Moving beyond Coltheart's *N*:
A new measure of orthographic similarity

TAL YARKONI AND DAVID BALOTA
Washington University, St. Louis, Missouri

AND

MELVIN YAP
National University of Singapore, Singapore

Visual word recognition studies commonly measure the orthographic similarity of words using Coltheart's orthographic neighborhood size metric (ON). Although ON reliably predicts behavioral variability in many lexical tasks, its utility is inherently limited by its relatively restrictive definition. In the present article, we introduce a new measure of orthographic similarity generated using a standard computer science metric of string similarity (Levenshtein distance). Unlike ON, the new measure—named *orthographic Levenshtein distance 20* (OLD20)—incorporates comparisons between all pairs of words in the lexicon, including words of different lengths. We demonstrate that OLD20 provides significant advantages over ON in predicting both lexical decision and pronunciation performance in three large data sets. Moreover, OLD20 interacts more strongly with word frequency and shows stronger effects of neighborhood frequency than does ON. The discussion section focuses on the implications of these results for models of visual word recognition.

Although visual word recognition seems relatively effortless for skilled readers, the high degree of similarity across many patterns makes it a remarkable skill. When a word is visually presented, it likely overlaps to varying degrees in orthography with other words, resulting in early partial activation of multiple orthographic representations. For example, the letter string C-A-T is likely to activate not only the orthographic representation for CAT, but also the representations of other words containing those letters—for example, CUT, BAT, CAN, CATS, and CAST. One might expect such overlap to present a major source of interference when attempting to identify words—a prediction made explicit in some influential models of visual letter–word processing (see, e.g., McClelland & Rumelhart, 1981). Yet skilled readers are able to uniquely identify most words in a fraction of a second, and a wealth of empirical evidence suggests that—if anything—words that are orthographically similar to many other words are recognized faster than are more distinctive words (for a review, see Andrews, 1997).

The most common measure of orthographic similarity in the psychological literature is Coltheart's *N* (ON; Coltheart, Davelaar, Jonasson, & Besner, 1977, cited nearly 600 times, according to the 2007 ISI Web of Science), defined simply as the number of words that can be produced by changing a letter in a word of the same length. Although ON has a demonstrable influence on measures of lexical access (Andrews, 1997), it is surely too restrictive a metric. As Davis (2006) pointed out, there is evidence of activation from TRIAL to TRAIL, WIDOW to WINDOW, and PLANE to PLANET, none of which would be neighbors by Coltheart's binary metric. To explain such effects, researchers have recently begun to develop a number of alternative orthographic coding schemes (for a review, see Davis & Bowers, 2006). These approaches have been useful in accounting for data from factorial experiments with relatively small sets of items but have yet to be extended to larger databases of words. Importantly, there have been no previous attempts to map similarity onto the full adult lexicon.

In the present study, we introduce a measure of orthographic similarity that is based on principles similar to those of ON, but that is less restrictive. The new measure—termed *orthographic Levenshtein distance 20* (OLD20)—overcomes two fundamental constraints that limit the predictive utility of ON. First, ON is a binary measure. Two words either are or are not neighbors; they cannot be more or less neighborly, despite the fact that perceptual similarity between words clearly varies in a graded manner. Second, ON restricts the definition of neighbors to pairs of words that can be generated by a single letter substitution, despite the aforementioned fact that insertion, deletion, or transposition operations can also result in highly similar words that strongly prime one another (e.g., WIDOW → WINDOW, PLANET → PLANE, or TRAIL → TRIAL). As a result, ON is of limited use for long words, because, as will be discussed below, most long words do not have any neighbors using the ON metric.

T. Yarkoni, tyarkoni@wustl.edu

The measure we introduce incorporates a graded and more flexible definition of similarity. It is based on *Levenshtein distance* (LD; Levenshtein, 1966), a standard computer science metric of string edit distance. The LD between two words is the minimum number of substitution, insertion, or deletion operations required to turn one word into the other. Although LD and related metrics play a central role in a wide range of practical applications (e.g., spell checking, speech recognition, and DNA analysis), they have not been systematically applied to the psychological study of visual word recognition. Using hierarchical regression analyses, we show that the new orthographic measure captures substantially more variance than does ON in behavioral measures of speeded pronunciation and lexical decision. Importantly, the advantages of the new measure over ON are greatest for longer multisyllabic words, enabling powerful investigations of similarity effects across the full adult lexicon.

METHOD

LD is defined as the number of insertions, deletions, and substitutions needed to generate one string of elements from another. For example, the LD from SMILE to SIMILES is 2, reflecting two insertions (I and S), and the LD from CHANCE to STRAND is 5, reflecting three substitutions (C → T, H → R, and C → D), an insertion (S), and a deletion (E). Our implementation initially assigned equal costs to the three operations (i.e., insertion = deletion = substitution = 1). However, because other weighting schemes could conceivably predict behavioral measures of lexical access more strongly, we also explored the effects of adding transposition as an elementary operation (e.g., treating TRIAL → TRIAL as one transposition rather than two substitutions; Damerau, 1964) or otherwise varying costs. When transposition was enabled, the resulting scores were virtually identical to the original scores ($r = .997$) and produced identical regression results. Results were similarly unaffected by systematic 20% reductions or increases in the relative cost of insertion, deletion, or substitution operations (e.g., assigning costs of insertion = 0.8, deletion = 1, substitution = 1). Across several permutations of operation costs, correlations with OLD20 were always near unity ($rs > .95$). Results are therefore reported only for the original measure, although future explorations of different weighting schemes should continue.

