Phonological Typicality Does Not Influence Fixation Durations in Normal Reading

Adrian Staub, Margaret Grant, and Charles Clifton, Jr.
University of Massachusetts at Amherst

Keith Rayner
University of California, San Diego

Using a word-by-word self-paced reading paradigm, T. A. Farmer, M. H. Christiansen, and P. Monaghan (2006) reported faster reading times for words that are phonologically typical for their syntactic category (i.e., noun or verb) than for words that are phonologically atypical. This result has been taken to suggest that language users are sensitive to subtle relationships between sound and syntactic function and that they make rapid use of this information in comprehension. The present article reports attempts to replicate this result using both eyetracking during normal reading (Experiment 1) and word-by-word self-paced reading (Experiment 2). No hint of a phonological typicality effect emerged on any reading-time measure in Experiment 1, nor did Experiment 2 replicate Farmer et al.'s finding from self-paced reading. Indeed, the differences between condition means were not consistently in the predicted direction, as phonologically atypical verbs were read more quickly than phonologically typical verbs, on most measures. Implications for research on visual word recognition are discussed.

Keywords: reading, word recognition, eye movements, language processing

The time it takes to recognize a printed word is reliably influenced by a range of factors, both in single-word recognition paradigms, such as lexical decision and naming (see Balota, Yap, & Cortese, 2007; Rastle, 2007, for recent reviews), and in normal reading, in which the duration of readers' eye fixations is regarded as an index of word-recognition time (Rayner, 1998). These factors include, among other things, the length of the word (e.g., Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Just & Carpenter, 1980), its frequency (e.g., Balota & Chumbley, 1984; Rayner & Duffy, 1986), the familiarity of the letter strings within the word (White, 2008), and the number of orthographic "neighbors" it has (i.e., the number of other words with similar spellings; e.g., Pollatsek, Perea, & Binder, 1999). There are also phonologically mediated effects on visual word recognition, such as an effect of the relationship between a word's spelling and its sound, as words with irregular spellings take longer to recognize (although this occurs primarily when the word is low in frequency; e.g., Jared, 2002), and an effect of a word's phonological neighborhood size (Yates, Locker, & Simpson, 2004).

Recently, Farmer, Christiansen, and Monaghan (2006) reported a novel effect of a word's sound on visual word recognition. Farmer et al. described four experiments apparently demonstrating that a word whose sound is relatively typical for its syntactic category (i.e., noun or verb) is processed more quickly than a word whose sound is relatively atypical for its category. They measured typicality by quantifying the phonological similarity[1] of each of a large collection of nouns and verbs to the other nouns and to other verbs, identifying nouns that are more similar to nouns in general than to verbs (e.g., *marble*), nouns that are more similar to verbs than to nouns (e.g., *insect*), and verbs that are similar to verbs or to nouns (e.g., *amuse* and *ignore*, respectively).

We concentrate on Farmer et al.'s Experiments 2 and 3, which measured the reading speed of the critical words in a word-by-word self-paced reading paradigm. In these experiments, Farmer et al. presented the words in sentence frames that were designed to generate a strong expectation of a word in the appropriate syntactic category: The nouns were presented following a transitively biased verb and the word *the*, as in 1a and 1b, and the verbs were presented in an infinitival clause after the word *to*, as in 2a and 2b; the critical words are italicized.

1. (a) The curious young boy saved the *marble* that he found on the playground.
   (b) The curious young boy saved the *insect* that he found in his backyard.

2. (a) The young girl had tried to *amuse* herself while waiting for her mother by working on a crossword puzzle.
   (b) The young girl had tried to *ignore* the boy that kept on pulling on her hair during recess.

[1] Phonological (dis)similarity between two words was measured by the Euclidean distance between the words in phonological feature space, after optimally lining up the segments of the words' onsets, nuclei, and codas. The mean distances between each word and all nouns, and between each word and all verbs, were then computed as a measure of the phonological typicality of the word as a noun and as a verb. For a detailed example, see Farmer et al. (2006, pp. 12203–12204).

Farmer et al. (2006) found that length-adjusted reading times for phonologically typical nouns and verbs were approximately 40 to 60 ms faster than for atypical nouns and verbs.

Arguably, these results have implications for a long-running debate between two contrasting positions about how people comprehend language. At one extreme, language users are claimed to have distinct and specialized cognitive systems for processing the different components of language—phonology, syntax, semantics, etc. At the other extreme, language users are claimed to have a single cognitive system that optimally integrates all available information about the message contained in an utterance. Each position has much to recommend it. The first, modular, position is buttressed by the success of linguistic analyses that posit distinct collections of principles for each component of language, and it receives support from analyses of cognitive function in terms of distinct visual, sensorimotor, etc. systems (see Fodor, 1983). The second, interactive, position directly addresses the efficient way in which language users integrate disparate sources of information in understanding what we read or hear. It receives support from the ability of interactive, distributed, connectionist systems to account for some intricacies of language, and it honors the fact that the distinct brain systems are densely interconnected, interacting intimately with each other (Elman et al., 1996; Tanenhaus & Trueswell, 1995).

