Title
Task effects reveal cognitive flexibility responding to frequency and predictability: Evidence from eye movements in reading and proofreading

Author
Levy, Roger

Publication Date
2014

Peer reviewed

Cognition 131 (2014) 1–27

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Task effects reveal cognitive flexibility responding to frequency and predictability: Evidence from eye movements in reading and proofreading ☆

Elizabeth R. Schotter [a,*], Klinton Bicknell [a], Ian Howard [a], Roger Levy [b], Keith Rayner [a]

[a] Department of Psychology, University of California, San Diego, United States
[b] Department of Linguistics, University of California, San Diego, United States

ARTICLE INFO

Article history:
Received 11 February 2013
Revised 22 November 2013
Accepted 30 November 2013

Keywords:
Reading
Eye movements
Task effects
Frequency
Predictability

ABSTRACT

It is well-known that word frequency and predictability affect processing time. These effects change magnitude across tasks, but studies testing this use tasks with different response types (e.g., lexical decision, naming, and fixation time during reading; Schilling, Rayner, & Chumbley, 1998), preventing direct comparison. Recently, Kaakinen and Hyönä (2010) overcame this problem, comparing fixation times in reading for comprehension and proofreading, showing that the frequency effect was larger in proofreading than in reading. This result could be explained by readers exhibiting substantial cognitive flexibility, and qualitatively changing how they process words in the proofreading task in a way that magnifies effects of word frequency. Alternatively, readers may not change word processing so dramatically, and instead may perform more careful identification generally, increasing the magnitude of many word processing effects (e.g., both frequency and predictability). We tested these possibilities with two experiments: subjects read for comprehension and then proofread for spelling errors (letter transpositions) that produce nonwords (e.g., *trcak* for *track* as in Kaakinen & Hyönä) or that produce real but unintended words (e.g., *trial* for *trail*) to compare how the task changes these effects. Replicating Kaakinen and Hyönä, frequency effects increased during proofreading. However, predictability effects only increased when integration with the sentence context was necessary to detect errors (i.e., when spelling errors produced words that were inappropriate in the sentence; *trial* for *trail*). The results suggest that readers adopt sophisticated word processing strategies to accommodate task demands.

1. Introduction

The processing of a word in a sentence is affected by a range of linguistic properties, across many tasks and experimental paradigms, but how does the cognitive system change the way it responds to these properties in different tasks? Two hallmark effects derive from the frequency of a word to be processed (high frequency words are processed more quickly than low frequency words) and the predictability of a word in its sentence context (more predictable words are processed more quickly than less predictable words; see Kutas & Federmeier, 2011; Rayner, 1998, 2009

* Corresponding author. Address: Department of Psychology, University of California, San Diego, La Jolla, CA 92093, United States. Tel.: +1 (858) 822 7813.
E-mail address: eschotter@ucsd.edu (E.R. Schotter).

0010-0277/$ - see front matter © 2013 The Authors. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cognition.2013.11.018

for reviews). While frequency and predictability effects are robust and well documented, the magnitudes of these effects vary across tasks and paradigms (even when equating the magnitude of the frequency or predictability manipulation). The fact that these effects change across tasks suggests that the way in which people approach a task can modulate the extent to which they are sensitive to specific linguistic properties of the words they read (even when held constant across tasks). In the present study, we investigated this cognitive flexibility in reading for comprehension and proofreading. While still poorly understood, proofreading is a useful task for elucidating how cognitive processing changes along with task demands because of its similarity to reading for comprehension in terms of stimuli and response measure. The only differences in experimental design between these two tasks are the instructions and the inclusion of sentences that contain an error. Thus, we can study how processing of sentences without errors changes when people are asked to process them in different ways: checking for errors or reading for understanding. In the remainder of this introduction, we briefly discuss frequency effects and predictability effects and existing evidence regarding how they change magnitude across tasks, then turn to theoretical and empirical aspects of proofreading and discuss the goals and design of the present study.

1.1. Frequency effects

Word frequency is one of the strongest linguistic aspects of a word that affects how easily it is processed, across many tasks (*lexical decision*, Stanners, Jastrzembski, & Westbrook, 1975; *word naming*, Berry, 1971; Forster & Chambers, 1973; reading a sentence, as indexed by *eye fixation times*, Inhoff & Rayner, 1986; Rayner & Duffy, 1986; see Rayner, 1998, 2009 for reviews; and *event related potentials*, King & Kutas, 1998; Polich & Donchin, 1988; see Kutas & Federmeier, 2011 for a review). In general, cross-experiment comparisons cannot convincingly test whether frequency effects change size across tasks because they use different stimuli (the magnitude of the effect on the response variable depends on the magnitude of the frequency manipulation) and different subjects (more skilled readers show smaller frequency effects than average readers; Ashby, Rayner, & Clifton, 2005).

The most direct indication that frequency effects change across tasks comes from studies by Schilling, Rayner, and Chumbley (1998; for a more recent similar study, see Kuperman, Drieghe, Keuleers, & Brysbaert, 2013) and Rayner and Raney (1996; see also Rayner & Fischer, 1996, as well as Murray & Forster, 2008). Schilling et al. used the same materials and subjects and compared frequency effects between word naming, lexical decision, and *gaze duration*[1] (how long the eyes remain on a word before leaving it) during reading. The sizes of the frequency effect on naming latencies, lexical decision latencies, and gaze durations were highly correlated (though Kuperman et al. (2013) reported generally lower correlations), but more importantly, were not equal across tasks (64 ms in naming, 149 ms in lexical decision, and 67 ms in gaze durations during reading). These tasks differ in the type of processing required (Schilling et al., 1998): naming emphasizes producing the sounds of the word (although this can be greatly facilitated by lexical and semantic access), lexical decision emphasizes how familiar the word is (Gernsbacher, 1984; which is highly related to word frequency), and reading emphasizes accessing the meaning of the word (but obviously involves processing the word's sounds and familiarity, as well). Rayner and Raney (1996); see also Rayner & Fischer, 1996) found that the frequency effect (which was 53 ms when subjects read for comprehension) went away (i.e., was only 1 ms) when subjects searched for a particular word in a passage (and responded when they had found it). Rayner and Raney suggested that reading for comprehension requires accessing meaning (dependent on lexical access) and searching for a word in a text can be performed by more surface-level matching and may not be sensitive to frequency. In a similar vein, during mindless reading (e.g., when the reader "zones out" and stops understanding the sentence but their eyes continue to move along the text) frequency effects are absent (Reichle, Rennenberg, & Schooler, 2010) or attenuated (Schad & Engbert, 2012).

Taken together, data on frequency effects across tasks suggest that when word identification does not occur (either because it is not necessary, as in word search, or shuts off, as in mindless reading) subjects are insensitive or minimally sensitive to word frequency, whereas when word identification is required or emphasized (as in reading and lexical decision) frequency influences how long it takes to do so, although the precise way in which it does so in these cases also depends on the particular requirements of the task.

1.2. Predictability effects

When a word is encountered in a sentence (as opposed to in isolation) the meaning of the other words in the sentence can help constrain and identify the target word. In fact, the predictability of a word (i.e., how expected the word is, given the prior context) has an effect on reading times and fixation probabilities (Balota, Pollatsek, & Rayner, 1985; Drieghe, Rayner, & Pollatsek, 2005; Ehrlich & Rayner, 1981; Kliegl, Grabner, Rolfs, & Engbert, 2004; Rayner, Slattery, Drieghe, & Liversedge, 2011; Rayner & Well, 1996; Zola, 1984; see Rayner, 1998, 2009 for reviews) as well as ERPs (Kutas & Hillyard, 1984; see Kutas & Federmeier, 2011 for a review).

Tests for predictability effects in isolated word processing tasks are rare. However, some studies have recorded response times to target words presented after a sentence context (in word naming: Stanovich & West, 1979, 1981; West & Stanovich, 1982; and lexical decision: Schuberth & Eimas, 1977) or when the target word is preceded by a single prime word (in naming: de Groot, 1985; Meyer & Schvaneveldt, 1971; and lexical decision: Schuberth & Eimas, 1977). Here, cross task comparisons reveal that the predictability effect for primed lexical decision (65 ms) is larger than for primed naming (38 ms; de Groot,

[1] Schilling et al. (1998) also analyzed first fixation duration and single fixation duration.

1985; cf. West & Stanovich, 1982), but these have not been directly compared to eye fixations in reading using the same materials and the same subjects. Therefore, as with frequency effects, discussed in Section 1.1, the degree to which subjects respond to inter-word information (i.e., predictability, or the target word's fit into the sentence context) is also modulated by the type of processing the task requires.

1.3. Proofreading

While the above studies suggest that frequency and predictability effects change across tasks, they are not the most direct test of such changes because the different tasks used (lexical decision, naming, reading) elicit different types of responses (e.g., button presses, vocal responses, eye fixation times, and EEG). Thus, comparisons between tasks, such as Schilling et al. (1998; de Groot, 1985; Kuperman et al., 2013; West & Stanovich, 1982) are suggestive of, but not conclusive about, how different tasks affect word processing, particularly with respect to how word properties are emphasized. Therefore, we turn to a pair of tasks that can utilize the same stimuli, subjects, *and* response measures: reading for comprehension and proofreading. Kaakinen and Hyönä (2010) did just this: they compared frequency effects while subjects were reading sentences for comprehension vs. proofreading for spelling errors. We will return to Kaakinen and Hyönä (2010) shortly. First, however, we discuss possible task differences introduced by proofreading, introduce a framework within which to understand and predict these task differences, and discuss previous studies investigating proofreading.

Unlike in ordinary reading, where the reader's goal is to generally apprehend the meaning of the text, in proofreading the goal of the proofreader is to search the text, identify errors (such as omissions, additions, and replacements), and (in some cases) potentially determine how the text should be changed in order to eliminate those errors while preserving (or restoring) what the proofreader infers to be the intended meaning of the text. At an intuitive level, it is plausible that there may be substantial differences in the linguistic processing performed during proofreading as compared with ordinary reading since the goals of the two tasks are substantially different: in particular, whereas in ordinary reading errors can generally be ignored so long as they do not interfere with apprehension of the text's intended meaning, in proofreading these errors are the focus of the task.

The errors existing in a text to be proofread can come in various forms: spelling errors, grammatical errors, semantic violations, etc. Most studies (including our present research) focus on misspellings, for which the error is localized to a specific word. Perhaps the most easily detectable of these errors are those that produce nonwords (*nonword errors*; e.g., *trcak* for *track*). Detection of these errors requires only the assessment of word status (i.e., whether the letter string is a known word; Daneman & Stainton, 1993; Levy, Newell, Snyder, & Timmins, 1986), and they can sometimes be identified from the surface features of the word alone (i.e., determining if the letter string follows orthographic rules of the language or can yield pronounceable output). Proofreading for these nonword (surface level) errors may be easiest because the proofreader need only check orthographic legality and/or word status and then stop (i.e., not try to integrate an error into the sentence). Thus, in these situations, linguistic processing beyond orthographic checking and basic word recognition may be reduced compared with what occurs in ordinary reading.

More subtle (and consequently less easily detected) errors are those that constitute real words (*wrong word errors*; e.g., replacing an intended word *trail* with *trial*) because these words would pass a cursory assessment of orthographic legality or word status. Consequently, to detect these types of errors, proofreaders may need to perform deeper processing than for nonword errors: they must know not only that a letter string is a word, but also what word it is, what its syntactic and semantic properties are, and whether some other word would have been appropriate instead, in order to decide whether it is an incorrect word. Note in particular that proofreading for wrong word errors thus generally requires not only checking the word itself, but also assessing the degree to which the word's meaning and grammatical properties are appropriate for the context, which requires integration of information across multiple words. Consequently, these errors should be different in at least two respects from nonword errors: (1) they should be more difficult to detect because they require processing at multiple levels (i.e., checking orthographic legality, determining word status, and checking inter-word compatibility); and (2) proofreading for wrong-word errors should involve less reduction of deeper linguistic processing (both lexical and sentence level).

1.3.1. Theoretical framework

With these considerations in mind, we now lay out a theoretical framework within which potential differences between various "reading" tasks, including normal reading for comprehension, proofreading to catch nonwords, and proofreading to catch wrong words, can be understood. This framework is agnostic as to the specific model of eye movement control in reading (e.g., Bicknell & Levy, 2010; Engbert, Nuthmann, Richter, & Kliegl, 2005; Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Rayner, & Pollatsek, 2003; Schad & Engbert, 2012) assumed, although it should be noted that any complete model of reading must ultimately be able to account for task differences in reading behavior.

Our starting desideratum is that any type of reading—be it normal reading, scanning (skimming the text to find keywords), or proofreading—must involve some combination of (1) identifying words and (2) combining the meanings of those identified words to recover sentence meaning. Each of (1) and (2) can be further broken into different components (Table 1). Word identification involves both recognition of word-form and access of lexical content. Word-form recognition can involve both decisions about whether or not the letter string is a word and, furthermore, what exact word it is. For example, *wordhood assessment*, which we define as recognizing whether the letter string has a legal (known) orthographic entry (similar to the "orthographic checking" process hypothesized by

Table 1
Framework of some processes involved in normal reading and our hypotheses about how they change when proofreading for the two types of errors involved in Experiment 1 (nonwords) and Experiment 2 (wrong words). Check marks in the reading column show that these processes are engaged and check marks for the proofreading columns show these processes are unchanged relative to normal reading. Arrows show relative increases or decreases in the importance of (and thus amount of predicted reader engagement with) each component process relative to normal reading. Multiple entries in a single cell indicate multiple possibilities as to how a component process will be modulated in the particular type of proofreading as compared with normal reading.

Process	Reading	Proofreading – nonwords	Proofreading – wrong words
Word identification			
Wordhood assessment	√	↑	√↓
Form validation	√	↑	↑
Content access	√	√↓	↑
Sentence processing			
Integration	√	√↓	↑√↓
Word-context validation	√	√↓	↑

Kaakinen & Hyönä, 2010) is most obviously relevant for proofreading, but is also relevant even for normal reading since the reader must be able to deal with novel words. We define *form validation*, on the other hand, as recognizing the specific sequence of letters constituting the word currently being read. Wordhood assessment and form validation are logically distinct. A reader may, for example, conclude that an incompletely identified letter string such as "qo###" is not a word (wordhood assessment without complete form validation), and may also correctly identify the exact letter sequence of a word such as "aortas" while failing to successfully match the sequence to an entry in his/her mental lexicon (correct form validation but incorrect wordhood assessment). *Content access* involves retrieving word meaning and grammatical properties. Sentence-level processing includes combining individual words' content into larger, phrasal units (*integration*) and also assessment of whether each individual word is compatible with the rest of the sentence (*word-context validation*; essential for many types of error correction). We assume that there is flexibility in the degree to which each of these component processes is engaged during reading and that there is some cognitive resource cost (perhaps minimal, perhaps not) associated with each such process. We further propose that readers adaptively shift the degree of engagement of each process so as to efficiently meet task goals (for further discussion see Section 1.4) without expenditure of undue amounts of cognitive resources (Table 1).

