

Single- Versus Dual-Process Models of Lexical Decision Performance: Insights From Response Time Distributional Analysis

Melvin J. Yap and David A. Balota
Washington University

Michael J. Cortese
College of Charleston

Jason M. Watson
University of Utah

This article evaluates 2 competing models that address the decision-making processes mediating word recognition and lexical decision performance: a hybrid 2-stage model of lexical decision performance and a random-walk model. In 2 experiments, nonword type and word frequency were manipulated across 2 contrasts (pseudohomophone–legal nonword and legal–illegal nonword). When nonwords became more wordlike (i.e., *BRNTA* vs. *BRANT* vs. *BRANE*), response latencies to nonwords were slowed and the word frequency effect increased. More important, distributional analyses revealed that the Nonword Type \times Word Frequency interaction was modulated by different components of the response time distribution, depending on the specific nonword contrast. A single-process random-walk model was able to account for this particular set of findings more successfully than the hybrid 2-stage model.

Keywords: lexical decision, nonword type, word frequency, visual word recognition

The study of the processes underlying isolated visual word recognition is a major endeavor in experimental psychology and has provided insights in domains as diverse as psycholinguistics, pattern recognition, computational modeling, attention, and neuroscience. Although many procedures have been developed for studying word recognition, the speeded lexical decision task (LDT; Rubenstein, Garfield, & Millikan, 1970) remains one of the most widely used tasks (e.g., Murray & Forster, 2004; Ratcliff, Gomez, & McKoon, 2004). In this task, participants are presented with a letter string and are required to decide as quickly as possible whether the string forms a word or nonword, most typically with a keypress response. Findings obtained in the LDT have been very

influential in informing models of word recognition (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001).

A number of models have been proposed to accommodate lexical decision performance (see Ratcliff et al., 2004, for a recent review). For example, the classic logogen model (Morton, 1969) posits word detectors (logogens) for every lexical entry. When a word (e.g., *DOG*) is presented, the logogen for *DOG* accumulates evidence until some threshold is reached and word identification takes place. The original logogen model could not handle nonwords, but more recent extensions to the model can carry out a lexical decision. For example, Grainger and Jacob's (1996) multiple read-out model (MROM) implements three processes that drive a lexical decision response. Word responses are produced when either 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. Nonword responses are produced when lexical activity does not reach threshold after some (variable) time deadline. The dual-route cascaded (DRC) model (Coltheart et al., 2001) adopts essentially the same principles to accommodate performance in the LDT. It is important to note, for these two examples, that the speed and accuracy of lexical decision responses are yoked to the activity of the word representations contained in the models' lexicons. Lexical decisions have also been instantiated in parallel distributed processing (PDP) reading models, which contain distributed orthographic and semantic representations (Plaut, 1997). In the latter framework, words or nonwords that are presented to the model generate varying degrees of stress values, which reflect how semantically familiar they are. A decision criterion can then be adopted that allows the model to discriminate between words (stress values higher than criterion) and nonwords (stress values lower than criterion). This approach assumes that lexical decision

Melvin J. Yap and David A. Balota, Department of Psychology, Washington University; Michael J. Cortese, Department of Psychology, College of Charleston; Jason M. Watson, Department of Psychology, University of Utah.

Michael J. Cortese is now at the Psychology Department, University of Nebraska at Omaha.

Melvin J. Yap and David A. Balota contributed equally to this project and share lead authorship. Portions of this research were presented at the 45th Annual Meeting of the Psychonomic Society, November 2004, Minneapolis, Minnesota. This research was supported by National Institute of Aging Grant AG03991 and National Science Foundation Grant BCS 0001801 to David A. Balota.

We thank Max Coltheart and Kathleen Rastle for their helpful comments on earlier versions of this article. We are also grateful to Roger Ratcliff and Jeff Rouder for assistance in various stages of this project. 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. E-mail: dbalota@artsci.wustl.edu

performance is driven by the activity of (distributed) representations.

In this article, we focus on two different approaches to lexical decision that emphasize the decision processes tied to lexical decision. Two models that specifically address the decision-making processes that mediate word identification and behavioral responses are the two-stage model of lexical decision performance (Balota & Chumbley, 1984; Balota & Spieler, 1999) and the diffusion model (Ratcliff et al., 2004). It is important to note that both frameworks have been used extensively in accommodating data in binary decision tasks, ranging from memory scanning (Atkinson & Juola, 1974) to episodic memory retrieval (Ratcliff, 1978). Moreover, both the two-stage model and the diffusion model have been shown to successfully accommodate basic lexical decision phenomena. It is interesting to note, though, that these two models are built on very different premises. The diffusion model assumes that a single process can drive lexical decision, whereas the two-stage model suggests that there are two qualitatively distinct processes. Whether lexical decision is better accommodated by a single- or a dual-process model is another instance of a broader distinction across a wide variety of domains (see, e.g., Yonelinas, 2002). Just as the debate between proponents of these two theoretical approaches has ramifications beyond psycholinguistics, the answer to the proposed question gives greater leverage in understanding how binary decisions are carried out in general. In this article, we explore this issue systematically. We begin by examining some interesting constraints in lexical decision performance, then describe the two classes of models, and finally quantitatively evaluate which approach accommodates the results from two experiments more successfully.

Interaction Between Nonword Type and Word Frequency

A critical variable that has been investigated in lexical decision performance is the similarity of the nonwords to real words. Nonwords can be pronounceable and orthographically legal (e.g., *FLIRP*), unpronounceable and orthographically illegal (e.g., *RPFLI*), or homophonous with real words (e.g., *BRANE*). As one might expect, nonword type powerfully modulates lexical decision latencies for word trials and also produces interactive effects with other variables that influence lexical decision performance. For example, the word-frequency effect (faster lexical decision latencies for frequently encountered words) is strongly modulated by the type of nonword context.

Stone and Van Orden (1993) systematically manipulated nonword type and word frequency in lexical decision and observed the pattern presented in Table 1. As nonwords become more similar to words, two trends are apparent. Lexical decision word latencies become slower, and, more intriguingly, the word-frequency effect becomes larger. Stone and Van Orden interpreted these results as consistent with both a pathway selection framework and a random-walk framework. The pathway selection framework proposes that the lexical processing system consists of independent processing modules that are interconnected by pathways. Manipulating the nature of the nonwords alters the task context, and the system strategically selects the pathways that optimize task performance. More relevantly for this article, the results were also accommodated within a random-walk framework. The random-walk model has been useful for describing various aspects of binary decisions

Table 1
Means of Participants' Mean Lexical Decision Response Times (in Milliseconds) as a Function of Word Frequency and Nonword Type

Word frequency	Illegal nonwords (<i>RPFLI</i>)	Legal nonwords (<i>FLIRP</i>)	Pseudohomophones (<i>BRANE</i>)
Low	578	697	867
High	542	621	707
Effect	36	76	160

Note. Data reproduced from "Strategic Control of Processing in Word Recognition," by G. O. Stone and G. C. Van Orden, 1993, *Journal of Experimental Psychology: Human Perception and Performance*, 19, p. 753. Copyright 1993 by the American Psychological Association.

(Ratcliff & Rouder, 1998) and is a member of a more general class of sequential sampling models. The random-walk perspective conceptualizes lexical decision as an evidence-accumulating process. When a stimulus is presented, noisy information is accumulated over time toward one of two possible decision boundaries, word or nonword in the case of LDT (see Figure 1). A word response is produced when the accumulation process reaches the word boundary; a nonword response is produced when the accumulation process reaches the nonword boundary. For the simplest random-walk model, two parameters are of interest: the signal strength and the response criterion. *Signal strength* refers to the rate of evidence accumulation and is greater for stimuli that are processed more efficiently (e.g., high-frequency words). The *response criterion* refers to the distance of the boundaries from the start point; increasing the response boundaries reflects more conservative response criteria.

Using this simple random-walk model, Stone and Van Orden (1993) argued that there is a linearly decreasing concave function between signal strength and the amount of time needed to reach criterion (see Figure 2), that is, the same change in signal strength has a greater impact on response times (RTs) when signal strength is lower compared with when signal strength is higher. The interaction between nonword type and word frequency is predicted by this function. Low-frequency words have lower signal strengths than high-frequency words. When nonwords become more word-like (e.g., from *BRONE* to *BRANE*), word-nonword discrimination becomes more difficult. The signal strengths of both low- and high-frequency words decrease, leading to longer decision times. It is important to note that because of the concave function, word-frequency effects are larger in the pseudohomophone condition than in the legal nonword condition, mimicking the pattern presented in Table 1. Putatively, this account also explains why word-frequency effects are larger in the legal nonword condition than in the illegal nonword condition. It is important to note, though, that Stone and Van Orden's data were examined at the level of the mean and there was no explicit implementation of this model. Hence, it was a descriptive account of the pattern observed in the means. As shown below, analyzing the same data at the level of distributional characteristics may yield further insights that are neither apparent nor intuitive.

An alternative account of the Nonword Type \times Word Frequency interaction is provided by Balota and Chumbley's (1984) two-process model, which is based on Atkinson and Juola's (1974)

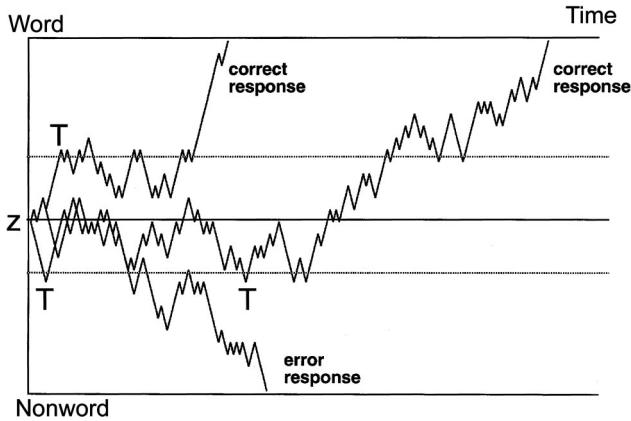


Figure 1. Illustration of the diffusion model. The different slopes reflect a diffusion process, which is the continuous version of the random-walk model. Z refers to the starting point, and the point at which the slope intersects with the word or nonword boundary is the time taken for a decision (word or nonword) to be made. T = the point at which processing terminates. From "Modeling Response Times for Two-Choice Decisions," by R. Ratcliff and J. F. Rouder, 1998, *Psychological Science*, 9, p. 348. Copyright 1998 by the American Psychological Society. Adapted with permission.

two-stage model of memory search. The application of this framework to lexical decision performance is displayed in Figure 3. This model was originally advanced as an account of task-specific effects in lexical decision (Balota & Chumbley, 1984) and was

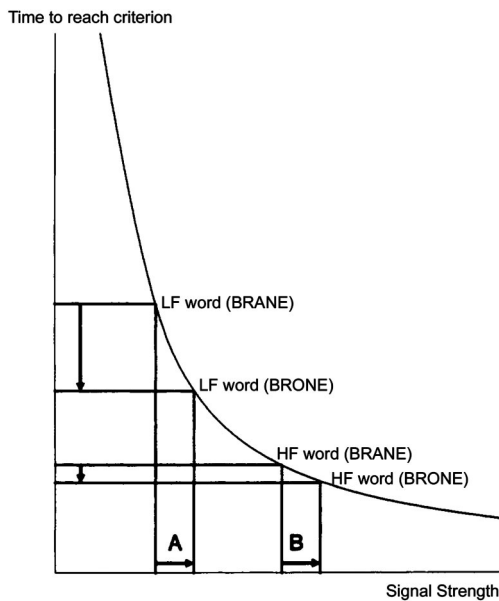


Figure 2. The concave function between signal strength and finishing times in the random-walk model. LF = low frequency; HF = high frequency. From "Strategic Control of Processing in Word Recognition," by G. O. Stone and G. C. Van Orden, 1993, *Journal of Experimental Psychology: Human Perception and Performance*, 19, p. 765. Copyright 1993 by the American Psychological Association. Adapted with permission.

motivated by the observation that frequency effects are different in size across lexical decision, naming, and category classification, three tasks that presumably tap the same word identification process. Balota and Chumbley found that frequency effects were largest in lexical decision and argued that this pattern was likely due to the fact that the frequency effect reflects both word identification processes and the word–nonword discrimination process that is specific to that task.

