Journal of Experimental Psychology:
Human Perception and Performance
2012, Vol. 38, No. 1, 53–79

0096-1523/11/$12.00 DOI: 10.1037/a0024177

Individual Differences in Visual Word Recognition: Insights From the English Lexicon Project

Melvin J. Yap
National University of Singapore

David A. Balota
Washington University in St. Louis

Daragh E. Sibley
Haskins Laboratories, New Haven, Connecticut

Roger Ratcliff
Ohio State University

Empirical work and models of visual word recognition have traditionally focused on group-level performance. Despite the emphasis on the prototypical reader, there is clear evidence that variation in reading skill modulates word recognition performance. In the present study, we examined differences among individuals who contributed to the English Lexicon Project (http://elexicon.wustl.edu), an online behavioral database containing nearly 4 million word recognition (speeded pronunciation and lexical decision) trials from over 1,200 participants. We observed considerable within- and between-session reliability across distinct sets of items, in terms of overall mean response time (RT), RT distributional characteristics, diffusion model parameters (Ratcliff, Gomez, & McKoon, 2004), and sensitivity to underlying lexical dimensions. This indicates reliably detectable individual differences in word recognition performance. In addition, higher vocabulary knowledge was associated with faster, more accurate word recognition performance, attenuated sensitivity to stimuli characteristics, and more efficient accumulation of information. Finally, in contrast to suggestions in the literature, we did not find evidence that individuals were trading-off their utilization of lexical and nonlexical information.

Keywords: visual word recognition, individual differences, response time distributional analysis, ex-Gaussian analysis, diffusion model

How do people recognize visually presented words? The ability to read is one of the towering achievements of human civilization and cognition, with word recognition being a focus of inquiry since Cattell's (1886) pioneering work. Insights from this field have informed a host of domains, including reading acquisition (e.g., Adams, 1990; Perfetti, 1994), literacy instruction (e.g., Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001), automatic versus attentional processing (e.g., Neely, 1977; Posner & Snyder, 1975), the neural correlates of lexical processing (e.g., Petersen, Fox, Posner, Mintun, & Raichle, 1989), social cognition (e.g., Bargh, Chen, & Burrows, 1996; Devine, 1989), and cognitive aging (e.g., Balota & Ferraro, 1996).

Although converging procedures have been developed to study the processes underlying word recognition, the tasks most often used to study isolated word recognition are speeded pronunciation and lexical decision. In speeded pronunciation, words (and sometimes, nonwords, e.g., *flirp*) are presented, and participants have to read these aloud. In lexical decision, participants discriminate between words and nonwords, typically with a button press. Using these tasks to provide a window into the processes involved in accessing and using lexical representations (Seidenberg, 1990), researchers have identified many properties of words that influence word recognition speed and accuracy. For example, more commonly encountered words (e.g., *world*) are recognized faster than less common words (e.g., *glitch*); this is known as the *word-frequency effect*. Similarly, words with relatively few syllables (Ferrand & New, 2003) and letters (New, Ferrand, Pallier, & Brysbaert, 2006) tend to be recognized faster and more accurately. Words also differ with respect to the number of orthographic neighbors (Coltheart, Davelaar, Jonasson, & Besner, 1977) they possess where orthographic neighbors refer to the number of words one can produce by changing a single letter in the target word (e.g., *sand*'s neighbors include *band, send, said,* and *sank*). The general finding is that words with many neighbors are recognized faster than words with few neighbors (see Andrews, 1997, for a review).

Most studies of word recognition have focused on group-level data that average across participants. Likewise, computational

This article was published Online First July 4, 2011.

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

This paper was based on parts of a doctoral study completed by M. J. Yap under the supervision of D. A. Balota, and was supported by NIA Grants PO1 AGO3991 and PO1 AGO26276, as well as NSF Grant BCS 0001801 to D. A. Balota. We thank Mark Steyvers, Cyrus Shaoul, and Chris Westbury for providing us with semantic measures. We also thank James Adelman, Colin Davis, Sachiko Kinoshita, and Eric-Jan Wagenmakers for very helpful comments on an earlier version of this article.

Correspondence concerning this article should be addressed to Melvin J. Yap, National University of Singapore, Department of Psychology, Faculty of Arts and Social Sciences, Block AS4, #02-07, 9 Arts Link, Singapore 117570. E-mail: melvin@nus.edu.sg

models of word recognition have rarely considered individual differences among skilled adult readers (but see Zevin & Seidenberg, 2006, for an exception). These efforts emphasize the characterization of the “prototypical” reader. However, an emerging body of work indicates substantial individual differences among readers. These studies have demonstrated processing differences related to reading skill (e.g., Ashby, Rayner, & Clifton, 2005; Chateau & Jared, 2000; Jared, Levy, & Rayner, 1999; Schilling, Rayner, & Chumbley, 1998; Unsworth & Pexman, 2003; Yap, Tse, & Balota, 2009), and it is possible that some of the inconsistencies in the literature may be driven by individual differences among participants (see Yap et al., 2009, for an example).

The present study harnesses the power of the English Lexicon Project (ELP; Balota et al., 2007) to explore individual differences in word recognition performance. The ELP is an online repository (http://elexicon.wustl.edu) of behavioral (speeded pronunciation and lexical decision) measures for 40,481 words. More important, the ELP contains trial-level data from 1,289 (470 for speeded pronunciation and 819 for lexical decision) participants across six universities. Each participant contributed approximately 2,500 pronunciation latencies or 3,400 lexical decision latencies. Data were collected over two sessions, separated at most by 1 week. Hence, the ELP contains data for almost four million word recognition trials, sampled across a large population of participants, making it the most comprehensive resource of its kind presently available. Examining data from such a large sample should provide an excellent way to estimate individual differences in word recognition behavior. Finally, vocabulary knowledge (Shipley, 1940; Zachary, 1992) was also measured for each participant, to estimate the integrity of that participant’s lexical representations (see Yap et al., 2009).

In the present study we use the ELP to address three broad questions. First, can subtle individual differences be reliably detected, and if so, how great is the variation in these differences? Second, what are the systematic relationships between an individual’s reading skill, word recognition efficiency, and sensitivity to lexical dimensions? For example, if a reader produces a large effect of word frequency, will he or she also produce a large effect of orthographic neighborhood size? Finally, there is evidence that moving beyond measures of central tendency when analyzing response time (RT) data provides finer grained insights into individual differences in performance (see Balota & Yap, 2011, for a review). Specifically, how is variability in the leading edge and tail of RT distributions systematically related to other outcomes in word recognition? Do individuals who produce large word-frequency effects produce more skewing in their RT distribution, and are they less efficient in accumulating evidence about the stimulus, as reflected by diffusion model (Ratcliff, 1978) parameters? The present investigation represents the first attempt to answer these interrelated questions in a unified manner, using a very large, well-characterized set of words and participants. We first consider the issue of reliability in word recognition performance before turning to a selective review of studies exploring individual differences in word recognition.

Are Word Recognition Measures Reliable?

Reliability is usually assessed in the domain of psychological testing, in which one determines if a measure of an individual is similar at different points in time (Anastasi & Urbina, 1997). Establishing test–retest reliability is a critical prerequisite for developing valid measures of individual differences in intelligence, aptitude, personality, interests, values, and attitudes. We find it surprising that the reliability of RT measures is rarely evaluated; instead, researchers’ confidence in a measure increases as effects replicate across different studies (Goodwin, 2009). Indeed, classic measures of cognitive performance, such as the Stroop task, have relatively low test–retest reliability (see Lowe & Rabbitt, 1998). Low reliability not only limits the sensitivity of an individual measure but also circumscribes the extent to which that measure can be expected to correlate with other measures (Lowe & Rabbitt, 1998). Furthermore, and more critical for our purposes, without first establishing reliability, it is unclear whether variability among readers reflects meaningful individual differences or measurement noise.

Hence, if one is going to make inferences about individual differences in word recognition processing, then it is critical that one has some estimate of the stability of pronunciation and lexical decision performance across time. For example, consider a participant who does the lexical-decision task on two separate occasions, with a different randomly selected large set of stimuli in each session. Will the participant exhibit qualitatively similar behavior at both points in time? Despite the intuitive importance of this question, no study, to our knowledge, has systematically explored the long-range stability or alternate-form reliability of different word recognition measures. The study that comes closest to doing this is one by Schilling et al. (1998), who examined performance on the speeded pronunciation, lexical decision, and eye tracking tasks. Each participant was randomly assigned to two of the three tasks, which were conducted at least 1 week apart. Although they could not estimate within-task test–retest reliability, they observed significant pairwise correlations in mean RT and frequency effects across tasks, attesting to the cross-task stability of these effects.

Hence, the first objective of the paper is to explore the within-task reliability of word recognition performance, reflected by both within-session reliability (assessed by correlation between performance for odd vs. even items) and between-session reliability (assessed by correlation between performance on different sets of stimuli at two time points across multiple days). At the simplest level, the reliability of mean RTs can be evaluated, which addresses the stability of overall processing speed. Also, beyond mean RTs, one can also examine the reliability of the RT distributional characteristics for each participant, and ascertain whether participants carry with them a distributional profile above and beyond simple mean speed. We can also estimate the reliability of different parameters from a computational model (the diffusion model) of lexical decision performance. Finally, and perhaps most important, we can assess the extent to which there is stability in participants’ sensitivity to theoretically important variables such as word frequency, length, and neighborhood size.

Vocabulary Knowledge and Word Recognition Performance

Reading skill is a complex, multifaceted construct encompassing word decoding efficiency (Perfetti, 1983, 1985), orthographic and phonological processing skill (Stanovich & West, 1989), and sentence comprehension (Chateau & Jared, 2000). Reading skill is related to *vocabulary knowledge* (i.e., knowledge of word forms and meaning) and *print exposure* (i.e., the amount of text a person reads), both of which strongly modulate word recognition performance. For example, participants with more vocabulary knowledge (as reflected by the number of words they know the meaning of) pronounce words faster (Butler & Hains, 1979). There is also evidence that performance on the lexical decision and speeded pronunciation tasks predicts vocabulary size (Katz et al., in press). Likewise, participants with more exposure to print (as reflected by the number of author names they recognize) are faster and more accurate on various lexical processing tasks, including speeded pronunciation, lexical decision, semantic classification, and nonword naming (Chateau & Jared, 2000; Lewellen, Goldinger, Pisoni, & Greene, 1993). Finally, vocabulary knowledge and print exposure have been shown to be statistically related, even when general cognitive ability is controlled for (Stanovich & Cunningham, 1992).

Studies have explored how reading skill, as reflected by vocabulary knowledge or print exposure, might interact with how stimulus characteristics affect recognition. Consider the general hypothesis that as readers acquire more experience with words, they become increasingly reliant on automatic lexical processing mechanisms (LaBerge & Samuels, 1974; Stanovich, 1980). In this case, it is possible that as automatic mechanisms develop, words may be less influenced by lexical characteristics (Butler & Hains, 1979) and even context (Stanovich & West, 1983; Yap et al., 2009). This hypothesis seems consistent with extant data. Specifically, readers with more vocabulary knowledge or print exposure are less sensitive to a number of lexical dimensions, including number of letters (Butler & Hains, 1979), word frequency (Chateau & Jared, 2000), and orthographic neighborhood size (Chateau & Jared, 2000).

These results are compatible with many other findings. For example, faster readers produce smaller regularity and lexicality effects in speeded pronunciation (P. Brown, Lupker, & Colombo, 1994), readers with more years of education produce smaller frequency effects in lexical decision (Tainturier, Tremblay, & Lecours, 1992), and rapid word decoders produce smaller frequency effects in speeded pronunciation (Schilling et al., 1998; Seidenberg, 1985). It is worth noting that all the studies listed above, with the exception of Butler and Hains (1979), did not control for overall processing speed when computing effects. That is, effects were based on actual, not standardized, RTs. We revisit this issue in greater depth later.

However, there are reports that readers who are more skilled do not always produce smaller effects. For example, Lewellen et al. (1993) compared low- and high-ability readers, who were contrasted based on their subjective familiarity ratings of words, print exposure, and vocabulary knowledge. They then measured participants' frequency and neighborhood density effects on speeded pronunciation and lexical decision. We find it interesting that frequency and neighborhood size effects were of the same size for low- and high-ability readers in both tasks, prompting the conclusion that such effects did not interact with reading skill. These findings obviously conflict with Chateau and Jared (2000), who reported smaller frequency effects in their low-print-exposure participants. Sears, Siakaluk, Chow, and Buchanan (2008) suggested that the discrepancy might be due to Lewellen et al. using legal nonwords (e.g., *brone*) as distracters in their lexical-decision task, and Chateau and Jared using *pseudohomophones* (i.e., nonwords that sound like real words, e.g., *brane*). Indeed, Sears et al. demonstrated that the moderating effects of print exposure on frequency and neighborhood size effects were reliable only when pseudohomophones, but not legal nonwords, were used. Consistent with this, Yap, Balota, Tse, and Besner (2008), who also used pseudohomophone distracters, observed smaller frequency effects for participants with more vocabulary knowledge.

According to Sears et al. (2008), low-print-exposure participants, compared to their high-print-exposure counterparts, possessed less efficient orthographic processing skills, which resulted in slower and less accurate recognition of words with lower quality representations, such as low-frequency words. However, when discriminating between words and legal nonwords, lower exposure participants could compensate for their relatively inefficient orthographic processing by switching to phonological processing, resulting in similar size word-frequency effects for low- and high-print-exposure participants. This phonological strategy is not viable in the pseudohomophone condition because phonology, by definition, cannot be used to discriminate between words and pseudohomophones. As a result, the low-print-exposure participants were slowed down, particularly for the low-frequency words, which exaggerated their word-frequency effect.

Although Sears et al. (2008) made a plausible attempt to reconcile the conflicting findings, there is a simpler alternative. In both Chateau and Jared (2000) and Sears et al.'s pseudohomophone condition, low-print-exposure participants were reliably slower on the lexical-decision task. In contrast, in Lewellen et al. (1993) and Sears et al.'s legal nonword condition, there was no significant (by-participant) effect of print exposure on lexical decision latencies. Faust, Balota, Spieler, and Ferraro (1999) pointed out that Group $\times$ Treatment interactions are difficult to interpret when the groups are not matched on overall latency (see also Cerella, 1990, 1991; Salthouse, 1985). Specifically, because a participant's overall processing time is positively correlated with the magnitude of his or her effect (i.e., larger effects for slower participants), slower participants could spuriously produce larger effects due to slowing. Hence, the larger frequency effects seen in low-print-exposure participants (e.g., Chateau & Jared, 2000) may simply result from these participants being slower. In contrast, when overall response latency was matched across groups, print exposure did not moderate word-frequency effects (e.g., Lewellen et al., 1993).

One way to rule out processing speed as a confound is to standardize raw latencies using a *z*-score transformation[1] (see Faust et al., 1999). Consistent with this, Butler and Hains (1979), who did standardize their effects, were also unable to detect a relationship between vocabulary knowledge and word-frequency effects. It is interesting to note however, although vocabulary

[1] However, note that the *z* transformation has mathematical assumptions about the form of distributions and is therefore not theory independent. A reviewer also pointed out that *z* scoring and standardized regression coefficients fully adjust for processing speed only if all participants are equally reliable, an assumption that seems intuitively untenable. Future work could explore individual differences in effects when the influence of reliability is controlled for.

knowledge was unrelated to the size of word-frequency effects, high-vocabulary-knowledge readers still showed smaller effects of length. In summary, the mixed findings make it unclear if better readers are indeed less sensitive to lexical characteristics. Furthermore, it is unclear if reader proficiency interacts with certain variables (e.g., length) but not others (e.g., word-frequency), and if these interactions are task modulated (e.g., lexical decision vs. speeded pronunciation).

