Journal of Experimental Psychology:
Learning, Memory, and Cognition
2015, Vol. 41, No. 3, 597–613

0278-7393/15/$12.00 http://dx.doi.org/10.1037/xlm0000064

Responding to Nonwords in the Lexical Decision Task: Insights From the English Lexicon Project

Melvin J. Yap
National University of Singapore

Daragh E. Sibley
Haskins Laboratories, New Haven, Connecticut

David A. Balota
Washington University in St. Louis

Roger Ratcliff
Ohio State University

Jay Rueckl
University of Connecticut

Researchers have extensively documented how various statistical properties of words (e.g., word frequency) influence lexical processing. However, the impact of lexical variables on nonword decision-making performance is less clear. This gap is surprising, because a better specification of the mechanisms driving nonword responses may provide valuable insights into early lexical processes. In the present study, item-level and participant-level analyses were conducted on the trial-level lexical decision data for almost 37,000 nonwords in the English Lexicon Project in order to identify the influence of different psycholinguistic variables on nonword lexical decision performance and to explore individual differences in how participants respond to nonwords. Item-level regression analyses reveal that nonword response time was positively correlated with number of letters, number of orthographic neighbors, number of affixes, and base-word number of syllables, and negatively correlated with Levenshtein orthographic distance and base-word frequency. Participant-level analyses also point to within- and between-session stability in nonword responses across distinct sets of items, and intriguingly reveal that higher vocabulary knowledge is associated with less sensitivity to some dimensions (e.g., number of letters) but more sensitivity to others (e.g., base-word frequency). The present findings provide well-specified and interesting new constraints for informing models of word recognition and lexical decision.

Keywords: visual word recognition, nonwords, individual differences, ex-Gaussian analysis, diffusion model

In the lexical processing literature, a prodigious amount of work has been directed at identifying the various statistical properties (e.g., word frequency, number of letters, number of orthographic neighbors, imageability) that influence how quickly and accurately participants can recognize visually presented words (see Balota, Yap, & Cortese, 2006, for a review). This wealth of findings has yielded rich insights into the mechanisms underlying visual word recognition and has stimulated the development of sophisticated computational models that are able to closely approximate human performance (e.g., Perry, Ziegler, & Zorzi, 2007). Although *word* recognition has been well studied, much less work (e.g., Whaley, 1978) has focused on the processes that underlie *nonword* responses, particularly in the context of the lexical decision task (LDT), in which participants are required to discriminate between real words and nonwords (e.g., FLIRP). Indeed, in a lexical decision study, experimenters have little interest in participants' nonword data and typically discard them. Importantly, a better specification of the processes driving nonword responses could help inform the mechanisms underlying lexical processing (Caramazza, Laudanna, & Romani, 1988). Specifically, in the LDT, information is accumulated over time for both words and nonwords, and the participant presumably relies on lexical processes to generate signals that can be used to discriminate words from nonwords. Indeed, one might even argue that nonwords may provide unique information regarding these early processes, because they are not contaminated by the influence of the word itself.

The present study leverages on the power of the megastudy approach to explore the influence of different nonword statistical properties on nonword lexical decision performance for almost

This article was published Online First October 20, 2014.

Melvin J. Yap, Department of Psychology, National University of Singapore; Daragh E. Sibley, Haskins Laboratories, New Haven, Connecticut; David A. Balota, Department of Psychology, Washington University in St. Louis; Roger Ratcliff, Department of Psychology, Ohio State University; Jay Rueckl, Department of Psychology, University of Connecticut.

This work was supported in part by National Institute on Aging Grants PO1 AGO3991 and PO1 AGO26276 as well as National Science Foundation Grant BCS 0001801, awarded to David A. Balota. Portions of this research were presented at the 52nd annual meeting of the Psychonomic Society, November 2011, Seattle, Washington. We thank Colin Davis, Steve Lupker, James Adelman, and Dennis Norris for their valuable comments on earlier versions of this manuscript.

Correspondence concerning this article should be addressed to Melvin J. Yap, Department of Psychology, National University of Singapore, Block AS4, #02-07, Singapore 117570. E-mail: Melvin@nus.edu.sg

37,000 nonwords from the English Lexicon Project (ELP; Balota et al., 2007; see also Balota, Yap, Hutchison, & Cortese, 2012, for a review). Using trial-level lexical decision data from the ELP from over 800 participants, we also assess the stability of nonword decision measures and the interrelationships between individual differences in vocabulary knowledge and nonword decision performance.

Effects of Psycholinguistic Variables on Nonword Lexical Decision Performance

Although studies based on the LDT have emphasized word processing, a number of characteristics has been shown to influence nonword lexical decision performance, including neighborhood density, base-word properties, and length (syllabic, morphemic, and letter). In their seminal study, Coltheart, Davelaar, Jonasson, and Besner (1977) examined the influence of *orthographic neighborhood size* (i.e., orthographic *N*, the number of words derivable by changing one letter while preserving the identity and position of the other letters). Orthographic *N* can be computed for both words and nonwords (e.g., FLIRP's only word neighbor is FLIRT). Although words with many neighbors (particularly low-frequency words) are classified more quickly than words with few neighbors (Andrews, 1989, 1992), nonwords with many neighbors are responded to more slowly (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Coltheart et al., 1977). More recently, the neural correlates underlying this dissociation have been explored by Holcomb, Grainger, and O'Rourke (2002), and they made the intriguing observation that distinct effects of orthographic *N* for words and nonwords are seen in behavioral response times (RTs), but not in event-related potential components. Specifically, for both words and nonwords, items from large neighborhoods, compared with words from small neighborhoods, elicited larger N400s, suggesting that orthographic *N* effects for words and nonwords implicate the same basic, response-independent processes.

Researchers have also investigated how the properties (e.g., word frequency) of the base word a nonword is derived from affect nonword lexical decision times. For example, KEAP is a *pseudohomophone* (i.e., nonword homophonous with a real word) that is derived from KEEP. In addition to using pseudohomophones, nonwords can also be created by replacing (e.g., FLIR*P* from FLIR*T*) or transposing (e.g., JU*GD*E from JU*DG*E) letters in the base word.[1] Interestingly, compared with orthographic *N* effects, the empirical evidence for base-word frequency effects in nonword lexical decision has been more equivocal (see Perea, Rosa, & Gómez, 2005, for a review). Although some studies indeed report a disadvantage for high-frequency nonwords (e.g., Andrews, 1996; Perea et al., 2005), other studies (e.g., Duchek & Neely, 1989; Ziegler, Jacobs, & Klüppel, 2001) yield the opposite pattern, and yet other studies (e.g., Allen, McNeal, & Kvak, 1992) find no effect.

In addition to word frequency, a nonword's base word is associated with other important lexical properties that could potentially influence the processing of that nonword. These properties include *letter length* (number of letters), *syllabic length* (number of syllables), and *morphemic length* (number of morphemes). For letter length, Balota et al. (2004) reported that participants took more time to reject nonwords with more letters (see also Whaley, 1978), consistent with the idea that the processing of nonwords in lexical decision is more likely to implicate serial processes (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) or peripheral visual input or articulatory output processes (Seidenberg & Plaut, 1998; but see Perry et al., 2007).

Turning to syllabic length, although there is support for the role of syllables in visual word recognition, particularly in languages with well-defined syllabic boundaries and a shallow orthography (Álvarez, Carreiras, & Taft, 2001; Carreiras, Álvarez, & de Vega, 1993; Conrad & Jacobs, 2004; Perea & Carreiras, 1998), whether syllables function as processing units in English is more contentious (see Yap & Balota, 2009, for a review). That said, work by Yap and Balota (2009) indicate that a word's syllabic length (see also Ferrand & New, 2003) is positively correlated with both speeded pronunciation and lexical decision latencies, after influential covariates such as letter length, phoneme length, word frequency, neighborhood size, and phonological consistency are controlled for (see also Butler & Hains, 1979; Muncer & Knight, 2012; New, Ferrand, Pallier, & Brysbaert, 2006). Interestingly, although syllabic length is a robust predictor of word lexical decisions, the impact of this variable on nonword lexical decisions is less clear. In a French lexical decision study, Ferrand and New (2003) did not observe a syllabic length effect for nonwords. Similarly, Muncer and Knight (2012) examined lexical decision responses to mono- and disyllabic five-letter nonwords in the British Lexicon Project (Keuleers, Lacey, Rastle, & Brysbaert, 2012) and failed to find a significant effect of syllabic length.[2]

Unlike syllabic length, there has been almost no work examining the impact of morphemic length on lexical decisions to nonwords. However, there are a number of studies demonstrating the influence of morphological structure on nonword lexical decisions. Specifically, nonwords are more difficult to reject when they are made up of existing morphemes, compared with when they are not. For example, participants respond more slowly to nonwords (both prefixed and non-prefixed) with a real stem (e.g., DEJUVENATE or JUVENATE) than those with pseudo stems (e.g., DEPERTOIRE or PERTOIRE) (Taft & Forster, 1975). In the same vein, Caramazza et al. (1988), using Italian stimuli, reported that nonwords that can be fully decomposed into morphemes (e.g., CANT-EVI) elicit longer RTs than nondecomposable (i.e., pseudo stem and nonsuffix ending, e.g., CANZ-OVI) nonwords. There is also evidence that morphologically complex nonwords are rejected more slowly than controls when morphemes are presented in their usual positions (e.g., GASFUL vs. GASFIL) but not when they are reversed (e.g., FULGAS vs. FILGAS), pointing to the position specificity of underlying suffix representations (Crepaldi, Rastle, & Davis, 2010). Collectively, these studies suggest that morphologically complex words (and nonwords) are decomposed into morphemes during word recognition, and consequently one would expect

[1] Interestingly, recent models of orthographic input coding, such as the spatial coding model (Davis, 2010), the open-bigram model (Grainger & Van Heuven, 2003), the SERIOL model (Whitney, 2001), and the overlap model (Gomez, Ratcliff, & Perea, 2008) help provide a principled explanation for why manipulating the form of nonwords in this manner might affect nonword decision performance.

[2] Syllabic length effect might be moderated by the difficulty of the nonword. Specifically, supplementary analyses by Muncer et al. (2012) indicate reliable syllabic length effects for nonwords with response times longer than the mean response times for words.

processing time to be longer for nonwords with more morphemes. In line with this, Muncer, Knight, and Adams (2013), using data from the British Lexicon Project, reported that nonwords containing an inflectional morpheme (e.g., -S, -ER, -EST, -ED) were more difficult to reject in lexical decision than nonwords without these morphemes.

The first objective of the present study was to use hierarchical regression analyses to examine and compare the unique influence of a comprehensive set of variables (neighborhood density, morphemic and syllabic length, base-word frequency) on nonword lexical decision times. Although the effects of the foregoing variables have been separately investigated across different studies, no study, to our knowledge, has examined all these variables at the same time on a common set of items. Doing so will allow us to assess the *relative* unique predictive power of the different variables, which will help provide finer-grained constraints for computational models. Specifically, instead of just regressing model latencies onto human latencies (Spieler & Balota, 1997), models can be tested more rigorously by assessing whether a model's latencies are affected to the same extent by the variables that influence human latencies (Perry, Ziegler, & Zorzi, 2010). Our analyses may also help shed light on some of the empirical discrepancies in the literature.

