The Effects of Predictability and Stimulus Quality on Lexical Processing: Evidence from the Coregistration of Eye Movements and EEG

The Effects of Predictability and Stimulus Quality on Lexical Processing: Evidence from the Coregistration of Eye Movements and EEG

A Thesis Presented

by

JON T. BURNSKY

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

Master of Science

February 2021

Psychological and Brain Sciences

The Effects of Predictability and Stimulus Quality on Lexical Processing: Evidence from the Coregistration of Eye Movements and EEG

A Thesis Presented

by

JON T. BURNSKY

Approved as to style and content by:

____
Adrian Staub, Chair

____
Lisa Sanders, Member

____
Brian Dillon, Member

____
Caren Rotello, Department Chair
Psychological and Brain Sciences

ABSTRACT

The Effects of Predictability and Stimulus Quality on Lexical Processing: Evidence from the Coregistration of Eye Movements and EEG

FEBRUARY 2021

JON T. BURNSKY, B.A., UNIVERSITY OF MARYLAND, COLLEGE PARK

M.S., UNIVERSITY OF MASSACHUSETTS, AMHERST

Directed by: Professor Adrian Staub

A word's predictability has been shown to influence its processing. Two methodologies have demonstrated this time and again: eye tracking while reading and Event Related Potentials (ERPs). In eye tracking while reading, words that are made predictable by their contexts (as operationalized by the cloze task; Taylor, 1953) receive shorter first fixation times (Staub, 2015, for a review) as well as shorter gaze duration and increased skipping rate. In ERPs, the N400 component's amplitude has also been shown to inversely correlate with a word's predictability (Kutas and Federmeier, 2011, for a review). Despite the similarities, there is much reason to suspect that these two measures are reflections of different underlying cognitive processes, both modulated by a word's predictability. We utilized the simultaneous collection of EEG and eye tracking data to investigate the differential effects of lexical predictability and stimulus quality on these measures. We found that these two manipulations had additive effects in the eye movement record, but yet only the manipulation of predictability influenced the N400 Fixation Related Potential (FRP) amplitude, with stimulus quality influencing neither the amplitude nor the latency of the N400. These findings provide no evidence for there being a role for predictability in early visual processing, and thus call into question the relative ordering of

lexical processing effects laid out in Staub and Goddard (2019). Our findings also suggest that the N400's underlying process is strictly temporally fixed and indexes the lexical processing difficulty left after there has already been a convergence of evidence towards the identity of the observed stimulus.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Lexical Predictability

As sentences unfold, they constrain what upcoming words are likely. Often, the constraint of the sentence makes some upcoming words so predictable that when subjects are asked to provide a likely next word, they will converge on one to a few completions. Example 1 is a constraining sentence for which many people provide "sugar" (Kutas and Hillyard, 1980)

1. I like to drink my coffee with cream and ____

Collecting completions of sentence fragments like this is referred to as the Cloze Task (Taylor, 1953). Researchers end up with a "cloze probability": a number between 0 and 1 that indicates the count of the responses for some target word divided by the count of all responses (1 means all participants provided the intended target word; .5 means half of the participants did, etc.). The cloze task can be untimed or with a time limit, and done in the lab or online, and it is by far the most widely used method for operationalizing a word's predictability in its context. A hypothesis of the underlying mechanism relating cloze responses to predictability is laid out in Staub, Grant, Astheimer, and Cohen (2015).

1.2 The Effect of Predictability on Eye Tracking while Reading Measures

Having an operational definition of a word's predictability given its context opens up a vast literature exploring how predictability affects lexical processing. One of the most widely used methodologies for doing this is eye tracking while reading. A well-established finding is that the first fixation on a word (the time from a reader directly fixating a word for the first time, until they saccade) is inversely related to that word's cloze probability (Ehrlich and Rayner, 1981; Kliegl, Grabner, Rolfs, and Engbert 2004; Staub, 2015; Frisson, Harvey, and Staub, 2017; a.o.). Examples from Rayner, Slattery, Drieghe, and Liversedge (2011) are provided below. *Restaurant* has a higher cloze probability in 2 than in 3, and so has shorter first fixation times when presented with the context in 2 than that in 3. Predictable words also receive shorter gaze duration times, and are skipped more frequently compared to unpredictable words.

2. Linda was very hungry and wanted to get food. Raul decided to take Linda to a *restaurant* downtown that served Italian food. (mean first fixation = 195ms)

3. Linda and Raul are close friends from college. Raul decided to take Linda to a *restaurant* downtown that served Italian food. (mean first fixation = 211ms)

A question that arises when considering the effect of predictability on the first fixation is whether this is an effect on the early orthographic processing of the word in question, or something later like accessing lexico-semantic information from memory. Staub and Goddard (2019) addressed this question using a parafoveal preview manipulation crossed with a predictability manipulation. With a parafoveal preview manipulation, the participant is denied seeing a target word until they fixate it directly. Rather, another word, or a non-word letter string (called an "invalid preview"), replaces the target word while they are fixating the previous word. The invalid preview changes into the target word as the participant saccades past an invisible boundary; participants rarely detect the change. The authors assume this manipulation only affects the earliest stages of lexical processing, those that take place in the parafovea where visual acuity is markedly diminished. Staub and Goddard found that when participants are provided with an invalid preview, the predictability effect on the first fixation on the target word is eliminated. Predictability effects in reading appear to be limited to cases where there is valid parafoveal preview of the target word. The interaction of predictability and preview validity suggests that they influence the same stage of lexical processing, and the authors argue that this stage is likely to be quite early.

This view of the predictability effect has been recently put to the test in Staub (2020). They manipulated the predictability of words in their context as well as the contrast between the text and the background of the screen (for a similar manipulation see White and Staub, 2012).The background of the screen was kept constant at a light shade of grey, while the color of the text was either black, or shade of grey only slightly darker than the background. The factors of predictability and stimulus quality did not produce a significant interaction; the predictability effect was the same size for both clearly visible and diminished text, suggesting, as the author states, that either the true interaction effect is smaller than any eye tracking study to date would have the power to detect, or not present at all.

The Dual Route Cascaded (DRC) Model of reading laid out in Coltheart, Rastle, Perry, Langdon, and Ziegler (2001) and the Bi-modal Interactive Activation Model (BIAM) laid out in Grainger and Holcomb (2009) can assist in understanding what process the first fixation may be indexing given the manipulations that affect it and the interaction of those manipulations. The DRC for instance proposes that the process of lexical access in reading can be broken down into cascaded stages: Visual Feature processing –> Letter (grapheme) processing –> Orthographic Word (word form) processing –> Semantic processing (there are later stages needed for language comprehension but these are outside of the realm of this model). Creating a slowdown at an early stage has downstream consequences, and thus the ultimately measured slowdown in the first fixation can be due to an effect at any earlier stage.

The EZ Reader Model (Reichle, Rayner and Pollatsek, 2003) similarly makes a cut between stages of lexical access in reading. L1 is the first stage which is lexical processing all the way up to word form processing (sometimes called the familiarity check). L1 can be broken into subprocesses such as the first 3 stages laid out in the DRC, although these models are designed for different purposes. After L1 is complete, L2 begins, which is full lexical access including access to the semantic content of the word, but not including any integrative processes.

Logically, the predictability of word may influence any one of these stages. Suppose a reader is reading sentence 1) and before reaching the target word, they expect that the speaker / message will likely mention "sugar". The reader can preactivate the meaning of the word "sugar", which may preactivate the word form "sugar", which may in turn preactivate its component letters {"s" "u" "g" "a" "r"}, which may finally preactivate the curves and lines that make those shapes recognizable as letters. Upon actually encountering "sugar", each of these processes could be easier due to this preactivation, as well as other later integrative processes not included in the model.

There is already reason to believe that predictability does not in fact influence all of these stages of processing. Staub and Goddard (2019), Kretzschmar, Schlesewsky, and Staub (2015) among others have reliably found that predictability and a words' frequency do not interact; they are instead additive effects. This additive relationship suggests that it cannot be ruled out that the two factors have independent effects on the computations reflected by the first fixation duration (Sternberg, 1969). That is, within the framework laid by Sternberg, given the additive effects and the assumption of successive stages of processing between encountering a stimulus and saccade planning, the likely relationship between the factors and the stages themselves is a non-overlapping one, where frequency influences the duration of one processing stage, and predictability influences another. As Sternberg points out, there are exceptions to this logic where two factors may influence the same stage but happen to do so orthogonally, resulting in additive RT effects. If this were the case, the two factors would interact with other factors in tandem; there would be no cases of some factor producing an interaction with predictability but not frequency, say. This however is ruled out by the fact that frequency effects persist even in cases where the reader receives an invalid preview (Staub and Goddard, 2019), while predictability effects don't. This demonstrates that what has been observed is not an exception to the additive factors logic laid out in Sternberg (1969), and indeed predictability and frequency may influence separate processing stages. Staub and Goddard argue their findings imply that predictability functions at lower levels of processing while frequency functions at higher up levels of representation. This would predict that predictability and manipulations that influence those earlier stages should interact.

Stimulus quality manipulations are assumed to selectively affect visual feature processing or L1 (Reingold and Rayner, 2006), as there is no a priori reason to suspect the faintness of text to affect semantic, word form, or letter processing over and above the difficulty of extracting visual features. Given this assumption, if there were an interaction between predictability and stimulus quality this would suggest that predictability affects this earliest stage as well. However, if the two manipulations do not interact but are both reflected in the first fixation, predictability and stimulus quality may not influence processing at the same stage, and may instead operate on separate processing stages which are both computed in succession to inform the decision to move one's eyes to the next word in reading. For instance, a reader may have to complete all stages up to word form processing in order to activate the plan for a saccade to the next word, with stimulus quality slowing the first stage, and predictability slowing the second or third.

Parafoveal preview of a word is not as neatly localized to hindering or aiding a specific stage. Veldre and Andrews (2018) provide an argument that parafoveally previewed words may be analyzed well into the semantic processing stage at least occasionally, based on data demonstrating that lexical features of the preview like semantic relatedness influence the reading time of the target. However, Staub and Goddard assume that preview manipulations influence only earlier stages, and thus conclude that predictability too influences, at least partially, these earlier stages as well. Regardless of which stage(s) preview validity influences, the fact remains that predictability and preview validity do interact which suggests that the two manipulations influence some stage concurrently and similarly. What stage(s) predictability and preview validity simultaneously influence exactly is an open question. The central goal of the current study is to attempt to localize and / or rule out processing stages that predictability influences, which has bearing on this larger question.

1.3 The Effect of Predictability on the N400

Similar to the first fixation, the N400 ERP component, a negative deflection in EEG recordings starting at around 250ms peaking between 350 ms and 500ms after a word's onset, is also inversely related to the cloze probability of that word. The higher the cloze, the more positive the N400 component. This too has been demonstrated time and again

(Kutas and Hillyard, 1980; Kutas and Hillyard, 1984; Nieuwland and Kuperberg, 2008; Lau, Phillips, and Poeppel, 2008; Lau, Holcomb, and Kuperberg, 2013; a.o.).

A key difference between these methodologies is how the words of a sentence are displayed. In eye tracking while reading, sentences are presented as a line of text, and it is up to the participant to move their eyes to view the next word, which allows the participant to read naturally. It is common for ERP experiments to display words one at a time in a centrally located position on the monitor for a fixed amount of time, which is usually around 500 ms, but can also be longer. This is called Rapid Serial Visual Presentation (RSVP). The reasons for RSVP are that it prevents the participant from having to move their eyes, which can contaminate the EEG recordings with eye movement artifacts, and it ensures that there is only one stimulus associated with the components in question; the N400 is time-locked to 400 ms after the current word and does not reflect processing of any other stimuli as none have intervened.

Already, it is obvious that there is tension between these two measures. First, the timing of the predictability effects is rather different. The N400 only begins to surface at around 250ms, whereas predictability effects in eye tracking occur much sooner in both first fixation durations, which are often around 200ms, and in skipping probability. In addition to this, Staub and Goddard (2019) demonstrated that the predictability effect on the first fixation is limited to the cases where the participant had valid parafoveal preview. In all standardly designed EEG experiments investigating predictability, there is no preview of the upcoming word whatsoever, as words are presented one at a time, in the middle of the screen. This would suggest that the N400 is showing one of two possible things. It could be showing a delayed effect of predictability from some stage that the first fixation had already captured. Or it could reflect a different process from the one reflected by the first fixation altogether; a stage still influenced by predictability nevertheless. Pinning down what process(es) the N400 reflects has been far from straightforward and is central to long-standing debates in the field; narrowing the possibilities down is the second goal of this study.

