Journal of Experimental Psychology: Human Perception and Performance

Form-To-Expectation Matching Effects on First-Pass Eye Movement Measures During Reading

Thomas A. Farmer, Shaorong Yan, Klinton Bicknell, and Michael K. Tanenhaus

Online First Publication, April 27, 2015. http://dx.doi.org/10.1037/xhp0000054

Journal of Experimental Psychology:
Human Perception and Performance

0096-1523/15/$12.00 http://dx.doi.org/10.1037/xhp0000054

Form-To-Expectation Matching Effects on First-Pass Eye Movement Measures During Reading

Thomas A. Farmer and Shaorong Yan
University of Iowa

Klinton Bicknell
Northwestern University

Michael K. Tanenhaus
University of Rochester

Recent Electroencephalography/Magnetoencephalography (EEG/MEG) studies suggest that when contextual information is highly predictive of some property of a linguistic signal, expectations generated from context can be translated into surprisingly low-level estimates of the physical form-based properties likely to occur in subsequent portions of the unfolding signal. Whether form-based expectations are generated and assessed during natural reading, however, remains unclear. We monitored eye movements while participants read phonologically typical and atypical nouns in noun-predictive contexts (Experiment 1), demonstrating that when a noun is strongly expected, fixation durations on first-pass eye movement measures, including first fixation duration, gaze duration, and go-past times, are shorter for nouns with category typical form-based features. In Experiments 2 and 3, typical and atypical nouns were placed in sentential contexts normed to create expectations of variable strength for a noun. Context and typicality interacted significantly at gaze duration. These results suggest that during reading, form-based expectations that are translated from higher-level category-based expectancies can facilitate the processing of a word in context, and that their effect on lexical processing is graded based on the strength of category expectancy.

Keywords: eye movements, reading, prediction, form-based expectations

Reading involves the coordination of linguistic, visual, and oculo-motor systems, all of which work together to facilitate a decision about whether to move the eyes to a new word, or to gather more information from the word that is currently fixated. This decision must be made extremely quickly given that the average fixation on a word is approximately 200 ms and it takes time to plan or cancel a saccade. How then do multiple systems coordinate to produce a decision on such a fast time-scale? One part of the answer to this question is that both readers and listeners rely heavily on knowledge about the structure of language to generate predictions for many aspects of upcoming linguistic input that serve to speed processing at multiple levels when incoming input exhibits properties consistent with expectancies (e.g., Altmann & Kamide, 1999; Arai & Keller, 2013; Bicknell, Elman, Hare, McRae, & Kutas, 2010; Brown, Salverda, Dilley, & Tanenhaus, 2011; DeLong, Urbach, & Kutas, 2005; Farmer, Christiansen, & Monaghan, 2006; Federmeier, 2007; Hale, 2001; Kamide, Altmann, & Haywood, 2003; Kimball, 1975; Levy, 2008; Staub & Clifton, 2006; van Berkum, Brown et al., 2005; see Kamide, 2008 for an overview of anticipatory effects in the sentence processing literature). But, how close to low-level perceptual processing do knowledge-driven expectancies reach before the sensory transduction of a newly fixated word?

Here, we pursue the hypothesis that higher-level expectancies can be translated into low-level form-based estimates of the visual information that is likely to be encountered during a subsequent fixation (e.g., Dikker et al., 2009; Tanenhaus & Hare, 2007). The availability of form-based expectations to sensory cortex may serve as a template that facilitates low-level perceptual processing of the visually transduced signal, or as the basis for the production of an error signal (i.e., "prediction error") upon encountering form-based properties that are inconsistent with higher-level expectancies (see also Carreiras, Armstrong, Perea, & Frost, 2014, for a review of work supporting the role of form-based expectations during reading). The matching of physical form to perceptual expectancies may play a central role in negotiating the delicate balance between staying on a word or leaving it on such a fast time-scale.

Thomas A. Farmer and Shaorong Yan, Department of Psychology, University of Iowa; Klinton Bicknell, Department of Linguistics, Northwestern University; Michael K. Tanenhaus, Department of Brain and Cognitive Sciences, University of Rochester.

We thank Anne Pier Salverda, Morten Christiansen, and Padraic Monaghan for their comments on and contributions to earlier versions of this work, as well as Danielle Reece and Francis Smith for their assistance with data collection. Additionally, we are grateful to Barbara Juhasz and four additional anonymous reviewers for detailed and cogent reviews that resulted in a much-improved article. This research was partially supported by NIH Grant HD27206 to MKT and by NIH training grant (T32 DC000035) to the Center for Language Sciences at the University of Rochester.

Correspondence concerning this article should be addressed to Thomas A. Farmer, Department of Psychology, 11 Seashore Hall E, Iowa City, IA 52242-1407. E-mail: thomas-farmer@uiowa.edu

First-pass eye movement measures such as *skip rate* (the probability that a word will be skipped), *first fixation duration* (amount of looking time for the initial fixation on a word), *gaze duration* (the sum of all fixation times on a word before the eyes leave the word for the first time, either to the left or the right), and *go-past time* (the total amount of time spent on a word before the eyes initially move past it to the right), demonstrate sensitivity to manipulations of form-based properties of a word, such as length, frequency, and familiarity (e.g., Inhoff & Rayner, 1986; Rayner & Duffy, 1986; Williams & Morris, 2004), among other lexical variables (e.g., Juhasz & Rayner, 2003). Some contextual variables also exert an influence on first-pass eye movement measures. When sentential context is highly predictive of a specific word, for example, participants are significantly more likely to skip the predictable word, and if the word is fixated, both first fixation and gaze durations are longer when the target word is unexpected (e.g., Altarriba, Kroll, Sholl, & Rayner, 1996; Ashby, Rayner, & Clifton, 2005; Balota, Pollatsek, & Rayner, 1985; Ehrlich & Rayner, 1981; Rayner, Ashby, Pollatsek, & Reichle, 2004; Rayner & Well, 1996; Kliegl, Grabner, Rolfs, & Engbert, 2004). Given the well-documented effects of lexical-level variables and lexical predictability on first-pass eye movement measures, form-based expectations—if they exist—would be most likely to influence these measures. We test this prediction of the form-to-expectation matching hypothesis by manipulating both the degree of category predictability associated with an upcoming word and the degree to which the physical form-based properties of that word are typical—as opposed to atypical—of other words in that given category, to determine whether these variables interact in their influence on first-pass eye movement measures.

Phonological Typicality Effects in Category Predictive Contexts

Context is rarely constraining enough to support the generation of expectations for a specific word (e.g., Jackendoff, 2002; Stanovich & West, 1979), although it is more likely to support the reliable generation of expectations for a word from a specific category.

(1a) The curious young boy saved the *marble* that he . . . (noun-like noun)

(1b) The curious young boy saved the *insect* that he . . . (verb-like noun)

Upon encountering a sentence frame such as "*The curious young boy saved the* . . ." (1), readers cannot accurately anticipate the specific lexical item that they are likely to next encounter (the context provided by this isolated sentence is not constraining enough). However, based on the strong bias of the main verbs to be followed by a direct object (DO) Noun Phrase (NP) (as indexed by verb-bias norms) and the presence of a determiner, readers can anticipate the word's likely grammatical category. For example, the next word is quite likely to be a noun but unlikely to be a verb.[1]

Nouns and verbs differ from one another across a number of stimulus dimensions inherent to their physical form (e.g., length, lexical stress, word onset properties, see Kelly, 1992 for an overview), providing probabilistic cues to grammatical category that can facilitate word category learning during acquisition (Cassidy & Kelly, 2001; Fitneva et al., 2009; Monaghan, Chater, & Christiansen, 2005) and aid in lexical processing (Arciuli & Monaghan, 2009; Kelly, 1998). Capitalizing on these probabilistic differences in the form-based properties of nouns and verbs, Farmer, Monaghan, and Christiansen (2006) created a phonological feature-based index of *phonological typicality,* which they defined as the degree to which the sound properties of an individual word are typical of other words in the same grammatical category. Nouns and verbs from the CELEX database (Baayen, Piepenbrock, & van Rijn, 1993) were converted into slot-based representations, and each phoneme was coded based on the presence, absence, or degree of 11 different phonemic features (taken from Harm & Seidenberg, 1999, and listed at the bottom of Table 6 in Appendix A). For each word, the average euclidean distance was calculated between its phonological representation and all nouns, and also between all verbs. An index of category typicality was created by subtracting the average distance in phonological feature space of a word to all other words in its own grammatical category from its average distance to all words in the opposite category. The resulting metric provides an index of the degree to which an individual noun or verb is "noun-like" or "verb-like" with respect to its physical form-based features (more details about the calculation of phonological typicality can be found in Appendix A). We note that because orthographic similarity and phonological similarity are highly correlated, we use phonological typicality as a proxy for orthographic typicality. Thus, while the metric quantifies phonological typicality, we will interpret the results as effects of visual form-typicality.

Farmer et al. (2006) found that reading times elicited in a word-by-word self-paced reading paradigm were faster on the target noun following the main verb and determiner when its form-based properties were typical, as opposed to atypical, of other nouns. In a separate experiment, participants read sentences with main verbs that were designed to facilitate a strong expectation for an infinitival complement structure containing a verb (2). Reading times were significantly slower when the verb that appeared in the strongly expected infinitival complement structure contained a constellation of phonemic features that was atypical of other verbs.

(2a) The very old man attempted to *assist* his . . . (verb-like verb)

(2b) The very old man attempted to *vary* his . . . (noun-like verb)

One interpretation of these results is that upon processing the bias-conferring main verb of these sentences, readers were able to generate expectancies not specific to the physical form-based properties of the target word itself, given that its specific identity could not be anticipated from context, but instead for the form-based properties that are typical of other words in the same heavily expected category. However, there is an alternative explanation that does not appeal to expectations generated from context. It is possible that category atypical words may take longer to recognize, regardless of contextual bias. Indeed, Farmer et al. (2006, Study 1) reported an effect of typicality on lexical naming latencies, providing some evidence that word-form typicality effects can occur without a category-biasing sentential context.

Several experiments suggest that category-based expectations modulate the strength of the aforementioned phonological typical-

[1] We note that other syntactic categories (e.g., adjectives) can also occur in this environment. However, use of prenominal modification is strongly conditioned on the referential context, that is, when modification is needed to distinguish among alternative salient referents (e.g., Sedivy, 2003).

ity effects. For example, in a self-paced reading and an eye-tracking experiment, Staub, Grant, Clifton, and Rayner. (2009) reported a failure to replicate the typicality effects reported in Farmer et al.'s (2006) Experiment 2 (noun) and Experiment 3 (verb), discussed above. Their design, however, deviated from Farmer et al.'s in that the materials from each separate experiment were intermixed as opposed to blocked. The sentential contexts for the noun- and verb-predictive contexts contained the same syntactic form up through the category-biasing main verb, producing an experimental context in which half of the items contained main verbs with a strong noun bias, and half with a verb structure bias. Farmer, Monaghan, Misyak, and Christiansen (2011) demonstrated that because of the strong structural overlap in the sentences up until the target word, participants learned, over the course of experience with the intermixed design, to down-weight their reliance on the verb-bias cue. The typicality effect was present for noun- and verb-predictive contexts early in the experiment, but decreased (in the case of noun frames) or flipped (in the verb frame case) over the course of the experiment. In a following rejoinder, Staub et al. (2011) acknowledged that grammatical knowledge is likely to facilitate expectations for grammatical properties of downstream information during sentence processing. However, they argued that grammatical knowledge was not likely to be malleable enough to facilitate a change in expectations as a function of experience with the set of items in the task. An increasing body of recent work, however, has demonstrated that higher-level expectancies are adjusted over the course of sentence processing experiments as participants learn about probabilistic contingencies in the experimental context (e.g., Fine, Jaeger et al., 2013; Fine, Qian et al., 2010; Kamide, 2012; Kaschak & Glenberg, 2004). Readers do seem to possess the ability to adapt their expectations for various properties of an unfolding linguistic signal in order to best approximate the distributions of linguistic information in a specific communicative context.