In addition to the traditional neighborhood density metrics (e.g., Coltheart et al., 1977), measures based on Levenshtein distance (Yarkoni, Balota, & Yap, 2008) are also explored. The Levenshtein measures (to be described later) incorporate comparisons between all pairs of words in the lexicon, including words of different length. They serve as an important complement to traditional neighborhood density measures, which have limited or no variance for long letter strings (e.g., a long nonword like TELECOMMANICUTIONS has no orthographic neighbors). The results of these analyses will provide a well-specified set of benchmark phenomena for informing models of word recognition and lexical decision. More specifically, a more complete description of the functional relationships between stimulus properties and nonword lexical decision performance can help shed more light on the mechanisms driving "nonword" responses. We now turn to a selective review of the nonword lexical decision modeling literature.

Modeling Nonword Lexical Decision Performance

According to the dual-route cascaded (DRC) model (Coltheart et al., 2001) and multiple read-out model (MROM; Grainger & Jacobs, 1996), the mechanism for making lexical decisions monitors lexical activity both *locally* (at the level of individual representations) and *globally* (summed activity across all representations). A word response is made when either local or global activity exceeds their prespecified respective thresholds. A nonword response is produced after some processing duration (or deadline) has elapsed, and a word response has not been made. To improve the efficiency of the system, the nonword deadline is flexible and is extended when the system detects more global lexical activity early on in processing (Coltheart et al., 2001).

Although a variable deadline can accommodate Coltheart et al.'s (1977) finding of longer mean latencies for more wordlike nonwords (i.e., nonwords with many word neighbors), it has more difficulty with the equivocal effects of base-word frequency in nonword lexical decision. Specifically, some studies find *shorter* latencies for nonwords derived from high-frequency base words (e.g., Duchek & Neely, 1989; Ziegler et al., 2001), or no effect (e.g., Allen et al., 1992). This has led to the proposal that distinct mechanisms drive the "no" response in lexical decision, and these produce opposing effects that could sometimes offset each other (Perea et al., 2005). Specifically, in addition to the variable deadline mechanism described earlier, there is a later verification procedure that detects deviations between nonwords and their respective base words (Paap, Newsome, McDonald, & Schvaneveldt, 1982). This verification is frequency ordered wherein nonwords with *higher* frequency base words will be checked (and rejected) earlier. However, it remains unclear how a combined deadline/verification procedure might produce morphemic or syllabic length effects.

The major current computational and quantitative models of lexical decision do not assume that nonword responses are driven by distinct and opposing processes. As described earlier, both the MROM (Grainger & Jacobs, 1996) and DRC model (Coltheart et al., 2001) rely on a variable temporal deadline for making a nonword decision. The deadline approach has been heavily criticized (see Ratcliff, Gomez, & McKoon, 2004; Wagenmakers, Ratcliff, Gomez, & McKoon, 2008). Specifically, empirical RT distributions are virtually always positively skewed, and a deadline model cannot predict this because deadline time is constrained to be normally distributed across trials (Ratcliff et al., 2004). Moreover, a deadline account is unable to generate fast responses to nonwords when a reasonable accuracy rate is required (Wagenmakers et al., 2008).

More recent approaches to modeling nonword lexical decision have likened it to the sequential sampling of noisy information in a diffusion process (Ratcliff et al., 2004) or have computed and compared the posterior probability that the input stimulus is a word versus a nonword (Norris, 2006). In particular, the Bayesian reader model (Norris, 2006, 2009) unifies lexical and decision processes within a common framework that assumes that readers behave like optimal Bayesian decision makers when carrying out lexical decisions. Specifically, the model computes the probability that the presented letter string is a word rather than a nonword, given the input, and it does this by deciding whether an input is more likely to have been generated by a word or by a nonword near the input. Indeed, an extended version of the Bayesian reader model that adds noise to the input (Norris, 2009) has been shown to be able to simulate empirical RT distributions well.

Recently, Dufau, Grainger, and Ziegler (2012) have described a leaky competing accumulator (LCA) model of lexical decision that can be attached as a response/decision module to *any* computational model of word recognition. First developed by Usher and McClelland (2001) as an alternative to the diffusion model, the LCA model possesses WORD and NONWORD decision nodes that are linked via mutually inhibitory connections. The former is driven by lexical input (i.e., lexical activity), whereas the latter is driven by a constant value minus the lexical input, and the model makes word or nonword decisions on the basis of noisy, leaky, and competing information accumulating over time. Although the full architecture of the model is beyond the scope of this article, Dufau et al. (2012) have demonstrated that the LCA model successfully simulates mean RTs and RT distributions for a number of benchmark experiments. Like the diffusion model (Ratcliff et al., 2004),

the LCA model is designed to be a stand-alone decision-making module, and its performance is constrained by the processing assumptions of the word recognition model it is attached to. When Ratcliff, Thapar, Smith, and McKoon (2005) fit data from a number of experiments to the diffusion model and the LCA model, they found no qualitative basis for selecting one model over the other, although the diffusion model, compared with the LCA model, fit the data better.

In sum, it is clear that the foregoing models are driven by the presence (or more precisely, absence) of a lexical input. Despite the sophistication of newer modeling approaches (e.g., diffusion model, LCA model), they are predicated on the simple premise that a single process drives lexical decision to nonwords. Specifically, nonwords that elicit more lexical activity (e.g., legal nonwords such as FLIRP) should be responded to more slowly than nonwords that elicit less lexical activity (e.g., illegal nonwords such as BRNTA). However, the specific influence and relative importance of the different dimensions that contribute to that signal remain unclear. Moreover, experimental findings where participants take *less* time to respond to nonwords associated with more lexical activity (e.g., nonwords derived from high-frequency base words) will be challenging for single-process models without positing an additional verification process (Perea et al., 2005).

Individual Difference in Nonword Decision

Despite compelling evidence that variation in reading skill predicts word recognition evidence (see Yap, Balota, Sibley, & Ratcliff, 2012, for a review), empirical studies and computational models have traditionally focused on group-level performance. Yap et al., (2012), using trial-level data from the ELP, examined individual differences in speeded pronunciation and lexical decision performance for over 1,200 participants. In addition to detecting considerable within- and between-session reliability in the data, their analyses also revealed a number of relationships between vocabulary knowledge and sensitivity to underlying lexical dimensions in word recognition performance. For example, participants with more vocabulary knowledge were associated with attenuated sensitivity to lexical characteristics, and were able to accumulate evidence about the lexicality of a letter string at a more rapid rate (i.e., steeper drift rates in the diffusion model). Yap, Tse, and Balota (2009) have suggested that readers' vocabulary knowledge could reflect the integrity of underlying lexical representations, and the extent to which readers are likely to rely on relatively more automatic processing mechanisms. To our knowledge, there is no work examining the reliability of nonword responses or the impact of individual differences on nonword processing.

Hence, in addition to identifying the effects of different variables on nonword lexical decision times, the secondary goal of the present study was to extend the work by Yap et al., (2012) by examining the role of individual differences in nonword responses. To what extent do individual differences in vocabulary knowledge systematically modulate different aspects of nonword decision performance? Like Yap et al., (2012), we examine trial-level RT data both at the level of mean RTs and at the level of underlying RT distributional characteristics (see Balota & Yap, 2011, for a review). Specifically, distributions of individual participants will be fitted to the ex-Gaussian distribution (Ratcliff, 1979), a theoretical distribution that approximates positively skewed empirical distributions well. An ex-Gaussian distribution contains three parameters; μ and σ, respectively, reflect the mean and standard deviation of the Gaussian distribution, whereas τ reflects the mean and standard deviation of the exponential distribution. Changes in μ are consistent with distributional shifting, whereas changes in τ reflect changes in the tail of the distribution. τ effects are of particular theoretical interest, because some researchers have suggested that lapses in attentional control are in some tasks related to modulations in the tail of the distribution (see also Tse, Balota, Yap, Duchek, & McCabe, 2010). In this light, it is interesting that the relationship between vocabulary knowledge and word recognition RTs is predominantly mediated by the slow tail of the distribution (Yap et al., 2012).

As an important adjunct to the ex-Gaussian parameters, individuals' RT distributional data will also be fitted to the diffusion model (Ratcliff et al., 2004), a process-oriented model of binary decision that likens lexical decision to the accumulation of noisy information over time from a starting point (z) toward one of two decision boundaries, word (a) or nonword (0). The mean rate at which information is accumulated is reflected by drift rate (v), whereas nondecision processes (encoding and response execution) are collectively captured by T_{er}. Vocabulary knowledge has also been found to be systematically related to diffusion model parameters. Specifically, participants with high-vocabulary knowledge are associated with steeper drift rates, more liberal response criteria, and a shorter nondecision component (Yap et al., 2012). Collectively, the results of these analyses will address an important empirical gap in the literature and help inform emerging lexical processing models that take individual differences into account.

Method

Data Set

All analyses reported in this article are based on archival trial-level data from the ELP (see Balota et al., 2007, for a full description of the data set). The analyses focused on the 819 participants who provided data for the LDT. These participants, who were all native English speakers, were recruited from six universities (see Table 1 of Balota et al., 2007, for descriptive statistics of participant demographics) that included private and public institutions situated in the Midwest, Northeast, and Southeast portions of the United States. Data were collected over two sessions on different days, separated by no more than 1 week. Across both sessions, each participant received approximately 3,374 lexical decision trials. Nonword stimuli were created by changing letters in word targets to produce pronounceable nonwords that did not sound like real words. Additional demographic information collected included vocabulary knowledge scores, based on the 40-item Vocabulary subscale of the Shipley Institute of Living Scale (Shipley, 1940), and circadian rhythm, based on the Morningness-Eveningness Questionnaire scores (Horne & Ostberg, 1976).

Predictor Variables

Length. Number of letters.

Orthographic neighborhood size. The number of words that can be obtained by changing a single letter in the target word, while holding the other letters constant (Coltheart et al., 1977).

Table 1
Means and Standard Deviations for Full Set of Predictors and Dependent Variables Explored in the Item-Level Regression Analyses

Variable	*M*	*SD*
Nonword LDT RT (*z*-score)	−.07	.35
Nonword LDT accuracy	.87	.13
Number of letters	7.88	2.40
Orthographic neighborhood size	1.91	2.27
Levenshtein orthographic distance	2.92	1.01
Average base-word frequency	6.25	2.21
Average base-word number of syllables	2.52	1.08
Number of affixes	1.01	.64

Note. N = 36,985. LDT = lexical decision task; RT = response time.

Levenshtein orthographic distance. *Levenshtein orthographic distance* (Yarkoni et al., 2008) refers to the average number of insertions, deletions, and substitutions needed to convert a nonword into its 20 closest word neighbors in the ELP. The Levenshtein measure is particularly useful for quantifying the orthographic distinctiveness of long letter strings, because these typically have few or no orthographic neighbors.

Average base-word frequency. This was obtained by first identifying all neighbors at Edit Distance 1 (i.e., one insertion, deletion, or substitution) from the target nonword. The average log HAL frequencies (Lund & Burgess, 1996) of these words was then computed.

Average base-word number of syllables. This was obtained by first identifying all neighbors at Edit Distance 1 from the target nonword. The average number of syllables of these words was then computed.

Number of affixes. This was provided by the Affix Detector program (Muncer et al., 2013), which counts the number of morphemelike elements (i.e., prefixes and suffixes) in a nonword, based on a comprehensive list of affixes listed in Fudge (1984).

