Words from the wizarding world: fictional words, context and domain knowledge

Ingram, Joanne; Hand, Christopher J.

Published in:
Journal of Experimental Psychology: Learning, Memory, and Cognition

DOI:
10.1037/xlm0000946

Publication date:
2020

Document Version
Author accepted manuscript

Citation for published version (Harvard):
Ingram, J & Hand, CJ 2020, 'Words from the wizarding world: fictional words, context and domain knowledge', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 46, no. 11, pp. 2179-2192. https://doi.org/10.1037/xlm0000946

Words from the wizarding world: fictional words, context and domain knowledge

Joanne Ingram[1] & Christopher J. Hand[2]

[1]School of Media, Culture and Society, University of the West of Scotland, Paisley, UK.

joanne.ingram@uws.ac.uk

[2]Department of Psychology, Glasgow Caledonian University, Glasgow, UK.

christopher.hand@gcu.ac.uk

Corresponding Author:

Dr. Joanne Ingram
School of Education and Social Science
University of the West of Scotland
Paisley Campus
Paisley
United Kingdom
PA1 2BE

tel: +44(0)141 849 4429
e-mail: joanne.ingram@uws.ac.uk

Abstract

The influence of domain knowledge on reading behaviour has received limited investigation compared to the influence of, for example, context and/or word frequency. The current study tested participants with and without domain knowledge of the *Harry Potter* (HP) universe. *Fans* and *Non-Fans* read sentences containing HP, high-frequency (HF), or low-frequency target-words. Targets were presented in contexts which were supportive or unsupportive within a 2 (Group: Fans, Non-Fans) × 3 (Context: HP, HF, LF) × 3 (Word Type: HP, HF, LF) mixed design. Thirty-two Fans and 22 Non-Fans read 72 two-sentence experimental items whilst eye-movement behaviour was recorded: initial sentences established context; second sentences contained target-words. Fans processed HP words faster than Non-Fans. No group difference was observed on HF or LF processing durations suggesting equivalent reading capabilities. In HP contexts, HP and LF targets were processed equivalently. Processing of HF and LF words was facilitated by their supportive context as expected. Non-Fans made more regressions into the target region in HP contexts and regressed more into HP targets than other targets; Fans regressed into target word regions equivalently across all context and word types. Results suggest that domain knowledge influences *early* but not *immediate* lexical access, whilst the processing effect of novelty was seen in regressive eye movements. These results are more supportive of modular accounts of linguistic processing and serial models of eye movement control. Words without grounding in reality, or true embodiment, were integrated into Fans’ mental lexicons.

Keywords: *Harry Potter; domain knowledge; fiction; eye movements; context*

1. Introduction

The independent effects of word frequency and supporting context on the reading of normal language have been extensively studied (see Hand, Miellet, O'Donnell & Sereno, 2010 for a review). To date, there has been a lack of investigation into reading processes associated with *fictional words*. Consider the works of, for example, J.R.R. Tolkien, Lewis Carroll, or C.S. Lewis: set in fantastical or parallel worlds wherein it is permissible to use fictional vocabulary to describe abstract ideas or chimeric objects. The meaning of the fictional words may be explained to readers or may be determined from the surrounding context during reading. Readers immersed in such literature may quickly become familiar with the fictional words. However, in alternative contexts, these words may become senseless, and for those who have not read the book or series in question the words are effectively meaningless. The current study examined the processing of fictional words – specifically words appearing in the *Harry Potter* book series. The experiment reported below examined the eye movement behaviour of readers familiar or unfamiliar with the Harry Potter series when encountering fictional words from this series, either in supportive or unsupportive contexts.

The popularity of the HP book series, as well as the abundance of new vocabulary found within these books, makes this series an ideal domain for research into how fictional words are processed. There are seven official books authored by J.K. Rowling which feature the character Harry Potter. Harry Potter is first introduced as an 11-year-old boy who, upon discovering that he is a wizard, is subsequently introduced to a parallel world where magic and wizards exist. Throughout the seven books – which span the seven years of his wizarding education – Harry is introduced to a range of magic spells, objects, animals and abstractions which have been given 'made-up', allegorical names. For the most part, the author has created these words to represent items and actions which can only exist or be embodied within the magical world. The longevity and popularity of the HP series means that the fictional words within these books are likely to be highly familiar to avid readers, whilst these words in their written form will be novel (or at least highly unfamiliar) to those who have not engaged with the series.

A major influence on eye movement behaviour during written language processing is word length (e.g., Just & Carpenter, 1980; Rayner, Sereno & Raney, 1996). After controlling for the effects of word length, two additional factors have extensively been documented to influence on-line processing – word frequency and support from preceding context (Rayner, 1998, 2009). Studies examining word-form frequency routinely demonstrate that high frequency (HF) words – typically those which occur >50 times per million words – are fixated for shorter durations than low frequency (LF) words – typically those occurring <10 times per million occurrences (e.g., Hand et al., 2010; Rayner, Ashby, Pollatsek & Reichle, 2004). The role of contextual support on target word processing has also been consistently reported – words which receive more support from the prior and evolving discourse context are processed for shorter durations and skipped more often than words which are less supported by context (e.g., Balota, Pollatsek & Rayner, 1985; Hand et al., 2010; Hand, O'Donnell & Sereno, 2012; Kliegl, Grabner, Rolfs & Engbert, 2004; Rayner et al., 2004; Sereno, Hand, Shahid, Yao & O'Donnell, 2018).

One factor which thus far has received minimal attention in relation to word processing is prior, or domain, knowledge. Prior knowledge of a topic in general has been shown to have a facilitating effect on comprehension for text (e.g., Anderson, 1984; Spilich, Vesonder, Chiesi & Voss, 1979). At the discourse level, it is thought that the integration of new information into a representation is facilitated by prior knowledge which acts as a schema (Voss, Vesonder & Spilich, 1980). However, relatively little research has assessed the relationship between domain knowledge and lexical access, with recent exceptions assessing domain related subordinate meanings of homographs (Rodd et al., 2016; Wiley, George & Rayner, 2018). Rodd et al. examined adults' lexical-semantic representations of ambiguous words when subordinate or domain specific meanings were experienced in both recent and long-term timeframes. Using the sport of rowing as the specific domain, participants with more rowing experience generated more rowing-specific meanings for relevant ambiguous words; participants who had engaged with the domain (rowed) on the day of the experiment provided more domain-specific responses. Wiley et al. (2018) used the domain of baseball to examine reading times on ambiguous words. They found that participants with high domain knowledge (baseball experts) were impaired in resolving an ambiguous word (e.g., *bats*) to its subordinate (non-baseball) meaning, even when a strong biasing context was provided. Together, the results of these studies provide strong evidence that domain knowledge can influence lexical access of ambiguous words. These results support the reordered access model of lexical processing (Duffy, Morris & Rayner, 1988) as domain knowledge, in addition to meaning frequency, has had a clear influence on the interpretation of ambiguous words. Within the current study we expanded upon these studies to assess if lexical access for *fictional*, domain-specific, words is influenced in the same way. A high level of domain knowledge (i.e., having read HP novels) was expected to facilitate the benefit provided by a supportive context when processing HP words. If domain knowledge does influence processing of fictional words, this would support interactive models of language processing (e.g., McClelland, 1987) which suggest that higher-order factors can directly influence lexical access. It is possible that encountering fictional words outside of their normal domain may have disrupted lexical access and obscured the effects of domain knowledge. Whilst the subordinate meanings of ambiguous words are less frequent, they still represent concepts or objects grounded in reality. That is, for the most part, participants have had personal real-world experience with both meanings of the word. In the case of fictional words, readers have had highly restricted experience of the words and / or no personal embodiment of these words and concepts. If the fictional nature of these words were to affect lexical access, higher reading times would be expected even when presented in a supportive context.