Closely related to the foregoing issues is the question of whether there are individual differences that reflect distinct types of proficient readers. One source of individual differences could arise from readers emphasizing different strategies or types of information during reading. If this was the case, one might expect to find trade-offs in a reader's sensitivity to different lexical characteristics. For example, are readers who are more sensitive to word frequency less sensitive to word length (and vice versa)? This trade-off could emerge if individuals are differentially emphasizing two distinct reading mechanisms, in which one mechanism is sensitive to frequency and the other is sensitive to length.

According to the dual-route perspective (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), word recognition is mediated by two pathways, a lexical pathway that directly maps words onto underlying lexical representations, and a nonlexical pathway that assembles the pronunciations of words via abstract spelling-to-sound rules. This dichotomy was in large part motivated by studies of acquired dyslexia that revealed a striking dissociation between the two pathways. Specifically, individuals with phonological dyslexia (Patterson, 1982) can pronounce real words but perform poorly for nonwords, whereas individuals with surface dyslexia (Shallice & Warrington, 1980) tend to regularize irregular words (e.g., pronounce PINT so that it rhymes with HINT). Because the two pathways make independent and unique contributions to reading, each pathway is sensitive to different properties of a word. The lexical pathway is sensitive to word frequency (but not to length), whereas the nonlexical pathway is sensitive to length (but not to word frequency).

Early work by Baron and Strawson (1976) suggested that there are different types of readers who selectively rely on these two pathways; *Phoenician* readers rely predominantly on nonlexical processing, while *Chinese* rely predominantly on lexical processing. If this is correct, one would expect a trade-off between word-frequency and length effects, which implies a negative correlation between sensitivity to these two variables. It is interesting to note that using Baron and Strawson's methodology, P. Brown et al. (1994) were unable to meaningfully separate their participants into Phoenicians and Chinese. There was no evidence for a trade-off between lexical and nonlexical processing (as suggested by Baron and Strawson's, 1976, work). Instead, participants who relied less on lexical processing also appeared to rely less on nonlexical processing. Although this pattern is difficult for an unembellished dual-route model to accommodate (see P. Brown et al., 1994, for more discussion), it is consistent with the simpler notion that faster, more skilled readers are simply less sensitive to a wide array of lexical variables. In summary, the second objective of this study is to more systematically explore the intriguing interplay between reader proficiency (as reflected by vocabulary knowledge), word recognition efficiency (as assessed by RT and error rate), and readers' sensitivity to different lexical characteristics.

RT Distributions, Word Recognition Performance, and Individual Differences

As pointed out in the Introduction, most word recognition studies are based on group-level data that aggregate across participants. In addition to averaging across participants, the typical word recognition experiment also collapses across trials in each condition to examine data at the level of mean RTs. There is increasing evidence that finer grained consideration of performance at the level of RT distributional characteristics (Balota & Yap, 2011; Hohle, 1965; Luce, 1986; Ratcliff, 1978, 1979) yields additional insights into important aspects of human cognition, including selective attention (Castel, Balota, Hutchison, Logan, & Yap, 2007; Heathcote, Popiel, & Mewhort, 1991; Spieler, Balota, & Faust, 2000; Tse, Balota, Yap, Duchek, & McCabe, 2010), visual search (Palmer, Horowitz, Torralba, & Wolfe, 2011), episodic memory (Hockley, 1984; Ratcliff, 1978; Rohrer & Wixted, 1994), priming (Balota, Yap, Cortese, & Watson, 2008; Kinoshita & Hunt, 2008; Lyons, Kellas, & Martin, 1995; Tse, Hutchison, & Li, 2011; Yap et al., 2009), eye fixation durations (Staub, White, Drieghe, Hollway, & Rayner, 2010), and isolated word recognition (Andrews & Heathcote, 2001; Balota & Spieler, 1999; Goh, Suárez, Yap, & Tan, 2009; Plourde & Besner, 1997; Ratcliff, Gomez, & McKoon, 2004; Yap & Balota, 2007; Yap, Balota, Cortese, & Watson, 2006; Yap et al., 2008).

Although there are many different ways to examine the characteristics of an RT distribution (see Luce, 1986; Van Zandt, 2000, for reviews), a relatively accessible method is to fit RT distributions to a theoretical distribution such as the ex-Gaussian (Ratcliff, 1979). The ex-Gaussian distribution is the convolution of a Gaussian (normal) and exponential distribution that approximates the positively skewed RT distribution often seen in empirical data. An ex-Gaussian distribution contains three parameters; μ and σ respectively reflect the mean and standard deviation of the Gaussian distribution, while τ reflects the mean and standard deviation of the exponential distribution. Using maximum likelihood procedures, the ex-Gaussian function can be fit to empirical data, and changes in μ are consistent with distributional shifting, whereas changes in τ reflect modulations in the tail of the distribution. A very useful aspect of ex-Gaussian analysis is that mean RTs are mathematically constrained to be the sum of μ and τ. Hence, effects in mean RTs can be partitioned such that one can evaluate the extent to which the effect is reflected in either distributional shifting or in the tail of the distribution. Also, it is possible for two conditions with identical mean RTs to be associated with different underlying RT distribution, due to trade-offs between μ and τ (see Balota et al., 2008; Heathcote et al., 1991, for more discussion).

Recently, researchers have used RT distributional analysis to explore questions in individual differences, particularly in the attention and working memory domain. For example, individuals with impaired attentional control systems are less able to maintain task goals and suppress irrelevant information, and may therefore experience lapses in control across time. More important, some researchers have suggested that these lapses are primarily reflected in the slowest RTs, that is, the tail of the distribution (see Coyle, 2003; Larson & Alderton, 1990, for more discussion). Tse et al.

(2010) compared young adults, healthy older adults, and individuals with very mild dementia of the Alzheimer's type (DAT) on three different tasks of selective attention (Stroop, Simon, and Task Switching). We find it interesting that age influenced all three parameters whereas slowing due to early stage DAT was primarily reflected in τ (the tail of the distribution). Structural equation modeling also revealed that changes in τ were most strongly related to a latent variable that reflected working memory. In a related study, Schmiedek, Oberauer, Wilhelm, Süβ, and Wittmann (2007) estimated ex-Gaussian parameters in eight choice RT tasks in young adults, and showed that τ (but not μ or σ) predicted performance on working memory, reasoning, and psychometric speed. Furthermore, the predictive power of τ appeared to generalize across different tasks.

At this juncture, it is worth clarifying that ex-Gaussian analysis primarily serves to provide an accurate and concise summary of empirical RT distributions (Schwarz, 2001). Ex-Gaussian parameters do not possess intrinsic theoretical interpretations (Matzke & Wagenmakers, 2009; Schmiedek et al., 2007), and it is dangerous to map them directly onto specific cognitive processes (cf. Hohle, 1965; McGill & Gibbon, 1965), unless one has a specific theoretical framework that makes predictions about the underlying RT distributions. To better understand the behavior of interest, researchers (e.g., Balota & Spieler, 1999; Schmiedek et al., 2007) have recommended linking static RT distributional characteristics to process-oriented theoretical models that explicitly specify the dynamics of information accumulation over time. The most well-known of these models is the diffusion model of binary decision (Ratcliff, 1978; Ratcliff et al., 2004), which is able to quantitatively fit data from binary decision tasks across various domains, including recognition memory, numerosity judgments, brightness and color discrimination, visual search, and lexical decision. More important, the diffusion model is able to fit RTs and error rates simultaneously, while accommodating the shapes of RT distributions for both correct and error responses.

The central premise of the diffusion model of lexical decision is that binary decision involves the accumulation of noisy information over time from a starting point (z) toward one of two decision boundaries, respectively word (a) or nonword (0). *Drift rate* (v) refers to the mean rate at which information is accumulated from the stimulus and is positively correlated with the quality of information yielded by processing the stimulus (Ratcliff et al., 2004). T_{er} refers to the nondecision component that collectively captures encoding and response execution. Other model parameters include s (variability in drift within each trial), η (variability in drift rate across trials), s_z (variability in starting point), and s_t (variability in the nondecision component). The inclusion of the variability parameters allows better fits of the model to lexical decision data. In their study, Ratcliff et al. (2004) applied the diffusion model to data from a number of lexical decision experiments and found that the effects of most variables (e.g., word frequency, nonword type, repetition) mainly influenced one parameter, drift rate, indicating that drift rate is the critical parameter for modeling lexical decision performance. This is perhaps unsurprising as drift rate and τ have been shown to be closely related. Effects of both parameters are most pronounced on the tail of the distribution, with steeper drift rates associated with smaller estimates of τ (see Spieler et al., 2000).

As in ex-Gaussian parameters, diffusion model parameters can also be estimated for individual participants. For example, Ratcliff, Thapar, and McKoon (2010) obtained diffusion model parameters for three groups of participants (college students, old, and very old) across three binary decision tasks (numerosity discrimination, item recognition, and lexical decision); participants also completed the Vocabulary and Matrix Reasoning subtests of the Wechsler Adult Intelligence Scale (3rd ed.; Wechsler, 1997). Ratcliff et al. (2010) observed a robust relationship between drift rate and IQ, with drift rates increasing as IQ increased; essentially, information accumulation processes were more efficient for the higher IQ participants. Boundary separation (a) and nondecision time (T_{er}) were relatively unaffected by IQ. Age had only intriguingly minimal effects on drift rates. Older adults were slower not because their drift rates were decreasing, but because they set more conservative response criteria and had longer nondecision times. This dissociation between age and IQ on different components of the diffusion model provide compelling evidence regarding the power of this approach.

Although we focus on how the diffusion model explains individual differences in lexical decision performance, it is important to note that there are important theoretical alternatives, such as the Bayesian reader model (Norris, 2006), which can also account for the effects of variables on RT distributions (see Norris, 2009). The Bayesian reader model unifies word recognition and decision-making processes within an integrated framework that assumes that readers behave like optimal Bayesian decision makers when processing words. That is, during word recognition, posterior probabilities for different words are computed (with their prior probabilities taken into account), and readers choose the word in the lexicon that best matches the input. In this light, people recognize high-frequency words faster because high-frequency words are associated with higher priors than low-frequency words. To carry out lexical decision, the model computes the probability that the presented letter string is a word versus a nonword, given the input. Although a detailed comparison of the diffusion and Bayesian approaches to lexical decision performance is beyond the scope of this paper (see Norris, 2009, for more discussion), future work could explore individual differences in word recognition performance by estimating Bayesian reader model parameters for each participant, and examining how variability in these parameters relate to other outcome measures.

To summarize, the extant literature points to stable, task-general individual differences in τ and diffusion model parameters that are systematically related to outcomes of interest. Hence, our third and final objective is to extend the individual difference work by Schmiedek et al. (2007), Tse et al. (2010), and Ratcliff et al. (2010) to the word recognition domain. As discussed earlier, readers become less reliant on controlled lexical processing mechanisms as they gain proficiency (LaBerge & Samuels, 1974; Stanovich, 1980). If controlled processing is marked by higher τ values and lower drift rates (Schmiedek et al., 2007; Tse et al., 2010), then one would expect fluent lexical processors, operationally defined by their vocabulary knowledge, to have lower values on τ but higher values on drift rate. In addition to testing this, we also will consider how τ and the central diffusion model parameters (i.e., drift rate, boundary separation, and nondecision time) relate to word recognition efficiency, vocabulary knowledge, and the extent to which readers are influenced by different lexical variables. The diffusion model parameters can only be estimated for the lexical decision data because the pronunciation task has not been modeled using this approach.

Method

Dataset

All analyses reported in this paper were based on archival trial-level data from the ELP. Because a full description of the methodological aspects of the ELP is available in Balota et al. (2007), we simply highlight some of the more salient aspects of the database. There were a total of 1,289 participants, with 470 providing data for the speeded pronunciation task and 819 providing data for the lexical-decision task. 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. Each participant took part in either the speeded pronunciation or lexical-decision task, and data were collected in two sessions on different days, separated by no more than 1 week. Across both sessions, each participant received approximately 2,530 speeded pronunciation trials or 3,374 lexical decision trials. Nonword trials on the lexical-decision task were legal nonwords that did not sound like real words. Evidence from previous work (e.g., Balota & Spieler, 1998) suggests that an individual participant can produce stable data for these numbers of stimuli. 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).

Results

We first excluded incorrect trials[2] 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 ease of exposition, we first describe the reliability analyses, before considering the relationships among participants' vocabulary knowledge, word recognition performance, and sensitivity to different lexical dimensions.

Analysis 1: Reliability Analyses

We first partitioned the data for each participant, so that trials were organized 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. Briefly, comparing S1 to S2 trials allows us to assess between-session reliability, while comparing odd- to even-numbered trials allows us to assess within-session reliability (as reflected by split-half correlations). For each participant, we then computed the mean and standard deviation of RTs, along with ex-Gaussian (μ, σ, τ) and diffusion model parameters[3] for S1 trials, S2 trials, odd-numbered trials, and even-numbered trials. Ex-Gaussian parameters were estimated for each participant using continuous maximum likelihood estimation in R[4] (R Development Core Team, 2004). Using Nelder and Mead's (1965) simplex algorithm, negative log-likelihood functions were minimized in the R statistics package (see 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 comprised 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 to minimize the value of chi-square (see Ratcliff & Tuerlinckx, 2002, for more information).

Table 1 presents the mean latency, its standard deviation, and ex-Gaussian parameters by task (speeded pronunciation and lexical decision) and trial type (overall, S1, S2, odd-numbered trials, even-numbered trials). Diffusion model parameters are also reported for the lexical-decision task. Participants' latencies tended to be faster and less variable on the speeded pronunciation task, compared to the lexical-decision task. Although speeded pronunciation performance was faster at the level of the mean, this mean difference was mediated fully by τ (change in the tail of the distribution) rather than by μ (shift in the modal portion). In fact, an individual's μ parameter for lexical decision was on average 39 ms faster than for pronunciation. These indicate that lexical decision RTs were associated with a faster leading edge that was offset by a heavier tail in the distribution. That is, for most words, participants could initiate a response more rapidly in lexical decision than in speeded pronunciation. However, the most difficult items in lexical decision produced more responses in the tail of the distribution, and this is consistent with the argument that the exaggerated tail in lexical decision reflects the postlexical decision processes engaged by the word–nonword discrimination demands of the task (Balota & Spieler, 1999; Ratcliff et al., 2004).

Turning to the diffusion model fits for the lexical decision data, the obtained parameters were in line with estimates reported elsewhere (e.g., Ratcliff et al., 2004). However, it is noteworthy that compared to Ratcliff et al.'s (2004) samples, the ELP participants were associated with larger boundary separation (a) and nondecision (T_{er}; i.e., encoding and response execution) components, suggesting that they were setting more conservative response criteria and taking longer to encode and respond to stimuli. More important, the mean word drift rate (v_{word}) was similar to Ratcliff et al.'s (2004) observed drift rates for low-frequency words (see their Table 6), consistent with the fact that the ELP participants were responding to a diverse set of items that included a relatively large proportion of difficult, lower frequency words.

Table 2 presents the Pearson correlations between each individual's response in S1 and S2 trials, and between odd- and even-numbered trials, for mean RT, standard deviation, ex-Gaussian parameters, and diffusion model parameters. For all correlations reported in this paper, bivariate outliers were detected by first computing Mahalanobis distance (D^2) for each pair of scores. This metric reflects the extent to which a particular participant is discrepant from the rest of the sample,

[2] Although incorrect trials were generally not analyzed, they were necessarily included in the diffusion model analyses.

[3] Satisfactory fits could be obtained for 780 of the 819 participants assigned to the lexical-decision task.

[4] QMPE 2.18 (http://www.newcl.org/software/qmpe.htm; S. Brown & Heathcote, 2003) is a free and user-friendly computer program that enables users to carry out distributional analyses on their RT data. In fact, QMPE can fit RT data with as few as 40 observations per condition.

and we excluded participants who produced D^2 with unusually low probability values (i.e., $p < .001$).