A more recent experiment by Dikker, Rabagliati, Farmer, and Pylkkänen (2010) provides additional support for contextually conferred category predictability as one strong determinant of phonological typicality effects. They demonstrated that the magnitude of the M100 response—an MEG component generated in visual cortex ~100 ms–120 ms poststimulus onset during reading that often reflects sensitivity to manipulations of low-level form-based properties of a visual stimulus, such as the size of letter strings (Tarkiainen, Helenius, Hansen, Cornelissen, & Salmelin, 1999)—was modulated by the degree of match between the phonological typicality of an individual lexical item and a contextually conferred expectation for grammatical category during reading. When a sentential context created a strong expectation that a noun would **not** appear, and yet a noun-like noun did appear (as in *soda* in *The tastelessly soda* . . .), the magnitude of the M100 response was larger than when there was a strong expectation for a noun, and the same noun-like noun appeared (*The tasteless soda* . . .). No effect of expectation occurred when the form of the target noun was neutral (its typicality score was roughly equal to zero, indicating that the physical form-based properties of the word were not, on average, more or less typical of words in either category). These results suggest that phonological typicality effects are at least partially contingent upon contextually conferred expectations for a word from a specific grammatical category.

The Present Experiments

The results of Dikker et al. (2010) suggest that the phonological typicality effect is replicable in experiments where sentential context is strongly predictive of a word's upcoming category, and are also consistent with the notion that form-based expectations corresponding to higher-level category-based expectancies can guide the processing of transduced visual stimuli at a relatively low level during reading. The word-by-word presentation with extended interstimulus intervals inherent to EEG/MEG paradigms is, however, atypical of natural reading. Additionally, Staub et al.'s failure to elicit a corresponding phonological typicality effect in an eye movement experiment leaves open the possibility that form typicality effects in category-predictive contexts are limited to paradigms that rely on unnatural word-by-word stimulus presentation. In light of these observations, the three experiments presented here were designed to address two remaining questions: (a) Does phonological typicality affect first-pass eye movement measures during natural reading?; and (b) Does the strength of expectation for a word from a specific grammatical category influence the strength of a phonological typicality effect on first-pass measures?

We focus our efforts on the noun typicality effect because of differences in the syntactic distributions of nouns and verbs. Whereas nouns tend to occur in a limited number of syntactic constructions, verbs are less predictable. The typicality effect originally reported for verbs included sentences with main verbs that conferred an expectation for a an *inf-comp* structure, although the probability with which the target verb occurs in an *inf-comp* construction as opposed to a wide array of other syntactic constructions in which it could occur was not controlled. Readers are acutely sensitive to distributional information associated with verbs (e.g., Garnsey, Pearlmutter, Myers, & Lotocky, 1997), even when presented in isolation (e.g., Linzen, Marantz, & Pylkkanen, 2013; see also Monaghan, Christiansen, & Chater, 2007 for a discussion of the effects of the differential distributional properties of nouns and verbs). Thus, nouns produce a more tractable testing ground than verbs for examining contextually mediated category-based word-form typicality effects.

Experiment 1

In Experiment 1, we tracked participants' eye movements as they read sentences that conferred a strong expectation for a noun, predicting an effect of form-based typicality—longer fixation durations for verb- as opposed to noun-like nouns—on the first-pass fixation measures discussed above. If readers generate form-based expectations for upcoming visual input, at least when context is strongly constraining for grammatical category, the effects of encountering word-form properties that are inconsistent with contextually conferred expectations should influence first-pass eye movement measures, where sensitivity to form-based and lexical-level properties of a newly fixated word has been repeatedly demonstrated.

Method

Participants. Thirty-eight native English-speaking undergraduates from the University of Rochester participated in this experiment for $10 in compensation. All participants reported normal or corrected-to-normal vision.

Materials. We created 18 experimental items (1), each containing a four-word preamble (*The curious young boy*), a strongly NP-biased verb (*saved*, in 1), a determiner, the target noun, and the remaining portion of the sentence. No significant difference existed in the number of words appearing after the target word in the noun-like ($M = 5.34$, $SD = 1.72$) and verb-like ($M = 5.78$, $SD = 1.31$) conditions ($p = .24$), although not all sentences contained the same posttarget word continuations due to semantic constraints imposed by the differing target words across each typicality condition (all materials used in this and the following two experiments are listed in Appendix B). Additionally, the length of the word appearing immediately after the target word did not differ across sentences in the noun-like ($M = 3.39$, $SD = 1.20$) or verb-like ($M = 3.72$, $SD = 1.81$) noun conditions ($p = .34$). One version of each item contained a noun-like target noun (*marble*, in 1a) and another contained a verb-like noun (*insect*, in 1b). Ten of the experimental items were taken directly from Farmer et al.'s (2006) Experiment 2. To increase the power of the design, we created eight new items by identifying another eight DO NP-biased verbs. For the 10 items taken from Farmer et al. (2006), the bias of a verb to be followed by a DO NP was quantified by examining verb norms provided by Connine et al. (1984), with a mean percentage of DO NP completion = 87.7% ($SD = .068$). The new bias-conferring verbs were identified by placing a set of verbs in sentence preambles such as the one in example (1), and asking 40 participants on Amazon's MTurk to complete the sentence fragment ($M = 96.88\%$ NP completion, $SD = 2.10$).

Noun-like and verb-like target words did not differ significantly on variables that often exert robust effects on visual word recognition, such as orthographic length, number of phonemes, log frequency, and number of neighbors. Additionally, we controlled for lexical-level properties of the target words that have been demonstrated to influence gaze duration, such as concreteness and age of acquisition (e.g., Juhasz & Rayner, 2003; all p's $> .12$, see Table 1 for descriptive statistics associated with control variables). The 36 sentences from the 18 experimental items were counterbalanced across two presentation lists such that each participant only saw one version of each item, but an equal number of trials per condition. Each list also contained 54 unrelated filler items along with two practice items. The filler sentences contained no psycholinguistic manipulation, but were matched roughly in length to the experimental sentences and included multiple types of syntactic constructions (, e.g., "*The sales clerk acknowledged that the error should have been detected earlier;*" or "*The minister blessed the food before the banquet, and the rabbi blessed it too.*").

Table 1
Means and Standard Deviations for Control Variables Across Each Typicality Condition in Experiment 1

	Noun-like noun		Verb-like noun	
Control	Mean	*SD*	Mean	*SD*
Orthographic length	6.2	1.4	5.8	1.0
Number of phonemes	5.9	0.6	5.6	1.0
Log-frequency[a]	9.5	1.3	9.1	1.2
Number of neighbors[b]	1.8	1.9	2.8	3.4
Concreteness[c]	4.6	0.3	4.3	0.8
Age of acquisition[d]	6.4	0.6	5.5	1.0

[a] Hyperspace Analogue to Language corpus (Burgess & Livesay, 1998). [b] English Lexicon Project (Balota et al., 2007). [c] Brysbaert, Warriner, and Kuperman (2014). [d] Kuperman, Stadthagen-Gonzalez, and Brysbaert (2012).

Procedure. Participants were randomly assigned to a presentation list, and the presentation order of items was randomly assigned per participant. Eye movements were recorded with an EyeLink 1000 eye-tracker at a sampling rate of 1,000 Hz. Viewing was binocular but data were only recorded from the right eye. Stimuli were presented in 14-pt Courier New font on a 19-inch ViewSonic CRT monitor with a $1{,}024 \times 768$ pixel resolution. Participants were seated ~60 cm from the screen, with their head positioned on a chin rest, such that 2.7 characters equaled 1 degree of visual angle. Participants were asked to read the sentences in a normal manner, and to press a button on a hand-held controller when finished. Participants answered a yes/no comprehension question (e.g., "Did the boy save something?" for the sentence in Example 1, above) about each sentence after it was read.

Results and Discussion

For the eye movement data, trials that contained either a blink or track loss were excluded, affecting less than 5.1% of total trials. Fixations less than 80 ms in duration and less than one character away from an adjacent fixation were incorporated into the nearest fixation. Fixations less than 80 ms that were more than one character away from an adjacent fixation were removed. No trials produced a fixation duration longer than 1,000 ms for the first-pass measures. Linear mixed-effects models (for fixation durations) and mixed logit models (for skipping and comprehension question accuracy) were adopted in analyzing the measures. The analyses were implemented with the lme4 package (Bates, Maechler, & Bolker, 2012) in the R environment (R Development Core Team, 2014).

Mean fixation times (or fixation probabilities) for each typicality condition, per measure, are presented in Table 2. The model for each dependent measure included typicality (noun-like or verb-like), word length, and log-frequency. Although target nouns were sampled from opposite ends of the noun-like to verb-like continuum, continuous typicality scores were entered into the model to address subtle variability in each noun's proximity to its respective end-point on the continuum. Length and frequency were controlled a priori across condition, but following Staub et al. (2011), we included terms for these variables in the models to account for subtle differences in the amount of variability that existed across conditions for each variable. These variables were not, however, included in the random effects structure of the models. All continuous variables, including the length and log frequency control variables, were centered. The maximum random effects structure supported by the data for both participants and items was identified for every model based on model comparison using log-likelihood ratio tests (e.g., Pinheiro & Bates, 2000). Maximal random effects structures for participants and items were first identified. We then removed slopes and intercepts that did not improve model fits. We report t-values (for linear mixed-effect models), z-values (for

Table 2
Mean Fixation Duration (or Probability of Fixation) and Standard Errors for Each First-Pass Measure Examined in Experiment 1

	Noun-like noun		Verb-like noun	
Eye movement measure	Mean	*SE*	Mean	*SE*
Skipping rate (%)	9	2.3	9	1.9
First fixation time (ms)	209	4.3	227	5.3
Gaze duration (ms)	241	7.2	260	8.5
Go-past time (ms)	254	7.7	275	9.5

mixed logit models), and corresponding *p* values[2] (for *t*-values, estimated from the *t*-value using the Sattherthwaite approximation for degrees of freedom, as implemented in the lmerTest package; Kuznetsova, Christensen, & Brockhoff, 2014) for the analysis of each DV in Table 3.

Results from a logistic mixed effects model on comprehension question accuracy rates revealed that participants were equally accurate at answering comprehension questions in the noun-like (M = 96.0, SE = 1.1) and verb-like (M = 97.5, SE = 1.0) conditions, t = 1.23, $p > .1$. Consistent with previous work on word length and skipping (e.g., Drieghe, Brysbaert, Desmet, & De Baecke, 2004), shorter words were skipped more frequently. Significant frequency effects occurred on gaze duration (marginal for first fixation duration), with longer fixations on less frequent words, also consistent with previously reported effects of frequency on fixation duration measures (e.g., Rayner & Duffy, 1986). Crucially, when sentential context conferred a strong expectation for a noun, first-fixation durations (β = 19.25 ms), gaze durations (β = 21.71 ms), and go-past times (β = 26.90 ms) were significantly longer for verb-like nouns relative to noun-like nouns.[3]

We note here that although length and log-frequency were controlled across each condition a priori, significant effects were still elicited, signaling that our other control variables may still exert potentially confounding effects with respect to our assessment of the statistical reliability of the typicality effect. As a result, we ran a series of control models to more rigorously test for potentially confounding effects of each control variable listed in Table 1. We maintained the structure of the original model reported above (including the presence of length and log-frequency as control variables), and reassessed the effect of typicality after entering one of the additional centered control variables into the model as well. Due to the heavily intercorrelated nature of many of these variables, one model was conducted for each control variable on each eye movement measure. For each fixation duration measure reported above, the typicality effect always remained significant. No variable that we controlled for exerted a significant independent effect on any measure, although there was a near significant effect of *age of acquisition* (AoA) on gaze durations, t = 1.85, p = .07, such that later acquired words were fixated longer.