The specific domain of *Harry Potter* has previously been used to assess the benefits of domain knowledge on sentence processing. Troyer and Kutas (2018) and Troyer, Urbahc & Kutas (2019) found reduced N400 amplitudes on target words for participants with high domain knowledge when presented at the end of a contextually-supported sentence. Target words were a combination of regular words and ‘fictional’ words from the HP domain. However, the effect of type of target word was not examined. In the initial study (Troyer & Kutas, 2018), the N400 context effects for HP sentences could be predicted by the participants’ level of HP knowledge. The latter study (Troyer et al., 2019) provided further support using a single-trial design to analyse trials where participants indicated if they knew the information in the sentence before reading it. These studies provided clear evidence that semantic processing can be quickly influenced by domain knowledge. In restricting the knowledge to a single domain, these authors were able to demonstrate how an individual’s experiences can affect semantic processing. Within the current study we made use of the same domain knowledge to examine how individuals’ experiences affected word-level processing, specifically in relation to those words rarely, if ever, encountered outside the *Harry Potter* universe.

Providing prior knowledge of context has also been shown to influence both discourse comprehension and lexical access in studies using passages with and without titles. Wiley and Rayner (2000) found that reading was faster, due to fewer regressions, when complex passages were presented with a descriptive title, demonstrating the effect of prior knowledge on general discourse comprehension. Wiley and Rayner also examined the effect of prior knowledge on the lexical access of ambiguous words. The authors found shorter gaze durations for dominant, balanced, and relatively frequent subordinate meanings of ambiguous words when presented in passages with a supporting title. Only lexical access of very infrequent subordinate meanings was not facilitated by prior knowledge. As prior knowledge was not sufficient to aid integration of very infrequent subordinate meanings, these results cannot be taken to support interactive accounts of lexical access (e.g., McClelland, 1987). However, given the processing benefit seen for relatively frequent subordinate meanings, nor can the results be said to support a strictly modular accounts (e.g., Fodor, 1983). Instead, Wiley and Rayner suggested that a context-sensitive approach (Paul, Kellas, Martin & Clark, 1992) was most appropriate. In a study examining level of prior knowledge and the influence of perspective instructions, Kaakinen and Hyönä (2007) found that higher prior knowledge reduced overall gaze duration on text in comparison to lower prior knowledge, and higher prior knowledge notably increased the probability of skipping words in the middle and at the end of a sentence. In relation to perspective instructions prior knowledge of a topic aided readers in identifying perspective-relevant text to integrate into the developing text representation, and to disregard perspective-irrelevant text which need not be incorporated.

Words from HP, and their supporting contexts, will be familiar to those who have read these books. However, for those who have not engaged with the HP series these words are novel (or at least grossly unfamiliar). Novel words have been demonstrated to be read more slowly than familiar words, to be regressed into more frequently than other words, and to lead to longer times spent reading and rereading the surrounding context (Williams & Morris, 2004). A study using prefixes to create novel words (e.g., "miscircled") found context to be less important to the reading of novel words (Pollatsek, Slattery & Juhasz, 2008). Reading times suggested that the meaning of the new word was established on first encountering it (or shortly after). This contrasted with the findings of Chaffin, Morris and Seeley (2001) who found that only about 25% of the additional cost of encountering a novel word was immediate. Chaffin et al. (2001) found that the processing of novel words was not initially different to that of low-familiarity words; rather, that differences in processing during development of the new word-meaning later in the sentence drove processing costs. Informative contexts which followed novel words were read for longer and novel words were reread and regressed into more often. However, definitional associates of target words, presented in a second sentence, appeared to be fully inferred from the earlier informative context regardless of the familiarity of the target word. Chaffin et al.'s (2001) findings demonstrate that the meaning of novel words can be learned rapidly, within a single sentence, but not immediately. Chaffin et al.'s initial processing time results are considered to be consistent with the modular E-Z Reader model of lexical access (e.g., Reichle, Pollatsek, Fisher & Rayner, 1998). That is, initial processing time of low familiarity and novel words did not significantly differ as eye movements may have been initiated prior to the completion of lexical access, based on alternative features (e.g., orthographic features) which allow the expected outcome of lexical access to be pre-calculated. Within the current study the supporting context was presented prior to the 'novel' word; however, without domain knowledge this is unlikely to affect reading of fictional (novel to non-fans) words. Therefore, where participants are not familiar with the domain of *Harry Potter*, initial processing of fictional HP words may not differ from that of low-frequency words, but later measures of eye movements may show processing costs.

To compare the pattern of eye-movements when reading fictional HP words to reading standard non-fictional words, we presented materials which were orthogonally manipulated across both target word type and contextual support. Prior studies of lexical processing typically include a measure of Cloze predictability (e.g. Rayner et al, 2004; Hand et al, 2010). However, since even Fans of HP would be unlikely to predict fictional words in a Cloze task, we have designed bespoke materials to provide strong contextual support (similar to Troyer & Kutas, 2018; Troyer et al., 2019). Eye-movements associated with processing HP words were compared with those for length-matched HF and LF words, presented in contexts which either supported or were neutral to the target words. Materials were read by two participant groups: those with a high level of HP domain knowledge (Fans) or those with very little domain knowledge (Non-Fans). Based on previous findings, it was expected that for those with stronger domain knowledge, a supportive context would facilitate the processing of fictional HP words. However, those with minimal domain knowledge would process these HP words as if they were novel.

2. Method

2.1. Participants

We recruited participants who were identified as highly familiar with the Harry Potter series (Fans) and those who were identified as unfamiliar (Non-Fans). Every participant in the Fan group had read at least one of the novels; Fans had read each of the entries in the book series between 1-6 times (M=2.24, SD=0.05); additionally, Fans had seen at least one entry in the film series (each film had been viewed between 1-6 times, M=3.60, SD=0.09). Non-Fans were much less familiar with the Harry Potter universe: no participant in the Non-Fan group had ever read any of the novels; however, there was evidence of exposure to the motion picture series, with some Non-Fans having seen some of the films (number of film views ranged from 0-4; M=0.66, SD=0.23).

In total, 32 Fans and 22 Non-Fans took part. Prior to recruitment, we established a target number of participants based on the effects observed in previous studies (e.g., Hand et al., 2012). A power analysis determined that 20 participants per group would yield approximately .90 power to detect the smallest anticipated (Cohen's d = 0.1) effect with α = .05. Volunteers were recruited from the communities of UK Higher Education institutions.

Participants were either paid £5 or received course credit for their efforts. All participants were native English speakers, with no diagnosed history of reading disorders (e.g., dyslexia). All participants had normal or corrected-to-normal vision.