The very high correlations (all rs $\geq .92$) between odd- and even-numbered trials point to impressive within-session reliability for the mean, standard deviation[5], and ex-Gaussian parameters. Within-session reliability was also high for the majority of diffusion model parameters[6], particularly for the central ones, that is, boundary separation (a), nondecision component (T_{er}), and drift rate (v). Moreover, when between-session reliability was assessed, these correlations were also very high for means (rs $\geq .87$), and relatively high for ex-Gaussian (rs from .51 to .94) and diffusion model (rs from .39 to .74) parameters. These results provide evidence that readers carry with them a particular RT distributional signature that goes beyond simple mean performance. Indeed, this signature is maintained across days of testing on different sets of stimuli.

There are a couple of other noteworthy observations. First, the tail (τ) of the RT distribution seems to be considerably more stable than the modal portion (μ and σ) of the distribution. In fact, the between-session reliability of τ (.940 and .872 for pronunciation and lexical decision, respectively) was comparable to that for the mean (.929 and .871 for pronunciation and lexical decision, respectively). The between- and within-session reliability of word drift rate (.692 and .814, respectively) were also relatively high. The stability of τ and drift rate is consistent with the idea that these two parameters serve as markers of individual differences (Ratcliff et al., 2010; Schmiedek et al., 2007; Tse et al., 2010). Within the present context, 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. Second, the between-session reliabilities for the parameters were higher in speeded pronunciation than in lexical decision, suggesting that speeded pronunciation performance may be inherently more stable than lexical decision performance.

Table 2
Correlations Between Session 1 and Session 2 Parameters, and Odd- and Even-Numbered Trial Parameters

	Speeded pronunciation		Lexical decision	
	S1–S2	Odd–Even	S1–S2	Odd–Even
M RT	.929***	.998***	.871***	.997***
SD	.958***	.992***	.924***	.993***
μ	.865***	.993***	.717***	.983***
σ	.732***	.920***	.509***	.921***
τ	.940***	.987***	.872***	.988***
a			.708***	.906***
z			.720***	.910***
T_{er}			.736***	.930***
η			.403***	.649***
s_z			.388***	.812***
s_t			.497***	.647***
v_{word}			.692***	.814***
$v_{nonword}$			.645***	.827***

*** $p < .001$.

Having established the reliability of RT distributional characteristics, we next considered the reliability of individuals' sensitivity to different lexical characteristics. For example, if a participant produces large frequency effects on S1, will he or she also produce large frequency effects on S2? The most obvious way to approach this question is to conduct multiple-regression analyses at the level of individual participants, and to estimate, for each participant, regression coefficients for the various lexical variables of interest[7] (see 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 first partialed out the effects of word initial phoneme by coding dichotomously for the following 13 articulatory features: affricative, alveolar, bilabial, dental, fricative, glottal, labiodental, liquid, nasal, palatal, stop, velar, and voiced (see Balota,

Table 1
Means, Standard Deviations, Ex-Gaussian Parameters, and Diffusion Model Parameters as a Function of Task and Trial Type

	Overall	Session 1	Session 2	Odd	Even
			Speeded pronunciation[a]		
M	727	731	723	727	727
SD	180	176	180	180	180
μ	559	570	556	560	559
σ	63	64	62	64	63
τ	168	162	167	168	168
			Lexical decision[b]		
M	767	780	753	766	767
SD	243	240	235	243	242
μ	520	535	516	519	520
σ	67	68	65	66	66
τ	247	246	237	247	247
a	.169	.171	.165	.169	.170
z	.093	.095	.091	.093	.094
T_{er}	.495	.506	.49	.496	.497
η	.165	.171	.167	.167	.169
s_z	.118	.114	.114	.118	.119
s_t	.169	.175	.165	.169	.171
v_{word}	.223	.224	.23	.225	.229
$v_{nonword}$	−.255	−.256	−.261	−.256	−.257

[a] $n = 470$. [b] $n = 819$.

[5] It is noteworthy that the standard deviation is reliable, indicating that the ELP behavioral dataset has reliable estimates of the amount of variability in each individual's RTs. This is relevant when one uses tests of statistical significance that incorporate estimates of this variability.

[6] The within-session reliabilities of the η (drift rate variability) and s_t (nondecision component variability) parameters are lower because the variability parameters depend on error RT distributions, which have far fewer observations; these parameters are therefore always more poorly estimated (see Ratcliff & Tuerlinckx, 2002, for more discussion). The reliability of diffusion model, compared to ex-Gaussian, parameters could also be lower due to the diffusion model having to fit two additional dependent variables, accuracy and error RTs.

[7] To include semantic neighborhood density and number of senses as predictors, we restricted our analyses to the 28,803 words in the ELP which are represented on both norms.

Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Chateau & Jared, 2003; Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995). Doing this helps control for biases associated with using the voice key for measuring vocal responses (Kessler, Treiman, & Mullennix, 2002; Rastle, Croot, Harrington, & Coltheart, 2005). After controlling for the onset, we examined the effects of the following lexical and semantic variables: log-transformed hyperspace analog to language (HAL; Lund & Burgess, 1996) frequency (henceforth word frequency); number of morphemes; number of syllables; number of letters; orthographic neighborhood size; phonological neighborhood size; orthographic Levenshtein distance 20[8] (OLD20, a measure of orthographic distinctiveness; Yarkoni, Balota, & Yap, 2008); phonological Levenshtein distance 20 (PLD20, a measure of phonological distinctivness; Yap & Balota, 2009); log-transformed number of senses (Steyvers & Tenenbaum, 2005), which reflects the number of meanings a word has in the WordNet database (Miller, 1990); and log-transformed semantic neighborhood density (Shaoul & Westbury, 2010), estimated using co-occurrence information from a billion-word Wikipedia corpus (Shaoul & Westbury, 2009). Semantic neighborhood density was estimated by the average radius of co-occurrence (ARC), which is based on the average semantic distance between a target word and its closest neighbors in high-dimensional semantic space (see Shaoul & Westbury, 2006). In other words, words with lower ARC values are associated with denser semantic neighborhoods.

Figures 1 and 2 present the distributions of standardized regression coefficients across participants as a function of lexical variable, for speeded pronunciation (see Figure 1) and lexical decision (see Figure 2), respectively. First, note the substantial variability in the magnitude of effects produced by participants.[9] For example, although virtually all participants produced negative regression coefficients for the word-frequency effect (see Figures 1 and 2), indicating faster latencies for higher frequency words, the coefficients were normally distributed. Second, the direction and relative magnitudes of participant-level effects in speeded pronunciation and lexical decision were generally consistent with item-level effects reported in the literature. That is, word frequency was the best predictor in both tasks, followed by number of letters, number of syllables, and number of morphemes. Generally, higher frequency words and words with more morphemes were recognized faster, whereas words with more syllables and letters were recognized slower.

One caveat of multiple regression analyses with many lexical predictors is that these predictors are often correlated (Baayen, Feldman, & Schreuder, 2006; Cutler, 1981). For example, longer words tend to be lower in frequency and have fewer neighbors; this collinearity in the predictor matrix (see Table 3) may cause regression coefficients to fluctuate in magnitude and direction, leading to estimates of individual regression coefficients that are unreliable due to large standard errors. To alleviate this problem, principal components analysis (Baayen et al., 2006) was used to reduce the full set of 10 lexical variables to a smaller set of orthogonal principal components. Principal components analysis was carried out using varimax rotation with Kaiser normalization. Three principal components, accounting for 85% of the variance, were extracted.[10] The rotated component matrix indicated that number of letters, number of syllables, OLD20, PLD20, and number of morphemes loaded on the first component (PC1);[11] orthographic and phonological neighborhood size loaded on the second component (PC2); and word frequency, number of senses, and ARC loaded on the final component (PC3); factor loadings are presented on Table 4. Thus, PC1 appears to capture the structural properties of words, PC2 neighborhood size, and PC3 word frequency/semantics. From the dual-route perspective, one could also say that PC1 seems to reflect the sublexical properties of words, whereas PC3 taps whole-word properties.

We next obtained principal component regression coefficients for each participant across all trials, S1 trials, S2 trials, odd-numbered trials, and even-numbered trials (see Table 5 and Figure 3).[12] We find it interesting that although the between-task differ-

[8] The Levenshtein-based measures are relatively new measures of orthographic and phonological distinctiveness that are optimized for longer words. They are based on Levenshtein distance (LD), a computer science metric reflecting the minimum number of substitution, insertion, or deletion operations required to convert one string of elements (either letters of phonemes) into another. For example, the LD from *kitten* to *sitting* is 3, reflecting two substitutions ($k \rightarrow s$, $e \rightarrow i$) and one insertion (insert *g* at the end). Yarkoni et al. (2008) computed the LD from each word in the ELP to every other word, and this was used to generate LD20 values for each word, defined as the mean LD between a word and its 20 closest neighbors. OLD20 is based on distances between word spellings, whereas PLD20 is based on distances between word pronunciations. Words with higher LD20 values are further from their closest neighbors, implying they are more orthographically or phonologically distinct. Traditional neighborhood size measures have limited utility for long words, which have few or no single-letter substitution neighbors. Measures based on LD circumvent this limitation by providing estimates of distinctiveness for even very long words. We find it interesting that Yarkoni et al. (2008) reported that LD-based measures of orthographic distinctiveness provide a significant advantage over traditional density-based measures in predicting performance on English word recognition tasks, particularly for longer words.

[9] According to a reviewer, the distribution of effects reflects both true and residual variance. In consequence then, the individual differences between participants may be partly mediated by their variability in residual variance. The extent to which individual differences reflect true versus residual variance is an interesting question that merits future investigation.

[10] It is worth noting that the original set of 10 predictors respectively explained 55.4% and 62.5% of the variance in speeded pronunciation and lexical decision RTs. The principal components respectively accounted for 48.2% and 56.4% of the RT variance, indicating that the dimensional reduction did not substantially compromise the predictive power of the lexical variables.

[11] Given that the Levenshtein measures (Yarkoni et al., 2008) have been described as metrics for capturing neighborhood size, one might find it surprising that OLD20 and PLD20 loaded on a component reflecting structural properties (PC1) instead of neighborhood size (PC2). However, Yarkoni et al. showed that although LD20 and neighborhood size are highly correlated for shorter, monosyllabic words, this relationship is considerably weaker for longer words with more syllables, due to range restriction for neighborhood size (long words have few or no substitution neighbors). Consistent with this, Table 3 shows that for the full set of words in our analyses, the LD20 measures are more highly correlated with length than with neighborhood size.

[12] For ease of exposition, our analyses focus on the principal component effects. However, supplementary tables reporting the descriptive statistics (Table A1) and reliabilities (Table A2) of effects of individual variables are presented in the Appendix. It is very clear that reliability estimates are substantially higher for principal component effects than for effects of individual variables (compare Tables A2 and 6). This attests to the utility of the PCA approach, which provides far more reliable estimates of how stimuli characteristics affect behavioral data.

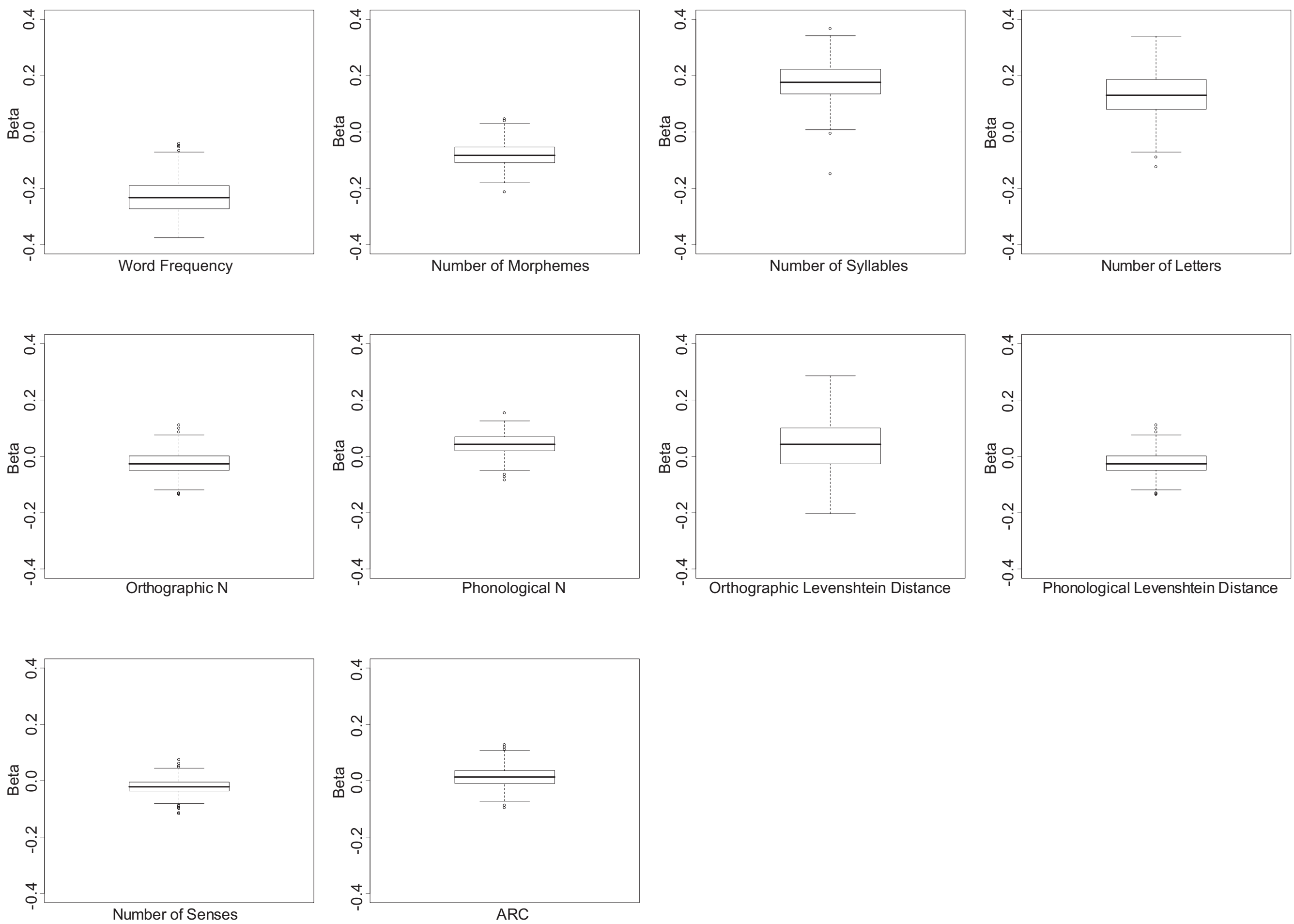


Figure 1. Distributions of standardized regression coefficients across participants as a function of lexical variable for speeded pronunciation. ARC = average radius of co-occurrence.

ences were in the correct direction (e.g., effects of word frequency/semantics were stronger in lexical decision, whereas effects of structural properties were stronger in speeded pronunciation), they were relatively subtle compared to those suggested by item-level regression analyses (e.g., Balota et al., 2004; Yap & Balota, 2009). Of course, those analyses were based on much smaller sets of words (2,428 monosyllabic words for Balota et al., 2004, and 9,639 monomorphemic multisyllabic words for Yap & Balota, 2009). For a more appropriate comparison, we also conducted parallel item-level regression analyses on the full set of items from the ELP; these are presented on an additional item column on Table 5. As can be seen, when one considers all the words in the ELP, the relative magnitudes of the participant- and item-level effects across tasks showed similar trends, reassuring us that the modest between-task differences are not simply artifacts of participant-level analyses. It is also worth noting that item-level effects were consistently larger than participant-level effects. Because item means average across participants, this might help to decrease measurement error and increase the size of the effect.

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 three principal component regression coefficients. Within- and between-session measures of reliability were quite high in both tasks,[13] with higher estimates observed in speeded pronunciation than in lexical decision. In addition, effects of PC1 (structural properties) and PC3 (word frequency/semantics) were more reliable than effects of PC2 (neighborhood size). To summarize, within- and between-session reliabilities are reassuringly high in word recognition performance, as reflected in RT distributional characteristics, diffusion model parameters, and in effects of principal components. Reliability was also higher for speeded pronunciation performance, and this might be related to the additional task-specific processes entailed by lexical decision (Balota & Chumbley, 1984).

