words. One obvious interpretation of these findings is that the additive effects of stimulus quality and word frequency may not be task independent but may instead reflect task operations that are specific to the LDT. However, because of the importance of this interaction in pronunciation performance, it is necessary to replicate and extend this pattern.

Experiment 3

Although a significant interaction was obtained with mean RTs, none of the parameters produced an interaction. Basically, the

[4] Because the same stimuli were used for lexical decision (Experiment 1) and speeded pronunciation (Experiments 2 and 3), the high- and low-frequency words were not matched in advance on phonological factors known to bias voice key RT measurement (see Kessler, Treiman, & Mullennix, 2002). It was assumed that this would not be a problem because the same items occur in both the clear and degraded conditions. However, to address a reviewer's concern that voice key effects may influence sensitivity to interactions, we controlled for onset characteristics (using the method described in Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004) in the speeded pronunciation experiments before carrying out the ANOVAs. After we controlled for phonological onset characteristics, the Stimulus Quality × Word Frequency interaction was still significant in both Experiments 2 and 3.

Table 4
Means of Participants' Speeded Pronunciation Response Time Means and Ex-Gaussian Parameter Estimates as a Function of Word Frequency and Stimulus Quality (Experiment 2—Between Participants)

Stimulus quality/word frequency	*M*	% of errors	μ	σ	τ
Clear words					
High-frequency words	516	0.8 (1.3)	468	74	48
Low-frequency words	533	1.5 (1.5)	478	79	56
Frequency effect	17	0.7 (0.2)	10	5	8
Mask-degraded words					
High-frequency words	597	1.7 (2.0)	493	69	105
Low-frequency words	635	2.2 (2.8)	506	74	129
Frequency effect	38	0.5 (0.8)	13	5	24
Difference of difference (interaction)	21	−0.2 (0.6)	3	0	16

Note. Means are given in milliseconds. Standard deviations are in parentheses.

interaction was exhibited to some extent in both μ and τ. Because of the theoretical importance of this interaction, it is necessary to both replicate this pattern of results with an independent sample and to increase the power of the design. Hence, Experiment 3 is essentially a replication of Experiment 2, with both stimulus quality and word frequency manipulated within participants, which should increase power.

Method

Participants

A total of 48 young adults (mean age = 21.3 years, *SD* = 3.40) participated in this study for course credit. All participants had normal or corrected-to-normal vision and were recruited from the undergraduate student population of Washington University in St. Louis. The participants had an average of 14.3 years of education (*SD* = 1.19) and a mean vocabulary age of 18.6 (*SD* = 1.03) on the Shipley vocabulary subtest. Data from 3 of the 48 participants were discarded because of excessively high error rates or slow latencies, using the same multivariate outlier procedure described in Experiment 1, leaving 45 participants. The apparatus, stimuli, and procedure were the same as those used in Experiment 2.

Design

A 2 × 2 factorial design was used: Both stimulus quality and word frequency were manipulated within participants. The original 200 stimuli were divided into two sets of 100 words (Set A and Set B), with each set containing 50 high- and 50 low-frequency words; the two sets were matched on length and word frequency.[5]

Results

Errors (1.8%) and extreme response latencies (faster than 200 ms or slower than 3,000 ms) were first excluded from the analyses, and the overall mean and standard deviation of each participant's word latencies were then computed on the remaining latencies. Response latencies 2.5 standard deviations above or below each participant's respective mean latency were removed. These criteria eliminated a further 2.6% of the pronunciation responses. ANOVAs were then carried out on the mean, accuracy, and the ex-Gaussian parameters of the RT data (see Table 5).

Mean response latencies and accuracy. For means, the main effect of stimulus quality, $F_p(1, 44) = 78.79$, $p < .001$, $MSE = 3{,}910.68$, $\eta^2 = .64$; $F_i(1, 198) = 185.86$, $p < .001$, $MSE = 3{,}773.73$, $\eta^2 = .48$, and word frequency, $F_p(1, 44) = 47.54$, $p < .001$, $MSE = 599.08$, $\eta^2 = .52$; $F_i(1, 198) = 15.00$, $p < .001$, $MSE = 4{,}305.12$, $\eta^2 = .07$, were significant by participants and items. More crucially, the Stimulus Quality × Word Frequency interaction was highly reliable by both participants and items, $F_p(1, 44) = 24.04$, $p < .001$, $MSE = 547.05$, $\eta^2 = .35$; $F_i(1, 198) = 8.00$, $p = .005$, $MSE = 3{,}773.73$, $\eta^2 = .039$, with a larger stimulus quality effect for low-frequency words than for high-frequency words. Turning to accuracy, the main effect of stimulus quality was significant, $F_p(1, 44) = 10.39$, $p = .002$, $MSE = 0.00099$, $\eta^2 = .19$; $F_i(1, 198) = 4.07$, $p = .045$, $MSE = 0.0029$, $\eta^2 = .02$, whereas the main effect of word frequency was not significant by participants ($F_p < 1$) but was significant by items, $F_i(1, 198) = 3.95$, $p = .048$, $MSE = 0.0030$, $\eta^2 = .020$. The interaction between stimulus quality and word frequency was not significant by participants or items.

Ex-Gaussian analyses. Turning to the ex-Gaussian parameters, for μ, both the main effects of stimulus quality, $F(1, 44) = 56.26$, $p < .001$, $MSE = 1{,}043.86$, $\eta^2 = .56$, and word frequency, $F(1, 44) = 11.87$, $p = .001$, $MSE = 1{,}180.91$, $\eta^2 = .21$, were significant. The interaction approached but did not reach significance ($p = .14$). Likewise, for σ, both the main effects of stimulus quality, $F(1, 44) = 12.07$, $p = .001$, $MSE = 725.50$, $\eta^2 = .22$, and word frequency, $F(1, 44) = 5.36$, $p = .025$, $MSE = 531.79$, $\eta^2 = .11$, were significant. The interaction was not significant ($F < 1$). Turning to τ, the main effect of stimulus quality was significant, $F(1, 44) = 19.66$, $p < .001$, $MSE = 4{,}982.45$, $\eta^2 = .31$. Neither the main effect of word frequency nor the Stimulus Quality × Word Frequency interaction was significant (*p*s > .10). It is noteworthy that in the within-participants degradation manipulation, there is no influence of word frequency on the τ component

[5] For the original 48 participants, half the participants saw Set A words clearly and Set B words with degradation (Order 1), and the other half of the participants saw Set B words clearly and Set A words with degradation (Order 2). The elimination of 3 participants resulted in 22 participants for the Order 1 condition and 23 participants for the Order 2 condition. The order variable did not interact with any of the experimental variables.

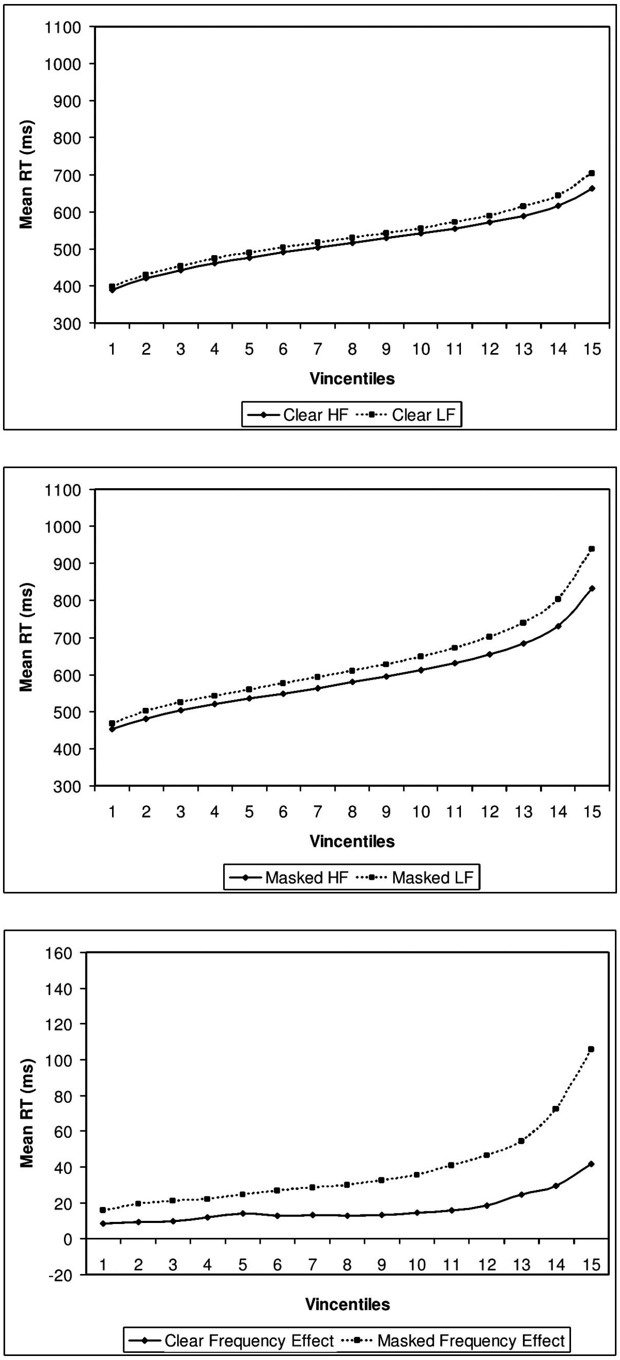

Figure 5. Experiment 2 vincentile means of participants' pronunciation response times as a function of stimulus quality and word frequency (between participants). RT = response time; HF = high frequency; LF = low frequency.

for the clear condition, replicating the pattern observed by Balota and Spieler (1999).

