AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology:
Human Perception and Performance

ISSN: 0096-1523

2021, Vol. 47, No. 1, 81–102
http://dx.doi.org/10.1037/xhp0000878

Contextual Constraint and Lexical Competition: Revisiting Biased Misperception During Reading

Jesse A. Harris, Stephanie Rich, and Ian Rigby
University of California, Los Angeles

In identifying and accessing lexical items while comprehending text, readers must rapidly select a word from visually similar words before integrating it into a sentence. It has been proposed that readers are likely to misperceive a low frequency word as a highly frequent orthographically similar alternative, particularly when the alternative is supported by previous context (Gregg & Inhoff, 2016; Perea & Pollatsek, 1998; Pollatsek, Perea, & Binder, 1999; Slattery, 2009). In such cases, the misperception may not be corrected until the reader encounters incongruent information. However, many of these studies place incongruent text directly after the critical word, confounding whether readers regress backward in text to resolve their misperception or to halt forward text progression in order resolve a lexical level conflict between the word form and its competitor. In 3 eye tracking while reading experiments, we adapted materials from previous studies to include a postcritical spillover region to address this possibility. Two of these experiments were designed to permit an ex-Gaussian analysis of the distribution of first pass reading prior to disambiguating information. The evidence suggests that postlexical competition-inhibition between orthographically similar forms can delay forward movement of the eyes as a competitor is inhibited. The possibility that misperception and postlexical competition-inhibition arise from the same set of mechanisms is discussed.

Public Significance Statement
This series of reading studies suggests that readers delay moving forward in text after encountering a word that is similar in form to another highly frequent word prior to disambiguation.

Keywords: misperception, eye tracking while reading, ex-Gaussian analysis

Supplemental materials: http://dx.doi.org/10.1037/xhp0000878.supp

Reading is no easy matter. Yet, many adult readers have acquired skills and strategies that make it remarkably efficient and phenomenologically effortless. In order to swiftly progress through text, readers must recognize, access, and integrate words into the sentence and discourse with incredible speed. Scores of reading studies over the decades have explored the cognitive processes that make such a demanding task seemingly so efficient, typically by generating situations in which normal reading strategies and interpretative processes conflict with the text in some way. We explore how the language processing system is designed to recognize and access a lexical item from among other orthographically similar word forms called *orthographic neighbors*, and the extent to which these forms are activated during normal reading (Perea, 2015 for review). As with previous natural reading studies on the topic (e.g., Gregg & Inhoff, 2016; Pollatsek et al., 1999; Sears, Campbell, & Lupker, 2006; Slattery, 2009; Veldre & Andrews, 2015, 2017; Warrington, McGowan, Paterson, & White, 2018; Williams, Perea, Pollatsek, & Rayner, 2006), the experiments below were designed to investigate whether the presence of a highly frequent orthographic neighbor disrupts the reading process, and, if so, at what point in sentence interpretation such a disruption becomes apparent. The theoretical issue we address is the extent to which disruption from a more frequent neighbor can be attributed to a complete *misperception* of the target word as opposed to *postlexical competition and inhibition* among neighboring words that slows forward movement of the eyes.

This article was published Online First October 22, 2020.

Jesse A. Harris, Stephanie Rich, and Ian Rigby, Department of Linguistics, University of California, Los Angeles.

Stephanie Rich is now at the Department of Linguistics, University of California, Santa Cruz. Ian Rigby is now at the Department of Linguistics, University of Southern California.

The research was completed in the Linguistics Department at the University of California, Los Angeles. We are indebted to Adrian Staub for his guidance on the ex-Gaussian analysis. We thank Adrian Staub and Patrick Plummer for their generous feedback on previous versions. We also thank Lexi Loessberg-Zahl for assisting with the eye-tracking experiments. Portions of the research were presented at the 30th Annual CUNY Conference on Human Sentence Processing in 2017 at the Massachusetts Institute of Technology, and we thank the audience there for their comments.

Correspondence concerning this article should be addressed to Jesse A. Harris, Department of Linguistics, University of California, Los Angeles, 3125 Campbell Hall, Los Angeles, CA 90095. E-mail: jharris@humnet.ucla.edu

We adopt the classic definition of an *orthographic neighbor* (ON), in which two words of the same length are orthographic neighbors of each other if they differ in just one letter (Coltheart, Davelaar, Jonasson, & Besner, 1977; Havens & Foote, 1963; but see also Landauer & Streeter, 1973 and Yarkoni, Balota, & Yap, 2008 for alternate definitions). For example, *coat* has *boat* as one of its ONs, but not *cut* (which is a phonological neighbor of *coat*) or a nonword, like *zoat.* Words further differ in terms of their relative *neighborhood sizes* or *densities* (denoted as N), which quantifies how many ONs a word has. Some words like *page* have large or dense Ns (*wage*, *sage*, *rage, pane*, *pale*, *pare*, etc.), whereas other words like *ghost* have no ONs at all.

Effects of Neighborhood Size and Frequency

Whether the effect of N on word recognition and lexical processing is *facilitatory* or *inhibitory* has been the subject of a long-lasting dispute. On the one hand, words with higher Ns are sometimes facilitated in lexical decision and naming tasks (Andrews, 1992, 1997; Carreiras, Perea, & Grainger, 1997; Huntsman & Lima, 1996; Sears, Hino, & Lupker, 1995). Such an effect is captured straightforwardly on single stage interactive models of lexical retrieval (e.g., McClelland & Rumelhart, 1981), in the which bottom-up processing of the visual input of a word activates its ONs by passively activating letters that words in the neighborhood have in common, inducing increased top-down activation of the stimulus. Thus, the more word-like the target is, the more overall activation is generated from the ONs, and the faster the word is recognized as a word.

However, there is also evidence that word recognition may also be *inhibited* for words with large Ns in lexical decision (Perea & Rosa, 2000a; Pollatsek et al., 1999) and other identification paradigms (e.g., Grainger & Segui, 1990; Snodgrass & Mintzer, 1993). Studies showing inhibition have usually controlled for the frequency of a word's neighbors, and argue for a distinct effect of high versus low frequency neighbors. Whereas a single highly frequent orthographic neighbor (HFON) disrupts lexical access, particularly for low Ns, words without a highly frequent neighbor may benefit from more neighbors (e.g., Paap & Johansen, 1994; Perea & Rosa, 2000a, 2000b). Moreover, recognizing a letter string as a word is not equivalent to accessing its lexical characteristics (e.g., Balota, 1994), and some semantic priming studies show a reduced effect of a large N when semantic properties need to be retrieved as well (see Perea & Rosa, 2000a for review).

The neighborhood frequency effect is theoretically compatible with multiple architectures of lexical retrieval. For example, in a single stage interactive-activation model, lexical representations inhibit each other according to several factors, such as orthographic overlap, salience, and frequency. Higher frequency words enjoy a higher resting level of activation and are thus able to be activated more easily. A HFON provides greater inhibition with the target than other words in the N. In contrast, multistage activation-verification models (e.g., Paap, Newsome, McDonald, & Schvaneveldt, 1982) distinguish between an initial *word recognition* stage, in which candidate word forms that are partially compatible with the input are activated, and a later *verification* stage, in which a (single) lexical entry is selected and accessed. The greater frequency of the HFON gives it a head start in the race for verification, which in turn escalates competition with the target word, thereby delaying the time that the winner reaches the decision threshold (e.g., Grainger & Jacobs, 1996).

Task Related Effects

The effect of N and a HFON may also depend on the task-specific strategies that allow word recognition or familiarity without necessarily requiring access to lexically specific information. Facilitation for words with large Ns is found in lexical decision tasks results when speed is emphasized over accuracy (Carreiras et al., 1997; Grainger & Jacobs, 1996) or when lexical targets are embedded among nonword fillers that are not very word-like (Andrews, 1989, 1997; Forster & Shen, 1996; Grainger & Jacobs, 1996; N. F. Johnson & Pugh, 1994; Sears et al., 1995). Similarly, the effect of a HFON may depend on which strategies subjects adopt in the experiment, for example, favoring speed over accuracy (Sears et al., 2006).

The dependency on task is especially important for considering how effects observed in isolated word recognition tasks might transfer to naturalistic reading tasks (Perea & Pollatsek, 1998; Pollatsek et al., 1999; Sears et al., 2006). Pollatsek, Perea, and Binder (1999) found that words with larger Ns facilitated responses times in a lexical-decision task, but inhibited early reading times when embedded into sentence frames in an eye movement study, arguing that the inhibitory effect was due to a HFON instead of absolute neighborhood size. Their interpretation was supported by Pollatsek et al.'s (1999) follow-up eye tracking study which again compared words with high and low N in words without HFONs. Words with larger Ns not only tended to be fixated longer in later rereading measures, but were also skipped twice as often in first pass reading compared to words with smaller Ns. Increased skipping rates were observed only for words with neighbors that were compatible with the pretarget context. They hypothesized that readers sometimes misidentified the target as one of its neighbors in parafoveal preview and subsequently skipped the target on the basis of partial word identification, so that readers were also more likely to return to and spend more time on targets with more neighbors to correct the misperception.

Biased Misperception

Slattery (2009) reasoned that readers would be more likely to misperceive a low frequency word as its ON when the neighbor was highly frequent (provided that the ON is also compatible with the prior sentence context) compared with words with only low frequency neighbors. We follow Gregg and Inhoff (2016) in referring to this idea as the *biased misperception hypothesis*. Slattery (2009) proposed that readers who initially misperceived the target word on first pass reading would be more likely to return to the word, spend more time processing it during rereading. As expected, readers regressed back to and spent more time rereading low frequency experimental words (*brunch*) with HFONs (*branch*) more often than matched control words (*buffet*) without a HFON, particularly when the prior context was compatible with the HFON (the *neutral context*).

(1) *Sentence pair from* Slattery (2009), where "|" marks analysis regions. The HFON "branch" is never shown.

a. *Neutral context:* Due to the freezing rain, the | (brunch/buffet) | was postponed a week.

b. *Biased context:* Everyone said the food at the | (brunch/buffet) | was simply magnificent.

More evidence for biased misperception comes from Gregg and Inhoff (2016), who presented readers with sentences with both high and low frequency members of the orthographic pair (e.g., high: *branch* vs. low: *brunch*).

The target word was followed by three different contexts differing according to which word they supported: congruent with target, congruent with the HFON, and neutral. A penalty for low frequency words was observed on the target region itself, but crucially was reduced when the posttarget context was congruent with the HFON. This asymmetry can be explained if readers are more likely to misperceive a low frequency target word (*brunch*) as its HFON (*branch*) than vice versa. In the event that the target word was misidentified, posttarget contexts compatible with its neighbor would have facilitated integrating the misperceived word into the sentence as its neighbor. Gregg and Inhoff (2016) also manipulated the extent to which *pretarget* material was contextually constrained to support the HFON in two separate experiments. A word frequency effect and its interaction with posttarget congruency was observed both in an experiment with only neutral pretarget contexts and in an experiment with only contexts biased toward the HFON. However, the effect of biased misperception was not greater in the study with contexts biased toward the HFON. In an oral reading follow-up of pretarget contexts biasing toward the HFON, Gregg and Inhoff (2016) observed that a low frequency word was misnamed as its HFON more often (6%) than a high frequency word was misnamed as its low frequency neighbor (1%). Naming errors on low frequency words further increased when the posttarget context biased toward the HFON.

The effects of a HFON in reading also been investigated with a gaze-contingent boundary change paradigm (Rayner, 1975). In this paradigm, a word presented in parafoveal preview is replaced by the target word (*preview* > *target*) once the reader crosses an invisible boundary (|) during a saccade, so that the preview is present only in parafoveal vision. Williams, Perea, Pollatsek, and Rayner (2006) found that a HFON preview (*sweet*) provided as much of an advantage in early reading measures on a low frequency target (*sleet*) word (*Mary was afraid of the* | *sweet* > *sleet* . . .) as an identical word preview (*sleet* > *sleet*) compared with a nonword with the same number of matching letters in the same position (*speet* > *sleet*). However, a low frequency orthographic neighbor did not reduce reading times on the high frequency target (*sleet* > *sweet*) any more than nonword controls. They proposed that while frequency aided the recognition of letter identity of a word form in parafoveal vision, lexical competition was delayed until a postlexical verification stage. In another boundary charge study, Veldre and Andrews (2015) observed that the effect of a HFON over a nonword preview was modulated by reading proficiency and neighborhood density. Whereas a HFON preview elicited a penalty for high proficiency readers for words in dense neighborhoods, it conferred an advantage for low proficiency readers, who spent longer rereading the word after the boundary change.[1]

Eye movement behavior on forced fixations is also highly compatible with biased misperception (Schotter & Leinenger, 2016; Schotter, Leinenger, & von der Malsburg, 2018, 2019). When a word is sufficiently recognized in parafoveal preview, the oculomotor control system may opt to skip over the word entirely. For example, in E-Z Reader (Reichle, 2011; Reichle, Pollatsek, Fisher, & Rayner, 1998, 2006) and SWIFT (Engbert, Nuthmann, Richter, & Kliegl, 2005) models of reading, saccadic planning consists of two stages: a labile stage (M1), in which a saccade may be cancelled if a signal is received before the temporal deadline, and a nonlabile stage (M2). If a saccade to the word cannot be cancelled during M1, the eye may be forced to briefly fixate on the word target, even as it prepares to make another fixation elsewhere. Gaze contingent display change studies have shown that highly frequent words in parafoveal preview may prompt forced fixations (Schotter & Leinenger, 2016). Readers may incorrectly identify the word as its preview, and fail to correct the misperception, despite fixating on the word directly. Additional evidence suggests that readers fully incorporate misperceived words into their representation of the sentence; when comprehension questions probed the identity of the target word directly, readers were more likely to report having seen the highly frequent preview when they had either fixated on the target only briefly or skipped it entirely (Schotter & Jia, 2016; Schotter et al., 2018). Similar findings have been observed when the preview word is highly plausible given preceding context (Schotter, von der Malsburg, & Leinenger, 2019; Veldre & Andrews, 2016, 2017).

Misperception and Lexical Competition-Inhibition

That readers may sometimes misperceive a word as its HFON is intuitive and well documented. Several accounts suggest that a misperception occurs when a word form is incorrectly recognized due to top down factors such as increased lexical activation from a highly frequent neighbor, perhaps in concert with highly supporting prior context, so that the HFON is subsequently integrated into the sentence as the target would have been (Gregg & Inhoff, 2016; Pollatsek et al., 1999; Slattery, 2009). A misperceived word would not be noticed or corrected until the reader encounters disconfirming evidence, which renders the HFON implausible. Complete misperception of this sort may be most likely in cases of word skipping, where the target word is never fixated on in first pass reading (see Warrington et al., 2018 for discussion) or is fixated very briefly (Schotter et al., 2018, 2019).