Results

We first excluded incorrect trials and trials with response latencies faster than 200 ms or slower than 3,000 ms. For the remaining correct trials, RTs more than 2.5 standard deviations away from each participant's mean were also identified as outliers. For the RT analyses, data trimming procedures removed 15.7% (12.7% errors; 3% RT outliers) of the trials. For ease of exposition, we first describe the effects of different lexical variables on nonword decision performance, followed by reliability analyses, before considering the relationships between participants' vocabulary knowledge, nonword decision performance, and sensitivity to different lexical dimensions. Table 1 presents descriptive statistics for the predictors and measures, whereas Table 2 presents the intercorrelations between the predictors and dependent variables being examined.

Analysis 1: Regression Analyses on Nonword Decision Performance

Item-level regression analyses were conducted on the 36,985 nonwords that possessed values for all relevant predictors and the two dependent measures, *z*-scored LDT RT and accuracy. As different participants received different subsets of nonwords, *z*-scored RTs were used to control for variation in processing speed across participants (Faust, Balota, Spieler, & Ferraro, 1999). There were a number of noteworthy observations. First, our six predictors accounted for almost 40% of the variance in nonword RTs (see Table 3). To provide a frame of reference, the analogous predictors accounted for approximately 61% of the variance in word RTs (see Table 3), in line with other word megastudies (e.g., Yap & Balota, 2009). Second, number of letters was by far the strongest predictor of nonword lexical decision performance; responses were slower and less accurate to longer nonwords. Third, nonwords that were *less* orthographically distinct, as reflected by having more orthographic neighbors or closer Levenshtein neighbors, were also responded to more slowly and less accurately. Fourth, participants found it more difficult to reject nonwords associated with more syllables and affixes.

Finally, and somewhat surprisingly, there was a small but reliable facilitatory effect of base-word frequency, wherein nonwords derived from *higher* frequency base words were rejected more quickly and accurately. Given the potential theoretical importance of this pattern, it was important to ensure that the facilitatory effect of base-word frequency was not simply an artifact of the regression analysis (e.g., through a misspecification of the functional form of other effects in the model). To address this, we conducted additional regression analyses (with the same six predictors) on subsets of the full data set in which we respectively restricted the range of number of letters and Levenshtein orthographic distance (a measure of neighborhood characteristics). This afforded the

Table 2
Correlations Between Full Set of Predictors and Dependent Variables Explored in the Item-Level Regression Analyses

Variable	1	2	3	4	5	6	7	8
1. Nonword LDT RT (*z*-score)	—	−.469***	.608***	−.212***	.518***	−.246***	.520***	.290***
2. Nonword LDT accuracy		—	−.039***	−.130***	.058***	−.008	−.024***	−.105***
3. Number of letters			—	−.507***	.887***	−.383***	.829***	.347***
4. Orthographic neighborhood size				—	−.525***	.286***	−.440***	−.228***
5. Levenshtein orthographic distance					—	−.428***	.768***	.209***
6. Average base-word frequency						—	−.297***	−.079***
7. Average base-word number of syllables							—	.379***
8. Number of affixes								—

Note. LDT = lexical decision task; RT = response time.
*** $p < .001$.

Table 3
Standardized RT and Accuracy Regression Coefficients of the Item-Level Regression Analyses

Predictor variable	Nonwords (N = 36,985) RT	Nonwords (N = 36,985) Accuracy
Number of letters	.624***	−.397***
Orthographic neighborhood size	.139***	−.175***
Levenshtein orthographic distance	−.029**	.349***
Avg. base-word frequency	−.040***	.033***
Avg. base-word number of syllables	.038***	−.001
Number of affixes	.094***	−.077***
R^2	.392***	.066***

Predictor variable	Words (N = 38,467) RT	Words (N = 38,467) Accuracy
Number of letters	−.055***	.485***
Orthographic neighborhood size	.068***	−.011*
Levenshtein orthographic distance	.299***	−.229***
Frequency	−.510***	.550***
Number of syllables	.299***	−.243***
Number of morphemes	−.053***	.170***
R^2	.608***	.330***

Note. RT = response time; Avg. = Average.
* $p < .05$. ** $p < .01$. *** $p < .001$.

creation of data sets that were more homogenous with respect to number of letters (leftmost panel of Figure 1) and Levenshtein orthographic distance (center panel of Figure 1). For both dimensions, base-word frequency effects remained reliably facilitatory for two of the three subsets, indicating that this intriguing pattern is not simply an artifact of model misspecification. To ascertain why facilitatory base-word effects were not reliable for *all* subsets, we also partitioned subsets by RTs (rightmost panel of Figure 1). This revealed that facilitatory effects were most evident in the slowest trials, consistent with the idea that these effects reflect a relatively late-frequency-ordered verification procedure (Paap et al., 1982).

In addition to the main effects explored above, we selected a number of theoretically important interactions to test (a) the Number of Letters × Base-Word Frequency interaction, (b) the Orthographic Neighborhood Size × Base-Word Frequency interaction, (c) the Base-Word Number of Syllables × Base-Word Frequency interaction, and (d) the Number of Affixes × Base-Word Frequency interaction. Regression interactions were explored using the method described in Cohen, Cohen, West, and Aiken (2003), whereby variables of interest and other control variables were first entered, followed by the interaction term in the following step. The four interactions listed above were all statistically reliable ($ps < .05$), and the respective simple slopes underlying each interaction are presented in Figure 2.

The results of the interaction analyses are easy to summarize. We observed that the inhibitory effects of number of letters, orthographic neighborhood size, base-word number of syllables, and number of affixes decreased as the nonword's base-word frequency increased. This pattern of results broadly mirrors the analogous interactions for responses to words (Yap & Balota, 2009). Specifically, for words, the influence of number of letters (inhibitory), orthographic neighborhood size (facilitatory), and base-word number of syllables (inhibitory) decreases as word frequency increases (see also Andrews 1989, 1992; Jared & Seidenberg, 1990; Weekes, 1997). The finding that qualitatively similar interactions are seen for words and nonwords is consistent with the idea that common lexical processes are recruited to generate a signal for word/nonword discrimination (Holcomb et al., 2002).

Analysis 2: Reliability Analyses

Trials for each participant were first partitioned into Session 1 (S1) trials, Session 2 (S2) trials, odd-numbered trials, and even-numbered trials; trial number reflects the order in which trials were presented. Using split-half correlations, comparing S1 with S2 trials allows the assessment of between-session reliability, whereas comparing odd- with even-numbered trials allows the assessment of within-session reliability.

For each participant, we then computed the mean and standard deviation of RTs, along with ex-Gaussian (μ, σ, τ) and diffusion model parameters for S1 trials, S2 trials, odd-numbered trials, and even-numbered trials. Ex-Gaussian parameters were estimated for

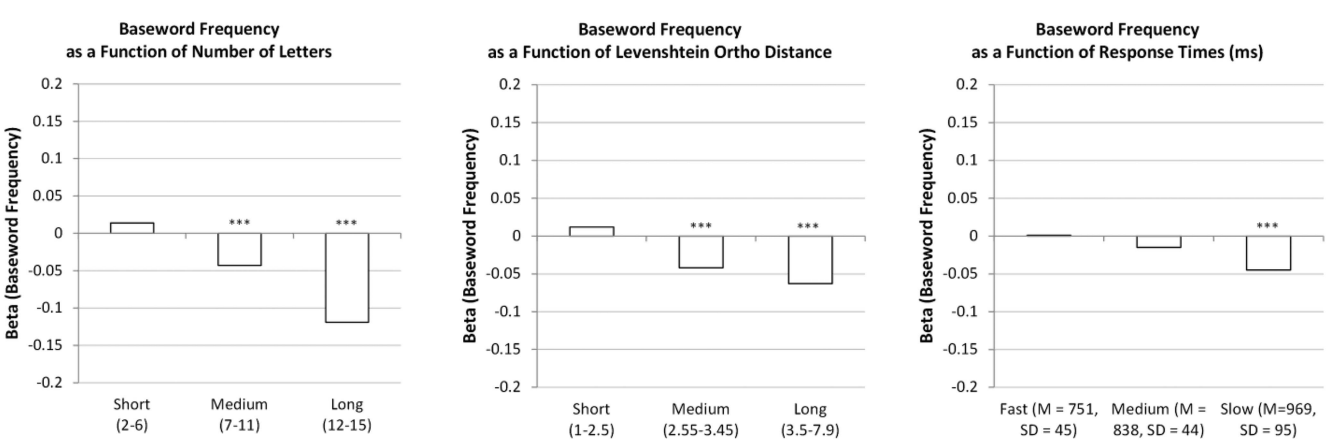

Figure 1. Base-word frequency effects as a function of number of letters (left), Levenshtein orthographic (Ortho) distance (center), and response times (right). *** $p < .001$.

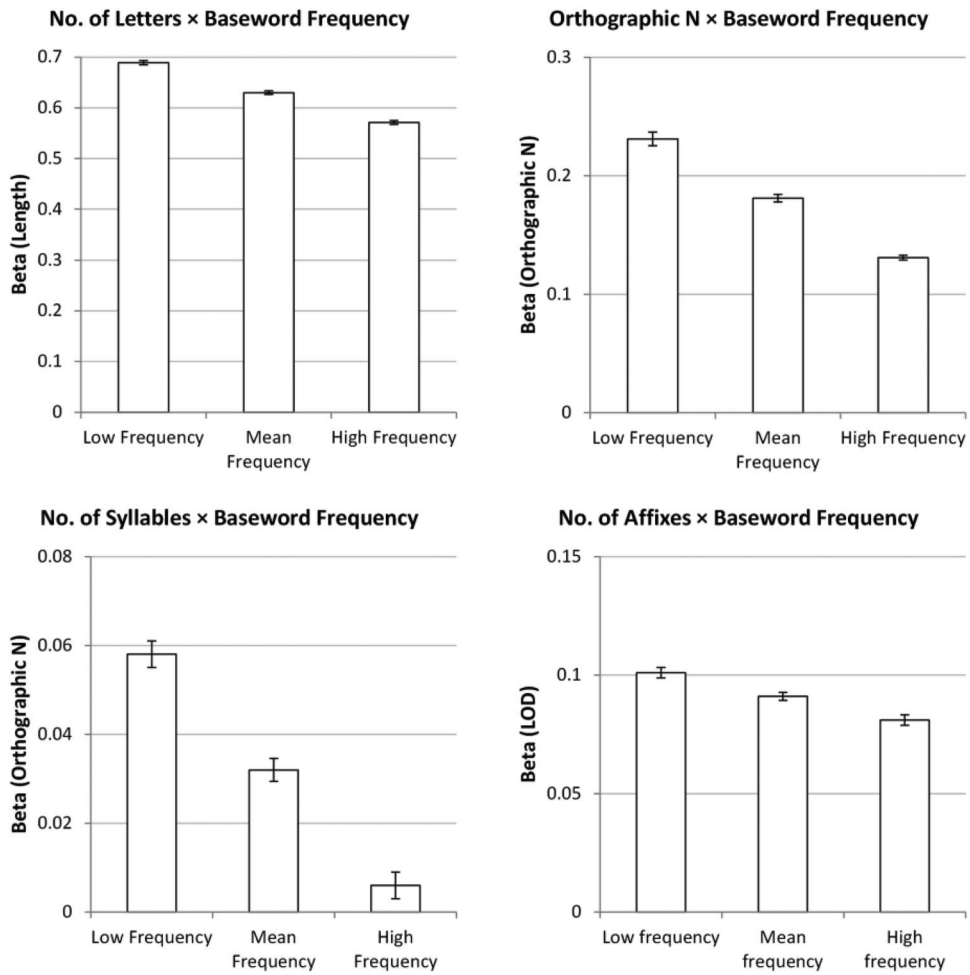

Figure 2. Interactions between base-word frequency and number (No.) of letters (upper left), base-word number of syllables (lower left), orthographic neighborhood size (upper right), and number of affixes (lower right). The bars represent the standardized regression coefficient for each variable as a function of low-, medium-, and high-frequency words. Error bars denote standard errors.

each participant using continuous maximum likelihood estimation in R (R Development Core Team, 2004). Using Nelder and Mead's (1965) simplex algorithm, negative log-likelihood functions were minimized in the R statistics package (Speckman & Rouder, 2004), with all fits successfully converging within 500 iterations. The diffusion model parameters were estimated simultaneously by fitting each participant's data to the model. The data for each participant were composed of the .1, .3, .5, .7, and .9 quantile RTs for correct and error responses, along with the corresponding accuracy values. A general SIMPLEX minimization routine was then used that adjusted the parameters of the model in order to minimize the value of chi-square (Ratcliff & Tuerlinckx, 2002). Table 4 presents the mean latency, standard deviation, ex-Gaussian parameters, and diffusion model parameters for nonword responses, as a function of trial type (overall, S1, S2, odd-numbered trials, even-numbered trials).