To generate an LD-based measure of orthographic similarity, we first calculated the LD from each word to every other word in the large set of well-described words contained in the English Lexicon Project (Balota et al., 2007). Words containing apostrophes were excluded from analysis. We then computed a quantity: OLD20, the mean LD from a word to its 20 closest orthographic neighbors. The number 20 was chosen on the basis of a cursory analysis indicating that the relationship between RTs and the number of words used to generate the LD measure was curvilinear. The increment in variance explained was smallest for very low or very high values (<5 or >50) and peaked around 10–20 words, depending on data set and task. The choice to use 20 words rather than, say, 10 was relatively arbitrary; however, choosing other values in the 5–50 range had relatively minimal effects on the present results (at most 0.02%–0.03% difference in explained variance) and produced no qualitative change in effects.

Table 1 provides an example of a word from an orthographically dense neighborhood and a word from an orthographically sparse neighborhood.[1] Note that words from orthographically dense neighborhoods have relatively low OLD20 scores and that words from orthographically sparse neighborhoods have relatively high OLD20 scores. Thus, OLD20 is coded in the opposite direction from ON (for which high values indicate greater similarity). OLD20 scores for 35,502 English words are now available at elexicon.wustl.edu.

Table 1
Twenty Closest Levenshtein Neighbors for CONDITION (Low OLD20, Orthographically Dense) and PISTACHIO (High OLD20, Orthographically Sparse)

CONDITION		PISTACHIO	
Levenshtein Neighbor	Pairwise Distance	Levenshtein Neighbor	Pairwise Distance
conditions	1	distraction	4
coalition	2	hibachi	4
cognition	2	mustache	4
conditional	2	mustached	4
conditioned	2	mustaches	4
conditioner	2	pigtail	4
conduction	2	pistil	4
contrition	2	pitch	4
conviction	2	pitched	4
recondition	2	pitcher	4
rendition	2	pitches	4
addition	3	pitching	4
audition	3	psychic	4
collation	3	psycho	4
collision	3	abstain	5
commotion	3	abstraction	5
conception	3	antacid	5
concoction	3	attach	5
concretion	3	attache	5
conditioners	3	attached	5
OLD20: 2.4		OLD20: 4.3	

Additionally, a Windows program for generating similar scores given arbitrary input lexicons and operation costs may be downloaded at artsci.wustl.edu/~tyarkoni/LD/.

To compare the predictive utility of OLD20 with that of ON, we conducted a series of hierarchical multiple regression analyses. Item-level pronunciation and lexical decision latencies (defined as the standardized mean RT for each item across subjects) were regressed on ON or OLD20 in three different data sets, including (1) 2,422 monosyllabic words (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004), (2) 9,266 monomorphemic mono- and multisyllabic words (Yap, 2007), and (3) 35,502 mono- and multimorphemic words that serve as the behavioral data set in the English Lexicon Project (Balota et al., 2007). Details regarding data collection, subject characteristics, and predictor variables for these three data sets are well described in Balota et al. (2004), Balota et al. (2007), and Yap (2007), and the data are available at elexicon.wustl.edu. Key subject characteristics are summarized in Table 2.

In each data set, a number of variables were controlled for prior to entering ON and OLD20. For the monosyllabic words, control variables included phonological onsets, feedforward and feedback consistency, word frequency, familiarity, and length (see Balota et al.,

Table 2
Participant Demographics for Monosyllabic (Balota et al., 2004) and Multisyllabic (Balota et al., 2007) Data Sets

	Data Set			
	Monosyllabic		Multisyllabic	
	M	SD	M	SD
	(n = 31)		(n = 444)	
Speeded pronunciation				
Age	22.6	5	23.5	9.3
Years of education	14.8	2	14.7	1.8
	(n = 30)		(n = 816)	
Lexical decision				
Age	20.5	2	22.9	6.9
Years of education	14.9	1.6	14.8	1.7

Table 3
Correlations Between Key Lexical Variables and Dependent Behavioral Measures

	1	2	3	4	5	6
Monosyllabic Words						
1. Length	–	−.648	.705	−.152	.396	.088
2. Orthographic N		–	−.925	.134	−.364	−.086
3. Orthographic Levenshtein distance 20			–	−.181	.386	.123
4. Frequency				–	−.275	−.586
5. Speeded pronunciation RT					–	.279
6. Lexical decision RT						–
Monomorphemic Words						
1. Length	–	−.617	.887	−.311	.590	.478
2. Orthographic N		1	−.642	.292	−.432	−.361
3. Orthographic Levenshtein distance 20			–	−.337	.637	.545
4. Frequency				–	−.539	−.689
5. Speeded pronunciation RT					–	.753
6. Lexical decision RT						–
Mono- and Multimorphemic Words						
1. Length	–	−.535	.868	−.327	.552	.557
2. Orthographic N		–	−.561	.295	−.366	−.343
3. Orthographic Levenshtein distance 20			–	−.371	.592	.612
4. Frequency				–	−.526	−.647
5. Speeded pronunciation RT					–	.794
6. Lexical decision RT						–

Note—For all correlations, $p < .001$.