However, complete misperception may not be the sole outcome of a HFON in reading (Sears et al., 2006; Veldre & Andrews, 2015; Williams et al., 2006). Another potential consequence is that the HFON competes with the target word for recognition during reading, resulting in processing delays until the competitor is inhibited and the target word can be integrated into the sentence. Competition and subsequent inhibition between the target word and a HFON might cause the oculomotor control system to slow

[1] A reviewer noted that the perceptual evidence given to readers in boundary change studies is very different than what is encountering in normal reading. Indeed, presenting a HFON as an orthographic prime in preview gives perceptual evidence for the HFON. In contrast, studies without a boundary change do not provide direct perceptual evidence in favor of the HFON, as the HFON word is never shown. Thus, the two paradigms may potentially induce different sources for activation.

down on the critical word or its spillover region to allow the word recognition system to catch up, as predicted by models of reading which encode some temporal dissociation between word recognition and saccadic generator systems (e.g., Engbert et al., 2005; Mitchell, Shen, Green, & Hodgson, 2008; Schad & Engbert, 2012). Studies in which disambiguating material immediately follows the target word typically were not designed to explicitly tease apart these two potential outcomes. Although many interpretations of our results are possible, we fully expect that both complete misperception and postlexical delay occur in language processing, though they might be influenced by different factors and varying by individual reading strategies (Veldre & Andrews, 2015).

In the General Discussion, we attempt to unite misperception and lexical competition-inhibition within a single account inspired in part by prior models of lexical access in which word recognition may be followed by a postlexical verification check (Balota & Spieler, 1999; Chumbley & Balota, 1984; Paap et al., 1982; see also Williams et al., 2006 for related discussion). In many models, a HFON competes with the input for word recognition. In cases where familiarity alone suffices for word recognition, a HFON would promote faster attentional shift to the next word, even as lexical processing of the word itself continues. In such instances, the HFON could be sometimes retrieved erroneously and integrated into the input in place of the actual word, corrected only when the reader encounters inconsistent information later in the sentence. In contrast, there may also be cases in which competition between the target and its HFON persists as the eye moves to the next fixation location, engaging a postlexical checking process in which additional attention is allocated toward discriminating the word target from its HFON. In such cases, forward progression of the eyes might halt in order to resolve competition between word forms, preventing the HFON from being fully integrated into the sentence. Assuming that the postlexical check is successful, the HFON is rejected and there is no misperception. Thus, it is possible that processing costs of words with a HFON would become evident *before* disambiguating information is encountered. Such findings would suggest that at least some of the increased reading time penalties observed by previous studies could be attributed to delays in postlexical processing, rather than initially misperceiving the target as its HFON, although both outcomes could be generated by the same underlying processing system.

The Present Study

The central goals for this study were to investigate two possible consequences of a HFON on the target word: first, an *inhibitory effect* of a HFON in unrestricted reading (e.g., Grainger, O'Regan, Jacobs, & Segui, 1992; Paap, Johansen, Chun, & Vonnahme, 2000; Paterson, Liversedge, & Davis, 2009), and second, whether any processing delays can be detected *prior to disambiguation*. Experiment 1 was designed to partially replicate the previously observed effect of pretarget contexts on words with a HFON, compared with control word lacking a HFON (but matched with the experimental target word on other relevant lexical characteristics, as in Slattery, 2009 and others). Although we followed previous manipulations closely in the first experiment, we departed in two primary ways. First, we added a pretarget contextual manipulation that biases toward the target as well as context that biases toward the HFON alternative, while also minimizing lexical and syntactic differences in biasing contexts between conditions. Second, we redesigned the materials to separate the critical target word from the disambiguation region with a spillover region compatible with the control word, the experimental word, and its HFON. Reading delays on the target word or on the following spillover region would support the claim that a HFON increases postlexical competition-inhibition without resulting in misperception of the word.

A good deal of literature has shown that strongly biasing context modulates the accessibility of lexical and word sense competitors. While words with multiple unrelated senses slow reading compared with unambiguous controls in early measures, supporting context appears to eliminate the penalty (e.g., Binder & Morris, 1995; Binder & Rayner, 1998; Duffy, Morris, & Rayner, 1988; Rayner & Duffy, 1987; Swinney, 1979; Tanenhaus, Leiman, & Seidenberg, 1979, among many others). Contextual support has been shown to further modulate the influence of orthographically related forms (e.g., Johnson, 2009). And yet, early effects of prior context on words with a HFON has been somewhat mixed; Slattery (2009) observed that the early penalty for words with HFONs in neutral contexts was mitigated in contexts biasing toward the HFON, but Gregg and Inhoff (2016) found only a trend across experiments.

The second two experiments were thus specifically designed to address the influence of contextual bias on words with a HFON in ex-Gaussian distributions (Ratcliff, 1979) of the first-pass reading, in addition to standard eye reading measures. An ex-Gaussian analysis can detect differences in the shape of the distribution that may not be evident in the mean (e.g., Balota & Yap, 2011). Experiment 2 compares (a) neutral contexts compatible with both the target and its HFON against (b) contexts biased toward the HFON. Experiment 3 compares (a) neutral contexts against (b) contexts biased toward the experimental word, that is, the word actually presented to the reader. We predicted that a HFON would influence fixation distributions on or immediately after the target word, especially if the HFON was congruent with the preceding context. In particular, we expected that context supporting a HFON would increase the rightward skew of fixation durations prior to reaching disambiguating information, by engaging a postlexical verification process arbitrating between orthographically similar word forms.

Experiment 1

We conducted an eye tracking while reading experiment to address the biased misperception hypothesis with a design intended to distinguish postlexical processing delays due to competition-inhibition from misperception of a word form as its HFON. If readers completely misperceive a word as its HFON, disruptions should appear only after they encounter disambiguating information. If words with a HFON also increase competition with (and subsequent inhibition of) the neighbor, additional disruption may appear relatively early in the reading record.

Method

Participants. Sixty-six self-reported native English speakers from the University of California, Los Angeles participated in the study for one course credit in sessions lasting no more than 30 min. Participants for all experiments were college-aged (approximately 18–24). We did not collect gender information, but estimate from names not linked to the data that approximately 70% of our participants were female in all the experiments reported here. All participants had normal or corrected-to-normal vision.[2]

Materials. Materials consisted of 30 sextets in a 2 × 3 design, crossing a target word (*Experimental* word with HFON vs. *Control* word without HFON) with the bias of the preceding context (*Neutral* vs. *Bias-Exp* vs. *Bias-HN*). The experiment consisted of 30 pairs of sentences manipulating the context preceding a target word. Each target word was matched with a high frequency orthographic neighbor (HFON). When possible, we used the same the targets and HFONs as Slattery (2009). As in other studies, control words with no HFONs were matched with the target words along several dimensions, including length, first letter, frequency, syntactic category, number of syllables, number of phonemes, and number of morphemes.

Frequency was estimated in terms of Hyper Analogue to Language (HAL; Lund & Burgess, 1996) and SUBTLEX (Brysbaert & New, 2009) values obtained from the English Lexicon Project (Balota et al., 2007). As intended, the log frequency of the HFON was significantly higher than its target neighbor, Log HAL: $t(29) = 13.29$, $p < .001$; Log SUBTLEX: $t(29) = 10.50$, $p < .001$, and the control word, Log HAL: $t(29) = 9.27$, $p < .001$; Log SUBTLEX: $t(29) = 10.79$, $p < .001$, in paired t tests with Bonferroni corrections in both measures. The difference in frequency of the control and the target words was not significant on either of these measures; see Table 1 for a summary of the lexical characteristics.

We replaced word pairs that did not hold part of speech constant in the sentence frame. Unlike previous research, differences in syntactic frame structure and length in the pretarget context were minimized, holding at least one to two pretarget words constant before the target word. Importantly, identical posttarget one- to two-word spillover regions (e.g., *would not*) followed the target region containing information that clearly disambiguated the target from the HFON competitor. Spillover words consisted solely of function words, such as auxiliaries, articles, negation, and connectives, and are highly unlikely to produce bias toward any word over another. Material in the disambiguating region was held consistent within control or the experimental word conditions, such that control words were always followed by one continuation (2a, c, e) and experimental words by another (2b, d, f). The HFON of the target is provided in parentheses in (2) below. All items are provided in the online supplementary materials. Analysis regions are indicated by a pipe symbol "|" and labeled as *pretarget*, *target*, *spillover*, and *disambiguation*.

Table 1
Lexical Characteristics for Experiment 1

Measure	Control	Target	HFON
Length	5.30 (0.10)	5.30 (0.10)	5.30 (0.10)
Log HAL	8.48 (0.22)	8.14 (0.19)	10.72 (0.17)
Log SUBTLEX	2.44 (0.08)	2.43 (0.08)	3.45 (0.08)
Phonemes	4.17 (0.13)	4.23 (0.12)	4.33 (0.13)
Syllables	1.50 (0.09)	1.27 (0.08)	1.30 (0.09)
Morphemes	1.03 (0.03)	1.07 (0.05)	1.03 (0.03)

Note. Only measures of word frequency (log HAL and log SUBTLEX) differed between conditions.

(2) *Sample item from Experiment 1*

a. *Bias-HN - Control:* $|_{\text{Pretarget}}$ The geometry teacher said that his right $|_{\text{Target}}$ elbow $|_{\text{Spillover}}$ would be $|_{\text{Final}}$ healed after a few days of rest.

b. *Bias-HN - Exp:* $|_{\text{Pretarget}}$ The geometry teacher said that his right $|_{\text{Target}}$ ankle (angle) $|_{\text{Spillover}}$ would be $|_{\text{Final}}$ fixed before the end of the week.

c. *Bias-Exp - Control:* $|_{\text{Pretarget}}$ The doctor quickly said that his right $|_{\text{Target}}$ elbow $|_{\text{Spillover}}$ would be $|_{\text{Final}}$ healed after a few days of rest.

d. *Bias-Exp - Exp:* $|_{\text{Pretarget}}$ The doctor quickly said that his right $|_{\text{Target}}$ ankle (angle) $|_{\text{Spillover}}$ would be $|_{\text{Final}}$ fixed before the end of the week.

e. *Neutral - Control:* $|_{\text{Pretarget}}$ The teacher quickly said that his right $|_{\text{Target}}$ elbow $|_{\text{Spillover}}$ would be $|_{\text{Final}}$ healed after a few days of rest.

f. *Neutral - Exp:* $|_{\text{Pretarget}}$ The teacher quickly said that his right $|_{\text{Target}}$/ankle (angle)/$|_{\text{Spillover}}$ would be $|_{\text{Final}}$ fixed before the end of the week.

Items were presented in a counterbalanced and individually randomized Latin Square design, along with 56 items from unrelated experiments and filler items.[3]

Continuation norming study. Prior to testing, a separate Internet ratings study ($N = 30$) presented a separate set of self-

[2] Computing the power for a linear mixed effects analysis requires estimating the degrees of freedom in those models, which is not trivial as the parameter estimates are calculated from the residual maximum likelihood, rather than from observed and expected mean squares familiar from ANOVAs (Pinheiro & Bates, 2006). Therefore, we initially estimated the sample size on the basis of previous eye tracking experiments addressing similar questions. However, we also performed a power analysis using estimates based on the F-distribution, taking Slattery's (2009) study as a starting place for comparison. Slattery collected eye movement data from 32 subjects in a 2 × 2-crossed design with 44 quartets. Assuming a standard significance level of $\alpha = 0.05$ and a power of $\beta = 80\%$, the effect size of his study is approximately $d = 0.3$. Experiment 1 reports data from 48 subjects in a 2 × 3-crossed design with 30 sextets. Assuming the same $d = 0.3$ effect size and $\alpha = 0.05$ significance level, the power of our study is $\beta = 0.99$ under traditional power analyses. However, repeated measures designs complicate power calculations (Brysbaert & Stevens, 2018; Judd, Westfall, & Kenny, 2017). As each condition in our study received fewer observations per subject compared to Slattery's, it was likely less highly powered than the original.

[3] As an exploratory factor probing individual reading skill and print exposure, Moore and Gordon's (2015) version of the author recognition task (ART; Stanovich & West, 1989) questionnaire was administered to participants after the reading section for each experiment (Experiment 1: $M = 14$, range = 0–46; Experiment 2: $M = 19$, range = 6–43; Experiment 3: $M = 19$, range = 6–43). ART scores did not improve linear and logistic mixed effect regression models of standard reading measures when added as an interactive predictor, or interact with experimental variables, although higher ART scores did associate with faster reading overall across experiments. In Experiment 3, the ex-Gaussian analysis of first fixation and first pass times on the target word revealed a moderate negative correlation ($r = -0.33$) in the normal Gaussian parameters, and a negative correlation ($r = -0.27$) in the exponential parameter, for the difference between ART and the effect of context. No other effects of ART were observed in any experiment. Given the lack of a consistent effect of ART, a more complete discussion is omitted from the text.

reported native English-speaking subjects with the sentence fragment followed by either the experimental (*angle*) or the control (*elbow*) word.

(3) a. *Neutral:* The teacher quickly said that his right . . . ANGLE/ELBOW

b. *Bias-HN:* The doctor quickly said that his right . . . ANGLE/ELBOW

c. *Bias-Exp:* The geometry teacher said that his right . . . ANGLE/ELBOW

In by-items paired *t* tests with Bonferroni corrections, the control word was judged to be as natural as experimental words in neutral contexts (3a) and in contexts biased to the HFON (3b), t's < 1. However, experimental words were rated as more natural continuations in contexts biased to the experimental word (3c) than control words were, $t_2(29) = -5.38$; $p < .001$. Results from all continuation rating norming studies are presented in Table 2 and Figure 1.

Procedure. Eye movements were monitored with an SR EyeLink 1000 Plus eye tracker. Sentences were presented and recorded using a Lenovo desktop computer running Windows 7 via the University of Massachusetts, Amherst EyeTrack software (https://blogs.umass.edu/eyelab/software/). Materials were presented to subjects seated 55 cm from a 24-in. Dell UltraSharp U2410 LCD monitor (55.88 cm width × 49.27 cm height) set at a 1024 × 768 resolution. We estimate that approximately two characters subtended one degree of visual angle. Viewing was binocular, but only data from the right eye were recorded. The text was displayed on a single line in 11-point fixed width proportion Monaco black font on a light gray background. Participants were calibrated using a three-point procedure on a single line. Participants were encouraged to take a break halfway through, or as frequently as needed, and were recalibrated as needed. All unnecessary software was turned off, and there was no connection to the Internet.