Vincentile analyses. The mean vincentiles for these data are plotted in the top two thirds of Figure 6, and the frequency effects for the clear and degraded conditions are plotted in the bottom third. As shown in the bottom third, the frequency effect for the degraded condition is considerably larger than that for the clear condition, and this trend appears to increase somewhat at the last vincentiles. The similarity between the bottom thirds of Figures 5 and 6 is quite striking.

Summary. Stimulus quality and word frequency produced interactive effects in means, with a larger stimulus quality effect for low-frequency words. Partitioning this interaction (see Table 5) suggested that the interaction was being mediated by both the μ and τ parameters; the interaction effect in means (35 ms) is attributable to both the μ (17 ms) and the τ component (18 ms). Coupled with the results in Experiment 2, this suggests that the interaction between stimulus quality and word frequency in pronunciation is due to the disproportionate shifting and skewing of visually degraded low-frequency words (although this must be qualified by the nonsignificant interaction in both μ and τ). This is further highlighted by the vincentile plot; the frequency effect is consistently larger for the degraded words and increases across the RT distribution.

Experiment 4

The results from the previous experiments provide an interesting dissociation between the effects of stimulus quality and frequency and reading task. Specifically, additive effects are obtained with lexical decision, whereas interactive effects are obtained with speeded pronunciation. In Experiment 4, we investigated whether interactive effects of degradation and frequency effects generalize to semantic categorization, a task that taps both word identification and meaning access processes (Forster & Shen, 1996). We have proposed that lexical decision's emphasis on familiarity is responsible for the additive effects in that task. Another binary decision task that does not allow decisions to be driven by familiarity information, such as semantic categorization, should therefore not produce additive effects. If we again observe interactive effects of degradation and frequency in semantic categorization, this would provide convergent support for the notion that the LDT is the outlier task. On the other hand, obtaining additive effects with semantic categorization would suggest that speeded pronunciation engages task-specific operations that are responsible for the interaction.

In Experiment 4, participants judged the animacy (i.e., animate or inanimate) of presented words; stimulus quality and word frequency were manipulated within participants. It is noteworthy that a number of researchers (e.g., Balota & Chumbley, 1984; Forster & Shen, 1996; Sears, Lupker, & Hino, 1999) have argued that data from nonexemplar trials are particularly useful in studying word identification, because there is no priming from the category on nonexemplar trials. In order to classify an item correctly, participants need to access its lexical entry and extract sufficient semantic information to drive a response. In addition, as recommended by Forster and Shen (1996), we used a single category throughout the experiment, and we used a single large, natural category to minimize postlexical identification processing and typicality effects (Monsell, 1991).

Method

Participants

A total of 45 young adults (mean age = 19.6 years, SD = 1.3) participated in this study for course credit. All participants had

Table 5
Means of Participants' Speeded Pronunciation Response Time Means and Ex-Gaussian Parameter Estimates as a Function of Word Frequency and Stimulus Quality (Experiment 3—Within Participant)

Stimulus quality/word frequency	*M*	% of errors	μ	σ	τ
Clear words					
High-frequency words	578	09 (3.1)	502	68	76
Low-frequency words	586	1.2 (2.7)	511	76	75
Frequency effect	8	0.6 (−0.4)	9	8	−1
Mask-degraded words					
High-frequency words	643	2.5 (5.2)	530	82	113
Low-frequency words	686	2.6 (4.8)	556	90	130
Frequency effect	43	0.1 (−0.4)	26	8	17
Difference of difference (interaction)	35	−0.5 (0.0)	17	0	18

Note. Means are given in milliseconds. Standard deviations are in parentheses.

normal or corrected-to-normal vision and were recruited from the undergraduate student population of Washington University in St. Louis. The participants had an average of 13.6 years of education ($SD = 1.3$) and a mean vocabulary age of 18.7 ($SD = 1.0$) on the Shipley vocabulary subtest. Data from 5 of the 45 participants were discarded because of excessively high error rates or slow latencies, assessed with the same multivariate outlier procedure described in Experiment 1, leaving 40 participants.

Apparatus and Stimuli

The apparatus was the same as Experiment 1. The stimuli for Experiment 4 consisted of 400 words, 200 animate words and 200 inanimate words extracted from Andrews and Heathcote's (2001) stimuli. Using the Lund and Burgess (1996) frequency norms, half the words in each animacy set were designated as high frequency (animate words mean counts per million = 115; inanimate words mean counts per million = 140), and the other half were designated as low frequency (animate words mean counts per million = 13; inanimate words mean counts per million = 20). For high-frequency words, the mean length was 6.20 letters for animate words and 6.23 for inanimate words, the mean orthographic neighborhood size was 1.82 for animate words and 1.83 for inanimate words, and the mean summed bigram frequency was 10,920.86 for animate words and 10,042.01 for inanimate words. For low-frequency words, the mean length was 6.22 for animate words and 6.33 for inanimate words, the mean orthographic neighborhood size was 1.69 for animate words and 1.73 for inanimate words, and the mean summed bigram frequency was 9,291.81 for animate words and 9,418.81 for inanimate words. There was no significant difference between high- and low-frequency words with respect to length or orthographic neighborhood size for both animate and inanimate words (all ts < 1).

Procedure

Participants were told that words would be presented at the center of the screen, and their task was to indicate as quickly and as accurately as possible via a button press on the keyboard (the apostrophe key for living words and the *A* key for nonliving words, or vice versa) whether the word was a living or nonliving object. Participants were presented with 20 practice trials, followed by five experimental blocks of 80 trials, with mandatory breaks occurring between blocks. The order in which stimuli were presented was randomized anew for each participant. Each trial consisted of the following order of events: (a) a fixation point (+) at the center of the monitor for 2,000 ms, (b) a blank screen for 650 ms, and (c) a stimulus centered at the fixation point's location. The stimulus remained on screen until a keyboard response was made. Responses were followed by a 1,000-ms delay. If the response was incorrect, at the onset of the 1,000-ms delay, there was a 170-ms tone that was presented simultaneously with the word *Incorrect* displayed for 450 ms slightly below the fixation point.

Design

A 2 × 2 factorial design was used: Both stimulus quality and word frequency were manipulated within participants. The original 200 stimuli were divided into two sets of 100 words (Set A and Set B), with each set containing 50 high- and 50 low-frequency words; the two sets were matched on length and word frequency. Half the participants saw Set A words clearly and Set B words with degradation, whereas the other half saw Set B words clearly and Set A words with degradation. In addition, the response keys used to make "animate" and "inanimate" responses were counterbalanced so that a particular key was used to indicate "animacy" for half the participants and "inanimacy" for the remaining participants.

Results

Errors (5.8%) and extreme response latencies (faster than 200 ms or slower than 3,000 ms) were first excluded from the analyses, and the overall mean and standard deviation of each participant's word latencies were then computed. Response latencies 2.5 standard deviations above or below each participant's respective mean latency were removed. These criteria eliminated a further 2.7% of the classification responses. ANOVAs were then carried out on the mean, accuracy, and ex-Gaussian parameters of the RT data (see Tables 6 and 7). To simplify the presentation of the results, we present the analyses for "inanimate" responses (see Table 6) first, followed by analyses for "animate" responses (see Table 7).