2.2. Materials and Design

Seventy-two target words across three categories of word were selected. Subsequently, 72 bespoke contextual frames were designed. This allowed for 24 words of each type (HP, LF, HF) to be presented in three distinct styles of context (supporting either HP, LF, or HF targets). Target words were matched exactly across conditions for word length. The British National Corpus (Davies, 2004) was used to ascertain frequencies of LF and HF targets; no such academic resource was available to derive HP frequencies. All LF targets occurred maximally 10 times per million tokens; HF targets occurred minimally 50 times per million tokens. Careful counterbalancing was used to ensure that all 72 target words were seen by each participant, but only in a single contextual iteration; similarly, distinct participant sub-groups were presented with different versions of the experimental materials. Example materials are presented in Table 1, and a full list of experimental stimuli is provided within Supplemental Material A. Each contextual frame could support one of the three target word ‘triplicates’, and depending on which frame was used, the target words would be supported to a greater or lesser extent. The semantic fit of targets to context frames was not specifically controlled but was addressed through counterbalancing. After counterbalancing, each participant saw eight experimental items in each of the 9 experimental conditions; thus, a 2 (Group: Fans, Non-Fans) × 3 (Context: HP, HF, LF) × 3 (Word Type: HP, HF, LF) mixed-factorial design was employed.

Table 1

Table 1: Example target stimuli

Context	Supported Target Material	Alternative Target
Harry Potter (HP)	Hagrid explained that the world is filled with magical and non-magical people. He said that it is often easy to spot a *muggle* at King's Cross station.	*person* | *robber* (HF) | (LF)
High Frequency (HF)	Peter had broken his wrist watch and was unsure if he was on time for the meeting. He would ask the next *person* that he saw if they knew the correct time.	*muggle* | *robber* (HP) | (LF)
Low Frequency (LF)	The detective took statements from the cashier as the alarm sounded in the bank. He did not expect to catch the *robber* as they had left little evidence.	*muggle* | *person* (HP) | (HF)

Note. Target words presented in *italics*. Example items contain supported target items. Target words were rotated in a Latin square design across three sub-groups such that every participant saw every target word but only once, and saw each bespoke sentence frame, but again only once. An equal number of items were seen by each participant across Word Type × Context conditions.

2.3. Apparatus

Participants' eye movements were monitored using an SR Research Desktop Mount EyeLink 1000 eye-tracker with a chin-forehead rest used to minimise head movements. The tracker sampled pupil position and corneal reflection at 1000 Hz, and the tracker has a resolution of .01°. Viewing was binocular, with eye movements recorded from the right eye. EyeTrack software (https://blogs.umass.edu/eyelab/software/) was used to present stimuli. A Dell Optiplex 745 Desktop PC with an Intel Core 2 processor controlled stimulus presentation and a Dell Trinitron CRT monitor (170 Hz) was used to display written stimuli. Experimental passages were displayed over two lines of display with double line spacing. Each line was a maximum of 75 characters long. Text (black letters on a white background) was displayed in Vera Sans Mono font, 14-point size. At a viewing distance of approximately 72cm, three characters subtended each degree of visual angle.

2.4. Procedure

The current study was designed and carried out to British Psychological Society standards (BPS, 2014), and ethical approval was granted by the institutional research ethics committee. Informed consent was obtained from each participant before testing began. Participants were not made aware of the '*Harry Potter*' aspect of the research prior to the eye-tracking session, however, they were made aware that certain words presented would be 'fictional'. After receiving written and verbal instructions, participants completed a written consent form. Participants were given detailed instructions about the eye-tracking session – that they were to read silently for comprehension, and that 'yes/no' questions would follow some passages. The eye-tracking session began with initial calibration of the eye-tracker (a nine-point calibration extending across the full horizontal and vertical ranges of the display).

Initial calibration was followed immediately by a calibration validation to ensure that all fixation points were $<0.5^{\circ}$ erroneous.

Each trial began with a central fixation point which was displayed until the experimenter accepted the position of the participant's eye. A black square then appeared on the left side of the screen corresponding to the first character of the passage; this remained until the display software accepted the accuracy of participant fixation. After reading each passage, participants were instructed to look at a red sticker on the bottom right corner of the computer monitor, and then to execute a manual response to clear the screen. Passages were either immediately followed by the central fixation point to begin a new trial cycle, or followed by a simple 'yes/no' comprehension question relating to the previous passage. Comprehension questions were displayed after $1/6^{th}$ of experimental trials and referred to the sentence in general (e.g. Were the creatures guarding a hospital?) rather than domain specific aspects so as to remain the same across conditions. Accuracy of comprehension questions was 90.38% for Fans and 87.73% for Non-Fans (three participants did not respond to one of the comprehension questions, data was missing for one Non-Fan). At the end of the eye-tracking task, participants were asked to provide data regarding their familiarity with the *Harry Potter* series, by indicating which novels and movies they had read/seen (if any) and how many times they had read or seen these, before being debriefed as to the purpose of the experiment.

3. Results

Several standard measures of eye movement behaviour were analysed (see Rayner, 1998, 2009 for a review). Prior to inferential testing, raw eye movement data was subject to standard cleaning processes (e.g., Hand et al., 2010; Hand et al., 2012; Ingram, Hand & Moxey, 2014). Individual fixations which were shorter than 50 ms but were within one character of another fixation were merged together. Remaining fixations which were less than 50 ms and more than three characters from another fixation were eliminated. Individual fixations longer than 750 ms were eliminated. Trials were eliminated if there was a track loss or blink in the target region – this accounted for 1.49% of total experimental trials (1.97% Fan data, 0.86% Non-Fan data). Target words were skipped during first-pass reading in 7.02% of remaining trials (7.31% Fan data, 6.66% Non-Fan data); target words were skipped entirely on 4.59% of trials (4.89% Fan data, 4.20% Non-Fan data).

3.1. Measures of Fixation Duration

Measures examined included: first fixation duration (FFD) – the duration of the initial fixation on a target word; gaze duration (GD) – the sum of the durations of all fixations prior to leaving the target region in any direction; total time (TT) – the sum of the durations of any fixations made in the target region, including regressions. Descriptive statistics are presented in Table 2.

Table 2: Descriptive statistics across conditions and fixation duration measures

		First Fixation Duration			Gaze Duration			Total Reading Time		
Context	**Group**	**HP word**	**HF word**	**LF word**	**HP word**	**HF word**	**LF word**	**HP word**	**HF word**	**LF word**
HP	Fans	250 (89)	226 (66)	253 (90)	382 (282)	270 (139)	325 (194)	480 (336)	389 (251)	487 (355)
	Non-Fans	254 (95)	233 (80)	259 (93)	515 (411)	279 (134)	394 (240)	780 (541)	381 (286)	487 (299)
HF	Fans	262 (113)	224 (66)	248 (82)	402 (290)	261 (108)	342 (220)	610 (480)	289 (141)	470 (334)
	Non-Fans	263 (105)	220 (61)	248 (89)	528 (399)	272 (109)	361 (226)	788 (529)	293 (123)	494 (312)
LF	Fans	272 (97)	235 (77)	240 (85)	426 (288)	284 (168)	284 (135)	636 (489)	393 (264)	339 (210)
	Non-Fans	281 (118)	239 (80)	238 (79)	584 (391)	283 (121)	311 (175)	885 (611)	366 (250)	373 (239)
95% CIs										
HP	Fans	[219,281]	[203,249]	[222,284]	[284,480]	[222,318]	[258,392]	[363,596]	[302,476]	[363,610]
	Non-Fans	[214,294]	[199,266]	[220,298]	[344,687]	[223,335]	[294,494]	[554,1006]	[261,500]	[362,612]
HF	Fans	[223,301]	[201,247]	[220,276]	[302,503]	[223,298]	[266,418]	[443,776]	[240,338]	[355,586]
	Non-Fans	[219,307]	[195,245]	[211,285]	[361,695]	[226,318]	[266,455]	[567,1009]	[241,344]	[364,625]
LF	Fans	[238,306]	[208,261]	[210,269]	[326,526]	[226,342]	[238,331]	[467,805]	[302,484]	[266,412]
	Non-Fans	[232,331]	[205,272]	[205,271]	[420,748]	[233,334]	[238,384]	[629,1140]	[262,471]	[273,473]

Note. (*SD*s) presented in ms rounded to the nearest whole number. 95% CIs = lower and upper confidence estimates.