[13] Within-session reliability was consistently higher than between-session reliability, consistent with other studies (e.g., Kimberley, Khandekar, & Borich, 2008). As the time-gap between two measurements increase, the less similar are the factors that contribute to error, attenuating correlations (Trochim & Donnelly, 2006). We tested this by examining if the number of days between sessions moderated the extent to which Session 1 performance predicted Session 2 performance. In both tasks, we found that reliability was generally significantly lower as the number of intervening days increased, although this trend was seen only in mean RTs and RT distributional characteristics, but not in effects of principal components.

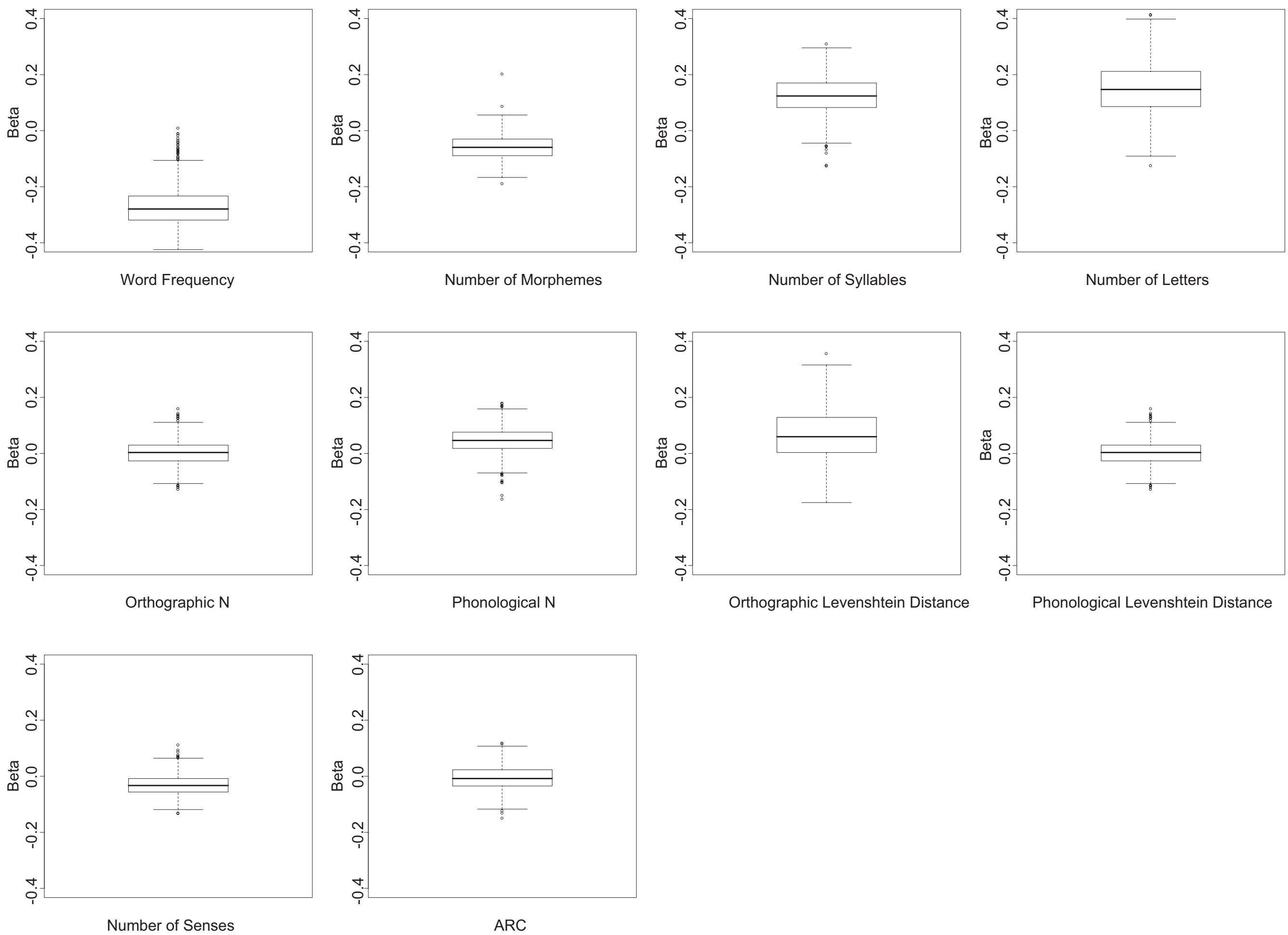


Figure 2. Distributions of standardized regression coefficients across participants as a function of lexical variable for lexical decision. ARC = average radius of co-occurrence.

Analysis 2: Vocabulary Knowledge and Word Recognition Performance

We now turn to the relationship between vocabulary knowledge and word recognition performance. As discussed, 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 (Yap et al., 2009). Higher vocabulary-knowledge participants should be associated with faster and more accurate word recognition performance. Specifically, vocabulary knowledge should be negatively correlated with RTs and positively correlated with accuracy. Figures 4 (RT) and 5 (accuracy) present the scatterplots between vocabulary knowledge (as assessed by the number of correct responses on the Shipley, 1940, vocabulary subscale) and word recognition performance, after excluding participants who were more than 1.5 interquartile ranges below the lower quartile on a boxplot (3.0% in speeded pronunciation and 3.3% in lexical decision). Vocabulary knowledge was negatively correlated with mean speeded pronunciation RTs ($r = -.402$) and lexical decision RTs ($r = -.323$), and positively correlated with speeded pronunciation accuracy ($r = .456$) and lexical decision accuracy ($r = .622$).

Schmiedek et al. (2007) and Tse et al. (2010) showed that the tail of the RT distribution is a strong predictor of controlled processing, suggesting that this component should be especially sensitive to variability in word recognition efficiency. If this is correct, then the correlation between vocabulary knowledge and RTs should be primarily mediated by τ. This was precisely the pattern observed (see Table 7). Although vocabulary knowledge was negatively correlated with both μ (leading edge of the modal distribution; *r*s were $-.218$ and $-.159$ for pronunciation and lexical decision, respectively) and τ (tail of the distribution; *r*s were $-.448$ and $-.335$ for pronunciation and lexical decision, respectively), the correlation was reliably stronger in τ than in μ in both speeded pronunciation ($p < .001$) and lexical decision ($p < .001$). This indicates that variation in recognition performance for difficult words (e.g., *antiestablishment*), relative to easy words (e.g., *chair*), is more systematically related to individual differences in skill. The predictive power of τ, which has so far been demonstrated using general binary decision tasks (Schmiedek et

Table 3
Correlation Matrix for the Ten Lexical and Semantic Predictors

Predictors	1	2	3	4	5	6	7	8	9	10
1. Word frequency (HAL)	—	−.276***	−.277***	−.360***	.297***	.305***	−.402***	−.383***	.440***	.681***
2. Number of morphemes		—	.603***	.699***	−.331***	−.362***	.528***	.551***	−.254***	−.146***
3. Number of syllables			—	.830***	−.479***	−.512***	.749***	.800***	−.257***	−.103***
4. Number of letters				—	−.541***	−.553***	.870***	.840***	−.278***	−.173***
5. Orthographic neighborhood size					—	.801***	−.567***	−.513***	.296***	.133***
6. Phonological neighborhood size						—	−.545***	−.561***	.292***	.138***
7. Orthographic Levenshtein distance							—	.913***	−.332***	−.210***
8. Phonological Levenshtein distance								—	−.312***	−.196***
9. Number of senses									—	.263***
10. Semantic neighborhood density										—

Note. HAL = hyperspace analog to language.
*** $p < .001$.

al., 2007) and tasks tapping selective attention (Tse et al., 2010), appears to generalize to lexical processing tasks as well.

We also examined the correlations between the diffusion model parameters and vocabulary knowledge, lexical decision RTs, and lexical decision accuracy (see Table 8). Although the full set of model parameters are presented, we are mainly interested in the correlations for boundary separation (a), nondecision time (T_{er}), and drift rate (v). Participants who were faster on the lexical-decision task produced smaller values for boundary separation and nondecision time, and larger values for drift rates. In other words, they were setting more liberal decision criteria, had a faster nondecision component, and could accumulate information more rapidly. Notably, individual differences in vocabulary knowledge were also systematically related to these three parameters. Specifically, higher vocabulary-knowledge participants were associated with lower values for boundary separation and nondecision time, and higher drift rates. At first blush, this might seem inconsistent with Ratcliff et al. (2010), who found that IQ (which taps vocabulary knowledge) was correlated with drift rate, but not with boundary separation or nondecision time. However, it should be noted that the correlations for boundary separation ($r = -.085$) and nondecision time ($r = -.218$) were significantly smaller (all ps $<$ .001) than for drift rate (.536 and −.448 for words and nonwords, respectively). As the present sample is more than four times larger than Ratcliff et al.'s (2010) sample, the present analyses have more power to detect the more subtle effects for boundary separation and nondecision time.

Next, we considered the correlations between diffusion model parameters and ex-Gaussian parameters (see Table 8). Consistent with Schmiedek et al. (2007), drift rates were more highly correlated with τ than with μ or σ, ps $< .001$. However, one should note that τ also was strongly and positively correlated with boundary separation, indicating that participants producing a large proportion of slow responses were setting more conservative response criteria. This indicates that there is only limited mapping between ex-Gaussian and diffusion model parameters, in line with Matzke and Wagenmakers (2009). Having said this, the relationship between μ and nondecision time was very strong, consistent with the idea that the leading edge of a participant's RT distribution primarily reflects encoding time and response execution, and is relatively independent of efficiency in accumulating information.

Analysis 3: Vocabulary Knowledge, Diffusion Model Parameters, and Effects of Lexical Variables

It has been generally assumed that reading proficiency (as reflected by print exposure or vocabulary knowledge) is negatively correlated with the effect of variables like frequency (Chateau & Jared, 2000) or length (Butler & Hains, 1979), such that better readers are less influenced by different variables. To explore this, Table 9 (see also Figure 6) presents the correlations among participant-level standardized principal component regression coefficients, word recognition RTs, and vocabulary knowledge (see Table A3 in the Appendix for correlations based on individual predictors).

In speeded pronunciation, higher-vocabulary-participants were less sensitive[14] to a word's structural properties (PC1), neighborhood size (PC2), and word frequency/semantics (PC3). Similarly, in the lexical-decision task, higher vocabulary-

Table 4
Rotated Component Matrix for the Three Principal Components Extracted

Predictors	Component 1	2	3
1. Number of letters	**.901**	−.284	−.146
2. Number of syllables	**.870**	−.253	−.071
3. Phonological Levenshtein distance	**.845**	−.314	−.195
4. Orthographic Levenshtein distance	**.819**	−.348	−.221
5. Number of morphemes	**.796**	−.050	−.130
6. Orthographic neighborhood size	−.293	**.886**	.134
7. Phonological neighborhood size	−.329	**.865**	.133
8. Semantic neighborhood density	−.065	−.042	**.883**
9. Word frequency	−.224	.138	**.873**
10. Number of senses	−.143	.308	**.551**

Note. Boldface numbers reflect principal component loadings.

[14] Note that a negative correlation between the structural component and vocabulary knowledge indicates that length effects decrease toward zero as vocabulary increases; length and RTs are positively related. In contrast, a positive correlation between the word-frequency/semantics dimension and vocabulary knowledge indicates that frequency effects increase toward zero as vocabulary increases; word-frequency/semantics and RTs are negatively related.

Table 5
Descriptive Statistics for Participant-Level Principal Component (PC) Effects as a Function of Task and Trial Type

Principal component effects	Overall	S1	S2	Odd	Even	Item
Speeded pronunciation[a]						
PC1 (length/OLD/PLD)	.262	.261	.261	.256	.260	.509
PC2 (ON/PN)	−.099	−.097	−.100	−.097	−.097	−.196
PC3 (frequency/semantics)	−.224	−.228	−.218	−.221	−.221	−.401
Lexical decision						
PC1 (length/OLD/PLD)	.251	.256	.252	.246	.251	.500
PC2 (ON/PN)	−.070	−.074	−.065	−.069	−.069	−.153
PC3 (frequency/semantics)	−.264	−.269	−.263	−.259	−.263	−.533

Note. S1 = Session 1; S2 = Session 2; OLD = orthographic Levenshtein distance; PLD = phonological Levenshtein distance; ON = orthographic neighborhood size; PN = phonological neighborhood size.
[a] $n = 470$. [b] $n = 819$.

participants produced smaller neighborhood effects. However, an individual's sensitivities to the structural and word-frequency/semantics stimulus dimensions were only marginally related to vocabulary knowledge. Moreover, the weak correlation between word-frequency/semantics effects and vocabulary knowledge indicate that effects were larger for participants with more vocabulary knowledge. These results challenge the assertion that word-frequency effects in lexical decision performance are negatively related to an individual's vocabulary knowledge when processing speed is controlled. The present results are compatible with earlier work by Butler and Hains (1979), Lewellen et al. (1993), and Sears et al. (2008). It remains unclear why this pattern is seen in lexical decision but not in speeded pronunciation, in which effects of all three

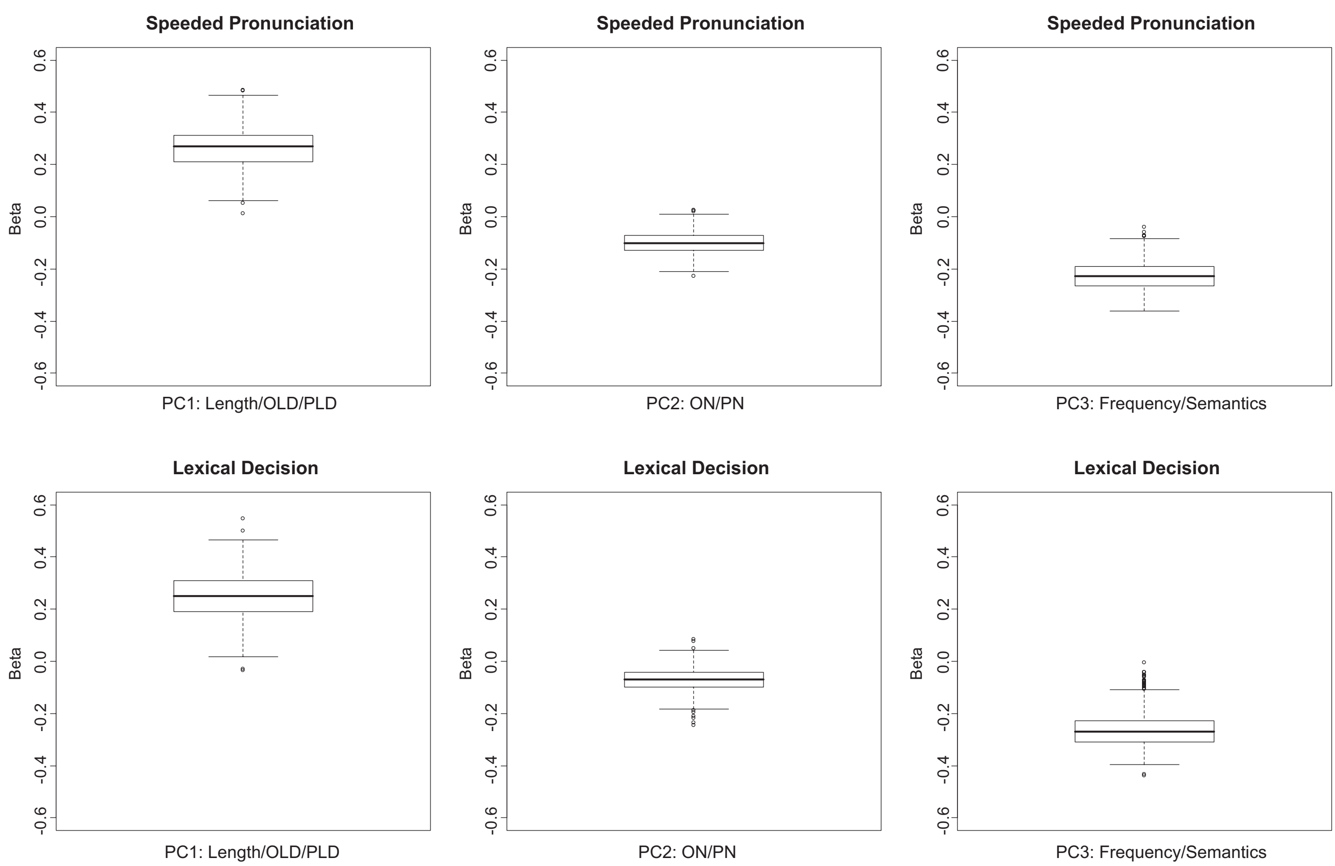


Figure 3. Distributions of standardized regression coefficients across participants as a function of principal component (PC) for speeded pronunciation (top) and lexical decision (bottom). OLD = orthographic Levenshtein distance; PLD = phonological Levenshtein distance; ON = orthographic neighborhood size; PN = phonological neighborhood size.