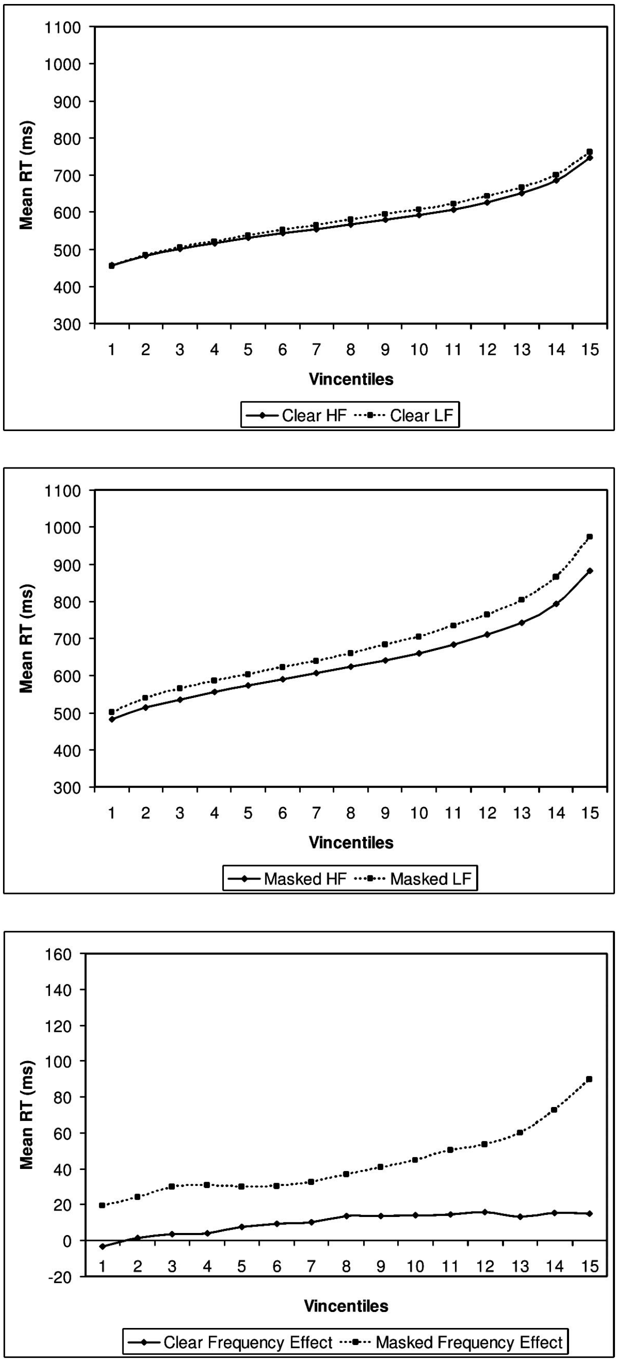

Figure 6. Experiment 3 vincentile means of participants' pronunciation response times as a function of stimulus quality and word frequency (within participants). RT = response time; HF = high frequency, LF = low frequency.

Stimulus Quality × Word Frequency (Inanimate)

Response latencies and accuracy. For means, the main effect of stimulus quality, $F_p(1, 39) = 55.08$, $p < .001$, $MSE = 7{,}309.59$, $\eta^2 = .59$; $F_i(1, 198) = 231.81$, $p < .001$, $MSE = 4{,}390.51$, $\eta^2 = .54$ and word frequency, $F_p(1, 39) = 67.07$, $p < .001$, $MSE = 1{,}267.87$, $\eta^2 = .63$; $F_i(1, 198) = 22.81$, $p < .001$, $MSE = 9{,}512.07$, $\eta^2 = .10$, were significant by participants and items. More crucially, the Stimulus Quality × Word Frequency interaction was reliable by participants and items, $F_p(1, 39) = 6.15$, $p = .018$, $MSE = 858.08$, $\eta^2 = .14$; $F_i(1, 198) = 4.10$, $p = .04$, $MSE = 4{,}390.51$, $\eta^2 = .02$, with a larger stimulus quality effect for low-frequency words than for high-frequency words. Turning to accuracy, the main effect of word frequency, $F_p(1, 39) = 5.57$, $p = .02$, $MSE = 0.0007$, $\eta^2 = .13$; $F_i(1, 198) = 3.13$, $p = .08$, $MSE = 0.013$, $\eta^2 = .016$, was significant. The main effect of stimulus quality was not significant by participants ($p = .32$) but was significant by items, $F_i(1, 198) = 5.32$, $p = .02$, $MSE = 0.003$, $\eta^2 = .03$. The Stimulus Quality × Word Frequency interaction was not significant by participants ($p = .18$) or by items ($p = .39$).

Ex-Gaussian parameters. Turning to the ex-Gaussian parameters, for μ, both the main effects of stimulus quality, $F(1, 39) = 41.03$, $p < .001$, $MSE = 2{,}091.84$, $\eta^2 = .51$, and word frequency, $F(1, 39) = 10.85$, $p = .002$, $MSE = 3{,}413.46$, $\eta^2 = .22$, were significant. The interaction did not reach significance ($F < 1$). For σ, none of the effects were significant. Turning to τ, the main effect of stimulus quality was significant, $F(1, 39) = 12.87$, $p = .001$, $MSE = 9{,}240.65$, $\eta^2 = .25$. The main effect of word frequency approached significance ($p = .09$), but the Stimulus Quality × Word Frequency interaction was not significant ($p = .27$).

Vincentile analyses. The mean vincentiles for these data are plotted in the top two thirds of Figure 7, and the mean frequency effects for the clear and degraded conditions across vincentiles are displayed in the bottom third of this figure. The data in the bottom third of Figure 7 clearly show that the frequency effect for the degraded condition is considerably larger than for the clear condition, and this trend appears to increase across vincentiles. These data appear most similar to the pronunciation data displayed in Figures 5 and 6.

Stimulus Quality × Word Frequency (Animate)

Mean response latencies and accuracy. For means, the main effects of stimulus quality, $F_p(1, 39) = 83.01$, $p < .001$, $MSE = 2{,}131.28$, $\eta^2 = .68$; $F_i(1, 198) = 93.64$, $p < .001$, $MSE = 4{,}836.51$, $\eta^2 = .32$, and word frequency, $F_p(1, 39) = 163.42$, $p < .001$, $MSE = 1{,}683.11$, $\eta^2 = .81$; $F_i(1, 198) = 78.08$, $p < .001$, $MSE = 11{,}028.92$, $\eta^2 = .28$, were significant by participants and items. The Stimulus Quality × Word Frequency interaction was not significant by participants or items (Fs < 1). Turning to accuracy, only the main effect of word frequency was significant, $F_p(1, 39) = 96.15$, $p < .001$, $MSE = 0.0017$, $\eta^2 = .71$; $F_i(1, 198) = 25.25$, $p < .001$, $MSE = 0.029$, $\eta^2 = .11$. Both the main effect of stimulus quality ($p = .20$) and the interaction effect ($F < 1$) were not significant.

Ex-Gaussian analyses. Turning to the ex-Gaussian parameters, for μ, both the main effects of stimulus quality, $F(1, 39) = 9.81$, $p = .003$, $MSE = 2{,}198.85$, $\eta^2 = .20$, and word frequency, $F(1, 39) = 60.46$, $p < .001$, $MSE = 1{,}785.05$, $\eta^2 = .61$, were significant. The interaction did not reach significance ($F < 1$). For σ, none of the effects were significant. Turning to τ, the main effect of stimulus quality, $F(1, 39) = 14.11$, $p = .001$, $MSE = 5{,}579.90$, $\eta^2 = .27$, and the main effect of word frequency, $F(1, 39) = 14.15$, $p = .001$, $MSE = 2{,}449.66$, $\eta^2 = .27$, were signif-

Table 6
Means of Participants' Semantic Classification Response Time Means and Ex-Gaussian Parameter Estimates as a Function of Word Frequency and Stimulus Quality (Experiment 4—Inanimate Responses)

Stimulus quality/word frequency	*M*	% of errors	μ	σ	τ
Clear words					
High-frequency words	764	3.5 (3.3)	577	54	185
Low-frequency words	799	5.0 (3.5)	605	58	194
Frequency effect	35	1.5 (0.2)	28	4	9
Mask-degraded words					
High-frequency words	853	4.5 (4.1)	621	64	232
Low-frequency words	910	4.9 (4.2)	654	70	256
Frequency effect	57	0.4 (0.1)	33	6	24
Difference of difference (interaction)	22	−1.1 (−0.1)	5	2	15

Note. Means are given in milliseconds. Standard deviations are in parentheses.

icant, but the Stimulus Quality × Word Frequency interaction was not significant ($F < 1$).