In order to test the independent and combined effects of Group, Context and Word Type on fixation time measures, we used the 'lme4' package in R (Bates, Mächler, Bolker, & Walker, 2015; R Development Core Team, 2016; http://www.r-project.org) to create a number of linear mixed effect (LME) models. Fixation time data were first visualised using Q-Q plots; the data was then log-transformed to improve the fit of residuals. For all models, the optimal random effect structure justified by the data was identified using forward model selection (see Barr, Levy, Scheepers, & Tily, 2013; Matuschek, Kliegl, Vasishth, Baayen & Bates, 2017). Fixed effects were tested using likelihood-ratio tests comparing full and reduced models. Post-hoc tests for main effects and significant interactions were conducted using the 'phia' package (De Rosario-Martinez, 2015), and our significance threshold was adjusted using the Bonferroni method.

[1]Across fixation duration measures, all models included random intercepts across subjects and items. After constructing intercept-only models, model fits across all fixation time measures were improved by adding random slopes of Group × Word Type across items (all χ^2s>61.49, all *p*s<.001). All other iterations either did not improve upon these models or failed to converge. A summary of effects and interactions across measures is presented in Table 3. Only the interactions between Group, Context and Word Type are described in detail in this section; for isolated fixed effects of each variable and follow-up comparisons (where appropriate), please see Supplemental Material B.

Table 3: Summary of effects on fixation duration measures

		FFD		GD		TT	
	df	χ^2	*p*	χ^2	*p*	χ^2	*p*
Group × Context × Word Type	4	<1		2.83	ns	6.29	ns
Group × Context	2	<1		<1		1.41	.494
Group × Word Type	2	<1		29.45	<.001	55.33	<.001
Context × Word Type	4	17.79	.001	16.83	.002	41.30	<.001
Group	1	<1		4.52	.034	3.21	.073
Context	2	5.79	.055	<1	ns	2.66	ns
Word Type	2	51.37	<.001	61.62	<.001	57.80	<.001

Note. χ^2 values rounded to 2DP; *p*-values rounded to 3DP. ns = $p>.15$.

3.1.1. Group × Context × Word Type

There was no evidence of a three-way interaction in any of the fixation duration measures (see Table 3).

3.1.2. Group × Context

There was no evidence of an interaction between participant Group and Context in any of the fixation duration measures (see Table 3).

3.1.3. Group × Word Type

A significant Group × Word Type interaction was observed in GD and TT (both χ^2s>29.45, both *p*s<.001); post-hoc comparisons are presented below in Table 4. Figure 1 provides an illustration of these interactions.

Table 4: Post-hoc comparisons for significant Group × Word Type interactions

Word Type	**Group**		**GD**	**TT**
HP	Fans vs. Non-Fans	diff.	138	241
		χ^2	17.96	26.20
		p	<.001	<.001
HF	Fans vs. Non-Fans	diff.	6	11
		χ^2	<1	<1
		p	ns	ns
LF	Fans vs. Non-Fans	diff.	38	18
		χ^2	2.56	<1
		p	ns	ns

Note. HP = Harry Potter; HF = high frequency; LF = low frequency. diff = mean difference in ms rounded to nearest whole number. χ^2 values rounded to 2DP; p-values rounded to 3DP. ns = p>.15. All dfs = 1.

[Insert Figure 1 about here]

In both GD and TT, HP Fans processed HP words significantly faster than Non-Fans (both ps<.001). Importantly, the non-significant differences between Fan and Non-Fan processing of HF and LF targets, across both GD and TT (all χ^2<2.56, all ps>.15), suggests that the between-group difference in processing time of HP words is specific to the Fans' domain knowledge, rather than a general difference in reading ability / processing time.

3.1.4. Context × Word Type

A significant interaction between Context and Word Type was observed across all measures (all χ^2s>16.83, all ps<.002) and associated post-hoc comparisons are summarised in

Table 5 below. Figures 2-4 illustrate the Context × Word Type interactions on FFD, GD and TT respectively.

Table 5: Post-hoc comparisons for Context × Word Type interactions across measures

Context	**Word Type**		**FFD**	**GD**	**TT**
HP	HP vs. HF	diff.	24	163	220
		χ^2	11.48	11.22	34.97
		p	.006	.007	<.001
	HP vs. LF	diff.	3	83	119
		χ^2	<1	<1	5.70
		p	ns	ns	ns
	HF vs. LF	diff.	27	80	101
		χ^2	16.60	14.34	14.51
		p	<.001	.001	.001
HF	HP vs. HF	diff.	39	189	394
		χ^2	32.57	39.18	116.47
		p	<.001	<.001	<.001
	HP vs. LF	diff.	14	104	204
		χ^2	2.62	5.42	19.00
		p	ns	ns	<.001
	HF vs. LF	diff.	25	85	190
		χ^2	15.62	21.58	48.47
		p	<.001	<.001	<.001
LF	HP vs. HF	diff.	40	204	354
		χ^2	33.05	24.75	80.07
		p	<.001	<.001	<.001
	HP vs. LF	diff.	37	192	383
		χ^2	26.54	19.23	80.95
		p	<.001	<.001	<.001
	HF vs. LF	diff.	3	12	29
		χ^2	<1	<1	<1
		p	ns	ns	ns

Note. HP = Harry Potter; HF = high frequency; LF = low frequency. diff = mean difference in ms rounded to nearest whole number. χ^2 values rounded to 2DP; *p*-values rounded to 3DP. ns = p>.15. All *df*s = 1.

[Insert Figure 2 about here]

[Insert Figure 3 about here]

[Insert Figure 4 about here]

For target words presented in HP contexts, HF words demonstrated a processing advantage over both HP and LF targets across all measures (all *p*s<.007; see Table 5, Figures 2-4). Interestingly, when presented in HP contexts, there were no significant differences between the processing of HP targets vs. LF targets across all measures (all χ^2s<5.70, all *p*s>.15; see Table 5, Figures 2-4). Thus, participants are able to utilise contextual information to equalize the processing of HP words relative to LF targets.

Considering contexts designed to support HF targets, it is evident that HF words are processed significantly faster than both HP and LF targets across all measures (all *p*s<.001; see Table 5, Figures 2-4). In FFD and GD, there were no observed differences between HP and LF targets (both χ^2s<5.42, both *p*s>.15; see Table 5, Figures 2 and 3), suggesting that participants can quickly resolve that both target types are equally unsuitable given the developing contextual frame. However, in TT – typically regarded as a measure reflective of later processing and integration, as opposed to access – HP words were processed significantly slower than LF words (*p*<.001, see Table 5 and Figure 4). This suggests that (collapsed across participant groups), processing and integration of real-world LF words is more straightforward than fictional HP words when both target types were unsupported by the contextual frame.

Furthermore, we examined the processing of HP, HF and LF words in contexts supportive of LF targets. The supportive context equalized the processing times of LF and HF targets across all measures (all χ^2s<1; see Table 5, Figures 2-4). Analyses revealed that LF targets presented in supportive LF contexts were processed significantly faster than HP words across all measures (all χ^2s>19.23, all *p*s<.001; see Table 5, Figures 2-4). Finally, HF words were processed significantly faster than HP words across all measures (all χ^2s>24.75, all *p*s<.001; see Table 5, Figures 2-4). This suggests that (collapsed across participant groups), processing and integration of real-world HF words is more straightforward than fictional HP words when both target types are unsupported by the contextual frame.