In addition to showing predictability effects, the N400 also has been demonstrated to be sensitive to orthographic neighborhood size (Laszlo and Federmeier, 2012), as well as, for certain long Inter-Stimulus Intervals (ISIs), misspellings (Ito, Corley, Pickering, Martin, and Nieuwland, 2016). These manipulations go no deeper than word form processing, but they suggest that, at certain long ISIs and with RSVP, the N400 may be able to reflect processing lower than lexico-semantic processing.

Holcomb (1993) conducted a study investigating what the N400 tracks by concurrently collecting ERP and lexical decision data for a simple single word priming study crossed with a second factor of stimulus quality. In contrast to White and Staub (2011), who achieved this by altering the contrast of the text against the background, this manipulation was instead achieved by randomly removing pixels that made up the text. Holcomb found that stimulus quality affected lexical decision times and moreover interacted with priming. Primed words were judged faster than unrelated words, and words in harder to read, faint text showed an even larger priming effect compared to their easier-to-read counterparts. Holcomb also found that while the N400 amplitude was sensitive to the priming manipulation, it was insensitive to the stimulus quality changes, and the two factors did not interact. Holcomb did however find that degraded text produced delays in the N400 peak's latency. These findings suggest that the N400 amplitude, at least in single word priming designs, may not be sensitive to the earliest visual or orthographic processing of the word, and that difficulties at these stages of processing are rather tracked by other measures, such as lexical decision. From the latency shift however, the underlying process(es) that the N400 does index might have to await the conclusion of some earlier stage of processing which is affected by this form of stimulus quality. It is worth highlighting that this was a single word design, and was neither intended to study the predictability effect on the N400 amplitude nor to mimic a natural reading task; it has the same lack of parafoveal preview as typical EEG experiments.

In the current study we are seeking to extend upon this line of research by searching for the effects of lexical predictability (as operationalized by cloze probability), stimulus quality and their interaction in eye tracking while reading measures and FRPs obtained during natural reading. To do so means facing the challenges of collecting EEG data timelocked to fixations rather than preset stimulus onset. Luckily, methodological breakthroughs outlined in the next section have made this possible.

1.4 The Coregistration of EEG and Eye Tracking

In recent years, the aforementioned discrepancy between presentation methods traditionally used in eye tracking and EEG experiments has been alleviated, using "co-registration:" the simultaneous collection of eye tracking while reading and EEG data (Henderson, Luke, Schmidt, and Richards, 2013). Here, participants are allowed to read sentences naturally, thus making it comparable to simple eye tracking while reading. The EEG waveform is time-locked to the time at which the participant began their fixation of a given word, producing a Fixation Related Potential (FRP).

It has been demonstrated that predictability effects can be observed using co-registration, both in the form of shorter first fixation times and a reduced N400 amplitude in the FRPs (Dimigen, Sommer, Hohlfeld, Jacobs, and Kliegl, 2011; and Kretzschmar, Schlesewsky, and Staub, 2015). The N400 FRP effects of predictability in these 2 studies have been broadly distributed both temporally and spatially. Kretzschmar et al. report the N400 component being temporally smeared: starting at as early as 150ms and lasting at least until 650ms post-fixation. Visual inspection of Dimigen et al.'s plots reveal a similar pattern of temporal smearing, particularly in the later portion of the FRP around 500ms. What's more, both studies found the predictability effect on the N400 located in centroparietal electrode sites centered around electrode Pz, in contrast with many RSVP ERP studies which report a more lateralized N400. Dimigen et al. reports no interaction between region of interest (laterality) and predictability in the N400 window, and Kretzschmar et al. find such an interaction present only in one 50ms slice of the N400 window.

This methodology allows us to maintain comparability between EEG and eye tracking findings as the presentation method is identical. Finally, this also allows us to build upon previous research and investigate what process(es) the N400 is indexing by observing what manipulations of the stimuli influence the N400 amplitude and latency.

CHAPTER 2

METHODS AND MATERIALS

2.1 Participants

34 undergraduate students at the University of Massachusetts, Amherst participated in exchange for monetary compensation. All participants were right-handed, native speakers of American English and had no language impairments. Ultimately, only the data from 25 participants was used for the full analysis. 3 participants' data could not be used because of experimenter error. Discussion of exclusion criteria are discussed in later sections.

2.2 Stimuli

The sentence materials were the same as those in Staub and Goddard (2019). There were 180 experimental sentences containing a target word. There were 90 target words; each target word was embedded in a context that either made it predictable or made it unpredictable. The word immediately preceding the target word was the same in the predictable and unpredictable sentence contexts. Each participant saw all 180 sentences. Each experimental sentence was presented on a single line. We crossed the predictability of the target word and the stimulus quality of the sentence resulting in a 2x2 design. The conditions are exemplified in Table 1. Each participant was exposed to 45 trials in each condition. Participants saw each target word twice, once in a predictable context and once in an unpredictable context. Participants also saw each target only once in a clear text condition and once in a faint text condition. The balancing of these as follows: if a participant saw the sentence in the upper left corner of Table 1, they would later see the sentence in the lower right corner, but no other sentences in Table 1. This corresponded to the list that a subject was assigned, of which there were 2, and which were counterbalanced between subjects. In the predictable contexts target words had a mean cloze probability of 0.93 (sd = 0.03, max = 0.99, min = 0.87). In the unpredictable contexts, the target words had a mean cloze of 0.004 (sd = 0.01, max = 0.05, min = 0). The target words were an average of 4.2 characters long (sd = 0.74, max = 6, min = 3). All target words were of moderate to high frequency with a mean Zipf frequency of 4.74 (sd = 0.61, max = 6.29, min = 3.36).

	Clear Text	Faint Text
Predictable Target	On Valentine's Day the woman received a single red **rose** from her secret admirer.	On Valentine's Day the woman received a single red **rose** from her secret admirer.
Unpredictable Target	The traffic cop finally admitted that the red **rose** that fell out of the car wasn't meant for him.	The traffic cop finally admitted that the red **rose** that fell out of the car wasn't meant for him.

Table 1: Example Stimuli. The target word "rose" (which is shared between the two) is bolded here but was not in the experiment. The target word is predictable in the contexts on the top row; not along the bottom row. The colors of the text and background in the table use the RGB values used in the experiment itself; however the achieved contrast will differ between monitors and printer settings.

The stimulus quality manipulation was achieved similar to White and Staub (2012), by maintaining a constant grey background, but altering the contrast of the text in the foreground. The RGB values of the background in the experiment were (RGB = 245 245 245). The clear text was generated with the text colored black (RGB = 0 0 0); the faint text was generated with the text just slightly darker than the background (RGB = 235 235 235).

2.3 Equipment

Eye movement data was collected using a desk-mounted Eyelink 1000. Participants sat and rested on a chin rest to reduce movement. EEG data were collected using a 128-channel electrode net (EGI, Eugene OR) at a sampling rate of 500Hz.

2.4 Procedure

Participants provided informed consent and then were fitted with an EEG cap upon arrival. All electrode impedances were kept below 50 kΩ for the experiment. The experiment was split into 2 blocks of 90 trials each, one where all the clear text items were presented and the other where all the faint text items were presented. Trial order was randomized within blocks. The block design was chosen over a completely randomized design to avoid modulating pupil sizes between successive trials. The ordering of the blocks was counterbalanced between participants. Between the blocks participants were given a break. Participants read the sentences silently to themselves as their eyes were tracked and answered YES / NO comprehension questions following 27% of the trials (12 trials of the 45 total for each condition). Responses were indicated with a left or right trigger press on a gaming controller. The correspondence between YES and NO responses and the trigger (left or right) was counterbalanced between participants. The experiment including calibration and clean-up took approximately 2 hours.

CHAPTER 3

RESULTS

3.1 Eye Movement Measures

All eye tracking data from the 25 subjects with complete data (EEG and eye tracking) were preprocessed using Robodoc and Eyedry (http://blogs.umass.edu/eyelab/software/). These data were then analyzed in the R environment (R Core Team, 2020) using the tidyverse (Wickham et al. 2019) package. Analyses focused primarily on the target word, with supplemental analyses performed for the word immediately preceding the target. The measures of interest were the first fixation duration (the duration of very first fixation which fell upon the target region of the text), the first pass time (the sum of all of the durations of fixations upon the target region before saccading to the left or the right of the region) the go past time (the sum of all of the durations of fixations upon the target region and previously fixated regions, if there were any regressive saccades from the target region, before the target region is exited to the right), and Skipping Probability (the proportion of trials for which the reader did not fixate the target word on first pass reading, but rather made a saccade past it).

3.1.1 First Fixation Duration on the Target Word

The first fixation duration is the eye tracking measure of most interest in this study. With the EZ reader model (Reichle, Rayner and Pollastek, 2003) as a linking hypothesis between reading times and processes, the first fixation is the measure from which the most definitive conclusions can be drawn. An interaction between predictability and stimulus quality in the first fixation duration would suggest these two factors influence the same stage of processing and that stage is computed as a part of computation that yields the decision to saccade. The lack of such an interaction indicates that we may not reject the hypothesis that these effects influence that overall computation independently.

The mean first fixation durations across subjects for each condition are plotted in Figure 1. Fixation times were centered but otherwise left untransformed for these analyses (see Liceralde and Gordon, 2018 for discussion on this topic). Trials for which participants did not fixate the target word were excluded from the analysis. First fixation durations were analyzed using a Linear Mixed Effects Model implemented using the lme4 package (Bates et al. 2015). The fixed effects were the predictability of the target word (predictable vs unpredictable) and the stimulus quality (contrast of the text against the background) (clear vs faint). The contrasts for these effects were coded using sum coding. The random effects structure of the model included random intercepts for both subjects and items, and random slopes for predictability and stimulus quality for subjects, but not their interaction. The maximal model was fit first, but after failing to converge, random slopes were dropped in order of the variance they accounted for, with the slope accounting for the least variance getting dropped first, following Barr, Levy, Scheepers, and Tily (2013). This was iterated until the model converged; that resulting model is what we report here. This procedure was utilized for the remaining eye tracking measures as well. A summary of this model is provided in Table 2. P-values were calculated using the lmerTest package (Kuznetsova, Brockhoff and Christensen, 2017). To correct for multiple comparisons introduced by each successive eye tracking measure, a Bonferroni correction for the significance level was applied following von der Malsberg and Angele (2017) to give a new significance threshold of $p = 0.0125$. As can be seen, there were statistically significant effects of both predictability and of stimulus quality on reading times ($p < 0.001$ in both cases) and these were in the expected direction, with predictable words receiving shorter fixations compared to unpredictable words, and clear text receiving shorter fixations compared to faint text. However, these two effects did not interact ($p = 0.911$). That is, the predictability effect on the first fixation was of the same magnitude in the clear condition as it was in the faint condition.

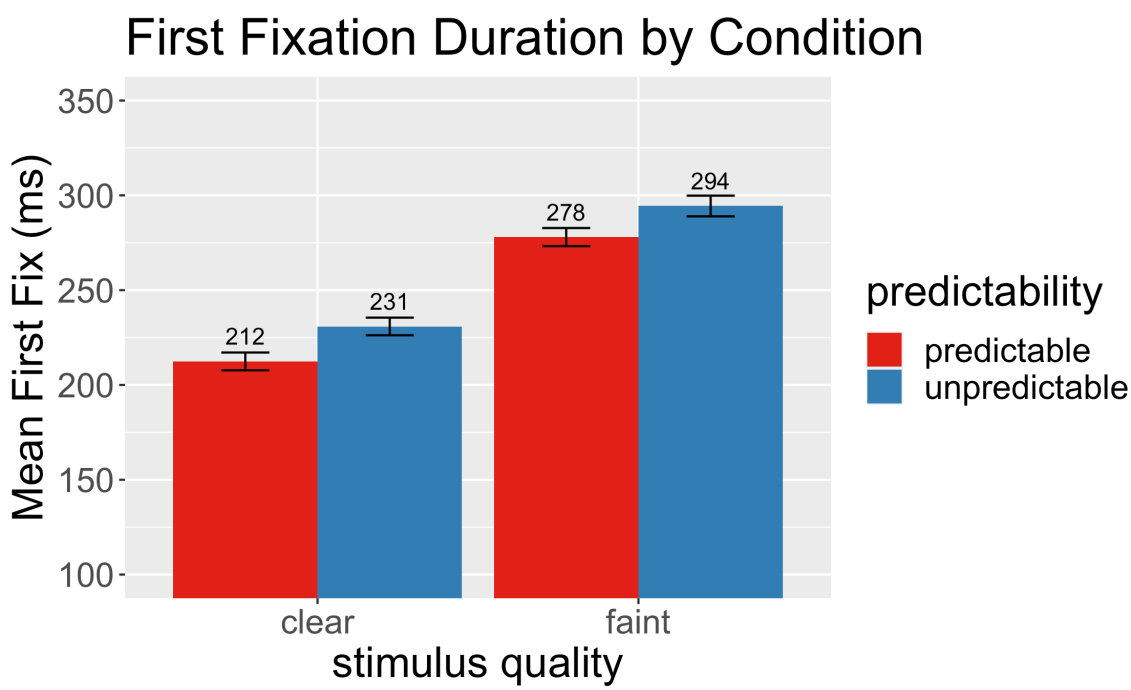

Figure 1: Mean First Fixation Duration of the Target Word by Condition. Error Bars represent 1 by subjects Standard Error.