Table 5 presents the Pearson correlations between each individual's nonword responses in S1 and S2 trials, and between odd- and even-numbered trials, for mean RT, standard deviation, ex-Gaussian parameters, and diffusion model parameters. The high correlations (all rs $\geq$ .87) between odd- and even-numbered trials indicate substantial within-session reliability for the mean, standard deviation, and ex-Gaussian parameters. Within-session reliability was also high for most of the diffusion model parameters. When between-session reliability was assessed, correlations were also relatively high for the mean and standard deviation (rs $\geq$ .87), ex-Gaussian parameters (rs from .39 to .77), and diffusion model parameters (rs from .39 to .72). These results support the idea that readers are associated with a specific RT distributional signature that applies to both word (see Yap et al., 2012) and nonword responses. Importantly, because no participant saw the same nonword twice, this signature holds up across different testing sessions and different sets of stimuli.

As shown in Table 5, it is also noteworthy that there is evidence for relatively high test–retest stability in drift rate and the tail (τ) of the RT distribution (see Yap et al., 2012, for a replication of this pattern with word responses), consistent with the proposal that these two parameters serve as important markers of individual

Table 4
Means, Standard Deviations, Ex-Gaussian Parameters, and Diffusion Model Parameters as a Function of Task and Trial Type

Variable	Overall	Session 1	Session 2	Odd	Even
M	840	855	825	840	840
SD	231	228	224	231	230
μ	607	625	602	607	607
σ	81	82	78	80	81
τ	233	230	223	233	233
a	0.169	0.171	0.165	0.169	0.17
z	0.093	0.095	0.091	0.093	0.094
T_{er}	0.495	0.506	0.49	0.496	0.497
η	0.165	0.171	0.167	0.167	0.169
s_z	0.118	0.114	0.114	0.118	0.119
s_t	0.169	0.175	0.165	0.169	0.171
v_{word}	0.223	0.224	0.23	0.225	0.229
$v_{nonword}$	−0.255	−0.256	−0.261	−0.256	−0.257

Note. Lexical decision (N = 780).

differences (Ratcliff, Thapar, & McKoon, 2010; Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007; Tse et al., 2010). Specifically, individuals associated with a lower drift rate or larger τ could be seen as less efficient lexical processors who rely more heavily on controlled word recognition processes.

Having established the reliability of RT distributional characteristics, we now turn to the reliability of individuals' sensitivity to different lexical characteristics. For example, if a participant produces large length effects on S1, will he or she also produce large length effects on S2? To address this, we conducted multiple regression analyses at the level of individual participants and looked at the reliability of the regression coefficients (Balota & Chumbley, 1984; Lorch & Myers, 1990). One might be concerned that the participant-level regression analyses were conducted on different sets of items, because participants were presented with different sublists of the full set of words in the ELP. However, the counterbalancing procedure ensured that the means, standard deviations, and ranges of different variables were similar across the different sublists.

For each participant, we examined the effects of the six predictors (number of letters, orthographic neighborhood size, Levenshtein orthographic distance, average base-word frequency, average base-word number of syllables, and number of affixes) explored in the overall item-level analyses. Figure 3 presents the distributions of standardized regression coefficients across participants as a function of lexical variable. First, note the substantial variability in the magnitude of effects produced by participants. For example, although the majority of participants produced positive regression coefficients for the number of letters effect, indicating longer latencies for longer nonwords, the coefficients were normally distributed. Second, the direction and relative magnitudes of participant-level effects were generally consistent with the item-level effects reported earlier. That is, number of letters was the best predictor, followed by orthographic neighborhood size, then by the other predictors. Generally, nonwords that were recognized more slowly were longer, less orthographically distinctive, possessed more affixes, and were derived from lower frequency base words with more syllables.

Turning to the reliability analyses, Table 6 presents the Pearson correlations between S1 and S2 trials, and between odd- and even-numbered trials, for the regression coefficients corresponding to the six effects of interest. With the exception of the effect of number of affixes, within- and between-session measures of reliability were generally moderate to high (.26 < rs < .46) for nonword lexical decision performance. Effects of structural properties (number of letters, orthographic neighborhood size) seem to be more reliable than those reflecting base-word properties (word frequency, number of syllables).

Analysis 3: Vocabulary Knowledge, Diffusion Model Parameters, and Nonword Decision Performance

We now turn to the relationship between vocabulary knowledge and nonword decision performance. As discussed earlier, the size of a reader's vocabulary could reflect the integrity of underlying lexical representations, and the extent to which readers rely on relatively more automatic processing mechanisms (Yap et al., 2009). Figure 4 presents the scatterplots between vocabulary knowledge (as assessed by the number of correct responses on the Shipley, 1940, Vocabulary subscale) and nonword decision RTs and accuracy, after excluding 71 (8.7%) participants who were more than 1.5 interquartile ranges below the lower quartile on a box plot. Vocabulary knowledge was negatively correlated with nonword RTs ($r = -.292$, $p < .001$) and positively correlated with accuracy ($r = .615$, $p < .001$). In addition, vocabulary knowledge was slightly more strongly correlated with τ (tail of the distribution; $r = -.276$, $p < .001$) than with μ (leading edge of distribution, $r = -.237$, $p < .001$).

We next consider the correlations between vocabulary knowledge and the central diffusion parameters. Vocabulary knowledge was negatively correlated with boundary separation (a) ($r = -.076$, $p = .037$), nondecision time (T_{er}) ($r = -.228$, $p < .001$), and nonword drift rate (v) ($r = -.432$, $p < .001$). In other words, higher vocabulary knowledge participants were setting a more liberal

Table 5
Correlations Between Session 1 (S1) and Session 2 (S2) Parameters, and Odd- and Even-Numbered Trial Parameters

	Lexical decision	
Variable	S1-S2	Odd-Even
Mean RT	.866***	.998***
SD	.938***	.994***
μ	.587***	.951***
σ	.392***	.878***
τ	.767***	.949***
a	.693***	.890***
z	.724***	.895***
T_{er}	.704***	.906***
η	.386***	.634***
s_z	.385***	.675***
s_t	.407***	.522***
v_{word}	.662***	.823***
$v_{nonword}$	.635***	.788***

Note. With the exception of mean RT and the diffusion model parameters, overall mean RT was partialed from each correlation. RT = response time.
*** $p < .001$.

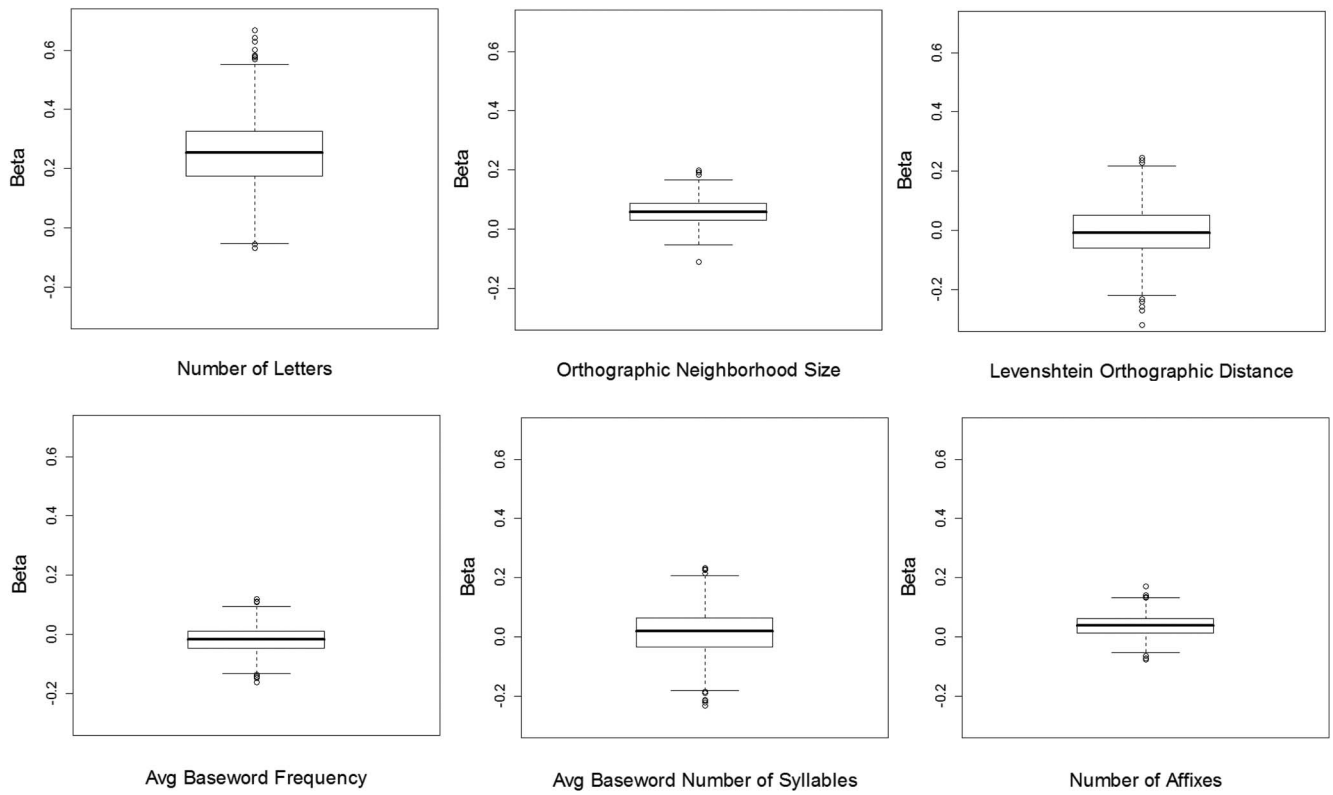

Figure 3. Distributions of standardized regression coefficients across participants as a function of lexical variable. Avg = Average.

decision criteria, had a faster nondecision component, and could accumulate information at a more rapid rate. However, it is worth noting that the relationship between vocabulary knowledge and boundary separation is relatively modest compared with vocabulary knowledge's correlations with the other two parameters (i.e., drift rate and nondecision time), mirroring the analyses conducted on word data (Yap et al., 2012).