Balota and Chumbley (1984) suggested that words and nonwords could be conceived as reflecting two underlying distributions that vary along a familiarity–meaningfulness (FM) dimension. Participants can use two types of information to make lexical decisions. The first is a relatively fast-acting familiarity-based signal and the second is a slower, more attention-demanding signal, which may involve explicitly checking the spelling of the stimulus. Low-frequency words are particularly sensitive to variables that modulate the checking process because low-frequency words are more likely to overlap with the nonwords on the FM dimension. Hence, as one increases the overlap between the two distributions by making the nonwords more similar to the words, this further increases the checking process for the low-frequency words, thereby slowing these items. Therefore, greater checking will occur for low-frequency words when these items are embedded in lists with pseudohomophones, compared with legal nonwords. Moreover, the smallest amount of checking will occur for low-frequency words when these items are embedded in lists with illegal nonwords because the nonword distribution will overlap very little with the word distribution. Hence, the two-stage model also accommodates Stone and Van Orden's (1993) Nonword Type \times Word Frequency interaction by assuming two distinct processes instead of a single random-walk process. In addition, the framework was able to qualitatively account for other lexical

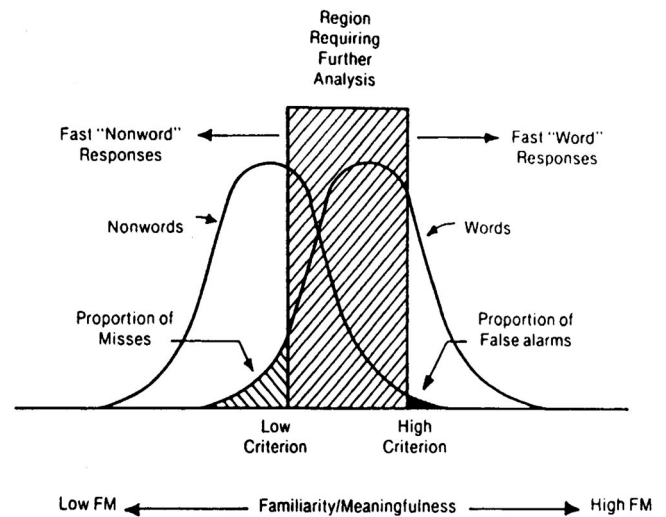


Figure 3. The two-stage model of lexical decision performance. FM = familiarity–meaningfulness. From "Are Lexical Decisions a Good Measure of Lexical Access? The Role of Word Frequency in the Neglected Decision Stage," by D. A. Balota and J. I. Chumbley, 1984, *Journal of Experimental Psychology: Human Perception and Performance*, 10, p. 352. Copyright 1984 by the American Psychological Association. Reprinted with permission.

decision effects (e.g., blocking effects and repetition effects) that were troublesome for certain extant word recognition models.

Beyond Measures of Central Tendency

In standard chronometric studies, a set of RTs for a particular experimental condition is collected for each participant. Typically, the mean of those response times (M_{RT}) is then computed, with M_{RT} providing an estimate of the central tendency for that condition. Of course, it is possible that variables do not simply shift the RT distribution, as implicitly assumed by analyses based on means; variables may also change the shape of the distribution. Hence, when possible, it is also useful to investigate the influence of a variable on the shape (e.g., variance, skew) of a distribution (Andrews & Heathcote, 2001; Heathcote, Popiel, & Mewhort, 1991; Yap & Balota, in press). For example, fitting the ex-Gaussian function to data (Hohle, 1965; Luce, 1986; Ratcliff, 1979) allows researchers to estimate how different variables shift, skew, or shift and skew RT distributions. An ex-Gaussian analysis characterizes an RT distribution by assuming an explicit model for the shape of the distribution. This model is a convolution of the normal (Gaussian) and exponential distributions and has 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. In addition to providing unusually good fits to positively skewed empirical RT distributions (Luce, 1986, p. 439), one useful consequence of ex-Gaussian analysis is that the algebraic sum of μ and τ is approximately equivalent to the mean when one estimates parameters from empirical data (μ and τ are exactly equal to the mean in the theoretical ex-Gaussian model). Briefly, this property allows differences in means to be conveniently partitioned into two components: a component that is associated with distributional shifting (μ) and a component that is associated with distributional skewing (τ). There are at least two other reasons why such a distributional analysis might be valuable.

First, Heathcote et al. (1991) pointed out that analyzing mean RTs can often be inadequate and misleading because such an analysis does not consider the shape of the RT distributions. For example, they examined Stroop color-naming performance with both traditional and ex-Gaussian analyses. On the basis of mean response latencies, there was no difference between the congruent (RED displayed in red) and baseline (XXX displayed in red) conditions. This suggests that congruency has no effect on color naming, relative to the baseline. However, 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 instance, congruency shifted the RT distribution leftward while increasing its skew. These two effects cancelled each other out, spuriously producing null effects of congruency (see Spieler, Balota, & Faust, 1996, for a replication of this trade-off).

Second, by exploiting more of the information available in a RT distribution, one can make increasingly sophisticated predictions about how a variable might modulate the shape of a distribution, rather than just asking whether a variable has an effect in mean RTs. This is useful when one is trying to adjudicate between two models. Models may be indistinguishable at the level of the mean

but make different predictions at the level of the RT distribution (see Hockley, 1984; Mewhort, Braun, & Heathcote, 1992).

The two experiments reported in this article represent an extension of Stone and Van Orden's (1993) Experiment 1, with nonword type (legal nonwords, illegal nonwords, and pseudohomophones) and word frequency (high and low) factorially manipulated in a lexical decision task. To obtain a sufficient number of observations to provide adequate estimates of RT distributional characteristics, we collected 100 observations for each of the cells for each participant. Experiment 1 examined the contrast between legal nonwords (e.g., *FLIRP*) and pseudohomophones (e.g., *BRANE*), whereas Experiment 2 examined the contrast between illegal nonwords (e.g., *RPFLI*) and legal nonwords. It is important to note that we examine the joint effects of the two variables on RT distributional properties, using both ex-Gaussian analysis and a nonparametric technique called *vincentizing*, described in the *Results* section of Experiment 1. Following the empirical section, we describe and implement the two modeling frameworks and then test which framework better accommodates the observed effects, both at the level of the mean and at the level of distributional characteristics.

Experiment 1

Method

Participants. A total of 75 young adults (mean age = 19.2 years, $SD = 1.65$) participated in this experiment for course credit. All participants had normal or corrected-to-normal vision and were recruited from the undergraduate student population of Washington University. The participants had an average of 13.7 years of education ($SD = 1.17$) and a mean vocabulary age of 18.2 years ($SD = 0.88$) on the Shipley Vocabulary subtest (Shipley, 1940).

Data from 7 of the 75 participants were discarded because of excessively high error rates and/or slow latencies via the following procedure. To identify outlier participants,¹ we combined each participant's response latencies and error rates into a vector of four scores (mean RTs for high- and low-frequency words; error rate for high- and low-frequency words), and then computed the Mahalanobis D^2 metric (Lattin, Carroll, & Green, 2003) 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., $ps < .05$) 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. In total, there were 35 participants in the legal nonword condition and 33 participants in the pseudohomophone (PsH) condition.

Apparatus. An IBM-compatible computer controlled stimulus presentation and collected response latencies, via the keyboard, to the nearest millisecond. The stimuli were displayed on a 17-in. (43.18-cm) Super VGA monitor.

Stimuli. The stimuli consisted of 200 words, 200 length-matched pronounceable nonwords, and 200 length-matched PsHs. Using the HAL (hyperspace analogue to language) frequency norms (Lund & Burgess, 1996), 100 words were designated high frequency (median log counts per 131 million = 11.09) and 100 words were designated low frequency

¹ To determine whether participant and RT screening procedures were influencing the results, we reanalyzed the data using all participants and less conservative screening criteria, that is, removing only latencies faster than 200 ms and slower than 3,000 ms. The pattern of results did not change.

(median log counts per 131 million = 7.76). Nonwords were constructed by changing one to three letters of the word items. Words and nonwords ranged from three to seven letters in length. For high-frequency words, 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 orthographic neighborhood size was 4.82, and the mean summed bigram frequency was 6,149.13. There was no significant difference between high- and low-frequency words with respect to both orthographic neighborhood, $t(198) = -0.08$, $p = .94$, and summed bigram frequency, $t(198) = 0.42$, $p = .67$. For the nonwords, the mean orthographic neighborhood size was higher for PsHs (mean $N = 4.41$) than for legal nonwords (mean $N = 3.38$; $p = .02$).² The mean summed bigram frequency was higher for PsHs (6,663.5) than for legal nonwords (5,984.7).

Procedure. Participants were tested individually in sound-attenuated cubicles. They were seated about 60 cm from the computer screen. Before the experimental trials began, participants completed a computer-administered Shipley Vocabulary subtest (Shipley, 1940).

For the LDT, participants were told that letter strings would be presented in the center of the screen and that their task was to indicate as quickly and as accurately as possible via a buttonpress on the keyboard whether the letter string was a word or nonword. Participants then received 20 practice trials and four experimental blocks of 100 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 250 ms, and (c) a letter string centered at the fixation point's location. The letter string remained on the screen until a response was made. Participants pressed the *slash* key for words and the *Z* key for nonwords. Responses were followed by a 1,500-ms delay. If the response was incorrect, 750 ms of that 1,500 ms was consumed by a 200-Hz tone and an *Incorrect Response* message was displayed.

Design. A 2×2 factorial design was used: Nonword type (PsH or legal) was manipulated between participants, whereas word frequency (high or low) was manipulated within participant.

Results

Errors (5.9% across both conditions) and response latencies faster than 200 ms or slower than 3,000 ms were first excluded

Table 2
Means of Participants' Mean Lexical Decision Response Times (in Milliseconds), Accuracy, and Ex-Gaussian Parameter Estimates as a Function of Word Frequency and Nonword Type (BRONE vs. BRANE) in Experiment 1

Nonword type and word frequency	<i>M</i>	% error	μ	σ	τ
Words					
Legal					
High frequency	564	1.7	459	40	105
Low frequency	635	7.0	504	58	131
Frequency effect	71	5.3	45	18	26
Pseudohomophone					
High frequency	619	2.1	483	45	136
Low frequency	722	10.4	525	61	197
Frequency effect	102	8.3	42	16	61
Difference of difference (interaction)	31	3.0	-3	-2	35
Nonwords					
Legal	679	4.4	551	59	127
Pseudohomophone	783	8.5	588	65	195
Nonword type effect	104	4.1	37	6	68

from the analyses. On the basis of the remaining observations, the overall mean and standard deviation of each participant's word and nonword latencies were computed. Response latencies 2.5 standard deviations above or below each participant's respective mean latency were removed. These criteria eliminated a further 5.0% of the lexical decision responses. Analyses of variance (ANOVAs) were then carried out on the mean response latencies and accuracies (both by participants and items) and also the ex-Gaussian parameters of the word and nonword RT data (see Table 2).

Word response latencies. For mean response latencies, the main effect of nonword type, $F_p(1, 66) = 10.42$, $p = .002$, $MSE = 16,467.33$, $\eta^2 = .14$; $F_i(1, 198) = 162.27$, $p < .001$, $MSE = 3,641.59$, $\eta^2 = .45$, was significant, as was the main effect of word frequency, $F_p(1, 66) = 433.09$, $p < .001$, $MSE = 594.80$, $\eta^2 = .87$; $F_i(1, 198) = 248.05$, $p < .001$, $MSE = 3,525.47$, $\eta^2 = .56$. More important, the Nonword Type \times Word Frequency interaction was also significant, $F_p(1, 66) = 13.91$, $p < .001$, $MSE = 594.80$, $\eta^2 = .17$; $F_i(1, 198) = 11.93$, $p = .001$, $MSE = 3,641.59$, $\eta^2 = .06$. As shown in Table 2, there were larger frequency effects in the presence of PsHs compared with legal nonwords.

Percentage correct. Turning to the accuracy data, the main effect of nonword type, $F_p(1, 66) = 8.51$, $p = .005$, $MSE = 0.0014$, $\eta^2 = .11$; $F_i(1, 198) = 3.35$, $p = .07$, $MSE = 0.0079$, $\eta^2 = .017$, was significant, as was the main effect of word frequency, $F_p(1, 66) = 184.08$, $p < .001$, $MSE = 0.0009$, $\eta^2 = .74$; $F_i(1, 198) = 112.06$, $p < .001$, $MSE = 0.0061$, $\eta^2 = .36$. The Nonword Type \times Word Frequency interaction was significant, $F_p(1, 66) = 8.99$, $p = .004$, $MSE = 0.0009$, $\eta^2 = .12$; $F_i(1, 198) = 3.43$, $p = .07$, $MSE = 0.0079$, $\eta^2 = .017$, with larger frequency effects observed in the PsH condition than in the legal nonword condition.

Ex-Gaussian analyses. Ex-Gaussian parameters (μ , σ , τ) were obtained for each participant using continuous maximum likelihood estimation in R (R Development Core Team, 2004). Continuous maximum likelihood estimation provides relatively efficient and unbiased parameter estimates (Van Zandt, 2000) and uses all of the available raw data (see Heathcote, Brown, & Mewhort, 2002, for an alternative approach). Using 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.