In a recent commentary, Tanenhaus and Hare (2007) enlisted the Farmer et al. (2006) results in support of this second position. They suggested that the phonological typicality results undermine the assumption that language comprehension respects a clear "form–content distinction" (p. 93), and that these results provide a "striking example of the importance of distributional patterns in language processing" (p. 94). One specific interpretation of the Farmer et al. results might hold that at the point of reaching the critical word in frames like 1 and 2, readers are expecting a word in a particular syntactic category (cf. Lau, Stroud, Plesch, & Phillips, 2006; Staub & Clifton, 2006), and that this is, in fact, a phonological expectation, in whole or in part, so that the reader activates in advance a specific region of phonological space. An alternate interpretation would not emphasize the role of expectations but, instead, would propose that word recognition is delayed when there is conflict between retrieved phonological information and syntactic category information.

It is not unexpected that phonological properties of words can influence aspects of lexical processing. Instances of sound symbolism are well attested (e.g., onomatopoeia), and there are many demonstrations that the phonological properties of words can influence their offline identification as function versus content words or as nouns versus verbs and their learning in childhood or in artificial languages (see Farmer et al., 2006, for references; see also Sereno, 1994; Sereno & Jongman, 2000, for highly relevant research).[2] What is unexpected is the apparent speed of the effect, showing up in the time it takes to read a word, and the size of the effect, approximately 50 ms. Even assuming somewhat slowed self-paced reading times (as discussed below, Farmer et al., 2006, did not report raw reading times), this result implies that highly abstract, and not introspectively apparent, differences in phonological typicality may account for more than 10% of the time readers spend on a word.

In our opinion, there is a need to examine the Farmer et al. (2006) findings more closely, for several reasons. First, self-paced reading is often substantially slower than normal reading (Rayner & Pollatsek, 1989), and slower reading may accentuate phonological influences, perhaps by increasing the incidence of subvocalization. In fact, Tanenhaus and Hare (2007) explicitly raised the question of how the Farmer et al. effect would be manifested in readers' eye movements, speculating that it would appear in first fixation duration, the very earliest measure of lexical processing. Second, the reported results are based on a rather small amount of data, considering the typical variability in self-paced reading measures. In each of the two relevant experiments, 22 subjects each provided five observations in each experimental condition (typical vs. atypical noun or verb), so that reading-time means were based on a maximum of 110 observations. Finally, no information was provided about the between-item variability in reading times (i.e., no statistical analyses were reported in which item was the random variable), so it is possible that the reported results were being driven by a small number of anomalous items.

The present article reports an attempt to verify the results reported by Farmer et al. (2006). In Experiment 1, we made several critical modifications in procedure. First, we measured eye movements during normal reading, to see if the critical finding could be replicated in a more naturalistic paradigm and, if so, to see which of several eye movement measures are sensitive to phonological typicality (see, e.g., Staub & Rayner, 2007, for discussion of various eye movement measures and their interpretation). Second, we used a within-subject design, so that the same subjects saw the nouns and the verbs, rather than varying word class between subjects (and between experiments) as Farmer et al. did. Third, we dramatically increased the number of observations contributing to each condition mean in two ways: increasing the number of subjects tested and allowing each subject to see all 10 of the words in each category examined by Farmer et al. (noun-like noun, etc.), rather than just half of these words. Fourth, rather than using length-adjusted reading times as the dependent measure, we used statistical analyses that independently assessed the contributions of length, frequency, and phonological typicality to reading times.[3]

Experiment 1 failed to find any hint of a phonological typicality effect. This raised the obvious question of whether the critical difference between Experiment 1 and the Farmer et al. (2006) study was the use of fixation-duration measures in normal reading, rather than button-press latency in self-paced reading; if so, this would point to a difference between methodologies in need of further investigation. To address this issue, in Experiment 2, we once again used self-paced reading with the same materials, while preserving the increased power and improved statistical analyses from Experiment 1.

---

[2] Whereas most of the cited research reports only offline effects, Sereno and Jongman (2000) did report an online effect of phonological typicality on auditory lexical decision time. However, their effect was modulated by word frequency. Having a back, as opposed to a front, vowel was more typical for high-frequency nouns than for high-frequency verbs, but no difference was apparent for low-frequency items. Lexical decision times mirrored these typicality effects.

[3] Even though the phonologically typical and atypical items did not differ significantly in mean length and frequency, the existing differences can bias statistical tests (J. L. Myers, personal communication, January 14, 2008).

Experiment 1

Method

Subjects. Thirty-six native speakers of English, who were members of the University of Massachusetts community, were given course credit or were paid $7 to participate in the experiment. Subjects were recruited using an online system maintained by the University of Massachusetts Psychology Department. All subjects had normal or corrected-to-normal vision and were naïve to the purpose of the experiment.