Vincentile analysis. The mean vincentiles for these data are plotted in the top two thirds of Figure 8, and the bottom third presents the word frequency effects for the clear and degraded conditions as a function of vincentile. As shown in the bottom third, these data appear a bit more noisy, and the frequency effect appears to be somewhat larger for the degraded conditions primarily at the middle vincentiles. Clearly, this pattern is quite different from the pattern displayed in Figures 3, 5, 6, and 7, in which reliable interactive effects were observed.

Discussion

In Experiment 4, using the semantic classification task, we obtained interactive effects of stimulus quality and word frequency, with larger stimulus quality effects for low-frequency words. This pattern is consistent with the results from speeded pronunciation (Experiments 2 and 3) and, more critically, is discrepant with the additive effects obtained in lexical decision (Experiment 1). The interaction is most clearly observed with the responses from the nonexemplar (i.e., inanimate) trials. For the animate trials, the two variables appeared to produce additive effects in means, although follow-up distributional analyses indicated that this additivity was not as clear as the type observed in Experiment 1 (cf. Figures 2 and 7).

These results reinforce the notion that exemplar trials may not be as useful for analyzing identification processes of isolated words (Balota & Chumbley, 1984; Forster & Shen, 1996; Sears et al., 1999) because of heightened activation of the exemplars within the category. Specifically, one might expect a trade-off between degradation and priming effects. As noted, there is an interaction between priming and word frequency, such that low-frequency words benefit more from priming than do high-frequency words. Hence, although low-frequency words are impaired more by degradation, as reflected by the inanimate trials, they also benefit more from being primed. Interestingly, evidence for this argument is available from Borowsky and Besner's (1993, Figure 8) primed lexical decision results. When word targets were preceded by unrelated primes, the frequency effect was much larger for degraded targets (regression coefficient = −61) than for clear targets (regression coefficient = −10). In contrast, this interaction goes away for related trials; the frequency effect was in fact very slightly larger for clear targets (regression coefficient = −13) than for degraded targets (regression coefficient = −6). Hence, the Stimulus Quality × Word Frequency interaction was considerably attenuated when words were preceded by a related prime in

Table 7
Means of Participants' Semantic Classification Response Time Means and Ex-Gaussian Parameter Estimates as a Function of Word Frequency and Stimulus Quality (Experiment 4—Animate Responses)

Stimulus quality/word frequency	*M*	% of errors	μ	σ	τ
Clear words					
High-frequency words	693	4.3 (4.1)	523	47	170
Low-frequency words	774	10.5 (6.0)	571	53	203
Frequency effect	81	6.2 (2.1)	48	6	33
Mask-degraded words					
High-frequency words	757	3.5 (3.6)	542	41	218
Low-frequency words	842	10.1 (6.1)	599	55	244
Frequency effect	85	6.6 (2.5)	57	14	26
Difference of difference (interaction)	4	0.4 (0.4)	9	8	−7

Note. Means are given in milliseconds. Standard deviations are in parentheses.

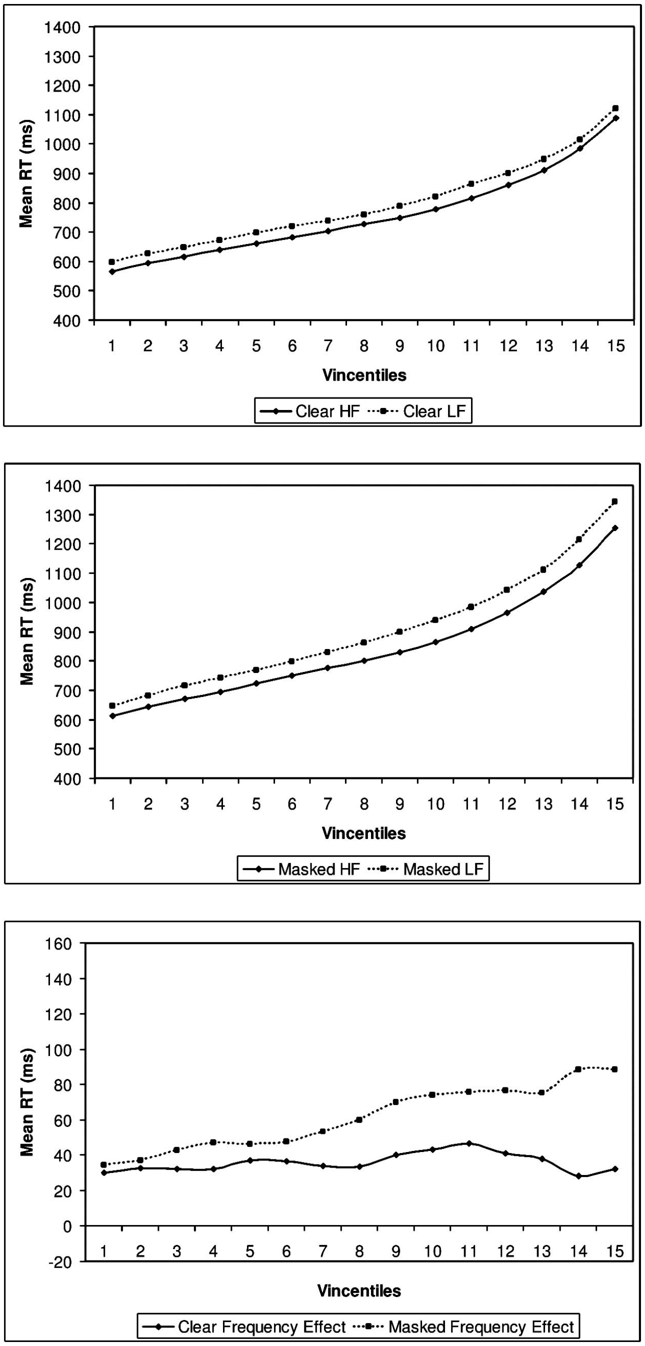

Figure 7. Experiment 4 vincentile means of participants' semantic classification response times as a function of stimulus quality and word frequency (inanimate responses). RT = response time; HF = high frequency; LF = low frequency.

Borowsky and Besner's study, mirroring our contrast between animate and inanimate responses.

Of course, this account is predicated on the assumption that the animate trials were the exemplar category and the inanimate trials were the nonexemplar category. In principle, both the animate and the inanimate may be represented as categories in this task, and so it is unclear which category should be defined as exemplar or nonexemplar. If there is indeed priming for a category as broad and inclusive as "animate," why should there not be priming for the "inanimate" category? Although it is possible that one makes a decision on the basis of an initial search of an inanimate category, there is evidence in the data that there may be a preference to

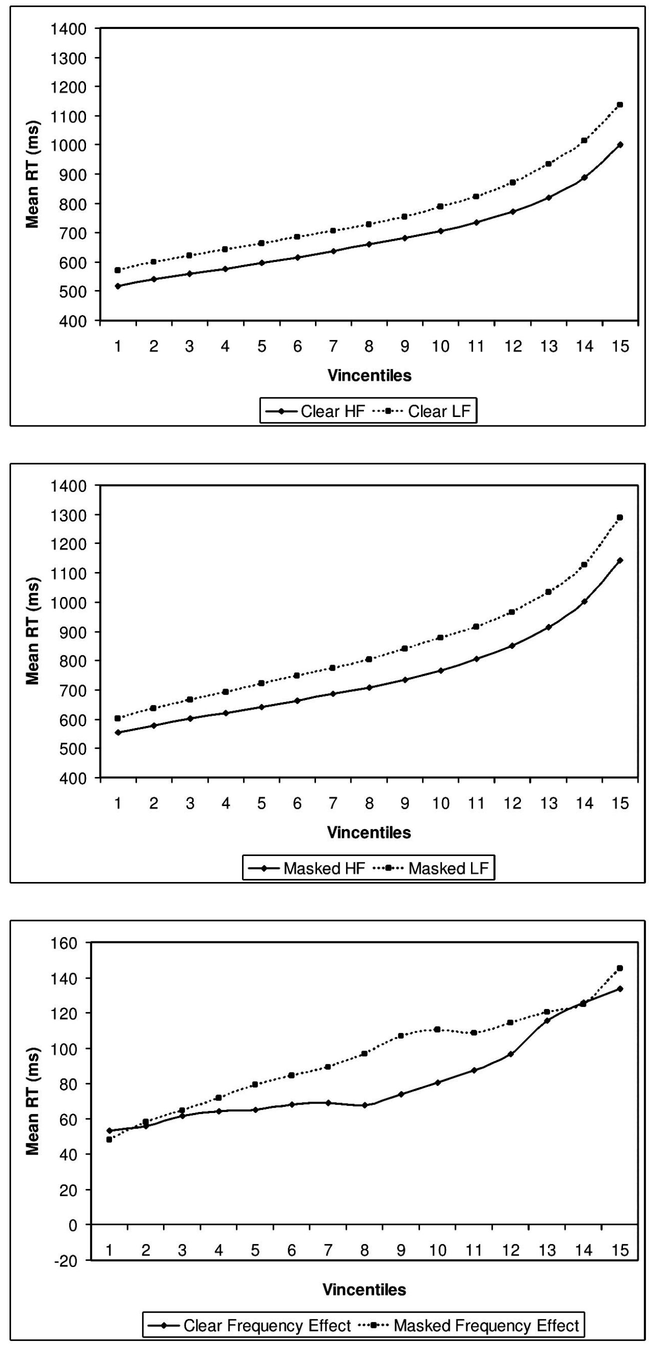

Figure 8. Experiment 4 vincentile means of participants' semantic classification response times as a function of stimulus quality and word frequency (animate responses). RT = response time; HF = high frequency; LF = low frequency.