It seems clear that all five of the above processes are relevant and have resources devoted to them during normal reading (hence the check marks in those cells in Table 1); we now turn to how, in different types of proofreading, they may differ in importance relative to normal reading. When proofreading for errors that produce *nonwords*, the most obvious change is that both processes related to surface form—wordhood assessment and form validation—increase in importance (hence the up arrows in those cells in Table 1). It is unlikely, on the other hand, that these proofreaders would need to access content, integrate that content across words, or expend resources on word-context validation as thoroughly as during normal reading, because errors could be detected based almost exclusively on surface features and engaging in these processes might unnecessarily slow the proofreader down. Nevertheless, if accessing content and performing sentence-level processing are not costly, it is possible that these processes would not be de-emphasized, since sentence-level context makes reading more efficient overall (Bicknell & Levy, 2012; Ehrlich & Rayner, 1981; Morton, 1964; Rayner & Well, 1996). Thus, we predict that during proofreading for nonwords these processes would be either unchanged (represented by check marks) or de-emphasized (represented by down arrows) as compared with normal reading.

Proofreading for errors that produce wrong words, in contrast, would lead to a different prioritization of component processes: fit into sentence context rather than surface features of words is the critical indicator of error status. This task would de-emphasize (or leave unaffected) wordhood assessment, since wrong words still match to lexical entries, but more heavily emphasize form validation and content access (essential, for example, to identify an erroneous instance of *trial* that should have been *trail*, or vice versa). This task would also more heavily emphasize word-context validation. However, it is unclear how sentence-level integration would be affected by proofreading for wrong words in comparison with normal reading (and so all three possibilities are represented): it might be enhanced by the need to perform effective word-context validation, it might be reduced since the depth of interpretation required for successful normal reading may not be necessary or worthwhile for adequate proofreading for wrong words, or it could remain unchanged.

Now let us consider how the two linguistic properties we focus on in the present study—frequency and predictability—are likely to be implicated in the processes we have described above, and the implications for how the two types of proofreading tasks may modulate the effects of these variables on eye movement behavior. Within word identification, increased emphasis on form validation is likely to slow the process overall during proofreading, so that readers obtain better input regarding word form, but is unlikely to modulate frequency or predictability effects, since visual input is ultimately the sole arbiter of the form of a string. Wordhood assessment and content access together are likely to implicate both frequency and predictability: frequent words may be easier to recognize as valid strings and to retrieve content for, and predictability effects reflect readers' anticipation of upcoming meanings and word forms.

Wordhood assessment and content access need to occur when a word is first encountered in order for understanding to proceed, hence their effects should not exclusively show up on late eye movement measures, but rather should appear during first pass reading. In sentence-level processing, however, predictability, which reflects degree of contextual fit, is likely to be far more important than frequency: words with higher predictability are likely to be easier to integrate syntactically (Hale, 2001; Levy, 2008) and semantically (Kutas & Hillyard, 1984), and easier to validate as being a valid word, given the context and the visual input (Levy, Bicknell, Slattery, & Rayner, 2009).

Our framework leaves open a number of possibilities, but it also makes three clear predictions: (1) overall speed is likely to be slower in proofreading than in normal reading provided that errors are reasonably difficult to spot and subjects proofread to a high degree of accuracy; (2) effects of proofreading for nonwords should show up (at least) in early eye-movement measures; and (3) predictability effects are more likely to be magnified in proofreading for wrong words than in proofreading for nonwords. We now turn to prior research on proofreading.

1.3.2. Existing data on proofreading

Existing data on proofreading are consistent with the above account, but are far from conclusive. Most studies of proofreading involve long passages and require subjects to circle, cross out, or indicate an error some way on-line during sentence reading. The major focus of these studies is whether certain types of errors are detected, indicating the success or failure of the process, but not how it is achieved. Additionally, to avoid ceiling effects in error detection, subjects in these studies were generally told to emphasize speed, potentially de-emphasizing some of the processes that would otherwise be involved in the proofreading task (as predicted by the framework described above).

From these studies, it is clear that the ability to detect spelling errors that are a result of letter substitutions or transpositions that produce nonwords (e.g., *exict* or *exsit* for *exist*; Healy, 1980, 1981; Holbrook, 1978a, 1978b) is fairly good, but depends on the similarity between wrong letters and intended letters (i.e., surface features of the word, relating to our hypothesized process of wordhood assessment). Furthermore, subjects are better able to detect nonword errors when the intended word is low frequency (e.g., *sleat* for *sleet*) than when it is high frequency (e.g., *grean* for *green*; Van Orden, 1991; see also Holbrook, 1978b; Jared, Levy, & Rayner, 1999), suggesting that subjects are more likely to coerce an errorful letter string into a real word if it is similar to a high frequency word (wordhood assessment and form validation may have been rushed and performed too cursorily). Less detectable are wrong word errors (Daneman & Stainton, 1993; Levy et al., 1986), which moreover show differences in the contribution of phonological similarity to the intended word: homophone substitutions (e.g., *mail* for *male*) are less detectable than spelling control substitutions (e.g., *mile*; Banks, Oka, & Shugarman, 1981; Jared et al., 1999), potentially implicating that phonological status may mediate content access. Perhaps in addition, it may be the case that spelling uncertainty, which coincides with homophony, mediates content access.

The proofreading studies mentioned above generally focused on detection of errors, in terms of accuracy and detection time and can only tell us about whether or not proofreading was successful, not about how it modulated fundamental component processes of reading. A deeper understanding of this latter issue requires investigating how the reading of error-free words and sentences is affected by the instructions to look for errors. The most direct assessment of this comes from the aforementioned study by Kaakinen and Hyönä (2010). They had native Finnish speakers perform two tasks with Finnish sentences: first, they read sentences for comprehension, answering occasional comprehension questions; then, they performed a proofreading task, in which they checked for misspellings of words that produced nonwords. They analyzed reading measures on sentences that did not contain errors, but did contain a frequency manipulation (as well as a length manipulation), finding an interaction between the frequency effect and task: frequency effects for gaze durations were larger in proofreading (141 ms for long words and 79 ms for short words) than in reading for comprehension (81 ms for long words and 30 ms for short words). They concluded that their task emphasized orthographic checking, which depends on word frequency (i.e., can be done faster when the word is more familiar).

There are two possible interpretations of Kaakinen and Hyönä's (2010) results. One is that, as suggested by Kaakinen and Hyönä, word processing works qualitatively differently in proofreading than in reading for comprehension. This account implies that readers can flexibly change how they read in response to task demands. That is, they perform *task-sensitive word processing* that leads to frequency information playing a larger role than in typical word identification when proofreading for nonwords, but other properties (i.e., predictability) might remain unchanged (because frequency would be sufficient for detecting errors). This account is consistent with the theoretical framework we laid out above. It is also possible, however, that readers may have less ability to selectively change the way they process words in response to task demands. Instead, proofreading could work in a qualitatively similar way as reading for comprehension but demand that subjects become more confident than usual in word identities (to rule out visually similar nonword neighbors). Thus, subjects would take advantage of all sources of information that would help them discern the identity of the word (e.g., the predictability of the word or its fit into the sentence context). Under this *more cautious reading* account, the amplification of the frequency effect in proofreading is just a result of the longer processing time required for higher confidence (e.g., the size of the effects may grow with increasing reading times) and we would expect to see similar changes in predictability effects in response to changes in task. This account would be inconsistent with the theoretical framework we laid out above, which predicts that subcomponent processes are differentially modulated by proofreading in general.

Thus, the task-sensitive word processing account predicts that proofreading for wrong words would amplify predictability effects whereas proofreading for nonwords would not. The more cautious reading account, on the other hand, predicts that predictability effects would be amplified across the board by proofreading, regardless of the type of proofreading task. Thus, finding differential effects of word predictability as a function of type of proofreading task would support the task-sensitive word processing account, and would imply that readers exhibit substantial cognitive flexibility in adapting reading behavior to task demands. On the other hand, if predictability effects increase in proofreading for both wrong word and nonword errors, it would lend support to the more cautious reading account and suggest that readers change how they process words in response to task demands in a global, less sophisticated way.

1.4. Goals of the present study

In the present study, we thus had three main goals. The first goal was to confirm the results of Kaakinen and Hyönä (2010) that frequency effects on non-error trials increase in proofreading for nonwords in another language (English). The second goal was to tease apart the task-sensitive word processing and more cautious reading accounts by determining whether predictability effects increase in the same way as frequency effects when subjects are proofreading for nonword errors. These first two goals are tested in Experiment 1. The third goal was to compare how different types of proofreading tasks change these effects (i.e., checking for nonwords like *trcak* for *track* in Experiment 1 vs. checking for wrong words like *trial* for *trail* in Experiment 2) and to compare those results against the predictions of the theoretical framework described in Section 1.3.1. In each experiment, we had subjects perform two tasks: reading for comprehension and then proofreading for spelling errors. Both tasks included sentences without errors that contained either a frequency or a predictability manipulation that we used to determine the extent to which subjects were sensitive to these word properties. In the first experiment, subjects checked for spelling errors that produced nonwords (e.g., *trcak* instead of *track*), similar to the subjects in Kaakinen and Hyönä's (2010) experiment.

2. Experiment 1

2.1. Method

2.1.1. Subjects

Forty-eight members of the University of California, San Diego community participated in the experiment for course credit, or monetary compensation ($10.00). Subjects were native English speakers who were unaware of the purpose of this experiment. They all had normal or corrected-to-normal vision with glasses or soft contacts. In this experiment, as in Experiment 2, the subjects ranged in age from 18 to 25 years old.

2.1.2. Apparatus

Eye movement data were recorded via an SR Research Ltd. Eyelink 1000 eye tracker in tower setup that restrains head movements with forehead and chin rests. Viewing of the monitor was binocular, but only the movements of the right eye were recorded, at a sampling frequency of 1000 Hz. Subjects were seated approximately 60 cm away from a 20-in. NEC MultiSync FP 1370 CRT monitor with a screen resolution of 1024×768 pixels and a refresh rate of 150 Hz. The sentences were presented in the center of the screen with black Courier New 14-point font on a white background and were always presented in one line of text with 3.8 characters subtending 1 degree of visual angle. Following calibration, eye position errors were less than 0.3°. Subjects' responses were recorded with a Microsoft controller using a directional pad and triggers.

2.1.3. Materials and design

The stimuli/materials were adopted from four published studies to create three sets of stimuli that were fully counterbalanced across subject and task in the experiments (see Table 2): filler items (error-free in the reading block and each item containing one error in the proofreading block; Appendix A), frequency items (high vs. low frequency; Appendix B), and predictability items (high vs. low predictability; Appendix C). Filler stimuli were 60 items taken from Johnson (2009), which investigated reading time on words that have a transposition letter neighbor (e.g., *trail*, which has the transposition neighbor *trial*) and control words that were matched on length, frequency, number of orthographic neighbors, number of syllables and fit into the sentence, but did not have a transposition letter neighbor (e.g., *track*). For the reading block, the sentences with the control word without a transposition letter neighbor were presented (e.g., "The runners trained for the marathon on the *track* behind the high school."). For the proofreading block, we adapted the target words to create error stimuli, introducing one word with a spelling error in these sentences. Error words were created by transposing two letters of the control words from Johnson (2009; e.g., *track* produced *trcak*; "The runners trained for the marathon on the *trcak* behind the high school."). We matched the location of the letter transposition in these words to the location in the word with a transposition letter neighbor. For example, *trail* differs from *trial* in that the third and fourth letters are transposed so we transposed the third and fourth letters in *track* to produce *trcak*. There were three exceptions, in which the to-be-transposed letters were identical (i.e., *eggs* and *cool*) or constituted a real word (i.e., *crab*[2] which would produce *carb*), in which case we transposed the closest two non-initial letters (i.e., *egsg*, *colo* and *crba*). Frequency stimuli (which did not contain any errors) were 60 items taken from Drieghe, Rayner, and Pollatsek (2008; e.g., "The inner components are protected by a black *metal*/*alloy* increasing its lifespan."); two items were slightly modified by changing or adding a word that was not the target. For the final set of items, target words were all five letters long; the high frequency words had a mean raw frequency of 94 per million (log frequency per million of 1.8 (SE = .05)) and low frequency words had a mean raw frequency of 7 per million (log frequency per

[2] The fact that this control word (*crab*) has a transposition letter neighbor (*carb*) was an error in Johnson's (2009) stimuli.

Table 2
Sample stimuli used in the experiments. Filler items, frequency items and predictability items were each counterbalanced across task and subject. For filler items, in the reading block only the error-free control word version was seen whereas in the proofreading block the nonword error was seen in Experiment 1 and the wrong word error was seen in Experiment 2. Target words are presented in italics.

Stimulus set	Task	Target word	Sentence
Filler	Reading	Control	The runners trained for the marathon on the *track* behind the high school.
	Proof reading	Error in Experiment 1	The runners trained for the marathon on the *trcak* behind the high school.
		Error in Experiment 2	The runners trained for the marathon on the *trial* behind the high school.
Frequency	Reading and proofreading	High frequency	The inner components are protected by a black *metal* increasing its lifespan.
		Low frequency	The inner components are protected by a black *alloy* increasing its lifespan.
Predictability	Reading and proofreading	High predictability	The skilled gardener went outside to pull up the *weeds* along the driveway.
		Low predictability	The skilled gardener went outside to pull up the *roses* along the driveway.

million of 0.6 (SE = .06)), estimated from the British National Corpus (BNC, 2007).

Predictability items (which also did not contain any errors) were taken from Rayner and Well (1996; 36 items) and Balota et al. (1985; 96 items; e.g., "The skilled gardener went outside to pull up the *weeds*/*roses* along the driveway."). We made minor changes to six items to make the sentences more plausible in the low predictability condition. We performed two kinds of norming on this set: (1) cloze norming ($N = 36$), and (2) fragment plausibility norming ($N = 50$), in which subjects rated the plausibility of the fragment up to and including the critical words on a scale of 1–9. To ensure the strength of the predictability manipulation with our subjects, we excluded any items for which more than one subject gave the low predictability completion in cloze. To ensure that the stimuli were not taken to be errors in the proofreading task, however, we also excluded any item that had plausibility lower than 6 in either condition. For the final set of 60 items (12 from Rayner and Well and 48 from Balota et al.), the high predictability condition had a mean cloze score of 0.64 (SE = .02) and a plausibility rating of 7.8 (SE = .1), and the low predictability condition had a mean cloze score of 0.008 (SE = .002) and a plausibility rating of 7.1 (SE = .1). The two conditions did not significantly differ in terms of frequency of the target words (high predictability, $M_{raw} = 46$ (SE = 9), $M_{log} = 1.29$ (SE = .08); low predictability, $M_{raw} = 47$ (SE = 8), $M_{log} = 1.32$ (SE = .08), again, estimated from the BNC).