Measure	Effect	Estimate	t-value	p-value
First Fixation	Predictability (P)	16.8033	5.123	**< 0.001** †
	Stimulus Quality (SQ)	64.7895	9.424	**< 0.001** †
	Interaction (P x SQ)	-0.7128	-0.112	0.911
First Pass	Predictability (P)	26.068	6.034	**< 0.001** †
	Stimulus Quality (SQ)	78.818	10.535	**< 0.001** †
	Interaction (P x SQ)	7.181	0.832	0.405
Go Past	Predictability (P)	63.83	5.214	**< 0.001** †
	Stimulus Quality (SQ)	96.14	7.839	**< 0.001** †
	Interaction (P x SQ)	24.84	1.015	0.31
Skipping Probability*	Predictability (P)	-0.08690	-0.717	0.4734
	Stimulus Quality (SQ)	-0.57306	-7.505	**< 0.001** †
	Interaction (P x SQ)	0.30281	1.987	**0.0469**

Table 2: Target Word Eye Tracking Statistical Tests. Mixed-Effects Model Parameter Estimates fitted to the data from different eye tracking measures for the target word. Bolded p-values are statistically significant at $p < 0.05$; p-values with a † have a $p < 0.0125$ (Bonferroni correction of 0.05 / 4 for the 4 eye tracking measures computed). * Skipping Probability returns a z- rather than a t-value.

3.1.2 First Pass Time on the Target Word

The means for the first pass time across subjects for each condition are plotted in Figure 2. Note that the first pass is an identical measure to the first fixation duration for trials in which participants fixated the target only once before exiting the target region. These times were also analyzed using a linear mixed effects model with sum contrast coding for the factors of predictability and stimulus quality, and random intercepts for subjects and items as well as random slopes for stimulus quality for subjects. Model estimates are provided in Table 2. Again, large and statistically significant effects were found for the main effects of predictability and stimulus quality ($p < 0.001$ in both cases), while the interaction was insignificant ($p = 0.405$).

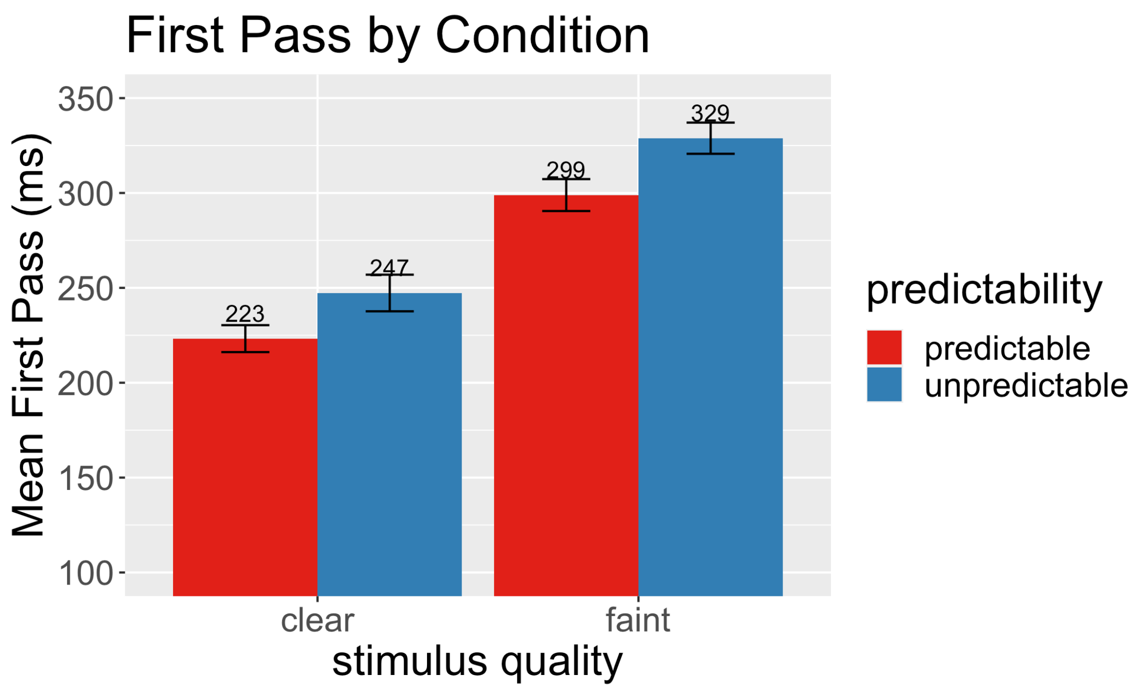

Figure 2: Mean First Pass Duration of the target word by Condition. Error Bars represent 1 by subjects Standard Error.

3.1.3 Go Past Time on the Target Word

The means for the go past times across subjects for each condition are plotted in Figure 3. The go past time is identical to the first pass measure for trials in which the participants did not make a regressive eye movement upon leaving the target region. These times were analyzed using a linear mixed effects model with sum contrast coding for the factors of predictability and stimulus quality, and random intercepts for subjects and items but no random slopes. This model's estimates are provided in Table 2. As with the other earlier measures, significant main effects of predictability and stimulus quality were found ($p < 0.001$ in both cases), while there was no significant interaction ($p = 0.31$).

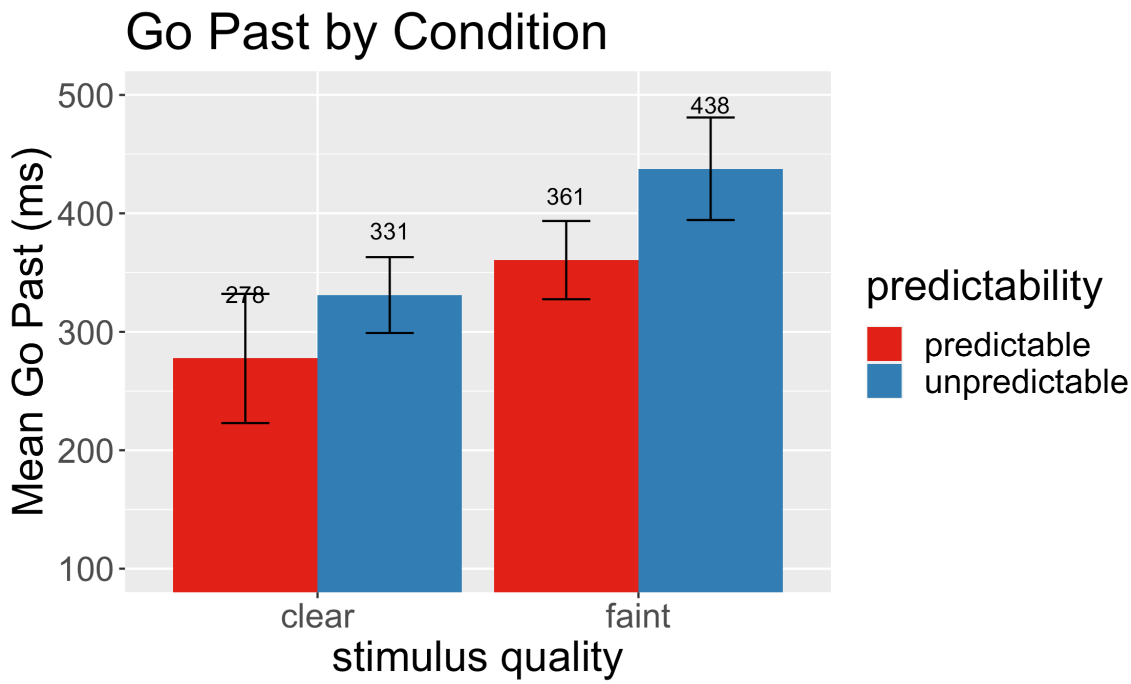

Figure 3: Mean Go Past Duration of the target word by Condition. Error Bars represent 1 by subjects Standard Error.

3.1.4 Skipping Probability of the Target Word

Finally, the probability of skipping the target word was analyzed. The mean skipping probability across subjects for each condition are plotted in Figure 4. These data were analyzed using a logistic mixed-effects model with sum contrast coding for the fixed effects of predictability and stimulus quality. There were random intercepts for subjects and items as well as random slopes for predictability for items. The model's parameter estimates are provided in Table 2. Stimulus quality had significant effect on the skip rate ($p < 0.001$), with clear text words being skipped more frequently compared to faint text words. In contrast to previous experiments, and most notably Staub (2020), no statistically significant main effect of predictability was found for the skip rate ($p = 0.4734$). The interaction of predictability and stimulus quality had a p-value of 0.0469. However, because of the correction for multiple comparisons, the interaction failed to reach statistical significance ($p < 0.0125$) despite patterning in the direction of there being a predictability effect on skipping in the clear text conditions but no effect whatsoever in the faint text conditions.

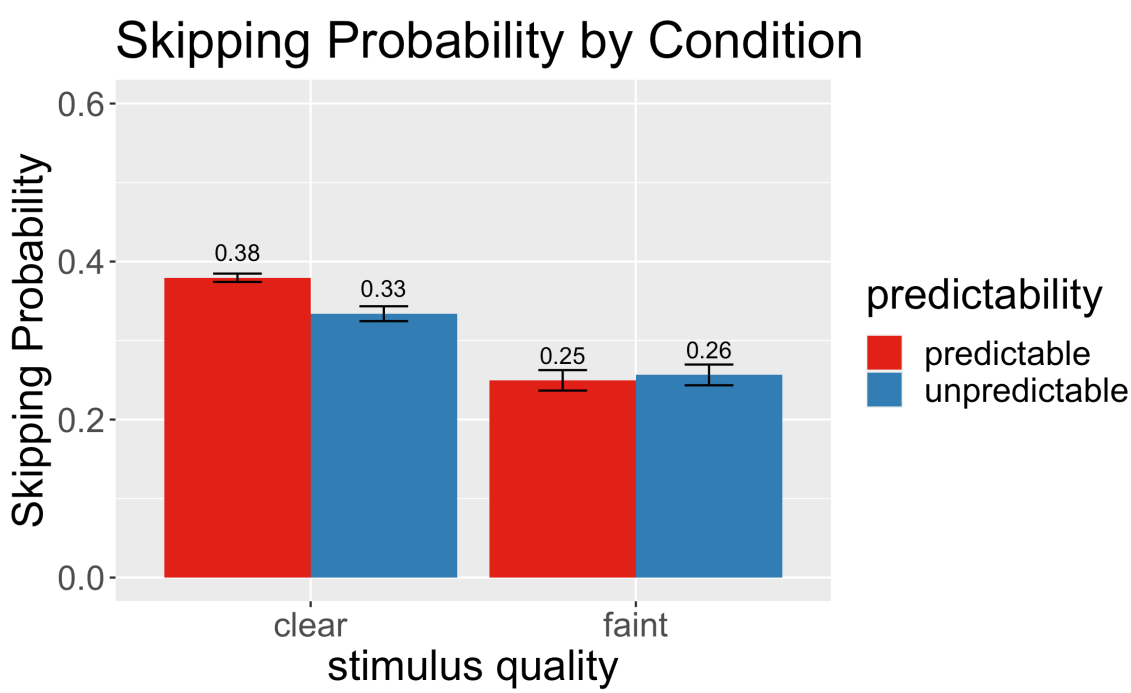

Figure 4: Mean Skipping Probability of the target word by Condition. Error Bars represent 1 by subjects Standard Error.

3.1.5 Eye Movement Measures of the Word Immediately Preceding the Target

The first fixation, first pass, go past and skipping probabilities were also calculated for the word that immediately preceded the target word in the sentence. This pre-target word was the same in the predictable and the unpredictable contexts; the only differences between conditions were the preceding material and therefore the predictability of the upcoming, parafoveal word. Figures 5, 6, 7 and 8 summarize these data, and Table 3 summarizes the statistical analyses.