2004). For monomorphemic multisyllabic words, control variables included phonological onsets, feedforward and feedback consistency, word frequency, length, and number of syllables. For the full set of mono- and multimorphemic words, control variables included word frequency, length, number of syllables, and number of morphemes.

To assess the convergent validity of OLD20, two additional analyses were conducted that were based on the interactive influences of word frequency and neighborhood frequency with ON (see Andrews, 1997). First, interactions between orthographic similarity and word frequency were modeled by entering the interaction term for OLD20 × frequency or ON × frequency into a hierarchical regression after controlling for the main effects of both variables as well as all other control variables. Second, effects of neighborhood frequency (NF) were compared for OLD20 and ON. ON NF was defined as the mean log frequency of a word's orthographic neighbors according to the hyperspace analogue to language (HAL) frequency norms (Lund & Burgess, 1996), which are derived from approximately 131 million words gathered from Usenet newsgroups. OLD20 NF was defined as the mean frequency of the 20 words closest to the target. Each NF measure was entered into a hierarchical regression after controlling for control variables, main effects of OLD20 and ON, and the other NF measure. Note that NF analyses could be performed only on words with $ON \geq 1$, reducing the number of items used in each data set, as specified below.

RESULTS

Table 3 presents the zero-order correlations for each data set between OLD20, the dependent measures, and the standard lexical variables of length, ON, and word frequency. Several points are worth noting. First, OLD20 was the single strongest predictor of speeded pronunciation latencies in the monomorphemic and full data sets, and it predicted lexical decision latencies almost as strongly as did frequency in all three data sets. Thus, if a single lexical variable is to be used to predict behavior, OLD20 arguably outperforms more traditional measures. Second, OLD20 was negatively correlated with ON and positively correlated with length in all three data sets. Importantly,

these relationships were modulated by the size of the corpus being examined. The negative relationship between OLD20 and ON was largest in the monosyllabic data set ($r = -.925$) and smallest in the full data set ($r = -.561$), suggesting that OLD20 and ON are functionally very similar for shorter monosyllabic words, but that they diverge significantly for longer words where ON has limited utility. Indeed, as shown in Figure 1, long words have very few—if any—ONs, whereas the OLD20 measure continues to be productive for long words. Third, as was expected, length was also strongly correlated with both ON and OLD20. Finally, correlations between OLD20 and word frequency were relatively modest ($rs < .38$).

To assess the contribution of OLD20 to the behavioral measures, a series of hierarchical regression analyses was conducted. Because OLD20, ON, and length were all highly intercorrelated, we focused on the unique contributions of each of the three variables. Table 4 presents the regression results for the monosyllabic, monomorphemic, and full data sets. Three separate hierarchical regression models were tested in each data set. Step 1 was identical in all three models and consisted of the control variables for each data set described earlier. Step 2 reflected the incremental contribution of two of the three remaining variables (i.e., ON and length, ON and OLD20, or length and OLD20). Finally, Step 3 reflected the unique contribution of ON, OLD20, or length after controlling for the remaining two variables.

An inspection of Step 3 coefficients across the different models indicates that the unique contribution of OLD20 to lexical decision and speeded pronunciation consistently exceeded that of ON and of length. The only exception is for monosyllabic words, where length alone made a meaningful unique contribution (1.0% of the variance in speeded pronunciation); neither OLD20 nor ON explained more than 0.1% of the variance in either be-