We note that typicality did not influence skip rate. One advantage of this category predictability manipulation is that given that only the grammatical category of the target word can be predicted, but not its identity, form-to-expectation matching effects can be assessed under conditions that are likely to minimize effects of parafoveal preview. During fixation, readers can glean visual information from the printed text occurring approximately four characters to the left of fixation and 14–15 character spaces to the right in English (e.g., McConkie & Rayner, 1975; McConkie & Rayner, 1976; Rayner, 1986; Rayner, Well, & Pollatsek, 1980), providing a potential source of bottom-up visual information to the reader before fixation, although the informational content that is available parafoveally may decrease as the distance from the current fixation increases (e.g., Hand, Miellet, O'Donnell, & Sereno, 2010; but see Slattery, Staub, & Rayner, 2012). Parafoveal preview effects are most heavily influenced by manipulations of form-based (orthographic and phonological) properties of the word in the parafovea, although in each case, lexical variables may also be assessed parafoveally (see Schotter, Angele, & Rayner, 2012, for an overview). Some experiments report only additive effects of predictability and form-related variables (e.g., Drieghe et al., 2004; Inhoff, Radach, Eiter, & Juhasz, 2003; Rayner, Slattery, Drieghe, & Liversedge, 2011), while others find that the probability of fixating a word is conditioned upon the degree to which form-based features of a target word are consistent with an expected word (e.g., White, Rayner, & Liversedge, 2005; Juhasz, White, Liversedge, & Rayner, 2008).

Parafoveal preview may provide enough information to facilitate a decision to skip, especially if the word is highly expected. In such a scenario, a predictable target word would be skipped more frequently than an unpredictable target word, creating unequal numbers of fixations on a target word across levels of a predictability manipulation. This differential skipping rate reduces statistical power (and can also induce sampling bias) when assessing the degree to which form-based manipulations of a target word may interact with contextual expectancy. In sentences such as (1), however, parafoveal information about the target word should be less robust. Upon processing the bias-conferring verb (i.e., *saved*, in 1), readers can easily anticipate a determiner, such that it is likely to be skipped, but may have limited parafoveal access to the entire target word. Additionally, the phonological typicality metric created by Farmer et al. produces an estimate of word-form, with respect to category membership, that is distributed across the entire word. Therefore, in this category predictability manipulation, readers should be less likely to skip the target word, facilitating a roughly equal number of fixations in each typicality condition. Indeed, the absence of a typicality effect on skip rate can be interpreted as evidence that readers experienced less parafoveal preview benefit than they would have if the n-1 fixation were closer to the target word. Given that we have not systematically manipulated the location of the n-1 fixation, however, future research will be necessary in order to better determine the relationship between skip rate, launch site, form-based typicality of a target word, and fixation durations on the target word.

[2] *p* values were retrieved by loading lmerTest when running mixed-effect models using lme4 package and reporting the *p* values included with model summaries.

[3] Although our analyses were conducted on raw fixation times per each fixation duration measure, a reviewer noted that, since the distribution of fixation durations are typically right-skewed, it may be more appropriate to analyze log durations. We note here that all analyses reported in this article produce the same pattern of results when log fixation durations are used instead of raw.

Table 3
Regression Coefficients and Test Statistics From Linear Mixed-Effects and Logistic Mixed-Effects Models for First-Pass Eye Movement Measures on the Target Noun in Experiment 1

	Skip rate			First-fixation time			Gaze duration			Go-past time		
	Est. β	*z*-value	*p*	Est. β	*t*-value	*p*	Est. β	*t*-value	*p*	Est. β	*t*-value	*p*
Intercept	−2.78	−9.95	<.001	217.84	44.81	<.001	249.32	31.85	<.001	263.20	29.74	<.001
Word length	−0.41	−2.59	0.010	−3.11	−1.14	0.258	4.10	0.93	0.354	7.81	1.58	0.119
Frequency (logHal)	0.15	1.27	0.203	−3.95	−1.71	0.091	−8.72	−2.33	0.022	−6.22	−1.49	0.139
Typicality	−0.21	−0.58	0.559	19.25	3.17	0.002	21.71	2.21	0.028	26.90	2.47	0.014

As evident in Figure 1, participants rarely skipped the bias-conferring verb or the target noun in either typicality condition, although the determiner was skipped on roughly 70% of trials. In another set of linear mixed effects models, we predicted first fixation duration, gaze duration, and go-past times on the target word. We included the same fixed effects structure as in the original model reported above, but included an additional predictor denoting whether or not the determiner was fixated (determiner fixation), and an interaction term between typicality and determiner fixation. A significant effect of determiner fixation occurred for each dependent measure, with shorter fixations on the target noun occurring on trials in which the determiner was fixated on the first-pass. Determiner fixation did not, however, interact with typicality for any measure, indicating that the typicality effect reported above was not driven by trials on which the determiner was fixated, and thus on trials where readers were likely to have more parafoveal preview of the target word.

Lastly, given that a subset of our items included words of differing length after the target word, we also conducted a control model, in the same manner as described above for the control variables listed in Table 1, that included the length of the word following the target noun. For each of the three fixation time measures, no significant effect of the length of the posttarget word occurred, and the typicality effect remained significant.

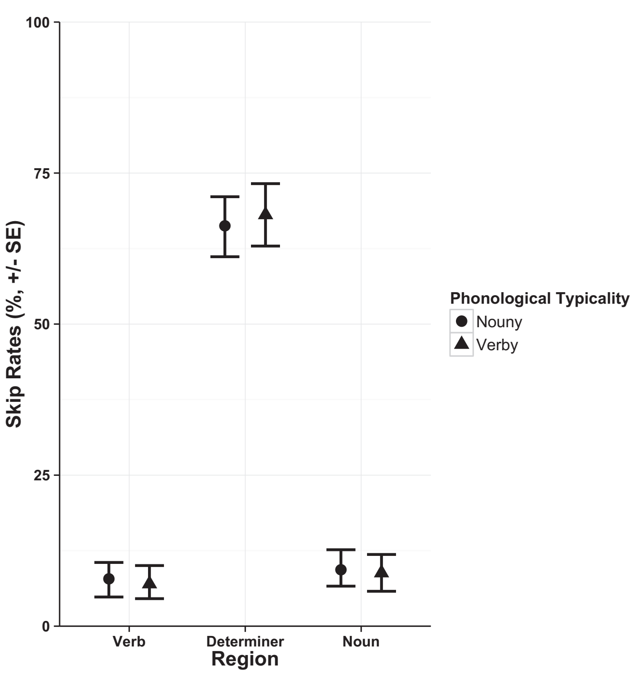

Figure 1. Percentage of trials (±*SE*) on which the bias-conferring verb (*saved*), the determiner (*the*), and the target noun (*noun-like or verb-like*) were skipped.

The typicality effect reported here is consistent with the effects of other form-based variables such as length and orthographic familiarity on first-pass eye movement measures (e.g., Kliegl et al., 2004; Rayner & Duffy, 1986), and the determiner fixation analysis provides evidence that participants were unlikely to be able to parafoveally assess, in a robust nature, every relevant form-based feature inherent to the target noun.

The pattern of results is consistent with the hypothesized role of form-based expectations in determining fixation duration on first-pass eye movement measures. We note, however, that the typicality effect reported here for nouns in noun-predictive sentential contexts does not rule out the possibility that the effect occurred solely as a result of differential levels of processing difficulty inherent to category atypical words, and thus without respect to contextual constraint. In Experiment 2, we address this issue by asking whether the *degree* of contextual support for a noun modulates the magnitude of the typicality effect, as would be predicted if the effects of typicality are driven by form-based expectations and not by word-specific form-based features.

Experiment 2

Using the same noun-like and verb-like noun pairings from Experiment 1, we manipulated contextual constraint by leaving half of these words in their *strongly noun-biased* contexts, and placing the other half in a *less noun-biased* context, as illustrated in (2).

(2) The second word on the sign was "marble"/"insect" and was spelled correctly.

Noun-like or verb-like nouns appeared in quotation marks, followed by at least three words (M = 5.75 words, SD = 2.71). Quotation marks were included around the target word for two reasons. First, they provide a strong visual cue to participants that, in principle, any word (from any category) could occur in between the quotation marks. Second, the quotation marks should attenuate surprisal effects associated with encountering words that may be interpreted as semantically incongruent given the lack of additional referential support. The goal of this experiment was to manipulate the strength of an expectation for a noun. The contextual manipulation employed here simply widens the distribution of words that could appear at the target location by increasing the likelihood of non-nouns in the distribution of predicted words. The net effect, we suggest, is that expectations for physical word-form features that are probabilistically associated with words from the noun category would be dampened by predictions for form-based properties of words from other categories.

If the typicality effect reported in the first-pass eye movement measures in Experiment 1 were the result of a phonological typicality effect on visual word recognition processes, and uninfluenced by expectancies generated from sentential context, then we would expect to see corresponding effects of word-from typicality on the first-pass eye movement measures. The size of the typicality effect, however, should not be modulated by the strength of a contextually conferred expectation for a noun. If, however, as we hypothesize, distributional widening will reduce the coherence of the form-based properties that could be predicted from context, then we would expect an interaction between typicality and context. Stronger expectations for a noun should elicit stronger typicality effects on first-pass eye movement measures.

Method

Participants. Ninety-six native English-speaking undergraduates from the University of Iowa participated in this experiment. All participants reported normal or corrected-to-normal vision and received research credit for their participation in an undergraduate psychology subject pool.

Materials.

Sentence norming. Although words from other categories could, in principle, occur within the quotation marks, the bias of the sentential context was estimated using a norming procedure in which participants were asked to rate the degree to which they anticipated that a noun versus a verb would appear. More nouns and verbs exist in English relative to words from other lexical categories, and the phonological typicality metric investigated here is computed based on the noun/verb distinction. As a result, we suggest that focusing on the noun/verb distinction is an appropriate way to generate norming values for each sentence frame given the goals of the contextual manipulation. We created 20 sentence frames that were designed to produce a less strong bias for a noun, as in (2), above. These frames were combined with 20 strongly biased sentence frames (the 18 from Experiment 1, with two newly created ones) as well as 10 unrelated filler items of different forms. Additionally, 10 sentence frames (from Farmer et al., 2006, Experiment 3) that produce a strong expectation for a verb (*The late student was required* _____, in which the verb *required* is biased to be followed by an *infinitival complement* such as *to assist* . . .) were included on the norming questionnaire to control for any increase in noun-bias caused by exposure to the noun-biased sentence frames during the norming procedure.

The text up to and including the bias-conferring verb (*saved*, in 1) was presented for both the noun- and verb-biased sentence frames. In the less noun-biased condition, sentence frames were presented up through the word before the intended location of the target word, and the filler items were truncated at variable points throughout the sentence. The text from each sentence fragment was followed by a short blank line (_____). For each sentence frame, participants were asked to circle, on an anchored scale of 1 (*noun*) to 7 (*verb*), the grammatical category from which they thought the word in the blank space would come. The order of item presentation was randomly assigned for each of the 40 participants. Less-biased contexts elicited significantly higher (and thus less noun biased) ratings (M = 3.28, SE = .09) than the strongly noun-biased sentence frames (M = 2.18, SE = .09), $t(39) = 5.86$, $p < .0005$. The 10 verb-biased frames elicited a mean response of 6.07 (SD = .96).

Experimental stimulus set. To create items in the less-biased context, we selected eight frames from the 20 less-biased frames included in the norming study (see Appendix B, mean norming rating = 2.96, SD = .60). We note that many of our less-biased sentence frames were created by including "*the word*" before the target word appearing in quotes. For our initial investigation into whether the typicality effect was dependent upon the strength of an expectation for a noun, we chose a smaller set of items that was more variable with respect to the way that the word in quotes was introduced in the sentential context (see Appendix B). For items in the strongly noun-biased context condition, we selected 16 out of the 18 items from Experiment 1, randomly excluding two of the original items in order to create an equal number of observations per cell of the design (mean norming rating = 2.01, SE = .08).