A linear mixed effects model was fit to the first fixation times. The random effects structure was determined using the same method as the target word analyses which resulted in a model with random slopes for predictability for items, random slopes for stimulus quality for subjects, and random intercepts for items and subjects. There was significant effect of stimulus quality ($p < 0.001$). The effect of predictability was greater than 0.0125 (the same Bonferroni correction for multiple eye tracking measures) with $p = 0.037$. The interaction was insignificant with $p = 0.076$.

Another linear mixed effects model was fit to the first pass times of the pre-target word. The model had random intercepts for items and subjects, but no random slopes. There were significant effects of stimulus quality ($p < 0.001$) and predictability ($p = 0.012$). The interaction was insignificant ($p = 0.31$)

A similar was fit to the go past times of the pre-target word. The model had random slopes for contrast for subjects, and random intercepts for items and subjects. There was a significant effect of stimulus quality ($p = 0.004$). The effect of predictability was insignificant ($p = 0.126$) as was the interaction ($p = 0.570$).

A logistic mixed effect model was fit to the skipping probability data for the pre-target word. The model had random slopes for predictability for items and stimulus quality for subjects as well as random slopes for items and subjects. No main effects or interactions reached significance.

Measure	Effect	Estimate	t-value	p-value
First Fixation	Predictability (P)	9.0863	2.119	**0.0372**
	Stimulus Quality (SQ)	57.9374	9.357	**< 0.001 †**
	Interaction (P x SQ)	-12.3714	-1.774	0.0762
First Pass	Predictability (P)	12.686	2.516	**0.0119 †**
	Stimulus Quality (SQ)	70.622	14.006	**< 0.001 †**
	Interaction (P x SQ)	-6.485	-0.642	0.5211
Go Past	Predictability (P)	22.682	1.532	0.12565
	Stimulus Quality (SQ)	74.780	3.140	**0.0043 †**
	Interaction (P x SQ)	16.847	0.569	0.56975
Skipping Probability*	Predictability (P)	-0.02116	-0.228	0.8196
	Stimulus Quality (SQ)	-0.15503	-1.171	0.2418
	Interaction (P x SQ)	-0.22671	-1.562	0.1182

Table 3: Pre-target Word Eye Tracking Statistical Tests. Mixed-Effects Model Parameter Estimates fitted to the data from different eye tracking measures for the pre-target word. Bolded p-values are statistically significant at $p < 0.05$; p-values with a † have a $p < 0.0125$ (Bonferroni correction of 0.05 / 4 for the 4 eye tracking measures computed). * Skipping Probability returns a z- rather than a t-value.

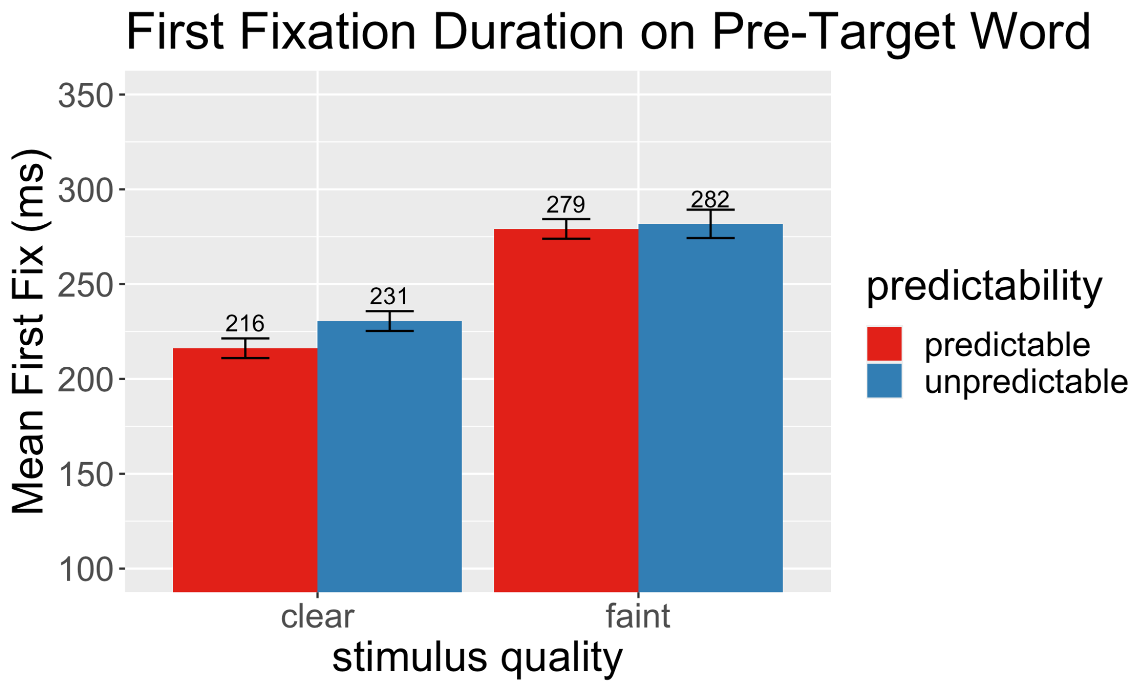

Figure 5: Mean First Fixation Duration of the Pre-Target Word by Condition. Error Bars represent 1 by subjects Standard Error.

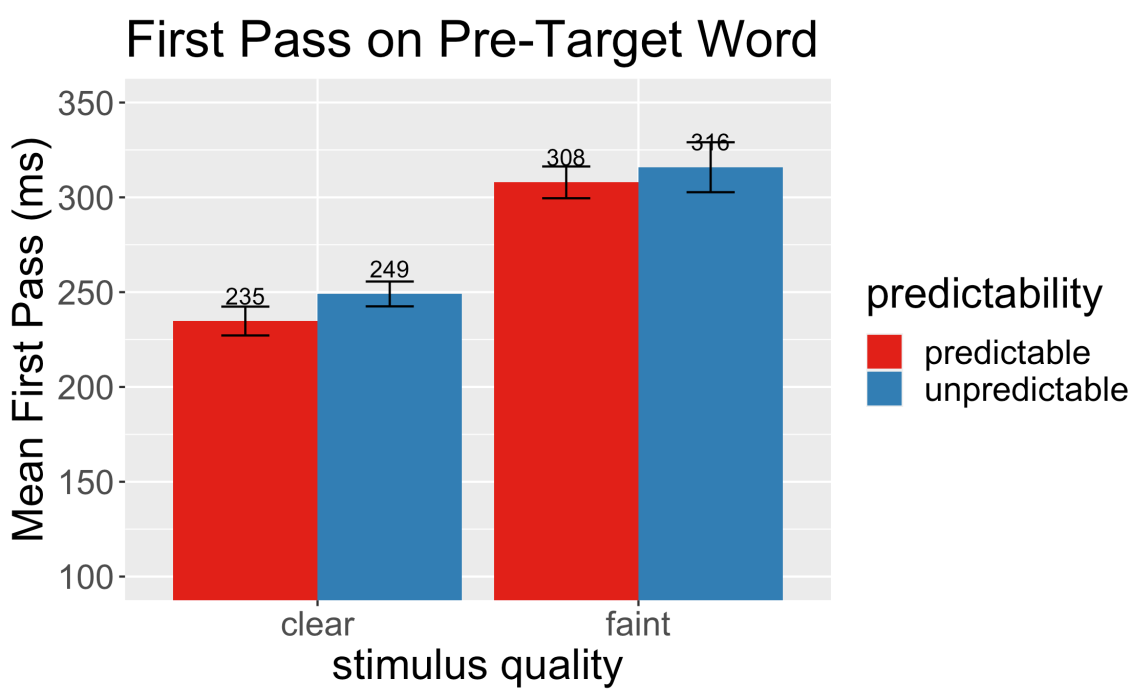

Figure 6: Mean First Pass Duration of the Pre-Target Word by Condition. Error Bars represent 1 by subjects Standard Error.

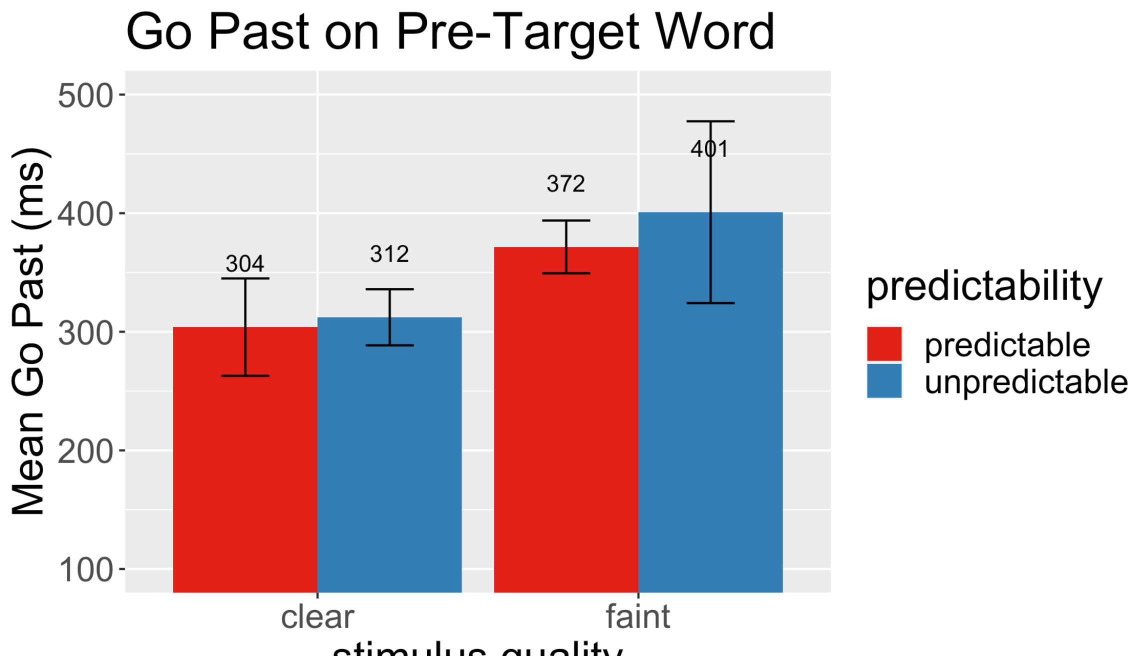

Figure 7: Mean Go Past Duration of the Pre-Target Word by Condition. Error Bars represent 1 by subjects Standard Error.

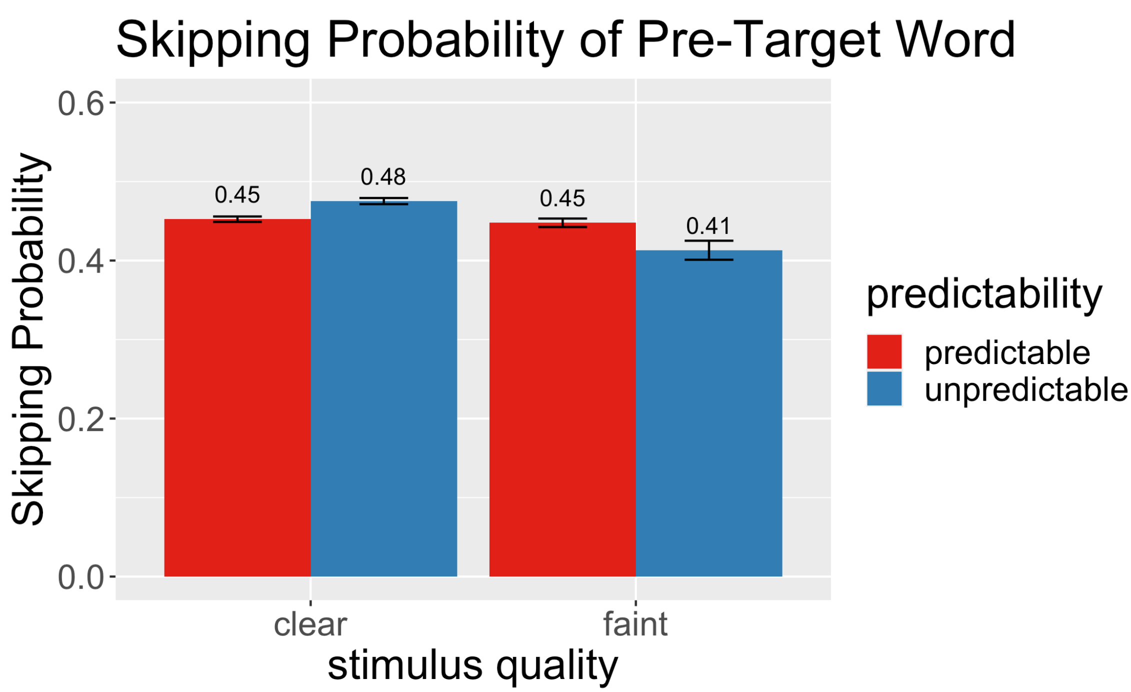

Figure 8: Mean Skipping Probability of the Pre-Target Word by Condition. Error Bars represent 1 by subjects Standard Error.