Fifteen participants were removed for excessive track loss or blinking during first pass reading of the target region, two were removed for poor performance on comprehension questions, and one was excluded due to an attention-deficit/hyperactivity disorder. All remaining participants in the final data set scored above 80% on comprehension questions, with an overall average score of 95% and an average score of 94% on questions corresponding to the relevant experimental items. Forty-eight participants equally distributed across counterbalancing lists were retained in the final data set. Prior to analysis, fixations under 40ms were merged with fixations occurring within one character of it, and other fixations under 80 ms were removed. Fixations above 1,200 ms were excluded. Data were cleared of major track losses and blinks in first pass reading of the target and post target regions using University of Massachusetts, Amherst EyeDoctor software (https://blogs.umass.edu/eyelab/software/). A total of 83 trials were removed from the final data set, leaving 1,357 trials remaining for analysis.

Eye movement measures. The standard reading measures reported in this analysis are *first fixation durations*, the duration of the first fixation in a region, *first pass times*, sum of all fixations in a region before exiting that region to the right or left, *go past times*, sum of all fixations beginning with the first fixation in a region before exiting that region to the right, including rereading previous regions, *second pass times*, rereading times in a region after it has been exited to the right, the *probability of regressions out* of and *regressions into* a region, and the rate at which a region was *skipped* in first pass reading (Rayner, 1998). For the biasing pretarget region, only rereading measures (regressions in and second pass times) are discussed. For roughly normally distributed measures (first fixation durations and first pass), observations above the top and bottom 5th percentile of first pass and first fixation measures were considered outliers, and transformed to the fifth and 95th percentile respectively via a symmetrical censoring procedure known as winsorization (Dixon, 1960; Tukey, 1962). Censoring values in this way retains a transformed representation of all data points in comparison to other forms of data trimming, and has been previously used in other eye tracking while reading studies (e.g., Sturt, Pickering, & Crocker, 1999; Van Dyke & McElree, 2011). Fixations over 4,000 ms for go past times were removed, accounting for less than 1% of the data. No data was transformed or removed from second pass reading measures, which are typically skewed due to the high proportion of zeroes indicating a failure to reread a region. First fixation, first pass, go past, and second pass time measures were analyzed as linear mixed effect regression models (Baayen, Davidson, & Bates, 2008). Regressions out, regressions in, and skipping rates are binomially distributed measures, and were modeled as logistic linear mixed effect regression models (Jaeger, 2008).

Results

All analyses were conducted in R (R Core Development Team, 2017). Means and standard errors for all measures of all four regions can be found in Table 3.

The data were modeled as a linear mixed effects regression models with fixed effects as random slopes and intercepts using lme4 (Bates, Mächler, Bolker, & Walker, 2015). As discussed in (Barr, Levy, Scheepers, & Tily, 2013), there is currently no universally adopted standard for computing *p* values in multilevel regression models. We estimated *p* values with the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017), using the

Table 2
Results of Norming Studies for Experiments 1–3

Experiment	Target	Context	Mean rating	Standard deviation
Experiment 1	Neutral	Control	5.11	1.80
		Experimental	5.35	1.72
	Bias-Exp	Control	4.72	2.01
		Experimental	6.05	1.46
	Bias-HN	Control	4.64	1.85
		Experimental	4.36	2.03
Experiment 2	Neutral	HFON	5.22	1.94
		Experimental	5.28	1.95
	Bias-HN	HFON	6.17	1.46
		Experimental	4.74	2.05
Experiment 3	Neutral	HFON	5.76	1.72
		Experimental	5.88	1.66
	Bias-Exp	HFON	3.59	2.31
		Experimental	6.32	1.43

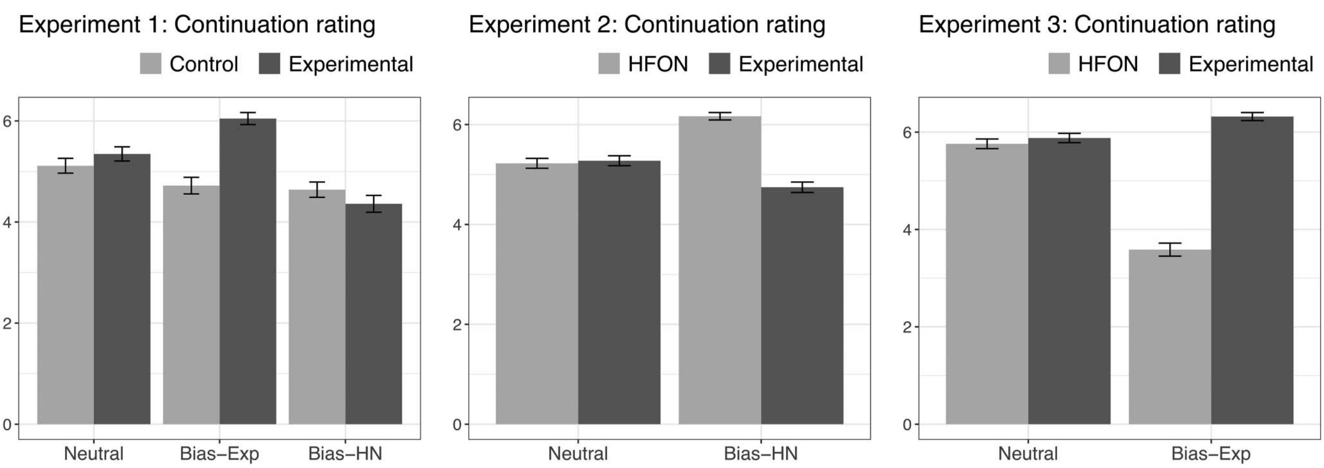

Figure 1. Means and standard errors for continuation rating norming studies in Experiments 1–3.

Satterthwaite method to approximate the degrees of freedom for the denominator in the *t*-statistic. Linear mixed effect regressions models were computed on standard eye movement measures, in which the factors of *bias*, *target*, and their interaction were included as fixed effects. The fixed effects contrasts were deviation (sum) coded so that each of the two biasing context conditions were compared to the neutral condition in the analysis. Models were first specified with by-subjects and by-item random slopes and intercepts (i.e., maximal random effect structures; Barr et al., 2013), but failed to converge on several measures. To present models with uniform random effects, all random effect structures were specified with random intercepts only in order to guarantee convergence. Results in the reported models followed the same pattern as those fit with maximal random effect structures.

Tests were limited to reduce the possibility of Type I error (von der Malsburg & Angele, 2017). Each reported measure was included in order to fully test the hypotheses and to facilitate comparison with related studies. Measures of first pass reading (first fixation, first pass time, and go past time) were reported for the target and spillover region only, to address the predictions of the competition-integration hypothesis and to permit close comparison to studies reporting ex-Gaussian analyses. Measures of second pass reading (second pass time and regressions in) were included for comparison with studies on biased misperception.[4] All significant effects observed are reported; see Table 4.

In reporting the findings, we differentiate the effect of pretarget context according to whether the reader had reached the disambiguation region or not. We predicted a general inhibitory effect of words with HFON and interactions between pretarget contextual bias and the target word prior to disambiguating information: an advantage for experimental words in supporting contexts, and a penalty for experimental words in contexts biasing toward the HFON competitor.

General penalty for words with a HFON. In support of an early inhibitory effect on words with a HFON, experimental words elicited marginally longer first fixation durations (diff = 4 ms, $t = 1.87$, $p = .06$) and a significant go past times penalty (diff = 22 ms, $t = 2.11$, $p < .05$) compared with controls on the target region. No general effects of the HFON were observed in the spillover region.

Contextual effects prior to disambiguation. Prior to disambiguation, we observed two interactions in go past times on the target region, depicted in Figure 2. First, reflecting a contextual congruency effect, experimental words elicited marginally shorter go past times in contexts that biased toward the experimental word (diff = 11 ms, $t = -1.94$, $p = .05$). Second, reflecting a contextual incongruency effect, experimental words were associated with differentially longer go past times in contexts that biased toward the HFON (diff = 83 ms, $t = 2.19$, $p < .05$). In the spillover region, contexts biased toward the experimental word elicited fewer regressions out following a word with a HFON over control words compared with the neutral context condition (diff = 5%, $z = -2.02$, $p < .05$).

Contextual effects in rereading measures. We observed robust evidence for contextual congruency and incongruency effects in regressions in and second pass rereading measures in keeping with Slattery (2009) and others. Contextual bias affected the difference between experimental and control words in different directions, shown in Figure 3. Contexts biasing to the experimental word resulted in fewer regressions into the pretarget region for experimental words compared to control words, 42% versus 54%, respectively, reflecting contextual congruency ($z = -2.78$, $p < .01$). In contrast, contexts biasing to the HFON elicited more regressions in for experimental words compared with control words, 58% versus 50%, respectively, reflecting contextual incongruency ($z = 2.95$, $p < .01$).

Interactions in second pass times supporting contextual congruency and contextual incongruency effects were observed in all three regions of interest. When the pretarget context biased toward the HFON, the penalty for second pass times for sentences con-

[4] Assuming that first fixation, first pass, and go past time measures are highly correlated, the significance level of go past times in Experiment 1 was corrected for multiple comparisons using Bonferroni corrections (adjusted $\alpha = .05/3 = .017$). The interaction between target word and context biased towards the experimental word on the spillover region remained significant ($p = .006$). The remainder of effects in go past times were not. Regression and second-pass measures were not corrected as their correlation with other measures is less direct. Corrections were not applied to Experiments 2–3, as only the minimal models are reported (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017).

Table 3
Means and Standard Errors for Experiment 1

Measures	Conditions	Regions			
		Pretarget	Target	Spillover	Final
First fixations	Bias-HN control	201 (3)	239 (5)	237 (4)	220 (3)
	Bias-HN experimental	209 (3)	234 (5)	234 (4)	225 (4)
	Neutral control	215 (3)	239 (5)	230 (4)	235 (4)
	Neutral experimental	210 (3)	236 (5)	236 (4)	222 (4)
	Bias-Exp control	212 (3)	239 (5)	237 (4)	229 (3)
	Bias-Exp experimental	212 (3)	225 (4)	231 (4)	221 (3)
First pass	Bias-HN control	1423 (30)	253 (6)	307 (8)	486 (15)
	Bias-HN experimental	1425 (28)	253 (6)	293 (8)	480 (12)
	Neutral control	1262 (23)	263 (7)	287 (8)	511 (15)
	Neutral experimental	1287 (29)	260 (7)	295 (8)	453 (12)
	Bias-Exp control	1491 (32)	253 (6)	299 (9)	515 (15)
	Bias-Exp experimental	1472 (32)	247 (6)	285 (7)	438 (11)
Go past	Bias-HN control	—	327 (16)	391 (17)	—
	Bias-HN experimental	—	410 (29)	425 (23)	—
	Neutral control	—	324 (14)	364 (17)	—
	Neutral experimental	—	353 (23)	417 (24)	—
	Bias-Exp control	—	351 (25)	404 (23)	—
	Bias-Exp experimental	—	340 (19)	350 (17)	—
Second pass	Bias-HN control	497 (44)	103 (13)	129 (17)	—
	Bias-HN experimental	422 (44)	90 (12)	126 (14)	—
	Neutral control	664 (55)	121 (13)	163 (14)	—
	Neutral experimental	703 (63)	159 (16)	167 (17)	—
	Bias-Exp control	456 (48)	124 (13)	128 (14)	—
	Bias-Exp experimental	417 (46)	73 (11)	96 (11)	—
Regressions out	Bias-HN control	—	17% (3)	15% (2)	18% (3)
	Bias-HN experimental	—	23% (3)	19% (3)	17% (2)
	Neutral control	—	17% (3)	15% (3)	19% (3)
	Neutral experimental	—	13% (3)	17% (3)	16% (2)
	Bias-Exp control	—	15% (3)	14% (2)	19% (3)
	Bias-Exp experimental	—	17% (3)	9% (2)	13% (2)
Regressions in	Bias-HN control	50% (3)	20% (3)	20% (3)	—
	Bias-HN experimental	58% (3)	26% (3)	23% (3)	—
	Neutral control	46% (3)	21% (3)	22% (3)	—
	Neutral experimental	44% (3)	22% (3)	18% (3)	—
	Bias-Exp control	54% (3)	18% (3)	23% (3)	—
	Bias-Exp experimental	42% (3)	14% (3)	16% (2)	—

taining the experimental word over the control word increased by 206 ms in the pretarget region ($z = 3.86$, $p < .001$), 56 ms in the target region ($z = 2.96$, $p < .01$), and 38 ms in the spillover region ($z = 2.85$, $p < .01$). In contrast, when the pretarget context biased toward the experimental word, second pass times for sentences containing the experimental word compared with the control word decreased 247 ms in the pretarget region ($z = -4.56$, $p < .001$), by 48 ms in the target region ($z = -4.33$, $p < .001$), and 67 ms in the spillover region ($z = -3.58$, $p < .001$).

Following Perea and Pollatsek (1998), subjects were split into two groups based on the median number of regressions made from the final region: A low group ($N = 25$) regressing in less than 17% of trials, and a high group ($N = 23$). No differences between groups were found on first fixation, first pass or go past times on the target word of spillover region.

Other effects of context. We observed a few main effects that were unrelated to questions of central theoretical interest here. There were more regressions out of the target region in contexts biasing to HFON ($M = 20\%$, $SE = 2$) compared with the neutral contexts ($M = 15\%$, $SE = 2$), regardless of whether the target region contained the experimental or control word ($z = 2.12$, $p <$ .05). There were also fewer regressions out of the spillover region in items with contexts biasing toward the experimental word ($M =$ 12%, $SE = 2$) compared to neutral contexts ($M = 16\%$, $SE = 2$; $z = -2.23$, $p < .05$).

Contexts biasing to the experimental word elicited fewer regressions into the target word region (diff = 6%, $z = -2.69$, $p < .01$). Contexts biasing to the HFON elicited increased regressions into the pretarget region (diff = 9%, $z = 3.08$, $p < .01$). Finally, contexts biasing toward the HFON elicited significantly longer second pass reading times on several regions: 160 ms longer in the pretarget region ($t = 2.89$, $p < .05$) and 24 ms longer in the target region ($t = 2.79$, $p < .05$). There was also a marginal 21 ms second pass time penalty for contexts biased toward the HFON on the spillover region ($t = 1.71$, $p = .087$). In addition, there was a small, but significant, 10 ms second pass time advantage for contexts biased toward the experimental word on the target region ($t = -2.08$, $p < .05$). There were no effects of target word or context on the skipping rate on the target word or spillover, including main effects or interactions. No other effects were observed.