Materials. The experimental materials consisted of the items included in Experiments 2 and 3 of Farmer et al. (2006).[4] These items included 10 words in each condition: noun-like noun, verb-like noun, noun-like verb, and verb-like verb. To have each subject read all 40 of the critical words, we devised novel preceding contexts for each item, in addition to those included in Farmer et al.'s experiments. These novel contexts were identical to Farmer et al.'s materials except for the specific lexical items used in the subject noun phrase. For example, the subject noun phrase *the curious young boy* in 1a and 1b was replaced with *the mischievous toddler*, and the subject noun phrase *the young girl* in 2a and 2b was replaced with *the small child*. The matrix verbs and determiners that immediately preceded the critical nouns were preserved from Farmer et al.'s materials, as were the matrix verbs and nonfinite *to* that preceded the critical verbs. The material following the critical word was also left unchanged from Farmer et al., as this material already varied between target words in their experiment.[5] The experimental items were counterbalanced so that each subject read half of the critical words with the preceding context from Farmer et al. and half with the novel preceding context. For instance, if a subject read the word *marble* presented in 1a, then he or she would read *insect* presented with the novel preceding context. The 40 experimental sentences were intermixed with 148 sentences from unrelated experiments.

The critical words used in the experiment ranged from four to eight letters in length. Table 1 shows mean length for each condition, as well as mean frequency for each condition, based on the Hyperspace Analogue to Language (HAL) corpus (Burgess & Livesay, 1998), obtained from the English Lexicon Project (Balota et al., 2002), as well as the CELEX frequencies (Baayen, Piepenbrock, & Gulikers, 1995) reported by Farmer et al. (2006). HAL frequency, rather than CELEX frequency, was entered into the statistical model, as CELEX frequency is primarily based on older texts written in British English, whereas both our study and Farmer et al.'s study were conducted on a U.S. population.

Procedure. Subjects were tested individually, and eye movements were recorded using an EyeLink 1000 (SR Research, Toronto, Ontario, Canada) eyetracker, interfaced with a PC computer. The sampling rate for recordings was 1,000 Hz. Stimuli were displayed on an Iiyama CRT monitor. Subjects were seated 55 cm from the computer screen. At this distance, 3.69 characters subtended 1° of visual arc. The angular resolution of the eyetracker is 10–30 min of arc. Viewing was binocular, but only the right eye was recorded. All sentences in this experiment were displayed on a single line.

Before the experiment began, subjects were instructed to read in their normal manner. Each subject read eight practice items before the experimental items were shown. Comprehension was checked

on 10% of all the critical trials by presenting the subject with yes/no questions. Subjects averaged 86% correct on these questions. Over all materials, including the 148 fillers, subjects answered comprehension questions on approximately 30% of trials; we do not report overall accuracy because some of these questions were designed to determine subjects' interpretation of ambiguous or semantically odd sentences associated with other experiments and, therefore, did not have a correct answer. The entire experiment lasted approximately 45 min.

Results

Three reading-time measures were computed for the critical word in each sentence (e.g., *marble* or *insect*). *First fixation duration* is the duration of the reader's first fixation on the critical word. *Gaze duration* is the sum of all fixations on the critical word before leaving the word for the first time, either to the left or to the right. These measures are both associated with lexical processing difficulty (e.g., Reichle, Rayner, & Pollatsek, 2003). *Go-past time* is the sum of all fixations from the first fixation on the critical word until the reader leaves the critical word to the right, including any time spent to the left of the critical word as a result of regressive eye-movements and any time spent rereading the critical word before moving on to the rest of the sentence.

Before analyses were performed, approximately 2.2% of all trials were excluded because of track losses. Fixations less than 80 ms in duration, and within one character of the previous or subsequent fixation, were incorporated into this neighboring fixation. Remaining fixations of less than 80 ms, and also fixations longer than 800 ms, were deleted (less than 1% of all fixations). Table 2 shows the means on each measure for the critical words in each condition, broken down by sentence frame (i.e., original Farmer et al., 2006, frame or new frame). On all measures, nouns were read somewhat faster than verbs, and noun-like target words were read somewhat faster than verb-like target words. Figure 1 displays the pattern of means on the gaze-duration measure, collapsed over the sentence frame factor. Analyses on the three eye-movement measures were performed using a linear mixed-effects model (Baayen, 2008; Baayen, Davidson, & Bates, 2008). This analysis allowed us to incorporate the categorical predictors of interest along with continuous predictors such as frequency and length. These analyses were carried out using R, an open-source programming language and environment for statistical computing (R Development Core Team, 2007) and, in particular, the lme4 package for linear mixed-effects models (Bates, 2005). Subjects and items were included as crossed random effects. The fixed effects included in the initial model were part of speech (noun or verb), phonological classification (noun-like or verb-like), sentence frame (original Farmer et al., 2006, frame or new frame), the interactions between these three factors, length in characters, and a log transformation of HAL frequency. However, there were no effects of the sentence

---

[4] The items used by Farmer et al. (2006) are available from http://www.pnas.org/cgi/content/full/0602173103/DC1

[5] Because of an error in the construction of materials, there were two items in each experimental list in which the material following the critical word was identical. Obviously, this could only have affected reading time after the critical word itself. This error was corrected prior to Experiment 2.