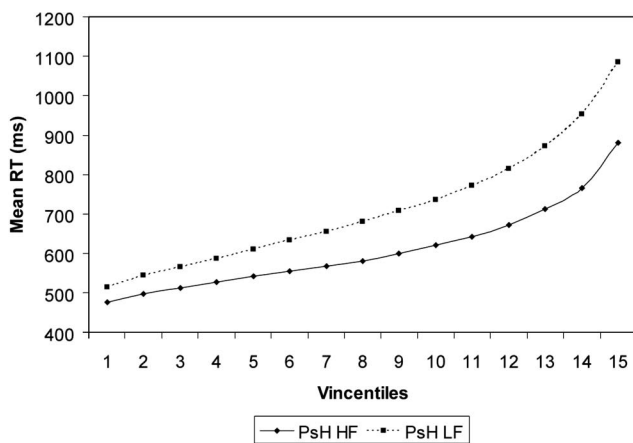
For μ , the main effect of nonword type approached significance, $F(1, 66) = 3.46$, $p = .07$, $MSE = 5,061.67$, $\eta^2 = .05$, whereas the word-frequency effect was significant, $F(1, 66) = 180.59$, $p < .001$, $MSE = 356.62$, $\eta^2 = .73$. The Nonword Type \times Word Frequency interaction was not significant ($F < 1$). Turning to σ , the main effect of word frequency was significant, $F(1, 66) =$

² The fact that the PsHs and legal nonwords were not matched on orthographic neighborhood size suggests that any observed nonword type effects cannot be unambiguously attributed to pseudohomophony; it is quite possible that orthographic neighborhood is also making a contribution. However, in Experiment 1, we are not claiming (and do not plan to claim) that pseudohomophony per se is responsible for the observed effects. Rather, we are primarily interested in manipulating the familiarity of the nonword context and examining how foils that are more similar to targets (i.e., PsHs) influence the responses to words. Clearly, familiarity in this instance is a multidimensional quantity that could encompass pseudohomophony and/or orthographic neighborhood size.

47.61, $p < .001$, $MSE = 207.26$, $\eta^2 = .42$. Neither the main effect of nonword type nor the interaction was significant ($F_s < 1$). For τ , the main effects of nonword type, $F(1, 66) = 11.72$, $p = .001$, $MSE = 6,776.14$, $\eta^2 = .15$, and word frequency, $F(1, 66) = 86.66$, $p < .001$, $MSE = 743.06$, $\eta^2 = .57$, were significant. The Nonword Type \times Word Frequency interaction was highly significant, $F(1, 66) = 13.90$, $p < .001$, $MSE = 743.06$, $\eta^2 = .17$, with larger frequency effects in the PsH condition. Thus, the Nonword Type \times Word Frequency interaction was localized in the τ parameter.

Vincentile analysis. 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; Ratcliff, 1979; Rouder & Speckman, 2004; Vincent, 1912; Yap & Balota, in press) to produce the RT distribution for a typical participant (see Figure 4 for an example). This approach does not rely on any 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 is first sorted (from fastest to slowest responses), and the first 10% of the data are then averaged, followed by the second 10%, and so on. Individual vincentiles are then averaged across participants and plotted. Plots of mean vincentiles are useful for investigating how different variables influence different regions of the RT distribution and provide a graphical, complementary perspective to ex-Gaussian analysis. For example, μ effects are reflected in additive changes in the vincentiles along the y-axis, whereas τ effects are reflected in the slowest (rightmost) vincentiles.

The mean vincentiles for the different experimental conditions are plotted in Figure 4. As shown here, the frequency effect increased systematically across vincentiles for both nonword type conditions. However, if we consider the later, slower vincentiles, the frequency effect for the PsH nonword condition was markedly larger than for the legal nonword condition.



Nonword response latencies. For mean response latencies, the main effect of nonword type was significant, $F_p(1, 66) = 12.39$, $p = .001$, $MSE = 14,803.50$, $\eta^2 = .16$; $F_i(1, 199) = 327.44$, $p < .001$, $MSE = 3,619.12$, $\eta^2 = .62$, with slower nonword responses to PsHs than to legal nonwords. For accuracy, the main effect of nonword type was also significant, $F_p(1, 66) = 11.81$, $p = .001$, $MSE = 0.0023$, $\eta^2 = .15$; $F_i(1, 199) = 13.17$, $p < .001$, $MSE = 0.0083$, $\eta^2 = .062$, with higher error rates for PsHs than for legal nonwords.

Turning to the ex-Gaussian parameters, for μ , the main effect of nonword type approached significance, $F(1, 66) = 3.67$, $p = .060$, $MSE = 6,187.09$, $\eta^2 = .053$; μ was larger in the PsH condition. For σ , the main effect of nonword type was not significant. Turning to τ , the main effect of nonword type was significant, $F(1, 66) = 14.56$, $p < .001$, $MSE = 5,293.59$, $\eta^2 = .18$; τ was larger in the PsH condition than in the legal nonword condition. The mean vincentiles for nonwords in the different experimental conditions are plotted in Figure 5. An effect of nonword type (i.e., faster latencies for legal nonwords compared with PsHs) is apparent across the vincentiles, although the difference between legal nonwords and PsHs becomes especially pronounced at the slowest vincentiles.

Discussion

Experiment 1 replicated and extended Stone and Van Orden's (1993) Nonword Type \times Word Frequency interaction; larger word-frequency effects are observed when PsHs are used. More intriguingly, the ex-Gaussian analyses and vincentile plots indicated that this interaction was localized in the exponential component (τ) of the distribution (see Table 2), suggesting that distributional skewing was responsible for producing the Word Frequency \times Nonword Type interaction in Experiment 1. For nonwords, the nonword type effect was mediated by both μ (to a lesser extent) and τ (to a greater extent), suggesting that both distributional shifting and, to a greater extent, skewing underlie the slower nonword responses to PsHs.

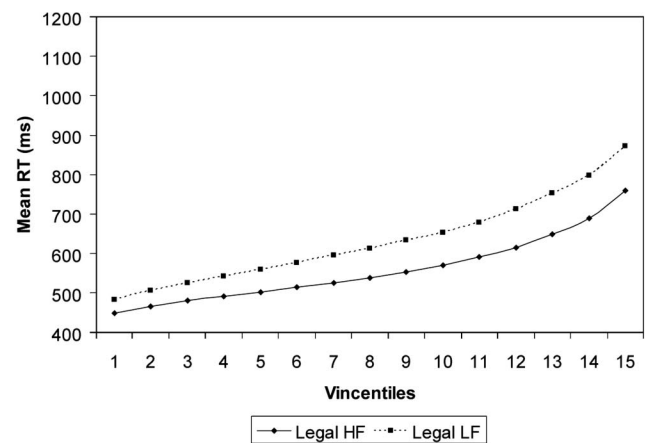


Figure 4. Vincentile means of the participants' word lexical decision response times (RTs; in milliseconds) as a function of nonword type (legal nonword vs. PsH) and word frequency. PsH = pseudohomophone; HF = high frequency; LF = low frequency.

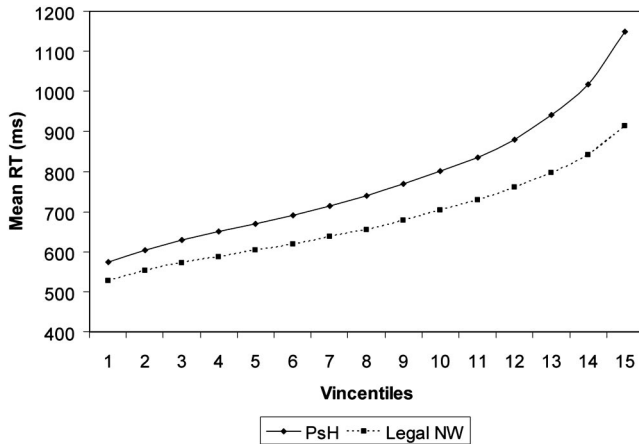


Figure 5. Vincintile means of the participants' nonword lexical decision response times (RTs; in milliseconds) as a function of nonword type (legal NW vs. PsH). PsH = pseudohomophone; NW = nonword.

Experiment 2

In the following experiment, we explored a different nonword contrast. In particular, we compared the word-frequency effect in the context of illegal nonwords versus the context of legal nonwords.

Method

Participants. A total of 77 young adults (mean age = 19.1 years, $SD = 1.20$) participated in this experiment for course credit. All participants had normal or corrected-to-normal vision and were recruited from the undergraduate student population of Washington University. The participants had an average of 12.9 years of education ($SD = 0.96$) and a mean vocabulary age of 18.5 years ($SD = 0.90$) on the Shipley Vocabulary subtest. Data from 3 of the 77 participants were discarded using the same multivariate outlier procedure described in Experiment 1. In total, there were 37 participants in the legal nonword condition and 37 participants in the illegal nonword condition.

Apparatus. An IBM-compatible computer running E-prime software (Schneider, Eschman, & Zuccolotto, 2001) was used to control stimulus presentation and to collect data. The stimuli were displayed on a 17-in. (43.18-cm) Super VGA monitor, and participants' responses were made on a computer keyboard.

Stimuli. The stimuli for the LDT consisted of the 200 words and 200 length-matched pronounceable nonwords used in Experiment 1, as well as another 200 orthographically illegal nonwords created by permuting the letters in the pronounceable nonwords. The mean orthographic neighborhood size was higher for legal nonwords (mean $N = 3.38$) than for illegal nonwords (mean $N = 0.15$; $p < .001$).³ The mean summed bigram frequency was higher for legal nonwords (5,984.7) than for illegal nonwords (3,346.8).

Procedure. The testing conditions were substantially the same as in Experiment 1. 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 for each participant. Stimuli were presented in 14-point Courier font. 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 200 ms, and (c) a stimulus centered at the fixation point's location. The stimulus remained on screen until a keyboard response was made. Participants pressed the *apostrophe* key for words and the *A* key for nonwords.

Responses were followed by a 1,600-ms delay. If the response was incorrect, 450 ms of that 1,600 ms was consumed by a 170-ms tone that was presented simultaneously with the word *Incorrect* displayed slightly below the fixation point.

Design. A 2×2 factorial design was used: Nonword type (legal or illegal) was manipulated between participants, whereas word frequency (high or low) was manipulated within participant.

Results

Errors (3.5% across both conditions) and 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 and nonword 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 2.5% of the lexical decision responses. ANOVAs were then carried out on the mean, accuracy, and ex-Gaussian parameters of the word and nonword response time data (see Table 3).

Response latencies. For the mean response latency data, the main effect of nonword type, $F_p(1, 72) = 21.28$, $p < .001$, $MSE = 13,013.35$, $\eta^2 = .23$; $F_i(1, 198) = 963.37$, $p < .001$, $MSE = 805.35$, $\eta^2 = .83$, was significant, as was the main effect of word frequency, $F_p(1, 72) = 115.62$, $p < .001$, $MSE = 364.40$, $\eta^2 = .62$; $F_i(1, 198) = 75.66$, $p < .001$, $MSE = 1,845.56$, $\eta^2 = .28$. More important, the Nonword Type \times Word Frequency interaction was also significant, $F_p(1, 72) = 26.02$, $p < .001$, $MSE = 364.40$, $\eta^2 = .27$; $F_i(1, 198) = 38.07$, $p < .001$, $MSE = 805.35$, $\eta^2 = .16$, with smaller frequency effects in the presence of illegal nonwords compared with legal nonwords.

Percentage correct. Turning to the analysis of the accuracy data, the main effect of nonword type, $F_p(1, 72) = 18.20$, $p < .001$, $MSE = 0.0017$, $\eta^2 = .20$; $F_i(1, 198) = 27.51$, $p < .001$, $MSE = 0.0028$, $\eta^2 = .12$, was significant, as was the main effect of word frequency, $F_p(1, 72) = 59.99$, $p < .001$, $MSE = 0.0009$, $\eta^2 = .45$; $F_i(1, 198) = 48.35$, $p < .001$, $MSE = 0.0055$, $\eta^2 = .20$. The Nonword Type \times Word Frequency interaction was also significant, $F_p(1, 72) = 14.67$, $p < .001$, $MSE = 0.0009$, $\eta^2 = .17$; $F_i(1, 198) = 13.82$, $p < .001$, $MSE = 0.0028$, $\eta^2 = .065$, with a smaller frequency effect observed in the illegal nonword condition compared with the legal nonword condition.

Ex-Gaussian analyses. The analysis for μ yielded main effects of nonword type, $F(1, 72) = 18.26$, $p < .001$, $MSE = 4,240.16$, $\eta^2 = .20$, and word frequency, $F(1, 72) = 111.83$, $p < .001$, $MSE = 216.47$, $\eta^2 = .61$. In contrast to Experiment 1, the Nonword Type \times Word Frequency interaction was significant for μ , $F(1, 72) = 22.47$, $p < .001$, $MSE = 216.47$, $\eta^2 = .24$, with smaller frequency effects in the illegal nonword condition. Turning to σ , the main effect of word frequency was significant, $F(1, 72) = 29.71$, $p < .001$, $MSE = 108.82$, $\eta^2 = .29$. Neither the main effect

³ In not matching legal and illegal nonwords on orthographic neighborhood, nonword type effects may be driven by both orthographic legality as well as orthographic neighborhood. Again, we do not intend to claim that the observed effects are driven solely by orthographic legality. Our variable of primary interest is the familiarity of the nonword context, and we think it is plausible that illegal nonwords are unfamiliar because they are both orthographically illegal and have few or no neighbors.