From this set of items, we constructed four presentation lists in the following manner. Lists 1 and 2 were created by randomly selecting eight of the strongly noun-biased items, and counterbalancing these across Lists 1 and 2 such that each participant only saw one version of each item, but an equal number of trials containing noun- and verb-like nouns. The remaining eight noun- and verb-like noun pairings from the other eight Experiment 1 items were placed in the less-biased contexts. Word-pairing (*insect* and *marble*, from Example 1) was maintained, such that both words from the pair appeared in the same less-biased context sentence, resulting in 16 sentences, half of which contained a noun-like noun and half of which contained a verb-like noun. The 16 words appearing in less-biased contexts were counterbalanced across each list such that each participant only saw one version of each item (or, one of the two words in each pairing), but an equal number of trials containing words from each typicality condition. Lists 3 and 4 were constructed by rejoining the eight noun- and verb-like word-pairings that appeared in the less-biased contexts on Lists 1 and 2 with their predictive contexts that were used in Experiment 1, and by removing the word-pairings that had appeared in their predictive contexts in Lists 1 and 2, placing them instead in the same less-biased contexts that were also used in Lists 1 and 2. Lists 3 and 4 were counterbalanced across each list as described for Lists 1 and 2.

This design was necessary because the predictive contexts are designed to accommodate, in an equally plausible manner, both the noun- and the verb-like nouns from a specific word pairing, such that other target words cannot be inserted into the predictive frames. Both words contained in each of the 16 noun- and verb-like word pairings appeared in each context. Each presentation list contained 64 unrelated filler items. Eight of these filler items contained less-biased contexts, taken from our norming experiment, that were never paired with target words on which typicality was manipulated, but contained category-unambiguous verbs. These items were included to eliminate a reader's ability to learn to anticipate a noun upon perceiving quotations marks in the sentence. The additional 54 filler items were the same filler items used in Experiment 1.

Procedure. The procedure was identical to that of Experiment 1.

Results and Discussion

The data were screened as described in Experiment 1, with 16.5% of trials removed for blinks or track loss. Fixations less than 80 ms in duration and less than one character away from an adjacent fixation were incorporated into the nearest fixation, and fixations less than 80 ms that were more than one character away from an adjacent fixation were removed. Less than 0.25% of the remaining trials were excluded for fixation durations longer than 1,000 ms. The analytic strategy utilized in Experiment 1 was adopted here, with word length, log-frequency, and the continuous typicality variable entered as fixed effects. Context (strongly or less noun-biased) was also entered as a fixed effect as a continuous predictor using the scores from the norming experiment. The crucial interaction between context and typicality was also included. All variables in these models were continuous, and thus, were centered around their mean.

Results from a logistic mixed effects model on comprehension question accuracy revealed no main effects of typicality or context on accuracy rates, although a significant Context × Typicality interaction occurred ($z = 2.55$, $p = .01$). This interaction is driven by higher accuracy rates for noun- versus verb-like nouns in the strongly noun-biased context, with no corresponding typicality effect in the less-biased context (see Figure 2). Table 4 provides a summary of the results for the analyses of the first-pass measures. Significant effects of word length appeared on gaze duration and go-past times, with longer words eliciting longer fixation durations. A significant effect of log-frequency occurred on each of the fixation time measures (marginal for first fixation duration), with higher frequency words eliciting shorter fixations.

Unlike in Experiment 1, a significant effect of typicality occurred on skip rate, such that category typical words were skipped more than category atypical words. (We note here, however, that this effect does not replicate in Experiment 3.) Consistent with Experiment 1, a significant effect of typicality was present on both the first fixation duration ($\beta = 16.17$ ms) and gaze duration ($\beta = 16.54$ ms) measures, and was marginally significant for go-past times ($\beta = 16.47$ ms), with noun-like nouns eliciting shorter fixations than verb-like nouns. A significant main effect of context occurred on gaze duration ($\beta = 23.53$ ms) and go-past time ($\beta = 32.91$ ms), with fixations lasting longer on target words as the noun-bias of the frame decreases. Readers were also significantly less likely to skip a target word in the less-biased context. These effects of context are consistent with work on lexical predictability manipulations demonstrating that when context is less constraining for a specific word, the target word is less likely to be skipped and it elicits longer fixation durations (Ehrlich & Rayner, 1981; Kliegl et al., 2004). We note here, however, that it could be argued that the presence of quotation marks around the target words might also account for this pattern of skipping and the effect of context on the fixation measures. We return to the issue of quotation marks at the end of this section. Crucially, context and typicality were found to significantly interact on gaze duration ($\beta = -23.70$ ms) as depicted in Figure 3. The typicality effect was robust when context conferred a strong expectation for a noun. As contextual bias for a noun decreased, however, gaze duration increased for noun-like nouns, such that the overall typicality effect decreased as context became less noun-biasing. No corresponding reliable Context × Typicality interaction was observed on the first fixation duration or go-past time measures.

As in Experiment 1, we ran separate models to assess the effects of each control variable by predicting both first fixation duration and gaze duration from the predictors in the model described above (including length and log-frequency), in addition to each individual centered control variable. Across each model for each control variable, typicality always remained a significant predictor, and no significant effect of a control variable was ever elicited. The interaction between typicality and context on the gaze duration measure also remained significant in the models containing nearest neighbors ($\beta = -22.94$ ms, $t = -1.95$, $p = .05$), concreteness ($\beta = -23.77$ ms, $t = -2.03$, $p = .04$), and was marginally significant in the model that contained age of acquisition as an additional covariate ($\beta = -20.50$ ms, $t = -1.73$, $p = .08$). We note however that the addition of AoA to the model did not result in a significant increase in model fit, $\chi^2(1) = 2.69$, $p = .10$.

The interaction between the strength of an expectation for a noun and word-form typicality on gaze duration in Experiment 2 provides evidence that the word-form typicality effect reported on early measures in Experiment 1 is dependent upon the strength of an expectation for a word from a given category. The effect of quotation marks around the target word in the *less*-biased contexts, however, remains unclear. The presence of quotation marks likely served as a cue to readers that, in principle, any word could occur in between the quotes, thus weakening expectations for nouns, along with the coherence of the form-based expectations that could be generated for words from the noun category. Only a few studies address how punctuation affects patterns of eye movements during reading (e.g., Hirotani, Frazier, & Rayner, 2006; Luo, Yan, & Zhou, 2013), and none specifically manipulated the presence of quotation marks. Three observations suggest, however, that the

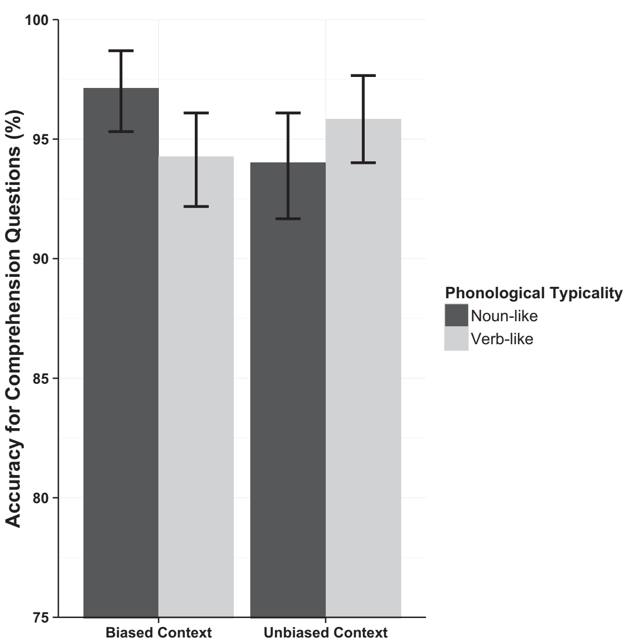

Figure 2. Accuracy rates (±*SE*) on comprehension questions, by condition, for Experiment 2.

Table 4
Regression Coefficients and Test Statistics From Linear Mixed-Effects and Logistic Mixed-Effects Models for First-Pass Eye Movement Measures on the Target Noun in Experiment 2

	Skip rate			First-fixation time			Gaze duration			Go-past time		
	Est. β	*z*-value	*p*	Est. β	*t*-value	*p*	Est. β	*t*-value	*p*	Est. β	*t*-value	*p*
Intercept	−3.58	−11.64	<.001	234.71	52.89	<.001	292.28	42.04	<.001	318.23	39.13	<.001
Word length	−0.27	−1.56	0.119	3.63	1.4	0.167	13.4	3.22	0.002	15.03	3.19	0.002
Frequency (logHal)	0.08	0.68	0.494	−4.16	−1.95	0.053	−10.62	−3.14	0.002	−10.51	−2.8	0.006
Typicality	−0.82	−2.19	0.029	16.17	2.81	0.005	16.54	1.97	0.049	16.47	1.79	0.073
Context	−1.13	−4.61	<.001	−2.11	−0.70	0.483	23.53	4.27	<.001	32.91	5.44	<.001
Context * Typicality	−0.38	−0.75	0.456	−9.69	−1.30	0.193	−23.70	−2.03	0.043	−18.88	−1.48	0.140

interaction at gaze duration was not artificially produced by the presence of the quotation marks. First, without respect to context, the results of Experiments 1 and 2 yielded similar effects of typicality in the predicted direction on first fixation duration, gaze duration, and go-past times (marginal on go-past times in Experiment 2). Second, the effect of frequency on gaze duration in Experiment 1 replicated in this experiment and again in Experiment 3, and the magnitude of the effect was similar in all three experiments. Although the consistency of frequency effects must be interpreted with caution given that frequency was included as a covariate and was not included in the random effects structure of our models, the reliability of frequency effects across experiments suggests that the presence of the quotation marks did not fundamentally alter reader's sensitivity to subtle differences in the frequency control variable at gaze duration. And third, with respect to fixation times, we note that the Context × Typicality interaction was significant for gaze duration but not for skip rate, first fixation duration, or go-past times, indicating that the interaction observed on gaze duration was not artificially produced by the presence of quotation marks (in which case one would expect to see the interaction appear elsewhere, and especially on go-past times).

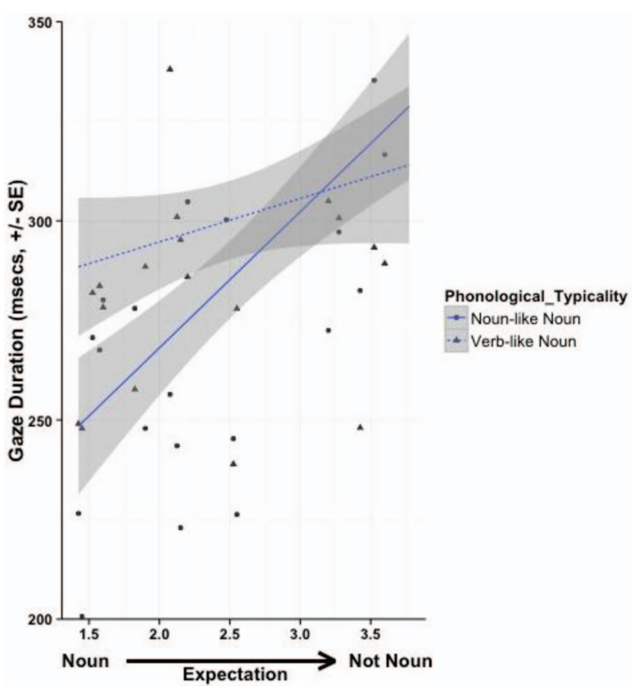

Figure 3. Gaze durations for noun- and verb-like nouns plotted against the degree of contextual constraint (represented by norming value) in Experiment 2. Values close to 1 are associated with items that conferred the strongest expectations for a noun, with sentence frames becoming progressively less noun-biased as the contextual value increases. Gray shaded regions represent 95% confidence intervals on the slopes. See the online article for the color version of this figure.