The experimental sentences were broken up into two blocks: reading for comprehension and proofreading. Both the reading and proofreading blocks consisted of 30 frequency stimuli (15 high frequency, 15 low frequency), 30 predictability sentences (15 high predictability, 15 low predictability) and 30 items from Johnson (2009), which served as fillers in the reading block (none contained errors) and errors in the proofreading block. In the proofreading block, one third of the items (30 trials) contained errors. These groups of items were fully counterbalanced in a Latin square design. The sentence presentation for each condition was randomized. Sentences in the reading block did not contain any spelling errors.

2.1.4. Procedure

At the start of the experiment, the eye-tracker was calibrated with a 3-point calibration scheme. Subjects started with the reading block and were told to read the sentences for comprehension and to respond to occasional comprehension questions. Subjects did so by pressing the left or right trigger on the Microsoft controller to answer yes or no, respectively. After each question, feedback was provided such that a correct answer would proceed to the next trial, whereas an incorrect response resulted in a screen presenting "INCORRECT!" for 3 s before advancing to a the next trial. Subjects received three practice trials before the reading block. In the proofreading block, subjects were instructed to proofread each sentence for spelling errors and after each sentence were prompted to respond whether or not there was a spelling error. There was feedback in proofreading the same as in reading. Subjects were instructed to proofread "looking for spelling errors only." At the beginning of this block, subjects received three practice trials (one of which had an error). Following Kaakinen and Hyönä (2010), the reading block was presented first to avoid carryover effects because starting with the proofreading block may have prompted subjects to continue proofreading in the reading block. Furthermore, subjects were unaware (during the reading block) that they would be proofreading in the experiment.

Each trial began with a fixation point in the center of the screen, which the subject was required to fixate until the experimenter started the trial. Then a fixation box appeared on the left side of the screen, located at the start of the sentence. Once a fixation was detected in this box, it disappeared and the sentence appeared. The sentence was presented on the screen until the subject pressed a button signaling they completed reading the sentence. Subjects were instructed to look at a target sticker on the right side of the monitor beside the screen when they

Table 3
Mean proofreading accuracy as a function of stimulus set in Experiments 1 and 2. Frequency and predictability stimuli accuracy constitutes correct rejections (i.e., correctly identifying that there is no error) and error stimuli accuracy constitutes hits (i.e., correctly identifying that there is an error). The d' measure is calculated from the hit rate and the false alarm rate (1-correct rejections).

Measure	Stimulus set	Experiment 1	Experiment 2
Accuracy	Frequency	.98 (.004)	.97 (.005)
	Predictability	.98 (.003)	.96 (.006)
	Error	.89 (.013)	.81 (.014)
d'		3.05 (.065)	2.53 (.073)

finished reading to prevent them from refixating a word as they pressed the button. In the reading block, the 30 filler sentences were followed by a comprehension question requiring a "yes" or "no" response with the corresponding triggers. In the proofreading block, every sentence was followed by a question asking, "Was there a spelling error?" After subjects finished proofreading each sentence they had to answer "yes" or "no" with the triggers. The experimental session lasted for approximately forty-five minutes to one hour.

2.2. Results and discussion

Data were analyzed using inferential statistics based on generalized linear mixed-effects models (LMMs). In the LMMs, task (reading vs. proofreading), target type (predictability item vs. frequency item, where applicable), and independent variable value (high vs. low, where applicable, or filler (error-free in the reading block) vs. error (in the proofreading block), where applicable) were centered and entered as fixed effects, and subjects and items were entered as crossed random effects, including intercepts and slopes (see Baayen, Davidson, & Bates, 2008), using the maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013). For models that did not converge before reaching the iteration limit, we removed random effects that accounted for the least variance and did not significantly improve the model's fit to the data iteratively until the model did converge.[3] In order to fit the LMMs, the lmer function from the lme4 package (Bates, Maechler, & Bolker, 2011) was used within the R Environment for Statistical Computing (R Development Core Team, 2009). For fixation duration measures, we used linear mixed-effects regression, and report regression coefficients (b), which estimate the effect size (in milliseconds) of the reported comparison, and the t-value of the effect coefficient. For binary dependent variables (accuracy and fixation probability data), we use logistic mixed-effects regression, and report regression coefficients (b), which represent effect size in log-odds space and the z value of the effect coefficient. Values of the t and z statistics greater than or equal to 1.96 indicate an effect that is significant at approximately the .05 level.

2.2.1. Question accuracy

Mean accuracy and error detection ability for proofreading are reported in Table 3. Overall, subjects performed very well both in the comprehension task (94% correct) and in the proofreading task (95% correct).

2.2.2. Eye movement measures

Fixations shorter than 80 ms were combined with a previous or subsequent fixation if they were within one character of each other or were eliminated. Trials in which there was a blink or track loss during first pass reading on the target word or during an immediately adjacent fixation were removed (1% of the original number of trials). For each fixation duration measure, durations greater than 2.5 standard deviations from the subject's mean (calculated separately across tasks) were also removed (less than 2% of the data from any measure were removed by this procedure). The remaining data were evenly distributed across conditions.

To assess the extent to which readers were sensitive to properties of specific words we analyzed reading measures on target words within the sentence.[4] We analyzed six standard fixation time measures (Rayner, 1998, 2009): first pass measures, such as *probability of making a first-pass fixation, first fixation duration* (the duration of the first fixation on the target, regardless of how many fixations are made), *single fixation duration* (the duration of a fixation on the target when only one fixation is made), *gaze duration* (the sum of the duration of all fixations made on the target before leaving it), as well as later measures, such as *total viewing time* (the sum of all fixations on the target, including rereading of it after first-pass reading) and *go-past time* (the sum of the duration of all fixations on the target and any rereading of words to the left of it until the target is passed to the right). In addition, we also analyzed the *probability of regressing into the target* and the *probability of regressing out of the target*.

[3] The models for which this was necessary were: regressions out of the target in Experiment 1 frequency stimuli (we removed the correlation between the intercept and the other random effects structure for items) and predictability stimuli (we removed the correlation between the intercept and the other random effects structure for subjects), total time analyses for predictability items in Experiment 2 (we removed the correlation between the intercept and the other random effects structure for items). For the models of the three-way interactions only the random slopes of the three-way interaction were used.

[4] In addition to the target word, we analyzed reading time measures on the pre-target (Appendix D) and post-target word (Appendix E). For the pre-target word, almost all measures showed an effect of task, with longer reading times when subjects were proofreading than when they were reading, reflecting the same pattern seen on the target and in global reading measures. For the main effects of frequency and predictability, significant effects only appeared for regressions into the pre-target word, suggesting that these effects were driven subjects directly reading the target word. For regressions into the pre-target word, frequency interacted with task but predictability did not in both Experiments. For the post-target word, there was a main effect of task for all measures in Experiment 2, but these effects were only significant for late measures in Experiment 1. There were no significant spillover effects of frequency for either Experiment. There was a significant spillover effect of predictability for regressions out of the post-target word in Experiment 1 and for all effects but gaze duration in Experiment 2. There were significant interactions between task and both of the manipulated variables for regressions out of the post-target word in Experiment 1.

Table 4
Local reading time measures (means and standard errors) on the target word with a frequency manipulation or predictability manipulation, separately across the two tasks (reading and proofreading) in Experiment 1.

Measure	Task	Frequency		Predictability	
		High	Low	High	Low
First fixation duration	Reading	214 (5.3)	229 (6.0)	202 (4.0)	212 (4.6)
	Proofreading	216 (4.3)	235 (4.9)	211 (3.8)	227 (5.0)
Single fixation duration	Reading	217 (5.6)	232 (6.2)	203 (4.3)	217 (5.7)
	Proofreading	218 (4.5)	247 (6.5)	215 (4.2)	231 (5.3)
Gaze duration	Reading	237 (6.8)	251 (7.0)	222 (5.5)	234 (7.3)
	Proofreading	236 (4.9)	281 (7.3)	234 (5.4)	250 (6.0)
Total time	Reading	257 (8.2)	273 (8.2)	233 (6.2)	255 (8.8)
	Proofreading	274 (7.5)	325 (11.3)	264 (7.8)	281 (8.8)
Go-past time	Reading	252 (8.6)	273 (8.7)	232 (6.6)	255 (8.9)
	Proofreading	269 (8.5)	326 (11.4)	250 (6.6)	273 (8.1)
Fixation probability	Reading	0.83 (0.02)	0.88 (0.02)	0.80 (0.02)	0.83 (0.02)
	Proofreading	0.87 (0.02)	0.91 (0.02)	0.84 (0.02)	0.88 (0.02)
Regressions in	Reading	0.10 (0.01)	0.07 (0.01)	0.04 (0.01)	0.08 (0.01)
	Proofreading	0.11 (0.01)	0.13 (0.01)	0.09 (0.01)	0.12 (0.02)
Regressions out	Reading	0.04 (0.01)	0.07 (0.01)	0.04 (0.01)	0.07 (0.01)
	Proofreading	0.08 (0.01)	0.10 (0.01)	0.06 (0.01)	0.07 (0.01)

Table 5
Results of the linear mixed effects regression models for fixation time measures in Experiment 1. All measures were centered before being entered into the analyses. The *b* value estimates the effect size (in milliseconds) and statistical significance is indicated by boldface.

Measure	Variable	*b*	SE	*t*
First fixation duration	Task	4.26	3.78	1.13
	Frequency	**17.72**	**3.43**	**5.17**
	Task * Frequency	3.92	5.84	0.67
	Task	**12.17**	**3.21**	**3.79**
	Predictability	**12.22**	**2.99**	**4.08**
	Task * Predictability	5.96	5.29	1.13
Single fixation duration	Task	**8.86**	**3.78**	**2.35**
	Frequency	**22.65**	**3.83**	**5.91**
	Task * Frequency	**13.12**	**6.35**	**2.07**
	Task	**13.53**	**3.45**	**3.93**
	Predictability	**14.95**	**3.53**	**4.23**
	Task * Predictability	2.77	5.75	0.48
Gaze duration	Task	**14.71**	**5.25**	**2.80**
	Frequency	**31.03**	**5.14**	**6.04**
	Task * Frequency	**29.91**	**9.56**	**3.13**
	Task	**14.15**	**4.60**	**3.08**
	Predictability	**13.71**	**4.21**	**3.25**
	Task * Predictability	3.92	7.45	0.53
Total time	Task	**34.25**	**7.40**	**4.63**
	Frequency	**35.43**	**7.77**	**4.56**
	Task * Frequency	**38.66**	**15.33**	**2.52**
	Task	**28.02**	**7.62**	**3.68**
	Predictability	**20.78**	**5.40**	**3.85**
	Task * Predictability	3.00	11.46	0.26
Go-past time	Task	**34.79**	**7.30**	**4.77**
	Frequency	**41.80**	**7.96**	**5.25**
	Task * Frequency	**34.86**	**14.67**	**2.38**
	Task	**17.97**	**7.00**	**2.57**
	Predictability	**22.71**	**5.24**	**4.33**
	Task * Predictability	1.78	9.33	0.19

To assess how subjects approached the task of proofreading, we analyzed reading time measures on target words that did not contain an error (in either the reading or proofreading block) but did contain either a frequency (e.g., "The inner components are protected by a black *metal*/*alloy* increasing its lifespan.") or predictability manipulation (e.g., "The skilled gardener went outside to pull up the *weeds*/*roses* along the driveway."). We analyzed local reading measures on the target words presented in italics above (but not presented in italics in the experiment; means and standard errors are in Table 4). For the following analyses, task (reading vs. proofreading) and independent variable (high vs. low) were entered as fixed effects in the LMMs. The LMMs were fit separately for frequency items and predictability items (except for test of the three-way interaction, see Section 2.2.2.3). An interaction between independent variable (high vs. low frequency or high vs. low predictability) and task (reading vs. proofreading) would indicate that subjects were changing their sensitivity to these word properties in order to perform the task.

2.2.2.1. Fixation time measures. Results of the linear mixed effects analyses on fixation time measures are reported in Table 5. There was a significant effect of task for all fixation time measures for sentences with a frequency manipulation (single fixation duration: $b = 8.86$, $t = 2.35$; gaze duration: $b = 14.71$, $t = 32.80$; total time: $b = 34.25$, $t = 4.63$; go-past time: 34.79, $t = 4.77$) with the exception of first fixation duration ($b = 4.26$, $t = 1.13$) and for sentences with a predictability manipulation (first fixation duration: $b = 12.17$, $t = 3.79$; single fixation duration: $b = 13.53$, $t = 3.93$; gaze duration: $b = 14.15$, $t = 3.08$; total time: $b = 28.02$, $t = 3.68$; go-past time: 17.97, $t = 2.57$), indicating that, when checking for nonword errors subjects spent longer on target words throughout their encounter with them (i.e., across all eye movement measures).

The effect of frequency was robustly found across all reading time measures (first fixation: $b = 17.72$, $t = 5.17$; single fixation duration: $b = 22.65$, $t = 5.91$; gaze duration: $b = 31.03$, $t = 6.04$; total time: $b = 35.43$, $t = 4.56$; go-past time: $b = 41.80$, $t = 5.25$) as was the effect of predictability (first fixation duration: $b = 12.22$, $t = 4.08$: single fixation duration: $b = 14.95$, $t = 4.23$; gaze duration: $b = 13.71$, $t = 3.25$; total time: $b = 20.78$, $t = 3.85$; go-past time: 22.71, $t = 4.33$).

Of more interest for our present purposes are the interactions between task and our manipulations of frequency and predictability. Here, the results are quite clear: frequency effects were reliably larger during proofreading than during reading across all measures (single fixation duration: $b = 13.12$, $t = 2.07$; gaze duration: $b = 29.91$, $t = 3.13$; total time: $b = 38.66$, $t = 2.52$, go-past time: 34.86, $t = 2.38$) with the exception of first fixation duration ($b = 3.92$, $t < 1$) whereas the effect of predictability was not modulated by task in any fixation time measure (all $ts < 1.14$). The interaction between task and the frequency effect in these data replicates Kaakinen and Hyönä's result (in a different language: English), showing that the effect of frequency becomes larger when proofreading for spelling errors that produce nonwords (see goal 1, in Section 1.4). In addition, the lack of an interaction with task for the predictability items helps to tease apart the possible interpretations of Kaakinen and Hyönä's finding (see goal 2, in Section 1.4). While the more cautious reading account predicted that there should be a similar interaction for the predictability materials, instead, these data support the task-sensitive word processing account, in which subjects process words in proofreading in a qualitatively different way that makes more use of frequency information but does not make more use of predictability. These data suggest that readers have a great deal of flexibility with respect to how they process words depending on their specific goal, making more or less use of each property of a word (e.g., its frequency or predictability from context) dependent on that feature's informativeness for the task at hand.