Table 3
Means of Participants' Mean Lexical Decision Response Times (in Milliseconds), Accuracy, and Ex-Gaussian Parameter Estimates as a Function of Word Frequency and Nonword Type (NBREO vs. BRONE) in Experiment 2

Nonword type and word frequency	M	% error	μ	σ	τ
Words					
Illegal					
High frequency	499	1.4	421	39	78
Low frequency	517	3.2	435	48	82
Frequency effect	18	1.8	14	9	4
Legal					
High frequency	570	2.4	456	37	114
Low frequency	620	8.0	493	47	127
Frequency effect	50	5.6	37	10	13
Difference of difference (interaction)	32	3.8	23		
Nonwords					
Illegal	527	2.0	448	46	79
Legal	657	4.7	526	49	132
Nonword type effect	130	2.7	78	3	53

of nonword type nor the interaction was significant ($F_s < 1$). For τ , the main effects of nonword type, $F(1, 72) = 12.41, p = .001, MSE = 4,968.36, \eta^2 = .15$, and word frequency, $F(1, 72) = 5.42, p = .023, MSE = 449.68, \eta^2 = .07$, were significant. In contrast to Experiment 1, the Nonword Type \times Word Frequency interaction was not significant ($p = .19$) in τ .

Vincentile analysis. The mean vincentiles for the different experimental conditions are plotted in Figure 6. As shown in the figure, there appears to be a larger frequency effect in the legal nonword condition compared with the illegal nonword condition that extends across all vincentiles. This is consistent with the interactive effects being located primarily in the μ component of the ex-Gaussian analyses.

Nonword response latencies. For mean response latencies, the main effect of nonword type was significant, $F_p(1, 72) = 41.54, p < .001, MSE = 7,525.80, \eta^2 = .37; F_t(1, 199) = 1232.06, p < .001, MSE = 1,442.52, \eta^2 = .86$, with faster nonword responses

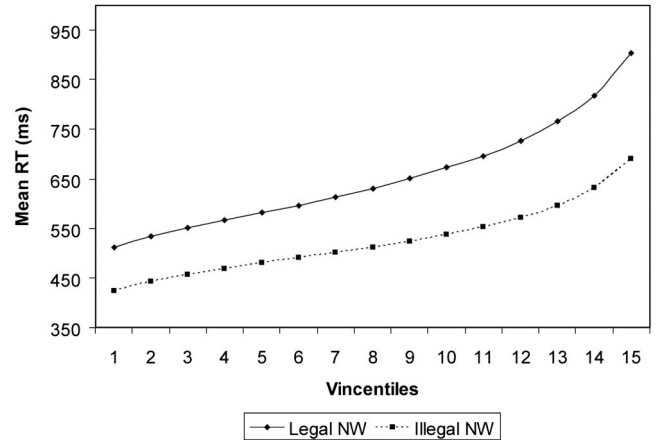
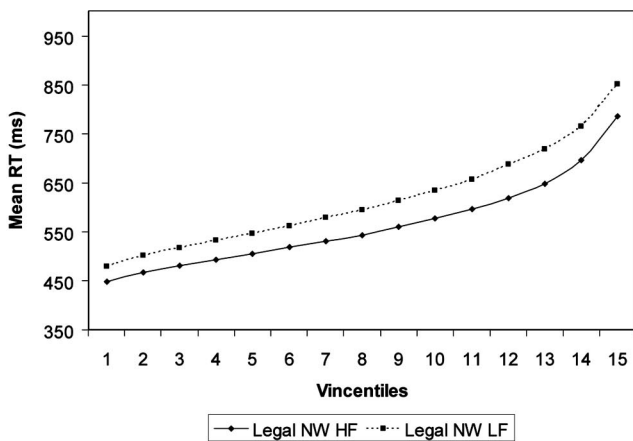


Figure 7. Vincentile means of the participants' nonword lexical decision response times (RTs; in milliseconds) as a function of nonword type (legal NW vs. illegal NW). NW = nonword.

when illegal nonwords were used. For accuracy, the main effect of nonword type was significant, $F_p(1, 72) = 16.28, p < .001, MSE = 0.0009, \eta^2 = .18; F_t(1, 199) = 13.14, p < .001, MSE = 0.0055, \eta^2 = .062$, with lower error rates when illegal nonwords were used.

Turning to the ex-Gaussian parameters, for μ , the main effect of nonword type was significant, $F(1, 72) = 45.20, p < .001, MSE = 2,447.89, \eta^2 = .39$; μ was smaller in the illegal nonword condition. For σ , the main effect of nonword type was not significant. For τ , the main effect of nonword type was significant, $F(1, 72) = 16.83, p < .001, MSE = 3,047.56, \eta^2 = .19$; τ was smaller in the illegal nonword condition. The mean vincentiles for nonwords in the different experimental conditions are plotted in Figure 7.

Discussion

The present results again indicated that as nonwords became more wordlike, frequency effects became larger. More important,

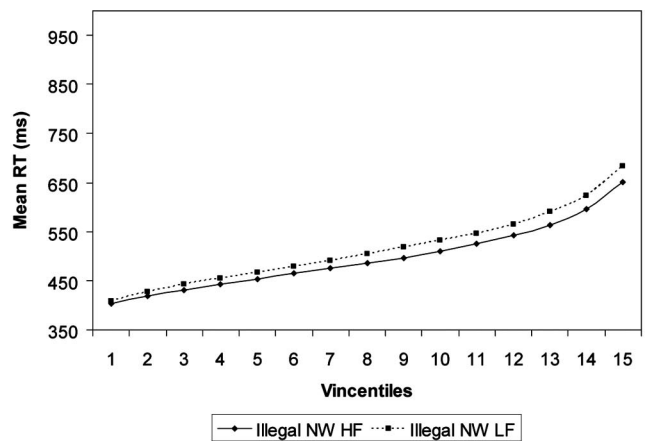


Figure 6. Vincentile means of the participants' word lexical decision response times (RTs; in milliseconds) as a function of nonword type (legal NW vs. illegal NW) and word frequency. NW = nonword; HF = high frequency; LF = low frequency.

the distributional analyses yielded the counterintuitive finding that the distributional characteristics of the Nonword Type \times Word Frequency interaction were modulated by a different component in this experiment compared with Experiment 1. In Experiment 1, word-frequency effects were larger in the presence of PsHs compared with legal nonwords, and this interaction was mediated by the τ component. For the nonword foils, nonword type effects (slower latencies for PsHs) were also mediated more strongly by the τ component (see Table 2). In Experiment 2, word-frequency effects were larger for the legal nonwords compared with the illegal nonwords, and the interaction in this case was mediated primarily by μ and, to a lesser extent, by τ . For the nonword foils, nonword type effects (slower latencies for legal nonwords) were mediated more strongly by the μ component (see Table 3). This intriguing pattern in the underlying distributional characteristics has never been reported and, as described below, provides a useful constraint for discriminating between single- and two-stage models of lexical decision performance. The current modeling endeavor is based on correct RTs and does not address error rates or error latencies. In any case, modeling errors is complicated by two factors: low error rates and the multiple factors that contribute to such errors (Balota & Spieler, 1999).

Two-Stage Hybrid Model of Lexical Decision Performance

Recently, Balota and Spieler (1999) extended the two-stage model, described earlier, to accommodate the effects of variables on the shape of RT distributions. So as to understand how the model captures characteristics of the response latency distributions, we briefly describe the results of Balota and Spieler's study, which manipulated frequency, lexicality, and repetition (repeated or nonrepeated). Using Figure 2, the predictions for the following effects are considered in turn: the main effect of frequency, the Frequency \times Repetition interaction, and the Lexicality \times Repetition interaction.

First, consider the main effect of frequency. Low-frequency words are more likely than high-frequency words to fall under the region requiring further analysis and hence are more likely to engage the second slow and attention-demanding process, resulting in an RT distribution that is more positively skewed. In ex-Gaussian terms, this increased skewing is indexed by a larger exponential component (τ). Second, turning to the Frequency \times Repetition interaction, repeating low-frequency words should increase their familiarity and push these items above the high criterion, making these items less likely to undergo the analytic check process. The distribution of repeated low-frequency words should therefore produce a smaller τ than nonrepeated low-frequency words. Repeating high-frequency words should modulate τ less because more high-frequency words are already above the high criterion. Finally, regarding the Lexicality \times Repetition interaction, repeated words should produce a decrease in τ because these items are pushed above the high criterion and hence increase the likelihood of the check process. In contrast, repeated nonwords should produce an increase in τ because these items are pushed above the lower criterion, subjecting them to more analytic checking.

The data supported the qualitative predictions made by the descriptive two-stage model. Balota and Spieler (1999) attempted

to implement a quantitative version of the model. After a few iterations, they settled on the hybrid model. This model assumes that Stage 1 (familiarity) item responses are Gaussian in shape, whereas Stage 1 + Stage 2 (familiarity + check) item responses are ex-Gaussian in shape. In Stage 1, the familiarity of items modulates the time to respond in a Gaussian fashion. Items of intermediate familiarity enter the Stage 2 check process and generate response times from an ex-Gaussian distribution. In this article, we test a two-stage model that is identical to the hybrid model described by Balota and Spieler to account for the Frequency \times Repetition and Lexicality \times Repetition interactions.

Random-Walk Model of Lexical Decision Performance

As discussed in the introduction, the random-walk model conceptualizes lexical decision as an evidence-accumulating process. At each time point, a unit of evidence is sampled, and this evidence is consistent with either a word or nonword response. Over time, evidence incrementally accrues for both responses but at a greater rate for the more probable response. When a criterion is reached, a word or nonword response is emitted. Random-walk finishing times plus some residual time for encoding and response are assumed to correspond to behavioral lexical decision times. The random-walk model instantiated in this article is conceptually similar to Stone and Van Orden's (1993) canonical model (see also Gordon's, 1983, resonance model) and is in fact identical to the model explored by Spieler, Balota, and Faust (2000) to accommodate attentional selection performance. The random-walk model is superficially similar to the counter model (Pike, 1973), which also tallies evidence for two different responses in separate counters, making a response when either tally exceeds an absolute criterion (Vickers, 1979). However, the random-walk model can be distinguished by its adoption of a relative rather than an absolute response criterion. A response is emitted only when the evidence for a response exceeds evidence for the other response by some criterion.

The model contains three parameters. The *signal strength* is the probability that the stimulus being processed is consistent with the response it maps on to, that is, in this case word or nonword. This parameter reflects the rate of evidence accumulation and is analogous to the diffusion model's drift rate. The *relative response criterion* is the amount of evidence a particular response must possess over its competitor before a response can be emitted; this is analogous to the diffusion model's boundary separation parameter. Finally, the *residual time* represents the nondecision components (i.e., encoding and output) of the response, similar to the diffusion model's T_{er} parameter.

We now turn to which of the two models more adequately accounts for the novel distributional effects observed in the Experiments 1 and 2.⁴ To reiterate, the results from Experiment 1 indicated that word-frequency effects were larger for PsHs than for legal nonwords, and this interaction was mediated almost entirely by the τ component. Nonword type effects for nonwords were more strongly mediated by τ . In Experiment 2, word-frequency effects were larger for legal than for illegal nonwords, and this

⁴ For the simulations, we are assuming that the legal nonword condition is invariant for the two contrasts. Empirically, of course, the legal nonword conditions in Experiments 1 and 2 yield similar but not identical results.

interaction was mediated primarily by μ and, to a smaller extent, by τ . Nonword type effects for nonwords were more strongly mediated by μ . Of course, in addition to capturing the effects of the variables on mean RTs, it is also important that the models accommodate how variables influence RT distributional shape, as reflected by the μ , σ , and τ parameters.

For the two-stage model, we assumed that as nonwords become more wordlike, the word–nonword overlap increases and, consequently, discrimination becomes more difficult. We therefore expect Stage 2 checking for low- and high-frequency words to increase as nonwords become more wordlike. The hybrid model contains a check parameter that determines the proportion of stimuli undergoing the Stage 2 analytic processes. In our simulation (see the Appendix for modeling details), we examined the distributional characteristics of high- and low-frequency words as a function of the check parameter. The check parameter was varied from -2.5 to 2.5 standard deviation units, with more negative values reflecting greater checking. As checking increases, the frequency effect in means (top two lines) increases, then decreases again at very high levels of checking (see Figure 8, top panel).

Turning to the predictions from the two-stage model for the ex-Gaussian parameters, the μ frequency effect becomes larger then smaller as checking increases. More interesting, checking strongly modulates the τ parameter. When there is little checking, the τ frequency effect is relatively small (see Figure 8, top panel, bottom two lines). As checking increases, it increases rapidly in size but becomes nonsignificant at very high levels of checking. The three vertical lines in the top panel of Figure 8 indicate possible checking parameters for the different nonword type conditions. For the PsH–legal nonword contrast, we selected two check parameter values to map onto the two experimental conditions (PsH check = 1.0; legal nonword check = 1.8). Plotting the model's predicted values for the two experimental conditions in Experiment 1 yielded the pattern shown in Table 4. Clearly, this accommodates the observed interaction nicely, where the larger frequency effect in the PsH condition is largely mediated by τ . Note, however, that the μ parameters for high- and low-frequency words in Table 2 are smaller in the legal nonword condition, compared with the PsH condition. The hybrid model predicts the opposite pattern (compare Tables 2 and 4).