Finally, we note that in the strongly biased contexts, a main verb and determiner occur before the target word, but items in the less-biased contexts have varying pretarget word formats. This raises the possibility that these differences could have affected the number of character spaces to the left of the target word from which the saccade to the target word was launched. Systematic differences in launch sites could then in principle contribute to the Context × Typicality interaction at gaze duration. Therefore, we conducted another linear mixed effects model predicting gaze duration from context, typicality, and their interaction (as before), and also including launch site (of the initial saccade to the target word) and its interaction with typicality. As in our previous control test models, we also included length and frequency as covariates. If launch site's interaction with typicality was driving our finding of an interaction of context and typicality, we should see a robust interaction of launch site and typicality in this model, and additionally, evidence for the interaction of context and typicality should be far less clear. Instead, we find no evidence for an interaction of launch site and typicality ($\beta = -1.3$ ms, $t = -0.6$, $p = .6$) and find that the coefficient estimate and t-value for the interaction of context and typicality is barely changed ($\beta = -23$ ms, $t = -1.96$, $p = .05$). This pattern of results strongly suggests that the typicality effect, and its interaction with context, did not arise as a result of differing pretarget word sentence formats that may have caused differences in the location of a launch site to the target word.

The goal of this experiment was to determine whether or not the strength of the typicality effect illuminated in Experiment 1 was dependent upon the strength of an expectation for a noun. The items were selected in order to produce an asymmetry with respect to category expectancy, such that we chose sentence frames possessing biases that were, for the most part, bimodally distributed (although there was some overlap in the middle of the continuum). Additionally, we utilized only a handful of our original norming frames in order to reduce overlap in the manner in which the word appearing in quotes was introduced. These choices produced a design in which each reader contributed observations to 16 different items, but contributed relatively few observations ($n = 4$) to each cell of the design. We compensated for the reduced power by testing a large number of participants. In Experiment 3, we relax the distribution associated with contextual biases of the sentence frames in order to sample broadly across the spectrum of contextual bias. This allows us to determine whether the Context × Typicality interaction on gaze duration replicates in a design in which power is increased because each participant now contributes 10 observations to each condition.

Experiment 3

Noun- and verb-like nouns appeared in strongly and less noun-biased contexts. We included all of the less-biased contexts for which we gathered norming data in relation to Experiment 2 and employed a different experimental design, described below, that substantially increased the number of observations contributed by each participant. We note here that this context variable is less bimodally distributed, such that more overlap existed in norming values elicited by strong- and less-biased sentence frames. As a result, although we use the terms "strongly noun-biased" (without quotation marks) and "less noun-biased" (with quotation marks) to maintain consistency with terminology utilized in Experiment 2, we sample here from a continuum of category-predictability, such that some items with quotes exhibit more noun-bias than others, and similarly for the items without quotes. The goal is the same as that of Experiment 2, namely to determine whether context and typicality interact in first-pass measures, but now in an experiment designed to allow participants to contribute a larger number of observations to each typicality condition and to sentences both with and without quotation marks around the target word.

Method

Participants. Forty-four native English-speaking undergraduates from the University of Iowa participated in this experiment. All participants reported normal or corrected-to-normal vision and received research credit for their participation in an undergraduate psychology subject pool.

Materials. We utilized the 18 strongly noun-biased sentential contexts from Experiments 1 and 2, along with the two newly created items included on the norming study described in Experiment 2. Additionally, we utilized all 20 less noun-biased sentence frames from the norming study described in Experiment 2.

The 40 sentences from the 20 experimental items in the strongly noun-biased sentence frames were counterbalanced across two presentation lists such that each participant only saw one version of each item, but an equal number of trials per condition. The same noun- and verb-like noun pairings per item (i.e., *marble* vs. *insect*) were randomly assigned to one less noun-biased sentence frame (as in example 2 above). The less-biased sentences containing the opposite word in the noun- and verb-like word pairing that occurred in the strongly noun-biased context on each list was then added to each list. For example, participants assigned to List 1 read the word *marble* in the strongly noun-biased context, but the word *insect* in the less noun-biased context, and vice versa for participants assigned to List 2. As a result, each participant read a total of 40 experimental sentences (20 noun-like and 20 verb-like nouns, half appearing with quotes and half without), thus affording each participant the opportunity to read both target words in each noun- and verb-like word-pairing. Each list also contained two practice items that participants read first, along with the 54 unrelated filler items utilized in Experiment 1. Additionally, each list contained 10 sentences appearing in contexts designed to confer a less-strong noun bias (with quotes), but with unambiguous verbs appearing within the quotes, in order to reduce a reader's ability to learn to anticipate a noun upon perceiving quotations marks in the sentence (see Table 5).

Procedure. The procedure was identical to that of Experiments 1 and 2.

Results and Discussion

The data were screened as described in Experiments 1 and 2, with 3.8% of trials removed for blinks or track loss or for fixation durations longer than 1,000 ms. Fixations less than 80 ms in duration and less than one character away from an adjacent fixation were incorporated into the nearest fixation, and fixations less than 80 ms that were more than one character away from an adjacent fixation were removed. We utilized the same analytic strategy described in Experiment 2.

A significant effect of length occurred on skip rate, with longer words being skipped less frequently, and on go-past times as in Experiment 2. Significant effects of log frequency occurred on each fixation time measure, but not on skip rate. As in Experiment 2, significant effects of context occurred on each measure except first fixation duration, such that fixation times decrease and words are more likely to be skipped as context becomes more constrain-

Table 5
Regression Coefficients and Test Statistics From Linear Mixed-Effects and Logistic Mixed-Effects Models for First-Pass Eye Movement Measures on the Target Noun in Experiment 3

	Skip rate			First-fixation time			Gaze duration			Go-past time		
	Est. β	*z*-value	*p*	Est. β	*t*-value	*p*	Est. β	*t*-value	*p*	Est. β	*t*-value	*p*
Intercept	−3.95	−12.24	<.001	240.47	51.92	<.001	302.52	37.93	<.001	320.68	34.96	<.001
Word length	−0.41	−2.54	0.011	0.52	0.23	0.822	6.36	1.59	0.116	9.62	2.42	0.017
Frequency (logHal)	0.14	1.12	0.261	−5.31	−2.66	0.009	−10.49	−3.00	0.003	−10.01	−2.91	0.004
Typicality	−0.25	−0.69	0.489	14.76	2.46	0.018	17.84	2.02	0.043	25.82	3.03	0.003
Context	−0.86	−3.15	0.002	−3.02	−0.81	0.420	18.39	3.76	<.001	17.76	2.68	0.011
Context * Typicality	0.59	1.28	0.201	−1.85	−0.28	0.776	−22.56	−2.02	0.044	−16.55	−1.53	0.127

ing for a noun. No significant effect of typicality occurred on skip rate, and verb-like nouns were fixated significantly longer than noun-like nouns on first fixation duration ($\beta = 14.76$ ms), gaze duration ($\beta = 17.84$ ms), and go-past times ($\beta = 25.82$ ms). Crucially, the interaction between context and typicality replicated on gaze duration ($\beta = -22.56$ ms), and the nature of the interaction (depicted in Figure 4) is qualitatively similar to the one observed in Experiment 2 (see Figure 3): the typicality effect decreases as sentential context becomes less predictive of a noun.

We tested for confounding effects of each of our control variables on first fixation and gaze duration measures as described in the previous experiments. The typicality effect occurred in each model that we conducted. Number of neighbors, concreteness, and AoA never exerted a significant independent effect on any fixation time measure. The Context × Typicality interaction remained significant at gaze duration for the control models including AoA ($\beta = -21.55$ ms, $t = -2.16$, $p = .031$), number of neighbors ($\beta = -20.86$ ms, $t = -2.08$, $p = .038$), and concreteness ($\beta = -21.67$ ms, $t = -2.16$, $p = .031$). Additionally, as for Experiment 2, in order to address the effect of systematic differences in pretarget sentence format across the sentences with and without quotation marks around the target word, we added launch site and its interaction with typicality to our mixed effects model. Similarly to the results for Experiment 2, we find no evidence for an interaction of launch site and typicality ($\beta = -1.0$ ms, $t = -0.4$, $p = .7$) and find that the coefficient estimate and t-value for the interaction of context and typicality is barely changed ($\beta = -22$ ms, $t = -1.86$, $p = .06$). This pattern of results again strongly suggests that the interaction of context and typicality is not mediated through launch site.

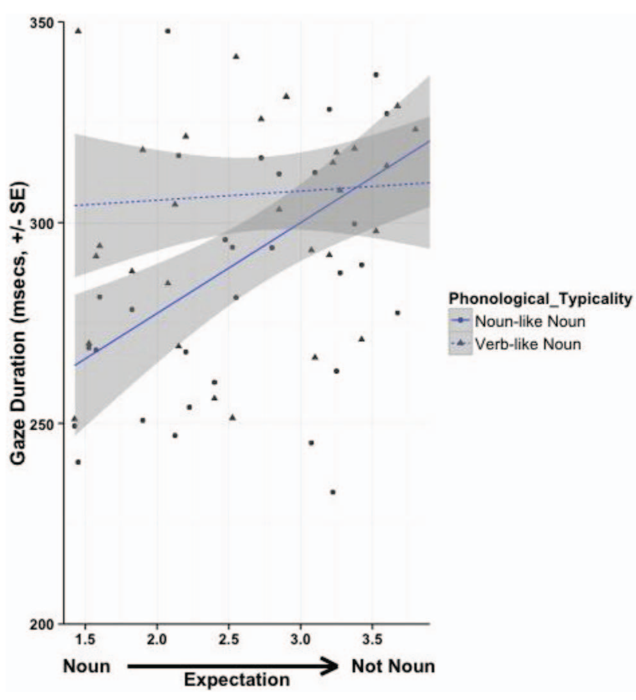

Figure 4. Gaze durations for noun- and verb-like nouns plotted against the degree of contextual constraint (represented by norming value) in Experiment 3. Values close to 1 are associated with items that conferred the strongest expectations for a noun, with sentence frames becoming progressively less noun-biased as the contextual value increases. Gray shaded regions represent 95% confidence intervals on the slopes. See the online article for the color version of this figure.

General Discussion

The first goal of the experiments presented here was to determine whether phonological typicality significantly influenced first-pass eye movement measures. Across three experiments, an effect of phonological typicality was reliably observed on first-fixation duration, gaze duration, and go-past times when sentential context was strongly predictive of a noun, but not predictive of the identity of the specific word-form. Our second goal was to explore the effect of category predictability. The results of Experiments 2 and 3 further demonstrate that the strength of the typicality effect on nouns increases as context becomes progressively more constraining for a word from the noun category. More specifically, gaze durations on noun-like nouns exhibited a strong decrease as context became more category-predictive, with the effect of context appearing less strong on verb-like nouns. Taken together, these results suggest that processing is especially facilitated when strong expectations for a word from a specific lexical category exist, and the form of the input is consistent with contextually conferred form-based estimates. These results provide a more direct test of the claim that word-form typicality effects are dependent upon the strength of category expectancy than was originally provided by Farmer et al. (2011). We note, however, that this interpretation of our results is dependent on the degree to which the presence of quotation marks around the target words facilitated less strong noun expectations while not interfering with patterns of eye-movements during reading. As noted in the discussion of the results of Experiment 2, evidence from multiple data points, all of which replicated in Experiment 3, suggest that the Context × Typicality interaction at gaze duration was not artificially produced by the presence of quotation marks. That said, further work on the effects of quotation marks around target words on the eye movement record will be necessary in order to rule out the possibility that their presence alone produced the interaction instead of the bias values, gauged by norming data, associated with noun expectation.