Table 1
Mean Length and Frequency Information for the Target Words in Each Condition in Experiments 1 and 2

Word type	Length		HAL frequency		Log HAL frequency		CELEX frequency	
	M	SD	M	SD	M	SD	M	SD
Noun-like nouns	6.3	1.4	19,122	21,972	9.21	1.24	546	482
Verb-like nouns	5.9	1.1	25,214	41,494	9.27	1.04	642	1,087
Noun-like verbs	5.5	0.8	16,577	16,904	9.05	1.52	492	435
Verb-like verbs	5.9	1.0	28,539	44,863	9.16	1.67	494	437

Note. HAL = Hyperspace Analogue to Language (Burgess & Livesay, 1998); CELEX (Baayen, Piepenbrock, & Gulikers, 1995).

frame factor or its interactions with other factors that approached significance (as can be seen in Table 2, the pattern of means is identical for the two sets of frames), so we eliminated these parameters from the model. Explicit model comparison based on maximum likelihood estimation also showed no reduction in fit in the resulting smaller model. The parameter estimates and *p* values for this model are shown in Table 3. We report *p*s estimated using posterior distributions for model parameters obtained by Markov chain Monte Carlo sampling (Baayen, 2008; Baayen et al., 2008). In all cases, *p*s based on the *t* statistic were similar to those reported (with no differences in the pattern of significant results).

For all three reading-time measures on the target word, there were highly significant effects of log frequency, as higher frequency words were fixated for shorter durations. However, there were no significant main effects of part of speech or phonological classification, and there was no hint of a significant interaction between these two factors, on any of the three measures.[6] We did not find significant effects of length on any of the three measures, which can be attributed to the small range of word lengths. As mentioned above, the critical words varied in length from four to eight letters, and in fact, 31 of the 40 words were between five and seven letters long. (We also tested a more complex model that allowed each subject a distinct length parameter, which is analogous to the procedure used by Farmer et al., 2006, of computing a separate regression equation for each subject relating word length to reading time. This modification resulted in no improvement in model fit.)

To assess whether the experimental manipulations might have had an effect while the target word was still to the right of fixation (i.e., a so-called parafoveal-on-foveal effect), or whether these manipulations might have affected spillover processing, we also examined the duration of the last fixation prior to fixating the target word and the duration of the first fixation on the region following the target word. No effects of the experimental manipulations approached significance on these measures; the means are shown in Table 2.

Discussion

The findings from this experiment are easily summarized. Although word frequency affected fixation durations on the critical word, there was no hint that phonological typicality had an effect on any measure. Noun-like words were read somewhat faster than verb-like words, and nouns were read somewhat faster than verbs, but neither of these effects approached significance. More important for present purposes, the critical interaction was completely absent.

There are three possible interpretations of the contrast between the present null finding and the result reported by Farmer et al.

(2006). (In the General Discussion, we entertain a fourth possibility.) First, the present result may simply be a Type II error. We regard this as unlikely, given the considerable statistical power of the experiment and the fact that the numerical pattern of means did not display the predicted interaction. Second, the Farmer et al. result may be a Type I error. This is somewhat more likely, given the small number of observations involved, and given that it is not known whether Farmer et al.'s results were significant across items. But most interesting, it is also possible that what is responsible for the different patterns of results is a genuine difference between self-paced reading and eyetracking. If so, this would suggest that factors that have no influence on fixation durations in normal reading may indeed have an effect when reading is slowed down and when highly skilled, relatively automatic eye-movement behavior is replaced by conscious, overt responses. We designed Experiment 2 to evaluate these possibilities.

Experiment 2

In Experiment 2, the materials used in Experiment 1 were presented to subjects in a self-paced, word-by-word reading paradigm. As in Experiment 1, we employed a within-subjects design, and as in Experiment 1, each subject read all 10 words in each category. Thus, we preserved two differences between Experiment 1 and the Farmer et al. (2006) study that should have provided a greater chance of detecting a real phonological typicality effect, if one exists. To make Experiment 2 as close as possible in all potentially relevant respects to the original Farmer et al. study, we also used the filler items and comprehension questions that were used in the original study.[7]

Method

Subjects. Twenty-four native speakers of English, who were members of the University of Massachusetts community, were given course credit or were paid $7 to participate in the experiment. Subjects were recruited using an online system maintained by the University of Massachusetts Psychology Department. All subjects had normal or

---

[6] As the pattern of means would suggest, conventional ANOVAs treating subjects (F_1) and items (F_2) as random factors, which did not include length or frequency as predictors, also failed to show a significant interaction between part of speech and phonological classification on any of the reading-time measures reported.

[7] Thanks to Thomas Farmer for providing these materials.