Mean ON As a Function of Length

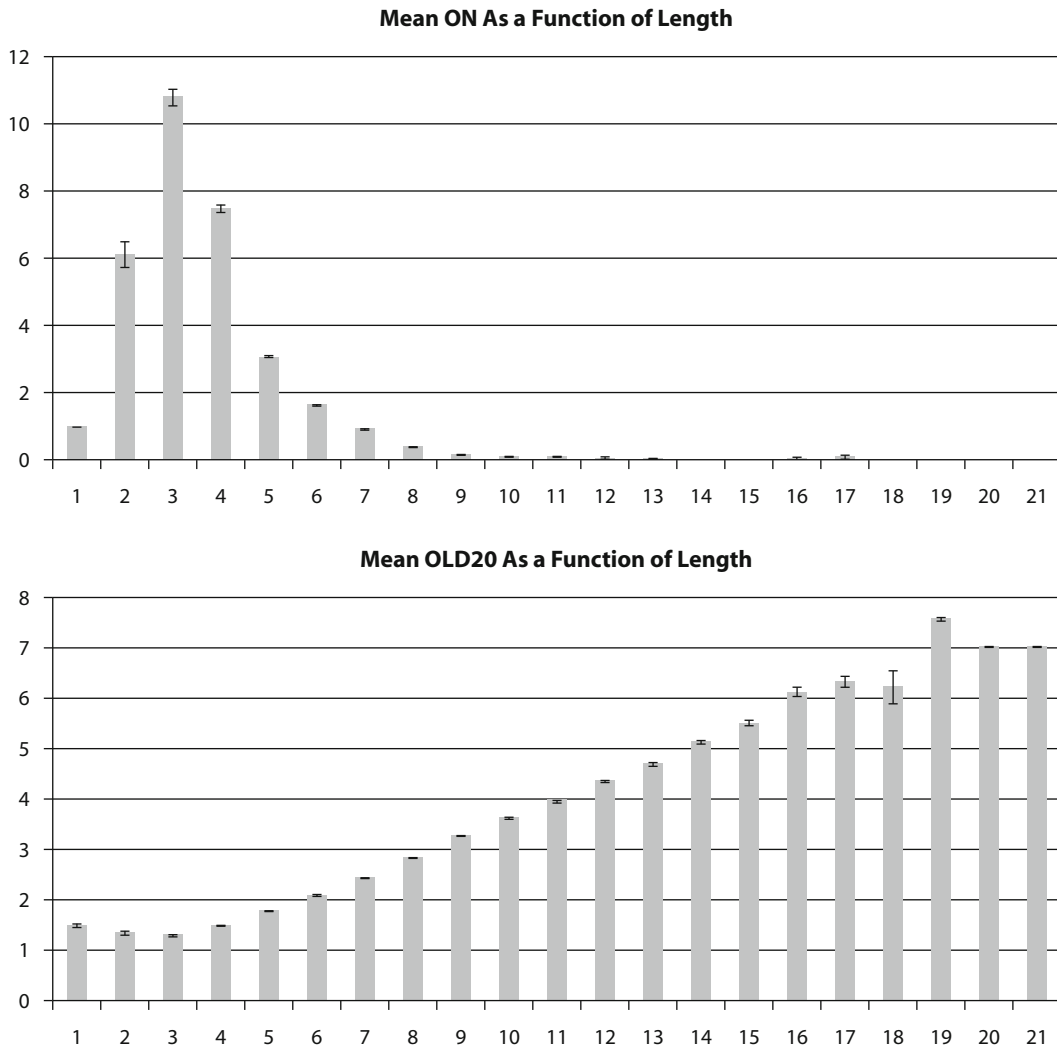**Mean OLD20 As a Function of Length**

Figure 1. Mean ON and OLD20 values as a function of word length. Error bars denote standard errors.

havioral variable. However, in the monomorphemic and full data sets, the unique contributions of OLD20 to the behavioral measures (1%–2.1%) were invariably larger than the contributions of length ($\leq$0.1%) or ON ($\leq$0.4%). Moreover, the direction of OLD20 effects was consistent across all models (orthographically distinct words always elicited slower responses), whereas the sign of ON and length coefficients varied, depending on analysis. Thus, the hierarchical regression results provide strong initial support for the notion that OLD20 is a more powerful metric of orthographic similarity than ON.

To further establish the validity of OLD20 as a measure of orthographic similarity, we now turn to the interactions between ON and word frequency and ON and the relative frequency of a word's orthographic neighbors (Andrews, 1989, 1992; Balota et al., 2004). It should be noted that efforts to reconcile such interactions with theoretical models of lexical access have been limited by the fact that effects tend to be relatively weak and often vary in direction across studies (Andrews, 1997). We reasoned that if OLD20 was

indeed a more powerful metric of orthographic similarity than ON, then stronger and more consistent frequency effects should be observed for OLD20 than for ON.

To test for interactions between orthographic similarity and word frequency, OLD20 $\times$ frequency and ON $\times$ frequency interactions were each estimated separately after controlling for the main effects of the two neighborhood measures and other control variables described previously. OLD20 and ON both interacted significantly with word frequency in all data sets (see Figure 2). For both variables, the influence of orthographic similarity became increasingly facilitatory as word frequency increased. However, OLD20 effects were facilitatory at all frequency levels, whereas ON effects varied in direction across frequency levels, with facilitation occurring for low-frequency words and inhibition occurring for high-frequency words. Importantly, the OLD20 $\times$ frequency interaction contributed more variance to lexical decision and speeded pronunciation than did the ON $\times$ frequency interaction in all three data sets.

Table 4
Standardized RT Regression Coefficients for OLD20, ON, and Length, After Onsets
and/or Lexical Variables Are Controlled for, in Three Different Data Sets