2.2.2.2. Fixation probability measures. Results of the logistic mixed-effects regression analyses on fixation probability measures are reported in Table 6. As with the reading time measures, in Section 2.2.2.1, fixation probability measures showed a robust effect of task, with a higher probability of fixating the target (frequency items: $z = 2.49$, $p = .01$; predictability items: $z = 3.77$, $p < .001$), regressing into the target (frequency items: $z = 3.77$, $p < .001$; predictability items: $z = 5.43$, $p < .001$) and regressing out of the target for frequency items ($z = 4.47$, $p < .001$) but not predictability items (all $ps > .24$).

Table 6
Results of the logistic regression mixed effects models for fixation probability measures in Experiment 1. All measures were centered before being entered into the analyses. Statistical significance is indicated by boldface.

Measure	Variable	*b*	*z*	*p*
Fixation probability	Task	**.38**	**2.49**	**.01**
	Frequency	**.68**	**4.24**	**<.001**
	Task * Frequency	.18	0.71	.48
	Task	**.48**	**3.77**	**<.001**
	Predictability	.23	1.78	.08
	Task * Predictability	.21	0.84	.40
Regressions into the target	Task	**.57**	**3.77**	**<.001**
	Frequency	.03	0.19	.85
	Task * Frequency	**.71**	**2.63**	**.009**
	Task	**.85**	**5.43**	**<.001**
	Predictability	**.78**	**4.65**	**<.001**
	Task * Predictability	**.90**	**2.36**	**.02**
Regressions out of the target	Task	**.81**	**4.47**	**<.001**
	Frequency	.21	1.20	.23
	Task * Frequency	.14	0.40	.69
	Task	.16	.91	.37
	Predictability	**.36**	**1.94**	**.05**
	Task * Predictability	.45	1.35	.18

Frequency yielded a main effect on probability of fixating the target ($z = 4.24$, $p < .001$) but not the probability of regressing out of the target ($p > .22$) or the probability of regressing into the target ($p > .84$). Predictability yielded a marginal effect on the probability of fixating the target ($z = 1.78$, $p = .08$), a significant effect on the probability of regressing into the target ($z = 4.65$, $p < .001$) and marginal effect on the probability of regressing out of the target ($z = 1.94$, $p = .05$). The only significant interactions between task and our manipulations of frequency and predictability were on regressions into the target (frequency items: $z = 2.63$, $p < .01$; predictability items: $z = 2.36$, $p < .001$); all other interactions were not significant (all $ps > .17$).

2.2.2.3. Do frequency and predictability interact with task similarly?. In addition to the analyses reported in Section 2.2.2.1, we tested whether the interaction in the frequency stimuli was significantly different from the null interaction in the predictability stimuli (i.e., the three-way interaction) in two key measures: gaze duration and total time. These measures have been taken to reflect the time needed for initial word identification (gaze duration) and to integrate the word into the sentence (total time). The results of these analyses revealed a significant three-way interaction for both gaze duration ($b = 11.95$, $t = 2.01$) and total time ($b = 19.93$, $t = 2.27$), confirming our analyses above in suggesting that the effect of predictability did not increase in proofreading while the effect of frequency did. Thus, our data do not show support for an account of proofreading in which subjects merely read more cautiously (and predictability effects would likewise increase) but rather support a qualitatively different type of task-sensitive word processing between reading for comprehension and proofreading.

3. Experiment 2

As discussed in Section 1.3.1, when proofreading for errors that produce real, wrong words, one must take into account the sentence context. Thus, one would expect that, when proofreading for wrong word errors, subjects may need to or want to take into account the predictability of a word more fully than they do when proofreading for nonword errors (as in Experiment 1 and Kaakinen & Hyönä, 2010). We might expect, then, that if subjects can adapt how they process words to the fine-grained demands of the task, then when proofreading for errors that produce actual words, subjects would show larger effects of predictability. Presumably, this would result from subjects' need to spend more time determining whether a word that is unlikely in context is an error.

To test whether subjects adapt how they process words based on the precise nature of the spelling errors included in the stimuli, we ran a second experiment, similar to Experiment 1 except that, during proofreading, subjects checked for spelling errors (letter transpositions) that produced real, wrong words (e.g., *trail* produced *trial*; "The runners trained for the marathon on the *trial* behind the high school."). We analyzed the same reading measures as in Experiment 1 to determine whether (a) frequency effects are increased when checking for errors that constitute real, but inappropriate words and (*b*) to determine whether predictability effects increase when task demands make predictability information useful. Additionally, we analyzed global reading measures and local reading measures on target words in the filler stimuli (fillers during the reading task and errors during the proofreading task), comparing them between the two experiments, to assess the relative difficulty of proofreading for nonword errors and proofreading for wrong word errors.

3.1. Method

The method of Experiment 2 was identical to the method for Experiment 1 with the following exceptions.

3.1.1. Subjects

A different set of 48 subjects, with the same selection criteria as Experiment 1 participated in Experiment 2.

3.1.2. Materials

The stimuli in Experiment 2 were identical to those in Experiment 1 except for the words that constituted errors in the proofreading task. Error stimuli were produced by selecting the transposition letter neighbor of the target word (from Johnson, 2009), which was inappropriate in the sentence context (e.g., *trail* produced *trial*; "The runners trained for the marathon on the *trial* behind the high school."). Using these items from Johnson (2009) in both experiments meant that the base words from which the errors were formed were controlled across experiments for length, frequency, number of orthographic neighbors, number of syllables and fit into the sentence. Thus, the only difference between experiments was whether the transposition error happened to produce a real word.

3.1.3. Procedure

The procedure was identical to Experiment 1 except that, in the proofreading block, subjects were instructed that they would be "looking for misspelled words that spell check cannot catch. That is, these misspellings happened to produce an actual word but not the word that the writer intended." and there were 5 practice trials (three errors) preceding the proofreading block instead of 3.

3.2. Results and discussion

3.2.1. Question accuracy

As in Experiment 1, subjects performed very well both on the comprehension questions (93% correct) and in the proofreading task (91% correct; Table 3). In addition to overall accuracy, we used responses in the proofreading task to calculate d' scores (the difference between the z-transforms of the hit rate and the false alarm rate; a measure of error detection) for each subject and compared them between experiments using an independent samples t test. Proofreading accuracy was significantly higher in Experiment 1 ($M = 3.05$, SE = .065) than in Experiment 2 ($M = 2.53$, SE = .073; $t(93) = 5.37$, $p < .001$), indicating that checking for real words that were inappropriate in the

Table 7
Local reading time measures (means and standard errors) on the target word with a frequency manipulation or predictability manipulation, separately across the two tasks (reading and proofreading) in Experiment 2.

Measure	Task	Frequency		Predictability	
		High	Low	High	Low
First fixation duration	Reading	216 (4.8)	224 (4.6)	206 (4.1)	213 (3.1)
	Proofreading	239 (4.7)	250 (5.9)	225 (4.2)	231 (4.3)
Single fixation duration	Reading	220 (5.0)	232 (6.5)	208 (4.6)	215 (3.1)
	Proofreading	249 (6.0)	266 (8.4)	223 (4.4)	239 (5.2)
Gaze duration	Reading	236 (6.2)	248 (7.4)	221 (5.9)	235 (4.6)
	Proofreading	275 (7.2)	312 (8.9)	260 (6.0)	287 (8.7)
Total time	Reading	285 (11.7)	293 (17.3)	257 (16.3)	278 (14.6)
	Proofreading	413 (26.3)	476 (32.2)	341 (16.6)	416 (24.5)
Go-past time	Reading	262 (8.5)	290 (13.9)	251 (9.1)	269 (9.9)
	Proofreading	335 (15.3)	400 (17.6)	309 (12.6)	350 (13.3)
Fixation probability	Reading	0.82 (.02)	0.85 (.02)	0.78 (0.02)	0.81 (0.02)
	Proofreading	0.86 (.02)	0.93 (.01)	0.85 (0.02)	0.87 (0.02)
Regressions in	Reading	0.14 (.02)	0.12 (.02)	0.08 (0.02)	0.14 (0.02)
	Proofreading	0.24 (.02)	0.23 (.02)	0.17 (0.02)	0.25 (0.03)
Regressions out	Reading	0.07 (.01)	0.11 (.02)	0.08 (0.01)	0.11 (0.02)
	Proofreading	0.12 (.02)	0.17 (.02)	0.11 (0.01)	0.16 (0.02)

Table 8
Results of the linear mixed effects regression models for fixation time measures in Experiment 2. All measures were centered before being entered into the analyses. The *b* value estimates the effect size (in milliseconds) and statistical significance is indicated by boldface.

Measure	Variable	*b*	SE	*t*
First fixation duration	Task	**24.14**	**4.40**	**5.49**
	Frequency	**10.35**	**3.96**	**2.61**
	Task * Frequency	3.98	6.68	0.60
	Task	**18.05**	**3.71**	**4.87**
	Predictability	**6.66**	**3.21**	**2.08**
	Task * Predictability	0.93	5.57	0.17
Single fixation duration	Task	**33.22**	**5.76**	**5.77**
	Frequency	**14.73**	**4.99**	**2.95**
	Task * Frequency	8.11	8.26	0.98
	Task	**19.73**	**3.99**	**4.95**
	Predictability	**11.04**	**3.54**	**3.12**
	Task * Predictability	7.44	6.72	1.11
Gaze duration	Task	**51.75**	**6.28**	**8.25**
	Frequency	**25.56**	**6.99**	**3.66**
	Task * Frequency	**26.51**	**9.46**	**2.80**
	Task	**44.79**	**6.41**	**6.99**
	Predictability	**20.95**	**5.06**	**4.14**
	Task * Predictability	13.17	9.65	1.36
Total time	Task	**155.25**	**27.15**	**5.72**
	Frequency	**36.53**	**15.67**	**2.33**
	Task * Frequency	**55.08**	**24.94**	**2.21**
	Task	**112.78**	**17.12**	**6.59**
	Predictability	**49.27**	**11.66**	**4.23**
	Task * Predictability	**57.60**	**21.18**	**2.72**
Go-past time	Task	**91.48**	**15.25**	**6.00**
	Frequency	**47.18**	**12.41**	**3.80**
	Task * Frequency	**41.51**	**18.88**	**2.20**
	Task	**69.06**	**11.36**	**6.08**
	Predictability	**29.94**	**9.56**	**3.13**
	Task * Predictability	21.10	15.77	1.34

sentence context was more difficult than checking for spelling errors that produce nonwords.

3.2.2. Eye movement measures

As with the analyses of Experiment 1 (when subjects were checking for nonwords) we analyzed reading measures on the target words in the frequency (e.g., *metal*/*alloy*) or predictability (*weeds*/*roses*) manipulation sentences when they were encountered in Experiment 2 (when subjects were checking for wrong words) to determine whether the type of error subjects anticipated changed the way they used different word properties (i.e.,

Table 9
Results of the logistic regression mixed effects models for fixation probability measures in Experiment 2. All measures were centered before being entered into the analyses. Statistical significance is indicated by boldface.

Measure	Variable	*b*	*Z*	*p*
Fixation probability	Task	**.82**	**4.92**	**<.001**
	Frequency	**.83**	**5.77**	**<.001**
	Task * Frequency	**.54**	**2.14**	**<.05**
	Task	**.62**	**5.41**	**<.001**
	Predictability	.19	1.77	.08
	Task * Predictability	.06	0.27	.79
Regressions into the target	Task	**.91**	**5.60**	**<.001**
	Frequency	.21	1.42	.16
	Task * Frequency	.30	1.33	.18
	Task	**.97**	**6.05**	**<.001**
	Predictability	**.84**	**5.35**	**<.001**
	Task * Predictability	.29	0.97	.33
Regressions out of the target	Task	**.60**	**3.64**	**<.001**
	Frequency	**.44**	**2.56**	**.01**
	Task * Frequency	.51	1.77	.08
	Task	**.62**	**4.15**	**<.001**
	Predictability	**.49**	**3.71**	**<.001**
	Task * Predictability	.15	0.52	.60

frequency and predictability; Tables 7–9). Experiment 1 revealed no evidence that the effect of the predictability of a word in the sentence differed in size between reading and proofreading (there was no interaction between predictability and task in any reading measure). Our interpretation of this result was that predictability information is not a more useful source of information when checking for nonwords as compared to when reading for comprehension. However, when the errors that must be detected are real, wrong words, the only way to detect an error is to determine whether the word makes sense in the sentence context, making predictability a more relevant word property for error detection. Thus, if our interpretation is correct that readers can qualitatively change the type of word processing they perform according to task demands, we may see the effect of predictability become larger in proofreading for wrong words (relative to reading). As with analyses of error-free items in Experiment 1, task (reading vs. proofreading) and independent variable (high vs. low) were entered as fixed effects in the LMMs. Separate LMMs were fit for frequency items and predictability items (except for the test of the three-way interaction, see Section 3.2.2.3).

3.2.2.1. Fixation time measures. There was a significant main effect of task for all fixation time measures for sentences with a frequency manipulation (first fixation duration: $b = 24.14$, $t = 5.49$; single fixation duration: $b = 33.22$, $t = 5.77$; gaze duration: $b = 51.75$, $t = 8.25$; total time: $b = 155.25$, $t = 5.72$; go-past time: $b = 91.48$, $t = 6.00$) and for sentences with a predictability manipulation (first fixation duration: $b = 18.05$, $t = 4.87$; single fixation duration: $b = 19.73$, $t = 4.95$; gaze duration: $b = 44.79$, $t = 6.99$; total time: $b = 112.78$, $t = 6.59$; go-past time: 69.06, $t = 6.08$), indicating that, when checking for spelling errors that produce wrong words subjects took more time, spending longer on the target words throughout their encounter with them (i.e., across all eye movement measures). Furthermore, the coefficients that estimate the effect size are notably larger in the second experiment, when subjects were checking for more subtle errors (letter transpositions that produced real words that were inappropriate in the context).

The effect of frequency was robustly found across all reading time measures (first fixation: $b = 10.35$, $t = 2.61$; single fixation duration: $b = 14.73$, $t = 2.95$; gaze duration: $b = 25.56$, $t = 3.66$; total time: $b = 36.53$, $t = 2.33$; go-past time: $b = 47.18$, $t = 3.80$) as was the effect of predictability (first fixation duration: $b = 6.66$, $t = 2.08$: single fixation duration: $b = 11.04$, $t = 3.12$; gaze duration: $b = 20.95$, $t = 4.14$; total time: $b = 49.27$, $t = 4.23$; go-past time: 29.94, $t = 3.13$).