search for animacy in our task. Specifically, if one examines Tables 6 and 7, it is apparent that the semantic classification times for animate trials were consistently faster than those for inanimate trials. Moreover, Andrews and Heathcote (2001) also reported faster semantic classification latencies for animate compared with inanimate stimuli, after controlling for word frequency and number of syllables. Collectively, this suggests that participants were using the presence (as opposed to absence) of animacy as the dimension by which semantic classifications were carried out.

The important finding of Experiment 4 is that when we considered RTs to nonexemplar words, we obtained the predicted interaction between stimulus quality and word frequency with a new set of stimuli. Specifically, stronger stimulus quality effects were observed for low-frequency words, a trend that is both consistent with the findings of Experiments 2 and 3 and predicted by extant models of word recognition.

General Discussion

The present series of experiments generated a number of notable observations. In Experiment 1, using the LDT, we found strong additive effects of stimulus quality and frequency in the means, ex-Gaussian parameters, and higher order moments. Experiments 2–4 were conducted to test whether these additive effects could be replicated in speeded pronunciation and semantic classification. Importantly, the latter tasks yielded interactive effects between word frequency and stimulus quality. We now turn to a discussion of the theoretical implications of these findings.

Implications for Models of Word Recognition

The major observation in this article is that stimulus quality and word frequency, which were strongly additive in lexical decision, were reliably interactive in speeded pronunciation and semantic classification. The additive effects of stimulus quality and word frequency have traditionally been interpreted as support for serial, independent stages (Sternberg, 1969a) and, as such, have often been cited as support for multistage models of word recognition (e.g., Becker & Killion, 1977; Borowsky & Besner, 1993; Forster, 1976; Paap et al., 1982). Conversely, it has been difficult to reconcile additive effects of stimulus quality and word frequency with models in which different variables influence a common localist (McClelland & Rumelhart, 1981; Morton, 1969) or distributed (Plaut et al., 1996) word representation.

The work described in this article provides some resolution to the conflict between the empirical findings and theoretical models. If the additive effects of stimulus quality and word frequency are indeed a faithful reflection of the word recognition system, then these effects should be task independent—that is, observed in lexical decision, speeded pronunciation, and semantic classification. The results clearly indicate that additive effects are not task independent but may be a function of the specific requirements of lexical decision. In contrast to the LDT, speeded pronunciation, using the same stimuli and the same degradation manipulation, produced a clear interaction between stimulus quality and word frequency, with larger word frequency effects in the degraded condition. Similarly, the nonexemplar responses in the semantic classification task also produced the same interaction. This interaction, which was replicated in three experiments, suggests that interactive effects of stimulus quality and word frequency may be a more general finding in lexical processing tasks than additive effects in lexical decision performance.

Interactive effects appear troublesome for models of lexical processing that invoke two stages. For example, Forster's (1976, 1992) bin model of lexical access posits two separate stages: a first stage in which candidates are identified and a second stage in which the candidates are evaluated through a serial, frequency-ordered process. The interaction between stimulus quality and word frequency in pronunciation and semantic classification undermines the necessity for two independent stages in lexical access and suggests that multiple-stage models (e.g., Borowsky & Besner, 1993; Forster, 1992) may be accounting for lexical decision performance rather than lexical access per se.

In contrast to the stage models, as discussed in the introduction, interactive effects are quite consistent with activation-class models. For example, Coltheart et al.'s (2001) computational DRC model of word recognition yields larger stimulus quality effects for low-frequency words in simulations (Reynolds & Besner, 2004). The following illustration provides a simple example of why one would predict such an interaction: Consider a high-frequency lexical entry that is 10 cycles away from threshold and a low-frequency entry that is 20 cycles away from threshold (i.e., a word frequency difference of 10 cycles). If the input to these two entries is degraded such that the input on each cycle decreases by half in the degraded condition (compared with the clear condition), then it should take twice as many cycles for the low-frequency entry to reach threshold. That is, the high-frequency entry will be recognized in 20 (2×10) cycles, whereas its low-frequency counterpart will be recognized in 40 (2×20) cycles. Hence, the original word frequency effect of 10 cycles will now yield a word frequency effect of 20 cycles in the degraded condition, thereby yielding the Stimulus Quality × Word Frequency interaction that we observed. The Stimulus Quality × Word Frequency interaction also appears consistent with the connectionist approaches to lexical processing. Specifically, the attractor network, distributed model described in Simulation 3 of Plaut et al. (1996) produces the same pattern of interactive effects (D. C. Plaut, personal communication, January 18, 2005).

In summary, the Stimulus Quality × Word Frequency interaction appears to be an important aspect of current computational models of word recognition. As we have emphasized, this is precisely why the additive effects of word frequency and stimulus quality in lexical decision are of particular interest.

Are Our Findings Specific to Our Method of Stimulus Degradation?

One potential criticism of the present experiments is that the findings may be specific to the type of stimulus quality manipulation used. Obviously, there are many possible ways to degrade a stimulus. Conventional methods include superimposing a random noise pattern (Balota & Abrams, 1995; Meyer, Schvaneveldt, & Ruddy, 1975; Stanners et al., 1975), contrast/luminance reduction (Becker & Killion, 1977; Borowsky & Besner, 1993; Plourde & Besner, 1997), and case alternation (Besner & McCann, 1987; Herdman et al., 1999). The type of stimulus degradation we used is somewhat different from these methods and involves rapidly alternating letter strings with a randomly generated mask of the same length. Hence, one might question whether the findings reported would have been observed using other methods of stimulus degradation. We believe that it is important to consider the type of degradation manipulation being used; specifying the boundary conditions for these effects is an important research question in its own right. However, we also believe that the interactive effects of stimulus quality and word frequency observed in Experiments 2, 3, and 4 are not merely artifacts of the degradation manipulation used for the following four reasons.

First, using our degradation manipulation, robust additive effects of stimulus quality and word frequency were observed in lexical decision, in line with researchers who have employed other forms of degradation manipulation (Borowsky & Besner, 1993; Becker & Killion, 1977; Stanners et al., 1975). Second, Plourde and Besner (1997) also used distributional analysis to explore the additive effects of stimulus quality and word frequency in lexical decision, using a within-participant design with a contrast reduction manipulation. The results of the present distributional analyses were virtually identical to Plourde and Besner's, with additivity in means, ex-Gaussian parameters, and higher order moments. Third, as noted in the introduction, there is already some indication in the literature that stimulus quality and word frequency interact in speeded pronunciation performance when other forms of stimulus degradation are employed. For example, Besner and McCann (1987) reported interactive effects when case alternation was used to degrade words, and Herdman et al. (1999) showed a clear trend towards an interaction when contrast reduction was used. More important, using the stimuli from the present study, O'Malley, Reynolds, and Besner (in press) recently observed an interaction between word frequency and stimulus degradation when a stimulus contrast manipulation was used. Fourth, as mentioned in the *Discussion* section of Experiment 4, there is already some indirect evidence supporting the interactivity of stimulus quality and word frequency in semantic classification. Specifically, our semantic classification findings are consistent with Borowsky and Besner's (1993) primed lexical decision results, in which contrast reduction was used to degrade items. When targets were preceded by unrelated primes, the frequency effect was much larger for degraded targets than for clear targets, whereas the frequency effects were of similar size for clear and degraded targets when they were preceded by related primes. Assuming that participants use prime–target relations to bias the word response in lexical decision performance (see Neely, 1991), the results from Borowsky and Besner's (1993) study with a contrast reduction manipulation are quite consistent with our semantic classification results.

Does Additivity Necessarily Denote Stages?