3.2. Measures of Regression Behaviour

We additionally considered the probability of readers making regressive saccades – either after initial processing of the target (Regressions Out) or back into the target region after later processing (Regressions In). Descriptive statistics are presented in Table 6.

Table 6: Descriptive statistics across conditions and measures of regression behaviour

		Regressions Out			Regressions In		
Context	**Group**	**HP word**	**HF word**	**LF word**	**HP word**	**HF word**	**LF word**
HP	Fans	.21 (.41)	.17 (.38)	.29 (.45)	.16 (.37)	.21 (.41)	.21 (.41)
	Non-Fans	.19 (.39)	.11 (.31)	.19 (.40)	.24 (.43)	.17 (.38)	.19 (.40)
HF	Fans	.21 (.41)	.12 (.32)	.20 (.40)	.22 (.41)	.08 (.27)	.25 (.44)
	Non-Fans	.18 (.38)	.15 (.35)	.21 (.41)	.22 (.42)	.05 (.22)	.14 (.35)
LF	Fans	.24 (.43)	.16 (.37)	.19 (.39)	.22 (.41)	.21 (.41)	.11 (.31)
	Non-Fans	.14 (.35)	.19 (.40)	.20 (.40)	.22 (.41)	.09 (.29)	.05 (.23)
95% CIs							
HP	Fans	[.07,.35]	[.04,.31]	[.13,.44]	[.03,.29]	[.07,.35]	[.07,.35]
	Non-Fans	[.02,.35]	[.00,.24]	[.03,.36]	[.06,.42]	[.01,.33]	[.03,.36]
HF	Fans	[.07,.35]	[.01,.23]	[.06,.34]	[.08,.36]	[.00,.17]	[.10,.41]
	Non-Fans	[.02,.34]	[.00,.29]	[.04,.38]	[.05,.39]	[.00,.14]	[.00,.29]
LF	Fans	[.09,.39]	[.03,.29]	[.05,.32]	[.08,.36]	[.07,.35]	[.00,.21]
	Non-Fans	[.00,.28]	[.03,.36]	[.03,.37]	[.04,.39]	[.00,.21]	[.00,.15]

Note. Regression probability means (*SD*s) rounded to 2DP. 95% CIs = lower and upper confidence estimates

The process of inferential analysis was essentially the same for regression data as for fixation duration measures. However, as regression data is binomial, we used a generalized linear mixed effects approach using the 'glmer' command and added the argument "*family = binomial*". Across regression measures, all models included random intercepts across subjects and items. A summary of effects and interactions across measures is presented in Table 7. Only the interactions between Group, Context and Word Type are described in detail in this section; for isolated fixed effects of each variable and follow-up comparisons (where appropriate), please see Supplemental Material.

Table 7: Summary of effects on regression probability measures

		Regressions Out		**Regressions In**	
	df	χ^2	*p*	χ^2	*p*
Group × Context × Word Type	4	6.77	.149	<1	ns
Group × Context	2	2.87	ns	10.33	.006
Group × Word Type	2	1.40	ns	17.54	<.001
Context × Word Type	4	4.67	ns	36.90	<.001
Group	1	1.67	ns	1.66	ns
Context	2	1.07	ns	12.80	.002
Word Type	2	12.80	.002	3.93	.140

Note. χ^2 values rounded to 2DP; *p*-values rounded to 3DP. ns = p>.15.

3.2.1. Regressions Out

After constructing intercept-only models, model fit across Regressions Out data improved by adding a random slope of Word across subjects (χ^2=11.83, p=.037). All other iterations either did not improve upon these models or failed to converge. No significant interactions between Group, Context and / or Word Type were observed (see Table 7).

3.2.2. Regressions In

After constructing intercept-only models, model fit across Regressions In data improved by adding a random slope of Word across items (χ^2=52.86, p<.001). All other iterations either did not improve upon these models or failed to converge. The three-way interaction between Group, Context and Word Type was non-significant (see Table 7).

3.2.2.1. Regressions In - Group × Context

A significant interaction between Group and Context was observed (see Table 7). Post-hoc comparisons are displayed in Table 8, and the interaction is illustrated in Figure 5.

Table 8: Post-hoc comparisons for Group × Context interactions

Group	**Context**		**RegIn**	**Group**	**Context**		**RegIn**
Fans	HP vs. HF	diff.	.00	Non-Fans	HP vs. HF	diff.	.06
		χ^2	<1			χ^2	9.18
		p	ns			p	.015
	HP vs. LF	diff.	.01		HP vs. LF	diff.	.08
		χ^2	<1			χ^2	13.57
		p	ns			p	.001
	HF vs. LF	diff.	.00		HF vs. LF	diff.	.02
		χ^2	<1			χ^2	<1
		p	ns			p	ns

Note. RegIn = Regressions In; HP = Harry Potter; HF = high frequency; LF = low frequency. diff = mean difference in probability rounded to 2DP. χ^2 values rounded to 2DP; p-values rounded to 3DP. ns = p>.15. All *df*s = 1.

[Insert Figure 5 about here]

When considering Fan Regressions In data, there were no significant differences between HP, HF or LF contexts (collapsing across target word types; all χ^2s<1, see Table 8).

However, Non-Fans made significantly more regressions into target regions in HP contexts (20.0%) than HF contexts (13.7%; χ^2=9.18, *df*=1, *p*=.015), and Non-Fans made significantly more regressions into target regions in HP contexts (20.0%) than LF contexts (12.0%; χ^2=13.57, *df*=1, *p*=.001). There was no difference between the proportion of regressions into target regions made by Non-Fans between HF and LF contexts (χ^2<1).

3.2.2.2. Regressions In - Group × Word Type

The Group × Word Type interaction was significant (see Table 7 and Figure 5). Post-hoc comparisons are displayed in Table 9. As with the Group × Context interaction, Fan Regressions In data showed no significant differences between HP, HF or LF target word types (collapsing across contexts; all χ^2s<3.32, all *p*s>.15; see Table 8). Non-Fans made significantly more regressions into target regions containing HP targets (22.7%) than HF targets (10.3%; χ^2=26.22, *df*=1, *p*<.001), and Non-Fans made significantly more regressions into target regions containing HP targets (22.7%) than LF targets (12.7%; χ^2=18.48, *df*=1, *p*<.001). There was no difference between the proportion of regressions into target regions made by Non-Fans between HF and LF targets (χ^2=1.23, *df*=1, *p*>.15).

Table 9: Post-hoc comparisons for Group × Word Type interactions

Group	Word Type		RegIn	Group	Word Type		RegIn
Fans	HP vs. HF	diff.	.03	Non-Fans	HP vs. HF	diff.	.13
		χ^2	3.32			χ^2	26.22
		p	ns			*p*	<.001
	HP vs. LF	diff.	.01		HP vs. LF	diff.	.10
		χ^2	<1			χ^2	18.48
		p	ns			*p*	<.001
	HF vs. LF	diff.	.02		HF vs. LF	diff.	.03
		χ^2	1.20			χ^2	1.23
		p	ns			*p*	ns

Note. RegIn = Regressions In; HP = Harry Potter; HF = high frequency; LF = low frequency. diff = mean difference in probability rounded to 2DP. χ^2 values rounded to 2DP; *p*-values rounded to 3DP. ns = *p*>.15. All *df*s = 1.