Table 2
Mean Reading Time in Experiment 1 (in ms) on Each Eye-Movement Measure, by Experimental Condition and by Original (Farmer et al., 2006) Versus New Sentence Frame

Word type	Prior fixation		First fixation		Gaze		Go-past		Spillover fixation	
	M	SD	M	SD	M	SD	M	SD	M	SD
Farmer et al. frames										
Noun-like nouns	235	5.9	243	6.0	264	7.7	308	13.5	231	5.6
Verb-like nouns	239	6.0	248	5.4	273	7.8	322	13.6	230	6.1
Noun-like verbs	228	5.3	250	6.0	280	9.8	319	13.3	233	5.3
Verb-like verbs	220	5.9	259	5.3	296	8.5	323	12.6	248	5.8
New frames										
Noun-like nouns	243	6.7	239	5.7	264	7.4	318	14.0	233	5.5
Verb-like nouns	237	5.5	250	6.9	275	9.0	341	15.4	230	5.2
Noun-like verbs	224	6.5	242	5.5	268	8.3	312	15.7	234	5.0
Verb-like verbs	229	5.4	251	6.4	276	8.0	306	11.6	242	5.1

corrected-to-normal vision and were naïve to the purpose of the experiment. None had participated in Experiment 1.

Materials. The critical materials for the present experiment were identical to those in Experiment 1 (but see footnote 5). For each subject, the 40 experimental items were randomly intermixed with 42 filler items taken from Farmer et al.'s (2006) Experiments 2 and 3 and were presented after six practice trials. In addition, comprehension questions for each of the experimental items and filler items used in the present experiment were taken from Farmer et al.'s study, modified only when required by one of the new sentence frames used in this experiment.

Procedure. Stimuli were presented using the Linger software package, running on a Windows operating system. All stimuli were presented in Courier font. Subjects read written instructions on the computer screen. These instructions emphasized that subjects should read at their natural pace, making sure that they understood what they read. Subjects were then given an opportunity to ask the experimenter any questions they had about the self-paced reading methodology. Six

practice items were presented to introduce the word-by-word moving window display before subjects began the experimental trials. Each practice and experimental trial consisted of a sentence on a single line. Before any words were displayed, dashes indicated the position of each word in the sentence. These dashes remained in view throughout the trial, except for the word currently being read. Subjects pressed the space bar on a keyboard to see the first word of the sentence and to advance through the sentence word by word. After each experimental sentence, a yes-or-no comprehension question was presented. Subjects used letter keys to indicate their responses. The entire experiment took about 20 min. Subjects averaged 94% correct on the comprehension questions associated with the experimental items and 90% correct overall.

Results

Visual inspection of the reaction time distributions revealed only a few long outliers for the critical word and the subsequent
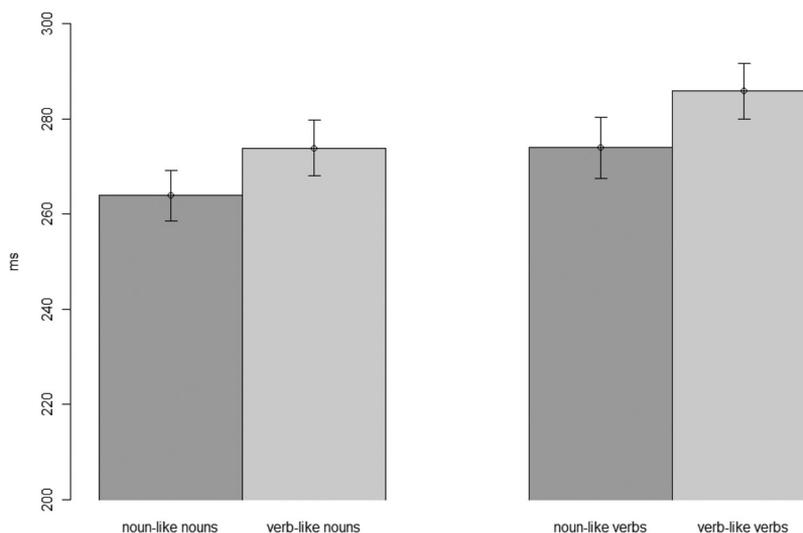

Figure 1. Gaze duration means in Experiment 1, collapsed across the sentence-frame factor. Error bars represent standard error of the mean.

Table 3
Regression Weights in Linear Mixed-Effects Model for Each of the Reading-Time Measures on the Critical Word in Experiment 1

	First fixation		Gaze		Go-past	
Variable	Estimate	p	Estimate	p	Estimate	p
Intercept	302.201	.0001	316.709	.0001	370.517	.0001
Part of speech	4.593	.459	13.902	.108	10.495	.562
Phonological classification	5.970	.315	9.443	.247	17.220	.317
Part of Speech × Phonological Classification	3.551	.661	0.281	.966	−22.801	.360
Length	−1.692	.398	2.594	.329	7.434	.216
Log frequency	−5.613	.002	−7.72	.001	−11.548	.023

word; trimming at 1000 ms eliminated three observations for each word position. The same mixed linear model that was fit to the data from Experiment 1 was applied here. The dependent measures were buttonpress latency on the critical word (where Farmer et al., 2006, found their effects) and on the following word. Reading-time means are presented in Table 4 and in Figure 2. It is apparent that, once again, the sentence-frame manipulation did not significantly affect reading times, nor were there notable interactions between this manipulation and the other experimental manipulations. As in Experiment 1, noun-like words were read slightly faster than verb-like words, but unlike in Experiment 1, verbs were read faster than nouns.