The hybrid model also accounts for the nonword responses in both experiments. We assumed that more wordlike nonwords, such as *BRANE*, are both more familiar and also more likely to undergo checking. Proceeding on these assumptions, we ran simulations for the hybrid model (see Figure 8 bottom panel; modeling details are provided in the Appendix) and plotted the model's predicted nonword values for the two experiments (see Tables 4 and 5). As one can see, nonword type effects for nonwords are more strongly mediated by τ than μ when comparing PsHs with legal nonwords and more strongly mediated by μ than τ when comparing legal nonwords with illegal nonwords.

However, the hybrid model is hard pressed to account for the word results in Experiment 2. Empirically, the frequency effect is considerably smaller when illegal nonwords are used, and this attenuation is reflected mainly in μ . As shown in Figure 8 (top panel), the illegal nonword condition is captured by the region to the left of the legal nonword condition (legal nonword check = 1.8; illegal nonword check = 2.5), where checking is decreased. At the level of the mean, the model predicts slightly larger fre-

quency effects for the legal nonword condition compared with the illegal nonword condition (see Table 5). However, the interaction is mediated by both μ and τ , instead of being predominantly localized in μ . Furthermore, in Experiment 2, the presence of illegal nonwords dramatically attenuated word-frequency effects, especially in μ . This trend is not evident in Figure 8, where word-frequency effects in μ remain fairly stable across the different levels of checking. Finally, the μ parameters for high and low-frequency words are substantially lower in the illegal nonword condition, compared with the legal nonword condition; the hybrid model fails to capture this (compare Tables 3 and 5). Thus, it appears that the hybrid model is only partially successful in accounting for the Nonword Type \times Word Frequency interaction. It accounts for aspects of the PsH–legal nonword contrast and the nonword data but is unable to accommodate the legal–illegal nonword contrast.

At this point, it is worth summarizing why the hybrid model behaves the way it does and how this allows it to successfully simulate some but not all of our data. Let us first consider the legal nonword–PsH contrast in Experiment 1. When PsHs, compared with legal nonwords, are used, the word–nonword overlap increases, and Stage 2 checking increases. Increased Stage 2 checking, which is reflected by having a greater proportion of item RTs sampled from an ex-Gaussian distribution, considerably increases the frequency effect in τ but not in μ , which explains why the Nonword Type \times Word Frequency interaction is purely mediated by τ when PsHs are compared with legal nonwords. Turning to Experiment 2, the presence of legal nonwords, compared with illegal nonwords, also increases checking. For this contrast, however, checking negligibly increases the size of the frequency effect in μ , which is inconsistent with our finding that the μ frequency effect increases dramatically when legal nonwords, compared with illegal nonwords, are used (compare Tables 3 and 5). Clearly, the model tends to localize Stage 2 checking effects in the τ component and has some trouble when the data, such as the results from Experiment 2, do not conform to this pattern.

Turning to the random-walk model, we assumed that high-frequency words start out with higher signal strengths than low-frequency words. Also, as nonwords become more wordlike, binary discrimination becomes more difficult, and the signal strength of both high- and low-frequency words decreases in magnitude (Ratcliff et al., 2004; Stone & Van Orden, 1993). Using our instantiation of the random-walk model (see the Appendix for modeling details), we plotted ex-Gaussian parameters for words and nonwords as a function of signal strength, with relative response criterion and residual time held constant (see Figure 9). The six experimental conditions are superimposed as vertical lines and schematically illustrate our predictions. As before, we plotted the model's predictions for the six experimental conditions (see Tables 6 and 7). As Tables 6 and 7 make clear, the random-walk model's predictions are remarkably consistent with the results we obtained in the two experiments. First, at the level of the mean, frequency effects systematically become larger as nonwords become more wordlike. More intriguingly, for the PsH–legal nonword contrast, the interaction effect is fully mediated by τ , whereas for the illegal–legal nonword contrast, the interaction is mediated by μ and, to a smaller extent, by τ . The random-walk model also correctly predicts lower μ parameters for legal words compared with PsHs and for illegal words compared with legal words.

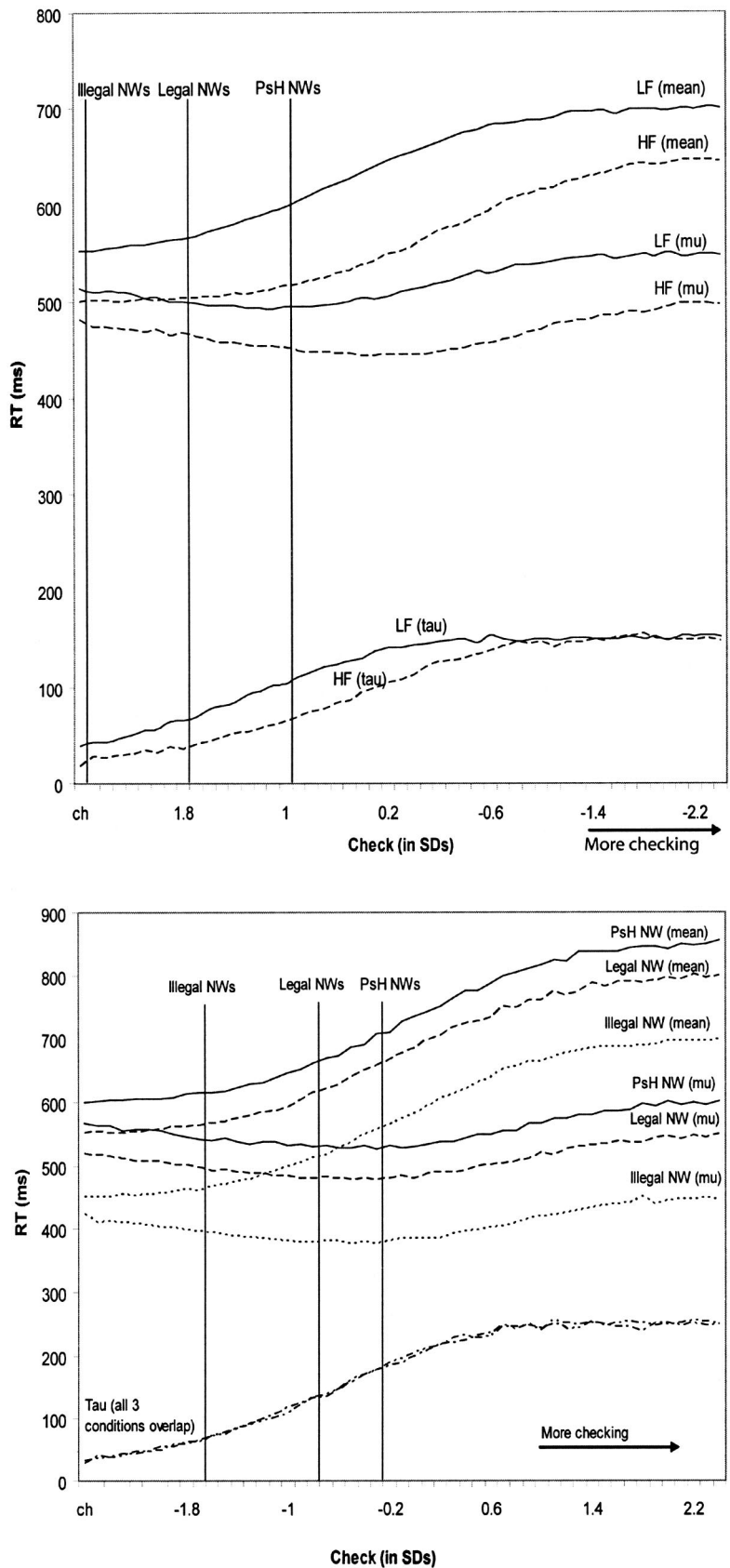


Figure 8. Mean, μ , and τ estimates from the response time (RT; in milliseconds) distributions of high-frequency words (HF), low-frequency words (LF), and nonwords (NWs), generated by the hybrid model when the check parameter is varied. PsH = pseudohomophone.

Table 4
Means of Hybrid Model's Mean Lexical Decision Response Times (in Milliseconds) and Ex-Gaussian Parameter Estimates as a Function of Word Frequency and Nonword Type (BRONE vs. BRANE)

Nonword type and word frequency	<i>M</i>	μ	σ	τ
Words				
Legal (check = 1.8)				
High frequency	503	464	51	38
Low frequency	564	500	48	64
Frequency effect	61	36	-3	26
Pseudohomophone (check = 1.0)				
High frequency	513	453	48	60
Low frequency	595	493	45	102
Frequency effect	82	40	-3	42
Difference of difference (interaction)	21	4	0	16
Nonwords				
Legal (check = -0.6)	618	478	41	140
Pseudohomophone (check = 0.2)	748	533	40	215
Nonword type effect	130	55	-1	75

The model was slightly less successful in accommodating nonword responses. Using the diffusion model parameters in Ratcliff et al. (2004) as a starting point, we made the simplifying assumption that the nonwords in a particular condition had negative signal strengths that were of similar magnitude to the low-frequency words in that condition (see Figure 9, bottom panel). Using these assumptions, we plotted the model's predicted nonword values for the two experiments (see Tables 6 and 7). As one can see, nonword type effects for nonwords are much more strongly mediated by τ than μ when comparing PsHs with legal nonwords. However, nonword type effects are only slightly mediated more by μ than τ when comparing legal nonwords with illegal nonwords. Although the random-walk model's prediction in this instance is qualitatively correct, it would have mimicked the empirical data more closely (see Table 3) if the difference between μ and τ were more pronounced.

Although the nonword simulations reveal some brittleness in our simple model, the random-walk model generally makes predictions that are consistent with the counterintuitive empirical findings, providing evidence that a simple random-walk framework in which only a single parameter is modulated can account for the effects observed in the two experiments. When we decompose the concave function of the random-walk model into μ and τ (see Figure 9), it is clear that as signal strength increases, τ decreases rapidly and curvilinearly whereas μ decreases gently and linearly. In particular, notice how skewed distributions are when signal strength is low, reflecting long random-walk finishing times when target-distractor discrimination is difficult. It is interesting to note that for low signal strengths, there is a marked difference in the slopes for the μ and τ parameters, with much steeper gradients for τ . As signal strength goes up, the slopes for the two parameters become increasingly parallel. These trends suggest that effects that reflect low signal strength processes (e.g., PsH vs. legal nonword) show more τ involvement than high signal strength processes, a view that is consistent with our findings.

General Discussion

The present studies generated a number of noteworthy findings. In two experiments, we replicated the classic Nonword Type \times Word Frequency interaction across two contrasts, showing that more wordlike nonword contexts produced slower word latencies and larger word-frequency effects. More important, the use of distributional analyses afforded evidence that the interactive effects of nonword context and word frequency were modulated by different components of the RT distribution. Specifically, when comparing PsHs and legal nonwords, the Nonword Type \times Word Frequency interaction was mediated totally by the τ (exponential) component. However, when comparing legal and illegal nonwords, the same interaction was mediated mainly by the μ (Gaussian) component. Qualitatively similar trends were observed with the nonword data. This pattern of results converged nicely with the vincentile plots and underscores how distributional analyses can serve to complement and extend traditional analyses of means. Differences that were not apparent at the level of the mean emerged at the level of the distributional analyses, thereby providing greater leverage in model adjudication.