Table 4
Linear Mixed Effects Regression Models for Experiment 1

Measure	Region	Predictor	Estimate	Std. error	*t/z* value
First fixations	Target	(Intercept)	235.03	4.52	51.94*
		Experimental	−3.42	1.83	−1.87†
		Bias-Exp	−2.86	2.60	−1.10
		Bias-HN	1.29	2.57	0.50
		Experimental × Bias-Exp	−3.16	2.60	−1.22
		Experimental × Bias-HN	1.29	2.57	0.50
	Spill over	(Intercept)	233.25	3.45	67.61*
		Experimental	−0.58	1.59	−0.37
		Bias-Exp	−0.32	2.25	−0.14
		Bias-HN	1.76	2.24	0.79
		Experimental × Bias-Exp	−2.53	2.25	−1.13
		Experimental × Bias-HN	−0.82	2.24	−0.37
First pass	Target	(Intercept)	253.74	5.77	43.98*
		Experimental	−1.18	2.34	−0.50
		Bias-Exp	−4.56	3.33	−1.37
		Bias-HN	−1.99	3.29	−0.60
		Experimental × Bias-Exp	−1.75	3.34	−0.52
		Experimental × Bias-HN	1.85	3.30	0.56
	Spill over	(Intercept)	290.57	8.12	35.78*
		Experimental	−2.77	2.96	−0.94
		Bias-Exp	−2.42	4.17	−0.58
		Bias-HN	7.02	4.16	1.69†
		Experimental × Bias-Exp	−4.48	4.18	−1.07
		Experimental × Bias-HN	−3.23	4.17	−0.77
Go past	Target	(Intercept)	346.43	17.48	19.82*
		Experimental	17.71	8.40	2.11*
		Bias-Exp	−5.87	11.96	−0.49
		Bias-HN	17.22	11.81	1.46
		Experimental × Bias-Exp	−23.20	11.98	−1.94†
		Experimental × Bias-HN	25.86	11.82	2.19*
	Spill over	(Intercept)	384.97	18.80	20.47*
		Experimental	6.64	7.76	0.86
		Bias-Exp	−15.31	10.95	−1.40
		Bias-HN	17.42	10.94	1.59
		Experimental × Bias-Exp	−34.23	10.96	−3.12*
		Experimental × Bias-HN	13.56	10.97	1.24
Regressions out	Target	(Intercept)	−1.85	0.16	−11.28*
		Experimental	0.06	0.09	0.75
		Bias-Exp	−0.08	0.12	−0.68
		Bias-HN	0.25	0.12	2.12*
		Experimental × Bias-Exp	0.05	0.12	0.40
		Experimental × Bias-HN	0.18	0.12	1.48
	Spill over	(Intercept)	−1.86	0.12	−15.37*
		Experimental	0.01	0.08	0.06
		Bias-Exp	−0.28	0.12	−2.23*
		Bias-HN	0.18	0.11	1.58
		Experimental × Bias-Exp	−0.25	0.12	−2.02*
		Experimental × Bias-HN	0.17	0.11	1.45
	Final	(Intercept)	−1.77	0.15	−12.16*
		Experimental	−0.15	0.07	−2.01*
		Bias-Exp	−0.06	0.11	−0.55
		Bias-HN	0.03	0.10	0.28
		Experimental × Bias-Exp	−0.08	0.11	−0.78
		Experimental × Bias-HN	0.08	0.10	0.77
Regressions in	Pretarget	(Intercept)	−0.10	0.18	−0.54
		Experimental	−0.05	0.06	−0.80
		Bias-Exp	−0.06	0.09	−0.64
		Bias-HN	0.27	0.09	3.08*
		Experimental × Bias-Exp	−0.24	0.09	−2.78*
		Experimental × Bias-HN	0.26	0.09	2.95*

(*table continues*)

Table 4 (*continued*)

Measure	Region	Predictor	Estimate	Std. error	*t/z* value
	Target	(Intercept)	−1.49	0.12	−12.70*
		Experimental	0.02	0.08	0.33
		Bias-Exp	−0.30	0.11	−2.69*
		Bias-HN	0.18	0.10	1.70†
		Experimental × Bias-Exp	−0.17	0.11	−1.50
		Experimental × Bias-HN	0.16	0.10	1.51
	Spill over	(Intercept)	−1.57	0.15	−10.62*
		Experimental	−0.10	0.07	−1.33
		Bias-Exp	−0.06	0.10	−0.62
		Bias-HN	0.08	0.10	0.76
		Experimental × Bias-Exp	−0.14	0.10	−1.31
		Experimental × Bias-HN	0.19	0.10	1.83†
Second pass	Pretarget	(Intercept)	524.64	60.76	8.63*
		Experimental	0.67	17.78	0.04
		Bias-Exp	17.16	25.13	0.68
		Bias-HN	72.70	25.15	2.89*
		Experimental × Bias-Exp	−114.65	25.13	−4.56*
		Experimental × Bias-HN	97.18	25.16	3.86*
	Target	(Intercept)	111.54	12.34	9.04*
		Experimental	7.05	4.98	1.42
		Bias-Exp	−14.65	7.03	−2.08*
		Bias-HN	19.36	7.04	2.75*
		Experimental × Bias-Exp	−30.49	7.04	−4.33*
		Experimental × Bias-HN	20.81	7.04	2.96*
	Spillover	(Intercept)	134.14	15.53	8.64*
		Experimental	−4.07	5.40	−0.75
		Bias-Exp	−4.68	7.63	−0.61
		Bias-HN	13.09	7.64	1.71†
		Experimental × Bias-Exp	−27.35	7.63	−3.58*
		Experimental × Bias-HN	21.77	7.64	2.85*

† Indicates a theoretically interesting trend. * Denotes significances at the .05 criterion.

Discussion

There was clear evidence that, regardless of contextual bias, words with a HFON were associated with slower reading on the target region than controls, consistent with previous research (e.g., Grainger, O'Regan, Jacobs, & Segui, 1989; Paterson et al., 2009; Slattery, 2009). A strong version of the biased misperception hypothesis predicts that any, or at least the majority, of penalties for incongruent pretarget contexts would be observed only after the disambiguation region, as subjects would have initially misperceived the target as its HFON. As predicted by this account, there were effects of both pretarget contextual congruency and incongruency in rereading measures.

However, a differential effect of pretarget context was also observed prior to disambiguation in go past times. In particular, there was a congruency advantage for experimental words in supporting contexts, but an incongruency penalty when the pretarget context supported the HFON. The overall pattern is consistent with a model in which competition-inhibition from a HFON sometimes results in delays of forward progression through text, and that such competition-inhibition is modulated by top-down contextual information.

Experiments with ex-Gaussian analyses. As in previous studies, Experiment 1 failed to find interactions with contextual bias on the target or spillover region in first fixation or first pass measures. While these studies relied on the standard analysis comparing mean values between conditions, an early effect of contextual bias might nonetheless be detected in an ex-Gaussian analysis, which offers a wide range of measures beyond the mean with which to explore response time patterns. The following two eye tracking studies were designed to further explore the timing of the contextual congruency and incongruency effects with an ex-Gaussian analysis of fixations on words with HFONs.

The ex-Gaussian distribution. Response time studies have long favored the mean as the primary measure of central tendency (Balota & Yap, 2011 for review). Nevertheless, the distribution of response time values can change without altering the mean value, and an overreliance on the mean has the potential to obscure important effects. Formally, an ex-Gaussian distribution is a mathematical integration (i.e., a convolution) of a normal and an exponential distribution (Burbeck & Luce, 1982; Luce, 1986; Ratcliff, 1979; Ratcliff & Murdock, 1976). Three parameters make up the ex-Gaussian distribution: the mode (represented as mu or "μ") and the standard deviation (represented as sigma "σ") together characterize the normal distribution component, and a measure of skew (represented as tau "τ"), which characterizes the exponential component. As the mean can be derived from the algebraic sum of μ and τ, if a manipulation equally affects the mode of the normal distribution μ and the exponential distribution τ in opposing directions, for example, by reducing μ and increasing τ by the same amount, the effect would therefore cancel out effects observable in the mean (as observed in color naming facilitation in the Stroop task, e.g., Heathcote, Popiel, & Mewhort, 1991; Spieler, Balota, & Faust, 1996). Studying the underlying components of the ex-

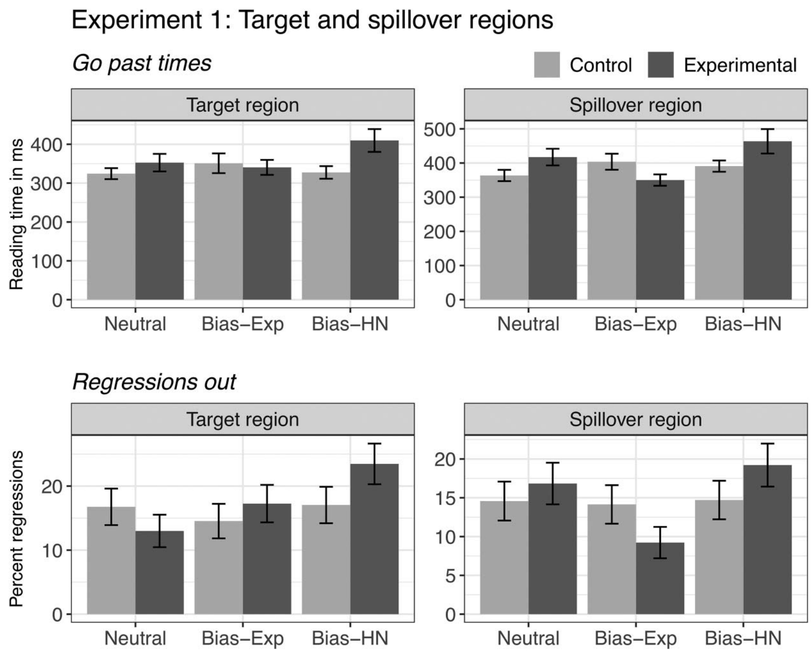

Figure 2. Experiment 1: Go past times and the percentage of regressions out on the target and spillover regions.

Gaussian distribution thus has the potential to reveal effects that would otherwise be masked (Balota, Yap, Cortese, & Watson, 2008).

In language processing research, ex-Gaussian analyses were initially conducted in single word recognition and priming studies, but have since been added to the arsenal of eye movement analyses (Staub, White, Drieghe, Hollway, & Rayner, 2010). Subsequent reading studies have shown that parameters in the ex-Gaussian distribution of response latencies are sensitive to distinct factors or properties of the stimulus, including *frequency* (Reingold, Reichle, Glaholt, & Sheridan, 2012; Staub et al., 2010), *predictability* (Sheridan & Reingold, 2012b; Staub, 2011; Staub & Benatar, 2013), *lexical ambiguity* (Sheridan & Reingold, 2012a), and *print quality* (White & Staub, 2012). Notably, different manipulations appear to affect distinct parameters of the distributions. Whereas increased word frequency increases both μ and τ (Staub et al., 2010), shifting the entire distribution rightward, predictability affects only μ (Staub, 2011).

There are, arguably, two main drawbacks to this method. The first is that analysis of the ex-Gaussian distribution requires a large number of observations to obtain stable parameter estimates (Heathcote, Brown, & Mewhort, 2002; Speckman & Rouder, 2004). In order to obtain such a large number of observations per participant, previous studies have presented all items from all conditions in each experimental session (e.g., Staub, 2011; Staub et al., 2010). The second is that early reading measures, such as first fixation and first pass times, are known to be highly sensitive to minute differences in lexical level qualities (Rayner, 1998 for

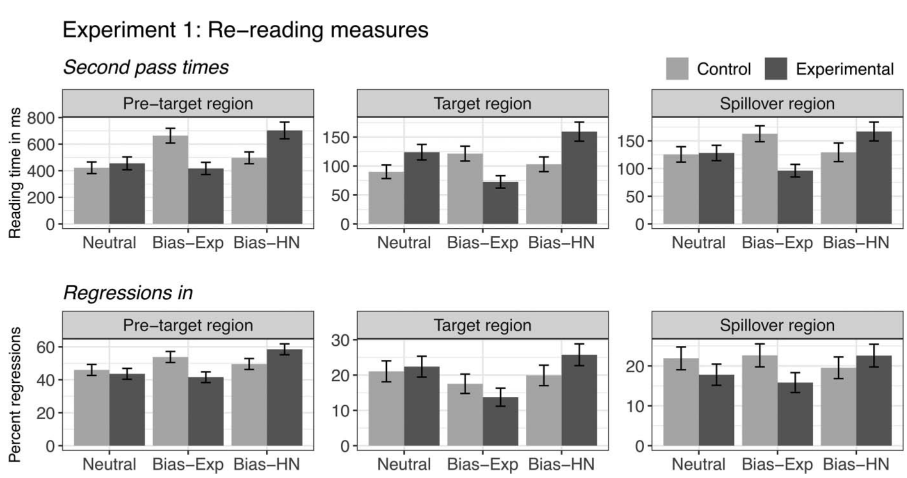

Figure 3. Experiment 1: Second pass times and percentage of regressions in on regions of interest.

review), which are likely to overshadow the distribution of early eye movement measures. To avoid this issue, the distribution is calculated on exactly the same linguistic content, preventing a direct comparison between a word with a HFON and a control word, as in Experiment 1.

We examined the contextual *incongruency* and *congruency* effects of context on a word with a HFON in two separate studies with ex-Gaussian analyses. Experiment 2 was designed to further test the contextual incongruency effect induced by pretarget contexts biased toward the HFON. Experiment 3 explored the contextual congruency effect, by determining whether the inhibition observed in Experiment 2 would be reduced or eliminated in contexts biasing toward the experimental word. We predicted that pretarget contextual information would affect the ex-Gaussian distribution in opposite directions prior to posttarget disambiguation. On the one hand, *incongruent* contexts that bias toward the HFON should increase lexical competition, slowing word recognition and increasing the rightward skew of fixation durations. On the other, *congruent* contexts that bias toward the experimental word should reduce lexical competition, resulting in faster word recognition and decreased skew.

Materials and methods. Materials for two eye tracking while reading experiments were created. Each experiment consisted of two conditions, which included a neutral pretarget context control (4a) that was compared against a contextual bias condition. The pretarget context biased toward a HFON in Experiment 2 (4b) and toward the experimental word in Experiment 3 (4c). The neutral contexts were made as similar to Experiment 1 as possible, with slight modifications when necessary to ensure that at least one to two pretarget words matched the biased context. Frames had identical posttarget one to two word continuations (*would be*), consistent with Experiment 1 whenever possible, followed by unique disambiguating material. Analysis regions are demarcated by a pipe symbol "|" and labeled with regions of interest. Occasionally, the disambiguation material changed across experiments in order to better fit the pretarget context, but such cases were kept to a minimum. Forty-eight such pairs were constructed in Experiment 2, and 40 in Experiment 3. A complete set of items for both experiments is provided in the online supplementary materials. Unlike Experiment 1, the disambiguating region always biased toward the target word.