Another potentially controversial issue concerns the interpretation of additive effects. As discussed in the introduction, separate stages imply additive effects, but additive effects do not necessarily imply separate stages. This is particularly critical because our theoretical account of a perceptual normalization process that is specific to lexical decision rests firmly on the assumption that additive effects are indicative of an additional stage of processing. However, as Roberts and Sternberg (1993) have pointed out, one can obtain approximately additive effects, as in Experiment 1, with a nonstage architecture like the cascade model.

Although the results from the present Experiment 1 do not allow us to conclusively adjudicate between these two alternatives, we would argue that there are several converging pieces of evidence that are easier to reconcile with a stage than a cascade model. First, the additivity of stimulus quality and word frequency in lexical decision holds in means and higher order moments. The Roberts and Sternberg (1993) simulations show that the cascade model predicts additivity at the level of the mean but not necessarily at the level of higher order moments. Second, if the additive effects in lexical decision indeed represent a cascadic process, it is unclear why such effects should be limited to lexical decision. The most intriguing aspect of these experiments is the qualitatively different empirical patterns across lexical processing tasks, and it is difficult to conceptualize a single mechanism that produces additive effects of two factors in one task but interactive effects in another. Third, it is unclear whether the additive effects of stimulus quality and word frequency and the interactive effects of stimulus quality and semantic context in lexical decision can be simultaneously accommodated without positing stages (see Borowsky & Besner, 2006; Plaut & Booth, 2006). The basic problem is that the sigmoid-based single-process explanation provided by Plaut and Booth (2000) appears to be unable to correctly simulate additive and interactive effects of the three factors when the RTs being simulated fall within the same range. In general, the results from Experiment 1 seem prima facie easier to reconcile with stages than with cascaded processing. Having said that, the validity of this claim can be fully evaluated only with appropriate simulations, and we look forward to such modeling work in the future.

Joint Effects of Stimulus Quality, Word Frequency, and Semantic Context

One of the motivations for this study was to better understand the interesting conundrum in which stimulus quality has additive effects with word frequency but interactive effects with semantic context. If word frequency and semantic context effects indeed reflect variations in the resting activation or activation threshold of unitary lexical representations, it is puzzling that semantic context and word frequency interact with stimulus quality in qualitatively distinct ways in lexical decision. Solutions to this problem include assuming that stimulus quality influences an encoding stage, word frequency influences a subsequent retrieval stage, and semantic context has effects on both stages (cf., the multistage activation model; Borowsky & Besner, 1993), or that additive and interactive effects represent different portions of a sigmoid input–output activation function (cf., Plaut & Booth, 2000).

The fact that additive effects of stimulus quality and word frequency are localized in lexical decision performance (at least for this series of experiments) suggests that the paradoxical joint effects of stimulus quality, word frequency, and semantic context may not apply to word recognition performance in general but are specific to the LDT. We have claimed that lexical decision emphasizes familiarity-based information, and degradation undermines this information, making normalization necessary. If this is the case, it is not immediately obvious why stimulus degradation should have differential effects on semantically primed versus unprimed words but equivalent effects on high- versus low-frequency words. For example, one might consider priming effects as reflecting some kind of lexical spreading activation mechanism (cf., Anderson, 1983; Collins & Loftus, 1975); *CAT* facilitates the recognition of *DOG* because activation spreads rapidly from *CAT* to its associates when the prime is presented. Clearly, this kind of account implicates a lexical locus for priming effects. However, we have already argued that for lexical decision, the early perceptual normalization stage is insensitive to top-down lexical feedback, making it unclear why degraded targets show a larger priming effect in lexical decision.

A study by Stolz and Neely (1995) may provide some resolution to this puzzle. In two lexical decision experiments, Stolz and Neely examined how the typically observed interaction between stimulus quality and semantic context was modulated by variables like relatedness proportion (the proportion of prime–target trials sharing a semantic relation) and prime–target stimulus onset asynchrony. Interestingly, they observed the standard overadditive Stimulus Quality × Semantic Context interaction only when the relatedness proportion was .50. In contrast, when the relatedness proportion was reduced to .25, additive effects of stimulus quality and relatedness were obtained. Importantly, these effects were observed with stimulus onset asynchronies of both 200 ms and 800 ms, indicating that these effects cannot be attributed entirely to expectancy-based processes.

To summarize, one obtains interactive effects of stimulus quality and semantic context when relatedness proportion is high but additive effects when relatedness proportion is low. We interpret these findings as being consistent with the idea that when the relatedness proportion is high (i.e., the payoff is high), participants have an incentive to attend to the semantic context (possibly due to checking for a prime–target relationship to bias the word response), and this is reflected by larger priming effects for the degraded targets. In contrast, when the relatedness proportion is low (i.e., the payoff is low), participants have less incentive to use the context in lexical decision. Instead, stimuli are perceptually normalized before lexical retrieval processes are engaged, and this is reflected by priming effects that are equivalent for clear and degraded targets. The fact that interactive effects are specific to priming conditions in which relatedness proportion is high suggests that strategic control may be associated with the typical Stimulus Quality × Semantic Context interaction. The important point is that when the participant has little incentive to attend to the semantic context, the perceptual normalization process in lexical decision does not seem to be sensitive to top-down lexical feedback.

Implications for Models of Lexical Decision Performance

Collectively, the results of the four experiments suggest that additive effects of word frequency and stimulus degradation, and their implication of an early stimulus normalization stage, are specific to lexical decision. Just as the main effect of frequency may be exaggerated by the discrimination component of the LDT (Balota & Chumbley, 1984), lexical decision may require degraded word stimuli to undergo perceptual normalization before they can be classified as a word or a nonword. This may reflect the emphasis on familiarity-based information to make the lexical decision (Balota & Chumbley, 1984). Specifically, because this familiarity-based information is undermined by degradation, the LDT encourages a normalization process to recover that information so that familiar words can be discriminated from unfamiliar nonwords. Interestingly, as discussed earlier, degradation also produces additive effects with stimulus set size in a memory scanning task. It is possible that in experimental contexts in which binary decisions are demanded and familiarity is a useful dimension to make the binary decision, there is an increasing emphasis on familiarity-based information to drive the response, making it necessary to normalize stimuli before familiarity-based information is readily available.

How might familiarity map onto the lexical processes in extant word recognition models? In lexical decision, the datum driving a "word" response can be based on a specific orthographic entry or all the entries in the orthographic lexicon. For example, both the DRC model (Coltheart et al., 2001) and the multiple read-out model (MROM; Grainger & Jacobs, 1996) make word lexical decisions when the activation level of a single lexical representation (local activity) or the summed activation levels of all lexical representations (global activity) exceed their respective thresholds. Global lexical activation would seem to map onto familiarity. In general, words will have higher global lexical activity than nonwords, making this a very diagnostic dimension for word–nonword discrimination. Indeed, for the lexical decision stimuli used in Experiment 1, the mean orthographic neighborhood size was higher for words (4.80) than for nonwords (3.38), where orthographic neighborhood size is defined as the number of words that can be created by changing a single letter of the target word (Coltheart et al., 1977). Orthographic neighborhood size is an approximate measure of the number of words that are orthographically similar to a target and functions as a good proxy for global lexical activity. It is possible that word–nonword discrimination relies heavily on this dimension. In contrast, it is clear that global lexical activity is not useful in either speeded pronunciation or semantic classification performance. For example, in the DRC and MROM simulations, global lexical activity drives "word" responses in lexical decision but not pronunciation or perceptual identification responses (Coltheart et al., 2001; Grainger & Jacobs, 1996). The system needs to locate a specific lexical entry before phonology can be initiated or an impoverished stimulus can be identified. We argue that this is also the case for the semantic classification task.

Thus far, we have sought to explain the between-task dissociations primarily in terms of lexical decision's emphasis on familiarity. This, of course, is overly simple. The LDT clearly is not a unitary task and has been likened to a signal-detection task that may recruit different types of information depending on the specific nonword context (Seidenberg, 1990). For example, when orthographically illegal words (e.g., *BRNTA*) are used as foils, discrimination can be based on orthographic information alone. In contrast, when orthographically legal words (e.g., *BRANT*) are used, then either phonological or semantic information needs to be consulted. Furthermore, Yap et al. (2006) have demonstrated that for the same set of high- and low-frequency words, the word frequency effect increases as nonword foils become more wordlike (i.e., *BRNTA* vs. *BRANT* vs. *BRANE*), and this Nonword Type × Word Frequency interaction is modulated by different components of the RT distribution, depending on the specific nonword contrast being examined.