Analysis 4: Individual Differences in Effects of Lexical Variables

The literature examining the relationship between lexical processing fluency (as reflected by print exposure or vocabulary knowledge) suggests that skilled lexical processors are less influenced by stimulus properties such as frequency (Chateau & Jared, 2000) and length (Butler & Hains, 1979). Yap et al. (2012) also reported that the influence of lexical variables was attenuated as vocabulary knowledge increased, although this trend was more clearly seen in speeded pronunciation, compared with lexical decision, performance.

Table 6
Correlations Between Session 1 (S1) and Session 2 (S2) Participant-Level Effects, and Odd- and Even-Numbered Trial Participant-Level Effects

Predictor variable	Lexical decision	
	S1-S2	Odd-Even
Number of letters	.437***	.464***
Orthographic neighborhood size	.333***	.269***
Levenshtein orthographic distance	.267***	.259***
Avg. base-word frequency	.341***	.381***
Avg. base-word number of syllables	.348***	.366***
Number of affixes	.128***	.167***

Note. Avg. = Average.
*** $p < .001$.

Table 7 presents the correlations between participant-level standardized regression coefficients and vocabulary knowledge and diffusion model parameters. It is important to point out that the correlations between the regression coefficients and the other variables cannot simply be attributed to processing speed, because these coefficients were standardized. Vocabulary knowledge was reliably correlated with every effect we examined. Figure 5 presents scatterplots describing the relationships between vocabulary knowledge and sensitivity to the different underlying lexical dimensions. For example, high-vocabulary knowledge participants were less influenced by the inhibitory effects of number of letters. Specifically, vocabulary knowledge *increased*, and individual-level regression coefficients for number of letters became *less* negative and became closer to zero. Likewise, high-vocabulary participants were less sensitive to the inhibitory effect of Levenshtein orthographic distance (i.e., slower responses to nonwords with closer Levenshtein neighbors). At the same time, participants with more vocabulary knowledge were *more* sensitive to the inhibitory effects of orthographic neighborhood size, base-word number of syllables, and base-word number of affixes; they were also facilitated by base-word frequency to a greater extent.

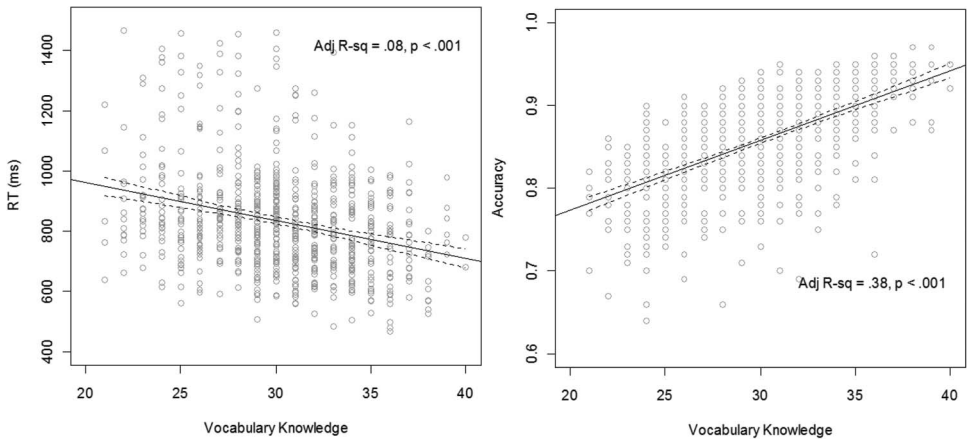

Figure 4. Scatterplots (with 95% confidence intervals) between vocabulary knowledge and nonword response times (RTs; left) and accuracy (right). Adj. R-sq = adjusted R^2.

Turning to the diffusion model parameters, we observed that participants who set more liberal response criteria (as reflected by lower values on *a*, the boundary separation parameter) were associated with larger inhibitory effects of orthographic neighborhood size ($r = -.12$), but smaller facilitatory effects of base-word frequency ($r = -.40$) and inhibitory effects of base-word number of syllables ($r = .21$) (see Figure 6). Participants who produced a shorter nondecision component (i.e., lower values on T_{er}) were associated with larger inhibitory effects of orthographic neighborhood size ($r = -.17$), but smaller inhibitory effects of number of letters ($r = .20$) and facilitatory effects of base-word frequency ($r = -.15$) (see Figure 7). Finally, and most importantly, participants who produced steeper nonword drift rates were associated with larger inhibitory effects of orthographic neighborhood size ($r = -.39$), facilitatory effects of base-word frequency ($r = .10$), inhibitory effects of base-word number of syllables ($r = -.21$), and inhibitory effects of number of affixes ($r = -.14$). These participants also produced smaller inhibitory effects of number of letters ($r = .23$) and inhibitory effects of Levenshtein orthographic distance ($r = -.14$) (see Figure 8).

For ease of understanding, Table 8 summarizes and organizes the results described above. Upward pointing arrows denote increased sensitivity to the influence of a variable, whereas downward pointing arrows denote decreased sensitivity. It is noteworthy that higher vocabulary knowledge and steeper drift rates are related to participant-level effects in the same manner. This suggests that participants who are more skilled in lexical processing (as reflected by more vocabulary knowledge and steeper nonword drift rates) are less sensitive to characteristics such as word length and Levenshtein orthographic distance, but are more sensitive to characteristics such as orthographic neighborhood size, number of affixes, base-word frequency, and number of syllables.

General Discussion

Using trial-level data from the ELP, the present study is the first large-scale investigation of influences on and individual differences in nonword decision performance. There were a number of noteworthy observations. First, the six predictors of interest (number of letters, orthographic neighborhood size, Levenshtein orthographic distance, average base-word frequency, average base-word number of syllables, number of affixes) successfully accounted for 39.2% and 6.6% of the item-level variance in response latencies and accuracy, respectively. Second, like responses to words (Yap et al., 2012), responses to nonwords showed relatively high between- and within-session reliability across different sets of stimuli, with respect to an individual's mean RT, RT distributional characteristics, diffusion model parameters, and sensitivity to underlying psycholinguistic dimensions. Third, vocabulary knowledge and diffusion model parameters were reliably and systemat-

Table 7
Correlations Between Participant-Level Standardized Regression Coefficients, Vocabulary Knowledge, and Diffusion Model Parameters

Lexical effect	Vocabulary knowledge	a	T_{er}	$v_{nonword}$
Number of letters	$-.355^{***}$	$-.057$	$.199^{***}$	$.232^{***}$
Orthographic neighborhood size	$.283^{***}$	$-.123^{***}$	$-.174^{***}$	$-.394^{***}$
Levenshtein orthographic distance	$.191^{***}$	$.032$	$-.006$	$-.141^{***}$
Avg. base-word frequency	$-.208^{***}$	$-.397^{***}$	$-.151^{***}$	$.100^{**}$
Avg. base-word number of syllables	$.295^{***}$	$.206^{***}$	$-.036$	$-.206^{***}$
Number of affixes	$.196^{***}$	$.040$	$.003$	$-.138^{***}$

Note. Avg. = Average.
$^{**}p < .01$. $^{***}p < .001$.

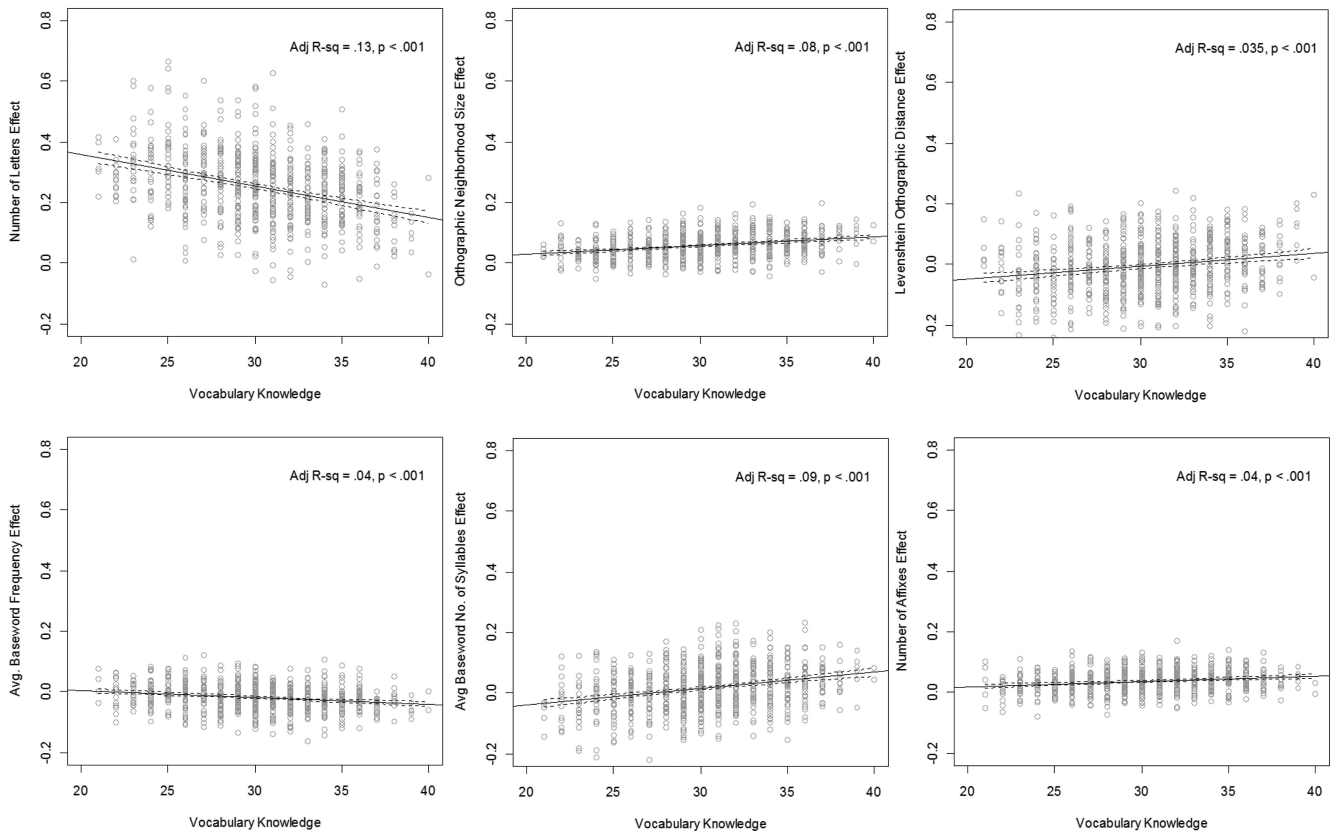

Figure 5. Scatterplots (with 95% confidence intervals) between vocabulary knowledge and participant-level effects. Adj. R-sq = adjusted R^2; Avg. = Average; No. = Number.

ically related to participant-level effects for the different predictors. We now turn to a discussion of these findings.

Item-Level Effects in Nonword Decision Performance

Our item-level regression analyses indicate that the six targeted predictors were able to account for a substantial proportion (39.2%) of the variance in nonword lexical decision latencies. Specifically, across all participants, RT was positively correlated with number of letters, number of orthographic neighbors, average base-word number of syllables, and number of affixes, and negatively correlated with Levenshtein orthographic distance and average base-word frequency. More importantly, although previous studies have assessed the effects of these variables separately, the present study allowed us to evaluate the *relative* predictive power of these factors on a very large, well-characterized set of nonwords. At the same time, these analyses can potentially shed light on extant empirical controversies (e.g., the influence of base-word frequency on nonword decision times).