3.2.2.3. Regressions In - Context × Word Type

A significant Context × Word Type interaction was observed (see Table 7). Table 10 details the associated post-hoc comparisons. This interaction is illustrated by Figure 6.

Table 10: Post-hoc comparisons for Context × Word Type interaction – Regressions In

Context	**Word Type**		**Regressions In**
HP	HP vs. HF	diff.	.01
		χ^2	<1
		p	ns
	HP vs. LF	diff.	.00
		χ^2	<1
		p	ns
	HF vs. LF	diff.	.01
		χ^2	<1
		p	ns
HF	HP vs. HF	diff.	.16
		χ^2	25.30
		p	<.001
	HP vs. LF	diff.	.03
		χ^2	<1
		p	ns
	HF vs. LF	diff.	.13
		χ^2	19.93
		p	<.001
LF	HP vs. HF	diff.	.07
		χ^2	4.98
		p	ns
	HP vs. LF	diff.	.14
		χ^2	22.81
		p	<.001
	HF vs. LF	diff.	.07
		χ^2	5.77
		p	.147

Note. HP = Harry Potter; HF = high frequency; LF = low frequency. diff = mean difference in probability rounded to 2DP. χ^2 values rounded to 2DP; *p*-values rounded to 3DP. ns = *p*>.15. All *df*s = 1.

[Insert Figure 6 about here]

When considering HP-supportive context frames, there were no differences between target word types in Regressions In data (all χ^2s<1; see Table 9 and Figure 6). When contexts were supportive of HF target words, participants made substantially fewer regressions into HF target regions (6.50%) than HP target regions (22%; χ^2=25.30, *df*=1, *p*<.001), and significantly fewer regressions into HF target regions (6.50%) than LF target regions (19.5%; χ^2=19.93, *df*=1, *p*<.001). When contexts were supportive of HF targets, there was no significant difference between the proportion of regressions made into HP or LF target regions (χ^2<1). Finally, when contexts were supportive of LF targets, participants made significantly fewer regressions into target regions containing LF targets (8.0%) than HP targets (22%; χ^2=22.81, *df*=1, *p*<.001); however, there was no significant difference between the proportion of regressions into targets containing LF words (8.0%) and HF targets (15.0%; χ^2=5.77, *df*=1, *p*=.147). Additionally, when contexts were constructed to be supportive of LF targets, there was no significant difference in the proportion of regressions into target regions containing HP (22.0%) and HF targets (15.0%; χ^2=4.98, *df*=1, *p*>.15).

4. Discussion

In this study we investigated the reading of words taken from the fictional domain of Harry Potter (HP). In an eye-tracking experiment we examined if familiarity with the HP domain would facilitate processing of fictional words from HP when presented in a supportive context. Evidence from gaze duration and total reading time measures show that those with high HP domain knowledge read HP words faster than those with low domain knowledge. These results provide evidence for an early, but not immediate, effect of domain knowledge on lexical processing as the same advantage was not seen in first fixation duration times. No differences were found between groups for other word types indicating that domain knowledge rather than differences in reading ability is responsible for this effect.

Differences in patterns of regressions also show the influence of domain knowledge on processing domain-specific information. When the context was supportive of an HP word Non-Fans made more regressions into target regions than in other contexts, and, regardless of context, Non-Fans made more regressions into HP targets than LF and HF targets. Fans showed no differences in regressions into HP and either HF or LF targets. Neither group showed differences in regressions into LF or HF contexts or targets, supporting a pattern of effects reliant on domain knowledge rather than reading behaviour.

When collapsing levels of domain knowledge and treating participants as a single group, domain-relevant information can still be seen to influence processing. When presented in a HP supportive context there was no significant difference in reading times between a HP target word and a LF target word. See Supplemental Material B and C for further comparison of HP and LF words in HP contexts. Reading of HP and LF targets appear to be equalised by a supportive context. However, when a sentence context supports a HF word, reading of HP words is slower than that of LF words in the later, total reading time measure. This suggests that domain-specific, fictional words have been reread when presented in HF supportive sentences. Where sentences supported LF targets, reading of LF and HF targets are equalised. In this context HP targets are read more slowly than other target words suggesting readers are using the context to constrain candidates, and that novel/familiar words are difficult to integrate.

Patterns of Regressions Out suggest similar effects. Readers made more regressions out of HP targets that HF targets, and more regressions out of LF targets than HF targets, but no difference was observed between regressions out of HP and LF targets. Regardless of reader and context both HP and LF targets may cause the reader to seek to re-read a section of text.

Within HP contexts, regressions into the target were equal across word types. In HF contexts HF targets were regressed into less frequently than HP and LF targets, which did not differ. In LF contexts, LF targets were regressed into less frequently than HP and HF targets, which did not differ. No difference in regressions into HP and LF targets in HF contexts contrasts with total reading times where there was a difference between these word types in this context. This result suggests that when an HP word is presented in a HF context, readers regress out at some point after the target and regress back into the passage at a point before the target word to allow rereading.

Domain knowledge has clearly facilitated processing. Fans of HP were faster to read words from the HP domain than Non-Fans. However, this effect was not immediate and was not necessarily reliant on contextual support. These differences align with previous evidence that real-world or specialist knowledge can influence early semantic processing (Filik & Leuthold, 2013; Hald, Steenbeek-Planting & Hagoort, 2007; Troyer & Kutas, 2018; Troyer et al., 2019). Where these previous studies have found this effect through a combination of world knowledge and contextual support, our results demonstrate that those with high domain knowledge read fictional HP words faster regardless of the preceding context. Previous research has also demonstrated that prior knowledge and discourse context can influence availability of word meanings. Access to the subordinate meaning of a homograph is slowed by strong domain knowledge which supports the dominant meaning (Wiley et al., 2018) and strong domain knowledge of the subordinate meaning can increase its availability (Rodd et al., 2016). The current study extends these findings, showing that knowledge of a fictional domain can benefit the processing of "fictional" words which are only relevant within a specific universe, generally have a lack of grounding in reality, and cannot be embodied.

It is less clear if domain knowledge and contextual support have combined to allow early lexical access to fictional HP words. In all reading time measures, when the context was supportive of HP target words, no significant difference was found between time spent on a LF target and a HP target. See Supplemental Material C for further consideration of differences between these reading times. This could suggest that when a domain-specific fictional word is presented in a supportive context, readers make immediate use of context allowing the time course of lexical access to the fictional word to equate to that of a low frequency word. If this were the case results would lend support to interactive (e.g., McClelland, 1987) as opposed to modular (e.g., Fodor, 1983) accounts of language processing. If context were not active in selecting target candidates, there would be clearer differences of HP vs. LF word contrasts in HP-supportive contexts in FFD (assumed to reflect lexical access; Rayner 1998, 2009). However, this model fails to explain why no differences were found across both Fan and Non-fan groups. One explanation is that the ubiquitous nature of Harry Potter allowed even those with low domain knowledge to recognise key information within contexts and target words. Indeed, it would be difficult to find someone who does not know the meaning of the word "*muggle*". This issue may have been exacerbated by the use of contextual 'support' rather than contextual 'predictability' within sentence items. Alternatively, Chaffin et al., (2001) found no early differences between processing of novel and low-familiarity words but noted that processing costs were seen later, in rereading and regressions, hence results were taken to be consistent with the E-Z Reader model of lexical access (e.g., Reichle et al., 1998). Regression patterns of Non-Fans reflect modular processing. Non-Fans regressed into HP targets more than LF targets in all contexts suggesting the equivalence of FFD times was based on extensive processing of basic word features such as orthography. In contrast, Fans showed no difference in regressions into HP and LF words. This remains consistent with Chaffin et al.'s results since Fans would be expected to treat HP words as low-familiarity rather than novel.