Of more interest for our present purposes are the interactions between task and our manipulations of frequency and predictability. Here, the results are quite clear: frequency effects were reliably larger during proofreading than reading across all measures (gaze duration: $b = 26.51$, $t = 2.80$; total time: $b = 55.08$, $t = 2.21$, go-past time: $b = 41.51$, $t = 2.20$) with the exception of first fixation duration ($b = 3.98$, $t = 0.60$) and single fixation duration ($b = 8.11$, $t = 0.98$) whereas predictability was not modulated by task in any reading measure (all *t*s < 1.37) except for total time ($b = 57.60$, $t = 2.72$). These data suggest that, when checking for spelling errors that produce real but inappropriate words, proofreaders still perform a qualitatively different type of word processing, which specifically amplifies effects of word frequency. However, while proofreaders do not appear to change their use of predictability during initial word recognition (i.e., first pass reading), later word processing does show increased effects of how well the word fits into the context of the sentence (i.e., during total time). We return to the issue of why this effect only appears on a late measure in Section 4.2.

3.2.2.2. Fixation probability measures. As with the reading time measures reported in Section 3.2.2.1, fixation probability measures showed a robust effect of task, with a higher probability of fixating the target (frequency items: $z = 4.92$, $p < .001$; predictability items: $z = 5.41$, $p < .001$), regressing into the target (frequency items: $z = 5.60$, $p < .001$; predictability items: $z = 6.05$, $p < .001$) and regressing out of the target (frequency items: $z = 3.64$, $p < .001$; predictability items: $z = 4.15$, $p < .001$) in the proofreading task than in the reading task. Frequency yielded a main effect on probability of fixating the target ($z = 5.77$, $p < .001$) and probability of regressing out of the target ($z = 2.56$, $p < .01$) but not probability of regressing into the target ($p > .15$). Predictability yielded a marginal effect on the probability of fixating the target ($z = 1.77$, $p = .08$) and a significant effect on the probability of regressing into the target ($z = 5.35$, $p < .001$) and regressing out of the target ($z = 3.71$, $p < .001$).

There was a significant interaction between task and frequency on the probability of fixating the target ($z = 2.14$, $p < .05$) and a marginal interaction on the probability of regressing out of the target ($z = 1.77$, $p = .08$). All other interactions were not significant (all $ps > .17$). Thus, it seems as if the interactions seen in total time in Experiment 2 were not due to an increased likelihood of making a regression into or out of the target word, but rather to the amount of time spent on the word during rereading.

3.2.2.3. Do frequency and predictability interact with task similarly?. As in Experiment 1, we tested for the three-way interaction between target type (frequency vs. predictability), independent variable value (high vs. low) and task (reading vs. proofreading) to evaluate whether the interactions between independent variable and task were different between the frequency stimuli and the predictability stimuli. As in Section 2.2.2.3, we tested for the three-way interaction in two key measures: gaze duration (Fig. 1) and total time (Fig. 2). The results of these analyses revealed that neither the three-way interaction for gaze duration ($b = 5.59$, $t < 1$) nor total time ($b = 2.26$, $t < 1$) were significant, suggesting that, when proofreading for wrong word errors, subjects processed words in a way that magnified the effects of both word frequency and predictability in a similar way. However, when gaze duration was analyzed separately by stimulus set, the task by frequency interaction was significant but the task by predictability interaction was not, and the three-way interaction, while not significant, does suggest a trend in that direction. Thus, the data suggest that, in first pass reading, subjects certainly demonstrated increased sensitivity to frequency information (discussed above) and demonstrated only slight increased sensitivity to predictability information (certainly more than they demonstrated increased sensitivity to predictability information when proofreading in Experiment 1). However, the substantial interaction between task and predictability does not emerge until further inspection of the word (i.e., total time, see Section 4.2).

3.2.3. Comparisons between the two experiments

The analyses reported in this section were performed on filler items from the reading task and items that contained errors in the proofreading task to assess the degree to which proofreading sentences that actually contain errors differs from reading error-free sentences for comprehension. When encountered in the reading block, sentences contained no errors and constituted the control sentences taken from Johnson (2009; i.e., "The runners trained for the marathon on the *track* behind the high school."). When encountered in the proofreading block, sentences contained errors; In Experiment 1 errors constituted nonwords (i.e., "The runners trained for the marathon on the *trcak* behind the high school.") and in Experiment 2 errors constituted wrong words (i.e., "The runners trained for the marathon on the *trial* behind the high school."). To investigate how errors were detected, we compared both global reading measures (reading time on the entire sentence) and local reading measures on the target word (shown in italics, above, but not italicized in the experiments) between the correct trials (when encountered in the reading block) and error trials (when encountered in the proofreading block). Task (reading vs. proofreading) and experiment (Experiment 1 vs. Experiment 2) were entered as fixed effects.

3.2.3.1. Global reading measures. We analyzed two global reading measures: *total sentence reading time* (*TSRT*; the total amount of time spent reading the sentence) and *reading*

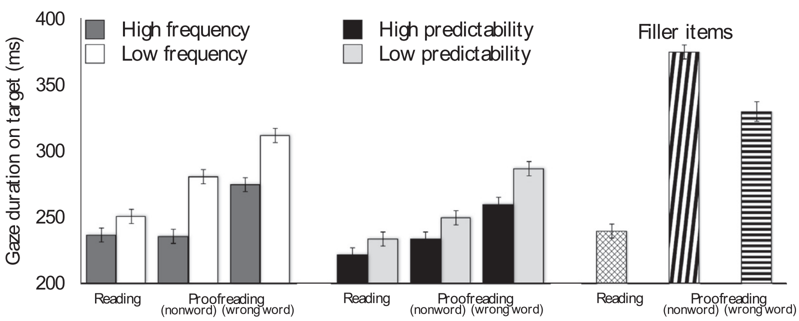

Fig. 1. Gaze duration on target words across the two experiments, separated by stimulus type. The left-hand panel represents frequency-manipulated stimuli, the central panel represents predictability-manipulated stimuli, and the right-hand panel represents filler stimuli (normal control words in the reading tasks, nonword spelling errors during the proofreading task in Experiment 1, and wrong word spelling errors in Experiment 2). Data for the reading task represent the average for a given condition across the two experiments.

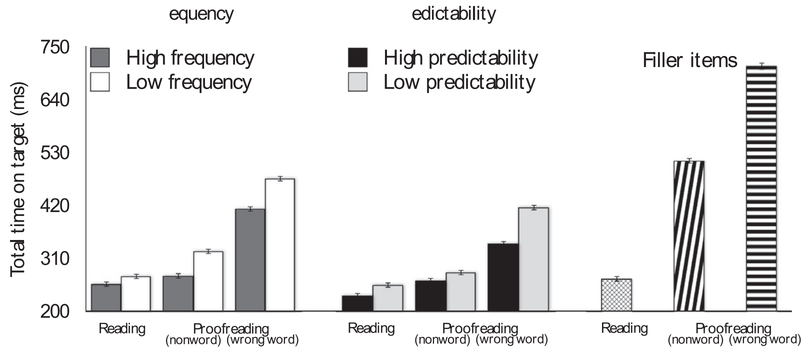

Fig. 2. Total time on target words across the two experiments, separated by stimulus type. The left-hand panel represents frequency-manipulated stimuli, the central panel represents predictability-manipulated stimuli, and the right-hand panel represents filler stimuli (normal control words in the reading tasks, nonword spelling errors during the proofreading task in Experiment 1, and wrong word spelling errors in Experiment 2). Data for the reading task represent the average for a given condition across the two experiments.

Table 10
Global and local reading time measures (means and standard errors) on the target word across the two experiments for filler trials (in reading) and error trials (in proofreading).

Measure	Reading		Proofreading	
	Experiment 1	Experiment 2	Experiment 1	Experiment 2
Global reading measures				
Total sentence reading time (ms)	2699 (95)	2970 (148)	2968 (102)	4320 (242)
Reading rate (words per minute)	327 (14)	304 (11)	299 (12)	226 (10)
Local reading measures				
First fixation duration	221 (4.5)	221 (4.4)	281 (7.0)	264 (4.7)
Single fixation duration	224 (5.0)	225 (4.9)	310 (9.6)	281 (7.0)
Gaze duration	240 (5.9)	240 (5.0)	375 (11.1)	330 (8.4)
Total time	263 (8.4)	273 (10.7)	513 (23.2)	710 (44.0)
Go-past time	268 (9.0)	274 (8.6)	451 (18.0)	465 (19.2)
Fixation probability	0.83 (0.02)	0.78 (0.02)	0.89 (0.02)	0.89 (0.01)
Regressions in	0.07 (0.01)	0.12 (0.02)	0.28 (0.02)	0.46 (0.03)
Regressions out	0.07 (0.01)	0.09 (0.01)	0.08 (0.01)	0.18 (0.01)

rate (words per minute: WPM), which index general reading efficiency (Rayner, 1998, 2009), to assess the general difficulty of the proofreading task, compared to the reading task, across the two experiments (see Table 10). More efficient reading is reflected by shorter total sentence reading time and faster reading rate (more words per minute). For the present analyses, data come from all stimuli, collapsed across target type. We entered task (reading vs. proofreading) and experiment (Experiment 1 vs. Experiment 2) as fixed effects in the LMMs.

The global reading measures confirmed the results of the accuracy analyses: The proofreading task was more difficult than the reading task, and this difference was more pronounced in the second experiment. Both measures revealed significant effects of task (TSRT: $b = 814.8$, $t = 7.99$; WPM: $b = -53.18$, $t = -9.74$), with the proofreading task leading to less efficient (slower) reading ($M_{TSRT} = 2986$ ms; $M_{WPM} = 299$ in Experiment 1 $M_{TSRT} = 4320$ ms; $M_{WPM} = 226$ in Experiment 2) than the reading for comprehension task ($M_{TSRT} = 2699$ ms; $M_{WPM} = 327$ in Experiment 1 $M_{TSRT} = 2970$ ms; $M_{WPM} = 304$ in Experiment 2). Both measures also revealed a significant effect of experiment (TSRT: $b = 801.7$, $t = 4.00$; WPM: $b = -47.84$, $t = -3.06$), with less efficient reading in the second experiment than in the first experiment. More importantly, there was a significant interaction in both measures (TSRT: $b = 1063.1$, $t = 5.23$; WPM: $b = -49.85$, $t = -4.62$), with the effect of task (reading vs. proofreading) larger in the second experiment (when proofreading involved checking for wrong words) than in the first experiment (when proofreading involved checking for nonwords).

3.2.3.2. *Local reading measures.* To assess how task demands change processing of the target words themselves (i.e., the only word that differed between tasks and between experiments in the proofreading task) we analyzed local reading measures (the same as mentioned above) on the filler trials; Tables 10–12. All analyses revealed a significant effect of task (for all fixation time measures, all $ts > 12$; for all fixation probability measures, all $ps < .001$) with longer reading times on and higher probabilities of fixating and regressing into or out of the target in the proofreading task than the reading task. There were significant differences between experiments in gaze duration and total time (both $ts > 2.09$), as well as the probability of regressing out of and into the target (both $ps < .001$), but not for any of the other fixation time measures (all $ts < 1.77$) or the probability of fixating the target ($p = .32$).

Most important for our purposes were tests for interactions between task and experiment. Analyses of fixation

Table 11
Results of the linear mixed effects regression models for fixation time measures separately for target words across the two experiments for filler trials (in reading) and error trials (in proofreading). All measures were centered before being entered into the analyses. The *b* value estimates the effect size (in milliseconds) and statistical significance is indicated by boldface.

Measure	Variable	*b*	SE	*t*
First fixation duration	Task	**52.00**	**4.21**	**12.36**
	Experiment	8.54	6.63	1.29
	Task * Experiment	−**19.24**	**8.56**	**2.25**
Single fixation duration	Task	**72.25**	**5.88**	**12.29**
	Experiment	15.05	8.56	1.76
	Task * Experiment	−**33.18**	**11.95**	**2.78**
Gaze duration	Task	**113.27**	**6.39**	**17.72**
	Experiment	**22.02**	**10.49**	**2.10**
	Task * Experiment	−**45.41**	**14.27**	**3.18**
Total time	Task	**343.11**	**26.12**	**13.14**
	Experiment	**103.79**	**30.39**	**3.42**
	Task * Experiment	**191.27**	**49.27**	**3.88**
Go-past time	Task	**187.14**	**11.95**	**15.67**
	Experiment	9.61	19.20	0.05
	Task * Experiment	7.83	25.11	0.31

Table 12
Results of the logistic mixed-effects regression models for fixation probability measures separately for target words across the two experiments for filler trials (in reading) and error trials (in proofreading). All measures were centered before being entered into the analyses. Statistical significance is indicated by boldface.

Measure	Variable	*b*	*z*	*p*
Fixation probability	Task	**1.02**	**9.13**	**<.001**
	Experiment	.21	1.00	.32
	Task * Experiment	.30	1.24	.22
Regressions into the target	Task	**2.15**	**15.96**	**<.001**
	Experiment	**.70**	**3.85**	**<.001**
	Task * Experiment	.37	1.43	.15
Regressions out of the target	Task	**.71**	**6.51**	**<.001**
	Experiment	**.67**	**3.71**	**<.001**
	Task * Experiment	**.67**	**2.92**	**<.005**

time measures revealed significant but qualitatively different interactions between task and experiment for early and late reading measures. There were significant interactions for early reading measures (first fixation duration: $b = -19.24$, $t = 2.25$; single fixation duration: $b = -31.18$, $t = 2.78$; gaze duration: $b = -45.41$, $t = 3.18$) with a larger increase in reading time in the proofreading block when checking for nonword errors (Experiment 1) than when checking for wrong word errors (Experiment 2; see Fig. 1). These data suggest that, upon initial inspection of the words, errors were detected (and produced longer reading times) more easily when they resulted in nonwords than when they resulted in wrong words. The pattern of results changed, though, in later measures. Here, reading time on the target increased more in the proofreading block when checking for wrong words (Experiment 2) than when checking for nonwords (Experiment 1) for total time on the target ($b = 191.27$, $t = 3.88$; see Fig. 2) but not significantly in go-past time ($t < .32$).

There was no significant interaction between task and experiment on the probability of fixating or regressing into the target (both *ps* > .14) but there was a significant interaction on the probability of regressing out of the target ($z = 2.92$, $p < .001$) with a small increase in regressions out of the target in Experiment 1 (.07 in reading compared to .08 in proofreading) and a large effect in Experiment 2 (.09 in reading compared to .18 in proofreading).

These data confirm that the proofreading task in Experiment 2 (checking for real, but inappropriate words for the context) was more difficult than the proofreading task in Experiment 1 (checking for nonwords). Early reading time measures increased more in Experiment 1 than Experiment 2, suggesting that these errors were easier to detect upon initial inspection. However, in later measures, reading time increased more in Experiment 2 than in Experiment 1, suggesting these errors often required a subsequent inspection to detect.

4. General discussion

Let us now consider these data in light of the theoretical framework laid out in the Introduction. Based on consideration of five component processes central to normal reading—*wordhood assessment*, *form validation*, *content access*, *integration*, and *word-context validation*—and how different types of proofreading are likely to emphasize or de-emphasize each of these component processes, this framework made three basic predictions regarding the outcome of our two experiments, each of which was confirmed. Additionally, several key patterns in our data were not strongly predicted by the framework but can be better understood within it. We proceed to describe these cases below, and then conclude this section with a brief discussion of the differences in overall difficulty of the two proofreading tasks.