3.2 Fixation Related Potentials

Preprocessing of the EEG data was performed in MATLAB (Mathworks, Inc.) using EEGLAB (Delorme and Makeig, 2004). From the eye tracking data we were able to extract the time that the participant fixated the target word, in ms from the start of the trial. We used this information to a create an event list to be imported alongside the raw EEG data to add triggers corresponding to those initial fixations post-hoc. By virtue of this method, trials where the participant skipped the target word were automatically excluded. A .3 to 30 Hz bandpass filter was applied to the data. The PREP Pipeline (Bigdely-Shamlo et al. 2015) was used to identify and interpolate bad channels. Manual artifact rejection followed to remove the artifacts not caught by the automated procedures. The data were then re-referenced to the mastoid electrodes, and finally grouped by conditions and subjects to get subject-specific average fixation-related-potentials (FRPs) for each condition.

3 participants were removed at this point because of excessive bridging between neighboring electrodes. 2 participants were removed for displaying extreme voltages. Finally, 1 participant was removed due to excessive data loss (fewer than 10 trials in at least one condition). Thus, after preprocessing, 25 participants remained in the set to be fully analyzed. The average number of trials per condition per subject which went into creating the FRPs after preprocessing was 29.14 (sd = 6.37, max = 41, min = 11).

To answer the question of what effects there are of stimulus quality and predictability on the amplitude of the N400, a standard timewindow of 250ms-500ms was chosen wherein the N400 amplitudes for each condition were compared. To answer the question of what effects there are of stimulus quality and predictability on the latency of the N400, a moving window using 50ms window durations was utilized, where the FRP amplitudes were again compared.

As per extant prior research, we expect that there will be an effect of predictability on the N400's amplitude. The key question here is whether there will be an effect of stimulus quality on the N400's amplitude and if these effects interact. They may interact not just in terms of amplitude but in N400 latency.

Given Holcomb's findings mentioned earlier, we might suspect that the predictability effect (and in fact, the entire N400) could be shifted backwards in time. This would result in an early point in time which marks the beginning of the predictability effect for the clear text (the onset of the N400), but at which point the N400 has not yet begun for the faint text due to a delay in processing. This would surface statistically as an interaction between stimulus quality and predictability early on in the FRPs. There would also be an interaction in the opposite direction at a later point in time, when the faint text predictability effect is ongoing, but the clear-text N400 is past its offset. If there is no N400 latency shift, contra Holcomb, and no interaction in the N400 amplitude itself, then we should observe no interaction between predictability and stimulus quality at any point.

3.2.1 Classic N400 Window Analysis

These data were analyzed further in the R environment. Following Luck and Gaspelin (2017), we sought to minimize our type 1 errors by defining a region of interest (ROI) informed by previous FRP studies, and collapsing across the electrodes within that ROI to avoid implicitly testing multiple comparisons. We identified 19 centroparietal electrodes for which we expect to see an FRP N400 effect which were identified using Kretzschmar et al. and Dimigen et al.'s previous studies. These electrodes include: C2 through Cz to C4, P3 through Pz to P4, and all remaining 13 electrodes in the resulting rectangular array. For each electrode of each participant, a condition-specific average baseline voltage from -200ms to 0ms was computed and removed to correct for baseline differences between conditions. The grand average FRP (collapsing across participants and electrodes in our predefined region of interest (ROI)) is presented as Figure 9.

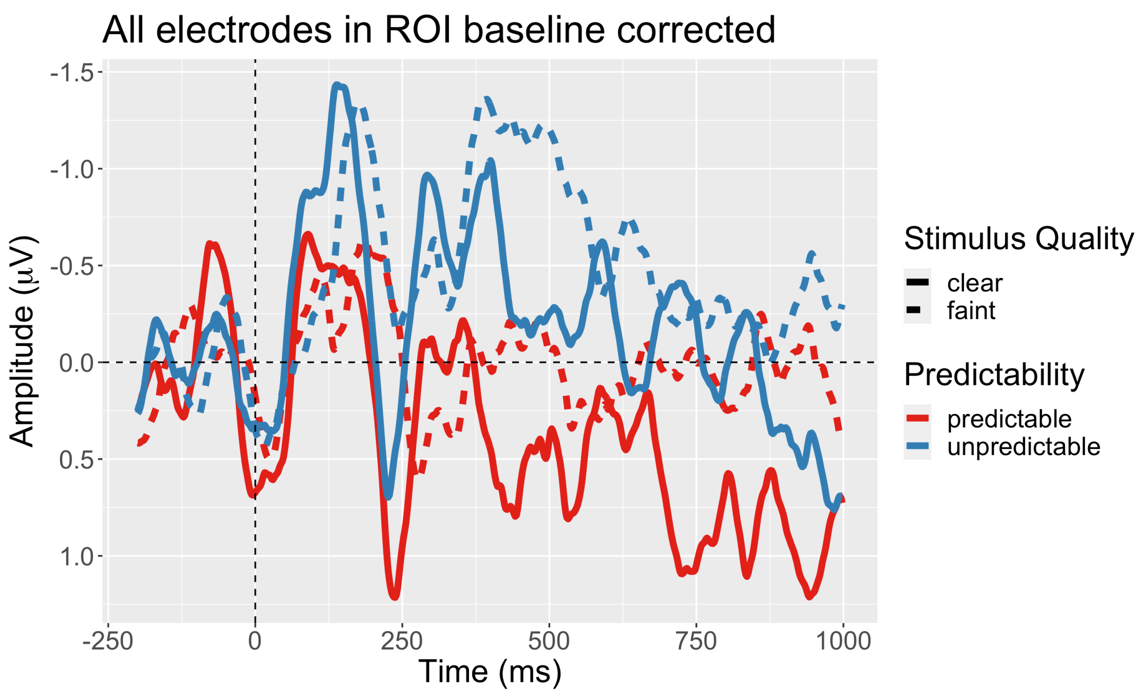

Figure 9: Grand Average Fixation Related Potentials. Negative voltages are plotted up. Time 0 is the beginning of the first fixation upon the target word.

Mean voltages for each condition for each participant were calculated for the 250ms to 500ms window following fixation of the target word. An ANOVA was run over these subject averages; the results are provided in Table 4. The statistical test showed that there was a significant main effect of predictability on the N400's amplitude ($F(1,24) = 13.68$, $p = 0.001$), however, there was neither a significant effect of stimulus quality ($F(1,24) = 1.393$, $p = 0.25$) nor a significant interaction ($F(1,24) = 0.044$, $p = 0.836$).

Effect	F-value	p-value
Predictability (P)	13.68	**0.001**
Stimulus Quality (SQ)	1.393	0.25
Interaction (P x SQ)	0.044	0.836

Table 4: FRP ANOVA Analysis. Output of the ANOVA over subject means voltages in the 250ms to 500ms timewindow for each condition. Bolded p-values are statistically significant at $p < 0.05$.

3.2.2 Difference Waves

To visualize the individual effects of predictability and stimulus quality on the FRPs, difference waves were calculated and are provided as Figures 6 and 7. For the effect of predictability, we split the data by stimulus quality and then calculated, for each electrode in our ROI at each timepoint for each participant, the difference between the baseline corrected amplitude for the predictable conditions and the unpredictable conditions. These results are shown in two difference waves: one for clear text predictable minus clear text unpredictable targets, and one for the faint text predictable minus faint text unpredictable targets. This is shown in Figure 10. For the effect of stimulus quality, the same general procedure was followed with the exception that the data were first partitioned into predictable and unpredictable conditions. We then ended up with two difference waves to represent the effect of stimulus quality: one for the predictable clear text minus predictable faint text targets, and one for the unpredictable clear text minus unpredictable faint text targets. This is shown in Figure 11.

These figures demonstrate that the effect of predictability is strikingly similar between the clear and faint text conditions, with a peak around 400ms and a smeared effect both earlier and later, mirroring the predictability effects that Dimigen et al. and Kretzschmar et al. observed. At just about all points, it is the unpredictable target that produces a more negative FRP. These figures also show that there appears to be some effects of stimulus quality, that is again strikingly similar regardless of the predictability of the target word. The faint text yields more negative FRPs at around 250ms and close to 500ms, though less convincing differences at the intermediate timepoints.

To investigate these effects and their onsets more thoroughly, Holcomb (1993) analyzed the latency of the peak amplitude. This analysis has however fallen out of fashion; as Luck (2005) puts it: "… differences in peak latency do not necessarily correspond to changes in component timing." Since ERPs as well as FRPs are the superposition of many underlying latent components, neither the observed amplitude peaks nor the observed latencies are independent from one another. For instance, a more positive P2 for one condition over another will often result in observing either a delayed onset of the next component or a more positive continuing trend. Accordingly, a moving window analysis was performed and is described in the next section.

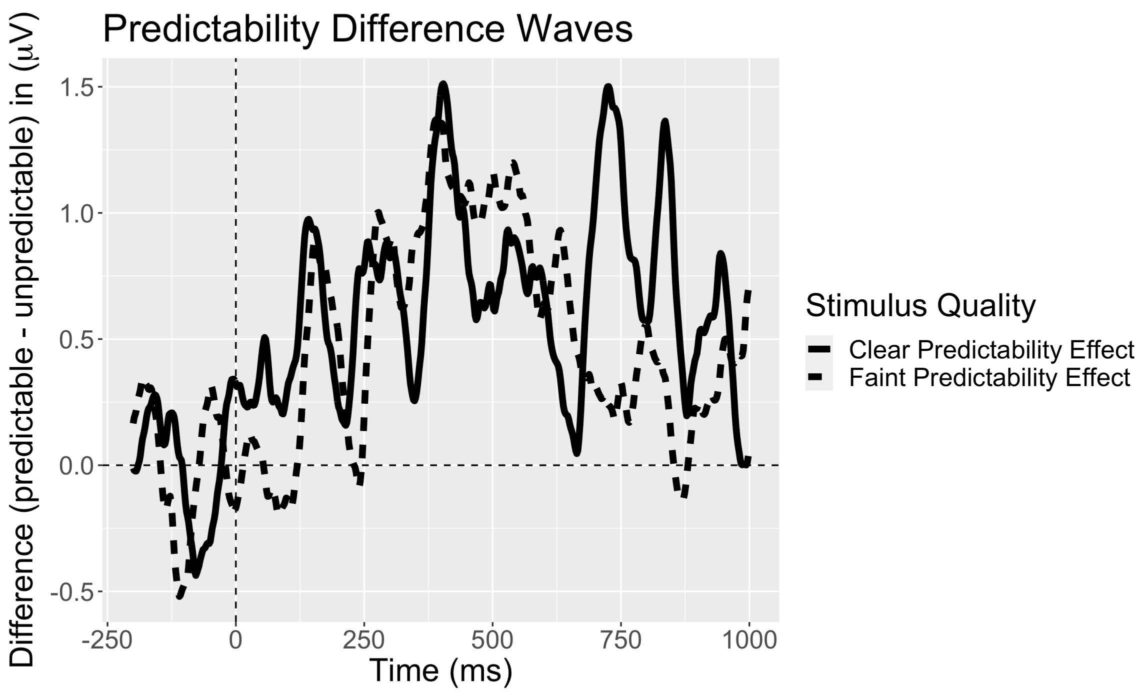

Figure 10: Predictability Difference Wave FRP generated by subtracting the amplitude of the unpredictable conditions from that of the predictable conditions.