It is clear that number of letters was, by far, the strongest predictor of nonword RTs; specifically, longer nonwords were rejected more slowly and less accurately. This could be seen as consistent with nonword processing being mediated by serial processes,[3] such as the sublexical mechanism in Coltheart et al.'s (2001) DRC model, which assembles pronunciations for nonwords grapheme by grapheme. Other factors that could contribute to longer latencies for longer nonwords include the decrease in visual acuity beyond the fixation point, the increased likelihood of refixations, and the increased overlap between nonwords and real words for longer nonwords (see New et al., 2006, for more discussion).

In addition to the influence of number of letters, nonwords with more orthographic neighbors and closer Levenshtein neighbors were responded to more slowly and less accurately, consistent with the notion that such nonwords elicit more global lexical activity and therefore take more time to reject (Coltheart et al., 2001). This finding can be accommodated by the LCA model (Dufau et al., 2012), in which the strength of the input to the nonword response node is inversely proportional to the strength of the lexical input. Of course, one might also argue that such results are consistent with decision mechanisms, which emphasize global familiarity-based signals to drive lexical decision performance (e.g., Balota & Chumbley, 1984; Ratcliff et al., 2004).

Lexical decisions were affected by the syllabic and morphological characteristics of the nonword stimuli. For example, nonwords

[3] Although a parallel processing mechanism can show sensitivity to word length (e.g., Simulation 3 of Plaut, McClelland, Seidenberg, & Patterson, 1996), such length effects are far too subtle (<1% of unique variance accounted for in model latencies) to be reconciled with the present pattern of results.

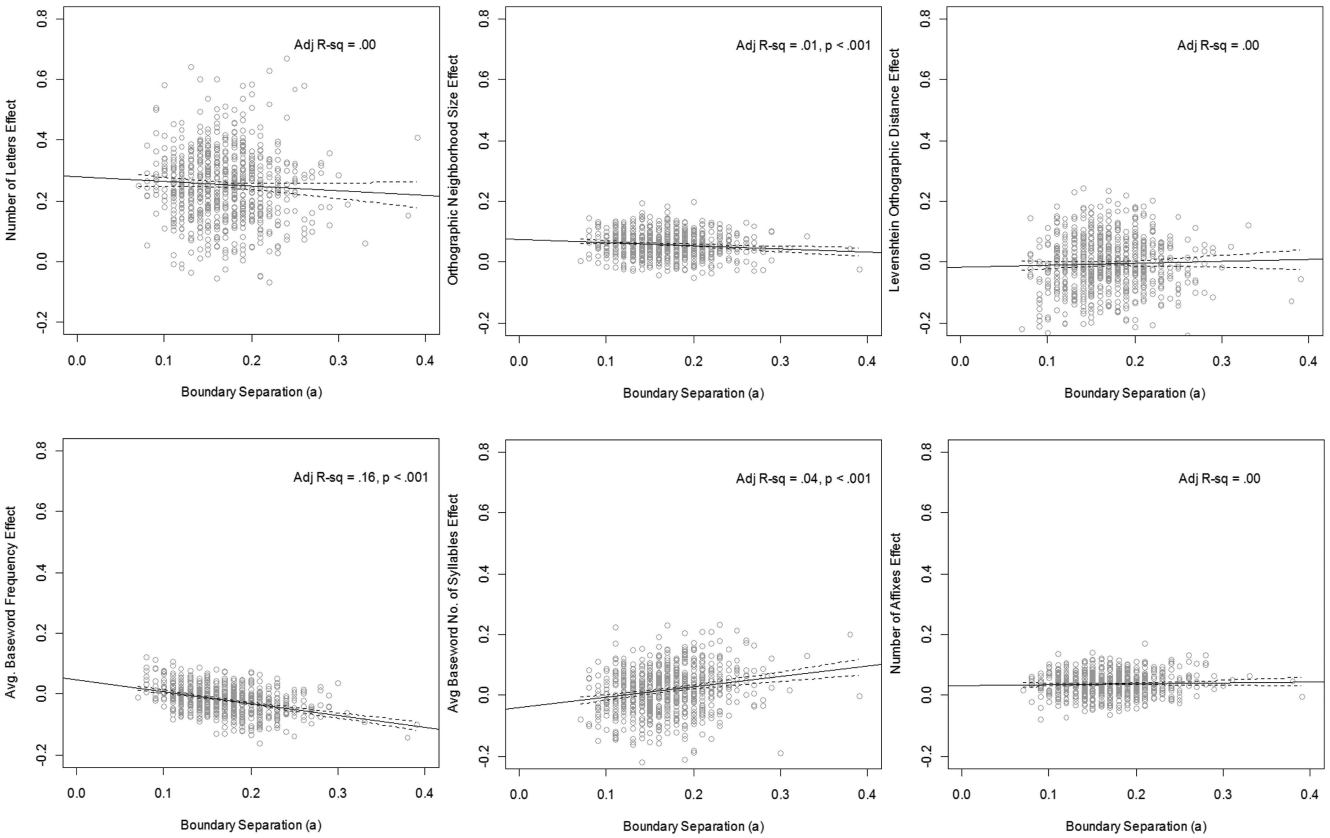

Figure 6. Scatterplots (with 95% confidence intervals) between boundary separation and participant-level effects. Adj. R-sq = adjusted R^2; Avg. = Average; No. = Number.

with more morphemic elements (as reflected by morphological prefixes and suffixes) took more time to reject. This is consistent with the study by Muncer et al. (2013) and supports the view that morphologically complex stimuli are decomposed at an early, relatively automatic stage in visual word recognition (Rastle & Davis, 2008; Rueckl & Aicher, 2008). Our data also indicate that nonwords with more syllables were rejected more slowly, a finding that fits well with the idea that the syllable is one of the sublexical codes mediating lexical access (see Yap & Balota, 2009, for more discussion). It is worth noting that syllabic length effects, although reliable in a very large data set, are relatively subtle, explaining why findings in the literature (e.g., Muncer & Knight, 2012) have been mixed.

Interestingly, we observed shorter latencies for nonwords derived from *higher* frequency base words; this trend was more pronounced for items that yielded longer RTs (see Figure 1). The effects described in the previous paragraph can be accommodated by activation-based perspectives; as the amount of lexical activity associated with a nonword *increases*, the strength of the input to the nonword node *decreases* (Dufau et al., 2012), hence lengthening lexical decision times (see also Balota & Chumbley, 1984). However, if one assumes that nonwords derived from high-frequency base words elicit more lexical activity, it is unclear how faster RTs for such nonwords can be accommodated. One possible solution is to augment an activation-based mechanism with a verification component (e.g., Ziegler et al., 2001). Specifically, high-frequency, compared with low-frequency, base words have more stable orthographic representations, allowing readers to verify more quickly deviations between a nonword and its respective base word (Paap et al., 1982).

Finally, the present study is the first to explore the joint effects of variables on nonword decision times. Briefly, we found that base-word frequency reliably moderated the influences of number of letters, orthographic neighborhood size, base-word number of syllables, and number of affixes; as baseword frequency increased, the effects of the above-mentioned variables decreased. Our results attest to qualitatively similar interactions for responses to words and nonwords and fit nicely with the perspective that common lexical processes are engaged to generate a signal for word/nonword discrimination.

In sum, the present study provides a finer-grained characterization of how nonword responses are modulated by various stimulus characteristics by exploring the relative and unique influence of a comprehensive array of variables. While providing additional support for better established findings (e.g., inhibitory effects of orthographic neighborhood density and number of letters), our results also shed light on effects that have been more equivocal (e.g., effects of base-word frequency and number of syllables). At the same time, these results yield a useful set of benchmark findings for informing computational models. As Perry et al. (2010) have argued, a strong correlation between model and human latencies is necessary but not suffi-

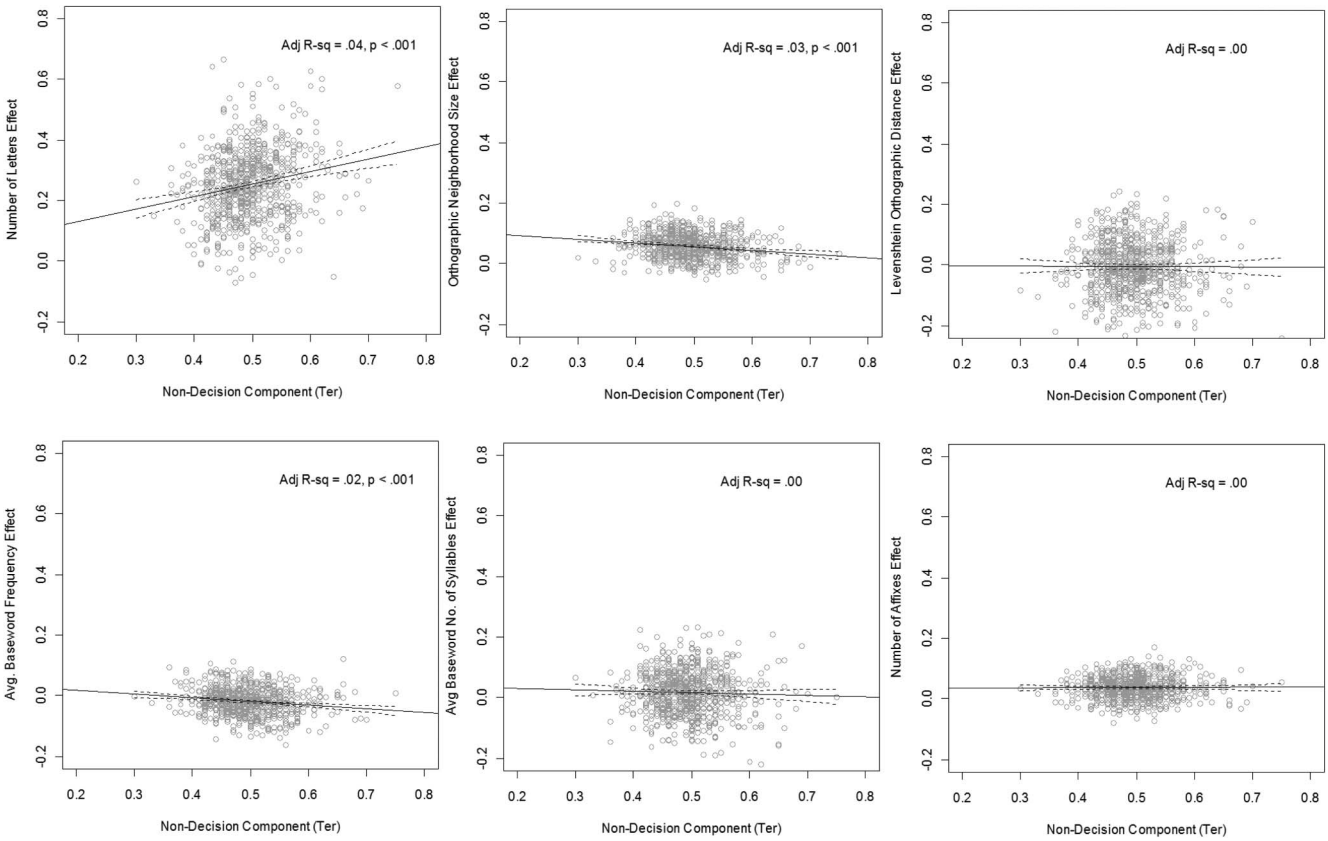

Figure 7. Scatterplots (with 95% confidence intervals) between nondecision time and participant-level effects. Adj. R-sq = adjusted R^2; Avg. = Average; No. = Number.

cient. It is also important for a computational model to correctly reproduce the *relative* proportions of variance accounted for by different variables in human data.