Table 6
Correlations Between Session 1 and Session 2 Participant-Level Principal Component Effects and Odd- and Even-Numbered Trial Participant-Level Principal Component Effects

Principal component effects	Speeded pronunciation		Lexical decision	
	S1–S2	Odd–Even	S1–S2	Odd–Even
PC1 (length/OLD/PLD)	.706***	.799***	.647***	.753***
PC2 (ON/PN)	.534***	.623***	.377***	.429***
PC3 (frequency/semantics)	.625***	.677***	.515***	.638***

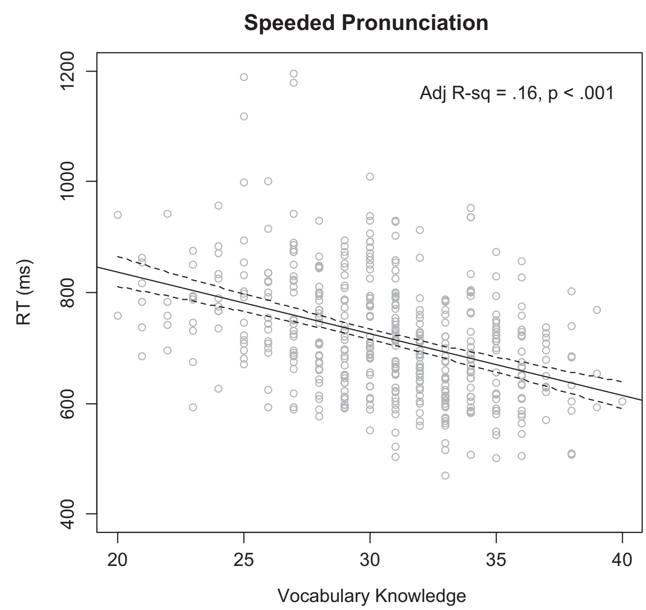
Note. S1 = Session 1; S2 = Session 2; OLD = orthographic Levenshtein distance; PLD = phonological Levenshtein distance; ON = orthographic neighborhood size; PN = phonological neighborhood size.
*** $p < .001$.

principal components were attenuated as vocabulary knowledge increased. Further discussion of this intriguing pattern will be postponed to the General Discussion section.

We now turn to the correlations between diffusion model parameters and principal component effects (see Table 10 and Figure 7; see Table A4 in the Appendix for correlations based on individual predictors) in lexical decision performance (because the diffusion model only applies to binary decision tasks). As before, our discussion will focus on the three parameters of greatest theoretical interest, that is, boundary separation (a), nondecision time (T_{er}), and drift rate (v). The results for the structural properties principal component (PC1) are relatively straightforward. Specifically, structural properties (PC1) effects were positively correlated with boundary separation and nondecision time, and negatively correlated with drift rate. In other words, individuals with a greater sensitivity to stimulus length show more conservative response criteria, longer nondecision times, and less efficient accumulation of evidence. Results for the other two principal component effects are less clear. Although boundary separation was reliably related to neighborhood (PC2) and word-frequency/semantics (PC3) effects, such that individuals with larger PC2 and PC3 effects produced larger values for boundary separation, the correlations were very low and should be interpreted with caution. Turning to nondecision time, PC2 and PC3 were negatively correlated, that is, participants with larger effects on these principal components yielded longer nondecision times. Finally, greater sensitivity to neighborhood size was associated with lower drift rates, but greater influence of word-frequency/semantics produced higher drift rates. Overall, these observations are compatible with the idea that larger effects of lexical variables are accompanied by lower drift rates and longer nondecision times. The major anomaly was the positive relationship between word-frequency/semantics effects and drift rate, wherein larger word-frequency/semantics effects were associated with larger drift rates. This finding will be explored in greater depth in the General Discussion.

Analysis 4: Relationships Among the Different Principal Component Effects

We now turn to the correlations between individual's sensitivity to different lexical dimensions, as reflected by their effects for the three principal components (see Table 11 and Figure 8). Contrary to Baron and Strawson's (1976) proposal, there was no evidence for a trade-off between nonlexical and lexical processes, which would be evident in sensitivity to length (PC1) and word-frequency/semantics (PC3) respectively. Individual differences in word recognition performance do not appear to reflect a trade-off between Phoenician and Chinese reading strategies, as Baron and Strawson suggested. Instead, all effects showed strong positive correlations, that is, being more sensitive to one lexical dimension is associated with a higher sensitivity to other dimensions. More important, it is worth pointing out that these correlations are unlikely to be entirely due to scaling or to general slowing, because



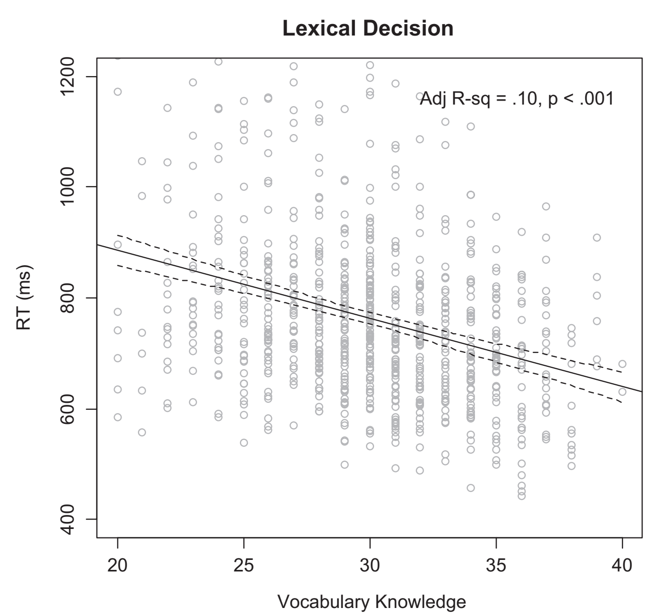


Figure 4. Scatterplots (with 95% confidence intervals) between vocabulary knowledge and speeded pronunciation (top) and lexical decision (bottom) response times (RT).

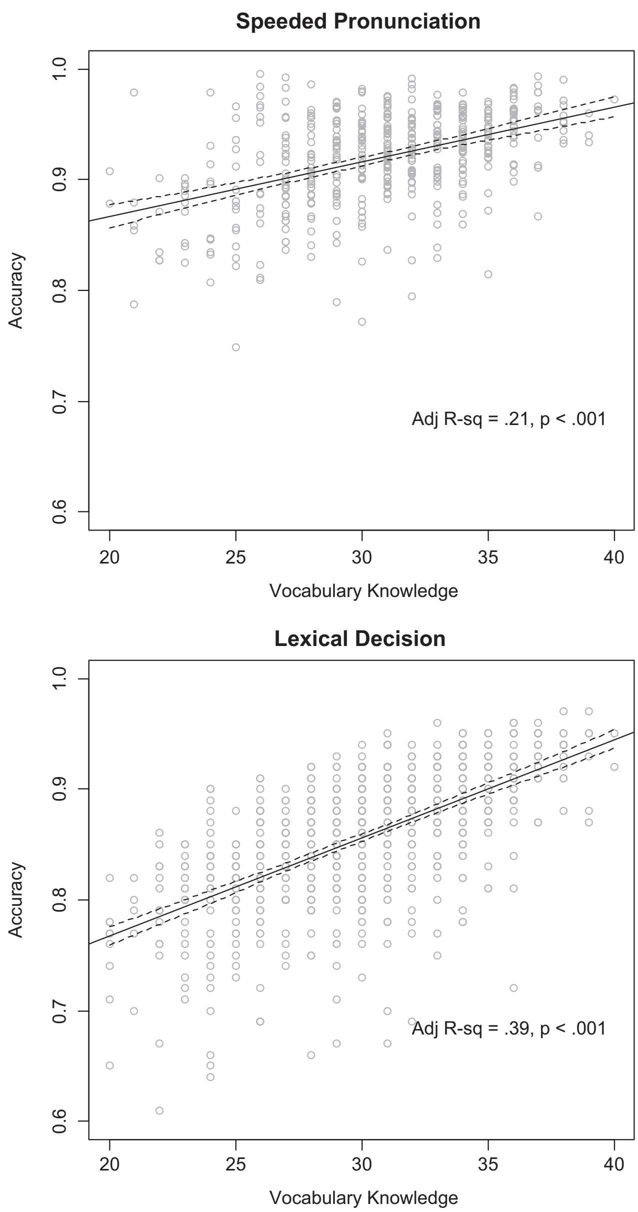


Figure 5. Scatterplots (with 95% confidence intervals) between vocabulary knowledge and speeded pronunciation (top) and lexical decision (bottom) accuracy.

they are based on standardized regression coefficients (but see Footnote 1 for caveats). In general, we found no evidence of distinct subtypes of readers or trade-offs that might reflect qualitatively distinct lexical processing strategies. Also, it is intriguing that the between-component correlations were considerably higher in speeded pronunciation than in lexical decision, which is compatible with earlier analyses suggesting that speeded pronunciation performance is more reliable.

General Discussion

The present study is the first large scale investigation of individual differences using trial-level data from the ELP. There were a number of noteworthy findings. First, across different sets of stimuli, between- and within-session reliability was relatively high for an individual's mean RT, RT distributional characteristics, diffusion model parameters, and sensitivity to underlying lexical dimensions. Second, higher vocabulary knowledge was associated with faster and more accurate word recognition performance, and generally attenuated sensitivity to underlying lexical characteristics, along with steeper drift rates, and shorter nondecision times. Third, there was no evidence for a trade-off between sensitivity to different types of lexical information. Instead, participants who showed more influence of one variable also showed more influence of other variables.

Variability and Reliability of Word Recognition Performance

Lexical decision and speeded pronunciation latencies are the most popular dependent variables in word recognition research. Our analyses established clear and substantial between-participants variability in word recognition performance, across virtually all the parameters (e.g., mean RTs, RT distributional characteristics) examined. Characterizing the scale of these individual differences within a large sample of participants across different universities should help constrain future work on individual differences, and also inform studies that attempt to study a "prototypical" reader using group-level analyses. For instance, the large individual differences among participants may also be responsible for some of the empirical inconsistencies in the literature (see, e.g., Yap et al., 2009).

We began by exploring the reliability of different measures of individual differences in word recognition performance. Understanding the reliability of the measures is a critical first step in the study of individual differences. Without first establishing reliability, it is unclear whether variability among readers reflects meaningful individual differences or measurement noise. Our analyses identified several measures of word recognition performance that are stable within an individual. Specifically, across distinct sets of words, within-session and between-session reliabilities were generally quite high with respect to mean RTs, standard deviations, distributional characteristics, diffusion model parameters, and sensitivity to underlying dimensions (as reflected by both individual predictors and principal components). Remarkably, cross-session correlations for the τ parameter estimates from the ex-Gaussian analyses (.940 and .872 for pronunciation and lexical decision, respectively) were as stable as those for the mean (.929 and .871, respectively), further underscoring the utility of RT distributional analyses.

Table 7
Vocabulary Knowledge as a Predictor of Mean RT and Ex-Gaussian Parameters in Speeded Pronunciation and Lexical Decision

Task	M	μ	σ	τ
Speeded pronunciation	$-.402^{***}$	$-.218^{***}$	$-.335^{***}$	$-.448^{***}$
Lexical decision	$-.323^{***}$	$-.159^{***}$	$-.241^{***}$	$-.335^{***}$

*** $p < .001$.

Table 8
Relationships Between Diffusion Model Parameters and Mean RT, Mean Accuracy, and Vocabulary Knowledge

Diffusion model parameters	*M* RT	*M* accuracy	Vocabulary knowledge	μ	Σ	τ
a	.744***	.127***	−.085*	.449***	.237***	.776***
z	.682***	.191***	−.083*	.388***	.170***	.731***
T_{er}	.636***	−.097**	−.218***	.891***	.459***	.350***
η	−.311***	.084*	.082*	−.106**	−.168***	−.338***
s_z	.359***	.018	−.053	.120***	.121***	.472***
s_t	.425***	−.334***	−.250***	.475***	.653***	.220***
v_{word}	−.476***	.580***	.536***	−.239***	−.343***	−.509***
$v_{nonword}$	.526***	−.568***	−.448***	.274***	.407***	.569***

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

More generally, these results suggest that participants carry with them a relatively stable distributional and processing profile that goes beyond one's average processing speed. The reliability of τ and drift rate (see Table 2) nicely inform claims (e.g., Schmiedek et al., 2007; Tse et al., 2010) that stable individual differences in these parameters are systematically related to controlled processing. Of course, τ is a descriptive parameter, whereas drift rate has the advantage of being an explanatory parameter that can be directly mapped onto an underlying process (Matzke & Wagenmakers, 2009).

Another striking finding is that reliability seems to be consistently higher in speeded pronunciation than in lexical decision, in terms of distributional characteristics and sensitivity to underlying lexical dimensions (see Tables 2 and 6). The reliability of speeded pronunciation performance is intriguing, given the possible extra sources of measurement noise in speeded pronunciation (e.g., variations in sensitivity of the voice-key to different speaker characteristics). However, in contrast to the pronunciation task, lexical decision may be driven by familiarity-based information and hence taps both lexical mechanisms as well as postlexical decision-based mechanisms (Balota & Chumbley, 1984; Ratcliff et al., 2004). In other words, lexical decision performance may be inherently less reliable because it is jointly influenced by variability in both lexical and postlexical processing. We comment further on this intriguing pattern in the next section.

The Interplay Between Vocabulary Knowledge, Word Recognition Performance, and Diffusion Model Parameters

A second issue addressed in this paper concerns the relationship between vocabulary knowledge and word recognition performance. Vocabulary knowledge was systematically related to many other measures of word recognition performance. For example, participants with higher vocabulary knowledge were associated with faster and more accurate speeded pronunciation and lexical decision performance. The ex-Gaussian analyses further revealed that the relationship between vocabulary and speed was predominantly mediated by τ (see Table 7), in both speeded pronunciation and lexical decision. In other words, vocabulary knowledge was more strongly correlated with slower, compared to faster, RTs. This pattern is consistent with the worst performance rule (Coyle, 2003), wherein IQ is most strongly related to the slowest RTs. Typically, this finding has been observed more strongly in measures of fluid intelligence.

Vocabulary knowledge was also related to the diffusion model parameters in interesting ways. In lexical decision, more vocabulary knowledge was associated with larger drift rates ($r = .536$), suggesting that as vocabulary increases, individuals process stimuli more efficiently (see Table 8). These results converge with Ratcliff et al.'s (2010) observation that IQ was strongly and positively related with drift rate, but not with other diffusion model

Table 9
Correlations Between Participant-Level Standardized Regression Coefficients, Vocabulary Knowledge, and Mean RT

Principal component effects	Speeded pronunciation[a]		Lexical decision[b]	
	Vocabulary knowledge	*M* RT	Vocabulary knowledge	*M* RT
PC1 (length/OLD/PLD)	−.200***	.033	−.060†	.205***
PC2 (ON/PN)	.323***	−.189***	.201***	−.063†
PC3 (frequency/semantics)	.280***	−.394***	−.083*	.125***

Note. RT = response time; PC = principal component; OLD = orthographic Levenshtein distance; PLD = phonological Levenshtein distance; ON = orthographic neighborhood size; PN = phonological neighborhood size.
[a] $n = 456$. [b] $n = 792$.
† $p < .10$. * $p < .05$. *** $p < .001$.