		Monosyllabic Words (n = 2,422)			Monomorphemic Words (n = 9,266)			All Words (n = 35,502)		
		β	ΔR^2	R^2	β	ΔR^2	R^2	β	ΔR^2	R^2
Speeded Pronunciation										
Step 1: Control variables		–	.446	.446	–	.548	.548	–	.445	.445
Step 2:										
Model 1	ON	−.102***	.048	.494	.01	.023	.571	−.016**	.022	.467
	length	.157***			.272***			.287***		
Model 2	ON	−.044	.04	.486	.053***	.044	.592	.006	.031	.476
	OLD20	.175***			.384***			.276***		
Model 3	length	.143***	.049	.495	.035*	.043	.591	.106***	.032	.477
	OLD20	.116***			.331***			.218***		
Step 3:										
Model 1	OLD20	.089*	.001	.495	.354***	.021	.592	.222***	.01	.477
Model 2	length	.142***	.01	.495	.049**	0	.592	.108***	.001	.477
Model 3	ON	−.03	0	.495	.057***	.001	.592	.010*	0	.477
Lexical Decision										
Step 1: Control variables		–	.418	.418	–	.561	.561	–	.55	.55
Step 2:										
Model 1	ON	.017	0	.418	.015	.005	.566	.044***	.014	.564
	length	−.004			.143***			.252***		
Model 2	ON	.077(*)	0	.418	.069***	.025	.586	.079***	.034	.584
	OLD20	.062			.302***			.307***		
Model 3	length	−.018	0	.418	−.094***	.024	.585	−.010***	.03	.58
	OLD20	.003			.327***			.277***		
Step 3:										
Model 1	OLD20	.072	0	.418	.351***	.021	.587	.304***	.02	.584
Model 2	length	−.016	0	.418	−.079***	.001	.587	.005	0	.584
Model 3	ON	.076(*)	0	.418	.062***	.002	.587	.080***	.004	.584

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

To test for NF effects, we conducted regression analyses comparing the unique contributions of OLD20- and ON-based NF measures to the behavioral measures after controlling for all variables in the earlier regression models (i.e., Steps 1–3 in Table 4; see Table 5). In the monosyllabic data set, neither NF variable made a significant contribution to the behavioral measures. In the monomorphemic data set, OLD20 NF contributed significantly to both measures, whereas ON NF did not contribute to either. Finally, in the full data set, both OLD20 NF and ON NF made significant contributions to both measures, but the contribution of OLD20 NF was substantially larger than that of ON NF. Thus, both the frequency interaction and NF analyses supported predictions that OLD20 should produce stronger and more consistent effects than ON.

DISCUSSION

The present study explored the effects of a new orthographic similarity metric—OLD20—on two standard measures of word recognition. This metric was based on LD, a measure with considerable utility in a broad spectrum of pattern recognition problems in computer science. OLD20 was shown to predict pronunciation and lexical decision performance more strongly than the widely used measure of ON, across three large databases. Moreover, OLD20 outperformed ON in secondary analyses testing interactions with frequency as well as NF effects. Im-

portantly, the advantage of OLD20 over ON increased as word length increased, making it particularly well suited for analyses of the full adult lexicon. We will now briefly discuss these results and consider the methodological and theoretical implications of OLD20.

Superiority of OLD20 Over ON

Clearly, OLD20 is superior to ON in a number of ways. It applies to a larger set of words, accounts for more unique variance, and is reliably facilitatory in its effect. These advantages appear to derive from two simple principles. First, the utility of ON is greatly reduced for longer words, which have few—if any—orthographic neighbors. Hence, similarity cannot be measured for these items with ON (see Figure 1). Second, ON is based on a dichotomous definition of neighbor relations; that is, a word is or is not a neighbor of another word. As was noted earlier, Davis (2006) recently reviewed considerable evidence indicating similarity effects of word pairs that are not traditional orthographic neighbors. OLD20 is able to capture the similarity between such pairs in a continuous manner, resulting in a more fine-grained index of orthographic similarity when applied to a full lexicon. Importantly, the benefits of OLD20 appear to carry over to derivative measures, such as OLD20-based neighborhood frequency. Thus, from a practical standpoint, there appear to be many advantages to using OLD20 as a complement or substitute for ON, and no apparent disadvantages.
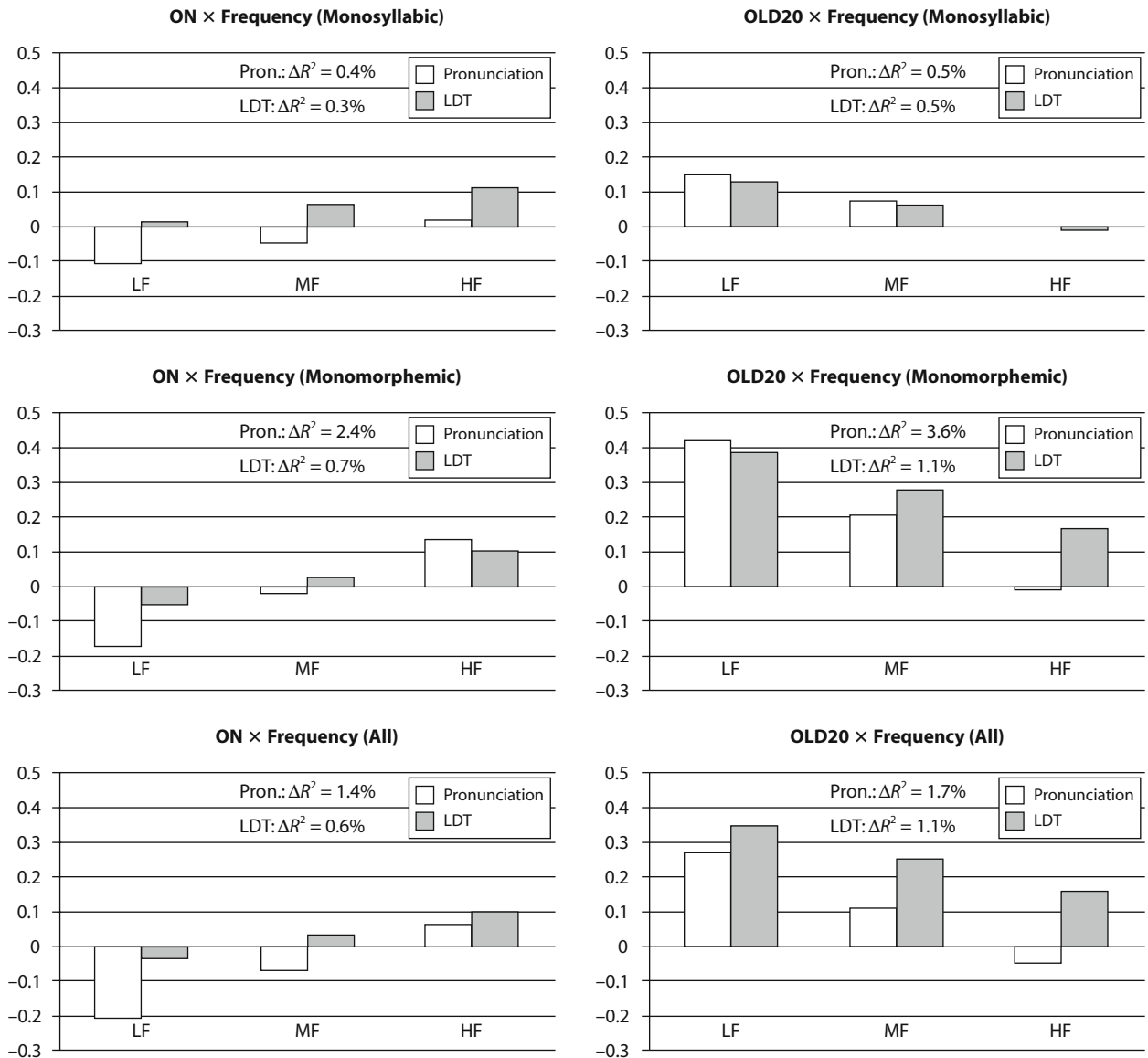
Figure 2. The ON × word frequency and OLD20 × word frequency interactions as a function of task and data set. All interactions are significant ($p < .001$). Frequency was a continuous variable in all analyses, and words are binned here solely for illustrative purposes. The y-axis represents standardized standard beta weights. Note that ON and OLD20 are coded in opposite directions; that is, large values of ON reflect orthographically similar words, whereas large values of OLD20 reflect orthographically distinct words. LF, low frequency; MF, medium frequency; HF, high frequency; LDT, lexical decision task.