Without respect to context, the typicality effect on the first-pass fixation measures is consistent with the effects of form-based and lexical level variables on first-pass measures, discussed throughout the article. It is also is consistent with EEG/ERP work demonstrating that form-based and lexical variables exert an influence shortly after sensory transduction occurs (~50 ms–80 ms in the visual modality, Nowak, Munk et al., 1995). Hauk, Davis, Ford, Pulvermuller, and Marslen-Wilson (2006) demonstrated, for example, that EEG responses to individually presented words reflect sensitivity to manipulations of form-related variables, such as word length, at ~90 ms–100 ms poststimulus onset, to lexical frequency at ~110 ms–160ms (see also Assadollhi & Pulvermuller, 2003; Hauk & Pulvermuller, 2004; Sereno, Rayner, & Posner, 1998), and to lexical status at ~150 ms.

With respect to context, even before a target word is presented, pretarget word patterns of EEG activity, in the form of increased theta-band activity, occur when a specific word can be predicted form context, relative to a condition in which context does not support an expectation for a specific word (Molinaro et al., 2013; see also Dikker et al., 2013). The degree to which a sentential context is strongly predictive of a specific lexical item influences the processing of a newly encountered visual word at ~100 ms–150 ms poststimulus onset in EEG/MEG experiments, suggesting sensitivity to unexpected word-form information in this time window. Moreover, manipulations of lexical predictability interact with manipulations of form-based and lexical-level variables, such as length and frequency, during the same time window (Federmeier & Kutas, 2001; Kim & Gilley, 2013; Lee, Liu, & Tsai, 2012; Molinaro, Barraza, & Carreiras, 2013; Penolazzi, Hauk, & Pulvermuller, 2007), suggestive of interactions between higher-level expectancies and form-based properties of incoming input, data with which the Context × Typicality interactions reported here are consistent.

We note that there is a lack of consensus on (a) the theoretical status of the prelexical versus postlexical distinction, and (b) whether or not first-pass eye movement measures such as first fixation and gaze duration truly reflect early as opposed to late processes. Therefore, the degree to which the Context × Typicality interaction at gaze duration occurs during prelexical or postlexical processing remains unclear.

In a recent fMRI experiment, Boylan, Trueswell, and Thompson-Schill (2014) demonstrated that patterns of activation in the left mid fusiform gyrus (sometimes referred to as the "visual word form area") in the ventral visual stream can be used to correctly classify whether participants had viewed a noun- versus a verb-predictive sentential context, even before they had a chance to see the target word. Pattern-classifier analyses were statistically robust at the vWFA, although Boylan et al. also found that patterns of activation in early visual cortex did not afford significant classification of category-predictive contexts. Although these results provide evidence for the existence of form-based expectations, at least in the ventral visual stream, they do not specifically address how early visual input is assessed with respect to those expectations after sensory transduction. Indeed, the strongest version of a form-based expectations account of the effects reported here would argue that form should be assessed against form-based expectations extremely early. Tanenhaus and Hare (2007), for example, raised the possibility that "first fixations might be influenced only by expectations that can be translated into form-based estimates of the information that a word is likely to contain. A word that is likely to be informative (i.e., unexpected) would be allotted more fixation time" (p. 94). Given that the reliable interactions between expectation strength and typicality only arose on gaze duration, the present experiments presented no clear evidence in favor of this hypothesis. Additionally, the inability of a pattern classifier to significantly discriminate noun- versus verb-predictive sentential contexts in early visual cortex reported by Boylan et al. suggests that form-based expectations influence the processing of visual information during reading only after some bottom-up processing of it has occurred.

We note that in the spoken language processing literature, there are clear effects of expectations on signal processing (Brown, 2014; Dahan & Tanenhaus, 2004; Magnuson, Tanenhaus, & Aslin, 2008; Revill, Tanenhaus, & Aslin, 2008). One hypothesis for why these effects might be stronger in spoken language processing than in natural reading is that the input arrives more slowly with speech. Spoken words unfold over time. Also, more words can be read than can be spoken in any given unit of time. Another alternative hypothesis is that the spoken language studies, particularly those using the visual world paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995) provide unusually constraining closed-world contexts. From our perspective, however, the reason may be that higher-level variables are reflected more directly in the signal in spoken language than in text (e.g., syllable duration is a strong cue to an upcoming word boundary; Salverda, Dahan & McQueen, 2003). But, discourse-based factors such as information structure can also affect syllable duration. Indeed, when the discourse context provides an alternative explanation for syllable duration, listeners down-weight duration as a cue to word recognition (Brown, Salverda, Gunlogson, & Tanenhaus, 2015). During natural reading, the category-indicative form-based properties of a word may not get assessed against word-form estimates until slightly later in the processing of the visual signal than in speech, only after some initial visual processing of a newly fixated word has occurred.

The results reported here are consistent with computational-level accounts of eye movements in reading that place a strong emphasis on predictions over distributions of words (e.g., Hale, 2001; Jurafsky, 1996; Levy, 2008). Smith and Levy (2010), for example, argue against the assumption that a word must be identified 100% accurately before higher-language processing can proceed, proposing instead that the processing system is tolerant of, and sensitive to, uncertainty about word identity. They created an average neighborhood surprisal (ANS) metric associated with an individual word that averages that word's contextual surprisal value together with the surprisals of its orthographic neighbors. This ANS value was significantly more predictive of first fixation times and the duration of the second of multiple initial fixations than raw surprisal values computed based only on lexical predictability, suggesting that the time allotted to these fixations is linked to the degree of visual similarity a target word has to other contextually permissible words.

In Experiments 2 and 3, lexical predictability was low, although category predictability varied from high to low. Under the Smith and Levy account, in highly noun predictive contexts, all possible nouns, and thus their associated form features, would be more contextually supported, producing increased processing difficulty upon encountering words that are visually inconsistent with nouns, such as verb-like nouns. As the distribution widens due to contexts that afford less category predictability, readers' estimates of visual similarity that are conditioned upon context become more noisy, reducing the advantage associated with noun-like nouns observed in the strongly predictive contexts. Although Smith and Levy made no appeal to form-based expectations, instead characterizing their results in the context of visual uncertainty given contextual predictability, we suggest that in many respects, we have provided more evidence in support of the predictive utility of form-based metrics such as ANS, as a contextualized index of form-based expectations, in predicting fixation durations on first-pass measures.

Conclusion

Accumulating evidence from multiple paradigms, manipulations, and methodologies suggests that higher-level expectancies—such as an expectation for a word from a specific grammatical category—can be translated into low-level form-based estimates of input that a sensory system is likely to encounter, and that those form-based estimates serve as something of a template against which to evaluate the incoming signal. Our experiments were not intended to test specific mechanistic claims about the processes that afford the generation of expectations from context, the processes by which they are translated into form-based estimates, and the degree to which the resultant form-based expectations are shaped by increasing patterns of activation as opposed to lateral inhibitory processes, though these are important questions for future research. We also did not test specific predictions of frameworks that make a strong theoretical distinction between pre and postlexical processes. Instead, we note that a number of recent theoretical models have been proposed to address the coupling of knowledge-based higher-level expectancies and perceptual-motor processes across multiple domains (e.g., Clark, 2013; Dell, 2014; Pickering & Clark, 2014; Pickering & Garrod, 2013; see also Kutas, DeLong, & Smith, 2011 for an early discussion of related points). We view our results as consistent with *data explanation* approaches to the analysis of incoming linguistic input. Clark (2013), for example, argued that a stream of hierarchically organized *generative models* propagates higher-level (more abstract) expectations to lower-level (progressively closer to perceptual cortex) models via feed-backward connections. At each level, as the incoming signal is intercepted, mismatch between input and expectation produces prediction error. These "error signals" contain information about the difference between what was predicted and the structure of the input contained in the arriving signal. The error signal feeds forward to higher levels of representation, potentially facilitating an adjustment of higher-level expectations such that the predictions generated at some given level of processing may be more precise in the future (e.g., Chang, Dell, & Bock, 2006; Fine, Jaeger, Farmer, & Qian, 2013; Jaeger & Snider, 2013; see Farmer, Brown, & Tanenhaus, 2013, for a more detailed discussion of the ramifications of Clark, 2013 for online language comprehension). We suggest that hierarchical predictive processing frameworks of online language comprehension will help elucidate how prior knowledge and top-down contextual information modulate the perception and interpretation of a physical signal during lower-level, and even sensory-based, processing of a linguistic signal.

Most generally, we think that hierarchical predictive processing frameworks offer the potential to unify prediction-based accounts of contextualized language processing by guiding work on questions related to what's being predicted, and when in the chain of processing those predictions are generated and assessed.

References

Altarriba, J., Kroll, J. F., Sholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory & Cognition, 24,* 477–492. http://dx.doi.org/10.3758/BF03200936

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73,* 247–264. http://dx.doi.org/10.1016/S0010-0277(99)00059-1

Arai, M., & Keller, F. (2013). The use of verb-specific information for prediction in sentence processing. *Language and Cognitive Processes, 28,* 525–560. http://dx.doi.org/10.1080/01690965.2012.658072

Arciuli, J., & Monaghan, P. (2009). Probabilistic cues to grammatical category in English orthography and their influence during reading. *Scientific Studies of Reading, 13,* 73–93. http://dx.doi.org/10.1080/10888430802633508

Ashby, J., & Rayner, K. (2004). Representing syllable information during silent reading: Evidence from eye movements. *Language and Cognitive Processes, 19,* 391–426. http://dx.doi.org/10.1080/01690960344000233

Assadollahi, R., & Pulvermuller, F. (2003). Early influences of word length and frequency: A group study using MEG. *NeuroReport, 14,* 1183–1187. http://dx.doi.org/10.1097/00001756-200306110-00016

Baayen, H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX database on CD-ROM*. Philadelphia, PA: Linguistic Data Consortium.

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology, 17,* 364–390. http://dx.doi.org/10.1016/0010-0285(85)90013-1

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes.

Bicknell, K., Elman, J. L., Hare, M., McRae, K., & Kutas, M. (2010). Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language, 63,* 489–505. http://dx.doi.org/10.1016/j.jml.2010.08.004

Boylan, C., Trueswell, J. C., & Thompson-Schill, S. L. (2014). Multi-voxel pattern analysis of noun and verb differences in ventral temporal cortex. *Brain and Language, 137,* 40–49. http://dx.doi.org/10.1016/j.bandl.2014.07.009

Brown, M. (2014). *Interpreting prosodic variation in context* (Doctoral dissertation). University of Rochester, Rochester, NY.

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2011). Expectations from preceding prosody influence segmentation in online sentence processing. *Psychonomic Bulletin & Review, 18,* 1189–1196. http://dx.doi.org/10.3758/s13423-011-0167-9

Brown, M., Salverda, A. P., Gunlogson, C., & Tanenhaus, M. K. (2015). Interpreting prosodic cues in context. *Language. Cognition and Neuroscience, 30,* 149–166. http://dx.doi.org/10.1080/01690965.2013.862285

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46,* 904–911.