It is interesting to note that the single-process random-walk model was able to account for these findings more naturally than the dual-process hybrid model. Although our simulations did not support the current instantiation of the hybrid model, they do not necessarily eliminate two-process models in general. However, if lexical decisions are indeed mediated by a familiarity-check process, then the specific assumptions underlying the current hybrid model have to be revised to accommodate the present results. One intriguing possibility is that the two stages map onto two separate accumulation processes (Diederich, 1997). A familiarity-based sequential sampling process may begin first, switching subsequently to a more strategic-analytic check process. The surprising finding is that a simple random-walk model can indeed account for the intriguing distributional changes without making such additional assumptions. A second possibility is that participants rely on a qualitatively different type of process when confronted with

Table 5
Means of Hybrid Model's Mean Lexical Decision Response Times (in Milliseconds) and Ex-Gaussian Parameter Estimates as a Function of Word Frequency and Nonword Type (NBREO vs. BRONE)

Nonword type and word frequency	<i>M</i>	μ	σ	τ
Words				
Illegal (check = 2.5)				
High frequency	500	482	58	18
Low frequency	553	514	53	39
Frequency effect	53	32	-5	21
Legal (check = 1.8)				
High frequency	503	464	51	38
Low frequency	564	500	48	64
Frequency effect	61	36	-3	26
Difference of difference (interaction)	8	4	2	5
Nonwords				
Illegal (check = -1.5)	464	394	44	70
Legal (check = -0.6)	618	478	41	140
Nonword type effect	154	84	-3	70

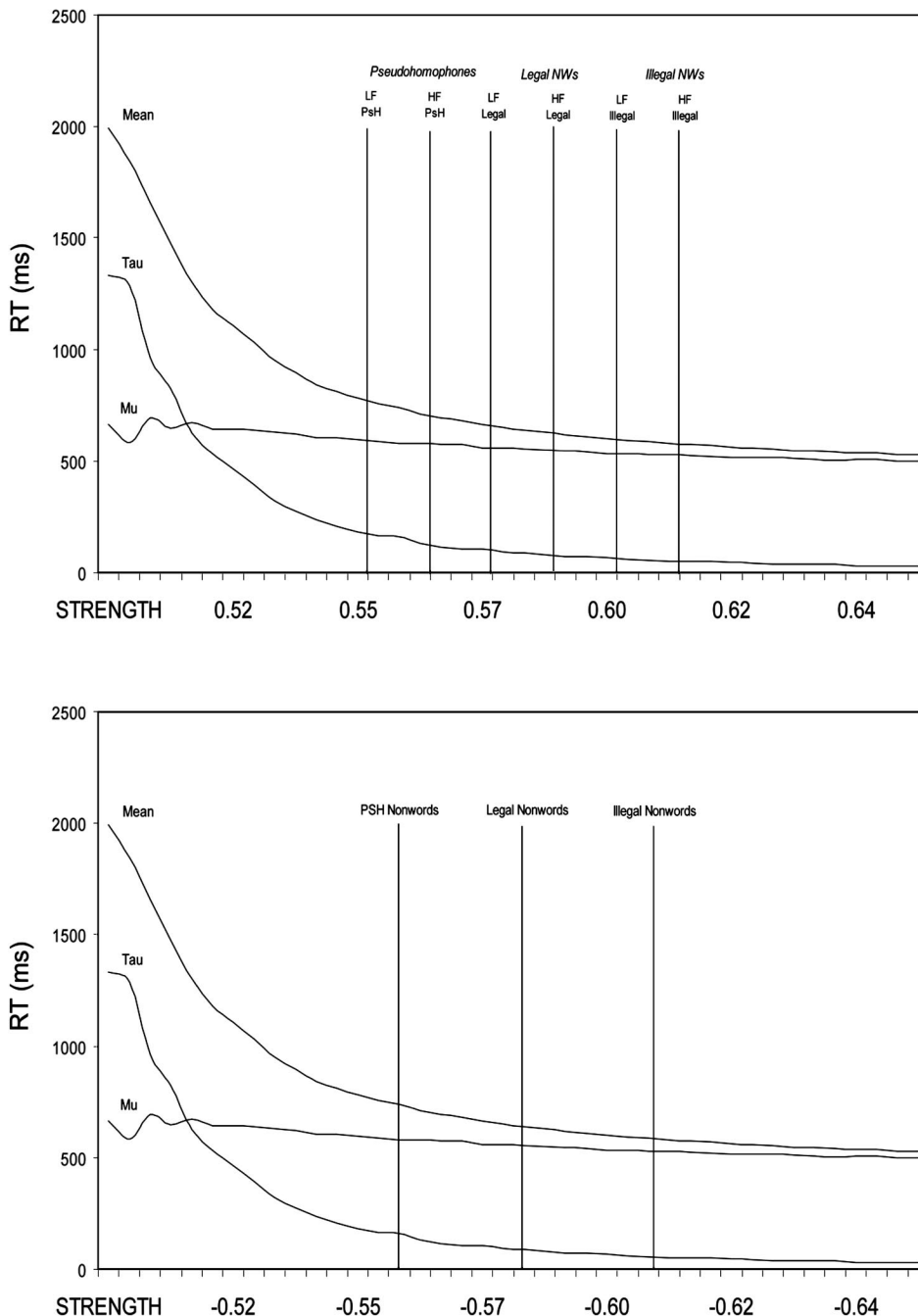


Figure 9. Mean, μ , and τ estimates from the response time (RT; in milliseconds) distributions of high-frequency words (HF), low-frequency words (LF), and nonwords (NWs), generated by the random-walk model when the signal strength parameter is varied. PsH = pseudohomophone.

illegal nonwords, relying on orthographic regularity in making the decisions, thereby minimizing lexical processing (as reflected by the diminished frequency effect). If this is the case, then the illegal nonword condition may be outside the scope of the two-process accounts of standard lexical decision processes, because lexical contributions have been minimized. Again, however, the random-walk model handles such a pattern without making such additional assumptions.

How Do Word Recognition Processes Map Onto Signal Strength–Drift Rate?

The random-walk simulations were carried out with two simple constraints. First, high-frequency words have larger signal strengths than low-frequency words. Second, increasing the word-likeness of nonwords decreases signal strengths for both high- and low-frequency words. These constraints are analogous to the mod-

Table 6
Means of Random-Walk Model's Mean Lexical Decision Response Times (in Milliseconds) and Ex-Gaussian Parameter Estimates as a Function of Word Frequency and Nonword Type (BRONE vs. BRANE)

Nonword type and word frequency	<i>M</i>	μ	σ	τ
Words				
Legal				
High frequency	650	559	43	91
Low frequency	680	576	52	104
Frequency effect	30	17	9	13
Pseudohomophone				
High frequency	755	590	57	165
Low frequency	810	605	63	205
Frequency effect	55	15	6	40
Difference of difference (interaction)	25	-2	-3	27
Nonwords				
Legal	680	576	52	104
Pseudohomophone	605	63	205	810
Nonword type effect	130	29	11	102

eling assumptions made by Ratcliff et al. (2004) in their diffusion model. It is interesting to note that modulating just signal strength, while holding other parameters constant, was sufficient for mimicking the complex pattern of observed distributional effects in the present experiments. As discussed earlier, signal strength is the probability that a stimulus is consistent with a particular response and reflects how rapidly evidence is accumulated. Ratcliff et al. also reported that variations in drift rate (the analogue of signal strength) could account for the effects of word frequency and nonword type, but they only modeled the frequency by legal-illegal nonword interaction, which produces effects mainly in μ . The present data indicate that the frequency by legal-PsH interaction is completely modulated by changes in τ . This pattern was also nicely handled by changes in signal strength in the random-walk model.

If we consider the classic models of word recognition, it is not obvious why high-frequency words might have higher signal strengths than low-frequency words. For example, in the classic logogen model (Morton, 1969), frequency has no influence on the rate of accumulation of evidence. In this model, a word detector (logogen) exists for every word in the reader's lexicon. Each logogen possesses a preset resting level of activation, and when a word is presented, the logogen for that word accumulates evidence until some threshold is exceeded, at which point word recognition takes place. Notice that word frequency modulates the recognition threshold of logogens. Because high-frequency words have logogens with lower thresholds, less evidence is required for recognition. Although the logogen model also assumes that evidence is accumulated over time, the critical point is that frequency does not modulate how rapidly that evidence is accumulating.

The classic interactive activation and competition (IAC) model (McClelland & Rumelhart, 1981) is also built on similar principles. The model has three processing levels (feature, letter, word) that are connected to each other by excitatory and inhibitory pathways, with every relevant unit represented by a node. A visual input first activates feature-level nodes, sending activation to letter-level nodes, and then on to word-level nodes, which in turn sends activation back to letter-level nodes. High-frequency words

start with higher resting levels of activation than low-frequency words, allowing them to inhibit competitors more rapidly. It is important to note that the rate of activation increase is determined mainly by the *net input*, which is defined as the summed excitatory and inhibitory influences of neighbors on a node. Despite differences (e.g., the logogen model is thresholded, the newer models are not), there is clearly a striking resemblance between the IAC model and the logogen model; the IAC model has indeed been described as a "hierarchical, nonlinear, logogen model" (McClelland & Rumelhart, 1981, p. 388) with interactivity and dynamical assumptions built in.

Clearly, the question of how signal strength or drift rate maps onto word recognition processes is an important one. The work by Ratcliff et al. (2004) and the findings described in this article suggest that frequency effects can be explained by variations in signal strength-drift rate. As the earlier discussion demonstrates, it is unclear how signal strength-drift rate maps onto the evidence accumulation processes in the logogen model or the IAC model, because the former assumes that there are differences in the rate of stimulus driven activation needed to surpass threshold for high- and low-frequency words and the latter assumes that high- and low-frequency words begin with different levels of resting activation. Obviously, these issues are also relevant to the DRC and MROM models, which, to a large extent, are built on the IAC framework. One possibility, suggested by Ratcliff et al., is that these models generate "wordness" values for incoming stimuli, and items with stronger lexical representations (i.e., high-frequency words) enter the lexical decision process with a higher signal strength-drift rate. Although this account works, it seems simplistic and assigns an unnecessarily marginal role to word recognition models.

We believe it is particularly important to consider how diffusion-type processes can be incorporated within the architecture of extant word recognition models. For example, although the latest instantiation of the DRC model (Coltheart et al., 2001) implements the lexical pathway using the IAC model, the activation dynamics are no longer identical to the IAC assumptions

Table 7
Means of Random-Walk Model's Mean Lexical Decision Response Times (in Milliseconds) and Ex-Gaussian Parameter Estimates as a Function of Word Frequency and Nonword Type (NBREO vs. BRONE)

Nonword type and word frequency	<i>M</i>	μ	σ	τ
Words				
Illegal				
High frequency	593	534	32	60
Low frequency	611	541	35	70
Frequency effect	18	7	3	10
Legal				
High frequency	650	559	43	91
Low frequency	680	576	52	104
Frequency effect	30	17	9	13
Difference of difference (interaction)	12	10	6	3
Nonwords				
Illegal	611	541	35	70
Legal	680	576	52	104
Nonword type effect	69	36	18	34

originally made by McClelland and Rumelhart (1981). Recall that in the original IAC model, word frequency is implemented by assigning higher baseline activation levels to nodes for high-frequency words. In the DRC model, word frequency has been moved to the equation that governs net input. Because net input effectively determines how rapidly activation is rising for a lexical entry, this implies that the activation for high-frequency words should therefore rise more rapidly than for low-frequency words, all other factors being equal (Coltheart et al., 2001). This suggests that there may be a mapping between the activation levels of entries in the orthographic lexicon and the signal strength–drift rate of the decision-making mechanism. Here, we describe one possible account of this mapping.

Specifically, we propose that the decision-making mechanism is continuously monitoring the activation of representations in the lexical system. There are actually two indexes within the DRC framework that could feed the decision-making mechanism. *Local activation* refers to the activation of individual lexical representations, whereas *global activation* refers to the sum of activations across lexical representations. Because we are arguing that the decision-making mechanism is constantly being updated across time, we believe that global activation is the more likely signal driving the decision-making mechanism. If local activation were being used as the stimulus unfolds across time, then the system would need to monitor each of the activated lexical representations and track the changes in activation across cycles. Although this is possible, this may be unnecessarily cumbersome, especially early in stimulus processing when many lexical candidates are receiving some activation. Hence, for simplicity, we consider how global processing could be used by the decision-making mechanism. It is also important to note here that although global activation includes all activated representations (e.g., orthographic neighbors) on a given cycle, it is clearly most influenced by the correct lexical candidate.

How might global activation be directly tied to the random-walk process to accommodate the present results? As activation is monitored over successive cycles, evidence is accumulated for the word response (i.e., value added to the word counter) on each cycle if the global activation index on the current cycle exceeds the global activation of the preceding cycle by some minimal amount. In contrast, if there is little change in global activation from Cycle N to Cycle $N + 1$, then evidence is accumulated for a nonword response (i.e., value added to the nonword counter).⁵ A word response is produced when the evidence accumulated in the word counter exceeds the evidence for the nonword counter by some threshold. Conversely, a nonword response is produced when the evidence accumulated for the nonword counter exceeds the evidence for the word counter by some threshold. This account predicts frequency effects in lexical decision because high-frequency words possess lexical entries with steeper rates of local activation (thereby also producing more global activation), which produce evidence for a word response at a more rapid rate across cycles.

It should also be noted that because frequency is represented in the strength of the connections between nodes within the PDP architecture of Seidenberg and McClelland's (1989) and Plaut, McClelland, Seidenberg, and Patterson's (1996) models, activation accumulates at different rates for high-frequency and low-frequency words, and so this could also be mapped onto drift rate. Hence, it appears quite reasonable that the more recently developed computational models of word recognition have parameters

that could accommodate the differences in drift rate for high- and low-frequency words.

Why Should Nonword Type Influence Drift Rate as Opposed to Response Criterion?

A conundrum produced by the present results is why changes in strength–drift rate across nonword contexts capture the distributional characteristics in the data instead of simple changes in criteria due to difficulty of the word–nonword discrimination. A priori, most extant models of word recognition assume that nonword type influences the response criterion that participants use to make a word response. Specifically, one would expect participants to become more conservative (i.e., set a higher criterion before a response is made) as the nonwords become more similar to words. In fact, this is precisely the account that Stone and Van Orden (1993) provided in their original explanation of the Nonword Type \times Word Frequency interaction. As nonwords become more wordlike (e.g., *BRANE* vs. *BRONE*), participants rely on a more conservative response criterion to make their decisions (see Figure 2). Note that this is conceptually similar to the lexical decision mechanisms adopted by the DRC and MROM models. As nonword foils become more wordlike, the summed activation of the orthographic lexicon is higher on nonword trials, and hence, the system adopts longer deadlines to avoid premature misses (Coltheart et al., 2001). The signal detection framework proposed by Seidenberg (1990; see also Plaut, 1997) provides complementary accounts. Because lexical decision involves speeded word–nonword discrimination, optimal decision criteria have to be established that permit fast responses while minimizing error rates. These criteria may shift depending on the nonword stimuli used in the task, allowing additional information (e.g., phonological and semantic) to play a role in the discrimination.