One might express concern that the advantages of OLD20 over ON—although real—are relatively modest, because the observed effect sizes were not large: The unique contribution of OLD20 to the behavioral measures was typically 1%–2% of the total variance. However, several points are important to note. First, although R^2 values of .02 may seem small, effects of this magnitude are typical of many areas of psychology (Meyer et al., 2001). Second, the estimates reported here are conservative; they represent the unique contribution of OLD20 after 40%–60% of the variance in behavior is already accounted for. Finally, given that ON and length together explained only 1%–2% of the variance in behavior in

many analyses (without OLD20 in the model), concerns about effect size are clearly not specific to OLD20, but apply more broadly. Indeed, the present results underscore the fact that the large effect sizes sometimes seen in factorial studies may reflect carefully controlled item selection and may not be representative of the lexicon as a whole.

Reconciling Facilitatory and Inhibitory Effects of Orthographic Similarity

There is ongoing debate as to whether orthographic similarity effects should be inhibitory—because of within-level lexical competition (see, e.g., McClelland &

Table 5
Standardized RT Regression Coefficients for Neighborhood Frequency Measures After
Controlling for Main Effects of Similarity Measures and Other Variables

			Monosyllabic Words (n = 2,274)			Monomorphemic Words (n = 5,275)			All Words (n = 14,407)		
			β	ΔR^2	R^2	β	ΔR^2	R^2	β	ΔR^2	R^2
						Speeded Pronunciation					
Steps 1 to 3:			–	.48	.48	–	.504	.504	–	.407	.407
Step 4:	Model 1	ON NF	−.016	0	.48	.048***	.002	.506	.024**	0	.407
Step 5:		OLD20 NF	−.022	0	.48	.174***	.007	.513	.176***	.01	.417
Step 4:	Model 2	OLD20 NF	−.027	0	.48	.181***	.009	.513	.165***	.01	.417
Step 5:		ON NF	−.01	0	.48	.012	0	.513	−.021**	0	.417
						Lexical Decision					
Steps 1 to 3:			–	.407	.407	–	.505	.505	–	.516	.516
Step 4:	Model 1	ON NF	.011	0	.407	.038**	.001	.506	.009	0	.516
Step 5:		OLD20 NF	−.03	0	.407	.144***	.004	.51	.114***	.004	.52
Step 4:	Model 2	OLD20 NF	−.02	0	.407	.149***	.005	.51	.103***	.003	.519
Step 5:		ON NF	.019	0	.407	.009	0	.51	−.020**	.001	.52