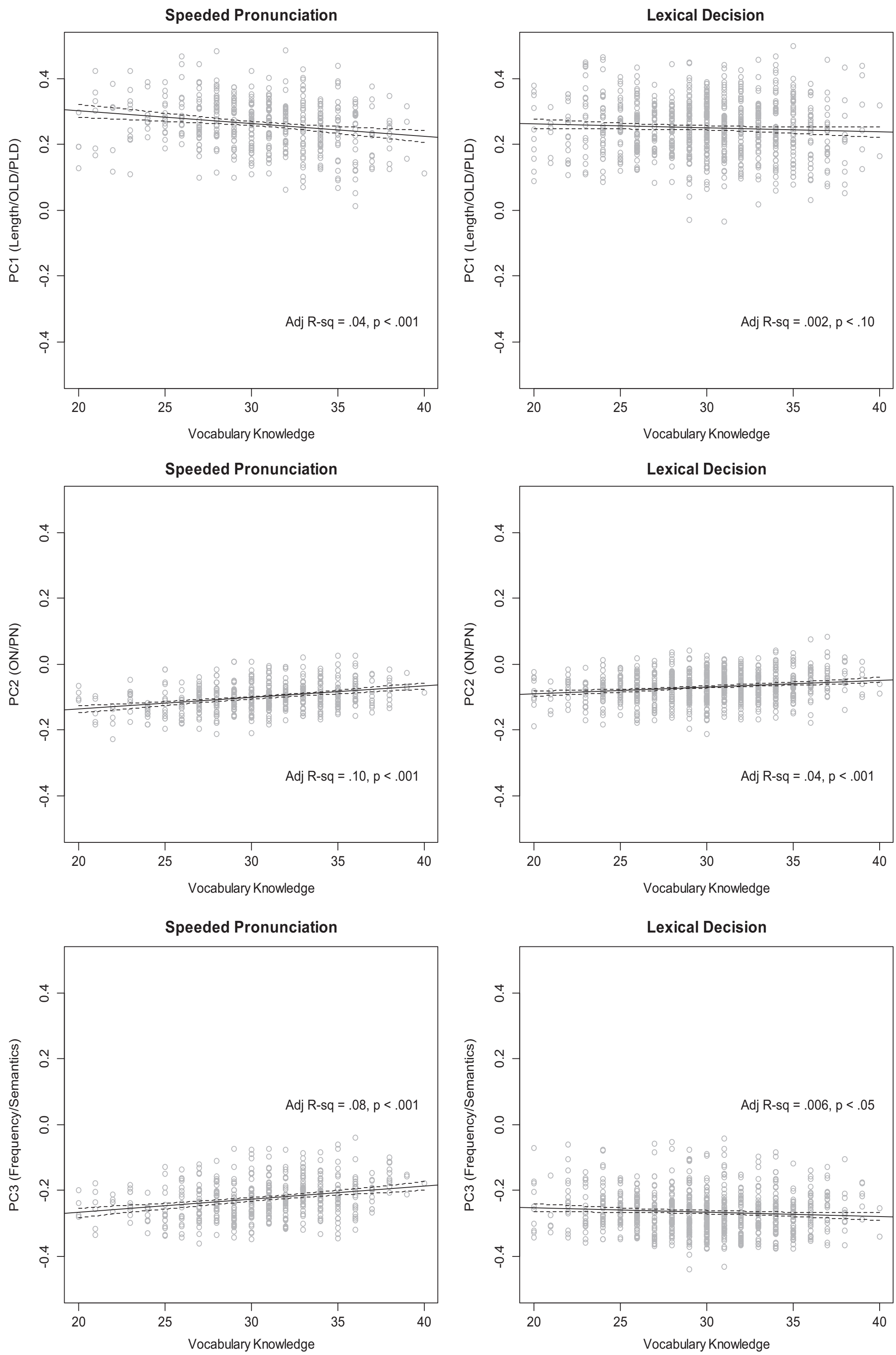


Figure 6. Scatterplots (with 95% confidence intervals) between vocabulary knowledge and principal component (PC) effects in speeded pronunciation (left) and lexical decision (right). OLD = orthographic Levenshtein distance; PLD = phonological Levenshtein distance; ON = orthographic neighborhood size; PN = phonological neighborhood size.

Table 10
Correlations Between Participant-Level Standardized Regression Coefficients and Diffusion Model Parameters

Principal component effects	a	z	T_{er}	η	s_z	s_t	v_{word}	$v_{nonword}$
PC1 (length/OLD/PLD)	.187***	.209***	.231***	−.044	.078*	.073*	−.103***	.066†
PC2 (ON/PN)	.074*	.072*	−.209***	−.032	.017	−.231***	.204***	−.144***
PC3 (frequency/semantics)	.111**	.032	−.122***	−.205***	.092**	.086*	−.145***	.284***

Note. PC = principal component; OLD = orthographic Levenshtein distance; PLD = phonological Levenshtein distance; ON = orthographic neighborhood size; PN = phonological neighborhood size.
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

parameters. However, vocabulary knowledge was also negatively related to boundary separation ($r = -.085$) and nondecision time ($r = -.218$), consistent with more liberal response criteria and shorter nondecision times. Obviously, these correlations were not as strong as the correlation for drift rate, but they suggest that vocabulary size can have additional effects on the parameters from the diffusion model. We find it interesting that although IQ did not reliably predict boundary separation and the nondecision component in Ratcliff et al.'s (2010) sample, their correlations (see their Table 5) were in the same direction as the corresponding correlations in the present study. Of course, the very large ELP sample affords more statistical power to detect smaller effects.

We were also interested in the relationship between vocabulary knowledge and sensitivity to underlying lexical dimensions. Our preliminary hypothesis was that larger effects of lexical variables should be produced by readers with less vocabulary knowledge. This was motivated by the perspective that readers who are less skilled should be more reliant on controlled lexical processing mechanisms, which ought to show a greater influence of lexical characteristics (LaBerge & Samuels, 1974; Stanovich, 1980). As discussed in the Introduction, a number of studies (e.g., P. Brown et al., 1994; Butler & Hains, 1979; Chateau & Jared, 2000; Schilling et al., 1998) support the view that skilled readers produce smaller effects of lexical variables. Of course, most of these studies did not control for overall processing speed, making it possible that these results simply reflect spurious Group × Treatment interactions that are driven by scaling differences (Faust et al., 1999).

To address this issue, we examined participants' standardized regression coefficients for the three principal components of interest, in which PC1 reflected structural properties, PC2 neighborhood size, and PC3 word frequency/semantics. The results from speeded pronunciation were very clear. There was a negative relationship between vocabulary knowledge and each of the three components, consistent with skilled readers being more reliant on relatively automatic lexical processing mechanisms, and hence showing less influence of word characteristics. The findings from lexical-decision task are more intriguing. Although high-vocabulary-participants indeed produced smaller neighborhood (PC2) effects, effects of structural properties (PC1) and word frequency/semantics (PC3) were only marginally related to vocabulary knowledge. These results suggest that word-frequency effects in lexical decision are not negatively related to print exposure/vocabulary knowledge when processing speed is controlled for, converging with reports by Butler and Hains (1979), Lewellen et al. (1993), and Sears et al. (2008).

The foregoing findings are also compatible with results from the Ratcliff et al. (2010) lexical decision study. Recall that their participants (young adults, older adults, very old adults) responded to high-frequency (HF), low-frequency (LF), and very-low-frequency (VLF) words in lexical decision, and that vocabulary knowledge was measured for each participant. We plotted the drift rates for the three frequency classes of words as a function of vocabulary knowledge and participant age (see Figure 9). As shown, drift rate was generally positively correlated with vocabulary knowledge, as described earlier. However, the relationships between drift rate and vocabulary knowledge for the different frequency classes were represented by parallel lines, indicating that the magnitude of word-frequency effects was not modulated by vocabulary knowledge.[15] In addition to the pattern described above, we also found that when participants showed more of an influence of lexical variables in lexical decision, this was accompanied by lower drift rates and longer nondecision times. The major exception to this trend was the positive relationship between word-frequency/semantics effects and drift rate, wherein larger word-frequency/semantics effects yielded steeper drift rates.

Why are word-frequency effects smaller for high-vocabulary-knowledge participants in speeded pronunciation, but not in lexical decision? In speeded pronunciation, the effect of word frequency is limited to lexical processes and possible production characteristics (Balota & Chumbley, 1985), but word frequency affects both lexical access and postlexical decision-making stages in lexical decision (e.g., Balota & Chumbley, 1984). If the negative relationship between word-frequency effects and vocabulary knowledge reflects lexical processes, but word-frequency effects in lexical decision performance predominantly reflect decision-making mechanisms, this might explain the dissociation between the two tasks. To further examine this in a more fine-grained manner, we examined the correlations between vocabulary knowledge and word-frequency effects at different regions of the RT distribution across both tasks (see Table 12). For each participant, we first obtained the following quantiles (.1, .3, .5, .7, .9; Ratcliff, 1979) for high-frequency (25% most frequent) and low-frequency (25%

[15] The only exception to this trend was seen in the line representing HF words in the young adult data, in which there appears to be no relationship between drift rate and vocabulary knowledge. However, this needs to be qualified by the instability of the HF word drift rate estimates for these young adults, which were driven by fitting problems associated with very low error rates and some problematic RTs (see Ratcliff et al., 2010; Ratcliff & Tuerlinckx, 2002).

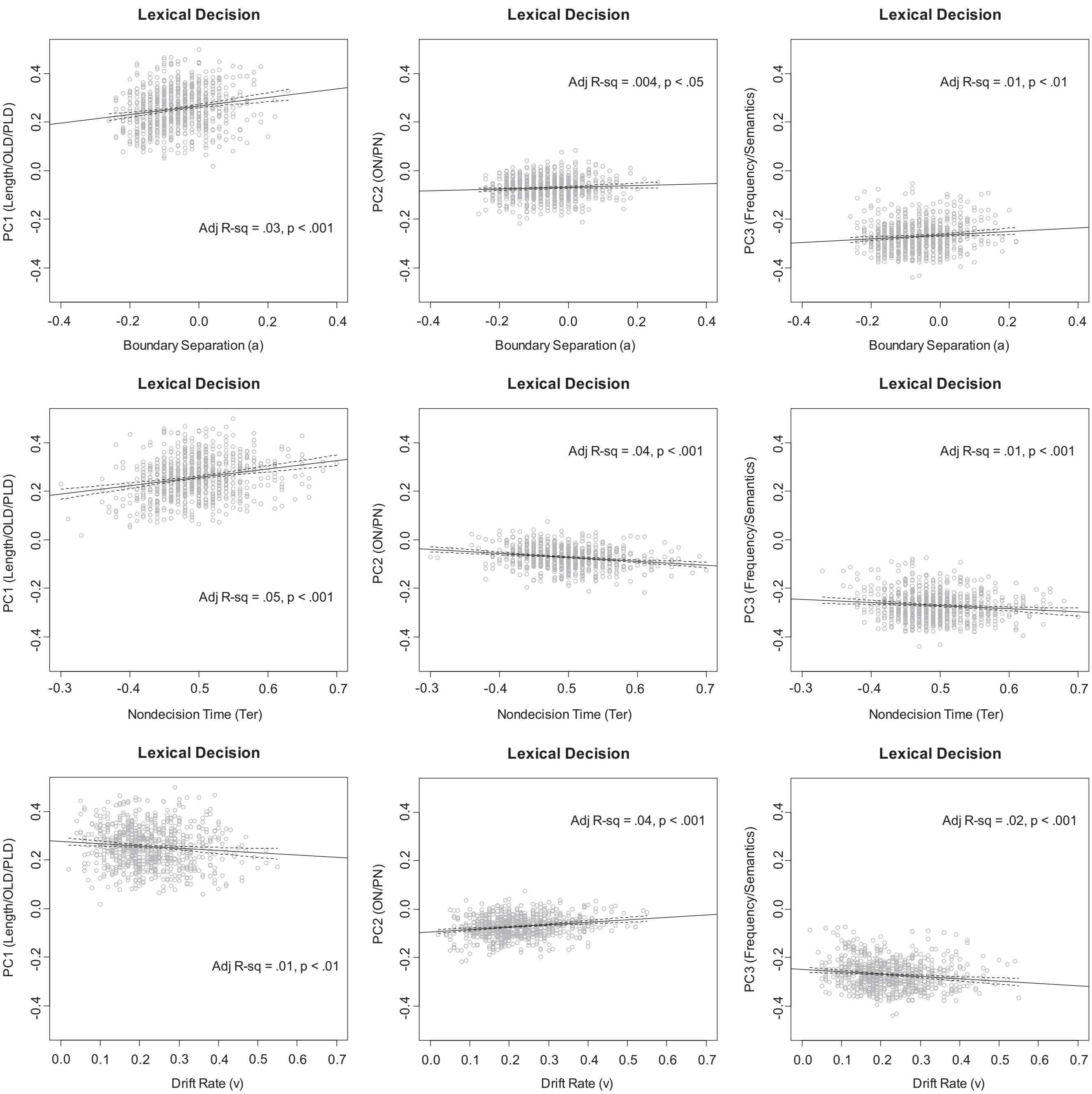


Figure 7. Scatterplots (with 95% confidence intervals) between principal component (PC) effects and diffusion model parameters in lexical decision performance. OLD = orthographic Levenshtein distance; PLD = phonological Levenshtein distance; ON = orthographic neighborhood size; PN = phonological neighborhood size.

least frequent) words, which yielded frequency effects for the fastest to the slowest responses in the RT distribution. Instead of raw RTs, we used standardized residuals as the dependent variable, which control for correlated variables (e.g., length and neighborhood size) and processing speed.

The between-task differences were quite striking. In speeded pronunciation, the correlations between vocabulary knowledge and word-frequency effects were negative and generally reliable throughout the RT distribution. In contrast, for lexical decision, reliable positive correlations between vocabulary knowledge and frequency effects were seen only in the fastest quantiles. These findings suggest that higher vocabulary participants, compared to individuals with less vocabulary, are better able to take advantage of familiarity-based information (such as word frequency) in lexical decision; this facilitates responses to high-frequency words, particularly those at the leading edge. This suggestion is compat-

Table 11
Correlations Between the Participant-Level Principal Component Effects

Principal component effects	PC1	PC2	PC3
	Speeded pronunciation[a]		
1. PC1	—	−.565***	−.516***
2. PC2		—	.547***
3. PC3			—
	Lexical decision[b]		
1. PC1	—	−.371***	−.411***
2. PC2		—	.197***
3. PC3			—

Note. PC = principal component.
[a] $n = 470$. [b] $n = 819$.
*** $p < .001$.

ible with the finding that steeper drift rates accompany larger word-frequency/semantics effects (see Figure 7). Specifically, participants who are more sensitive to familiarity-based information (which can be used to drive lexical decisions) are also able to accumulate information about a letter string more rapidly.

Obviously, these explanations are speculative and post hoc, and future investigations are needed. However, these results suggest that for the ELP dataset, speeded pronunciation performance is more consistently related to vocabulary knowledge in a predictable manner than lexical decision performance. This is surprising as one may have expected lexical processing to be diluted in pronunciation due to the reliance on the sublexical pathway. In this light, it is noteworthy that the ELP included a large number of low-frequency long words, which might have attenuated the influence of sublexical processing.