Of course, it is possible that by changing the response criteria within our random-walk model one could also capture the present results. We tested this idea with a series of random-walk simulations in which we kept signal strength constant while varying the response criterion. The results were unambiguous. Altering the response criterion only influences the μ component and has virtually no effect on τ , a pattern that was also reported by Spieler et al. (2000). In general, this is consistent with predictions from random-walk and diffusion-type models, that is, that shifts in

⁵ We have conducted DRC simulations with a subset of our stimuli that confirm that global lexical activity can be used to discriminate between the words and nonwords used in our experiments. However, it is plausible that a lexical decision system relying only on global lexical activity may have problems when it encounters words with few orthographic neighbors and nonword foils with many orthographic neighbors (M. Coltheart, personal communication, February 10, 2005). Because global lexical activity is sensitive to the orthographic neighborhood size of items, it may turn out that global activations in this situation may be higher for the nonwords than for the words. We would argue that, under these circumstances, where the experimenter directly manipulates a variable to work against global activation, other sources of information (e.g., local lexical activation or evidence from phonology and/or semantics) may play a more prominent role. It would also be useful to examine the distributional characteristics when global activation is truly equated across words and nonwords to determine whether an additional component is evidenced in these distributions.

criterion primarily shift the RT distribution instead of changing its shape and hence are primarily reflected in changes in μ . However, this is inconsistent with our data, where we clearly demonstrate that nonword type influences both μ and τ , depending on the difficulty of the discrimination.

In this light, one of the most intriguing questions of the present results is why nonword type manipulations should modulate the signal strength of words. After all, participants are responding to the same set of word stimuli across the different nonword contexts. Why would evidence accumulate more rapidly for *DOG* when it is paired with *NBREO* than when it is paired with *BRANE*? It does not seem reasonable to assume that the rate at which activation accumulates within a word recognition system built on a lifetime of experience should be so easily modified by the local nonword context.

One possible way to envisage the influence of nonword type is to consider the utility of the rate at which activation is accumulating within the word–nonword response systems. In the context of a difficult discrimination (e.g., in the context of PsHs), a single unit of information supportive of the word response (i.e., an increment in global activation between Cycle N and Cycle $N + 1$) is less informative than in the context of an easy discrimination (e.g., in the context of illegal nonwords). Hence, participants may use the difficulty of the discrimination to differentially weight the rate of activation within the lexicon. That is, the mapping between activation rate and signal strength–drift rate is not invariant but can be modulated by the difficulty of the discrimination process. When discrimination is easy (e.g., illegal foils), the mapping between activation and drift rate is weighted more heavily. When discrimination is difficult (e.g., PsH foils), the mapping between activation and drift rate is weighted less heavily. How does the system know whether discrimination is difficult or easy? An obvious source of information is the global (i.e., summed) lexical activity produced by nonwords across trials. More wordlike nonwords (e.g., *BRANE*) produce more global activity. Hence, the average global lexical activity for nonwords will be higher as discrimination difficulty increases. This information can then be used to adjust the weight between local lexical activity and signal strength–drift rate. It is important to note that because the system has to apply the same weight to both word and nonword trials, this explains why illegal nonwords (which should produce a stronger weight) possess a steeper signal strength–drift rate than PsHs (which should produce a weaker weight).

All extant models have mechanisms that accommodate discrimination difficulty, and typically these models adjust criteria arbitrarily on the basis of such difficulty. Our proposal is that discrimination difficulty, which is influenced by the global activation for nonwords, modulates the utility of activation rate; this activation rate is directly available from the output of word recognition models. To recapitulate, we believe that a plausible account of the present results is that word–nonword overlap modulates the utility of the global activation rate for high- and low-frequency words, and this is reflected in the signal strength parameter within our random-walk model of the decision process. Very simply, if the difficulty of the word–nonword discrimination is relatively easy, then the participant can rely heavily on activation building up in the lexicon to drive a decision, whereas, in a difficult discrimination, the activation building up will be relatively less informative because on some difficult nonword trials (e.g., for PsHs), activation will be building up at a relatively high rate even though the

stimulus is a nonword. This perspective has clear similarity to Ratcliff et al.'s (2004) account of lexical decision performance but more specifically indicates how this might be incorporated within available word recognition models.

To make our foregoing discussion more concrete, Figure 10 schematically illustrates how signal strength or drift rate vary as a function of word frequency and nonword context. As one can see, the rate of evidence accumulation for both words and nonwords becomes less steep as discrimination difficulty increases. Signal strength or drift rate also increases for higher frequency words. The response criteria in this figure represent the relative difference between the word and nonword accumulators. As noted above, we would argue that a simple way of connecting this framework to the DRC–MROM model is to argue that the slope of the drift rates is modulated by the global activation that accrues across the nonword trials, which in turn reflects the difficulty of the word–nonword discrimination. It is clear from Figure 10 that such changes in the rates modulate the word-frequency effect (as a function of nonword type) in the predicted fashion.

It is again important to emphasize that there are alternative ways to accommodate nonword type effects. For example, PDP-class models use signal detection to carry out lexical decisions, and the signal can either be orthographic (Seidenberg & McClelland, 1989) or semantic (Plaut, 1997) in nature. Nonword type effects arise because the model may consult different kinds of information across different nonword type contexts to optimize performance (Seidenberg, 1990). For example, consider how words and nonwords differ along several dimensions, any of which can be used to make a lexical decision. Words tend to have more familiar orthographic, phonological, and semantic patterns than nonwords. When orthographically illegal nonwords (e.g., *NBREO*) are used as foils, discrimination can be made on the basis of orthographic information alone. When orthographically legal nonwords (e.g., *BRONE*) are used, either phonology or semantics can reliably be used to carry out the discrimination because nonwords are not homophonous with real words. One major limitation of the PDP approach is that, unlike the DRC and MROM models, PDP models do not simulate lexical decision response latencies. For example, both Plaut (1997) and Harm and Seidenberg (2004) simulated asymptotic lexical decision accuracy rather than RTs (see Rastle & Coltheart, in press). To properly evaluate PDP models against our data, it is necessary to address the following two issues. First, how does nonword type context modulate the type of information (orthographic, phonological, semantic) that is being recruited to perform the lexical decision? Second, current PDP models generate either orthographic or semantic stress as their primary dependent variable. It is as yet unclear how stress values map onto RTs to produce the observed RT distributions, although Plaut (1997) has suggested that lexical decision latencies can be modeled by yoking the signal detection procedure to a diffusion process, with the assumption that stimuli with stronger signals produce steeper drift rates. To our knowledge, this has not yet been implemented.

In the foregoing discussion, we have attempted to make contact with the available models, through consideration of the decision mechanisms used in binary tasks, and have attempted to tie this to a general model of the binary decision process. It is still quite possible that the current effects of nonword type may be accommodated by changing response criteria mechanisms, or other mechanisms, within the available models. In this light, we look

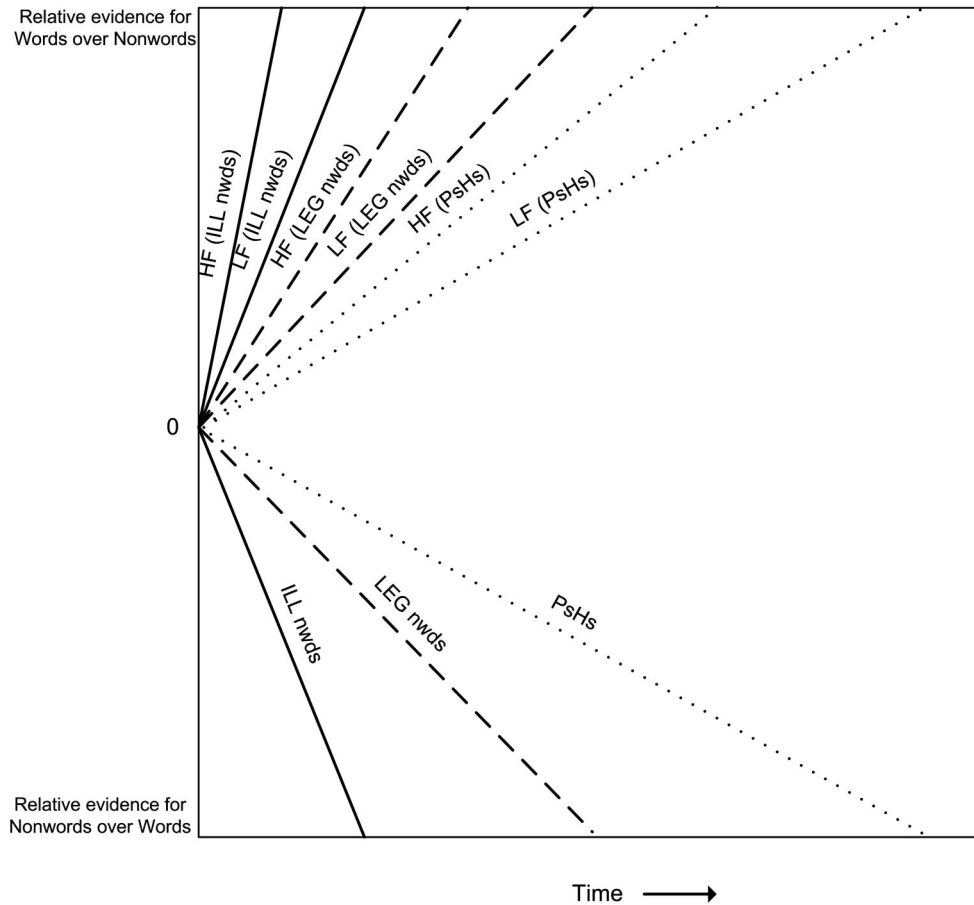


Figure 10. Signal strength–drift rate as a function of word frequency and nonword type. HF = high-frequency words; ILL = illegal; nwns = nonwords; LF = low-frequency words; LEG = legal; PsHs = pseudohomophones.

forward to specific implementations of existing word recognition models at the level of RT distributions.

Lexical Decision, Nonword Type, and the Effect of Distinct Codes

As discussed earlier, words and nonwords differ along several distinct dimensions, any of which can be used to make a lexical decision. In this light, it is at least plausible that different sources of information (e.g., orthographic, phonological, and semantic) contribute to a unidimensional quantity (wordness) that maps onto signal–drift rate (Ratcliff et al., 2004). Furthermore, it is noteworthy that there is evidence that distinct types of information appear to come online as a function of the word–nonword discrimination difficulty. For example, James (1975) obtained concreteness effects when legal nonwords were used but not when illegal nonwords were used. Joordens and Becker (1997) also observed the strongest semantic priming effects when PsH foils were used, compared with legal nonwords or illegal nonwords. Similarly, Waters and Seidenberg (1985) found stronger regularity effects in lexical decision when orthographically strange words (e.g., *AISLE*) were included. These findings reinforce the notion that semantics

or phonology are emphasized when the orthographic code becomes less useful for discrimination. Ultimately, such qualitatively distinct effects as a function of the word–nonword discrimination are also consistent with the notion that lexical decision latencies are modulated by a flexible lexical processor (Balota, Paul, & Spieler, 1999; Balota & Yap, in press), in which local task demands and contexts determine the extent to which attention is directed toward various lexical processing pathways. Within the DRM–MROM framework described earlier, such local task demands and contexts could also determine the extent to which local and global activation contribute toward the random-walk process.

It is indeed intriguing that nonword context modulates both signal strength and the salience of different stimulus dimensions. Clearly, signal strength–drift rate are able to provide only a partial account of lexical decision phenomena. Although it accounts for behavior (i.e., RT distributions and accuracy) remarkably well, it makes no predictions about the specific lexical information–processes that are contributing to the random-walk–diffusion process. If indeed a sequential sampling process is the best description of lexical decision performance, then we believe it is important, from a psycholinguistic perspective, to develop unified frame-

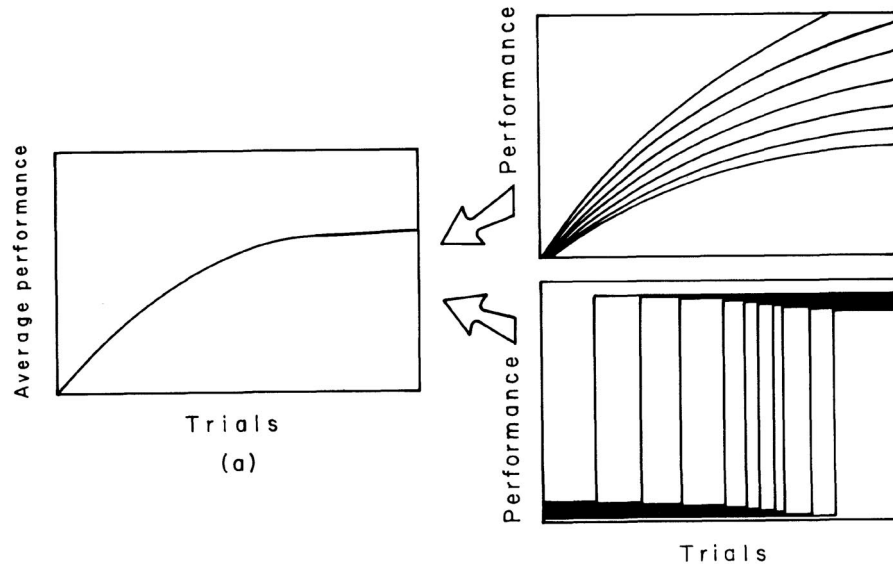


Figure 11. Illustration of how data averaging across participants creates an incremental learning curve, regardless of the actual learning function at the level of the individual participant. From *Principles of Learning and Memory* (p. 267), by R. G. Crowder, 1976, Hillsdale, NJ: Erlbaum. Copyright 1976 by Lawrence Erlbaum Associates. Reprinted with permission.

works that explicitly consider the lexical processing pathways that are brought online and how these pathways drive drift rate.