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. http://dx.doi.org/10.3758/BF03200655

Carreiras, M., Armstrong, B. C., Perea, M., & Frost, R. (2014). The what, when, where, and how of visual word recognition. *Trends in Cognitive Sciences, 18,* 90–98. http://dx.doi.org/10.1016/j.tics.2013.11.005

Cassidy, K. W., & Kelly, M. H. (2001). Children's use of phonology to infer grammatical class in vocabulary learning. *Psychonomic Bulletin & Review, 8,* 519–523. http://dx.doi.org/10.3758/BF03196187

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113,* 234–272. http://dx.doi.org/10.1037/0033-295X.113.2.234

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36,* 181–204. http://dx.doi.org/10.1017/S0140525X12000477

Connine, C., Ferreira, F., Jones, C., Clifton, C., & Frazier, L. (1984). Verb frame preferences: Descriptive norms. *Journal of Psycholinguistic Research, 13,* 307–319. http://dx.doi.org/10.1007/BF01076840

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6,* 84–107. http://dx.doi.org/10.1016/0010-0285(74)90005-X

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Evidence from immediate effects of verb-based constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 498–513. http://dx.doi.org/10.1037/0278-7393.30.2.498

Dell, G. S. (2014). Phonemes and production. *Language, Cognition and Neuroscience, 29,* 30–32. http://dx.doi.org/10.1080/01690965.2013.851795

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8,* 1117–1121. http://dx.doi.org/10.1038/nn1504

Dikker, S., & Pylkkänen, L. (2013). Predicting language: MEG evidence for lexical preactivation. *Brain and Language, 127,* 55–64. http://dx.doi.org/10.1016/j.bandl.2012.08.004

Dikker, S., Rabagliati, H., Farmer, T. A., & Pylkkänen, L. (2010). Early occipital sensitivity to syntactic category is based on form typicality. *Psychological Science, 21,* 629–634. http://dx.doi.org/10.1177/0956797610367751

Dikker, S., Rabagliati, H., & Pylkkänen, L. (2009). Sensitivity to syntax in visual cortex. *Cognition, 110,* 293–321. http://dx.doi.org/10.1016/j.cognition.2008.09.008

Drieghe, D., Brysbaert, M., Desmet, T., & De Baecke, C. (2004). Word skipping in reading: On the interplay of linguistic and visual factors. *European Journal of Cognitive Psychology, 16,* 79–103. http://dx.doi.org/10.1080/09541440340000141

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior, 20,* 641–655. http://dx.doi.org/10.1016/S0022-5371(81)90220-6

Farmer, T. A., Brown, M., & Tanenhaus, M. K. (2013). Prediction, explanation, and the role of generative models in language processing. *Behavioral and Brain Sciences, 36,* 211–212. http://dx.doi.org/10.1017/S0140525X12002312

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. http://dx.doi.org/10.1073/pnas.0602173103

Farmer, T. A., Monaghan, P., Misyak, J. B., & Christiansen, M. H. (2011). Phonological typicality influences sentence processing in predictive contexts: Reply to Staub, Grant, Clifton, and Rayner (2009). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1318–1325. http://dx.doi.org/10.1037/a0023063

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44,* 491–505. http://dx.doi.org/10.1111/j.1469-8986.2007.00531.x

Federmeier, K. D., & Kutas, M. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 202–224. http://dx.doi.org/10.1037/0278-7393.27.1.202

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE, 8,* e77661. http://dx.doi.org/10.1371/journal.pone.0077661

Fine, A. B., Qian, T., Jaeger, T. F., & Jacobs, R. A. (2010). Syntactic adaptation in language comprehension. *Proceedings of the 2010 Workshop on Cognitive Modeling and Computational Linguistics* (pp. 18–26). Uppsala, Sweden: Association for Computational Linguistics.

Fitneva, S. A., Christiansen, M. H., & Monaghan, P. (2009). From sound to syntax: Phonological constraints on children's lexical categorization of new words. *Journal of Child Language, 36,* 967–997. http://dx.doi.org/10.1017/S0305000908009252

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language, 37,* 58–93. http://dx.doi.org/10.1006/jmla.1997.2512

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies,* 1–8.

Hand, C. J., Miellet, S., O'Donnell, P. J., & Sereno, S. C. (2010). The frequency-predictability interaction in reading: It depends where you're coming from. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 1294–1313. http://dx.doi.org/10.1037/a0020363

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review, 106,* 491–528. http://dx.doi.org/10.1037/0033-295X.106.3.491

Hauk, O., Davis, M. H., Ford, M., Pulvermuller, F., & Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *NeuroImage, 30,* 1383–1400. http://dx.doi.org/10.1016/j.neuroimage.2005.11.048

Hauk, O., & Pulvermuller, F. (2004). Neurophysiological distinction of action words in the fronto-central cortex. *Human Brain Mapping, 21,* 191–201. http://dx.doi.org/10.1002/hbm.10157

Hirotani, M., Frazier, L., & Rayner, K. (2006). Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language, 54,* 425–443. http://dx.doi.org/10.1016/j.jml.2005.12.001

Inhoff, A. W., Eiter, B. M., Radach, R., & Juhasz, B. (2003). Distinct subsystems for the parafoveal processing of spatial and linguistic information during eye fixations in reading. *The Quarterly Journal of Experimental Psychology Section A, 56,* 803–827. http://dx.doi.org/10.1080/02724980244000639

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics, 40,* 431–439. http://dx.doi.org/10.3758/BF03208203

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford, England: Oxford University Press.

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition, 127,* 57–83. http://dx.doi.org/10.1016/j.cognition.2012.10.013

Juhasz, B. J., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 1312–1318. http://dx.doi.org/10.1037/0278-7393.29.6.1312

Juhasz, B. J., White, S. J., Liversedge, S. P., & Rayner, K. (2008). Eye movements and the use of parafoveal word length information in reading. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1560–1579. http://dx.doi.org/10.1037/a0012319

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science, 20,* 137–194. http://dx.doi.org/10.1207/s15516709cog2002_1

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass, 2,* 647–670. http://dx.doi.org/10.1111/j.1749-818X.2008.00072.x

Kamide, Y. (2012). Learning individual talkers' structural preferences. *Cognition, 124,* 66–71. http://dx.doi.org/10.1016/j.cognition.2012.03.001

Kamide, Y., Altmann, G., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49,* 133–156. http://dx.doi.org/10.1016/S0749-596X(03)00023-8

Kaschak, M. P., & Glenberg, A. M. (2004). This construction needs learned. *Journal of Experimental Psychology: General, 133,* 450–467.

Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review, 99,* 349–364. http://dx.doi.org/10.1037/0033-295X.99.2.349

Kelly, M. H. (1998). Rule- and idiosyncratically derived denominal verbs: Effects on language production and comprehension. *Memory & Cognition, 26,* 369–381. http://dx.doi.org/10.3758/BF03201147

Kim, A. E., & Gilley, P. M. (2013). Neural mechanisms of rapid sensitivity to syntactic anomaly. *Frontiers in Psychology, 4,* 45. http://dx.doi.org/10.3389/fpsyg.2013.00045

Kimball, J. (1975). Predictive analysis and over-the-top parsing. In J. Kimball (Ed.), *Syntax and semantics: Vol. 5. Predictive analysis and over-the-top parsing* (pp. 155–179). New York, NY: Academic Press.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology, 16,* 262–284. http://dx.doi.org/10.1080/09541440340000213

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44,* 978–990. http://dx.doi.org/10.3758/s13428-012-0210-4

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain* (pp. 190–207). Oxford, UK: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195395518.003.0065

Kuznetsova, A., Christensen, R. H. B., & Brockhoff, P. B. (2014). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). *R package version, 2*. 0–6.

Lee, C. Y., Liu, Y. N., & Tsai, J. L. (2012). The time course of contextual effects on visual word recognition. *Frontiers in Psychology, 3,* 285. http://dx.doi.org/10.3389/fpsyg.2012.00285

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106,* 1126–1177. http://dx.doi.org/10.1016/j.cognition.2007.05.006

Linzen, T., Marantz, A., & Pylkkanen, L. (2013). Syntactic context effects in visual word recognition: An MEG study. *The Mental Lexicon, 8,* 117–139. http://dx.doi.org/10.1075/ml.8.2.01lin

Luo, Y., Yan, M., & Zhou, X. (2013). Prosodic boundaries delay the processing of upcoming lexical information during silent sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 915–930. http://dx.doi.org/10.1037/a0029182

Magnuson, J. S., Tanenhaus, M. K., & Aslin, R. N. (2008). Immediate effects of form-class constraints on spoken word recognition. *Cognition, 108,* 866–873. http://dx.doi.org/10.1016/j.cognition.2008.06.005

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics, 17,* 578–586. http://dx.doi.org/10.3758/BF03203972

McConkie, G. W., & Rayner, K. (1976). Asymmetry of the perceptual span in reading. *Bulletin of the Psychonomic Society, 8,* 365–368. http://dx.doi.org/10.3758/BF03335168

Molinaro, N., Barraza, P., & Carreiras, M. (2013). Long-range neural synchronization supports fast and efficient reading: EEG correlates of processing expected words in sentences. *NeuroImage, 72,* 120–132. http://dx.doi.org/10.1016/j.neuroimage.2013.01.031

Monaghan, P., Chater, N., & Christiansen, M. H. (2003). Inequality between the classes: Phonological and distributional typicality as predictors of lexical processing. *Proceedings of the 25th Annual Conference of the Cognitive Science Society,* 810–815.

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorisation. *Cognition, 96,* 143–182. http://dx.doi.org/10.1016/j.cognition.2004.09.001

Monaghan, P., Christiansen, M. H., Farmer, T. A., & Fitneva, S. A. (2010). Measures of phonological typicality: Robust coherence and psychological validity. *The Mental Lexicon, 5,* 281–299.

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology, 55,* 259–305. http://dx.doi.org/10.1016/j.cogpsych.2006.12.001

Nowak, L. G., Munk, M. H. J., Girard, P., & Bullier, J. (1995). Visual latencies in areas V1 and V2 of the macaque monkey. *Visual Neuroscience, 12,* 371–384. http://dx.doi.org/10.1017/S095252380000804X

Penolazzi, B., Hauk, O., & Pulvermuller, F. (2007). Early semantic context integration and lexical access as revealed by event-related brain potentials. *Biological Psychology, 74,* 374–388. http://dx.doi.org/10.1016/j.biopsycho.2006.09.008

Pickering, M. J., & Clark, A. (2014). Getting ahead: Forward models and their place in cognitive architecture. *Trends in Cognitive Sciences, 18,* 451–456. http://dx.doi.org/10.1016/j.tics.2014.05.006

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36,* 329–347. http://dx.doi.org/10.1017/S0140525X12001495

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York, NY: Springer.

Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology, 41,* 211–236. http://dx.doi.org/10.1016/0022-0965(86)90037-8

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z reader model. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 720–732. http://dx.doi.org/10.1037/0096-1523.30.4.720

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition, 14,* 191–201. http://dx.doi.org/10.3758/BF03197692

Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 514–528. http://dx.doi.org/10.1037/a0020990

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review, 3,* 504–509. http://dx.doi.org/10.3758/BF03214555

Rayner, K., Well, A. D., & Pollatsek, A. (1980). Asymmetry of the effective visual field in reading. *Perception & Psychophysics, 27,* 537–544. http://dx.doi.org/10.3758/BF03198682

R Development Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, Vienna, Austria.

Revill, K. P., Tanenhaus, M. K., & Aslin, R. N. (2008). Context and spoken word recognition in a novel lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1207–1223. http://dx.doi.org/10.1037/a0012796

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition, 90,* 51–89. http://dx.doi.org/10.1016/S0010-0277(03)00139-2

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, & Psychophysics, 74,* 5–35. http://dx.doi.org/10.3758/s13414-011-0219-2

Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research, 32,* 3–23. http://dx.doi.org/10.1023/A:1021928914454

Sereno, S. C., Rayner, K., & Posner, M. I. (1998). Establishing a time-line of word recognition: Evidence from eye movements and event-related potentials. *Neuroreport, 9,* 2195–2200. http://dx.doi.org/10.1097/00001756-199807130-00009

Slattery, T. J., Staub, A., & Rayner, K. (2012). Saccade launch site as a predictor of fixation durations in reading: Comments on Hand, Miellet, O'Donnell, and Sereno (2010). *Journal of Experimental Psychology: Human Perception and Performance, 38,* 251–261. http://dx.doi.org/10.1037/a0025980

Smith, N. J., & Levy, R. (2010). Fixation durations in first-pass reading reflect uncertainty about word identity. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1313–1318). Austin, TX: Cognitive Science Society.