4.1. Basic predictions of the framework

Our framework made three basic predictions, each confirmed in our data. First, overall speed should be slower in proofreading than in normal reading, provided that errors are reasonably difficult to spot and that readers proofread accurately. The errors we introduced into our stimuli all involved single word-internal letter swaps expected *a priori* to be difficult to identify, and our readers achieved very high accuracy in proofreading—higher in Experiment 1 (95%) than in Experiment 2 (91%). Consistent with our framework's predictions under these circumstances, overall reading speed (e.g., TSRT – total sentence reading time) was slower during proofreading than during normal reading in both experiments. (Note, however, that the two types of proofreading interacted in slightly different ways with word frequency; see Section 4.2 for further discussion.) However, we must also note that even tasks that should be less onerous than reading (e.g., x-string scanning) can lead to longer reading times (Rayner & Fischer, 1996).

Second, our framework predicted that effects of proofreading for nonwords should not show up exclusively in late measures, since proofreading for nonwords should emphasize word identification processes, which must occur upon first encountering a word. Consistent with this prediction, in Experiment 1 we found effects of task on early measures including fixation probability, first fixation duration, single fixation duration, and gaze duration; and interactions of task with word frequency on single-fixation duration and gaze duration.

Third, our framework predicted that predictability effects should be magnified more in proofreading for wrong words than in proofreading for nonwords, since proofreading for wrong words emphasizes processes that intrinsically implicate the degree of fit between a word and the rest of the sentence, (e.g., word-context validation and integration), but proofreading for nonwords does not. Indeed, whereas when proofreading for nonwords (Experiment 1) the task (reading vs. proofreading) never interacted with predictability, when proofreading for wrong words (Experiment 2) task and predictability interacted in regressions into and total time on the target word.

With respect to interpretation of Kaakinen and Hyönä's previous results on proofreading, our new results overall favor our framework's *task-sensitive word processing* account, in which component sub-processes of reading are differentially modulated by change of task, over the *more cautious reader account*, in which proofreading simply involves processing words to a higher degree of confidence. In the more cautious reading account, sensitivity to each word property that we manipulated (frequency and predictability) should be affected similarly by both types of proofreading—frequency and predictability effects would have been magnified across the board. Instead, we see different effects on predictability in proofreading for nonwords vs. proofreading for wrong words, consistent with our framework.

4.2. Further results interpretable within the framework

The other major results in our data, though not directly predicted by our framework, can be readily understood within it. First, Experiment 1 affirms Kaakinen and Hyönä's (2010) original result that frequency effects are larger in proofreading for nonwords, showing that the pattern they found in Finnish also holds in English. Experiment 2 extended this result to the case of proofreading for spelling errors that produce real words. These results were supported by interactions between frequency effects and task (in both early and late reading measures) for error-free trials.

Importantly, effects of word frequency were modulated differently in the two proofreading tasks. In Experiment 1 (proofreading for nonwords), task did not affect reading speed on high frequency words, but did slow reading of low frequency words. In Experiment 2, on the other hand, proofreading slowed reading on all words (including high frequency words). To investigate this, we performed analyses separately on high frequency words and low frequency words, testing for the effects of task (reading vs. proofreading), experiment, and the interaction between them (with linear mixed effects models with the maximal random effects structure) and follow-up paired comparisons between reading times on either high frequency words or low frequency words (analyzed separately) as a function of task. For gaze duration, the main effect of task among only high frequency words was not significant in Experiment 1 ($t = 0.13$) but was significant in Experiment 2 ($t = 5.61$), confirming that high frequency words were unaffected by proofreading for nonwords (the same pattern of data was observed for other reading time measures).

For gaze duration for low frequency words, the main effect of task was significant in both Experiment 1 ($t = 3.72$) and Experiment 2 ($t = 7.89$), confirming that they were always affected by task, regardless of what type of proofreading was being performed (the same pattern of data was observed for all other reading time measures except the effect of task was not significant on first fixation duration for Experiment 1 or go-past time in Experiment 2). Although this difference is not directly predicted within our framework, it is compatible with it: the result implies that wordhood assessment, the sole frequency-sensitive process emphasized in proofreading for nonwords, is of only minimal difficulty for high frequency words but that content access, the sole frequency-sensitive process emphasized in proofreading for wrong words, is of nonminimal difficulty even for high frequency words.

Third is the question of why predictability effects were *unchanged* in proofreading for nonwords, rather than being magnified (to a lesser degree than in proofreading for wrong words) or reduced. Any of these results would have been compatible with our framework; recalling Table 1 and Section 1.4, predictability may be implicated in wordhood assessment and/or content access, and is certainly implicated in integration and word-context validation. Thus, our result implies either that none of content access, integration, or word-context validation is actually diminished during nonword proofreading, or that predictability is involved in wordhood assessment. Although our data do not distinguish between these two possibilities, the latter seems highly plausible, especially considering previous results that visual sentence context can strongly modulate explicit visual lexical decision times (Wright & Garrett, 1984). We also consider it unlikely that sentence-level integration is completely shut off even during proofreading for nonwords; given that one basic consequence of linguistic context in reading is that it allows for reading to proceed more quickly (Bicknell & Levy, 2012; Ehrlich & Rayner, 1981; Morton, 1964; Rayner & Well, 1996), it makes sense that subjects would continue to use linguistic context during word processing (and thus show predictability effects) in order to proofread more quickly. Interestingly, one of the differences between our (and Kaakinen & Hyönä's, 2010) proofreading paradigm and the other proofreading studies described in Section 1.3.2 is that the other experiments often emphasized speed as opposed to accuracy (to avoid ceiling effects since their dependent measure was percent detection). It would be worth investigating in future studies whether and how the effects we have found here would change if speed were emphasized as opposed to accuracy.

We must also address the fact that predictability effects were modulated only for *late* measures, not for *early* measures, in Experiment 2. Once again, this result is not directly predicted by our framework, but is compatible with it. One possibility is that subjects in our study may have been hesitant to flag an unpredictable word as an error until they see the context words to the right (or reread context to the left). Because subjects received feedback on every trial (a subjectively annoying 3 s timeout with the word "INCORRECT!" displayed on the screen), we assume they were highly motivated to avoid responding incorrectly. This happened not only after misses (i.e., failing to respond that there was an error when there was one) but also after false alarms (i.e., responding that there was an error when there was not). Thus, subjects may have been reluctant to prematurely (i.e., in first-pass reading) respond without seeing whether words after the target would make the word fit into context. For example, the error "The marathon runners trained on the *trial*..." could be salvaged with a continuation such as "...*course* behind the high school." Obviously, subjects would not know this without reading the rest of the sentence and may, for all sentences, continue reading to become more confident whether the sentence contained an error or not. Once subjects know both the left and right context of the word, they then evaluate the word's fit into the sentence context, and it is this latter process that produces large effects of word predictability in total time.

Finally, we note that several aspects of our data confirm that proofreading is more difficult when spelling errors produce wrong words (e.g., *trial* for *trail*) compared to when they produce nonwords (e.g., *trcak* for *track*). First, d' scores for proofreading accuracy when checking for wrong words (Experiment 2) were lower than d' scores when checking for nonwords (Experiment 1; see Table 1). Furthermore, this difference was driven by poorer performance correctly identifying errors (81% in Experiment 2 compared to 89% in Experiment 1) rather than performance correctly identifying error-free sentences (98% vs. 97%). Second, reading efficiency (total sentence reading time and reading rate) was poorer when proofreading for wrong words than when proofreading for nonwords. Our framework made no direct predictions regarding this result, but it follows naturally from consideration of what information sources are required to detect each type of error. As discussed in Section 1.3.1, nonword spelling errors may be more easily detectable based on surface features (e.g., *trcak* violates rules of English orthography while *trial* does not). Identifying a nonword error requires only successful wordhood assessment—which can be done without regard for context but which context may nevertheless be helpful for—while identifying a wrong word error requires successful word-context validation. Thus, more information sources support nonword identification than support wrong word identification.

In this vein, the question naturally arises to what extent readers were using orthographic or phonological well-formedness to identify nonwords, as opposed to a full check against the lexicon or against context. To investigate this question, we coded each error item in Experiment 1 as being either pronounceable or unpronounceable in English. Even though approximately half of the words were pronounceable and half were not, this distinction did not affect detection accuracy (88% vs. 89%; $z < 1$, $p > .94$). These data suggest that subjects were primarily assessing wordhood through a full check against the lexicon or against context, rather than purely checking surface features such as pronounceability. As mentioned above, though, the errors in Experiment 1 were easier to detect than those in Experiment 2, suggesting that the need to integrate the word with the sentence context in order to identify whether it is an error was likely what made the proofreading task in Experiment 2 more difficult.

5. Conclusion

The results of our study, combined with the experiments discussed in the introduction (Section 1), suggest that word and sentence processing during reading is highly adaptive and responsive to task demands. That is, our subjects' proofreading performance involved not just a more cautious version of normal reading, but rather a qualitative readjustment of different component sub-processes of overall reading so as to efficiently achieve high accuracy in identifying errors. We saw that the size of the frequency effect increased when proofreading for any type of spelling error, reflecting the fact that word frequency is useful for detecting violations of word status (i.e., nonwords do not have a detectable word frequency), which might be a first step in checking for spelling errors. Likewise, when the relationship between words was crucial to identify spelling errors (in Experiment 2), we saw that the magnitude of the predictability effect increased, as well. Relative to the previous literature, our data show an especially clear demonstration of the flexibility and precise control subjects have over the component processes involved in reading, since we used tasks with the same subjects, stimuli, and response measures. Our findings imply that, in the future, researchers should anticipate the way in which the instructions they give to subjects and the types of questions they ask of them might change the way they approach the task of reading and subsequently the way in which they process words and sentences.

Our interpretation that subjects can have such fine-grained control over how they perform linguistic processing in response to subtle differences in task demands is quite consistent with other extant data. As another example from the reading domain, Radach, Huestegge, and Reilly (2008) presented data suggesting that frequency effects are larger when readers expect comprehension questions than when they expect word verification questions (although the interaction was not significant). Wotschack and Kliegl (2013) also reported modulation of both frequency and predictability effects in response to differential question difficulty. Taken together, these results and ours fit naturally with claims that readers optimize how they read for their particular goals (Bicknell & Levy, 2010; Lewis, Shvartsman, & Singh, 2013) and that reading behavior can be well described as adaptive. The general framework we introduced for understanding task-specific modulations in different component processing of reading, which predicted several of the key findings of our experiments and shed light on several more, may prove to be of further use in understanding modulations of reading behavior with other tasks, such as different types of proofreading (e.g., word-position errors) and scanning for keywords. More generally, our findings broaden the range of examples of the adaptability of cognition, and point to the remarkable potential of the human mind to shape the details of even very highly practiced cognitive processing to the precise demands of the task and the agent's particular goals.

Acknowledgments

This research was supported by Grant HD065829 and training Grant DC000041 from the National Institutes of Health as well as Grant IIS0953870 from the National Science Foundation. Portions of these data were presented at the CUNY Conference on Human Sentence Processing (2012; New York, NY) and the Annual Meeting of the Psychonomic Society (2012; Minneapolis, MN). We thank Gerry Altmann, Reinhold Kliegl, Wayne Murray, and an anonymous reviewer for their comments on an earlier version.

Appendix A

List of filler stimuli (taken from Johnson, 2009) used in both experiments. Words presented in italics below were used as target words in the reading task (not in italics in the experiment). Words in boldface to the right (not bolded in the experiment) replaced the target words and were used as errors in the proofreading task (first column are nonword errors from Experiment 1, second column are wrong word errors from Experiment 2).

1	For Halloween, Julie dressed up like an *alien* and impressed the neighbors.	**aline**	**angle**
2	John used the mathematical device to measure the exact *slope* of the steel beam.	**sloep**	**angel**
3	The gardening tools were stored in the wooden *shed* for safe keeping.	**sehd**	**bran**
4	Judith is an alcoholic and visits all of the local *pubs* in town quite frequently.	**pbus**	**bras**
5	Kaitlyn was continually tormented by the thought of a mean *bully* confronting her.	**bulyl**	**beats**
6	During the symphony, I thought that many of the loud *tones* were a bit off.	**tonse**	**beast**
7	When the paint spilled, it made a large *blob* of color on the drop cloth below.	**bolb**	**bolt**
8	The ocean harbor was full of white *ships* rocking with the waves.	**shisp**	**boast**
9	The carpenter removed the rusty *nail* and replaced it with a shiny new one.	**nial**	**blot**
10	The healthy cereal contained more *oats* than the other leading competitor.	**otas**	**barn**
11	Phyllis carefully fastened the straps on the cotton *bibs* to secure them.	**bbis**	**bars**
12	The musical instrument that was for sale was a silver *flute* and so I bought it.	**fltue**	**bulge**
13	Unfortunately, the swelling from the heat caused the *lumps* in the tire.	**lupms**	**bugle**
14	Janice found that the waters were surprisingly *cool* that July morning.	**colo**	**clam**
15	The man in the meat department used the knife to carefully *slice* the turkey.	**silce**	**crave**
16	While walking along the beach, Gail picked up a sandy *crab* to show her husband.	**crba**	**calm**
17	The expert said that the damage was caused by a *clog* that had gone undetected.	**colg**	**colt**
18	The train was powered by natural *fuel* and was fairly energy efficient.	**fule**	**cola**
19	The oceanfront property overlooked the beautiful *beach* and the crashing waves.	**beahc**	**coats**

(*continued on next page*)