Revisiting Single Versus Multiprocess Distinctions in Models

The present results suggest that the Nonword Type \times Word Frequency interaction is better accommodated by a single-process model than the complex two-stage hybrid model. Of course, one must be cautious because we have only tested a single instantiation of the hybrid model. Moreover, some additional challenges need to be overcome before a more definitive answer is available. For example, one could argue that even RT data at the distributional level may be somewhat limited. Attempting to discriminate between a single-process and a multiprocess model may well be akin to discriminating between gradual and all-or-none learning in quantitative models of learning in the 1960s and 1970s. Specifically, researchers were interested in whether the rate of learning was better reflected by a negatively accelerated curve (i.e., learning is gradual and continuous) or a step function (i.e., learning is all-or-none). Crowder (1976) elegantly demonstrated how averaging data across participants creates an incremental learning curve (see Figure 11), regardless of the actual learning function at the level of the individual participant.

It is at least plausible that the present techniques of averaging over many trials and participants may indeed smooth potentially discrete changes in lexical processing. Using only group RT distributions across many trials, it may not be possible to conclusively tell whether decision performance is driven by one or two or more processes. As Estes (2002) argued, more definitive converging evidence may be uncovered by cognitive neuroscience methods. For example, it is possible that event-related potential experiments may reveal dissociations between variables that, in principle,

should selectively influence different stages. Specifically, there may be spatially and temporally distinct event-related potential signals that are associated with qualitatively different processing stages.

Conclusions

In the present article, we used RT distributional analysis to evaluate two general frameworks for interpreting binary decision processes, that is, the two-process hybrid model and a simple random-walk model. The targeted behavioral phenomenon was the Nonword Type \times Word Frequency interaction in lexical decision performance. We demonstrated that the random-walk model provided a better account of the observed effects via simple changes in signal strength. *Prima facie*, this is consistent with Ratcliff et al.'s (2004) recent argument that the burden of word recognition models is to generate drift rates for different experimental conditions and to feed this information into a decision process. Of course, a model of the decision process is only one target for word recognition researchers. Ultimately, it is critical to have an understanding of the representations and processes that feed into the drift rates and to consider the interplay among lexical structures, processes, and decision making.

References

- 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.
- 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 psychol-*

- ogy: *Learning, memory, and thinking* (Vol. 1, pp. 242–293). San Francisco: Freeman.
- Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 340–357.
- 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. (in press). Attentional control and flexible lexical processing: Explorations of the magic moment of word recognition. In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing*. Hove, England: Psychology Press.
- 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.
- Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.
- Diederich, A. (1997). Dynamic stochastic models for decision making with time constraints. *Journal of Mathematical Psychology*, *41*, 260–274.
- Estes, W. K. (2002). Traps in the route to models of memory and decision. *Psychonomic Bulletin & Review*, *9*, 3–25.
- Gordon, B. (1983). Lexical access and lexical decision: Mechanisms of frequency sensitivity. *Journal of Verbal Learning and Verbal Behavior*, *22*, 24–44.
- Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, *103*, 518–565.
- Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*, 662–720.
- 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.
- 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.
- Hohle, R. H. (1965). Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology*, *69*, 382–386.
- James, C. T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 130–136.
- Joordens, S., & Becker, S. (1997). The long and short of semantic priming effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 1083–1105.
- Lattin, J. M., Carroll, J. D., & Green, P. E. (2003). *Analyzing multivariate data*. Pacific Grove, CA: Brooks/Cole.
- 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., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Pt. 1. An account of basic findings. *Psychological Review*, *88*, 375–407.
- Mewhort, D. J. K., Braun, J. G., & Heathcote, A. (1992). Response-time distributions and the Stroop task: A test of Cohen, Dunbar, and McClelland's (1990) parallel distributed processing model. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 872–882.
- 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.
- Nelder, J. A., & Mead, R. (1965). A simplex algorithm for function minimization. *Computer Journal*, *7*, 308–313.
- Pike, A. R. (1973). Response latency models for signal detection. *Psychological Review*, *80*, 53–68.
- 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., 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.
- R Development Core Team. (2004). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rastle, K., & Coltheart, M. (in press). Is there serial processing in the reading system and are there local representations? In S. Andrew (Ed.), *From inkmarks to ideas: Current issues in lexical processing*. Hove, England: Psychology Press.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.
- Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Review*, *86*, 446–461.
- Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, *111*, 159–182.
- Ratcliff, R., & Rouder, J. F. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*, 347–356.
- 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.
- Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, *9*, 487–494.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2001). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools.
- 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.
- Spieler, D. H., Balota, D. A., & Faust, M. E. (2000). Levels of selective attention revealed through analyses of response time distributions. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 506–526.
- Stone, G. O., & Van Orden, G. C. (1993). Strategic control of processing

- in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 744–774.
- Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, 7, 424–465.
- Vickers, D. (1979). *Decision processes in visual perception*. New York: Academic Press.
- Vincent, S. B. (1912). The function of vibrissae in the behavior of the white rat. *Behavioral Monographs*, 1, (Whole No. 5).
- Waters, G. S., & Seidenberg, M. S. (1985). Spelling-sound effects in reading: Time course and decision criteria. *Memory & Cognition*, 13, 557–572.
- Yap, M. J., & Balota, D. A. (in press). Additive and interactive effects on response time distributions in visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46, 441–517.

Appendix

Implementing the Hybrid Two-Stage Model and the Random-Walk Model

Implementing the Hybrid Two-Stage Model

Determining the Familiarity–Meaningfulness (FM) Values of Items

First, to determine the FM values of items, we generated a Gaussian distribution ($\mu = 500$, $\sigma = 100$), and values were sampled from this distribution. The FM value of an item was the algebraic sum of this value and a boost parameter. If an item's FM value was above the mean of the Gaussian distribution (i.e., greater than 500), then the item was assumed to undergo the Stage 1 (familiarity) processes. If the FM value was below the mean of the FM distribution (i.e., less than 500), then the item was assumed to undergo Stage 1 + Stage 2 (familiarity + check) processes. In our simulations, the boost parameters for high- and low-frequency words were set at 36 and –48, respectively, which affords lower FM values and more checking for low-frequency words.

Generating Response Times (RTs) for Stage 1 (Familiarity) Processes

Items that undergo Stage 1 produce RTs that were sampled from a Gaussian distribution of specified μ and σ . In this stage, the familiarity of items is assumed to be related to the time needed to respond to them. Higher frequency words (i.e., more familiar items) produce responses from a Gaussian distribution with a smaller μ . To instantiate this, high-frequency words that enter Stage 1 produce RTs that were sampled from a Gaussian distribution in which $\mu = 500$ and $\sigma = 60$. Low-frequency words, in contrast, were sampled from a Gaussian distribution in which $\mu = 550$ and $\sigma = 60$.

Generating RTs for Stage 1 + Stage 2 (Familiarity + Check) Processes

Items that go through both stages produce RTs that were sampled from an ex-Gaussian distribution of specified μ , σ , and τ . The Gaussian parameters (μ and σ) reflect familiarity-based Stage 1 processes, whereas the exponential parameter (τ) reflects Stage 2 checking processes. High-frequency words that enter Stage 2 produce RTs that were sampled from an ex-Gaussian distribution in which $\mu = 500$, $\sigma = 60$, and $\tau = 150$. Stage 2 low-frequency words were sampled from an ex-Gaussian distribution in which $\mu = 550$, $\sigma = 60$, and $\tau = 150$.

Simulating Checking

The extent of checking in the model was simulated by manipulating the proportion of items that undergo Stage 2 processes. This was implemented

by shifting the Gaussian mean (i.e., 500) that FM values are evaluated against. Checking is increased when the Gaussian mean becomes larger (more FM values fall below threshold) and is decreased when it becomes smaller (more FM values fall above threshold).

Convolution of Stage 1 and Stage 2 RTs

To produce the final distribution, we convoluted the RTs for Stage 1 (Gaussian) and Stage 1 + Stage 2 (ex-Gaussian) processes to create a distribution that is well captured by the ex-Gaussian function. For the simulations described in this article, 10,000 replicates were used in each run to generate the targeted distribution.

Simulating Nonword Responses

To simulate nonword responses, we assumed that the different nonword types varied on familiarity and should therefore be sampled from different Stage 1 (Gaussian) distributions. Illegal nonword RTs were sampled from a Gaussian distribution with $\mu = 450$ and $\sigma = 60$, legal nonwords were sampled from a distribution with $\mu = 550$ and $\sigma = 60$, and pseudohomophones were sampled from a distribution with $\mu = 600$ and $\sigma = 60$. Nonwords that entered both stages produced RTs that were sampled, respectively, from an ex-Gaussian distribution with $\mu = 450$, $\sigma = 60$, and $\tau = 250$ (illegal); a distribution with $\mu = 550$, $\sigma = 60$, and $\tau = 250$ (legal); or a distribution with $\mu = 550$, $\sigma = 60$, and $\tau = 250$ (pseudohomophone). We also manipulated the proportion of nonwords undergoing checking and assumed that greater checking would be present for nonwords that were more wordlike. The check parameters for illegal nonwords, legal nonwords, and pseudohomophones were set at –1.5, –0.6, and 0.2, respectively (more positive values imply greater checking).

Implementing the Random-Walk Model

The random-walk model had three parameters: signal strength, relative response criterion, and residual time. For the simulations described in this article, relative response criterion and residual time were held constant at 45 units and 400 ms, respectively. Signal strength alone was varied, and we examined how RT distributional characteristics were modulated by signal strength.

Generating the RT for a Single Trial

To generate the RT for a single trial, incoming evidence is allocated to a response counter (word or nonword) at each cycle, with each cycle assumed to have a duration of 1 ms. We implemented this by sampling from a normal distribution ($\mu = 0$, $\sigma = 1$). If the sampled value was greater

than some specified criterion from this normal distribution, then a unit of evidence was allocated to the word response; otherwise, it was allocated to the nonword response. Signal strength is reflected by the criterion from the normal distribution that defines whether a randomly selected unit is added to the word or nonword counter. Greater signal strengths lower the criterion, making it more probable that evidence is allocated to the word response counter. This process was repeated until the evidence for a response exceeded its competitor by some specified amount. In our simulations, the random-walk process terminated when the difference between the word and nonword response counters exceeded 45 units. The number of cycles taken to achieve this difference is the *decision latency*, which was added to the residual time to produce the overall RT for that trial.

Generating RT Distributions

The description above only produces a single RT. To simulate RT distributions, it is necessary to repeat the process to generate multiple RTs. For our simulations, we generated 3,000 RTs for each run and fitted the resulting RT

distributions to an ex-Gaussian function. Just as Spieler et al. (2000) observed, these distributions were well fit by the ex-Gaussian function.

Simulating Nonword Responses

To simulate nonword responses, we assumed that the nonwords in the three nonword type conditions had negative signal strengths that were of similar magnitude to the low-frequency words in the respective conditions (on the basis of diffusion-model parameters in Ratcliff et al., 2004). For example, in the pseudohomophone condition, if the signal strength of low-frequency words was 0.548, then we assumed that pseudohomophones therefore had a signal strength of -0.548 . Obviously, this is merely a very crude first approximation.

Received July 12, 2005

Revision received February 17, 2006

Accepted February 19, 2006 ■

Members of Underrepresented Groups: Reviewers for Journal Manuscripts Wanted

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write to the address below. Please note the following important points:

- To be selected as a reviewer, you must have published articles in peer-reviewed journals. The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.
- To be selected, it is critical to be a regular reader of the five to six empirical journals that are most central to the area or journal for which you would like to review. Current knowledge of recently published research provides a reviewer with the knowledge base to evaluate a new submission within the context of existing research.
- To select the appropriate reviewers for each manuscript, the editor needs detailed information. Please include with your letter your vita. In the letter, please identify which APA journal(s) you are interested in, and describe your area of expertise. Be as specific as possible. For example, "social psychology" is not sufficient—you would need to specify "social cognition" or "attitude change" as well.
- Reviewing a manuscript takes time (1–4 hours per manuscript reviewed). If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.

Write to Journals Office, American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242.