** $p < .01$.   *** $p < .001$.

Rumelhart, 1981)—or facilitatory—because of the summing of lexical activation across orthographically similar words (for a review, see Andrews, 1997). Previous studies have produced mixed findings, with some studies reporting facilitation for orthographically similar words, and others reporting inhibition. Some researchers have attributed such discrepancies to task-specific differences in response criteria; for example, Grainger and Jacobs (1996) suggested that facilitatory ON effects in lexical decision tasks reflect greater reliance on overall lexical activation (which should be higher for words with large neighborhoods) than on word-specific activation. However, task-specific accounts can be ruled out in the present study, which identified simultaneous inhibitory and facilitatory effects for both lexical decision and speeded pronunciation tasks. In both cases, OLD20 exerted a facilitatory effect (i.e., orthographically similar words produced faster responses), whereas ON and OLD20 NF exerted an inhibitory effect.

We propose that the presence of both facilitatory and inhibitory effects of orthographic similarity can be parsimoniously explained by supposing that the two kinds of effects arise at different stages of processing. The core suggestion is that OLD20 predominantly reflects early, more general ("global") similarity, whereas ON predominantly reflects late, more specific ("local") similarity. In connectionist terms, the global facilitatory process is postulated to reflect the initial "pull" of an attractor basin containing orthographically similar words, whereas the local inhibitory process reflects mutual competition between highly similar words within the basin. A similar idea was expressed by Andrews (1997), who noted that "there may be a functionally equivalent trade-off between the stronger connections developed for more frequently occurring patterns and the overlap of the attractors for similar words" (p. 457). For example, a low-OLD20 word, like STAB, is likely to benefit from the fact that it shares spelling patterns with many other words (e.g., STATION, TABLE, STACK), because frequently presented patterns should be processed

more efficiently, producing a well-characterized attractor basin. Once within the basin, however, there are more highly similar competitors to contend with (e.g., STAR or SLAB), resulting in slower identification. Moreover, to the extent that such competitors are high in frequency, the processing of STAB will be further slowed due to increased lateral inhibition.

A global/local distinction also parsimoniously explains the observed OLD20-frequency and ON-frequency interactions. Although OLD20 effects are generally facilitatory, they are stronger for low-frequency words, because these weakly represented words benefit more from initial attraction into the appropriate orthographic basin. For the ON × frequency interaction, one observes facilitation for low-frequency words and some inhibition for high-frequency words. As was discussed before, the inhibition for high-frequency words may reflect late, local competition when the system is disambiguating between highly similar candidates. These inhibitory effects may be less apparent for low-frequency words, because such words take so long to get from the global to local neighborhood that the early facilitatory effects overshadow the later inhibitory effects, due to competition. In other words, all words experience competition at the local level, but these effects are far more salient for high-frequency words, which enter the appropriate orthographic basin relatively quickly, due to frequency of exposure.

Relationship Between Length and OLD20

An intriguing finding in the present study was that word length accounted for little or no variance in behavioral latencies after controlling for OLD20. This finding suggests that the putative influence of length on lexical access may actually derive from a more fundamental effect of orthographic similarity. That is, longer words may produce slower response latencies because they tend to be more orthographically distinctive than medium-length words. In this connection, note that the relation between length and response latencies in lexical decision and pronuncia-

Speeded Pronunciation (Monomorphemic)

Lexical Decision (Monomorphemic)