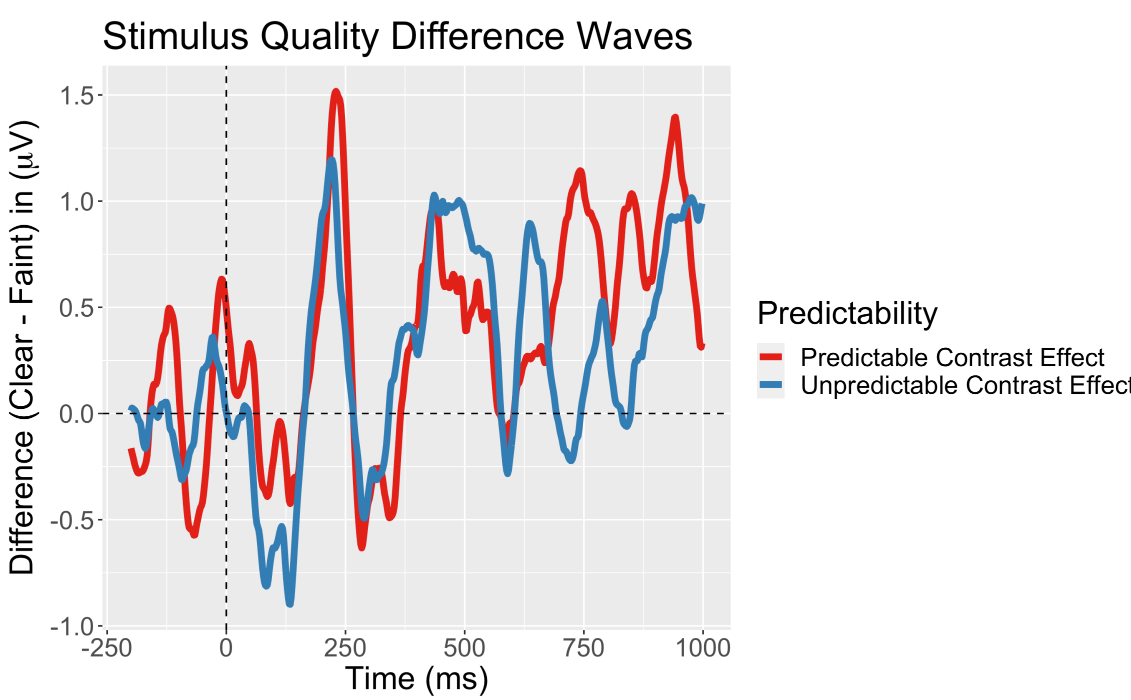

Figure 11: Stimulus Quality Difference Wave FRP generated by subtracting the amplitude of the faint text conditions from that of the clear text conditions.

3.2.3 Moving Window Analysis

We had an additional question that the classic analysis of the large N400 window could not speak to. We wanted to see if we were failing to detect an interaction in the 250ms to 500ms window because the size of the window is large enough to include both the onset and offset of the N400. Additionally, the difference waves suggested that there were earlier effects of predictability and suggested there may be effects of stimulus quality, especially around the P2 component. We adopted Kretzschmar et al.'s moving window / time-bin analysis to investigate these questions. We partitioned the data into 50ms bins (0ms to 50ms, 50ms to 100ms, and so on) for each subject's average FRP. The mean voltage for each condition for each subject within each bin was calculated and then for each bin, an ANOVA was run over those subject means. The results are provided in Table 5. After correcting for multiple comparisons, the data show no significant effect of stimulus quality in the N400 window, and no significant interaction between predictability and stimulus quality at any point. We did however find there to be a significant effect of stimulus quality on the FRP in the 200ms to 250ms window. This is discussed in the Discussion section.

Time Bin	Predictability (P)	Stimulus Quality (SQ)	Interaction (P x SQ)
0ms – 50ms	0.9898120	0.21960453	0.391966150
50ms – 100ms	0.2118576	3.41795815	0.952408720
100ms – 150ms	**4.3858629**	3.26700226	1.136777086
150ms – 200ms	**9.4923695**	0.68350669	0.044158948
200ms – 250ms	1.8254283	**11.49486534 †**	0.414489637
250ms – 300ms	**12.4680004 †**	0.29721816	0.000108882
300ms – 350ms	**9.9408453**	0.92574199	0.077661432
350ms – 400ms	**12.2312592 †**	0.86162950	0.081569608
400ms – 450ms	**13.8123449 †**	**4.57911214**	0.009349814
450ms – 500ms	**5.3661267**	**5.78540830**	0.318766028
500ms – 550ms	**5.7868241**	3.75821542	0.178025984
550ms – 600ms	**5.3780810**	0.07216173	0.009451251

Table 5: FRP Bin Analysis. F-values Output by the ANOVA over subject mean voltages in consecutive 50ms bins for each condition. Bolded F-values are statistically significant at $p < 0.05$; F-values with a † have a $p < 0.004$ (Bonferroni correction of 0.05 / 12 for the 12 time bins shown here).

CHAPTER 4

DISCUSSION

4.1 Discussion of Eye Movement Results

In the first fixation, first pass and go past measures, there were large main effects of both predictability and stimulus quality. In the skipping probability measure we found only a main effect of stimulus quality. These main effects replicate Staub (2020), a.o. and suggest that the computations performed in order to inform the decision to move one's eyes are directly impacted by these two manipulations. The lack of an interaction between these two manipulations is in agreement with Staub (2020). The additive relationship between predictability and stimulus quality suggests that it cannot be ruled out that the two factors have independent effects on the decision to move one's eyes (Sternberg, 1969). There is of course a potential problem of statistical power in detecting an interaction. However, Staub (2020) also found a null effect and were powered to find an interaction eff of 10ms (a modestly sized interaction). I will continue on assuming that the true state of the world is that we have found converging evidence against there being any sizable interaction between stimulus quality and predictability. Discussion of what this implies about predictability is laid out in the General Discussion section.

The analyses of the pre-target word revealed significant effects of stimulus quality and a small but significant effect of predictability on the first pass times. This is surprising and at odds with many other studies; "parafoveal-on-foveal effects" are strikingly rare and the effects of predictability are consistently isolated to the currently fixated word and its predictability. However, these pre-target words were very short which in part certainly helps explain their high skipping probability. It stands to reason that if this is not simply a type 1 error, it is due to "mislocated fixations" or the reader undershooting the intended landing site of their eye on the target word (similar to Drieghe, Rayner and Pollatsek, 2008). That is, on some proportion of trials, when the reader did fixate the pre-target word, they did so intending to skip it, and could thus allocate attentional resources to the parafovea, creating the small effect. Additionally, given the fact that the pre-target words were short and often function words, processing of them was likely quite far along if not complete by the time the reader fixated them, again allowing for extra allocation of attentional resources to the parafovea by reducing the foveal load.

Our eye tracking findings on the target word demonstrate a different pattern than that observed in Holcomb 1993's lexical decision data. In eye tracking we find an additive relationship between our two manipulations while Holcomb's lexical decision data showed an interaction. These measures are not thought to reflect the exact same underlying process, but for there to be an interaction in the lexical decision data suggests that the priming manipulation and stimulus quality manipulation operate on some shared stage. It is worth noting that in Holcomb's experiments, the source of expectation for upcoming words is strictly through associative priming which is in contrast to the current study which provides subjects with sentence contexts. Priming and predictions generated from sentence contexts are not reducible to the same thing (Otten and Van Berkum, 2008; Nieuwland and Van Berkum, 2006). Similarly, the stimulus quality manipulation was achieved by removing pixels at random, rather than altering the contrast between the text and the background. It's possible that Holcomb found and interaction because of the details of this stimulus quality manipulation. With a stimulus quality manipulation like ours, once the reader has observed the stimulus for long enough, there is no uncertainty about the identity of the lines and curves forming letters that make up the stimulus; a faint presentation of “sugar” and clear one are ultimately unambiguously “sugar.” If pixels are removed from the characters themselves, there may remain uncertainty about the underlying “true” lines and curves that make up the intended letters. For instance, “o” can appear similar to “e” or “c”, “r” can appear similar to “n”, “g” to “q”, and so on. Thus, Holcomb’s manipulation may not isolate visual feature processing as ours does.

4.2 Discussion of FRP Results

There were several key findings from the FRP data. First, in accordance with Dimigen et al. and Kretzschmar et al. we found a significant effect of predictability on the amplitude of the N400 FRP, further demonstrating the methodological abilities of coregistration experiments. Additionally, we failed to find any significant effect of stimulus quality on the amplitude and on the latency of the N400 FRP. Finally, the two manipulations did not produce any interactions in the N400 component. These findings have considerable consequences for understanding what the N400 reflects and the architecture of the language comprehension system; both are discussed in the General Discussion.

It is again worth addressing the discrepancy between the current findings and Holcomb’s. In the current study, we found no significant shift in the N400’s latency with poor stimulus quality, while Holcomb (1993) found that such a manipulation shifted the N400 back in time. To start, one might wonder if we are underpowered. Our 25 subjects is more than double the 12 subjects run in the Holcomb experiments, and our 45 items per condition is greater than Holcomb’s 40. However, given the small number of observations in Holcomb’s experiments, it is difficult to evaluate the true effect size of the latency shift.

Holcomb's estimate of the effect size is likely an overestimate due to his small sample (see Vasishth, Mertzen, Jäger and Gelman, 2018 for discussion); a study with more observations would likely find a smaller but more realistic estimate of the true effect. If the true effect size is small, we may still be underpowered to detect it.

As per Luck (2005)'s criticism, it may be that the differences between our claims and Holcomb's lie in the conclusions drawn from different analyses. Holcomb's peak latency analysis has since been discouraged to argue for latency differences, and thus casts some doubt on the effect. Inspection of Holcomb's difference wave figures suggests that the onset of N400 priming effect is the same for clear and degraded text, with the effect lasting longer for degraded words. However, this could be due to exactly the criticism that Luck points out: the elongation N400 may be due to neighboring components such as the P600.

It is widely held that within college-aged populations (Kutas and Iragui, 1998), the N400's latency is remarkably stable (Kutas and Federmeier, 2001). Holcomb is one of only a handful of studies that report any shift in the N400's latency. However, even if Holcomb's reported latency effect is real, while we find no such shift, experimental differences can again demonstrate that this is not necessarily contradictory. Holcomb's experiments were single word designs where time 0 in the ERPs corresponds to the subject's absolute first encounter with the word. In our study, time 0 of the FRP corresponds to the subjects first fixation on the word, but not necessarily the beginning of their processing of it because they are granted parafoveal preview. That is, some processing of the word may have already taken place in the parafovea for clear or faint words. What's more, this preview effect may further smear our effects given that the preview benefit varies between subjects

(Veldre and Andrews, 2015), effectively moving around the true time 0 at which each subject first started processing the word. This may explain some of the temporal smearing of the N400 that is reported in coregistration experiments. Additionally, in the current study, by 400ms after fixating the target word, the subject has almost always moved their eyes and has thus begun processing some new material. These are methodological concerns for all coregistration experiments.

There was a statistically significant effect of stimulus quality observed in the FRPs at around 200ms where the P2 component was observed. This was not a region of the FRPs we intended to investigate. Previous research has claimed that the P2 is an index of the interaction of orthographic and phonological representations of a target (Kramer and Donchin, 1987), and of orthographic processing alone (Kong et al. 2012). It may be that stimulus quality influences a stage of processing that is indexed by the P2, but surprisingly not by the N1, which is generally taken to index attention and visual structural analysis (Korinth, Sommer and Breznitz, 2012). However, with relatively scant investigation of the P2 it is hard to conclude what this may mean, and a more targeted experiment would have to be designed to investigate this.

4.3 General Discussion

This study had 2 direct questions: what processes of lexical access are influenced by a word’s predictability and what processes of lexical access are indexed by the N400? There is of course the larger goal of characterizing the entire series of computations involved in lexical processing looming over these. Here I will review the key findings from the current study and draw conclusions about the direct questions asked as well as about the larger picture.

The key findings of the current experiment are: 1) predictability had a main effect on first fixation times (as well as other early eye tracking measures), 2) stimulus quality had a main effect on first fixation times and other early eye tracking measures, 3) these two effects had an additive rather than an interactive relationship, 4) predictability had a main effect on the amplitude of the N400, 5) stimulus quality had no impact on the amplitude of the N400, and 6) stimulus quality had no impact on the latency of the N400. These findings will be used individually to address the direct questions of the current study, but will all be used together to provide a unified account of lexical processing and the measures used to investigate it.

First, the eye tracking results lend further support to there being stages of processing that are influenced by stimulus quality and predictability, as both manipulations resulted in main effects. It however cannot be ruled out that the stages these manipulations influence are distinct. There is no single stage influenced by both stimulus quality and predictability simultaneously, as there was no interaction between the two manipulations. This additive pattern is in contrast to Staub and Goddard who crossed predictability and preview validity. While no hard conclusion can be drawn comparing the current study to theirs, I speculate from this discrepancy that stimulus quality is not akin to invalid preview, and preview validity and stimulus quality target non-overlapping processes. Under the assumption that stimulus quality influences only early stages of lexical processing (most likely restricted to visual feature processing), this would mean that preview validity influences at least some later stages of processing (as has been argued for in Veldre and Andrews, 2018), and that that later stage is one place where the predictability of a word is also used to facilitate processing. A fully crossed 2x2x2 design would be needed to investigate this directly.