Given previous research (e.g., Troyer & Kutas, 2018; Troyer et al., 2019) we would have expected to have seen a stronger interaction between domain knowledge and context on reading times of domain-specific target words, particularly in later, if not immediate measures. The absence of this interaction may reflect the binary nature of domain knowledge applied in this study, that is, participants were allocated to Fans and Non-Fan groups based on self-reported incidence of book reading and movie viewing. However, the negligible incidence of reading/viewing in the Non-Fans group assumes very low knowledge of the Harry Potter domain. For ambiguous words, Wiley and Rayner (2000) found that prior knowledge aided lexical access of all but very infrequent subordinate meanings, whilst Kaakinen and Hyönä (2007) found higher prior knowledge facilitated processing in first-pass measures of gaze duration and probability of skipping a word. Perhaps the integration of Harry Potter knowledge into popular culture is such that even those without direct contact have enough prior knowledge to aid processing. Troyer and colleagues used more in-depth examination of HP knowledge to establish low and high domain knowledge participant groups. Furthermore, in single-trial ERP analysis Troyer et al. (2019) used a measure of domain knowledge, via a trivia quiz, as a fixed effect within their analysis model. Given the wide-reaching nature of the Harry Potter universe, the use of a continuous, rather than discrete, binary measure of domain knowledge may lead to a clearer picture of the combined effects of domain knowledge and context.

An interesting question, beyond the scope of this current study, is how the fictional HP words may have become incorporated into the lexicon? As it has been consistently demonstrated that children in early school years learn new words through incidental exposure such as through stories (Akhar, 2004), and that from mid-childhood onwards new words are learned by reading (Nagy, Anderson & Herman, 1987), it seems likely that those in the Fan group learned the words through reading (or *hearing*) the HP series of books. As no participant in the Non-Fan group reported having read a HP novel, most exposure to these words would be without context or (internal) auditory support. Collecting a more in-depth measure of domain knowledge might allow exploratory analysis on *how* exposure relates to word processing. Furthermore, a measure of how old each participant in the Fan group was when first reading, or being read, each novel could help establish an age of acquisition effect for fictional words, since the influence of age of word learning on speed of word processing is well established (e.g., Brysbaert & Ghysekinck, 2006; Johnston & Barry, 2006; Juhasz, 2005).

When examining HF and LF in isolation it is apparent that an LF-supportive context facilitated the processing of LF words such that the typical word frequency effect was not evident – significant differences in fixation durations between HF and LF targets were only evident when HF words were presented in an HF-supportive context. This pattern is consistent with previous eye-tracking research in which the relationship between contextual support and word frequency was under specific investigation (e.g., Hand et al., 2010; Sereno et al., 2018). It is possible that differences between processing of HF and LF words in a HF context have arisen through issues with the plausibility of LF target words in HF contexts. Rayner, Warren, Juhasz and Liversedge (2004) found words which were implausible in a given context led to greater processing difficulty in later fixation measures. More recently, implausible words have been shown to affect earlier fixation measures (Staub, Rayner, Pollatsek, Hyönä, & Majewski, 2007) however, evidence suggests that whilst the effects of plausibility may be detectable in early fixation measures, the influence of plausibility on reading may be post-lexical (Abbott & Stuab, 2015; Veldre, Reichle, Wong & Andrews, 2020). When testing our target words for plausibility (see Supplemental Material D for full details of this process and results) we found that each target word was rated most-plausible within its own supportive context. Whilst plausibility could lead to a difference in fixation measures, when a less plausible word is presented in a non-supportive context the well-documented immediate effect of word frequency on processing is more likely to be responsible for this discrepancy (see Hand, Millet, O'Donnell & Sereno, 2010 for a review). Indeed, Abbott and Staub (2015) found that their LF plausible words were rated as significantly less plausible than their HF counterparts; they noted that such a difference should be expected and presented evidence that plausibility ratings of LF plausible words correlated with log frequency of LF words, whereas no such correlation was found for other word types. Further, in eye movement analysis, previous research (Rayner et al., 2004; Staub et al., 2007; Abbott & Staub 2015; Veldre et al., 2020) generally treats plausibility as a binary factor where a threshold is applied during norming to determine *implausible* targets. Whilst some of our target words may be less plausible in certain context frames, ratings for non-HP targets (i.e., HF and LF targets) do not suggest these words were implausible. Controlling for plausibility across Contexts and Word Types may be prudent in future research but is unlikely to have driven the relationship between contextual support and word frequency in the current study.

4.1. Conclusion

Factors which affect the processing of written words have been extensively studied. The reading of niche fiction, based in alternate reality, has become increasingly popular. This novel study assesses the effect of domain knowledge on the lexical access of words from fiction in supportive and unsupportive contexts. Fictional words from alternate realities are processed faster by readers with higher domain knowledge. The nature and time-course of the combined effects of domain knowledge and contextual support in the current study provided more support for modular accounts of linguistic processing (e.g., Fodor, 1983), and align well with serial models of eye movement control (e.g., *E-Z Reader*; Reichle et al., 1998). Patterns of regressive eye movements suggested pronounced effects of domain knowledge on reading. Overall, our results illustrate that both domain knowledge and contextual support play a role in processing of fictional words, although these factors may not influence *immediate* lexical access.

Competing Interests

Neither author has any competing interests to declare.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

[1]Footnote

We additionally performed a separate analysis of our eye movement data, considering nested / multilevel models of our data, specifically, treating Group as a level two variable / effect. The pattern of significance outcomes found in the multilevel models was identical across measures and analyses to the results presented below – for simplicity we have included only our single-level omnibus analysis.

References

Abbott, M.J., & Staub, A. (2015). The effect of plausibility on eye movements in reading: Testing E-Z Reader's null predictions. *Journal of Memory and Language, 85*, 76-87. https://doi.org/10.1016/j.jml.2015.07.002

Akhtar, N. (2004). Contexts of early word learning. In O. G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon* (pp. 485-507). Cambridge, MA: MIT Press.

Anderson, R.C. (1984). Role of the reader's schema in comprehension, learning, and memory. In R. C. Anderson, J. Osborn & R. J. Tierney (Eds.). *Learning to read in American schools: Basal readers and content texts* (pp. 243–257). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Balota, D.A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, *17*(3), 364-390. https://doi.org/10.1016/0010-0285(85)90013-1

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*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Brysbaert, M., & Ghyselinck, M. (2006). The effect of age of acquisition: Partly frequency related, partly frequency independent. *Visual Cognition, 13*(7-8), 992-1011. https://doi.org/10.1080/13506280544000165

British Psychological Society (2014). *Code of Human Research Ethics*. Leicester: BPS.

Chaffin, R., Morris, R.K., & Seely, R.E. (2001). Learning new word meanings from context: study of eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(1), 225–235. https://doi.org/10.1037/0278-7393.27.1.225

Davies, M. (2004-). BYU-BNC. (Based on the British National Corpus from Oxford University Press). Available online at http://corpus.byu.edu/bnc/.

De Rosario-Martinez, H. (2015). Phia: Post-hoc interaction analysis. R package version 0.2.1. Retrieved from https://CRAN.R-project.org/package=phia

Duffy, S.A., Morris, R.K., & Rayner, K. (1998). Lexical ambiguity and fixation times in reading. Journal of Memory and Language, 27(4). 429-446. https://doi.org/10.1016/0749-596X(88)90066-6.