Variability and Reliability of Nonword Decision Performance

In line with the word data described in Yap et al., (2012), the present analyses support the variability and reliability of lexical decision performance of nonwords (see Figure 3). Across distinct sets of nonwords, we found relatively high within-session and between-session reliabilities with respect to mean RTs, standard deviations, ex-Gaussian parameters, and diffusion model parameters (see Table 5). Participants also demonstrated within- and between-session stability in their sensitivity to underlying lexical characteristics (see Table 6). These results indicate that participants carry with them a stable RT distributional and processing profile that applies to both word *and* nonword responses and that the variability in nonword decision performance reflects systematic and meaningful individual differences rather than just measurement noise. This provides further assurance that nonword response times data help provide meaningful and complementary insights into the lexical processing architecture.

Individual Differences and Nonword Decision Performance

The present study is the first to systematically explore the relationship between individual differences and nonword decision performance. First, consider the influence of vocabulary knowledge, which has been argued to tap the integrity of underlying lexical representations (Yap et al., 2009). Unsurprisingly, participants who possessed higher vocabulary knowledge were faster and more accurate in rejecting nonwords (see Figure 4). When we used the diffusion model to explore this relationship in a more differentiated manner, we observed that the better performance for the higher vocabulary knowledge participants was mediated by a more liberal decision criteria, a faster nondecision component, and a more rapid rate of accumulation of evidence (i.e., drift rate) about the nonword stimulus. Of these three parameters, vocabulary knowledge was most strongly correlated with drift rate, consistent with Ratcliff et al.'s (2010) demonstration that IQ is more strongly related to drift rate than to any other parameter in the diffusion model (see also Ratcliff, Thapar, & McKoon, 2011).

The close link between vocabulary knowledge and drift rate is also evident in Table 8, where these two variables predicted participant-level effects in the same way. The results broadly indicate that skilled lexical processors, who are associated with more vocabulary knowledge and steeper nonword drift rates, are

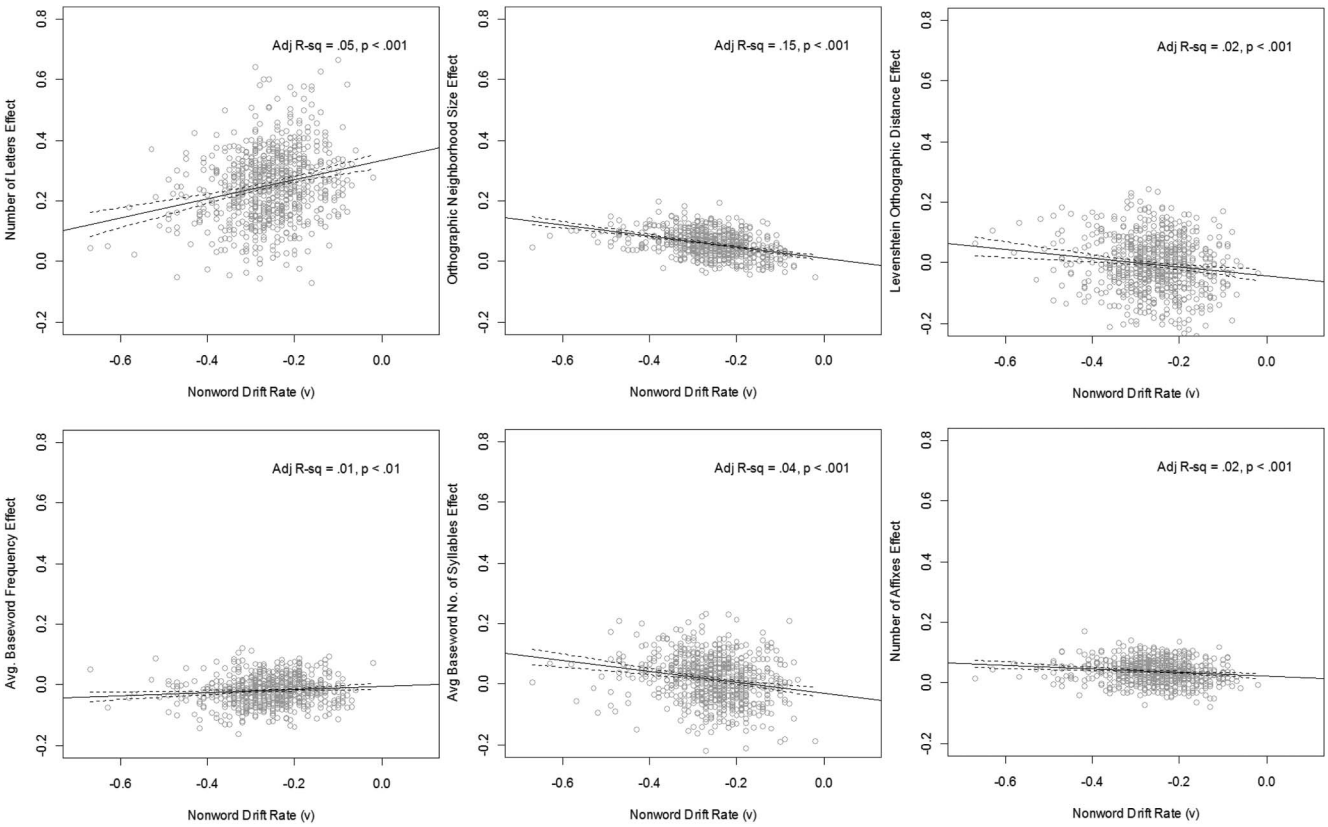

Figure 8. Scatterplots (with 95% confidence intervals) between nonword drift rate and participant-level effects. Adj. R-sq = adjusted R^2; Avg = Average; No. = Number.

less sensitive to characteristics such as word length and Levenshtein orthographic distance, but are more sensitive to characteristics such as orthographic neighborhood size, number of affixes, and base-word frequency and number of syllables. Although our data support the idea that fluent lexical processors can handle long letter strings more efficiently (Butler & Hains, 1979), it is not the case that skilled lexical processors are simply influenced to a lesser extent by *all* kinds of stimulus properties. Instead, we have a dissociation wherein highly skilled participants are *less* sensitive to some dimensions but are *more* sensitive to others.

These results seem most consistent with the notion of a flexible lexical processor (Balota, Paul, & Spieler, 1999; Balota & Yap, 2006), in which attentional control systems modulate the processing pathways between orthography, phonology, and semantics, so as to optimize performance on any given task. Although number of letters was closely matched between words (M = 8, SD = 2.46) and nonwords (M = 8, SD = 2.46) in the ELP, nonwords (M = 1.78, SD = 2.22) possessed more orthographic neighbors than words (M = 1.29, SD = 2.73), making orthographic neighborhood size a viable dimension for discriminating between words and

Table 8
Relationships Between Participant-Level Regression Coefficients and Vocabulary Knowledge, Boundary Separation, Nondecision Component, and Drift Rate

Lexical effect	Higher vocabulary knowledge	Lower boundary separation (a)	Shorter nondecision component (T_{er})	Steeper drift rate ($v_{nonword}$)
Number of letters (inhibition)	↓		↓	↓
Orthographic neighborhood size (inhibition)	↑	↑	↑	↑
Levenshtein orthographic distance (inhibition)	↓			↓
Avg. base-word frequency (facilitation)	↑	↓	↓	↑
Avg. base-word number of syllables (inhibition)	↑	↓		↑
Number of affixes (inhibition)	↑			↑

Note. Upward pointing arrows denote increased sensitivity to the influence of a variable while downward pointing arrows denote decreased sensitivity. Avg. = average.

nonwords. Hence, highly skilled lexical processors may emphasize the processing of density-based information that aid in such word/nonword discrimination.

These skilled participants are also more likely to carry out syllabic and morphological decomposition of nonword stimuli, and more likely to rely on procedures that verify the spellings of nonwords (Ziegler et al., 2001). To test this, we carried out a median split of participants based on vocabulary knowledge and compared high- and low-vocabulary knowledge participants on their sensitivity with base-word number of syllables, number of affixes, and base-word frequency. High-vocabulary knowledge participants, compared with their low-vocabulary knowledge counterparts, were higher on effects of base-word number of syllables (.04 vs. .00), number of affixes (.04 vs. .03), and base-word frequency (−.03 vs. −.01).

Implications for Models of Nonword Lexical Decision

The findings we report represent a well-specified set of benchmarks for constraining models of word recognition and lexical decision. Not surprisingly, word recognition models have emphasized speeded performance for words, and there has been relatively little consideration of the mechanisms that drive nonword responses. The major models that accommodate lexical decision, such as the DRC model (Coltheart et al., 2001) and the MROM (Grainger & Jacobs, 1996), are predicated on the assumption that nonword responses are produced after a variable temporal deadline that is modulated by global lexical activity. As discussed, this proposal has been criticized (see Ratcliff et al., 2004; Wagenmakers et al., 2008). More sophisticated approaches based on the diffusion model (Ratcliff et al., 2004) or the Bayesian reader model (Norris, 2006, 2009) provide a better fit for nonword RT data, but the solutions proposed by the latter perspectives are less straightforward (see Dufau et al., 2012, for more discussion). Recently, Dufau and colleagues have also described a hybrid model of nonword lexical decision that implements a variable deadline via the accumulation of noisy, leaky, and competing information over time.

The present results help provide additional constraints for any framework (e.g., LCA, diffusion, Bayesian reader) that drives lexical decisions via a single process. Although extant single-mechanism perspectives should be able to accommodate inhibitory effects of neighborhood density in a straightforward manner, it is unclear whether they predict an influence of morphological and syllabic structure, or whether they can produce facilitatory effects of base-word frequency (i.e., shorter latencies for nonwords derived from high-frequency base words) without invoking an additional verification-based mechanism. Of course, extant models are also generally mute on how diffusion model parameters are modulated by stimulus characteristics or how individual differences in lexical processing proficiency might moderate responses to nonwords. These are intriguing questions that can be pursued in future research.

Limitations and Concluding Remarks

In the present study, we examined the influence of various measures on approximately 37,000 nonwords in the ELP for over 800 participants. In spite of considerable *across*-participant variability in nonword decision performance, *within*-participant stability was reassuringly high. Individual differences in vocabulary knowledge were also systematically and interestingly related to an individual's sensitivity to the different underlying dimensions in a nonword. At a more profound level, the relationships between vocabulary knowledge/drift rate and sensitivity to different lexical characteristics are pertinent to the question of how changes in reading ability are associated with changes in the grain size that people use when reading. There are several empirical lines of evidence that converge on this conclusion (e.g., Ziegler & Goswami, 2005), and the present individual differences findings potentially help inform the issue of what it means to be a good reader. Related to this, the analyses of individual differences also provide an important goal for computational models to aim for. That is, they should be able to explain how learning is producing the present effects through changing representations and processes within the lexical system.