(4) *Sample item from Experiments 2 and 3. The HFON is in parentheses.*

a. *Neutral:* $|_{\text{Initial}}$ His teacher quickly said that his right $|_{\text{Target}}$ ankle (angle) $|_{\text{Spillover}}$ would be $|_{\text{Final}}$ swollen in a couple days.

b. *Bias-HN:* $|_{\text{Initial}}$ The geometry teacher said that his right $|_{\text{Target}}$ ankle (angle) $|_{\text{Spillover}}$ would be $|_{\text{Final}}$ fixed within a week.

c. *Bias-Exp:* $|_{\text{Initial}}$ The doctor said that the swelling in his $|_{\text{Target}}$ ankle (angle) $|_{\text{Spillover}}$ would be $|_{\text{Final}}$ fixed within a week.

As in Experiment 1, experimental words and their HFONs were matched for length, first letter, syntactic category, number of phonemes, number of syllables, and number of morphemes. As intended, the log frequency of the target word was significantly lower than its HFON according to HAL and SUBTLEX measures. See Table 5 for the lexical characteristics of items from Experiments 2–3.

All conditions were presented to each subject in order to obtain the observations needed for an ex-Gaussian analysis. Two pseudorandomized lists were used to ensure adequate distance between different conditions from the same item, and conditions were separated into two distinct, counterbalanced blocks, such that the biased to target condition of an item was never in the same block as the neutral condition of the same item. An equal number of subjects were assigned one of two lists, which differed only in the order in which the blocks were presented. To avoid overtaxing participants while achieving the high number of observations required, no filler sentences were included (as in related studies, e.g., Staub, 2011; Staub et al., 2010). Participants were presented with eight practice items to familiarize themselves with the procedure. Comprehension questions were presented after half of the sentences.

Norming studies. Two separate continuation ratings study testing the naturalness of the experimental word and its HFON in the pretarget context were conducted using the same method and procedure as Experiment 1, summarized in Figure 1 and Table 2. For items in Experiment 2 (N = 32), the HFON was rated as a more natural continuation in contexts biased to the HFON than the experimental word, $t_2(47) = 9.02$, $p < 0.001$. In contrast, there were no differences between the continuation words in neutral sentence contexts. For items in Experiment 3 (N = 41), the experimental word was rated as a more natural continuation in contexts biased to the experimental word than its HFON, $t_2(39) = 9.75$, $p < 0.001$. Again, the differences between the continuations in neutral sentence contexts were not significant. The norming studies confirmed that biasing contexts were more compatible with the intended word than the relevant orthographic neighbor and that neutral contexts were compatible with either continuation.

In addition, two variants of the cloze task (Taylor, 1953) were conducted post hoc to estimate the predictability of the target word in each condition. The first variant (N = 35) employed a standard cloze design, in which subjects provide the next word in a sentence fragment, for example, *His doctor*

Table 5
Lexical Characteristics for Experiments 2–3

Measure	Experiments 2 Target	Experiments 2 HFON	Experiment 3 Target	Experiment 3 HFON
Length	5.05 (0.11)	5.08 (0.12)	5.05 (0.11)	5.05 (0.11)
Log HAL	8.08 (0.2)	10.47 (0.15)	8.08 (0.20)	10.45 (0.15)
Log SUBTLEX	2.47 (0.09)	3.44 (0.07)	2.47 (0.09)	3.43 (0.07)
Phonemes	4.10 (0.14)	4.15 (0.14)	4.10 (0.14)	4.10 (0.14)
Syllables	1.20 (0.06)	1.27 (0.07)	1.20 (0.06)	1.25 (0.07)
Morphemes	1.05 (0.03)	1.05 (0.03)	1.05 (0.03)	1.05 (0.03)

Note. Only measures of word frequency (log HAL and log SUBTLEX) differed between conditions.

quickly said that his right __________. Misspellings were corrected and morphologically related forms were counted as successes as long as the stem of the word provided was identical to the target or HFON. Overall, cloze probabilities were low, but showed sensitivity to the manipulation. Subjects provided the target word in 8% of the time in contexts biased toward the target (range: 0–50% of target words), but rarely or never provided target words in neutral contexts (0% of target words) or contexts biased to the HFON (range: 0–20% of target words). Similarly, subjects provided the HFON word in 7% of the time in contexts biased toward the HFON (range: 0–90% of HFON words), but rarely or never provided target words in neutral contexts (range: 0–10% of HFON words) or contexts biased to the target. Although the low cloze values suggest that our contexts did not impose highly restrictive constraints on lexical identity of the continuation word, the sentence completion ratings indicate that critical words were greatly preferred over experimental alternatives as continuations to the sentence fragment.

In the second variant of the cloze task ($N = 31$), the first letter or consonant cluster (invariant between the target and the HFON) was provided for the subject to complete with the first natural word that could continue the fragment, for example, *His doctor quickly said that his right an*__________. We reasoned that the first letter or consonant cluster would be readily available in parafoveal preview prior to the critical word, and would serve to constrain estimates about word identify before the word itself was fixated (Johnson, Perea, & Rayner, 2007; White, Johnson, Liversedge, & Rayner, 2008). As both the target word and its HFON shared the initial letter of consonant cluster, it is unlikely that any lexical bias was introduced favoring one completion over the other, at least at the word level. There was a clear asymmetry showing that biasing contexts provided moderate constraint toward the expected word. For Experiment 2, the experimental word was provided 34% of the time in supporting contexts, 1% of the time after contexts biasing toward the HFON, and 2% of the time after neutral contexts. For Experiment 3, the HFON was provided 42% of the time when appearing after contexts biasing toward the HFON, 1% of the time after contexts biasing toward the target, and 8% of the time after neutral contexts.

Participants. For each experiment, a disjoint set of 50 self-reported native speakers of English from the University of California, Los Angeles participated for one course credit in sessions lasting no more than 30 min. All participants scored over 80% on comprehension questions, with an average performance of 91% correct in Experiment 2 and 95% correct in Experiment 3. All participants had normal or corrected-to-normal vision.

Procedure. Eye movements were recorded as before on the same equipment and procedure, except that they were presented on a 19" Mitsubishi Diamond Pro 900u flat-screen CRT monitor (35.56 cm width × 27.94 cm height) at 1,024 × 768 resolution set to a 170 Hz refresh rate. Approximately three characters subtended 1 degree of visual angle. Data were cleaned of blinks, long fixations, and track losses automatically using University of Massachusetts, Amherst software. In both experiments, less than 5% of the data were removed (Experiment 2: 133 trials removed, 4,667 remaining; Experiment 3: 161 trials removed, 3,839 remaining).

Experiment 2

Experiment 2 examined the contextual *incongruency* effect—that is, the effect of context biasing toward a HFON, theoretically increasing the possibility of misperception or delay due to lexical competition-inhibition. We predicted that contexts compatible with a HFON would disrupt reading even before the reader encountered disambiguating text, as the reader attempts to resolve the incompatibility between her prior expectations and the lexical input. An ex-Gaussian analysis permits investigation into how disruption might affect different parameters of the distribution. A shift in the entire distribution would result in an increased or decreased value for the μ parameter, whereas a selective change in the right tail of the distribution would result in an increased τ parameter. Assuming a theoretical account of biased misperception in which a HFON is recognized as familiar more quickly (increasing forward progression until encountering inconsistent information), fixations on the target word should be reduced when context biases toward the HFON, resulting in a decreased μ parameter and reduced reading times on the experimental word. However, if a HFON can also result in delay due to increased postlexical competition-inhibition on select trials, then contexts biased toward the HFON should be associated with a longer right tail on the target word or the spillover region, increasing the τ parameter.

Method

Standard eye movement measures. Results are presented in Table 6. All reading measures were fit to (logistic) linear mixed effects models with random slopes and intercepts for subjects and items using sum coding, in which the neutral condition was treated as the statistical baseline; see Table 7. Outlier removal was conducted as in Experiment 1. The order of presentation was included as an additive and interactive fixed-effect predictor in separate models. In all cases, the second occurrence of the experimental word was associated with faster reading or decreased regressions into a region, resulting in significantly better model fits. However, the order of presentation did not eliminate or interact with the

Table 6
Means and Standard Errors for Eye Movement Measures in Experiment 2

		Regions			
Measure	Condition	Pretarget	Target	Spill over	Final
First fixation	Neutral	206 (1)	245 (1)	238 (1)	234 (1)
	Bias-HN	219 (1)	246 (1)	242 (1)	237 (1)
First pass	Neutral	1082 (8)	265 (2)	308 (3)	1005 (9)
	Bias-HN	1153 (8)	268 (2)	314 (3)	955 (9)
Go-past	Neutral	—	323 (5)	414 (4)	—
	Bias-HN	—	347 (5)	445 (5)	—
Second pass	Neutral	295 (12)	85 (4)	86 (4)	—
	Bias-HN	367 (13)	105 (4)	98 (4)	—
Regressions out	Neutral	—	12% (1)	14% (1)	35% (1)
	Bias-HN	—	15% (1)	18% (4)	41% (1)
Regressions in	Neutral	34% (1)	19% (1)	16% (1)	—
	Bias-HN	42% (1)	22% (1)	18% (1)	—

Note. Means for go past measures are identical to first pass reading on the initial region and have been omitted.

Table 7
Linear Mixed Effects Regression Models for Experiment 2

Measure	Region	Predictor	Estimate	Std. error	*t/z* value
First fixation	Target	(Intercept)	245.16	3.57	68.70*
		Bias-HN	0.03	0.95	0.04
	Spillover	(Intercept)	239.37	3.29	72.80*
		Bias-HN	2.21	0.86	2.56*
First pass	Target	(Intercept)	265.15	4.56	58.19*
		Bias-HN	1.51	1.21	1.24
	Spillover	(Intercept)	307.38	7.38	41.68*
		Bias-HN	3.07	1.67	1.83
Go past	Target	(Intercept)	331.75	11.13	29.81*
		Bias-HN	11.28	3.44	3.28*
	Spillover	(Intercept)	420.18	18.86	22.28*
		Bias-HN	15.39	4.23	3.64*
Regressions out	Target	(Intercept)	−2.07	0.13	−16.07*
		Bias-HN	0.13	0.05	2.66*
	Spillover	(Intercept)	−1.84	0.11	−16.17*
		Bias-HN	0.14	0.04	3.27*
	Final	(Intercept)	−0.60	0.18	−3.39*
		Bias-HN	0.16	0.03	4.75*
Regressions in	Pretarget	(Intercept)	−0.58	0.18	−3.21*
		Bias-HN	0.24	0.03	6.85*
	Target	(Intercept)	−1.53	0.11	−14.27*
		Bias-HN	0.10	0.04	2.57*
	Spillover	(Intercept)	−1.77	0.13	−14.16*
		Bias-HN	0.09	0.04	2.14*
Second pass	Pretarget	(Intercept)	331.10	48.81	6.78*
		Bias-HN	36.31	7.21	5.04*
	Target	(Intercept)	95.47	11.30	8.45*
		Bias-HN	9.91	2.55	3.89*
	Spillover	(Intercept)	91.98	11.47	8.02*
		Bias-HN	5.84	2.43	2.40*

* Denotes significance at the .05 level.

effect of context, and is not presented here. All significant effects are reported.

Prior to disambiguation. There were significantly longer go past times on target words following a context biased to the HFON (diff = 24 ms, $t = 3.28$, $p < .01$), as well as more regressions out from the target word (diff = 3%, $z = 2.66$, $p < .01$), compared with the neutral context. In the spillover region, items with contexts biased to the HFON elicited longer first fixation durations (diff = 4 ms, $t = 2.56$, $p < .05$) and go past times (diff = 31 ms, $t = 3.64$, $p < .001$). There were also more regressions out of the spillover region in contexts biased to the HFON compared with neutral contexts (diff = 4%, $z = 3.27$, $p < .01$). Subjects skipped the target word at exactly the same rate ($M = 21\%$) regardless of contextual bias.

After disambiguation. In the final, disambiguating region, there were more regressions out to previous regions (diff = 6%, $z = 4.75$, $p < .001$) compared with neutral contexts. In second-pass measures, there was a persistent cost for pretarget contexts supporting the HFON in all regions prior to the final region: a 72 ms cost in the pretarget region, $t = 5.04$, $p < .001$; a 20 ms cost in the target region, $t = 3.89$, $p < .001$; and a 12 ms cost in the spillover region, $t = 2.40$, $p < .05$.

A median split analysis on the number of regressions out of the final region was conducted as in Experiment 1. There was no effect on the first fixation durations or first pass times on the target or spillover region. In contexts biased toward the HFON, the low regression group ($N = 25$, fewer than 32% regressions out) elicited shorter go pass times on the target ($\hat{\beta} = -6.76$, $SE = 3.43$, $t = -1.97$, $p < .05$) and spillover ($\hat{\beta} = -7.99$, $SE = 4.23$, $t = -1.89$, $p = .06$) region.

Predictability effects. To determine if differences in predictability could account for the reading patterns reported above, standard cloze values for both the target and its HFON (obtained from the first variant of the cloze norming task) were added as a fixed effect predictor to the contextual manipulation in the linear mixed regression model for each reading time measure. Random effect structures were modeled with by-subjects and by-items random intercepts after several random slope models failed to converge. Models with the additional predictors were compared against those reported above, and evaluated for model fit in terms of AIC scores. Adding cloze values produced significantly better models on some measures. On the target word, higher cloze values for the HFON resulted in longer first pass reading times ($\hat{\beta} = 51.81$, $SE = 21.77$, $t = 2.38$, $p < .05$). A similar effect of higher cloze values toward the HFON was observed on first fixation durations in the spillover region ($\hat{\beta} = 73.49$, $SE = 27.75$, $t = 2.65$, $p < .01$), interacting with the contextual bias predictor by significantly reducing the effect of context ($\hat{\beta} = -89.36$, $SE = 25.98$, $t = -3.44$, $p < .001$). In these models, neither the cloze value of the target word nor the general effect of contextual bias was significant. In all other measures, models with cloze values did not fit the data better.

Ex-Gaussian analysis. Using the QMPE software Version 2.18 (Cousineau, Brown, & Heathcote, 2004; Heathcote et al., 2002), we conducted by-subjects ex-Gaussian analyses of the empirical distribution of first fixation durations and first pass times on the target word and the spillover region.[5] Parameters for all data converged based on exit codes provided by the QMPE v.2.18 manual (Brown, Cousinau, & Heathcote, n.d.). All first fixation and first pass fixations were included in the ex-Gaussian analysis.