Model parameters and *p* values are presented in Table 5. There were no fully significant effects of any experimental manipulations. The effect of part of speech was marginally significant on the critical word. As in Experiment 1, there was no hint of the critical interaction effect. Unlike in Experiment 1, neither the frequency effect nor the length effect was significant, but both went in the predicted direction (i.e., reading times were longer on words that were longer and lower in frequency).

Discussion

The central result of Experiment 2 is simply the null finding of no phonological typicality effect. But there are a few data patterns in Experiment 2 that are worth discussing briefly. First, Experiment 2 found, once again, that noun-like items were read nonsig-

nificantly faster than verb-like items, for both nouns and verbs. If this is indeed a reliable effect, it would appear on its surface to be phonological in nature (although not a typicality effect). However, there are other possible explanations. For example, White (2008) and Staub, Hollway, White, and Rayner (2008) found that orthographic familiarity, as measured by the summed token frequency of letters, bigrams, and trigrams within a word, affects fixation durations in normal reading and response time in single word-recognition paradigms. Thus, as phonological differences between nouns and verbs are likely to correspond to spelling differences, it is plausible that orthographic familiarity may play a role in the present finding. Second, unlike in Experiment 1, reading times in Experiment 2 were marginally slower for nouns than for verbs. The context that preceded the critical word varied systematically between the nouns and verbs, and it is quite likely that this variation interacted differently with the different tasks. For example, the one-character difference in the length of the immediately preceding word ("the" or "to") is likely to have affected the probability that this word was directly fixated in normal reading (Drieghe, Rayner, & Pollatsek, 2005) and, hence, likely to have affected the eyes' landing position on the critical word, whereas no similar effect would have arisen in self-paced reading. Finally, it may appear surprising that no frequency effect was obtained in Experiment 2. But although frequency effects are very robust in the eye-movement record (Inhoff & Rayner, 1986; Rayner & Duffy, 1986) and in single-word paradigms (Balota & Chumbley, 1984, 1985), we know of no convincing demonstration of frequency effects in self-paced reading.

Naming and Lexical Decision Data

In addition to their reading-time studies, Farmer et al. (2006) also investigated, in their Experiment 1, the question of whether phonological typicality explains a significant portion of variance in existing naming latency data. Using a hierarchical regression analysis, they examined naming data for a total of 370 nouns and 70 verbs (obtained from Spieler & Balota, 1997). For both nouns and verbs, Farmer et al. found significant effects on naming latency of phonological distance to other words in the same syntactic category. However, Farmer et al. did not report naming latency data for the specific 20 nouns and 20 verbs that they used in their reading-time studies. As a final check on the reliability of our null findings, we obtained both naming and lexical decision latencies for each of the critical words from the English Lexicon Project database (Balota et al., 2002); the means for each word category are presented in Table 6.

Table 4
Mean Reading Time in Experiment 2 (in ms) for Critical Word and for Subsequent Word, by Experimental Condition and by Original (Farmer et al., 2006) Versus New Sentence Frame

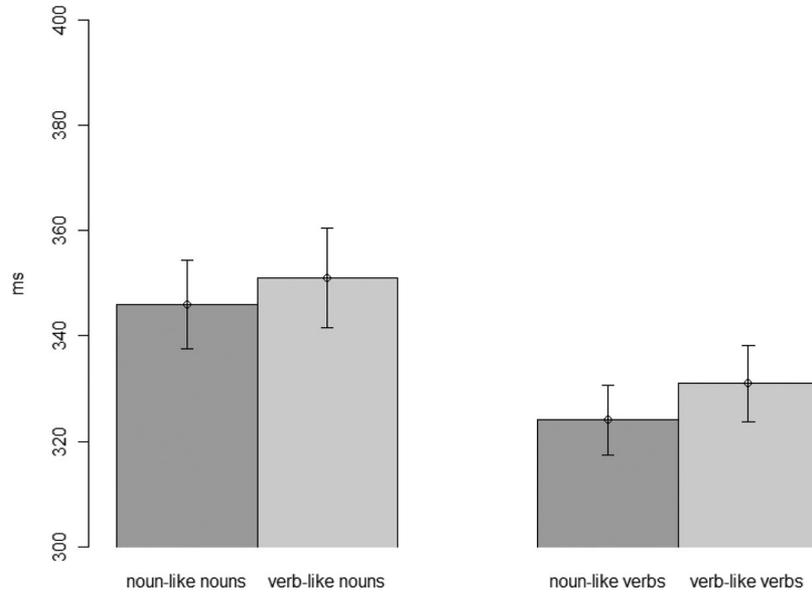|	Critical word		Subsequent word	
Word type	M	SD	M	SD
Farmer et al. frames				
Noun-like nouns	342	11.6	322	10.5
Verb-like nouns	356	13.9	344	12.2
Noun-like verbs	326	9.6	324	9.9
Verb-like verbs	328	9.6	331	10.0
New frames				
Noun-like nouns	351	12.5	342	12.4
Verb-like nouns	347	12.9	347	10.5
Noun-like verbs	322	9.1	326	10.9
Verb-like verbs	333	11.1	345	11.4

Figure 2. Reading-time means on the critical word in Experiment 2, collapsed across the sentence-frame factor. Error bars represent standard error of the mean.