Speeded Pronunciation (All)
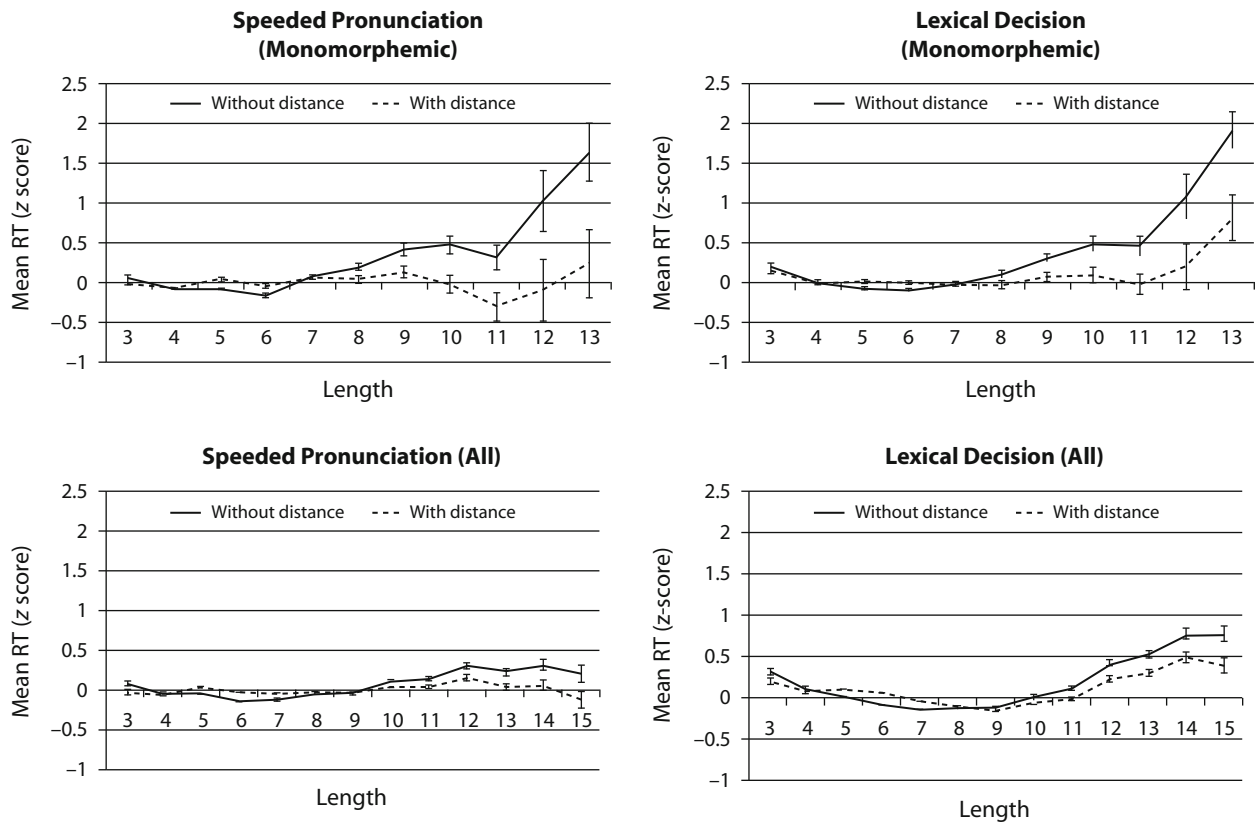

Lexical Decision (All)

Figure 3. Pronunciation and lexical decision latencies as a function of word length before and after controlling for OLD20 in the monomorphemic (top) and full (bottom) data sets. Error bars denote standard errors. RT, response time.

tion performance appears to be quadratic, with the fastest response latencies around 5 to 7 letters (New, Ferrand, Pallier, & Brysbaert, 2006; Yap, 2007). OLD20 scores should also be disproportionately low in the 5- to 7-letter range, because there are simply more 5- to 7-letter words in English than there are shorter or longer words, thereby reducing the orthographic distance from a target word to others in the same length range. Thus, OLD20 might be expected to explain not only much of the linear length effect, but also at least part of the quadratic length effect. A post hoc analysis confirmed this prediction. In both the monomorphemic and full data sets, quadratic length accounted for substantially more unique variance in speeded pronunciation and lexical decision performance when controlling for all variables except OLD20 than when OLD20 was also controlled for (unique variance without OLD20 = 1.1%–1.6%; unique variance with OLD20 = 0.0%–0.9%; mean reduction in variance explained = 78%; see Figure 3).

Conclusions

The present study introduced a new measure of orthographic similarity to the psychological literature. Regression analyses demonstrated that this measure explains substantially more variance in behavioral measures of lexical access than does the de facto standard of ON, interacts more strongly with frequency, and produces more

consistent NF effects. The results provide a novel theoretical perspective on orthographic similarity and furnish researchers with a large set of norms for use in future investigations.

REFERENCES

ANDREWS, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 802-814.

ANDREWS, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 234-254.

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

BALOTA, D. A., CORTESE, M. J., SERGENT-MARSHALL, S. D., SPIELER, D. H., & YAP, M. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, **133**, 283-316.

BALOTA, D. A., YAP, M. J., CORTESE, M. J., HUTCHISON, K. A., KESSLER, B., LOFTIS, B., ET AL. (2007). The English Lexicon Project. *Behavior Research Methods*, **39**, 445-459.

COLTHEART, M., DAVELAAR, E., JONASSON, J. T., & BESNER, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535-555). Hillsdale, NJ: Erlbaum.

DAMERAU, F. J. (1964). A technique for computer detection and correction of spelling errors. *Communications of the ACM*, **7**, 171-176.

DAVIS, C. J. (2006). Orthographic input coding: A review of behavioural data and current models. In S. Andrews (Ed.), *From inkmarks to ideas: Challenges and controversies about word recognition and reading* (pp. 180-206). New York: Academic Press.

DAVIS, C. J., & BOWERS, J. S. (2006). Contrasting five different theories of letter position coding: Evidence from orthographic similarity effects. *Journal of Experimental Psychology: Human Perception & Performance*, **32**, 535-557.

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

LEVENSHTEIN, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady*, **10**, 707.

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

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

MEYER, G. J., FINN, S. E., EYDE, L. D., KAY, G. G., MORELAND, K. L., DIES, R. R., ET AL. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, **56**, 128-165.

NEW, B., FERRAND, L., PALLIER, C., & BRYSBAERT, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review*, **13**, 45-52.

YAP, M. J. (2007). *Visual word recognition: Explorations of megastudies, multisyllabic words, and individual differences.* Unpublished doctoral dissertation, Washington University.

NOTE

1. As shown in Table 1, some of the LD close neighbors have morphological similarity to the target item (e.g., include CON or TION). This is expected, because morphological units are common spelling constituents. Note, however, that partialing out the number of morphemes in the hierarchical regression analyses did not modulate the results. Moreover, as described in the Results sections, analyses from monomorphemic data sets (Data Set 2) resulted in results very similar to those in the full data set (Data Set 3).