There was an average of 37 fixations on the target and 42 fixations on the spillover regions per subject. The manipulation did not affect ex-Gaussian parameters in first fixation durations or in first pass times on the target region. However, in the spillover region, there was a shift in ex-Gaussian parameters in first fixation durations: Contexts biased toward the HFON elicited marginally decreased values for μ, $t_1(49) = -1.88$, $p = .06$, and significantly decreased values for σ, $t_1(49) = -2.39$, $p < .05$, but resulted in longer values for the skew τ, $t_1(49) = 2.90$, $p < .01$, compared with neutral contexts (see Figure 4). No differences in first pass times were observed. Table 8 summarizes the results.

Discussion

Pretarget contexts were biased to the HFON of the target word, thereby creating a mismatch between high-level, top-down contextual bias and the lexical input. Although the mismatch disrupted reading, as predicted by the biased misperception hypothesis, a reading penalty for this mismatch was detected on the spillover region in first fixations, first pass, go past, and regressions out measures even before the region containing disambiguating information was reached. In the ex-Gaussian analysis, the mismatch produced effects in first fixation times on the spillover region that patterned in opposite directions for different parts of the distribution: a reduction for the normal component (the μ and σ parameters), but an increase in rightward skew (the τ parameter). Although the reduction in μ is compatible with a strong version of biased misperception, the fact that readers also spent more time on regions before disambiguation is not. While these two findings initially seem incompatible, they may instead reveal two distinct kinds of eye movement responses to contextual mismatch, depending on whether the reader regressed from the region or not.

To assess this possibility, a post hoc regression-contingent analysis of first fixation durations on the spillover region was conducted (Altmann, Garnham, & Dennis, 1992).[6] As expected, trials without first pass regressions out of the spillover region, accounting for over 80% of the trials, patterned closely with the analysis conducted on all the data: contexts biased to the HFON exhibited a marginally decreased μ, $t_1(49) = -1.91$, $p = .06$; a significantly decreased σ, $t_1(49) = -2.47$, $p < .05$ parameters; and an increase in τ, $t_1(49) = 2.46$, $p < .05$. However, including only the trials with first pass regressions from the spillover region revealed that contexts biased to the HFON significantly increased μ, $t_1(44) = 2.05$, $p < .05$; marginally increased σ, $t_1(44) = 1.89$, $p = .07$; but did not affect tau, $t < 1$. The pattern indicates that the increase in first fixation durations associated with contexts biased toward the HFON was due to an increase in τ in the spillover region, but only for cases in which the reader did not regress. On the subset of trials in which readers regressed out of the spillover, readers appeared to have dwelled longer on the region, suggesting two types of eye movement responses to incongruous contexts.

In any event, pretarget contexts biased toward a HFON disrupted forward progression immediately after the experimental word was encountered. These results cast doubt on the explanation that the observed penalties are *solely* due to an initial misperception of the input; the results instead suggest that such delays may sometimes reflect postlexical competition-inhibition between the

Table 8
Ex-Gaussian Analysis of First Fixation Durations and First Pass Times of the Target Word and Spill Over Region in Experiment 2 in Milliseconds

Measure	Condition	μ	σ	τ
		Target		
First fixation	Neutral	195.47	41.20	53.78
	Bias-HN	190.63	42.99	59.24
	Context effect	−4.84	1.79	5.46
	t-value	−1.07	0.53	1.25
First pass	Neutral	186.73	39.04	86.12
	Bias-HN	185.60	42.08	90.01
	Context effect	−1.13	3.04	3.89
	t-value	−0.26	0.82	0.74
		Spillover		
First fixation	Neutral	184.39	36.18	57.53
	Bias-HN	177.56	30.45	69.56
	Context effect	−6.83	−5.73	12.03
	t-value	−1.88†	−2.39*	2.90*
First pass	Neutral	190.98	46.62	123.71
	Bias-HN	190.66	44.59	130.46
	Context effect	−0.32	−2.03	6.75
	t-value	−0.44	−0.37	0.88

† Indicates a nonsignificant trend. * Denotes significance in a paired *t*-test at the .05 level.

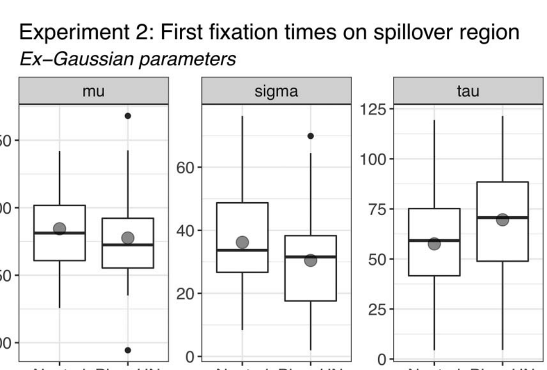

Figure 4. Experiment 2: Box and whisker plots for ex-Gaussian parameters obtained from QMPE for first fixation durations on the spillover region. Mean values are depicted as grey circles.

[5] Some studies report ex-Gaussian analyses of first pass durations (e.g., Hoedemaker & Gordon, 2017; Staub et al., 2010; White & Staub, 2012), whereas others do not (e.g., Staub, 2011; Staub & Benatar, 2013). For present purposes, we remain agnostic as to whether a 'composite' measure like first pass time, which may consist of a mixture of single and multiple fixations, is amenable to an ex-Gaussian analysis, and simply report both first fixation and first pass measures.

[6] We thank a reviewer for suggesting the regression-contingent analysis. As not all readers regressed from the spillover region, fewer subjects (45 out of 50) were included in the analysis of trials with regressions.

word and a more frequent neighbor, and that, in our materials, the resulting conflict tended to be resolved before the eye progressed through text.

Experiment 3

Experiment 3 further explores the time course of the *contextual congruency effect* observed in Experiment 1, comparing the effect of pretarget contexts biasing toward the experimental word with neutral contexts. We predicted that pretarget contexts biased toward the experimental word would elicit a reading time advantage before disambiguating information was encountered, indicating that pretarget contextual constraint can reduce the inhibitory effect of a HFON. If our interpretation of the increased τ parameter in Experiment 2 is correct, pretarget contexts biasing toward the experimental word should reduce the amount of rightward skew, decreasing the τ parameter, compared with neutral contexts.

Method

Standard eye movement measures. Results are presented in Table 9. The analysis, outlier removal, and model evaluation procedure were conducted as in Experiment 2. The order of presentation again did not eliminate or interact with the effect of context. All significant effects are reported in Table 10.

Prior to disambiguation. Prior to disambiguation, contexts biasing to the experimental word were associated with faster reading times in several regions in multiple reading measures. On the target word, contexts biasing toward the experimental word elicited a 5 ms advantage in first fixation durations, $t = -2.67$, $p < .01$, and a 8 ms advantage in first pass times, $t = -2.28$, $p < .05$. In the spillover region, a similar advantage was observed in go past times, diff = 44 ms, $t = -4.88$, $p < .001$; and regressions out, diff = 7%, $z = -6.39$, $p < .001$. Finally, readers skipped the target word marginally more often in first pass reading in contexts biased toward the experimental word ($M = 24\%$, $SE = 1$) compared with neutral contexts ($M = 21\%$, $SE = 1$), $\hat{\beta} = 0.19$, $SE = 0.10$, $z = 1.88$, $p = .06$.

Table 9
Means and Standard Errors for Eye Movement Measures in Experiment 3

Measure	Condition	Pretarget	Target	Spillover	Final
		Region			
First fixation	Neutral	200 (1)	242 (2)	234 (1)	229 (1)
	Bias-Exp	209 (1)	237 (2)	235 (1)	231 (1)
First pass	Neutral	1123 (9)	270 (2)	312 (3)	572 (6)
	Bias-Exp	1243 (9)	262 (2)	310 (3)	610 (6)
Go-past	Neutral	—	342 (5)	437 (5)	—
	Bias-Exp	—	333 (5)	393 (5)	—
Second pass	Neutral	368 (15)	111 (5)	109 (5)	—
	Bias-Exp	308 (14)	64 (3)	90 (4)	—
	Bias-Exp	168 (11)	27 (2)	61 (4)	—
Regressions out	Neutral	—	17% (1)	18% (1)	19% (1)
	Bias-Exp	—	17% (1)	11% (1)	20% (1)
Regressions in	Neutral	41% (1)	23% (1)	21% (1)	—
	Bias-Exp	36% (1)	14% (1)	19% (1)	—

Note. Means for go past measures are identical to first pass reading on the pretarget region and have been omitted.

Table 10
Linear Mixed Effects Regression Models for Experiment 3

Measure	Region	Parameter	Estimate	Std. error	t/z value
First fixation	Target	(Intercept)	239.04	3.24	73.80*
		Bias-Exp	−2.82	1.06	−2.67*
	Spillover	(Intercept)	233.36	3.20	72.90*
		Bias-Exp	0.10	0.88	0.11
First pass	Target	(Intercept)	262.69	4.84	54.22*
		Bias-Exp	−4.77	1.45	−3.28*
	Spillover	(Intercept)	307.53	8.18	37.59*
		Bias-Exp	−1.37	1.81	−0.76
Go past	Target	(Intercept)	333.19	11.31	29.46*
		Bias-Exp	−5.91	3.58	−1.65
	Spillover	(Intercept)	408.57	16.63	24.56*
		Bias-Exp	−21.59	4.57	−4.72*
Regressions out	Target	(Intercept)	−1.81	0.15	−12.45*
		Bias-Exp	−0.01	0.05	−0.15
	Spillover	(Intercept)	−1.98	0.12	−15.89*
		Bias-Exp	−0.30	0.05	−5.93*
	Final	(Intercept)	−1.66	0.15	−10.85*
		Bias-Exp	0.01	0.04	0.20
Regression in	Pretarget	(Intercept)	−0.59	0.18	−3.19*
		Bias-Exp	−0.13	0.04	−3.38*
	Target	(Intercept)	−1.61	0.11	−15.30*
		Bias-Exp	−0.31	0.05	−6.52*
	Spillover	(Intercept)	−1.59	0.13	−11.88*
		Bias-Exp	−0.07	0.04	−1.53
Second pass	Pretarget	(Intercept)	339.78	52.51	6.47*
		Bias-Exp	−29.04	8.34	−3.48*
	Target	(Intercept)	87.59	10.53	8.31*
		Bias-Exp	−23.60	2.60	−9.09*
	Spillover	(Intercept)	99.81	13.06	7.64*
		Bias-Exp	−9.78	2.70	−3.62*

* Denotes significances at the .05 criterion.

After disambiguation. Contexts biased toward the experimental word elicited fewer regressions into the pretarget (diff = 5%, $z = -3.38$, $p < .001$) and target (diff = 9%, $z = -6.52$, $p < .001$) region, as well as shorter second pass times in pretarget (diff = 60 ms, $t = -3.48$, $p < .001$), target (diff = 47 ms, $t = -9.09$, $p < .001$), and spillover (diff = 19 ms, $t = -3.62$, $p < .001$) regions.

Predictability effects. As before, cloze values for the target and the HFON were added as predictors to separate linear mixed effects regression models with by-subjects and by-items random intercepts for all relevant measures. The cloze value was a significant predictor for several reading measures, resulting in a better model fit in some cases. We only report better-fitting models that diverge from the findings reported above. In first pass times on the spillover region, higher cloze values for both the target ($\hat{\beta} = -106.49$, $SE = 54.23$, $t = -1.96$, $p < .05$) and the HFON ($\hat{\beta} = -19.24$, $SE = 9.52$, $t = -2.02$, $p < .05$) resulted in faster reading times. As before, an advantage for contexts biased to the target was observed in first pass times ($\hat{\beta} = -5.78$, $SE = 2.43$, $t = -2.57$, $p < .05$). No other effects of cloze predictability were observed.

Ex-Gaussian analysis. The ex-Gaussian parameters for all data converged in the QMPE software. There was an average of 30 fixations on the target and 35 fixations on the spillover region per subject. Consistent with the previous finding that more predictive contexts result in a leftward shift of the overall distribution (Sheridan & Reingold, 2012b; Staub, 2011), contexts biased toward the experimental word elicited a decrease of the μ parameter in both first fixation, $t_1(49) = -2.31$, $p < .05$, and first pass, $t_1(49) = -2.12$, $p < .05$, measures on the target word, shown in Figure 5, but had no significant effect on any other parameters in this region. For first past times in the spillover region, however, contexts biased toward the experimental word increased μ, $t_1(49) = 2.72$, $p < .01$, and σ, $t_1(49) = 2.68$, $p < .01$, but decreased τ, $t_1(49) = -2.87$, $p < .01$, shown in Figure 6. Results are summarized in Table 11.

Discussion

We interpret the eye movements results as indicating two central effects. First, there was a *predictability advantage* on the experimental word region when prior context biased toward the input, resulting in faster reading times, an increased skipping rate, and a leftward shift of the μ parameter of the ex-Gaussian distribution. Second, there was a *facilitatory effect of biasing context*, resulting in a reduced rightward skew in first pass times on the spillover region. While this region also showed an increase in both μ and σ parameters, we speculate that the increased richness of the biasing context may have slowed reading overall, as readers updated their semantic or discourse representations of the text with the predicted word. Note that such an effect would not be predicted in Experiment 2, as the prior context failed to predict the word that was actually encountered. It is unclear whether a strong version of biased misperception would make similar predictions.

In both experiments, more predictable continuations (as indexed by cloze values) appeared to reduce processing load. Crucially, however, the effect of contextual bias reported above remained significant in these models. Therefore, predictability alone cannot account for the effects we have attributed to postlexical competition-inhibition.

General Discussion

Previous research has observed that processing a word is strongly influenced by its orthographic neighbor in reading tasks. In word recognition and identification paradigms, words appear to benefit from larger Ns, especially if the task requires limited access to lexical level properties beyond the orthography. However, the facilitatory effect is modulated by the presence of a highly frequent orthographic neighbor, which appears to *inhibit* lexical access. These results are highly dependent on task (e.g., Andrews, 1997), and have not translated directly into studies of natural reading. Although several studies found a processing cost for words with HFONs, the effects were delayed until relatively late in the eye movement record, largely limited to measures of rereading (Gregg & Inhoff, 2016; Pollatsek et al., 1999; Slattery, 2009; Warrington et al., 2018). Previously observed patterns support the predictions of a strong version of biased misperception, in which words with HFONs are often initially misperceived as their HFON, eliciting disruptions only when encountering text that is inconsistent with the misperceived word. This interpretation accords very well with models of reading that allow the eye to skip a word when it is recognized in parafoveal preview or fixated on very briefly while

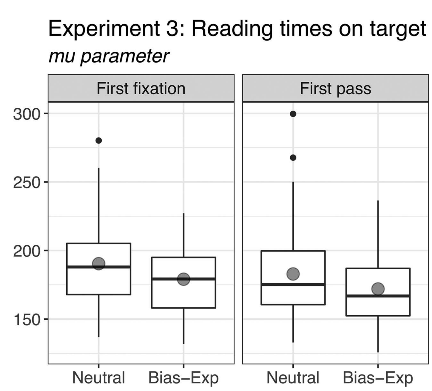

Figure 5. Experiment 3: Box and whisker plots for the μ parameter obtained from QMPE for first fixation and first pass measures on the target region. Mean values are depicted as grey circles.