As in both experiments presented here, there is a trend in the lexical decision and naming data toward faster responses for noun-like items. We conducted 2 × 2 analyses of variance (ANOVAs), with part of speech and phonological classification as fixed factors and items as the random factor, on both the naming and lexical decision data. In the naming data, there was a marginal effect of part of speech, with nouns named faster than verbs, $F(1, 36) = 3.35$, $p = .08$. There was no hint of an interaction between part of speech and phonological classification, $F(1, 36) = .60$, $p = .44$. No latency effects approached significance in the lexical decision data. Also reported in Table 6 are mean accuracy for both naming and lexical decision tasks, for each category of word. On the basis of ANOVA, naming accuracy was marginally higher for nouns than for verbs, $F(1, 36) = 3.71$, $p = .06$, but no other effects approached significance. There was a significant interaction effect on lexical decision accuracy, $F(1, 36) = 5.14$, $p = .03$, as responses to noun-like nouns were more accurate than responses to verb-like nouns, and responses to verb-like verbs were more ac-

curate than responses to noun-like verbs, apparently showing the first evidence we have uncovered of a phonological typicality effect. However, if the ANOVA with percentage correct as the dependent measure is replaced by logistic regression (as Jaeger, 2008, convincingly argued is required for accuracy data; see also Baayen, 2008), the apparent interaction is no longer significant ($p = .11$). Thus, there is no convincing evidence that naming or lexical decision latency or accuracy is affected by phonological typicality, for the specific items used by Farmer et al. (2006) and in the present experiments. We admit that the present analysis has little power to detect such an effect, given the small number of items involved and the variability associated with naming and lexical decision data, and we do not have reason to doubt the results that Farmer et al. obtained using a larger regression model.

General Discussion

The results of the present experiments are very easily summarized. In neither the eye movement record (Experiment 1) nor self-paced reading (Experiment 2) were we able to find a phonological typicality effect. It appears that phonological typicality does not affect measures of lexical processing in normal reading, either the early measure of first fixation duration (cf. Tanenhaus & Hare, 2007) or the somewhat later measures of gaze duration and go-past time. Moreover, it appears that the original self-paced reading result is not reliable. This raises the question of how Farmer et al.'s (2006) findings should be explained. Prior to conducting Experiment 2, we entertained the hypothesis that the difference in methodology was relevant, specifically speculating that the self-paced reading procedure may encourage subjects to read in a somewhat unusual way, perhaps engaging in more explicit subvocalization than in normal reading. But after conducting Experiment 2, it appears more likely that the Farmer et al. result is simply a Type I error.

Table 5
Regression Weights in Linear Mixed-Effects Model for Reading Time on the Critical Word and the Subsequent Word in Experiment 2

Variable	Critical word		Subsequent word	
	Estimate	p	Estimate	p
Intercept	335.915	.0001	335.070	.0001
Part of speech	−18.665	.065	−7.129	.456
Phonological classification	7.090	.475	13.293	.144
Part of Speech × Phonological Classification	−1.594	.930	−.448	.971
Length	4.719	.142	1.099	.702
Log frequency	−2.090	.439	−.998	.679

Table 6

Mean Lexical Decision and Naming Reaction Time and Accuracy for Critical Words, Based on English Lexicon Project Database (Balota et al., 2002)

Word type	Lexical decision reaction time		Naming reaction time		Lexical decision accuracy		Naming accuracy	
	M	SE	M	SE	M	SE	M	SE
Noun-like nouns	625	21.8	595	7.0	.973	.009	1.000	0
Verb-like nouns	626	18.1	627	10.8	.958	.010	.992	.005
Noun-like verbs	630	15.7	612	17.0	.941	.014	.970	.015
Verb-like verbs	655	16.2	624	12.6	.973	.005	.989	.006

There is yet another possibility, namely that the critical difference between the experiments reported here and the experiments reported by Farmer et al. (2006) relates to the intermixing of the two sets of experimental materials in a single session: It is possible that a phonological typicality effect appears when a subject is exposed either to typical and atypical nouns *or* to typical and atypical verbs, but not to both. If so, this would suggest that the phonological typicality effect reflects task-dependent strategic factors, as opposed to the processes involved in normal word recognition.

Finally, it is important to note that Farmer et al. (2006) conducted one experiment not discussed here, in which they found an effect of phonological typicality on the resolution of a noun/verb syntactic category ambiguity. We remain agnostic about this result.