A number of questions are worth exploring in future work. One, the ELP nonwords were created by replacing one or two letters in a corresponding target word, while ensuring that the nonword remained pronounceable. Nonwords can also be created by using computer programs (e.g., Wuggy; Keuleers & Brysbaert, 2010) that match generated nonwords to the target word in terms of subsyllabic structure and transition frequencies. It is likely that the procedure used to create nonwords may have some impact on the observed results. Of course, this is related to the types of information participants bring online in the lexical decision process, which will be in part based on the overlap of features of the words and nonwords. Two, due to the size of the ELP data set, each nonword's "frequency" and number of syllables were estimated by computing the average frequency and number of syllables from the nonword's closest Levenshtein word neighbors. To examine base-word effects more precisely in future work, one could focus on nonwords that are unambiguously derived from a specific word (e.g., voltage → VOLT*I*GE). Finally, and in a similar vein, the literature has emphasized base-word properties such as word frequency, but it is also possible to examine the semantic properties of the base word, such as imageability, number of features, and the like (see Pexman, 2012, for a review), and to assess whether semantics play a role in nonword decision.

References

Allen, P. A., McNeal, M., & Kvak, D. (1992). Perhaps the lexicon is coded as a function of word frequency. *Journal of Memory and Language, 31,* 826–844. doi:10.1016/0749-596X(92)90041-U

Álvarez, C. J., Carreiras, M., & Taft, M. (2001). Syllables and morphemes: Contrasting frequency effects in Spanish. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 545–555. doi:10.1037/0278-7393.27.2.545

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 802–814. doi:10.1037/0278-7393.15.5.802

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 234–254. doi:10.1037/0278-7393.18.2.234

Andrews, S. (1996). Lexical retrieval and selection processes: Effects of transposed-letter confusability. *Journal of Memory and Language, 35,* 775–800. doi:10.1006/jmla.1996.0040

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. doi:10.1037/0096-1523.10.3.340

Balota, D. A., Cortese, M. J., Sergent-Marshall, S., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133,* 283–316. doi:10.1037/0096-3445.133.2.283

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, United Kingdom: Psychology Press.

Balota, D. A., & Yap, M. J. (2006). 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* (pp. 229–258). New York, NY: Psychology Press.

Balota, D. A., & Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry: The power of response time distributional analyses. *Current Directions in Psychological Science, 20,* 160–166. doi:10.1177/0963721411408885

Balota, D., Yap, M. J., & Cortese, M. J. (2006). Visual word recognition: The journey from features to meaning (A travel update). In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 285–375). Amsterdam, the Netherlands: Academic Press. doi:10.1016/B978-012369374-7/50010-9

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39,* 445–459. doi:10.3758/BF03193014

Balota, D. A., Yap, M. J., Hutchison, K. A., & Cortese, M. J. (2012). Megastudies: What do millions (or so) of trials tell us about lexical processing? In J. S. Adelman (Ed.), *Visual word recognition volume 1: Models and methods, orthography and phonology* (pp. 90–115). Hove, United Kingdom: Psychology Press.

Butler, B., & Hains, S. (1979). Individual differences in word recognition latency. *Memory & Cognition, 7,* 68–76. doi:10.3758/BF03197587

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition, 28,* 297–332. doi:10.1016/0010-0277(88)90017-0

Carreiras, M., Álvarez, C. J., & de Vega, M. (1993). Syllable frequency and visual word recognition in Spanish. *Journal of Memory and Language, 32,* 766–780. doi:10.1006/jmla.1993.1038

Chateau, D., & Jared, D. (2000). Exposure to print and word recognition processes. *Memory & Cognition, 28,* 143–153. doi:10.3758/BF03211582

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

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: LEA.

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. doi:10.1037/0033-295X.108.1.204

Conrad, M., & Jacobs, A. M. (2004). Replicating syllable frequency effects in Spanish in German: One more challenge to computational models of visual word recognition. *Language and Cognitive Processes, 19,* 369–390. doi:10.1080/01690960344000224

Crepaldi, D., Rastle, K., & Davis, C. J. (2010). Morphemes in their place: Evidence for position-specific identification of suffixes. *Memory & Cognition, 38,* 312–321. doi:10.3758/MC.38.3.312

Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological Review, 117,* 713–758. doi:10.1037/a0019738

Duchek, J. M., & Neely, J. H. (1989). A dissociative word-frequency X levels-of-processing interaction in episodic recognition and lexical decision tasks. *Memory & Cognition, 17,* 148–162. doi:10.3758/BF03197065

Dufau, S., Grainger, J., & Ziegler, J. C. (2012). How to say "no" to a nonword: A leaky competing accumulator model of lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 1117–1128. doi:10.1037/a0026948

Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin, 125,* 777–799. doi:10.1037/0033-2909.125.6.777

Ferrand, L., & New, B. (2003). Syllabic length effects in visual word recognition and naming. *Acta Psychologica, 113,* 167–183. doi:10.1016/S0001-6918(03)00031-3

Fudge, E. (1984). *English word stress*. London, England: Allen & Unwin.

Gomez, P., Ratcliff, R., & Perea, M. (2008). The overlap model: A model of letter position coding. *Psychological Review, 115,* 577–600. doi:10.1037/a0012667

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103,* 518–565. doi:10.1037/0033-295X.103.3.518

Grainger, J., & Van Heuven, W. (2003). Modeling letter position coding in printed word perception. In P. Bonin (Ed.), *The mental lexicon* (pp. 1–24). New York, NY: Nova Science Publishers.

Holcomb, P. J., Grainger, J., & O'Rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience, 14,* 938–950. doi:10.1162/089892902760191153

Horne, J. A., & Ostberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology, 4,* 97–110.

Jared, D., & Seidenberg, M. S. (1990). Naming multisyllabic words. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 92–105. doi:10.1037/0096-1523.16.1.92

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods, 42,* 627–633. doi:10.3758/BRM.42.3.627

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods, 44,* 287–304. doi:10.3758/s13428-011-0118-4

Lorch, R. F., Jr., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research: A comparison of three different methods. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 149–157. doi:10.1037/0278-7393.16.1.149

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers, 28,* 203–208. doi:10.3758/BF03204766

Muncer, S. J., & Knight, D. C. (2012). The bigram trough hypothesis and the syllable number effect in lexical decision. *Quarterly Journal of Experimental Psychology, 65,* 2221–2230. doi:10.1080/17470218.2012.697176

Muncer, S. J., Knight, D. C., & Adams, J. W. (2013). Lexical decision and the number of morphemes and affixes. *Scandinavian Journal of Psychology, 54,* 349–352. doi:10.1111/sjop.12064

Nelder, J. A., & Mead, R. (1965). A simplex algorithm for function minimization. *Computer Journal, 7,* 308–313. doi:10.1093/comjnl/7.4.308

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Re-examining word length effects in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review, 13,* 45–52. doi:10.3758/BF03193811

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review, 113,* 327–357. doi:10.1037/0033-295X.113.2.327

Norris, D. (2009). Putting it all together: A unified account of word recognition and reaction-time distributions. *Psychological Review, 116,* 207–219. doi:10.1037/a0014259

Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation-verification model for letter and word recognition: The word superiority effect. *Psychological Review, 89,* 573–594. doi:10.1037/0033-295X.89.5.573

Perea, M., & Carreiras, M. (1998). Effects of syllable frequency and syllable neighborhood frequency in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 134–144. doi:10.1037/0096-1523.24.1.134

Perea, M., Rosa, E., & Gómez, C. (2005). The frequency effect for pseudowords in the lexical decision task. *Perception & Psychophysics, 67,* 301–314. doi:10.3758/BF03206493

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review, 114,* 273–315. doi:10.1037/0033-295X.114.2.273

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. *Cognitive Psychology, 61,* 106–151. doi:10.1016/j.cogpsych.2010.04.001

Pexman, P. M. (2012). Meaning-based influences on visual word recognition. In J. S. Adelman (Ed.), *Visual word recognition volume 2: Meaning and context, individuals and development* (pp. 24–43). Hove, United Kingdom: Psychology Press.

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. doi:10.1037/0033-295X.103.1.56

Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes, 23,* 942–971. doi:10.1080/01690960802069730

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86,* 446–461. doi:10.1037/0033-2909.86.3.446

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review, 111,* 159–182. doi:10.1037/0033-295X.111.1.159

Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology, 60,* 127–157.

Ratcliff, R., Thapar, A., & McKoon, G. (2011). Effects of aging and IQ on item and associative memory. *Journal of Experimental Psychology: General, 140,* 464–487. doi:10.1037/a0023810

Ratcliff, R., Thapar, A., Smith, P. L., & McKoon, G. (2005). Aging and response times: A comparison of sequential sampling models. In J. Duncan, P. McLeod, & L. Phillips (Eds.), *Measuring the mind: Speed, control, and age* (pp. 3–32). Oxford, England: Oxford University Press. doi:10.1093/acprof:oso/9780198566427.003.0001

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review, 9,* 438–481. doi:10.3758/BF03196302

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

Rueckl, J. G., & Aicher, K. (2008). Are CORNER and BROTHER morphologically complex? Not in the long term. *Language and Cognitive Processes, 23,* 972–1001. doi:10.1080/01690960802211027

Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H.-M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General, 136,* 414–429. doi:10.1037/0096-3445.136.3.414

Seidenberg, M. S., & Plaut, D. C. (1998). Evaluating word-reading models at the item level: Matching the grain of theory and data. *Psychological Science, 9,* 234–237. doi:10.1111/1467-9280.00046

Shipley, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *Journal of Psychology, 9,* 371–377. doi:10.1080/00223980.1940.9917704

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. doi:10.3758/BF03196613

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science, 8,* 411–416. doi:10.1111/j.1467-9280.1997.tb00453.x

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior, 14,* 638–647. doi:10.1016/S0022-5371(75)80051-X

Tse, C.-S., Balota, D. A., Yap, M. J., Duchek, J. M., & McCabe, D. P. (2010). Effects of healthy aging and early-stage dementia of the Alzheimer's type on components of response time distributions in three attention tasks. *Neuropsychology, 24,* 300–315. doi:10.1037/a0018274

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review, 108,* 550–592. doi:10.1037/0033-295X.108.3.550

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language, 58,* 140–159. doi:10.1016/j.jml.2007.04.006

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 50A* 439–456. doi:10.1080/713755710

Whaley, C. (1978). Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior, 17,* 143–154. doi:10.1016/S0022-5371(78)90110-X

Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin & Review, 8,* 221–243. doi:10.3758/BF03196158

Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language, 60,* 502–529. doi:10.1016/j.jml.2009.02.001

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognitions: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Human Perception and Performance, 38,* 53–79. doi:10.1037/a0024177

Yap, M. J., Tse, C.-S., & Balota, D. A. (2009). Individual differences in the joint effects of semantic priming and word frequency: The role of lexical integrity. *Journal of Memory and Language, 61,* 303–325. doi:10.1016/j.jml.2009.07.001

Yarkoni, T., Balota, D. A., & Yap, M. J. (2008). Beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review, 15,* 971–979. doi:10.3758/PBR.15.5.971

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131,* 3–29.

Ziegler, J. C., Jacobs, A. M., & Klüppel, D. (2001). Pseudohomophone effects in lexical decision: Still a challenge for current word recognition models. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 547–559. doi:10.1037/0096-1523.27.3.547

Received April 14, 2014
Revision received July 9, 2014
Accepted July 24, 2014 ■