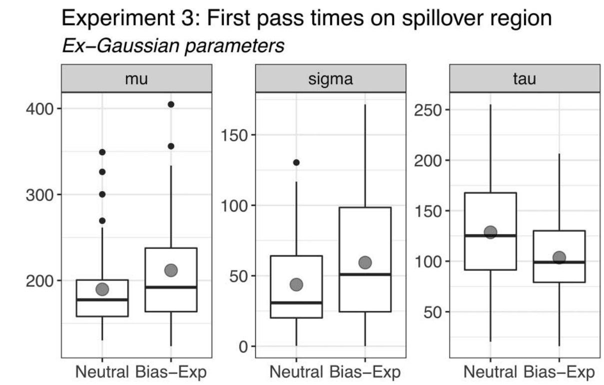

Figure 6. Experiment 3: Box and whisker plots for ex-Gaussian parameters obtained from QMPE for first pass times on the spillover region. Mean values are depicted as grey circles.

Table 11
Ex-Gaussian Analysis of First Fixation Durations and First Pass Times of the Target Word and Spillover Region in Experiment 3 in Milliseconds

Measure	Condition	μ	σ	τ
		Target		
First fixation	Neutral	190.38	38.10	55.35
	Bias-Exp	179.07	32.83	62.40
	Context effect	−11.31	−5.27	7.05
	t-value	−2.31*	1.45	1.50
First pass	Neutral	182.89	35.37	93.46
	Bias-Exp	171.98	29.43	93.94
	Context effect	−10.91	−5.94	0.48
	t-value	−2.12*	−1.38	0.08
		Spillover		
First fixation	Neutral	178.99	32.41	58.90
	Bias-Exp	181.30	31.41	57.30
	Context effect	2.31	−1.00	−1.60
	t-value	0.60	−0.45	−0.44
First pass	Neutral	189.66	43.71	128.61
	Bias-Exp	211.72	59.33	103.53
	Context effect	22.06	15.62	−25.08
	t-value	2.72*	2.68*	−2.87*

* Denotes significance in a paired by-subjects *t*-test at the .05 level.

an eye movement is being planned. In these studies, however, the disambiguating information was positioned directly after the critical word. This design makes it difficult to discern whether the processing cost manifested after the target word because the reader had initially misperceived the word and had only later encountered information which required correcting the misperception, or because postlexical competition-inhibition temporarily halted progression through the sentence.

The present experiments were intended to distinguish between these two possibilities, adding a spillover region on which to observe delays unrelated to correcting misperception. In the first study, we observed a cost for words with a HFON, but at an earlier time course than would be predicted by complete misperception alone. We also found a strong and persistent effect of context: Whereas contexts supporting the experimental word reduced the inhibition penalty early in processing, contexts supporting the HFON produced an additional penalty in somewhat delayed measures of reading. However, there was ample evidence that both effects also appeared before the reader encountered incompatible information, as effects were observed directly on the experimental word or the following spillover region.

The second two experiments were specifically designed to permit an ex-Gaussian analysis of the distribution of fixation durations on the critical word and the spillover region. Pretarget context was found to modulate the rightward skew of the distribution, increasing τ when the context biased toward the HFON and decreasing τ when it biased toward the target word. Together, the three studies support a model of reading in which context modulates postlexical competition-inhibition, whose effects appear on τ in a subset of trials.

However, we do not claim that context, in general, can or should only affect τ. Indeed, in Experiment 3, contexts biasing toward the experimental word decreased the μ parameter of the first fixation distribution on the word itself. This finding is compatible with previous studies on predictability and contextual constraint (Sheridan & Reingold, 2012b; Staub, 2011), and with the claim that supporting context can reduce mean reading times on early measures traditionally associated with lexical access (e.g., Duffy et al., 1988).[7]

Although our results ultimately argue against the strongest version of biased misperception, the results presented here, along with the wealth of previous literature and our own subjective experience, convinces us that readers sometimes do misperceive words. Misperception and delay from postlexical competition-inhibition may simply represent two different outcomes produced by the same underlying model. To illustrate, we take a two-stage cascaded model of lexical decision, in which a relatively automatic process of word recognition is followed by a slower attentional-strategic check, where lexical candidates are more closely discriminated against the signal (e.g., Balota & Spieler, 1999; Chumbley & Balota, 1984; Paap et al., 1982). The second, checking stage is presumably cancelled if word recognition completes within a scheduled deadline, in which case the oculomotor system stalls forward progression until the language processing system is ready to reinitiate normal routines (e.g., Engbert et al., 2005; Mitchell et al., 2008; Schad & Engbert, 2012). As checking would therefore only be engaged on a subset of trials, its effects are predicted to manifest primarily in the right tail of the distribution. This explanation is similar in spirit to Williams et al.'s (2006) discussion of how a two-stage model could be merged with E-Z Reader to produce early advantages for words with HFON, but a later inhibitory effect at a postlexical verification stage in eye movements.

Misperception would result when a HFON facilitates word recognition by virtue of its increased frequency. If word recognition occurs quickly enough, the oculomotor system cancels the initial stage of saccadic planning and allocates attention to another word, increasing the likelihood of misperception. As in previous accounts, disruptions in reading would appear after the reader encountered text that conflicted with the misidentified word. In contrast, postlexical delay would result when increased competition between lexical candidates prevents the initial stage of word recognition from completing, and the secondary checking mechanism is engaged, slowing forward progression until the correct word form is verified. Crucially, the HFON would be inhibited and there would be no misperception in such cases.

In this account, contextual bias toward the HFON could affect the extent that lexical candidates compete, but might elicit different outcomes depending on the strength of the context. Whereas *strong contextual bias* toward the HFON might result in early, but error prone, word recognition, *weak contextual bias* toward the HFON might instead increase competition enough to prevent word recognition from completing before postlexical checking and inhibition begins. The results of the current study clearly support the second outcome, as HFONs elicited processing delays prior to disambiguating information. In Experiment 2, contexts biased toward the HFON increased distributional skew, compatible with the claim that postlexical checking only affects a subset of trials. In contrast, contexts biasing toward the experimental target word reduced skew in Experiment 3, as would be expected if contextual support against the HFON reduced competition between word forms.

As there is consistent and robust evidence for a greater occurrence of misperception in very similar materials, it is quite possible that our pretarget contexts were less constraining than in other studies. As indicated by the cloze tasks, our materials were only moderately constraining, perhaps due to the more stringent lexical and syntactic constraints placed on the sentence fragment leading up to the target. On the basis of our items, readers may have adopted more conservative reading strategies that avoided generating strong predictions about word identity, which in turn might have led to fewer cases of misperception overall. Experiment 2 presented contexts designed to increase the likelihood that the reader would use prior contextual constraint to incorrectly identify the target word as the HFON competitor, possibly skipping the word as a result of an early familiarity check. Even in these cases, there was no evidence for increased word skipping on the critical word. The only effect on skipping rates was observed in Experiment 3, where the sentence fragment context biased toward the experimental word, and presumably made accurate identification of the word easier. As previous studies do not report cloze values

[7] More specifically, while contextual support in favor of a word might well speed lexical access on the majority of trials without a HFON, the effect of context on words with HFON in particular would depend on how contextual bias modulates post-lexical competition between the target word and its neighbor. In our case, the important finding is that context affected τ differently according to whether the experimental word or its HFON received contextual support.

of their materials, a more direct comparison is not possible. We expect that a careful comparison of the materials across experiments may well reconcile the differences between the two sets of findings.

If both misperception and competition-inhibition are possible outcomes from the same underlying set of processes, a number of additional factors may have contributed to the different findings across experiments. For example, readers may have employed different reading strategies based on proficiency level. This possibility is supported by Veldre and Andrews' (2015) display change study, in which HFONs in dense neighborhoods produced an early inhibitory effect among highly proficient readers, but an initial facilitation among less proficient readers, who were disrupted primarily during rereading. They argued that more precise representations resulted in deeper processing for skilled readers, which in turn generated early reading delays from lexical competition between the target and its HFON. In contrast, lower proficiency readers would be more likely to misperceive the target as its HFON, revising upon encountering incompatible information later in the sentence. Similarly, Perea and Pollatsek (1998) proposed that an "impulsive" reading strategy prompt early termination of word recognition by lowering the threshold for word recognition, allowing a forward eye movement to be programmed earlier (see also Pollatsek et al., 1999; Sears et al., 2006, and Warrington et al., 2018 for discussion of differences between readers).

Such an account also has the potential to unite the multiple reader-based and linguistic factors that have been argued to influence misperception within a unified mechanism. Readers who are less attentive or engage in riskier reading strategies might rely more on top-down contextual information in deciding how to progress through text. Such readers might be especially susceptible to biased misperception, as increased contextual support for a HFON could prompt early termination of word recognition processes, allowing a saccade to be programmed earlier. Conversely, the same contexts would increase lexical competition-inhibition when the threshold for word recognition remains relatively high, slowing reading until the correct word form is identified. The seemingly contradictory results from different studies might each reflect increased competition between a word and its HFON, resulting in either misperception or postlexical delay as a function of the word recognition threshold.

This account of misperception and postlexical delay naturally extends to other orthographically similar pairings. A highly related question addresses the effect of word neighbors that differ only in the order of two adjacent, transposed letters (TL), as in *clam* and *calm*. Words with TL neighbors elicit slower response times and higher error rates than controls in single word lexical decision tasks (Andrews, 1996; Chambers, 1979). They also elicit slower reading times when embedded in neutral sentence contexts (Acha & Perea, 2008; Johnson, 2009), an effect which is eliminated when pretarget context biases toward the word presented (Johnson, 2009; see also Luke & Christianson, 2012). Much like most research on HFONs, Johnson (2009) reported no effects in first pass or go past reading of the target word, instead finding a penalty in later measures (total times, refixations, regressions in, and second pass times) on the target word. However, the penalty appeared on the posttarget spillover region, regardless of whether the pretarget context was biased toward the word presented or was not, in keeping with the results reported here. Johnson, Staub, and Fleri (2012) found that words with TL neighbors elicited slower word naming latencies and resulted more errors (6.5% vs. 0.2% on controls), the majority (86%) of which produced the TL neighbor itself. An ex-Gaussian analysis of the distribution of naming latencies revealed that the effect was due to an increased rightward skew, affecting a subset of long trials, rather than affecting the distribution as a whole. The result was interpreted as evidence of misidentification of word identity, instead of global lexical competition between similar word forms.

Although this interpretation is plausible, it should be noted that the ex-Gaussian analysis excluded incorrect responses. If the increased skew were due to an actual misperception of the word, we would have expected the effect to have been limited to just those trials in which the TL neighbor word was erroneously produced. An alternative account is that the TL effect in skew reflects postlexical checking in which incorrect forms are inhibited on a subset of trials, whereas the effect in naming errors represents genuine misperception. Whatever the correct explanation is, the effect of TL and HFON neighbors, as well as other orthographically related forms, on word identification during reading is likely to draw on the same set of processing mechanisms.

In previous reading studies, the target and its HFON were contextually distinguished in a region that immediately followed the critical region. This kind of design cannot distinguish between two hypotheses regarding the influence of a HFON in normal reading: biased misperception or postlexical competition-inhibition between visually similar word forms. Our primary concern has been to provide evidence that the effects of a HFON may be observed prior to encountering posttarget disambiguation, which we argue reflects postlexical competition-inhibition rather misperception per se, though we have offered an explanation in which the two are fully compatible. Given the ample evidence that readers do sometimes misperceive text, a remaining challenge is to determine the circumstances in which postlexical competition-inhibition prevails.

References

Acha, J., & Perea, M. (2008). The effect of neighborhood frequency in reading: Evidence with transposed-letter neighbors. *Cognition, 108,* 290–300. http://dx.doi.org/10.1016/j.cognition.2008.02.006

Altmann, G. T. M., Garnham, A., & Dennis, Y. (1992). Avoiding the garden path: Eye movements in context. *Journal of Memory and Language, 31,* 685–712. http://dx.doi.org/10.1016/0749-596X(92)90035-V

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

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

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

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review, 4,* 439–461. http://dx.doi.org/10.3758/BF03214334

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of*

Memory and Language, 59, 390–412. http://dx.doi.org/10.1016/j.jml.2007.12.005

Balota, D. A. (1994). Visual word recognition: The journey from features to meaning. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 303–358). New York, NY: Academic Press.

Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General, 128,* 32–55. http://dx.doi.org/10.1037/0096-3445.128.1.32

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

Balota, D. A., Yap, M. J., Cortese, M. J., & Watson, J. M. (2008). Beyond mean response latency: Response time distributional analyses of semantic priming. *Journal of Memory and Language, 59,* 495–523. http://dx.doi.org/10.1016/j.jml.2007.10.004

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

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255–278. http://dx.doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1–48. http://dx.doi.org/10.18637/jss.v067.i01

Binder, K. S., & Morris, R. K. (1995). Eye movements and lexical ambiguity resolution: Effects of prior encounter and discourse topic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1186–1196. http://dx.doi.org/10.1037/0278-7393.21.5.1186

Binder, K. S., & Rayner, K. (1998). Contextual strength does not modulate the subordinate bias effect: Evidence from eye fixations and self-paced reading. *Psychonomic Bulletin & Review, 5,* 271–276. http://dx.doi.org/10.3758/BF03212950

Brown, S., Cousinau, D., & Heathcote, A. (n.d.). *Technical manual for QMPE v2.18: Fortran code to fit response time distributions*. Retrieved from https://www.researchgate.net/publication/237371107_Technical_Manual_for_QMPE_v218_Fortran_code_to_fit_Response_Time_Distributions

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41,* 977–990. http://dx.doi.org/10.3758/BRM.41.4.977

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition, 1,* 9. http://dx.doi.org/10.5334/joc.10

Burbeck, S. L., & Luce, R. D. (1982). Evidence from auditory simple reaction times for both change and level detectors. *Perception & Psychophysics, 32,* 117–133. http://dx.doi.org/10.3758/BF03204271

Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of the orthographic neighborhood in visual word recognition: Cross-task comparisons. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 857–871. http://dx.doi.org/10.1037/0278-7393.23.4.857

Chambers, S. M. (1979). Letter and order information in lexical access. *Journal of Verbal Learning and Verbal Behavior, 18,* 225–241. http://dx.doi.org/10.1016/S0022-5371(79)90136-1

Chumbley, J. I., & Balota, D. A. (1984). A word's meaning affects the decision in lexical decision. *Memory & Cognition, 12,* 590–606. http://dx.doi.org/10.3758/BF03213348

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale, NJ: Erlbaum. https://ci.nii.ac.jp/naid/10018074200/

Cousineau, D., Brown, S., & Heathcote, A. (2004). Fitting distributions using maximum likelihood: Methods and packages. *Behavior Research Methods, Instruments, & Computers, 36,* 742–756. http://dx.doi.org/10.3758/BF03206555

Dixon, W. J. (1960). Simplified estimation from censored normal samples. *Annals of Mathematical Statistics, 31,* 385–391. http://dx.doi.org/10.1214/aoms/1177705900

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language, 27,* 429–446. http://dx.doi.org/10.1016/0749-596X(88)90066-6

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112,* 777–813. http://dx.doi.org/10.1037/0033-295X.112.4.777

Forster, K. I., & Shen, D. (1996). No enemies in the neighborhood: Absence of inhibitory neighborhood effects in lexical decision and semantic categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 696–713. http://dx.doi.org/10.1037/0278-7393.22.3.696

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

Grainger, J., O'Regan, J. K., Jacobs, A. M., & Segui, J. (1989). On the role of competing word units in visual word recognition: The neighborhood frequency effect. *Perception & Psychophysics, 45,* 189–195. http://dx.doi.org/10.3758/BF03210696

Grainger, J., O'Regan, J. K., Jacobs, A. M., & Segui, J. (1992). Neighborhood frequency effects and letter visibility in visual word recognition. *Perception & Psychophysics, 51,* 49–56. http://dx.doi.org/10.3758/BF03205073

Grainger, J., & Segui, J. (1990). Neighborhood frequency effects in visual word recognition: A comparison of lexical decision and masked identification latencies. *Perception & Psychophysics, 47,* 191–198. http://dx.doi.org/10.3758/BF03205983

Gregg, J., & Inhoff, A. W. (2016). Misperception of orthographic neighbors during silent and oral reading. *Journal of Experimental Psychology: Human Perception and Performance, 42,* 799–820. http://dx.doi.org/10.1037/xhp0000193

Havens, L. L., & Foote, W. E. (1963). The effect of competition on visual duration threshold and its independence of stimulus frequency. *Journal of Experimental Psychology, 65,* 6–11. http://dx.doi.org/10.1037/h0048690

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review, 9,* 394–401. http://dx.doi.org/10.3758/BF03196299

Heathcote, A., Popiel, S. J., & Mewhort, D. J. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin, 109,* 340–347. http://dx.doi.org/10.1037/0033-2909.109.2.340

Hoedemaker, R. S., & Gordon, P. C. (2017). The onset and time course of semantic priming during rapid recognition of visual words. *Journal of Experimental Psychology: Human Perception and Performance, 43,* 881–902. http://dx.doi.org/10.1037/xhp0000377

Huntsman, L. A., & Lima, S. D. (1996). Orthographic neighborhood structure and lexical access. *Journal of Psycholinguistic Research, 25,* 417–429. http://dx.doi.org/10.1007/BF01727000

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434–446. http://dx.doi.org/10.1016/j.jml.2007.11.007

Johnson, N. F., & Pugh, K. R. (1994). A cohort model of visual word recognition. *Cognitive Psychology, 26,* 240–346. http://dx.doi.org/10.1006/cogp.1994.1008

Johnson, R. L. (2009). The quiet clam is quite calm: Transposed-letter neighborhood effects on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 943–969. http://dx.doi.org/10.1037/a0015572

Johnson, R. L., Perea, M., & Rayner, K. (2007). Transposed-letter effects in reading: Evidence from eye movements and parafoveal preview. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 209–229. http://dx.doi.org/10.1037/0096-1523.33.1.209

Johnson, R. L., Staub, A., & Fleri, A. M. (2012). Distributional analysis of the transposed-letter neighborhood effect on naming latency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 1773–1779. http://dx.doi.org/10.1037/a0028222

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology, 68,* 601–625. http://dx.doi.org/10.1146/annurev-psych-122414-033702

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*. Advance online publication. http://dx.doi.org/10.18637/jss.v082.i13

Landauer, T. K., & Streeter, L. A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior, 12,* 119–131. http://dx.doi.org/10.1016/S0022-5371(73)80001-5

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

Luke, S. G., & Christianson, K. (2012). Semantic predictability eliminates the transposed-letter effect. *Memory & Cognition, 40,* 628–641. http://dx.doi.org/10.3758/s13421-011-0170-4

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

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94,* 305–315. http://dx.doi.org/10.1016/j.jml.2017.01.001

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88,* 375–407. http://dx.doi.org/10.1037/0033-295X.88.5.375

Mitchell, D. C., Shen, X., Green, M. J., & Hodgson, T. L. (2008). Accounting for regressive eye-movements in models of sentence processing: A reappraisal of the selective reanalysis hypothesis. *Journal of Memory and Language, 59,* 266–293. http://dx.doi.org/10.1016/j.jml.2008.06.002

Moore, M., & Gordon, P. C. (2015). Reading ability and print exposure: Item response theory analysis of the author recognition test. *Behavior Research Methods, 47,* 1095–1109. http://dx.doi.org/10.3758/s13428-014-0534-3

Paap, K. R., & Johansen, L. S. (1994). The case of the vanishing frequency effect: A retest of the verification model. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1129–1157. http://dx.doi.org/10.1037/0096-1523.20.6.1129

Paap, K. R., Johansen, L. S., Chun, E., & Vonnahme, P. (2000). Neighborhood frequency does affect performance in the Reicher task: Encoding or decision? *Journal of Experimental Psychology: Human Perception and Performance, 26,* 1691–1720. http://dx.doi.org/10.1037/0096-1523.26.6.1691

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

Paterson, K. B., Liversedge, S. P., & Davis, C. J. (2009). Inhibitory neighbor priming effects in eye movements during reading. *Psychonomic Bulletin & Review, 16,* 43–50. http://dx.doi.org/10.3758/PBR.16.1.43

Perea, M. (2015). Neighborhood effects in visual-word recognition and reading. In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of reading* (pp. 76–87). New York, NY: Oxford University Press.

Perea, M., & Pollatsek, A. (1998). The effects of neighborhood frequency in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 767–779. http://dx.doi.org/10.1037/0096-1523.24.3.767

Perea, M., & Rosa, E. (2000a). The effects of orthographic neighborhood in reading and laboratory word identification tasks. *Revista de Psicología, 21,* 327–340.

Perea, M., & Rosa, E. (2000b). Repetition and form priming interact with neighborhood density at a brief stimulus onset asynchrony. *Psychonomic Bulletin & Review, 7,* 668–677. http://dx.doi.org/10.3758/BF03213005

Pinheiro, J., & Bates, D. (2006). *Mixed-effects models in S and S-PLUS*. New York, NY: Springer Science & Business Media.

Pollatsek, A., Perea, M., & Binder, K. S. (1999). The effects of "neighborhood size" in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1142–1158. http://dx.doi.org/10.1037/0096-1523.25.4.1142

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

Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review, 83,* 190–214. http://dx.doi.org/10.1037/0033-295X.83.3.190

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology, 7,* 65–81. http://dx.doi.org/10.1016/0010-0285(75)90005-5

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422. http://dx.doi.org/10.1037/0033-2909.124.3.372

Rayner, K., & Duffy, S. A. (1987). Eye movements and lexical ambiguity. In J. K. O'Regan & A. Levy-Schoen (Eds.), *Eye movements: From physiology to cognition* (pp. 521–529). Amsterdam, the Netherlands: North-Holland/Elsevier Science. http://dx.doi.org/10.1016/B978-0-444-70113-8.50073-4

R Core Development Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Reichle, E. D. (2011). Serial-attention models of reading. In Simon Liversedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 767–786). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/oxfordhb/9780199539789.013.0042

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105,* 125–157. http://dx.doi.org/10.1037/0033-295X.105.1.125

Reichle, E. D., Pollatsek, A., & Rayner, K. (2006). E–Z Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research, 7,* 4–22. http://dx.doi.org/10.1016/j.cogsys.2005.07.002

Reingold, E. M., Reichle, E. D., Glaholt, M. G., & Sheridan, H. (2012). Direct lexical control of eye movements in reading: Evidence from a survival analysis of fixation durations. *Cognitive Psychology, 65,* 177–206. http://dx.doi.org/10.1016/j.cogpsych.2012.03.001

Schad, D. J., & Engbert, R. (2012). The zoom lens of attention: Simulating shuffled versus normal text reading using the SWIFT model. *Visual Cognition, 20,* 391–421. http://dx.doi.org/10.1080/13506285.2012.670143

Schotter, E. R., & Jia, A. (2016). Semantic and plausibility preview benefit effects in English: Evidence from eye movements. *Journal of Experi-*

mental Psychology: Learning, Memory, and Cognition, 42, 1839–1866. http://dx.doi.org/10.1037/xlm0000281

Schotter, E. R., & Leinenger, M. (2016). Reversed preview benefit effects: Forced fixations emphasize the importance of parafoveal vision for efficient reading. *Journal of Experimental Psychology: Human Perception and Performance, 42,* 2039–2067. http://dx.doi.org/10.1037/xhp0000270

Schotter, E. R., Leinenger, M., & von der Malsburg, T. (2018). When your mind skips what your eyes fixate: How forced fixations lead to comprehension illusions in reading. *Psychonomic Bulletin & Review, 25,* 1884–1890. http://dx.doi.org/10.3758/s13423-017-1356-y

Schotter, E. R., von der Malsburg, T., & Leinenger, M. (2019). Forced fixations, trans-saccadic integration, and word recognition: Evidence for a hybrid mechanism of saccade triggering in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45,* 677–688. http://dx.doi.org/10.1037/xlm0000617

Sears, C. R., Campbell, C. R., & Lupker, S. J. (2006). Is there a neighborhood frequency effect in English? Evidence from reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 1040–1062. http://dx.doi.org/10.1037/0096-1523.32.4.1040

Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 876–900. http://dx.doi.org/10.1037/0096-1523.21.4.876

Sheridan, H., & Reingold, E. M. (2012a). The time course of contextual influences during lexical ambiguity resolution: Evidence from distributional analyses of fixation durations. *Memory & Cognition, 40,* 1122–1131. http://dx.doi.org/10.3758/s13421-012-0216-2

Sheridan, H., & Reingold, E. M. (2012b). The time course of predictability effects in reading: Evidence from a survival analysis of fixation durations. *Visual Cognition, 20,* 733–745. http://dx.doi.org/10.1080/13506285.2012.693548

Slattery, T. J. (2009). Word misperception, the neighbor frequency effect, and the role of sentence context: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 1969–1975. http://dx.doi.org/10.1037/a0016894

Snodgrass, J. G., & Mintzer, M. (1993). Neighborhood effects in visual word recognition: Facilitatory or inhibitory? *Memory & Cognition, 21,* 247–266. http://dx.doi.org/10.3758/BF03202737

Speckman, P. L., & Rouder, J. N. (2004). A comment on Heathcote, Brown, and Mewhort's QMLE method for response time distributions. *Psychonomic Bulletin & Review, 11,* 574–576. http://dx.doi.org/10.3758/BF03196613

Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 461–479. http://dx.doi.org/10.1037/0096-1523.22.2.461

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly, 24,* 402–433. http://dx.doi.org/10.2307/747605

Staub, A. (2011). The effect of lexical predictability on distributions of eye fixation durations. *Psychonomic Bulletin & Review, 18,* 371–376. http://dx.doi.org/10.3758/s13423-010-0046-9

Staub, A., & Benatar, A. (2013). Individual differences in fixation duration distributions in reading. *Psychonomic Bulletin & Review, 20,* 1304–1311. http://dx.doi.org/10.3758/s13423-013-0444-x

Staub, A., White, S. J., Drieghe, D., Hollway, E. C., & Rayner, K. (2010). Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 1280–1293. http://dx.doi.org/10.1037/a0016896

Sturt, P., Pickering, M. J., & Crocker, M. W. (1999). Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language, 40,* 136–150. http://dx.doi.org/10.1006/jmla.1998.2606

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18,* 645–659. http://dx.doi.org/10.1016/S0022-5371(79)90355-4

Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior, 18,* 427–440. http://dx.doi.org/10.1016/S0022-5371(79)90237-8

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *The Journalism Quarterly, 30,* 415–433. http://dx.doi.org/10.1177/107769905303000401

Tukey, J. W. (1962). The future of data analysis. *Annals of Mathematical Statistics, 33,* 1–67. http://dx.doi.org/10.1214/aoms/1177704711

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language, 65,* 247–263. http://dx.doi.org/10.1016/j.jml.2011.05.002

Veldre, A., & Andrews, S. (2015). Parafoveal lexical activation depends on skilled reading proficiency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41,* 586–595. http://dx.doi.org/10.1037/xlm0000039

Veldre, A., & Andrews, S. (2016). Is semantic preview benefit due to relatedness or plausibility? *Journal of Experimental Psychology: Human Perception and Performance, 42,* 939–952. http://dx.doi.org/10.1037/xhp0000200

Veldre, A., & Andrews, S. (2017). Parafoveal preview benefit in sentence reading: Independent effects of plausibility and orthographic relatedness. *Psychonomic Bulletin & Review, 24,* 519–528. http://dx.doi.org/10.3758/s13423-016-1120-8

von der Malsburg, T., & Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of Memory and Language, 94,* 119–133. http://dx.doi.org/10.1016/j.jml.2016.10.003

Warrington, K. L., McGowan, V. A., Paterson, K. B., & White, S. J. (2018). Effects of aging, word frequency, and text stimulus quality on reading across the adult lifespan: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44,* 1714–1729. http://dx.doi.org/10.1037/xlm0000543

White, S. J., Johnson, R. L., Liversedge, S. P., & Rayner, K. (2008). Eye movements when reading transposed text: The importance of word-beginning letters. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1261–1276. http://dx.doi.org/10.1037/0096-1523.34.5.1261

White, S. J., & Staub, A. (2012). The distribution of fixation durations during reading: Effects of stimulus quality. *Journal of Experimental Psychology: Human Perception and Performance, 38,* 603–617. http://dx.doi.org/10.1037/a0025338

Williams, C. C., Perea, M., Pollatsek, A., & Rayner, K. (2006). Previewing the neighborhood: The role of orthographic neighbors as parafoveal previews in reading. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 1072–1082. http://dx.doi.org/10.1037/0096-1523.32.4.1072

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

Received August 30, 2018
Revision received July 29, 2020
Accepted August 23, 2020 ■