Although our analyses focused on vocabulary knowledge, it is important to remember that reading skill ultimately is a complex, multidimensional concept that subsumes decoding speed (Perfetti, 1985; Stanovich, 1980, 1986), orthographic (Ehri, 2005) and phonological (Rayner et al., 2001) processing, and sentential-level comprehension. Ideally, reading ability should be assessed using multiple constructs, but the archival nature of the ELP dataset constrained us to assess reading skill rather narrowly by vocabulary knowledge, that is, knowledge of word forms and meaning. Nonetheless, despite its limitations, the utility of vocabulary knowledge for predicting word recognition performance in the present study is quite impressive.

Relationships Between Ex-Gaussian and Diffusion Model Parameters

In the present paper, we fit individual-level RT data from the lexical-decision task to the ex-Gaussian distribution and to the diffusion model, and were able to obtain both ex-Gaussian and diffusion model parameters for each participant. Consistent with the notion that a participant's slowest RTs reflect attentional lapses (Coyle, 2003; Larson & Alderton, 1990), a number of researchers (e.g., Schmiedek et al., 2007, & Tse et al., 2010) have suggested that τ may be more related to controlled processing (see Matzke & Wagenmakers, 2009, for more examples). Although ex-Gaussian parameters provide a finer grained characterization of RTs, diffusion model parameters additionally map onto psychological processes. In the preceding section, we considered how vocabulary knowledge relates to ex-Gaussian and diffusion model parameters, but the relationships between the two sets of parameters are also interesting. There is evidence from simulations (e.g., Spieler, 2001) that of the three ex-Gaussian parameters, τ is most closely associated with drift rate, whereas μ is most closely associated with the nondecision component. These trends were replicated in the present analyses (see Table 8). Having said that, we also observed a relationship between τ and boundary separation, in line with Matzke and Wagenmakers' (2009) claim that ex-Gaussian parameters cannot be uniquely ascribed to specific parameters of the diffusion model. Given this qualification, the very strong positive correlation between μ and nondecision time ($r = .89$) is at least consistent with the possibility that the leading edge of a participant's lexical-decision task RT distribution primarily reflects a nondecision component, whereas the tail of the distribution (i.e., τ) serves as a marker of decision processes. Clearly, as with all dependent measures, converging evidence is needed to understand the relationship between the parameters and underlying mechanisms.

Are There Distinct Types of Readers?

Baron and Strawson (1976) were the first to propose that there were distinct types of readers who selectively relied on lexical or sublexical processing, a perspective that predicts that participants who show larger frequency effects should show smaller length effects, and vice versa. Our study makes it clear that although there is undoubtedly a great deal of variation in how people recognize words, there is no clear evidence for a trade-off between readers' sensitivities to different lexical characteristics. Instead, we found strong positive correlations between effects of the three principal components after overall processing speed is controlled for, which is more compatible with P. Brown et al.'s (1994) finding that increased reliance on lexical processing is associated with increased reliance on sublexical processing. Of course, this pattern can be easily accommodated by the general claim that as readers get better, they rely increasingly less on controlled processing (Stanovich, 1980), which is reflected in a decreased sensitivity to the various characteristics of a word. Also, it is noteworthy that there may be a subset of participants that show trade-offs, but clearly these participants are not prevalent in this database.

Modeling Individual Differences

The overarching goal of this work was to characterize individual differences in word recognition behavior. For instance, we found that an individual's vocabulary knowledge was related to their sensitivity to lexical characteristics (see Table 9). This characterization is generally consistent with the proposal that better readers utilize a more automatic reading process (Stanovich, 1980). However, this obviously does not explain how or why some individuals become better readers, how or why they shift to an automatic reading process, and how or why more automatic processes are less sensitive to lexical variables. Building a theory that addresses these questions will require convergent methodologies.

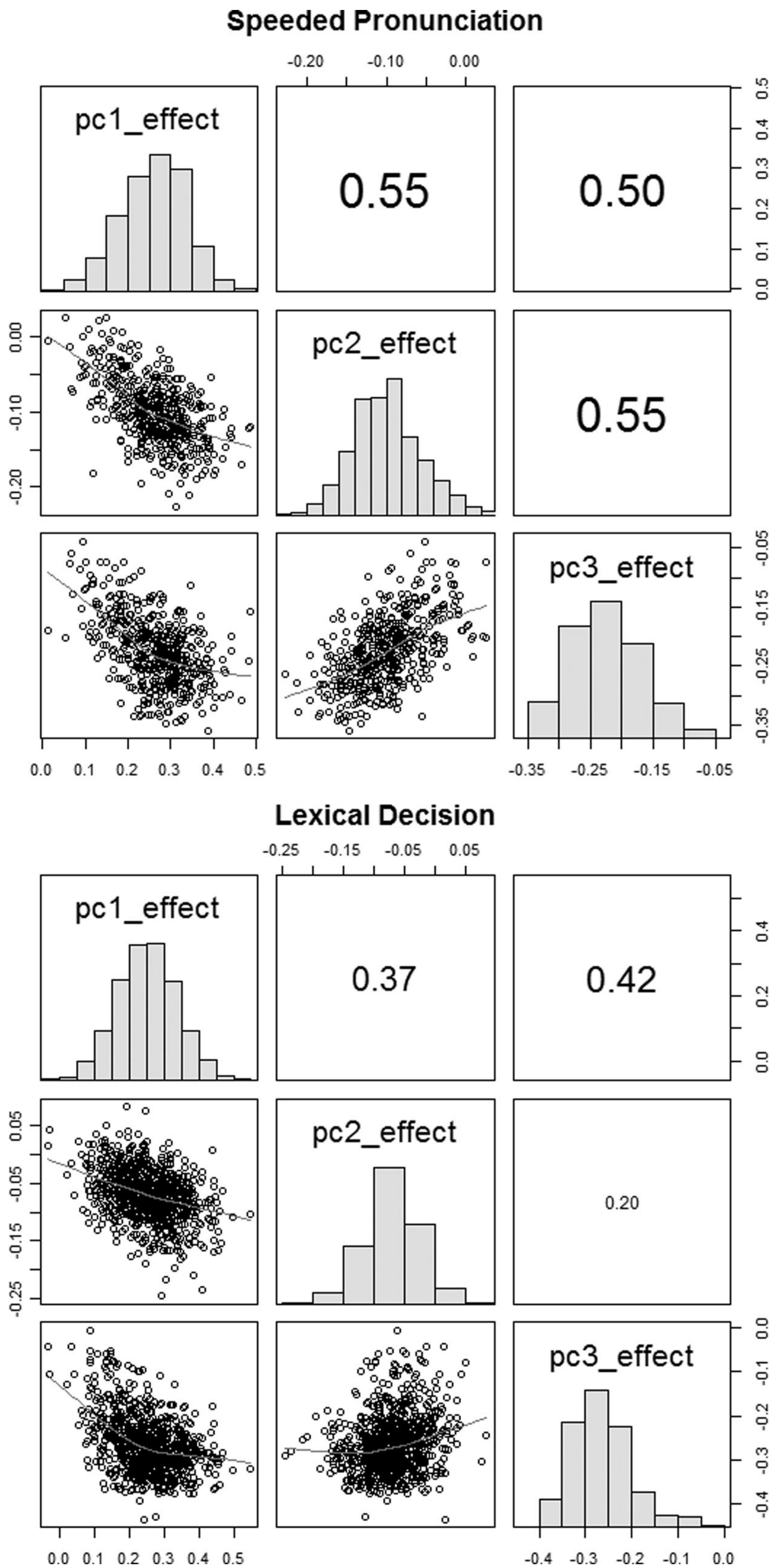


Figure 8. Scatterplots between principal component (PC) effects in speeded pronunciation (top) and lexical decision (bottom).

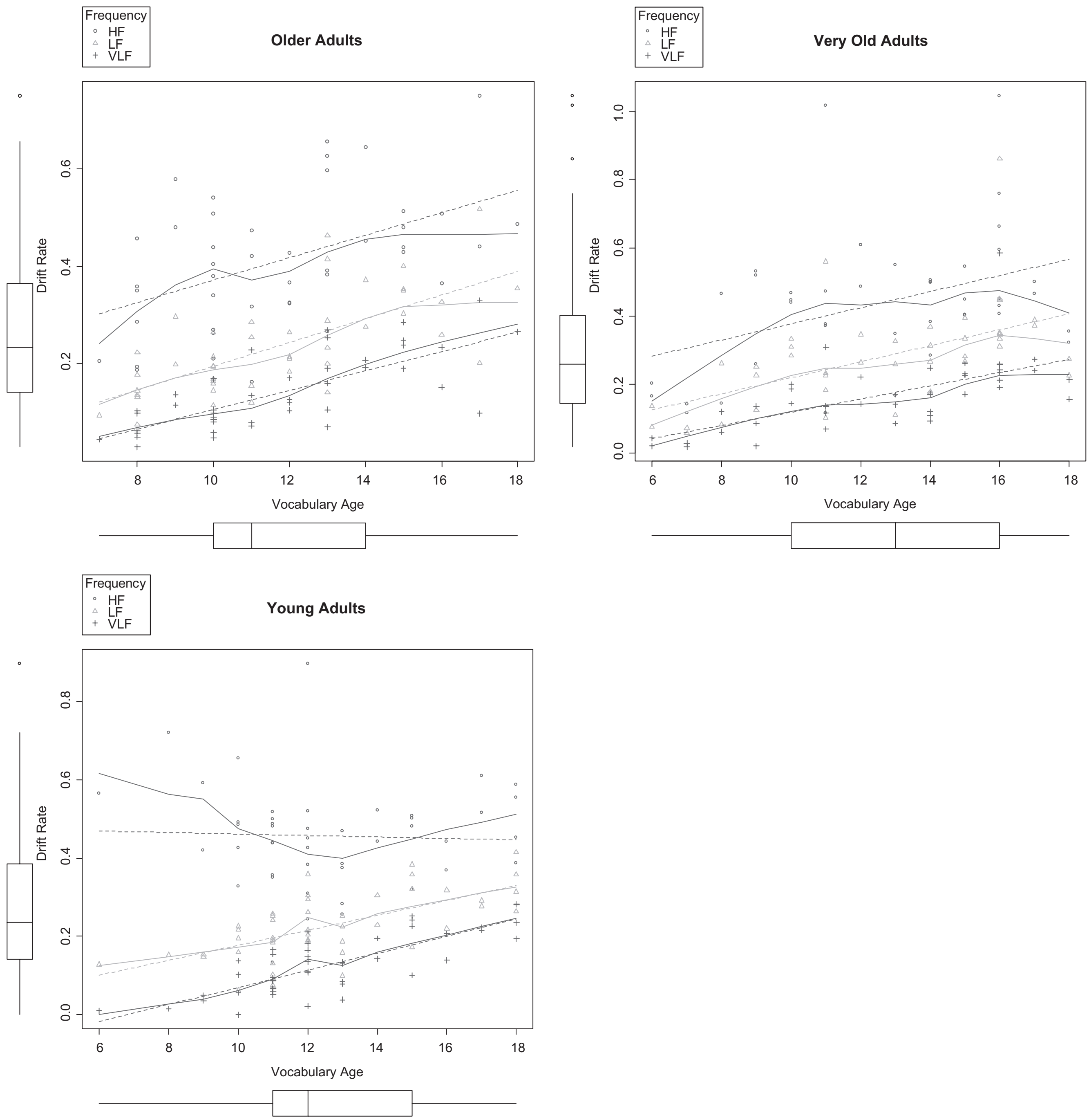


Figure 9. Scatterplots between vocabulary knowledge and drift rates for high-frequency (HF), low-frequency (LF), and very-low-frequency (VLF) words in young adults, older adults, and very old adults. Data are from Ratcliff et al.'s (2010) study.

Traditional experimental (e.g., Lewellen et al., 1993; Sears et al., 2008) and neurobiological (e.g., Frost et al., 2009; Pugh et al., 1997) studies will continue to inform theories of individual differences. In addition, computational modeling may be particularly well suited to addressing the behavioral patterns we presented. To date, models have focused on normative and dyslexic reading behavior (e.g., Coltheart et al., 2001; Harm & Seidenberg, 1999; McClelland & Rumelhart, 1981; Perry, Ziegler, & Zorzi, 2007, 2010; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989; Sibley & Kello, 2004), with relatively less attention given to individual differences within a typically developing population (but see Zevin & Seidenberg, 2006).

Computational modeling has offered numerous insights about reading behaviors, and can potentially contribute to theoretical

Table 12
Correlations Between Vocabulary Knowledge and Word-Frequency Effects at Different Regions of the RT Distribution for Speeded Pronunciation and Lexical Decision

	Quantiles				
	Fastest RTs → Slowest RTs				
Task	.1	.3	.5	.7	.9
Speeded pronunciation	−.073	−.216***	−.315***	−.389***	−.319***
Lexical decision	.141***	.093***	−.017	.004	.051

Note. RT = response time.
*** $p < .001$.

accounts of individual differences in word recognition. To do this, models will have to go beyond capturing group-level effects and instead help us understand the source of these individual differences. For example, can the individual differences reported in the present paper be simulated by training a population of models, such that different models are imbued with slightly different cognitive mechanisms or capacities, or are exposed to different learning experiences (e.g., Rueckl, 2010)? Alternatively, one could also use the computational modeling approach to test the hypothesis that individual differences phenomena reflect varying positions along the developmental trajectory of a single model. Clearly, the present results will provide fodder for capturing the nature of individual differences in standard word recognition tasks.

Limitations and Future Directions

In the present study, we conducted trial-level analyses of the ELP, an online behavioral repository of nearly four million speeded pronunciation and lexical decision trials from over 1,200 participants. We found evidence that word recognition performance is reliable, and that there are systematic relationships between vocabulary knowledge and word recognition performance. Of course, there are a number of limitations inherent in the present work. For example, due to the archival nature of the dataset, our only measure of reading skill is vocabulary knowledge. In addition, we only considered the linear effects of lexical predictors that were available for most (if not all) the ELP words. Hence, we did not model nonlinear effects (Baayen et al., 2006) nor did we include variables that have been theoretically influential, such as phonological consistency (Chateau & Jared, 2003) and morphological characteristics (Baayen et al., 2006; see Yap & Balota, 2009, for further discussion). Third, despite the scope of the ELP, one could argue that a sample based on college students, who are selected in part for their vocabulary knowledge, will show a restricted range of vocabulary knowledge relative to readers in general. Hence, we might be underestimating the strength of the relationships between vocabulary knowledge and word recognition performance. Finally, it would be useful to have multiple lexical processing measures on the same participants to determine the consistency of lexical processing across tasks including eye fixations and possibly semantic categorization (see Schilling et al., 1998, for a step in this direction).

In summary, the present study provides the first comprehensive analysis of individual differences in speeded word recognition tasks across a large set of stimuli and participants. The results produced clear evidence of stable lexical processing characteristics at the individual level. In considering how to best conceptualize lexical processing (an important and critical step in literacy), we believe it will be increasingly important to develop models that not only capture mean level performance but also capture the variability across individuals. The present study provides a stable set of empirical relationships that will be necessary for the next step in understanding visual word recognition.

References

Adams, M. J. (1990). *Beginning to read: Think and learning about print.* Cambridge, MA: MIT Press.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review, 4,* 439–461.

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

Ashby, J., Rayner, K., & Clifton, C., Jr. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *Quarterly Journal of Experimental Psychology, 58A,* 1065–1086.

Baayen, R. H., Feldman, L. F., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory & Language, 53,* 496–512.

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., & Chumbley, J. I. (1985). The locus of word frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory & Language, 24,* 89–106.

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,* 336–345.

Balota, D. A., & Ferraro, F. R. (1996). Lexical, sublexical, and implicit memory processes in healthy young and healthy older adults and in individuals with dementia of the Alzheimer type. *Neuropsychology, 10,* 82–95.

Balota, D. A., & Spieler, D. H. (1998). The utility of item-level analyses in model evaluation: A reply to Seidenberg and Plaut. *Psychological Science, 9,* 238–241.

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

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.

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.

Balota, D. A., Yap, M. J., Cortese, M. J., & Watson, J. M. (2008). Beyond mean response latency: Response time distributional analyses of semantic priming. *Journal of Memory & Language, 59,* 495–523.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71,* 230–244.