20	Jennifer went to the clothing store to buy some woolen *socks* for her relatives.	**socsk**	**coast**
21	Kelli drank all of the cherry-flavored *soda* and then asked for a refill.	**soad**	**coal**
22	The little horse turned out to be a small *foal* that belonged to my neighbor.	**faol**	**clot**
23	Seriously, $200 per hour was the *rate* that the lawyer charged for advice.	**reat**	**cots**
24	You should never be jealous of others or *covet* something that they have.	**cvoet**	**carve**
25	Rachel entered the information into the *chart* before she forgot about it.	**cahrt**	**dairy**
26	The cold weather made Philip's *toes* feel like they were going to fall off.	**teos**	**eras**
27	Since the beginning of time, the earth has seen many *eons* of development.	**enos**	**ears**
28	The company was going through many changes so the manager *hired* another employee.	**hried**	**fried**
29	In order to show his dissatisfaction, Karl stuck out his *chin* and complained.	**chni**	**fits**
30	My impatient mother told me about the *bout* of anger that she had last week.	**botu**	**fist**
31	To cook the chicken, my grandmother *baked* it and then served it over rice.	**bkaed**	**fired**
32	The woman purchased a sexy lace *halter* for her wedding and honeymoon.	**hlater**	**grater**
33	Go to the kitchen and get the new cheese *slicer* to prepare the cheddar.	**silcer**	**garter**
34	Bill's improper behaviors caused Judy to call him a big *jerk* behind his back.	**jrek**	**lair**
35	The government came out with more ridiculous *laws*, deeply angering the citizens.	**lwas**	**leis**
36	The animal that I saw turned out to be a sleepy *lamb* resting in the shade.	**lmab**	**loin**
37	When I couldn't find my wallet, I concluded that it would likely be *gone* forever.	**goen**	**lots**
38	The musical composition was a graceful *sonata* written in the 18th century.	**sonaat**	**minute**
39	The lawyer worked on the project for one *decade* before giving up.	**decaed**	**minuet**
40	Timothy saw the yellow bird sitting in an old *pine* in his backyard.	**pien**	**nets**
41	The fisherman loaded the boat up with hooks, *rods*, and plenty of bait.	**rosd**	**nest**
42	The small table was unsteady because the steel *base* had been bent.	**baes**	**pots**
43	The artist used the kiln to heat the ceramic *jars* before painting them.	**jasr**	**post**
44	The children were told to be extremely *quick* and out of obedience, they were.	**quikc**	**quite**
45	Unfortunately, the antique was covered in *mold* and had to be thrown away.	**modl**	**ruts**
46	The workers tried to fix the road by filling the deep *dips* with lots of asphalt.	**disp**	**rust**
47	The music in the worship service was truly *divine* and touched many people.	**dviine**	**scared**
48	The predator made the small animal feel very *uneasy* as it hid in the brush.	**uenasy**	**sacred**
49	In order to complete the recipe, the baker added *eggs* to the batter.	**egsg**	**slat**
50	The sheets were made of an expensive *silky* material that felt good to the touch.	**sliky**	**stain**
51	The document will be official once the president *edits* the copy on his desk.	**edtis**	**sings**
52	The proud athlete won a shiny *golden* trophy for his performance.	**gloden**	**sliver**
53	Since my aunt is not at all shy, she fearlessly *chats* loudly in public.	**chtas**	**signs**
54	To finish the construction project, Hank needed to buy another long *slab* of wood.	**salb**	**salt**
55	The thunderstorm caused the trunk of the tree to *crack* down the middle.	**crcak**	**spilt**
56	Tim hoped that cleaning the blanket would remove the horrible *scent* that it had.	**secnt**	**satin**
57	The runners trained for the marathon on the new *track* behind the high school.	**trcak**	**trial**

58	I went to the courthouse to see the *judge* that I had heard so much about.	**jugde**	**trail**
59	The wedding ceremony was intended to *unify* the couple as husband and wife.	**unfiy**	**untie**
60	The humid weather caused the cover of the book to *curl* and be permanently damaged.	**crul**	**wrap**

Appendix B

List of stimuli with a frequency manipulation (taken from Drieghe et al., 2008) used in both experiments. Target words are presented in italics with the high frequency word first and the low frequency word second. For each sentence the high or low frequency words were counterbalanced across task (reading for comprehension vs. proofreading).

1	Because his wife was ill, he visited the local *beach/swamp* without her.
2	The main tourist attraction was the local *beach/swamp* which was well-known.
3	Anxiously the happy *bride/groom* awaited the arrival of the rest of the family.
4	The idiot took a picture of the wrong *bride/groom* standing in front of the church.
5	Outside in the neglected yard was a brown *chair/canoe* which he had never thrown away.
6	The surrealistic painting depicted a black *chair/tulip* standing on top of a red table.
7	He gave the waitress an extra *check/bonus* rewarding her for her excellent service.
8	On stage there was a young *child/tenor* singing with the most amazing voice.
9	When he opened the door he saw a young *child/puppy* walking across the street.
10	The woman chose the wrong *color/broom* because nobody helped her in the shop.
11	The heavy metal fan had a black *cross/skull* tattooed on his chest.
12	The t-shirt displayed a black *cross/skull* surrounded by a menacing snake.
13	To his outrage they did the wrong *dance/psalm* during the ceremony.
14	Barbara never wore the brown *dress/pants* again after that disastrous party.
15	Betsy always wore a thick *dress/apron* because the thin ones wore out too fast.
16	Sam was disappointed they didn't sell his usual *drink/snack* anywhere on campus.
17	The waiter brought the wrong *drink/cider* annihilating his chances for a big tip.
18	She never had a worse *enemy/crush* before the day she met Andy.
19	The teacher was shocked when the quiet *girls/pupil* suddenly disrupted class.
20	George placed a clean *glass/plate* accidentally back into the dishwasher.
21	After killing the bison the locals cooked the fresh *heart/liver* celebrating the hunt.
22	The crime rate in urban *homes/jails* continues to rise each year.
23	The population in urban *homes/jails* rises for the fourth year in a row.
24	In the zoo he saw a young *horse/camel* clearly suffering from the cold weather.
25	The cage contained a small *horse/tiger* which had a serious injury on its head.
26	We stopped at a rural *hotel/motel* hoping that they accepted credit cards.
27	From the top of the hill you can see the large *house/manor* where the celebrity lives.
28	Maria and Leo would like to vacation at a quiet *house/lodge* during the summer.
29	On display was a small *image/photo* portraying the young artist during the war.
30	Inside the book, a small *image/photo* depicting her late husband was hidden.
31	Tim and Adrian had never seen such a young *judge/juror* before this trial.
32	Joe was the first *judge/clown* nominated for the prize in fifty years.
33	On the table in the restaurant was a large *knife/spoon* which was not very clean.
34	The inner components are protected by a black *metal/alloy* increasing its lifespan.
35	The time traveler arrived in the wrong *month/epoch* causing quite some problems.
36	It only took one short *month/cycle* before the washing machine was broken again.
37	When Jenny heard the awful *music/carol* again she almost started to weep.
38	After Richard heard the awful *music/carol* twice he decided to go to another church.
39	Annoyed by the bad atmosphere he went to the other *party/rodeo* without his friend.
40	Accidentally he ended up going to the other *party/rodeo* where nobody knew him.
41	At the shop she bought a white *phone/scarf* knowing she would regret it later.
42	It was a painting of a white *phone/scarf* lying on a kitchen table.
43	I don't think the white *plant/lotus* which I got for my birthday will last very long.
44	For Valentine's Day, Tom bought his wife a white *plant/lotus* which was quite expensive.
45	The general said that without a basic *radio/sonar* using the boat was pointless.
46	Rumor was that the museum displayed the wrong *rifle/spear* during the exhibition.
47	He made the statement without his usual *smile/irony* which surprised his audience.
48	You could see on their faces that his usual *smile/irony* displeased them.

(*continued on next page*)

49 The victim was killed by a sharp *stone*/*spear* thrown at him from quite some distance.
50 In the hospital there was a small *store*/*kiosk* where you could buy flowers or candy.
51 Written on the bathroom wall was a short *story*/*rhyme* stating Bob's love for beer.
52 Natalie quickly washed the dirty *table*/*stove* before the party guests arrived.
53 Jack quickly cleaned the dirty *table*/*stove* before his parents arrived.
54 The police examined the empty *train*/*vault* hoping the criminals left fingerprints.
55 The detective went over the empty *train*/*vault* without finding any extra clues.
56 It was not a quiet *watch*/*clock* because the noise it made sometimes even woke me up.
57 On the table there was a small *watch*/*clock* which David bought at the antique shop.
58 The mechanic wasn't sure the fixed *wheel*/*brake* would get them home.
59 The psychiatrist said he never saw a happy *woman*/*thief* throughout his career.
60 The cowboy made sure the young *woman*/*squaw* never ran out of food.

Appendix C

List of stimuli with a predictability manipulation (items 1-12 were taken from Rayner & Well, 1996; items 13-60 were taken from Balota et al., 1985) used in both experiments. Target words are presented in italics with the high predictability word first and the low predictability word second. For each sentence the high or low predictability words were counterbalanced across task (reading for comprehension vs. proofreading).

1 They were startled by the sudden *noise*/*voice* from the next room.
2 New clues led to the criminal's *arrest*/*pardon* years after the crime.
3 The girl crept slowly toward the *door*/*edge* without anyone hearing her.
4 The postman opened the package to inspect its *contents*/*packing* before sending it.
5 To keep the animals out of the garden, he put up a *fence*/*hedge* to block it off.
6 The hikers slowly climbed up the *mountain*/*hillside* to get a better view.
7 Jill looked back through the open *window*/*curtain* to see if the man was there.
8 He scraped the cold food from his *plate*/*spoon* before washing it.
9 The man decided to shave his *beard*/*chest* before the operation.
10 The worker was criticized by his *boss*/*help* but only behind his back.
11 The friends were not talking because they had a *fight*/*scheme* last semester.
12 The woman died after a prolonged *illness*/*surgery* to everyone's dismay.
13 The doctor told Fred that his drinking would damage his *liver*/*heart* very quickly.
14 Late at night, the loud crying of the hungry *baby*/*girl* woke the neighbors.
15 The women at the church were busy sewing a *quilt*/*cover* to sell at the carnival.
16 The banker loaned the businessman some more *money*/*stuff* for his new project.
17 For Halloween, Liz dressed up as an ugly old *witch*/*ghost* and went to the party.
18 The baby laughs and giggles when she shakes her new *rattle*/*bottle* for her father.
19 The skilled burglar stole all the gold and *silver*/*copper* trinkets in their house.
20 The lightening from the storm struck a large limb of a *tree*/*bush* in our backyard.
21 The skilled gardener went outside to pull up the *weeds*/*roses* along the driveway.
22 The sailor encouraged his younger brother to join the *Navy*/*Army* after high school.
23 He is a grouch in the morning until he has had his *coffee*/*shower* and read the paper.
24 It is clear that the tornado with its damaging *winds*/*rains* demolished the town.
25 The swift mugger snatched the elderly lady's new *purse*/*scarf* from her hands.
26 She scoured her pots and *pans*/*sink* before her company arrived for the lasagna dinner.
27 Since the wedding was today, the baker rushed the wedding *cake*/*pies* to the reception.
28 The young actress rehearsed her new lines for the *play*/*talk* at the university.
29 The child started crying when the doctor gave her a *shot*/*pill* to stop the fever.
30 He went to the local post office to pick up his *mail*/*wife* after a long day at work.
31 The busy bartender gave the regular customer her *drink*/*glass* and took her money.
32 When the class went to the zoo, the loud roar of the *lion*/*bear* frightened the child.
33 The hunter carefully aimed his expensive rifle at the *deer*/*bird* and fired at it.
34 On Sundays, he watches the football game while drinking *beer*/*wine* and eating pretzels.
35 The industrious farmer gets all his fresh milk from the *cows*/*goat* on his dairy farm.
36 When the electricity in the house went out, she lit the *candles*/*lantern* to get light.
37 The pregnant woman had a craving for ice cream and *pickles*/*bananas* nearly every night.
38 She used a needle and *thread*/*string* to sew the torn hem in her skirt.

39 The young nervous paratrooper jumped out of the *plane*/*chair* when he heard the shots.
40 On the subway, one often has to push and *shove*/*press* through the crowd.
41 The fat opera singer has meticulously trained her *voice*/*child* for a number of years.
42 Her sweater was made of *wool*/*silk* imported from a village in Northern Ireland.
43 Because it snowed, the boy was anxious to use his sled on the *hill*/*yard* before it melted.
44 The baker put the *bread*/*pizza* in the hot oven to bake for 50 min.
45 Everyone in the court rose to their feet when the *judge*/*queen* entered the room.
46 At the circus, the audience laughed at the funny *clown*/*tiger* that ran around the stage.
47 The little boy enjoys having some cookies and *milk*/*soda* for a snack at night.
48 The little boy is going on a fishing trip with his *father*/*mother* this coming weekend.
49 My younger brother has brilliantly composed a new *song*/*tune* for the school play.
50 The large gulls could be seen far out over the *ocean*/*river* when the fog cleared away.
51 Common reptiles, such as *lizards*/*turtles* and snakes often make little girls squeamish.
52 The pan pizza was made with a nice thick *crust*/*dough* covered with sauce.
53 Mary's husband gave her a beautiful new pearl *necklace*/*bracelet* for their anniversary.
54 The thirsty runner drank a big glass of ice cold *water*/*juice* after the marathon.
55 On every job, the photographer takes his *camera*/*sister* with him.
56 For Valentine's Day, she received some flowers and *candy*/*fruit* from her husband.
57 We saw the old nun and the *priest*/*bishop* arrive together for the meeting this morning.
58 We saw the boy run down the *street*/*avenue* after his dog when its leash broke loose.
59 When the waiter brought the dinner the banker put salt and *pepper*/*butter* on his potato.
60 He pounded the nail into the plaster wall with a *hammer*/*wrench* and hung the picture.

Appendix D

Means and standard errors (aggregated by subject) and results of LMM models for reading measures on the pre-target word.