Stanovich, K. E., & West, R. F. (1979). Mechanisms of sentence context effects in reading: Automatic activation and conscious attention. *Memory and Cognition, 7,* 77–85.

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. http://dx.doi.org/10.1037/0278-7393.32.2.425

Staub, A., Grant, M., Clifton, C., Jr., & Rayner, K. (2009). Phonological typicality does not influence fixation durations in normal reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 806–814. http://dx.doi.org/10.1037/a0015123

Staub, A., Grant, M., Clifton, C., Jr., & Rayner, K. (2011). Still no phonological typicality effect on word reading time (and no good explanation of one, either): A rejoinder to Farmer, Monaghan, Misyak, and Christiansen. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1326–1328. http://dx.doi.org/10.1037/a0024193

Tanenhaus, M. K., & Hare, M. (2007). Phonological typicality and sentence processing. *Trends in Cognitive Sciences, 11,* 93–95. http://dx.doi.org/10.1016/j.tics.2006.11.010

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632–1634. http://dx.doi.org/10.1126/science.7777863

Tarkiainen, A., Helenius, P., Hansen, P. C., Cornelissen, P. L., & Salmelin, R. (1999). Dynamics of letter string perception in the human occipitotemporal cortex. *Brain: A Journal of Neurology, 122,* 2119–2132. http://dx.doi.org/10.1093/brain/122.11.2119

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 443–467.

White, S. J., Rayner, K., & Liversedge, S. P. (2005). Eye movements and the modulation of parafoveal processing by foveal processing difficulty: A reexamination. *Psychonomic Bulletin & Review, 12,* 891–896. http://dx.doi.org/10.3758/BF03196782

Williams, R., & Morris, R. (2004). Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology, 16,* 312–339. http://dx.doi.org/10.1080/09541440340000196

(*Appendices follow*)

Appendix A

Computing Phonological Typicality

Table 6
Calculation of the Distance Between Two Words in Phonemic Feature Space

Position in word	Phoneme	Phonological features	Phoneme	Phonological features	Sum of squared differences
1	—	{−1, −1, −1, −1, −1, −1, −1, −1, −1, −1, −1}	—	{−1, −1, −1, −1, −1, −1, −1, −1, −1, −1, −1}	0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0
2	k	{−1, 1, −1, −1, 1, −1, −1, −1, −1, −1, 0}	P	{−1, 1, −1, −1, 1, 1, 0, −1, 1, 0, 0}	0 + 0 + 0 + 0 + 0 + 4 + 1 + 0 + 4 + 1 + 0
3	ε	{1, −1, 1, 0, −1, −1, 0, −1, −1, −1, −1}	Iə	{1, −1, 1, 0, −0.5, −1, 0, −1, −0.5, −0.5, −1}	0 + 0 + 0 + 0 + 0.25 + 0 + 0 + 0 + 0.25 + .25 + 0
4	I	{0.5, 0, 1, 0, −1, −1, 1, −1, −1, 1, 0}	ɹ	{0.5, 0, 1, 0, −1, −1, −1, 1, 1, −1, −1}	0 + 0 + 0 + 0 + 0 + 0 + 4 + 4 + 4 + 4 + 1
5	P	{−1, 1, −1, −1, 1, 1, 0, −1, 1, 0, 0}	—	{−1, −1, −1, −1, −1, −1, −1, −1, −1, −1, −1}	0 + 4 + 0 + 0 + 4 + 4 + 1 + 0 + 4 + 1 + 1

Note. FD = Σ√(Sum of Squared Differences = 12.51. Phonemic features: sonorant, consonantal, degree, voice, nasal, labial, palatal, pharyngeal, radical, round, tongue.

The calculation of the phonological typicality (PT) metric was originally presented by Monaghan, Chater, and Christiansen (2003). Here, we report the details of its calculation as employed in the analysis of PT effects on lexical naming times and in the sentence processing experiments presented in Farmer et al. (2006). This explanation is not meant to be exhaustive, but is instead intended to provide readers with the details necessary in order to understand what aspects of a word's internal structure make it noun-like or verb-like with respect to other nouns and verbs in the English vocabulary. More precise details about the calculation of this version of the PT metric, coupled with a more detailed exploration of alternative calculations of PT, can be found in Monaghan, Christiansen, Farmer, and Fitneva (2010).

In order to calculate the PT of a word with respect to other words in its lexical category, category unambiguous nouns and verbs were identified from the CELEX database. They were then converted into slot-based representations, with three slots for onset, two for nucleus, and three for coda. Each phoneme was coded in terms of the presence, absence, or degree of 11 phonemic features listed below Table 6. Most features were binary (coded as −1 or +1) or ternary (coded as −1, 0, or +1), although the sonorant feature was continuous. We also note that the features were not weighted. This coding scheme was originally employed in Harm and Seidenberg (1999), and we refer readers to that paper for an in depth discussion of each feature and of the feature coding scheme employed here. When empty slots occurred, each of the 11 phonemic features was coded as −1 for absent. This phonemic feature-based coding format produces a representation of a word in a high-dimensional phonemic feature space.

The slot-based representations for each word were compared to the slot-based representations of every other category-unambiguous noun and verb in the dataset. For example, Table 6 depicts the comparison between the noun *kelp* and the noun *peer*. These two words are both monosyllabic, and so were aligned at the nucleus. The sum of the squared difference in phonemic features across the aligned words was then computed, as illustrated in the sixth column of Table 6. The square root of the sum of squared difference for each phoneme slot comparison was calculated, and those resulting values were summed. The summed square root value represents the distance between two words in the high-dimensional phonemic feature space inherent to this method of calculating PT. After the feature distance values were calculated between a word and all other words in the comparison vocabulary, the average distance in phonemic feature space was calculated between the target word and all other nouns, and between the target word and all other verbs, producing an estimate of the average distance in phonemic feature space between a target word and words from both grammatical categories. Finally, the average distance of a target word to a verb was subtracted from the average distance of the word to nouns. The resulting value represents where an individual word falls on the continuum of nouny- to verby-ness with respect to its constellation of phonemic features. In this sense, a noun-like noun is a word with a constellation of phonemic features distributed across its phonemes that is most similar to the average constellation of features inherent to other words in the noun category. The constellation of phonemic features distributed across verb-like nouns, on the other hand, is more similar to the average distribution of those features of word in the verb category. More negative values are associated with more noun-like phonemic features, more positive values are associated with more verb-like features, and values close to zero indicate that the constellation of phonemic features inherent to the word is neither typical nor atypical of members from either category.

(*Appendices continue*)

Appendix B

Materials

The target noun on which Typicality was manipulated appears in **bold**. The first sentence in each pair contains the noun-like noun, and the second sentence contains the verb-like noun. The typicality score for each word appears in parenthesis next to each sentence. The value appearing in brackets before each sentence (or sentence pair) represents the norming value elicited by each sentence frame, as described in the materials section of Experiment 2.

Items from Experiment 1

[1.6] The curious young boy saved the **marble** that he found on the playground. (−.564)

The curious young boy saved the **insect** that he found in his backyard. (.243)

[2.15] The very little girl imitated the **laughter** of the old woman with a tone of mockery. (−.482)

The very little girl imitated the **infant** as soon as it began to cry. (.279)

[2.53] The very angry man described the **neighbor** as a menace to society. (−.491)

The very angry man described the **theft** to the policeman soon after it had occurred. (.327)

[3.08] The group of friends discussed the **movies** that they had just gone to see. (−.488)

The group of friends discussed the **scenes** from the movie that they found most humorous. (.344)

[1.53] The terrible car accident blocked many **drivers** from the main entrance to the shopping mall. (−.380)

The terrible car accident blocked many **lanes** of the town's only major highway. (.318)

[1.43] The extremely generous woman bought her **daughter** many expensive gifts for her birthday. (−.523)

The extremely generous woman bought her **friends** dinner at an expensive restaurant. (.319)

[2.4] The quiet college student read the **bible** during times of intense stress. (−.547)

The quiet college student read the **text** assigned by his history professor. (.317)

[2.56] The conservative political commentator criticized the **lawyers** for defending the killer. (−.437)

The conservative political commentator criticized the **airlines** for overcharging. (.236)

[1.6] The company truck driver unloaded the **cargo** from his truck onto the loading dock. (−.399)

The company truck driver unloaded the **trunks** from his truck into his client's office. (.264)

[1.45] The moving company employees carried the **sofa** from the van into the house. (−.431)

The moving company employees carried the **chest** from the van into the house. (.332)

[2.13] The very cautious tourist followed the **leader** of the tour group. (−.514)

The very cautious tourist followed the **streets** exactly as they appeared on the map. (.407)

[2.2] The economically savvy businessman established the **journal** in hopes of making more money. (−.446)

The economically savvy businessman established the **estate** in the name of his mother. (.360)

[2.2] The newly opened zoo received the **tiger** as a present from India. (−.567)

The newly opened zoo received the **birds** as a present from the benefactor. (.335)

[1.83] The overly concerned parent visited the **teacher** about his son. (−.527)

The overly concerned parent visited the **priest** about his son. (.429)

[3.43] The world famous coach taught the **bowler** how to hold the ball. (−.496)

The world famous coach taught the **teams** how to play the sport. (.324)

[1.58] The department store clerk accepted the **dollar** from the shopper. (−.439)

The department store clerk accepted the **gloves** for return. (.255)

[1.9] The world renowned scientist found the **data** from an old experiment. (−.471)

The world renowned scientist found the **genes** that are responsible for Parkinson's disease. (.314)

[2.13] The respected university president persuaded the **writer** to publish the controversial novel. (−.432)

The respected university president persuaded the **experts** to rank the university even higher. (.354)

Quoted Contexts from Experiment 2

[3.53] The teacher wrote the word "________"on the blackboard.

[3.78] The second word on the sign was "________" and was spelled correctly.

[2.48] John's e-mail password is "________" and he never forgets it.

[3.2] Paul had the word "________" tattooed on his left arm.

[3.28] The answer to 3-down on the crossword puzzle was "________" but Jason couldn't figure it out.

[2.23] The e-mail was entitled "________" but turned out to be spam.

[2.08] The new TV show was called "________" and first aired last fall.

[3.6] Sally played the word "________" during the Scrabble game but only got a few points for it.

(*Appendices continue*)

Additional Quoted and Nonquoted Contexts from Experiment 3

Strongly Noun-Biased Items

[2.73] The very wealthy businessman resented the **farmer** for not accepting the deal. (−.608)

The very wealthy businessman resented the **prince** for not accepting the deal. (.287)

[3.1] The brilliant performer answered the **scholar** with a quick reply. (−.366)

The brilliant performer answered the **applause** with a humble bow. (.265)

Less Noun-Biased Contexts

[3.68] Kathy had to spell the word "________" in the first grade spelling bee.

[3.38] Brad was typing the word "________" when the phone started to ring.

[2.85] Students were asked the meaning of the word "________" on the vocabulary quiz.

[3.53] The 4-year-old just said the word "________" for the first time.

[4.53] Andy had to act out the word "________" during the charades game.

[2.9] The sixth word on the wordlist that Sarah had to remember was "________" but she forgot it.

[3.25] The secret word used to get into the clubhouse was "________" but no one told Samantha.

[3.23] Someone etched the word "________" into the drying cement.

[3.25] Cindy Googled the word "________" and was surprised by how many hits she got.

[3.8] Mary finished typing the word "________" before she answered the phone.

[3.18] Alex thought the second word on the t-shirt was "________"but he turned out to be wrong.

[2.8] The poet read her new poem entitled "________"at open mic night yesterday.

Received September 18, 2014
Revision received March 4, 2015
Accepted March 5, 2015 ■