Baron, J., & Strawson, C. (1976). Use of orthographic and word-specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 386–393.

Brown, P., Lupker, S. J., & Colombo, L. (1994). Interacting sources of information in word naming: A study of individual differences. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 537–554.

Brown, S., & Heathcote, A. (2003). QMLE: Fast, robust and efficient estimation of distribution functions based on quantiles. *Behavior Research Methods, Instruments & Computers, 35,* 485–492.

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

Castel, A. D., Balota, D. A., Hutchison, K. A., Logan, J. M., & Yap, M. J. (2007). Spatial attention and response control in healthy younger and older adults and individuals with Alzheimer's disease: Evidence for disproportionate selection impairments in the Simon task. *Neuropsychology, 21,* 170–182.

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind, 11,* 63–65.

Cerella, J. (1990). Aging and information processing rate. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (3rd ed., pp. 201–221). San Diego, CA: Academic Press.

Cerella, J. (1991). Age effects may be global, not local: Comment on Fisk and Rogers (1991). *Journal of Experimental Psychology: General, 120,* 215–223.

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

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.

Coyle, T. R. (2003). A review of the worst performance rule: Evidence, theory, and alternative hypotheses. *Intelligence, 31,* 567–587.

Cutler, A. (1981). Making up materials is a confounded nuisance: Or will we be able to run any psycholinguistic experiments in 1990? *Cognition, 10,* 65–70.

Devine, P. G. (1989). Stereotypes and prejudice. Their automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5–18.

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading, 9,* 167–188.

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.

Ferrand, L., & New, B. (2003). Syllabic length effects in visual word recognition and naming. *Acta Psychologica, 113,* 167–183.

Frost, S. J., Landi, N., Mencl, W. E., Sandak, R., Fulbright, R. K., Tejada, E. T., . . . Pugh, K. R. (2009). Phonological awareness predicts activation patterns for print and speech. *Annals of Dyslexia, 59,* 78–97.

Goh, W. D., Suárez, L., Yap, M. J., & Tan, S. H. (2009). Distributional analyses in auditory lexical decision: Neighborhood density and word-frequency effects. *Psychonomic Bulletin & Review, 16,* 882–887.

Goodwin, C. J. (2009). *Research in psychology: Methods and design* (6th ed.). Hoboken, NJ: Wiley.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review, 106,* 491–528.

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 time as a function of foreperiod duration. *Journal of Experimental Psychology, 69,* 382–386.

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., Levy, B. A., & Rayner, K. (1999). The role of phonology in the activation of word meanings during reading: Evidence from proofreading and eye movements. *Journal of Experimental Psychology: General, 128,* 219–264.

Katz, L., Brancazio, L., Irwin, J., Katz, S., Magnuson, J., & Whalen, D. (in press). What lexical decision and naming tell us about reading. *Reading & Writing: An Interdisciplinary Journal.*

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

Kimberley, T. J., Khandekar, G., & Borich, M. (2008). fMRI reliability in subjects with stroke. *Experimental Brain Research, 186,* 183–190.

Kinoshita, S., & Hunt, L. (2008). RT distribution analysis of category congruence effects with masked primes. *Memory & Cognition, 36,* 1324–1334.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 293–323.

Larson, G. E., & Alderton, D. L. (1990). Reaction time variability and intelligence: A "worst-performance" analysis of individual differences. *Intelligence, 14,* 309–325.

Lewellen, M. J., Goldinger, S. D., Pisoni, D. B., & Greene, B. G. (1993). Lexical familiarity and processing efficiency: Individual differences in naming, lexical decision, and semantic categorization. *Journal of Experimental Psychology: General, 122,* 316–330.

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.

Lowe, C., & Rabbitt, P. (1998). Test/re-test reliability of the CANTAB and ISPOCD neuropsychological batteries: Theoretical and practical issues. *Neuropsychologia, 36,* 915–923.

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

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

Lyons, K. E., Kellas, G., & Martin, M. (1995). Inter- and intra-individual differences in semantic priming among young and older adults. *Experimental Aging Research: An International Journal Devoted to the Scientific Study of the Aging Process, 21,* 221–237.

Matzke, D., & Wagenmakers, E.-J. (2009). Psychological interpretation of ex–Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review, 16,* 798–817.

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.

McGill, W. J., & Gibbon, J. (1965). The general-gamma distribution and reaction times. *Journal of Mathematical Psychology, 2,* 1–18.

Miller, G. A. (1990). WordNet: An on-line lexical database. *International Journal of Lexiography, 3,* 235–312.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106,* 226–254.

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

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.

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review, 113,* 327–357.

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

Palmer, E. M., Horowitz, T. S., Torralba, A., & Wolfe, J. (2011). What are the shapes of response time distributions in visual search? *Journal of Experimental Psychology: Human Perception and Performance, 37,* 58–71.

Patterson, K. (1982). The relation between reading and phonological coding: Further neuropsychological observations. In A. W. Ellis (Ed.), *Normality and pathology in cognitive functioning* (pp. 77–112). London, England: Academic Press.

Perfetti, C. A. (1983). Individual differences in verbal processes. In F. F. Dillon & R. R. Schmeck (Eds.), *Individual differences in cognition* (Vol. 1, pp. 65–104). New York, NY: Academic Press.

Perfetti, C. A. (1985). *Reading ability.* New York, NY: Oxford University Press.

Perfetti, C. A. (1994). Psycholinguistics and reading ability. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 849–894). San Diego, CA: Academic Press.

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.

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.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1989). Positron emission tomographic studies of the processing of single words. *Journal of Cognitive Neuroscience, 1,* 153–170.

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

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

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 55–85). Hillsdale, NJ: Erlbaum.

Pugh, K., Shaywitz, B. A., Shaywitz, S. E., Shankweiler, D. P., Katz, L., Fletcher, J. M., . . . Gore, J. C. (1997). Predicting reading performance from neuroimaging profiles: The cerebral basis of phonological effects in printed word identification. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 299–318.

Rastle, K., Croot, K., Harrington, J., & Coltheart, M. (2005). Characterizing the motor execution stage of speech production: Consonantal effects on delayed naming latency and onset duration. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1083–1095.

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

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.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest, 2,* 31–74.

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

Rohrer, D., & Wixted, J. T. (1994). An analysis of latency and interresponse time in free recall. *Memory & Cognition, 22,* 511–524.

Rueckl, J. (2010, November). *Are canonical models really canonical? And what if they are not?* Paper presented at the Pre-Psychonomics Visual Word Recognition Conference, St. Louis, MO.

Salthouse, T. A. (1985). Speed of behavior and its implication for cognition. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (2nd ed., pp. 400–426). New York, NY: Van Nostrand Reinhold.

Schilling, H. E. H., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition, 26,* 1270–1281.

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.

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

Sears, C. R., Siakaluk, P. D., Chow, V. C., & Buchanan, L. (2008). Is there an effect of print exposure on the word frequency effect and the neighborhood size effect? *Journal of Psycholinguistic Research, 37,* 269–291.

Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition, 19,* 1–30.

Seidenberg, M. S. (1990). Lexical access: Another theoretical soupstone? In D. Balota, G. 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.

Shallice, T., & Warrington, E. K. (1980). Single and multiple component central dyslexic syndromes. In M. Coltheart, K. E. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 119–146). London, England: Routledge and Kegan Paul.

Shaoul, C., & Westbury, C. (2006). Word frequency effects in high-dimensional co-occurrence models: A new approach. *Behavior Research Methods, 38,* 190–195.

Shaoul, C., & Westbury, C. (2009). *A USENET corpus (2005–2009).* Edmonton, Canada: University of Alberta. Retrieved from http://www.psych.ualberta.ca/~westburylab/downloads/usenetcorpus.download.html

Shaoul, C., & Westbury, C. (2010). Exploring lexical co-occurrence space using HiDEx. *Behavior Research Methods, 42,* 393–413.

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

Sibley, D. E., & Kello, C. T. (2004). A computational exploration of double dissociations: Modes of processing instead of components of processing. *Cognitive Systems Research, 6,* 61–69.

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. (2001). Modeling age-related changes in information processing. *European Journal of Cognitive Psychology, 13,* 217–234.

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.

Stanovich, K. E. (1980). Towards an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, 16,* 32–71.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21,* 360–406.

Stanovich, K. E., & Cunningham, A. E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory & Cognition, 20,* 51–68.

Stanovich, K. E., & West, R. F. (1983). On priming by a sentence context. *Journal of Experimental Psychology: General, 112,* 1–36.

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly, 24,* 402–433.

Staub, A., White, S. J., Drieghe, D., Hollway, E. C., & Rayner, K. (2010). Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 1280–1293.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science, 29,* 41–78.

Tainturier, M. J., Tremblay, M., & Lecours, A. R. (1992). Education level and the word frequency effect: A lexical decision investigation. *Brain & Language, 43,* 460–474.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General, 124,* 107–136.

Trochim, W., & Donnelly, J. P. (2006). *The research methods knowledge database* (3rd ed.). Mason, OH: Atomic Dog.

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.

Tse, C.-S., Hutchison, K. A., & Li, Y. (2011). Effects of contextual similarity and target-repetition proportion on negative priming in RT distributional analyses. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 180–192.

Unsworth, S. J., & Pexman, P. M. (2003). The impact of reader skill on phonological processing in visual word recognition. *The Quarterly Journal of Experimental Psychology, 56A,* 63–81.

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

Wechsler, D. (1997). *The Wechsler Adult Intelligence Scale III.* San Antonio, TX: Psychological Corporation, Harcourt Brace.

Yap, M. J., & Balota, D. A. (2007). Additive and interactive effects on response time distributions in visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 274–296.

Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory & Language, 60,* 502–529.

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

Yap, M. J., Balota, D. A., Tse, C.-S., & Besner, D. (2008). On the additive effects of stimulus quality and word frequency in lexical decision: Evidence for opposing interactive influences revealed by RT distributional analyses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 495–513.

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 & Language, 61,* 303–325.

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.

Zachary, R. A. (1992). *Shipley Institute of Living Scale–Revised, manual.* Los Angeles, CA: Western Psychological Services.

Zevin, J. D., & Seidenberg, M. S. (2006). Consistency effects and individual differences in nonword naming: A comparison of current models. *Journal of Memory & Language, 54,* 145–160.

(*Appendix follows*)

Appendix A

Supplementary Tables Reporting Analyses Based on Effects of Individual Predictors

Table A1
Descriptive Statistics for Participant-Level Standardized Regression Coefficients as a Function of Task and Trial Type

Predictor effects	Overall	S1	S2	Odd	Even	Item
Speeded pronunciation[a]						
Word frequency	−.229	−.234	−.223	−.227	−.226	−.410
Number or morphemes	−.080	−.078	−.082	−.081	−.078	−.139
Number of syllables	.179	.177	.180	.177	.177	.431
Number of letters	.133	.134	.134	.132	.132	.158
Orthographic neighborhood size	−.024	−.022	−.026	−.022	−.025	−.063
Phonological neighborhood size	.044	.044	.044	.042	.045	.075
Orthographic Levenshtein distance	.039	.031	.047	.039	.039	.162
Phonological Levenshtein distance	.093	.101	.088	.092	.092	.119
Number of senses	−.022	−.023	−.020	−.022	−.021	−.030
Semantic neighborhood density	.013	.012	.014	.013	.015	.016
Lexical decision[b]						
Word frequency	−.270	−.276	−.270	−.267	−.270	−.545
Number or morphemes	−.058	−.057	−.061	−.059	−.056	−.103
Number of syllables	.124	.128	.126	.122	.124	.325
Number of letters	.150	.150	.153	.150	.148	.200
Orthographic neighborhood size	.002	−.001	.005	.002	.001	−.024
Phonological neighborhood size	.046	.047	.048	.046	.047	.092
Orthographic Levenshtein distance	.066	.065	.068	.066	.064	.207
Phonological Levenshtein distance	.080	.078	.086	.078	.080	.127
Number of senses	−.031	−.031	−.030	−.032	−.028	−.061
Semantic neighborhood density	−.007	−.005	−.010	−.007	−.007	−.021

Note. S1 = Session 1; S2 = Session 2.
[a] $n = 470$. [b] $n = 819$.

Table A2
Correlations Between Session 1 (S1) and Session 2 (S2) Participant-Level Standardized Regression Coefficients, and Odd- and Even-Numbered Trial Participant-Level Standardized Regression Coefficients

	Speeded pronunciation		Lexical decision	
Predictor effects	S1–S2	Odd–Even	S1–S2	Odd–Even
Word frequency	$.685^{***}$	$.753^{***}$	$.577^{***}$	$.742^{***}$
Number of morphemes	$.238^{***}$	$.305^{***}$	$.135^{***}$	$.155^{***}$
Number of syllables	$.320^{***}$	$.472^{***}$	$.257^{***}$	$.284^{***}$
Number of letters	$.416^{***}$	$.493^{***}$	$.459^{***}$	$.483^{***}$
Orthographic neighborhood size	$.194^{***}$	$.179^{***}$	$.087^{*}$	$.068^{\dagger}$
Phonological neighborhood size	$.129^{**}$	$.151^{**}$	$.113^{***}$	$.112^{**}$
Orthographic Levenshtein distance	$.353^{***}$	$.355^{***}$	$.101^{**}$	$.146^{***}$
Phonological Levenshtein distance	$.121^{**}$	$.133^{**}$	.038	.036
Number of senses	$.126^{**}$	$.124^{**}$	$.103^{**}$	$.084^{*}$
Semantic neighborhood density	$.078^{\dagger}$	.027	.010	−.017

$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

(*Appendix continues*)

Table A3
Correlations Between Participant-Level Standardized Regression Coefficients, Vocabulary Knowledge, and Mean RT

	Speeded pronunciation[a]		Lexical decision[b]	
Predictor effects	Vocabulary knowledge	*M* RT	Vocabulary knowledge	*M* RT
Word frequency	.298***	−.446***	−.080*	.186***
Number of morphemes	.236***	−.106*	−.009	.035
Number of syllables	−.143**	.180***	.071*	−.106**
Number of letters	−.128**	−.120*	−.057	.281***
Orthographic neighborhood size	.148**	−.154**	.099**	.120***
Phonological neighborhood size	.043	.024	.183***	−.141***
Orthographic Levenshtein distance	−.094*	−.214***	−.066†	−.154***
Phonological Levenshtein distance	−.066	.077	−.063†	.062†
Number of senses	.011	−.134**	−.094**	.038
Semantic neighborhood density	−.165***	−.016	−.093**	−.024

Note. RT = response time.
[a] $n = 456$. [b] $n = 792$.
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table A4
Correlations Between Participant-Level Standardized Regression Coefficients and Diffusion Model Parameters

Predictor effects	a	z	T_{er}	η	s_z	s_t	v_{word}	$v_{nonword}$
Word frequency	.151***	.077*	−.054	−.237***	.099**	.136***	−.157***	.328***
Number or morphemes	.032	−.005	.024	−.019	.058	.085*	−.041	.089*
Number of syllables	.010	.026	−.020	.122***	.031	−.154***	.136***	−.200***
Number of letters	.219***	.221***	.195***	−.162***	.052	.127***	−.162***	.168***
Orthographic neighborhood size	.190***	.190***	.010	−.024	.096**	−.066†	.071*	−.064†
Phonological neighborhood size	−.039	−.002	−.088*	.006	−.053	−.207***	.153***	−.181***
Orthographic Levenshtein distance	−.126***	−.110**	−.095**	.069†	−.062†	−.062†	.047	−.064†
Phonological Levenshtein distance	.031	.033	.054	−.012	.051	−.056	−.078*	.053
Number of senses	.036	.014	−.027	−.033	.069†	−.074*	−.078*	.099**
Semantic neighborhood density	−.060†	−.047	−.058	.007	−.032	.030	−.105**	.072*

† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Received January 3, 2011
Revision received April 7, 2011
Accepted April 27, 2011 ■