The additive effect we observed suggests that under natural reading conditions, lexical items are predicted but they are effectively not predicted down to the level of detail of their visual features. Predicting “sugar” as an upcoming word may not entail the corresponding prediction of the curves and lines that make up the letters “s”, “u”, “g”, “a”, and “r”, but rather only more abstract expectations which are described at levels higher than visual features. In the framework of the Dual Route Model, the level(s) of representation that predictability influence(s) can then be as low as graphemes, but may also be higher, such as at the word form or semantic processing stages. Relevant to this, is the fact that frequency and predictability have been demonstrated to also have additive effects (see Staub, 2015 for a review). Again, using the terminology of the DRC, frequency must operate at the graphemes, the word form or the semantic stage, with predictability having no concurrent influence in that stage but instead in one or both of the others. Speaking more generally, and model-agnostically, frequency operates at some level of representation higher than visual features, and predictability operates at another, distinct level, also higher than visual features.

There is another possibility which is that there is indeed some preactivation of these visual features, but in concert with simultaneously preactivated concepts (many of which will be expressed using words that have complementary composite graphemes) (Frisson et al. 2017), the set of predicted shapes corresponding to one particular word is no more preactivated than any other. However, skepticism for the claim that sentence contexts lead to predictions all the way down to visual word forms and perceptual units has grown in recent years (Nieuwland, 2019). Nieuwland suggests that it is more likely that predictability only measurably influences some higher-level representation(s). To foreshadow slightly, predictability must be present in at least the semantic stage as predictability effects are observable across modalities (Holcomb and Neville, 1990; Federmeier and Laszlo, 2009) suggesting that it is some shared content between written and spoken words that is made predictable by the context, e.g. a meaning-based representation. What's more, predictions in sentence contexts arise by way of predictions about sentence and discourse level meanings (Nieuwland and Van Berkum, 2006; Kuperberg and Jaeger, 2016). Many hypotheses about the source of the N400 claim that the N400, which shows predictability effects, indexes semantic processing. Localizing the stage(s) that predictability influence(s) is an ongoing process and future research will be needed to whittle down the possibilities further, but the current study suggests that a word's predictability plays no significant role in early visual processing.

Moving on to the FRP results, the current experiment demonstrated that the amplitude of the N400 is sensitive to a word's predictability but is not modulated by stimulus quality. It also demonstrated that the N400 is not significantly pushed back in time when the stimulus quality is poor. The finding that there was no effect of stimulus quality on the amplitude of the N400 suggests that the N400 is a component indexing processing that takes place after visual feature processing. This finding is in line with many accounts of the N400 (Kutas and Federmeier, 2000; Lau et al. 2008; Hagoort et al. 2009; a.o.) which propose the N400 indexes either lexico-semantic retrieval, but not prior processing, or a later integrative process, but not prior processing. These current findings cannot offer more or less support to either the lexico-semantic or integrative account; they simply demonstrate that the process or set of processes indexed by the N400 does not contain the computation that is influenced by stimulus quality manipulations: early visual processing.

Our findings additionally suggest that the processing difficulty at these earlier stages is not reflected in subsequent stages; lexical processing difficulty does not percolate upwards. This was not discernable from the eye tracking data as each measure showed an effect of stimulus quality. Since eye tracking measures build upon each other (the first fix is a submeasure of the first pass), it is hard to detect effects that arise early, but disappear in later processing. The N400 amplitude however showed no significant effect of stimulus quality. This means that the process(es) indexed by the N400 itself is / are not made more difficult by poor evidence at earlier stages. This could arise through two ways: either there is simply a convergence of evidence for one particular input / candidate by the time the process indexed by the N400 is underway, or there is a winner-take-all mechanism in place at the conclusion of a processing stage prior to the N400. Both of these will be explained in more detail below.

In keeping with the DRC's stages and cascaded architecture each stage can be thought of as a kind of series of evidence accumulation processes. Starting with the visual feature processing stage, each sample of the visual scene provides evidence for certain lines, curves and shapes that ultimately make up graphemes and then word forms. Evidence for any representation at this lowest level is percolated upwards as evidence for certain graphemes, and so on. Following this pattern, there will be broad activation for semantic content early on in lexical processing, but the entropy will decrease as time goes on and there will ultimately be a convergence. The fact that the N400 shows no sign of difficulty due to stimulus quality manipulations suggests that by the time the N400 has started, at around 250ms, this convergence has already occurred (for some level of representation that is relevant for the N400). Moreover, eye movement control models often take eye tracking measures such as the first fixation to index some convergence (or near convergence) of evidence, and so this too would have to have occurred. Thus, the only necessary modifications to the DRC would be to temporally constrain the model to guarantee this early convergence.

The other alternative involves a winner-take-all activation-based mechanism. EZ Reader for instance postulates that L1 is complete when some representation of word identity reaches a high enough level of activation to warrant confidence that lexical access is imminent. Evidence accumulation models of similar decision-making processes like the Linear Ballistic Accumulator (LBA) (Brown and Heathcote, 2008) can be useful for thinking of the process leading up to the N400. In the LBA, candidates receive activation on a linear scale proportional to their supported from the current input. Once some candidate passes a threshold activation level, its identity is passed forward. If such a process is what feeds into the N400, it would explain the lack of an effect of stimulus quality on the amplitude of the N400. This however would predict on the face of it that there would be a delay in the process indexed by the N400, as difficulty in evidence accumulation is operationalized as the temporal delay of the winning candidate passing the threshold.

However, stimulus quality had no significant effect on the latency of the N400. This suggests that the process indexed by the N400 does not depend on the conclusion of earlier stages to be initiated. It would appear at first that the same process indexed by the N400 can be done in a shorter time with faint text compared to clear text to make up for the delays experienced during earlier subprocesses of lexical identification. This would assume that the N400 and the decision to move one's eyes are temporally dependent, and they need not be. The current pattern rather suggests that the decision to move one's eyes is temporally independent from the feeding forward of lexical information to the process indexed by the N400. This is schematized in Figure 12.

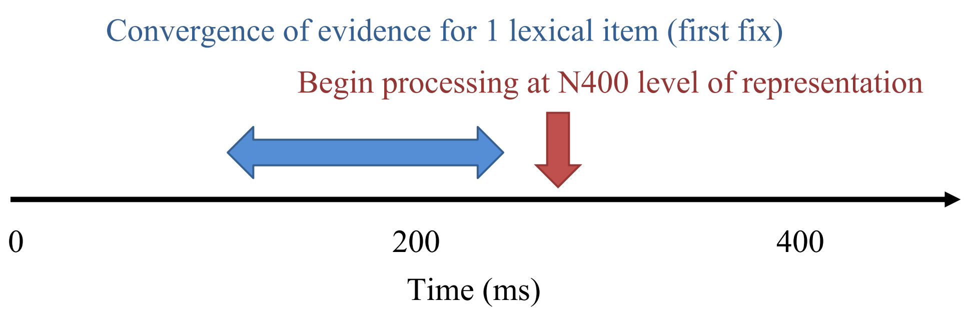

Figure 12: Lexical processing timeline demonstrating the stable point in time at which the N400 begins and the variable time beforehand that results in the first fixation effects.

The N400 has been notoriously time-locked, in that the onset of the component occurs at around 250ms regardless of the peak amplitude (Federmeier and Laszlo, 2009), save for Holcomb (1993)'s findings, and Kutas (1987). The consistency of the onset of the N400 has led some to speculate that it indexes a process that proceeds at a set time after stimulus onset (Federmeier and Laszlo, 2009). Federmeier and Laszlo refer to this process as a "binding" process that takes all currently online linguistic information in and draws upon it to create a semantic representation of the recently encountered stimulus.

Though not often thought of together, this claim about the N400's fixed timing is somewhat echoed in the literature on neural oscillations. Meyer (2018) reviews recent EEG findings that focus on increased power within certain frequency bands of neural activity rather than on time-locked ERPs. Li et al. (2017) found that words made predictable by their context had a reduced N400 amplitude. They also looked at the sentence as a whole and found that sentence constraint (constraining towards a predictable word versus unconstraining throughout) had effects on the theta wave (4Hz-7Hz) power observed with constraining sentences eliciting decreased power. The presentation rate was 700ms per word, yet at a fixed rate between 4 to 7 times a second, a network sensitive to contextual predictability was activated. Given the lower end of this frequency (4Hz, or every 250ms), this may be seen as complimentary to the view that there is a fixed onset for some abstract level of processing.

The current findings suggest that the computation underlying the N400 proceeds at a fixed point in time, when semantic processing must be initiated. This further distinguishes the N400 from eye tracking measures such as the first fixation. While the first fixation is thought of in EZ Reader as reflecting the attainment of some threshold of activation, the network underlying the N400 requires no threshold to be met before processing initiates. This allows there to be cases in which the process underlying the N400 is initiated without there having been a convergence towards one lexical item (Kutas and Federmeier, 2011). Additionally, since the reader has often decided to move their eyes before the N400 is underway, there must sometimes be a convergence of evidence for some lexical item well before the N400's processing begins.

Again, the current study cannot differentiate between the lexico-semantic retrieval account and the integrative account of the N400, which is often seen as the central debate in the N400 literature. The current study does however deepen our understanding of the full architecture underlying lexical access by demonstrating that some stage proceeds at a fixed point in time regardless of the difficulty encountered at earlier stages, and that this process uses lexical identity as input, but not a purely visual representation of the stimulus.

Our lab has run another coregistration experiment that directly followed up on Staub and Goddard's eye tracking experiments investigating the effects of predictability and preview validity. This can add some useful insights into thinking about the N400 FRP more generally. In this experiment, both an effect of predictability and of preview validity was observed on the N400's amplitude. Again, this points out a difference between preview validity and stimulus quality manipulations.

In the invalid preview conditions, the reader has recently encountered both the target word and the invalid preview. Schotter, Leinenger and von der Malsberg (2017) found that readers will sometimes report seeing only the preview and not the target word that replaces it in these paradigms, even in cases where they fixate the target directly. Thus, there is still some representation of the preview that is active even when the target is fixated. From these findings and our own, I will speculate that information is fed forward to the N400 stage of processing even if there still remains uncertainty about the right word to access: the true target or the preview. This scenario did not arise in the current experiment, but it can be seen as an explanation of the pattern observed in the other coregistration experiment that our lab has run. That is, in our current experiment, there was no uncertainty about what lexico-semantic content to look up at 250ms after encountering the stimulus regardless of the stimulus quality; faint "sugar" leads you to look up the semantic content of "sugar" just as well as clearly presented "sugar". However, in cases where two words have been encountered in rapid succession (as in a preview manipulation) there is still some activation from both the preview and target leading the reader to look up semantic content for more than just one lexical item at the time of the N400. Since the preview was semantically unrelated to the target, the lexico-semantic content being retrieved is very much non-overlapping which explains the effect on the N400's amplitude. Ideally, a separate ERP or FRP study manipulating preview relatedness and predictability would be needed to test this idea directly.

Finally, there were effects of predictability on the amplitude of the N400 and duration of the first fixation. These have been discussed in the previous sections, and as these measures reflect different cognitive states and processes, it is not the case that the two predictability benefits are perfectly redundant; lexical predictability is operating on two or more levels. Staub and Goddard assumed that preview validity affected early processing stages of lexical access, and so the interaction between it and predictability suggested that predictability is also influencing those very early stages of processing as well as some later stage that the N400 is sensitive to. The fact that frequency effects survived with invalid preview was taken to show that frequency effects are occurring later; at some higher level of representation, but not so high as the level that the N400 tracks, as frequency effects are not observed in the N400 in sentence contexts (Van Petten and Kutas, 1990). If the assumption about preview validity were true, stimulus quality should have worked similarly to preview validity. Yet it didn't. The current findings suggest that these two processing stages affected by predictability (one tracked by the first fixation duration and the other by the N400) may not be a low-level stage and a high level stage interrupted by an intermediate stage that isn't affected by predictability, as is the claim in Staub and Goddard (2019).

Removing the assumption that parafoveal preview manipulations operate solely on the earliest stages of word identification allows for the current findings and those of Staub and Goddard's experiments to be explained by an account in which frequency influences some stage(s) of processing which are earlier than those affected by predictability. What's more this removes the unintuitive ordering of lexical processing effects, whereby predictability affects some early processes, lies dormant, and reemerges to affect some late processes. Assuming lexical predictions are generated first at the speaker-meaning / discourse level (Nieuwland and Van Berkum, 2006; Kuperberg and Jaeger, 2016) and trickle down to preactivate lower level representations from lexical potentially down to perceptual, this proposed architecture avoids the necessity of skipping intermediate representations when preactivating lower ones. Finding the exact lowest point in the representational hierarchy that is preactivatable via sentence context is an ongoing endeavor, and whether tasks modulate this boundary will be an important additional consideration.