Filik, R., & Leuthold, H. (2013). The role of character-based knowledge in online narrative comprehension: Evidence from eye movements and ERPs. *Brain Research, 1506*, 94-104. https://doi.org/10.1016/j.brainres.2013.02.017

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

Hald, L.A., Steenbeek-Planting, E. G., & Hagoort, P. (2007). The interaction of discourse context and world knowledge in online sentence comprehension. Evidence from N400. Brain Research, 1146, 210-218. https://doi.org/10.1016/j.brainres.2007.02.054

Hand, C.J., Miellet, S., O'Donnell, P.J., & Sereno, S.C. (2010). The frequency-predictability interaction in reading: It depends where you're coming from. *Journal of Experimental Psychology: Human Perception and Performance, 36*(5)*,* 1294-1313. https://doi.org/10.1037/a0020363

Hand, C.J., O'Donnell, P.J., & Sereno, S.C. (2012). Word-initial letters influence fixation durations during fluent reading. *Frontiers in Psychology: Language Sciences, 3,* 1-19. https://doi.org/10.3389/fpsyg.2012.00085

Ingram, J., Hand, C.J., & Moxey, L.M. (2014). Processing inferences drawn from the logically equivalent frames half full and half empty. *Journal of Cognitive Psychology, 26*(7). pp. 799-817. https://doi.org/10.1080/20445911.2014.956747

Johnston, R.A., & Barry, C. (2006). Age of acquisition and lexical processing. *Visual Cognition, 13*(7-8). 789-845. https://doi.org/10.1080/13506280544000066

Juhasz, B.J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin, 131*(5), 684-712. https://doi.org/10.1037/0033-2909.131.5.684

Just, M.A., & Carpenter, P.A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87,* 329-354. https://doi.org/10.1037/0033-295X.87.4.329

Kaakinen, J.K., & Hyönä, J. (2007). Perspective effect in repeated reading: An eye movements study. *Memory & Cognition, 35*(6), 1323-1336. https://doi.org/10.3758/BF03193604

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*(1-2)*,* 262-284. https://doi.org/10.1080/09541440340000213

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

McClelland, J.L. (1987). The case for interactionism in language processing. In M. Coltheart (Ed.), *Attention & performance XII: The psychology of reading* (pp. 1–36). London: Erlbaum.

Nagy, W.E., Anderson, R.C., & Herman, P.A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal, 24*(2), 237-270. https://doi.org/10.2307/1162893

Paul, S.T., Kellas, G., Martin, M., & Clark, M.B. (1992). The influence of contextual features on the activation of ambiguous word meanings. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*(4), 703-717. https://doi.org/10.1037/0278-7393.18.4.703

Philips, N.D. (2018). YaRrr! The Pirate's Guide to R. https://bookdown.org/ndphillips/YaRrr/

Pollatsek, A., Slattery, T.J., & Juhasz, B.J. (2008). The processing of novel and lexicalized prefixed words in reading. *Language and Cognitive Processes, 23*, 1133–1158. https://doi.org/10.1080/01690960801945484

R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R foundation for Statistical Computing. Available from https://www.R-project.org/

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

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62,* 1457-1506. https://doi.org/10.1080/17470210902816461

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E.D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z Reader model. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 720-730. https://doi.org/10.1037/0096-1523.30.4.720

Rayner, K., Sereno, S.C., & Raney, G.E. (1996). Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance, 22*(5)*,* 1188-1200. https://doi.org/10.1037/0096-1523.22.5.1188

Reichle, E.D., Pollatsek, A., Fisher, D.L., Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*(1), 125-157. https://doi.org/10.1037/0033-295X.105.1.125

Rodd, J.M., Cai, Z.G., Betts, H.N., Hanby, B., Hutchinson, C., & Adler, A. (2016). The impact of recent and long-term experience on access to word meanings: Evidence from large-scale internet-based experiments. *Journal of Memory and Language, 87*, 16–37. https://doi.org/10.1016/j.jml.2015.10.006

Sereno, S.C., Hand, C.J., Shahid, A., Yao, B., & O'Donnell, P.J. (2018). Testing the limits of contextual constraint: Interactions with frequency and parafoveal preview during fluent reading. *Quarterly Journal of Experimental Psychology*, *71*(1), 302-313. https://doi.org/10.1080/17470218.2017.1327981

Spilich, G.J., Vesonder, G.T., Chiesi, H.L., & Voss, J.F. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning & Verbal Behavior, 18*(3), 275–290. https://doi.org/10.1016/S0022-5371(79)90155-5

Staub, A., Rayner, K., Pollatsek, A., Hyönä, J., & Majewski, H. (2007). The time course of plausibility effects on eye movements in reading: Evidence from noun–noun compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(6), 1162-1169. https://doi.org/10.1037/0278-7393.33.6.1162

Troyner, M., & Kutas, M. (2018). Harry Potter and the Chamber of *What*?: The impact of what individuals know on word processing during reading. *Language, Cognition and Neuroscience*, https://doi.org/10.1080/23273798.2018.1503309

Troyer, M., Urbach, T.P., & Kutas, M. (2019). Lumos!: Electrophysiological tracking of (wizarding) world knowledge use during reading. *Journal of Experimental Psychology: Learning, Memory and Cognition.* https://doi.org/10.1037/xlm0000737

Veldre, A., Reichle., E.D, Wong, R., Andrews, S. (2020). The effect of contextual plausibility on word skipping during reading. *Cognition, 197*, 104184. https://doi.org/10.1016/j.cognition.2020.104184

Voss, J.F., Vesonder, G.T., & Spilich, G.J. (1980). Text generation and recall by high-knowledg e and low-knowledge individuals. *Journal of Verbal Learning and Verbal Behavior, 19,* 651–667. https://doi.org/ 10.1016/S0022-5371(80)90343-6

Wiley, J., George, T., & Rayner, K. (2018). Baseball fans don't like lumpy batters: Influence of domain knowledge on the access of subordinate meanings. *Quarterly Journal of Experimental Psychology, 71*(1), 93-102. https://doi.org/10.1080/17470218.2016.1251470

Wiley, J., & Rayner, K. (2000). Effects of titles on the processing of text and lexically ambiguous words: Evidence from eye movements. Memory & Cognition, 28(6), 1011-1021. https://doi.org/10.3758/BF03209349

Williams, R.S., & Morris, R.K. (2004). Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology, 16*(1-2), 312-339. https://doi.org/10.1080/09541440340000196

Figure Legends

Figure 1: Group × Word Type interactions across Gaze Duration and Total Time. HP = Harry Potter; HF = high frequency; LF = low frequency. Error bars represent standard error of mean.

Figure 2: Context × Word Type interaction – First Fixation Duration. HP = Harry Potter; HF = high frequency; LF = low frequency. Each line represents a different Word Type. Error bars represent standard error of mean.

Figure 3: Context × Word Type interaction – Gaze Duration. HP = Harry Potter; HF = high frequency; LF = low frequency. Each line represents a different Word Type. Error bars represent standard error of mean.

Figure 4: Context × Word Type interaction – Total Time. HP = Harry Potter; HF = high frequency; LF = low frequency. Each line represents a different Word Type. Error bars represent standard error of mean.

Figure 5: Group × Context and Group × Word Type interactions – Regressions In. HP = Harry Potter; HF = high frequency; LF = low frequency. Error bars represent standard error of mean.

Figure 6: Context × Word Type interaction – Regressions In. HP = Harry Potter; HF = high frequency; LF = low frequency. Each line represents a different Word Type. Error bars represent standard error of mean.