Experiment 1

	Frequency				Predictability			
	Proof		Read		Proof		Proof	
	Hi	Low	Hi	Low	Hi	Low	Hi	Low
ffd	226 (4.2)	227 (4.4)	209 (3.9)	214 (3.9)	217 (4.7)	222 (4.7)	209 (4.2)	205 (3.9)
gzd	257 (6.4)	263 (7.1)	220 (4.5)	232 (5.5)	238 (6.1)	251 (8.0)	226 (4.7)	222 (5.1)
sfd	230 (4.6)	234 (5.9)	212 (4.1)	218 (4.2)	219 (5.1)	226 (5.1)	210 (4.6)	204 (4.6)
tvt	393 (25.0)	418 (28.4)	264 (13.5)	276 (14.2)	303 (13.2)	328 (15.8)	261 (15.5)	248 (8.8)
gopast	302 (11.2)	311 (13.5)	261 (11.6)	268 (8.3)	290 (12.5)	306 (17.7)	262 (11.8)	254 (10.0)
skip	0.87 (0.02)	0.88 (0.02)	0.85 (0.02)	0.85 (0.02)	0.51 (0.02)	0.52 (0.02)	0.54 (0.02)	0.52 (0.02)
regin	0.12 (0.01)	0.14 (0.01)	0.05 (0.01)	0.07 (0.01)	0.06 (0.01)	0.09 (0.01)	0.04 (0.01)	0.08 (0.01)
regout	0.07 (0.01)	0.06 (0.01)	0.08 (0.01)	0.07 (0.01)	0.04 (0.01)	0.04 (0.01)	0.04 (0.01)	0.04 (0.01)

Experiment 2

	Frequency				Predictability			
	Proof		Read		Proof		Read	
	Hi	Low	Hi	Low	Hi	Low	Hi	Low
ffd	216 (4.5)	214 (4.2)	209 (4.6)	210 (4.7)	209 (4.6)	204 (3.8)	199 (4.7)	204 (5.3)
gzd	238 (5.7)	244 (6.8)	233 (6.7)	232 (6.9)	228 (5.5)	219 (4.4)	210 (6.3)	214 (6.9)
sfd	217 (4.5)	220 (4.8)	214 (5.3)	214 (5.2)	212 (5.1)	206 (3.9)	199 (5.0)	205 (5.5)
tvt	276 (8.0)	288 (10.1)	253 (8.3)	249 (8.6)	247 (7.5)	247 (6.5)	219 (6.6)	226 (7.9)
gopast	254 (7.6)	258 (8.4)	254 (9.3)	250 (9.0)	251 (8.7)	247 (7.5)	227 (7.4)	231 (8.4)
skip	0.90 (0.01)	0.88 (0.02)	0.82 (0.02)	0.85 (0.02)	0.57 (0.02)	0.53 (0.02)	0.51 (0.02)	0.51 (0.02)
regin	0.24 (0.03)	0.29 (0.03)	0.11 (0.02)	0.13 (0.02)	0.14 (0.02)	0.20 (0.02)	0.10 (0.02)	0.12 (0.02)
regout	0.11 (0.01)	0.12 (0.02)	0.11 (0.02)	0.11 (0.02)	0.07 (0.01)	0.07 (0.01)	0.05 (0.01)	0.06 (0.01)

LMM outputs

Measure	Variable	Experiment 1		Experiment 2	
		b	*T*	*b*	*t*
First fixation duration	Task	5.20	1.68	**14.36**	**4.73**
	Frequency	.38	<1	2.74	<1
	Task * Frequency	2.39	<1	2.80	<1
	Task	3.65	1.16	**10.33**	**2.49**
	Predictability	−0.43	<1	1.36	<1
	Task * Predictability	9.33	1.44	12.12	1.84
Gaze duration	Task	**8.85**	**2.05**	**33.62**	**5.84**
	Frequency	4.19	1.26	**9.16**	**2.21**
	Task * Frequency	9.23	1.38	6.42	<1
	Task	6.27	1.05	**21.31**	**3.64**
	Predictability	1.79	<1	2.51	<1
	Task * Predictability	9.00	1.05	**20.21**	**2.01**
Total time	Task	**30.81**	**4.39**	**136.63**	**5.46**
	Frequency	4.72	<1	19.64	1.93
	Task * Frequency	17.02	1.69	14.92	<1
	Task	**16.60**	**2.28**	**54.66**	**3.46**
	Predictability	6.97	1.32	5.08	<1
	Task * Predictability	1.88	0.26	29.15	1.29
Regressions in (*z* & *p*)	Task	**6.83**	**<.001**	**6.96**	**<.001**
	Frequency	**3.87**	**<.001**	1.24	.21
	Task * Frequency	**1.99**	**.047**	**2.00**	**.046**
	Task	.74	.46	**4.28**	**<.001**
	Predictability	**3.39**	**<.001**	**3.48**	**<.001**
	Task * Predictability	1.05	0.29	.33	.74
Regressions out (*z* & *p*)	Task	.31	.76	.57	.56
	Frequency	1.49	.14	1.87	.06
	Task * Frequency	.50	.62	.02	.99
	Task	.72	.47	1.24	.22
	Predictability	**2.27**	**<.05**	1.30	.20
	Task * Predictability	.50	0.62	1.35	.18

Appendix E

Means and standard errors (aggregated by subject) and results of LMM models for reading measures on the post-target word

Experiment 1

	Frequency				Predictability			
	Proof		Read		Proof		Read	
	Hi	Low	Hi	Low	Hi	Low	Hi	Low
ffd	227 (5.1)	226 (4.4)	225 (5.9)	227 (6.5)	214 (5.3)	217 (5.2)	209 (6.1)	215 (5.9)
gzd	270 (6.4)	266 (5.6)	257 (7.6)	274 (9.5)	237 (7.4)	239 (6.8)	226 (7.1)	231 (6.6)
sfd	234 (5.9)	232 (4.9)	234 (7.3)	240 (7.7)	212 (5.3)	221 (5.7)	212 (6.2)	222 (6.8)
tvt	334 (12.6)	342 (11.3)	293 (9.6)	306 (11.3)	264 (9.8)	278 (10.7)	234 (7.5)	246 (8.4)
gopast	326 (11.5)	335 (12.4)	294 (11.0)	306 (12.4)	279 (13.0)	286 (14.1)	238 (8.2)	268 (14.7)
skip	0.86 (0.02)	0.85 (0.02)	0.85 (0.02)	0.87 (0.02)	0.52 (0.02)	0.52 (0.02)	0.48 (0.02)	0.51 (0.02)
regin	0.16 (0.02)	0.13 (0.02)	0.10 (0.01)	0.08 (0.01)	0.09 (0.01)	0.10 (0.01)	0.08 (0.01)	0.11 (0.02)
regout	0.08 (0.01)	0.12 (0.01)	0.07 (0.01)	0.07 (0.01)	0.05 (0.01)	0.06 (0.01)	0.02 (0.01)	0.05 (0.01)

Experiment 2

	Frequency				Predictability			
	Proof		Read		Proof		Read	
	Hi	Low	Hi	Low	Hi	Low	Hi	Low
ffd	238 (5.3)	240 (4.9)	225 (4.7)	222 (4.0)	219 (4.1)	226 (4.8)	214 (5.6)	220 (6.1)
gzd	289 (7.2)	295 (8.1)	266 (5.3)	267 (6.4)	252 (7.3)	255 (7.4)	228 (6.6)	238 (7.6)
sfd	245 (6.0)	253 (6.5)	232 (5.5)	229 (4.4)	219 (4.6)	230 (6.6)	213 (5.8)	223 (6.5)
tvt	471 (27.1)	447 (21.6)	332 (17.6)	335 (15.7)	338 (16.6)	367 (17.7)	270 (16.8)	282 (12.9)
gopast	397 (14.3)	381 (13.8)	320 (13.1)	314 (9.7)	326 (17.9)	341 (22.7)	270 (16.2)	338 (33.8)
skip	0.90 (0.01)	0.88 (0.02)	0.83 (0.02)	0.82 (0.02)	0.54 (0.02)	0.56 (0.02)	0.49 (0.02)	0.48 (0.02)
regin	0.25 (0.02)	0.26 (0.03)	0.16 (0.02)	0.18 (0.02)	0.19 (0.02)	0.21 (0.02)	0.15 (0.02)	0.16 (0.02)
regout	0.15 (0.02)	0.11 (0.01)	0.11 (0.01)	0.10 (0.01)	0.06 (0.01)	0.10 (0.01)	0.04 (0.01)	0.08 (0.01)

LMM outputs

Measure	Variable	Experiment 1		Experiment 2	
		b	*T*	*b*	*t*
First fixation duration	Task	.38	<1	**15.48**	**3.53**
	Frequency	2.21	<1	.47	<1
	Task * Frequency	3.04	<1	5.10	<1
	Task	.53	<1	5.53	1.17
	Predictability	6.97	1.74	**8.54**	**2.14**
	Task * Predictability	1.17	<1	3.13	<1
Gaze duration	Task	3.16	<1	**27.34**	**3.61**
	Frequency	6.39	1.15	3.01	<1
	Task * Frequency	**22.31**	**2.01**	2.32	<1

(*continued on next page*)

(*continued*)

Measure	Variable	Experiment 1		Experiment 2	
		b	*T*	*b*	*t*
	Task	1.79	<1	**14.67**	**2.11**
	Predictability	4.77	<1	9.92	1.90
	Task * Predictability	2.18	<1	2.92	<1
Total time	Task	**37.68**	**2.76**	**127.60**	**5.48**
	Frequency	8.25	1.07	11.72	1.00
	Task * Frequency	9.62	<1	30.45	1.62
	Task	**17.46**	**1.76**	**60.91**	**3.44**
	Predictability	11.17	1.94	**25.45**	**2.39**
	Task * Predictability	.39	<1	24.04	1.07
Regressions out (*z* & *p*)	Task	**3.53**	**<.001**	**3.07**	**<.005**
	Frequency	1.49	.13	1.25	.21
	Task * Frequency	**1.99**	**.047**	0.35	.73
	Task	**3.14**	**<.005**	**1.92**	**.055**
	Predictability	**3.28**	**<.005**	**3.52**	**<.001**
	Task * Predictability	**2.30**	**<.05**	.94	.34

References

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

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

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology, 17*, 364–390.

Banks, W. P., Oka, E., & Shugarman, S. (1981). Receding of printed words to internal speech: Does receding come before lexical access? In O. J. L. Tzeng & H. Singer (Eds.), *Perception of print* (pp. 137–170). Hillsdale, NJ: Erlbaum.

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.

Bates, D. M., Maechler, M., & Bolker, B. (2011). *Lme4: Linear mixed-effects models using S4classes. R package version 0.999375-42*. <http://CRAN.R-project.org/package.lme4>.

Berry, C. (1971). Advanced frequency information and verbal responses times. *Psychonomic Science, 23*, 151–152.

Bicknell, K., & Levy, R. (2010). A rational model of eye movement control in reading. In J. Hajič, S. Carberry, S. Clark, & J. Nivre (Eds.), *Proceedings of the 48th annual meeting of the Association for Computational Linguistics (ACL)* (pp. 1168–1178). Uppsala, Sweden: Association for Computational Linguistics.

Bicknell, K., & Levy, R. (2012). Word predictability and frequency effects in a rational model of reading. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the Cognitive Science Society* (pp. 126–131). Austin, TX: Cognitive Science Society.

The British *National Corpus*, version 3 (BNC XML Edition) (2007). *Distributed by Oxford University Computing Services on behalf of the BNC consortium*. <http://www.natcorp.ox.ac.uk/>.

Daneman, M., & Stainton, M. (1993). The generation effect in reading and proofreading: Is it easier or harder to detect errors in one's own writing? *Reading and Writing: An Interdisciplinary Journal, 5*, 297–313.

De Groot, A. M. B. (1985). Word-context effects in word naming and lexical decision. *The Quarterly Journal of Experimental Psychology, 37*, 281–297.

Drieghe, D., Rayner, K., & Pollatsek, A. (2005). Eye movements and word skipping during reading revisited. *Journal of Experimental Psychology: Human Perception and Performance, 31*, 954–969.

Drieghe, D., Rayner, K., & Pollatsek, A. (2008). Mislocated fixations can account for parafoveal-on-foveal effects in eye movements during reading. *The Quarterly Journal of Experimental Psychology, 61*, 1239–1249.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior, 20*, 641–655.

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

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12*, 627–635.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General, 113*, 256–281.

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

Healy, A. F. (1980). Proofreading errors on the word *the*: New evidence on reading units. *Journal of Experimental Psychology: Human Perception and Performance, 6*, 45–57.

Healy, A. F. (1981). The effects of visual similarity on proofreading for misspellings. *Memory & Cognition, 9*, 453–460.

Holbrook, M. B. (1978a). Effect of subjective verbal uncertainty on perception of typographical errors in a proofreading task. *Perceptual and Motor Skills, 47*, 243–250.

Holbrook, M. B. (1978b). Effect of subjective interletter similarity, perceived word similarity, and contextual variables on the recognition of letter substitutions in a proofreading task. *Perceptual and Motor Skills, 47*, 251–258.

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

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

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.

Kaakinen, J., & Hyönä, J. (2010). Task effects on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 1561–1566.

King, J. W., & Kutas, M. (1998). Neural plasticity in the dynamics of human visual word recognition. *Neuroscience Letters, 244*, 61–64.

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

Kuperman, V., Drieghe, D., Keuleers, E., & Brysbaert, M. (2013). How strongly do word reading times and lexical decision times correlate? Combining data from eye movement corpora megastudies. *The Quarterly Journal of Experimental Psychology, 66*, 563–580.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the Event-related Brain Potential (ERP). *Annual Review of Psychology, 62*, 621–647.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161–163.

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

Levy, B. A., Newell, S., Snyder, J., & Timmins, K. (1986). Processing changes across reading encounters. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 467–478.

Levy, R., Bicknell, K., Slattery, T. J., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Science, 106*, 21086–21090.

Lewis, R., Shvartsman, M., & Singh, S. (2013). The adaptive nature of eye-movements in linguistic tasks: How payoff and architecture shape speed-accuracy tradeoffs. *Topics in Cognitive Science, 5*, 581–610.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*, 227–234.

Morton, J. (1964). The effects of context upon speed of reading, eye movements and eye-voice span. *The Quarterly Journal of Experimental Psychology, 16*, 268–281.

Murray, W. S., & Forster, K. I. (2008). The rank hypothesis and lexical decision: A reply to Adelman & Brown (2008). *Psychological Review, 115*, 240–251.

Polich, J., & Donchin, E. (1988). P300 and the word frequency effect. *Electroencephalography and Clinical Neurophysiology, 70*, 33–45.

Radach, R., Huestegge, L., & Reilly, R. (2008). The role of global top-down factors in local eye-movement control in reading. *Psychological Research, 72*, 675–688.

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

Rayner, K. (2009). The Thirty Fifth Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62*, 1457–1506.

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

Rayner, K., & Fischer, M. H. (1996). Mindless reading revisited: Eye movements during reading and scanning are different. *Perception & Psychophysics, 5*, 734–747.

Rayner, K., & Raney, G. (1996). Eye movement control in reading and visual search: Effects of word frequency. *Psychonomic Bulletin & Review, 3*, 245–248.

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

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

R Development Core Team. (2009). R: A language and environment for statistical computing [Computer software]. Retrieved from http://www.R-project.org.

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.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26*, 445–476.

Reichle, E. D., Rennenberg, A. E., & Schooler, J. W. (2010). Eye movements during mindless reading. *Psychological Science, 21*, 1300–1310.

Schad, D. J., & Engbert, R. (2012). The zoom lens of attention: Simulated shuffled versus normal text reading using the SWIFT model. *Visual Cognition, 20*, 391–421.

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

Schuberth, R. E., & Eimas, P. D. (1977). Effects of context on the classification of words and nonwords. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 27–36.

Stanners, R. F., Jastrzembski, J. E., & Westbrook, A. (1975). Frequency and visual quality in a word–nonword classification task. *Journal of Verbal Learning and Verbal Behavior, 14*, 259–264.

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

Stanovich, K. E., & West, R. F. (1981). The effect of sentence context on ongoing word recognition: Tests of a two-process theory. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 658–672.

Van Orden, G. C. (1991). Phonologic mediation is fundamental to reading. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 7–103). Hillsdale, NJ: Erlbaum.

West, R. F., & Stanovich, K. E. (1982). Source of inhibition in experiments on the effect of sentence context on word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 385–399.

Wotschack, C., & Kliegl, R. (2013). Reading strategy modulates parafoveal-on-foveal effects in sentence reading. *The Quarterly Journal of Experimental Psychology, 66*, 549–562.

Wright, B., & Garrett, M. (1984). Lexical decision in sentences: Effects of syntactic structure. *Memory & Cognition, 12*(1), 31–45.

Zola, D. (1984). Redundancy and word perception during reading. *Perception & Psychophysics, 36*, 277–284.