We conclude by offering an observation about the current state of research on lexical processing in reading. We would point out that models of eye-movement control in reading, such as EZ-Reader (Reichle et al., 2003) and SWIFT (Engbert, Nuthmann, Richter, & Kliegl, 2005), have successfully accounted for a large portion of the variance in word-reading times, emphasizing factors such as frequency, length, and lexical predictability. There are some plausible sources of variance that are still relatively unexplored, such as variance due to the syntactic context in which a word appears (e.g., Levy, 2008). But we think it is rather unlikely that there are yet-to-be-discovered lexical variables that have large (i.e., on the order of 50 ms) effects on word-reading times. We suspect that future progress in identifying relevant lexical variables will proceed in small steps. Indeed, it was the magnitude of the effect reported by Farmer et al. (2006) that originally raised a red flag.

Finally, we turn to the underlying theoretical question of whether distributional regularities in the relationship between a word's sound and its meaning (or syntactic function) affect word recognition. On the basis of the present study, the answer appears to be no, at least with respect to reading. However, we think it is plausible that this question will ultimately receive an affirmative answer in the domain of auditory word recognition (cf. Sereno & Jongman, 2000).

References

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics.* Cambridge, England: Cambridge University Press.

Baayen, R. H., Davidson, D. H., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412.

Baayen, R., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database.* Philadelphia: Linguistic Data Consortium.

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: Access and/or production? *Journal of Memory and Language, 24,* 89–106.

Balota, D. A., Cortese, M. J., Hutchison, K. A., Neely, J. H., Nelson, D., Simpson, G. B., Treiman, R. (2002). *The English Lexicon Project: A web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords.* Washington University in St. Louis, MO. Available from: http://elexicon.wustl.edu/

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

Balota, D. A., Yap, M. J., & Cortese, M. J. (2007). Visual word recognition: The journey from features to meaning (A travel update). In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 285–376). London: Elsevier.

Bates, D. M. (2005). Fitting linear mixed models in R: Using the lme4 package. *R News: The Newsletter of the R Project, 5*(1), 27–30.

Burgess, C., & Livesay, K. (1998). The effect of corpus size in predicting reaction time in a basic word recognition task: Moving on from Kučera and Francis. *Behavior Research Methods, Instruments & Computers, 30,* 272–277.

Drieghe, D., Rayner, K., & Pollatsek, A. (2005). Eye movements and word skipping during reading revisited. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 954–959.

Elman, J., Bates, E., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development.* Cambridge, MA: MIT Press.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112,* 777–813.

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences of the United States of America, 103,* 12203–12208.

Fodor, J. A. (1983). *Modularity of mind.* Cambridge, MA: MIT Press.

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics, 40,* 431–439.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434–446.

Jared, D. (2002). Spelling-sound consistency and regularity effects in word naming. *Journal of Memory and Language, 46,* 723–750.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87,* 329–354.

Lau, E., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain and Language, 98,* 74–88.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106,* 1126–1177.

Pollatsek, A., Perea, M., & Binder, K. (1999). The effects of neighborhood size in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1142–1158.

Rastle, K. (2007). Visual word recognition. In G. Gaskell (Ed.), *Oxford handbook of psycholinguistics* (pp. 71–88). Oxford, England: Oxford University Press.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory and Cognition, 14,* 191–201.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading.* Englewood Cliffs, NJ: Prentice Hall.

R Development Core Team. (2007). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. Available from: http://www.R-project.org

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26,* 445–526.

Sereno, J. A. (1994). Phonosyntactics. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound symbolism* (pp. 263–275). Cambridge, England: Cambridge University Press.

Sereno, J. A., & Jongman, A. (2000). Phonological and form class relations in the lexicon. *Journal of Psycholinguistic Research, 19,* 387–404.

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science, 8,* 411–416.

Staub, A., & Clifton, C., Jr. (2006). Syntactic prediction in language comprehension: Evidence from *Either. . . or. Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 425–436.

Staub, A., Hollway, E. C., White, S. J., & Rayner, K. (2008). *Comparing distributional effects of lexical frequency and orthographic familiarity on eye fixations, lexical decision, and semantic categorization.* Manuscript submitted for publication.

Staub, A., & Rayner, K. (2007). Eye movements and on-line comprehension processes. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 327–342). Oxford, England: Oxford University Press.

Tanenhaus, M., & Hare, M. (2007). Phonological typicality and sentence processing. *Trends in Cognitive Sciences, 11,* 93–95.

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.), *Handbook of perception and cognition: Speech, language, and communication* (2nd ed., Vol. 11, pp. 217–262). San Diego, CA: Academic Press.

White, S. J. (2008). Eye movement control during reading: Effects of word frequency and orthographic familiarity. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 205–223.

Yates, M., Locker, L., & Simpson, G. (2004). The influence of phonological neighborhood on visual word perception. *Psychonomic Bulletin & Review, 11,* 452–457.