Such findings reinforce the idea that lexical decision performance for words is strongly modulated by the specific nonword context. It is quite plausible that when foils are very similar to words (homophonous with a real word, e.g., *BRANE*), the LDT is more likely to emphasize fine-grained, letter-by-letter individuation compared with when foils are very distinct from words (orthographically illegal and unpronounceable, e.g., *BRNTA*). In the present lexical decision experiment, the words possessed denser orthographic neighborhoods than the nonwords, making global lexical activity (or familiarity) a viable dimension for word–nonword discrimination. Degradation attenuated this global activation/familiarity dimension, making normalization necessary for the signal to be recovered. Of course, this also implies that when global lexical activity is less useful for discriminating between words and nonwords (e.g., when global lexical activity of words and nonwords are perfectly matched), the emphasis on familiarity-based information is reduced, and normalization may no longer be mandatory. This leads to the intriguing prediction that the effects of stimulus quality and word frequency in lexical decision may become interactive when global activity is less useful for discriminating words from nonwords (D. Besner, personal communication, January 3, 2006). For example, when pseudohomophones (e.g., *BRANE*) are used as foils, there is more overlap of global lexical activity/familiarity between the words and distractors, making it more likely that the system will rely on local lexical activity to drive decisions. The central issue here is whether additive effects of stimulus quality and word frequency in lexical decision generalize to situations in which words and nonwords overlap heavily in familiarity. This is an important empirical question that, as far as we know, has not been addressed in the literature.

To summarize, the claim is that for the stimuli used in Experiment 1, global lexical activity is diagnostic for word–nonword discrimination. However, this creates a curious situation. Thus far, we have argued that global lexical activity plays more of a role in lexical decision than in pronunciation or semantic classification, in which a specific lexical candidate has to be identified. However, given that our high- and low-frequency words were matched on orthographic neighborhood size (a good proxy for global lexical activity), one would expect relatively attenuated frequency effects in lexical decision if global lexical activity were the primary source of information consulted for lexical decision. In contrast, one would predict larger frequency effects for speeded pronunciation, which relies mainly on local lexical activity. Nevertheless, for the same set of words,[6] we obtained larger frequency effects in lexical decision compared with pronunciation, a trend that is very consistent with the literature (see, e.g., Andrews & Heathcote, 2001; Balota & Chumbley, 1984). There is a relatively straightforward response to this apparent discrepancy. First, even if high- and low-frequency words are perfectly matched on orthographic neighborhood size, high-frequency words will still posses higher global lexical activation than low-frequency words, because of the frequency of the target word. Second, lexical decision may engage a postlexical decision process that is highly sensitive to such differences in activation, and this decision component can exaggerate the word frequency effect either through an additional attention-demanding check process (e.g., Balota & Chumbley, 1984; Balota & Spieler, 1999) or through a noisy evidence-accumulating process (e.g., Ratcliff et al., 2004). Again, because word frequency effects are strongly modulated by the type of nonwords, it is important to keep in mind that frequency effects in lexical decision reflect both the properties of lexical access and decision mechanisms (see Yap et al., 2006, for further discussion).

Interestingly, the robust Stimulus Quality × Lexicality interaction observed in Experiment 1 (larger stimulus quality effects for nonwords) also makes it clear that words are more resistant to degradation than nonwords (see Table 3). As pointed out in the *Discussion* section of Experiment 1, this interaction may reflect a more conservative response criterion for degraded nonwords. In any case, it is clear that the effects of stimulus quality, and the variable's intriguing relationship with word frequency and lexicality, have not been adequately considered by extant models of lexical decision performance. For example, an early perceptual normalization stage is outside the scope of extant models that accommodate lexical decision performance, such as the MROM (Grainger & Jacobs, 1996), the DRC model (Coltheart et al., 2001), the two-stage model of lexical decision performance (Balota & Spieler, 1999), and Ratcliff et al.'s (2004) diffusion model.

In particular, Ratcliff and colleagues have proposed that all the important phenomena in lexical decision could be modeled using a diffusion process. In the diffusion model, a type of random-walk model, binary decisions are driven by the accumulation of noisy information from a stimulus over time. The diffusion process starts from point Z, and information accumulates toward either of two decision criteria, reflected by a (positive response boundary) and 0 (negative response boundary), respectively. When a criterion is reached, a response is made. The rate at which information accumulates is the *drift rate*. More wordlike words (e.g., high-frequency words) have a steeper drift rate than less wordlike words (e.g., low-frequency words) and thus reach the "word" criterion earlier. Analogously, less wordlike nonwords (e.g., orthographically illegal nonwords, *NBRO*) have a steeper drift rate than more wordlike nonwords (e.g., orthographically legal nonwords, *BRON*), and they also reach the "nonword" criterion earlier. The predicted drift rates for the different kinds of words and nonwords can be computed from a simple two-dimensional representation of the strings' wordlikeness, the two dimensions being lexical strength and orthographic wordlikeness. The difference among drift rates between high-frequency words and low-frequency words is also exaggerated when the similarity between words and nonword foils is decreased, as when orthographically illegal foils like *HGQPA* are used, compared with pronounceable foils such as *FLIRP*.

Ratcliff et al. (2004) reported that drift-rate variations were sufficient for modeling a variety of lexical decision phenomena. Intuitively, however, it is unclear how the diffusion model could simulate both the additive effects of stimulus quality and word frequency coupled with the interactive effects of stimulus quality and lexicality using drift rate alone. For example, stimulus degradation might decrease the drift rates of high- and low-frequency words. This, however, would predict larger frequency effects for degraded words, which is inconsistent with the lexical decision data. Without carrying out actual fitting with the diffusion model, there is no immediately obvious way for accommodating our findings within the diffusion framework. It is at least conceivable that additive effects of stimulus quality and word frequency, coupled with the interactive effects of lexicality and degradation, challenge the claim that the diffusion model provides a unified

[6] A different set of words was used for semantic classification.

account of lexical decision phenomena (Ratcliff et al., 2004). Of course, this is not only a potential problem for the diffusion model; we believe that this intriguing pattern is a challenge for virtually all extant models of lexical decision performance.

Utility of Distributional Analyses and Cross-Task Convergence

These experiments underscore how distributional analyses serve as a valuable complement to traditional analyses of means. Both ex-Gaussian fitting and vincentile plots are accessible, intuitive tools that can exploit the entire RT distribution, allowing existing data to be scrutinized at a more fine-grained level. Clearly, there may be alternative ways of considering a variable's influence on RT distributions that are preferable (see Van Zandt, 2000, for several examples). For example, RT distributions have also been modeled with the Weibull (Logan, 1995) and ex-Wald (Schwarz, 2001) distributions. In this study, we have demonstrated how ex-Gaussian analyses allow mean differences to be partitioned into two components: one that reflects distributional shifting and one that reflects distributional skewing. Importantly, ex-Gaussian analysis allows researchers to evaluate aspects of models that are not testable with traditional analyses of means (Hockley, 1984). Moreover, in our analyses, assumption-free vincentile plots were presented alongside ex-Gaussian parameter estimates (cf., Andrews & Heathcote, 2001), providing converging validation of parameter estimates and a representation of the raw data. As Roberts and Sternberg (1993) have forcefully argued, these converging distributional approaches are particularly critical in adjudicating between interactive and stage models.

The present findings also highlight the importance of cross-task convergence and the importance of separating task-specific and task-general operations. In agreement with Jacoby's (1991) argument in the memory literature, we argue that it is unlikely there are process-pure measures of lexical processing. Participants bring to tasks online processes that maximize performance under a given set of conditions (Balota, Paul, & Spieler, 1999). In this light, there is probably no general, unitary lexical retrieval mechanism. Instead, lexical identification processes are likely to be modulated by the task context in which they are embedded.

References

Abrams, R. A., & Balota, D. A. (1991). Mental chronometry: Beyond reaction time. *Psychological Science, 2,* 153–157.

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22,* 261–295.

Andrews, S., & Heathcote, A. (2001). Distinguishing common and task-specific processes in word identification: A matter of some moment? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 514–544.

Ashby, F. G. (1982). Deriving exact predictions from the cascade model. *Psychological Review, 89,* 599–607.

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: Learning, memory, and thinking* (Vol. 1, pp. 242–293). San Francisco, CA: Freeman.

Balota, D. A., & Abrams, R. A. (1995). Mental chronometry: Beyond onset latencies in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1289–1302.

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 and Performance, 10,* 340–357.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133,* 283–316.

Balota, D. A., Paul, S., & Spieler, D. H. (1999). Attentional control of lexical processing pathways during word recognition and reading. In S. Garrod & M. Pickering (Eds.), *Language processing* (pp. 15–57). East Sussex, England: Psychology Press.

Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General, 128,* 32–55.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. I., Kessler, B., Loftis, B., et al. (in press). The English Lexicon Project. *Behavior Research Methods.*

Becker, C. A. (1979). Semantic context and word frequency effects in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 252–259.

Becker, C. A., & Killion, T. H. (1977). Interaction of visual and cognitive effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 389–401.

Besner, D. (1983). Basic decoding components in reading: Two dissociable feature extraction processes. *Canadian Journal of Psychology, 37,* 429–438.

Besner, D., & McCann, R. S. (1987). Word frequency and pattern distortion in visual word identification and production: An examination of four classes of models. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 201–219). Hillsdale, NJ: Erlbaum.

Borowsky, R., & Besner, D. (1993). Visual word recognition: A multistage activation model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 813–840.

Borowsky, R., & Besner, D. (2006). Parallel distributed processing and lexical-semantic effects in visual word recognition: Are a few stages necessary? *Psychological Review, 113,* 181–195.

Broadbent, D. E. (1984). The Maltese cross: A new simplistic model for memory. *The Behavioral and Brain Sciences, 7,* 55–94.

Collins, A., & Loftus, E. (1975). A spreading activation theory of semantic processing. *Psychological Review, 82,* 407–428.

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

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

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica, 30,* 412–431. (Original work published 1868)

Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. C. T. Walker (Eds.), *New approaches to language mechanisms* (pp. 257–287). Amsterdam, the Netherlands: North-Holland.

Forster, K. I. (1992). Memory-addressing mechanisms and lexical access. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 413–434). Amsterdam, the Netherlands: North-Holland.

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

Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution.* New York: Basic Books.

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

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review, 9,* 394–401.

Heathcote, A., Popiel, S. J., & Mewhort, D. J. K. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin, 109,* 340–347.

Herdman, C. M., Chernecki, D., & Norris, D. (1999). Naming case alternated words. *Memory & Cognition, 27,* 254–266.

Hockley, W. E. (1984). Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 598–615.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30,* 513–541.

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

Lattin, J. M., Carroll, J. D., & Green, P. E. (2003). *Analyzing multivariate data.* Pacific Grove, CA: Brooks Cole.

Logan, G. D. (1995). The Weibull distribution, the power law, and the instance theory of automaticity. *Psychological Review, 102,* 751–756.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* New York: Oxford University Press.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers, 28,* 203–208.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86,* 287–330.

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

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1975). Loci of contextual effects on visual word-recognition. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 98–118). New York: Academic Press.

Monsell, S. (1991). The nature and locus of word frequency effects in reading. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 148–197). Hillsdale, NJ: Erlbaum.

Morton, J. (1969). The interaction of information in word recognition. *Psychological Review, 76,* 165–178.

Murray, W. S., & Forster, K. I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review, 111,* 721–756.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 236–264). Hillsdale, NJ: Erlbaum.

Nelder, J. A., & Mead, R. (1965). A simplex algorithm for function minimization. *Computer Journal, 7,* 308–313.

Norris, D. (1984). The effects of frequency, repetition and stimulus quality in visual word recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 36*(A), 507–518.

O'Malley, S., Reynolds, M., & Besner, D. (in press). Qualitative differences between the joint effects of stimulus quality and word frequency in reading aloud and lexical decision: Extensions to Yap and Balota (2006). *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

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.

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes, 12,* 765–805.

Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review, 107,* 786–823.

Plaut, D. C., & Booth, J. R. (2006). More modeling but still no stages: Reply to Borowsky and Besner. *Psychological Review, 113,* 196–200.

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

Plourde, C. E., & Besner, D. (1997). On the locus of the word frequency effect in visual word recognition. *Canadian Journal of Experimental Psychology, 51,* 181–194.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review, 111,* 159–182.

R Development Core Team. (2004). *R: A language and environment for statistical computing* [Computer software]. Vienna, Austria: R Foundation for Statistical Computing.

Reynolds, M., & Besner, D. (2004). Neighborhood density, word frequency, and spelling-sound regularity effects in naming: Similarities and differences between skilled readers and the dual route cascaded computational model. *Canadian Journal of Experimental Psychology, 58,* 13–31.

Roberts, S. (1987). Evidence for distinct serial processes in animals: The multiplicative-factors method. *Animal Learning and Behavior, 15,* 135–173.

Roberts, S., & Sternberg, S. (1993). The meaning of additive reaction-time effects: Tests of three alternatives. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 611–653). Cambridge, MA: MIT Press.

Rouder, J. N., & Speckman, P. L. (2004). An evaluation of the vincentizing method of forming group-level response time distributions. *Psychonomic Bulletin & Review, 11,* 419–427.

Sanders, A. F. (1990). Issues and trends in the debate on discrete versus continuous processing of information. *Acta Psychologica, 74,* 123–167.

Schneider, W., Eschman, A., & Zuccolotto, A. (2001). *E-Prime user's guide.* Pittsburgh, PA: Psychology Software Tools.

Schwarz, W. (2001). The ex-Wald distribution as a descriptive model of response times. *Behavior Research Methods, Instruments, & Computers, 33,* 457–469.

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

Seidenberg, M. S. (1990). Lexical access: Another theoretical soupstone? In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 33–71). Hillsdale, NJ: Erlbaum.

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

Shipley, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *Journal of Psychology, 9,* 371–377.

Speckman, P. L., & Rouder, J. N. (2004). A comment on Heathcote, Brown, and Mewhort's QMLE method for response time distributions. *Psychonomic Bulletin & Review, 11,* 574–576.

Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance in normal older adults and individuals with senile dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 461–479.

Stanners, R. F., Jastrzembski, J. E., & Westbrook, A. (1975). Frequency and visual quality in a word-nonword classification task. *Journal of Verbal Learning and Verbal Behavior, 14,* 259–264.

Sternberg, S. (1967). Two operations in character recognition: Some evidence from reaction-time measurements. *Perception and Psychophysics, 2,* 45–53.

Sternberg, S. (1969a). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica, 30,* 276–315.

Sternberg, S. (1969b). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist, 57,* 421–457.

Sternberg, S. (1998). Discovering mental processing stages: The method of additive factors. In D. Scarborough & S. Sternberg (Eds.), *Methods, models, and conceptual issues: An invitation to cognitive science* (pp. 703–863). Cambridge, MA: MIT Press.

Stolz, J. A., & Neely, J. H. (1995). When target degradation does and does not enhance semantic context effects in word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 596–611.

Townsend, J. T. (1984). Uncovering mental processes with factorial experiments. *Journal of Mathematical Psychology, 28,* 363–400.

Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes.* London: Cambridge University Press.

Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review, 7,* 424–465.

Vincent, S. B. (1912). The function of vibrissae in the behavior of the white rat. *Behavioral Monographs, 1*(Whole No. 5).

Wilding, J. M. (1988). The interaction of word frequency and stimulus quality in the lexical decision task: Now you see it, now you don't. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 40(A),* 757–770.

Yap, M. J., Balota, D. A., Cortese, M. J., & Watson, J. M. (2006). Single versus dual process models of lexical decision performance: Insights from RT distributional analysis. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 1324–1344.

Received March 6, 2006
Revision received August 7, 2006
Accepted August 10, 2006 ■

ORDER FORM

Start my 2007 subscription to ***Journal of Experimental Psychology: Learning, Memory, and Cognition!*** ISSN: 0278-7393

___ $148.00, APA MEMBER/AFFILIATE ________
___ $316.00, INDIVIDUAL NONMEMBER ________
___ $825.00, INSTITUTION ________
In DC add 5.75% / In MD add 5% sales tax ________
TOTAL AMOUNT ENCLOSED $ ________

Subscription orders must be prepaid. (Subscriptions are on a calendar year basis only.) Allow 4-6 weeks for delivery of the first issue. Call for international subscription rates.

AMERICAN PSYCHOLOGICAL ASSOCIATION

SEND THIS ORDER FORM TO:
American Psychological Association
Subscriptions
750 First Street, NE
Washington, DC 20002-4242

Or call **800-374-2721**, fax **202-336-5568**.
TDD/TTY **202-336-6123**.
For subscription information, e-mail:
subscriptions@apa.org

Check enclosed (make payable to APA)

Charge my: VISA MasterCard American Express

Cardholder Name ________
Card No. ________ Exp. Date ________

____
Signature (Required for Charge)

BILLING ADDRESS:

Street ________
City ________ State ________ Zip ________
Daytime Phone ________
E-mail ________

MAIL TO:

Name ________
Address ________

____
City ________ State ________ Zip ________
APA Member # ________ *XLMA17*