CHAPTER 5

CONCLUSION

Through the simultaneous collection of eye tracking and EEG data we have been able to gain some key insights into the mechanisms underlying lexical processing and the measures commonly used to investigate them. While there were main effects of predictability and stimulus quality on the eye tracking data (most notably the first fixation duration), the two manipulations did not interact. This demonstrates that there is no compelling evidence that linguistic predictions are fleshed out down to the detail of visual features. This in turn allows for the possibility of a contiguous series of lexical processing stages affected by predictability rather than an interrupted series of stages. The FRP analyses revealed that only predictability modulated the amplitude of the N400; stimulus quality had no effect on the component's amplitude nor its latency. The discrepancy between the eye tracking and the FRP results again highlights the differences in the measures' sources. The N400 patterns observed here suggest that the N400 does indeed reflect the neural activity underlying a process or set of processes that occur after the perceptual processing of visual features. Perhaps more surprisingly, this is a process that in natural reading occurs abruptly at a fixed starting point, rather than being delayed by earlier difficulties. With more data from studies that use these two sources (eye tracking and EEG) such as our own, a more fleshed out and precise model of lexical and sentence processing can be developed.

APPENDIX

Below are the items used in the experiment. Each item has two variants, a constraining variant in which the target word is predictable (pred), and an unconstraining variant in which the target is unpredictable (unpred).

Item Number	Condition	Sentence
1	Pred	For breakfast Jim wanted bacon and eggs with a side of homefries.
1	Unpred	We finally decided that pasta and eggs probably wouldn't be her favorite dinner.
2	Pred	They turned in their project on the date it was due for the first time this semester.
2	Unpred	If she was as ambitious as her sister she would know it was due the day before yesterday.
3	Pred	They sat together without speaking a single word because she was mad at him.
3	Unpred	Looking outside, he didn't expect to see a single word written on his windshield.
4	Pred	The limping horse was obviously in much pain after the race.
4	Unpred	Samantha believes that the program doesn't offer much pain relief to the lower back.
5	Pred	The cheap pen ran quickly out of ink before the student finished the test.
5	Unpred	It annoys Carl when the store runs out of ink during the week.
6	Pred	Susan bought a dress so now she just needed new shoes to go with her outfit.
6	Unpred	The mechanic was awfully surprised when he found the new shoes that his wife bought him.
7	Pred	Carolyn couldn't start her car without the right keys but she tried anyways.
7	Unpred	In order to avoid confusion, we should mark the right keys with her initials.
8	Pred	She went to the bakery for a loaf of bread for the second time today.
8	Unpred	None of my friends know I found a backpack full of bread last week.
9	Pred	She went to the beauty parlor to perm her hair without her mother's permission.
9	Unpred	Her boyfriend would always forget that her hair is sensitive to ocean water.

10	Pred	To hang the picture Ted needed a hammer and nail from the garage.
10	Unpred	I did not expect a rock and nail to do the trick so well.
11	Pred	When the alarm rang the firefighter slid down the pole as quickly as they could.
11	Unpred	The interior decorator wanted to remove the pole in the middle of the room.
12	Pred	When driving you should keep your eyes on the road to avoid accidents.
12	Unpred	Sometimes my brother Rob says that the road in our neighborhood is quite daunting.
13	Pred	When the two met, one of them held out his hand a bit aggressively.
13	Unpred	The boy's mother refused to look in his hand because she didn't want to know the truth.
14	Pred	The children went outside to play in the snow.
14	Unpred	Finding out if they're able to play at the concert is not so important.
15	Pred	He loosened the tie around his neck right when he left work.
15	Unpred	While eating dinner, the candidate splashed his neck with spaghetti sauce.
16	Pred	Dan gathered more wood for the fire so we could roast more s'mores
16	Unpred	We unanimously decided that the fire logo would be the best fit.
17	Pred	Ray fell down and skinned his knees on the pavement.
17	Unpred	Shelby's old boss had his knees replaced at age 51.
18	Pred	Joe did not like his outfit and decided to change into something more comfortable.
18	Unpred	It's really not surprising that he wanted to change the channel to watch football on Thanksgiving.
19	Pred	Her new shoes were the wrong size so she had to return them.
19	Unpred	When he told me that the wrong size fit, I got confused.
20	Pred	The farmer spend the morning milking his cows and mucking the stalls.
20	Unpred	Evan doesn't like talking to him about his cows because he doesn't know anything about agriculture.
21	Pred	Without her sunglasses the sun hurt Erika's eyes more than she'd expected.

21	Unpred	The professor wouldn't look at Erika's eyes while discussing serious matters.
22	Pred	She graduated at the top of her class in high school but struggled in college.
22	Unpred	Helen quite often worried about her class because it was so difficult.
23	Pred	The teacher wrote the problem on the board so that the class could try to solve it.
23	Unpred	He finally found his car keys on the board of the last class he was in.
24	Pred	It was windy enough to fly a kite through the sky.
24	Unpred	The boy's favorite gift was a kite that his aunt got for him.
25	Pred	The maid dusted the books on the shelf once a month.
25	Unpred	The rambunctious boy accidentally broke the shelf in the living room.
26	Pred	Jessie ran the race at a slower pace than the rest of her friends.
26	Unpred	Even if there is a slower pace in this race it is still the same distance.
27	Pred	He turned the page of his favorite book with a big smile on his face.
27	Unpred	All of his closest friends forgot his favorite book was Don Quixote.
28	Pred	She preheated the oven and greased the pan to prepare dinner for the night.
28	Unpred	If Robert had been able to clean the pan we wouldn't need to buy a new one.
29	Pred	The dentist recommends brushing your teeth twice a day to avoid problems.
29	Unpred	My Uncle's favorite thing in the world is a day at the beach.
30	Pred	After hitting the iceberg the ship began to sink at a dramatic rate.
30	Unpred	My son insisted that it's impossible to sink his canoe.
31	Pred	In the shower he washed his skin with soap he purchased from the mall kiosk.
31	Unpred	I can't believe I went the whole day with soap in my pocket.
32	Pred	You would need a raincoat to avoid getting wet during the thunderstorm.
32	Unpred	John didn't believe he would be getting wet so he didn't bring an umbrella.

33	Pred	After every meal it’s good to brush your teeth in order to avoid cavities.
33	Unpred	My roommate said that it's important to check your teeth before an interview.
34	Pred	He wore a heavy jacket because it was cold and he didn't want to get sick again.
34	Unpred	Nobody really knew why Meadow didn't believe it was cold some nights in the summer.
35	Pred	Karen awoke after a bad dream and couldn't fall back to sleep.
35	Unpred	Her students didn't believe that a bad dream could be a sign of a healthy brain.
36	Pred	At dinner he cut his food with a knife and fork.
36	Unpred	Isabella was the first to buy a knife from the strange man at the fair.
37	Pred	Chris needed a belt to hold up his pants or he could just buy new pants.
37	Unpred	Hank decided to tell all his friends about his pants with the funky patterns.
38	Pred	Surfers are scared of getting bitten by a shark after the accident last year.
38	Unpred	I never thought I would see a shark outside of an aquarium.
39	Pred	Dan was asked to be the new coach of the team but he declined.
39	Unpred	It's unlikely that my girlfriend will ask about the team since she doesn't like sports.
40	Pred	He brought his bait to the lake to catch fish every Sunday morning.
40	Unpred	Billy's mother was worried that he would catch fish in the forbidden lake.
41	Pred	On Valentine’s Day the woman received a single red rose from her secret admirer.
41	Unpred	The traffic cop finally admitted that the red rose that fell out of the car wasn't meant for him.
42	Pred	Walking through the dark room I accidentally stubbed my toe on the dresser.
42	Unpred	When it happened I asked Scott about my toe since he is in medical school.
43	Pred	Sarah saw animals from around the world at the zoo in central park.
43	Unpred	I wasn't surprised to see my coworker Tim at the zoo because he always talks about going there.
44	Pred	Amber went to the dealership to purchase a new car after she got a raise.

44	Unpred	Last week when I was arriving on campus a new car stole my parking spot.
45	Pred	The little girl left Santa a plate of cookies and milk before she went to bed.
45	Unpred	The maid was very upset to find dirt and milk stains on the brand new carpet.
46	Pred	Although Keith bowled well he did not have the highest score on the team.
46	Unpred	The stalwart marine was not aware that the highest score for his platoon was in the 90th percentile.
47	Pred	Charles dunked the basketball through the hoop and won the game for his team.
47	Unpred	After tryouts the coaches decided that the hoop was too wide for a high school team.
48	Pred	I realized I had no umbrella as it began to rain all over my new sweater.
48	Unpred	The spectators doubted the magician when he said it had to rain for his trick to work.
49	Pred	He was so sure the racehorse would win he made a bet larger than his wife allowed.
49	Unpred	Little did the officer know, his previous sergeant had a bet that he was going to be transferred.
50	Pred	It was cold in the room so they turned on the heat despite their father's wishes.
50	Unpred	Not even the smartest electrician could figure out where the heat was coming from.
51	Pred	In the night sky it is easier to see all the stars when in an area with less light pollution.
51	Unpred	The professor expressed his ignorance to the class about the stars in relation to Greek mythology.
52	Pred	She put the pot on the stove so the water would boil and she could cook dinner.
52	Unpred	The rest of the group came to the conclusion that it would boil quicker if we added salt.
53	Pred	Because of his driving ticket the man had to pay a fine of more than $300.
53	Unpred	To my dismay, I discovered that there was going to be a fine for parking where I did.
54	Pred	The baby birds were ready to leave the nest once they learned how to fly.
54	Unpred	The last easter egg was hidden in the nest next to the birdhouse my Dad built.
55	Pred	Because there was lightning she could not go to the pool to swim with her friends.
55	Unpred	Last year I made a New Year's resolution that I would try to swim but I never did

56	Pred	To keep the dogs out of the yard he put up a fence around the perimeter.
56	Unpred	I talked with my husband and he and I decided that a fence would be worth the money.
57	Pred	She could tell he was mad by the tone of his voice and his rotten demeanor.
57	Unpred	It's surprising that he can't believe what is happening to his voice since the doctor warned him.
58	Pred	They paid for their meals but forgot to leave a tip for the pleasant waiter.
58	Unpred	Her day got a lot better when she found a tip on her desk.
59	Pred	She didn’t have her watch so she asked for the time from a stranger.
59	Unpred	We wouldn't be in this situation if Julia had the time to meet yesterday.
60	Pred	After the argument Ann went to her room and slammed the door so hard that it broke.
60	Unpred	The chef lacked the necessary skills to be able fix the door to the freezer.
61	Pred	He mailed the letter without a stamp attached to the envelope.
61	Unpred	They reminded us to bring a stamp or two on the trip.
62	Pred	Expecting Jeff's call, she waited for the phone to ring until she fell asleep.
62	Unpred	It shouldn't be so difficult to find someone to ring the pizza shop and ask for a refund.
63	Pred	The package was sent through the mail two weeks ago.
63	Unpred	If by 5PM nobody claims the mail then it will be thrown away.
64	Pred	She lied about losing her report card to hide her bad grades from her parents.
64	Unpred	The sous chef that we met was convinced that bad grades do not correlate with intelligence.
65	Pred	The genie promised the man he would grant one wish if he rubbed the lamp three times.
65	Unpred	The contractor told me that if he had one wish he would go to Italy to see Venice.
66	Pred	He cashed his new paycheck at the bank down the street.
66	Unpred	The problem that he has with the bank is that they are always busy.
67	Pred	The doctor's suitcase was worn and obviously very old because of the cracks in the leather.

67	Unpred	They don't like that mall because it's very old compared to the one just a few towns over.
68	Pred	Spring was Jo’s favorite season of the year so she was sad that it was ending.
68	Unpred	For some reason Devon tells everyone the year of the dragon is his favorite zodiac sign.
69	Pred	His boss refused to give him a raise despite how hard he worked.
69	Unpred	Her neighbors thought she got a raise because of her new car.
70	Pred	The birthday card was funny and made me laugh until my stomach hurt.
70	Unpred	Finding my shoes in the hallway makes me laugh because I know my dog brought them there.
71	Pred	The princess would someday become a queen and rule the nation.
71	Unpred	Her brother firmly believes that a queen is usually a better ruler than a king.
72	Pred	To promote their album the band went on tour last summer.
72	Unpred	Rick told me about what he did on tour with the band.
73	Pred	The learn about their ancestors they drew a family tree on the chalkboard.
73	Unpred	My girlfriend's cousin is too young to understand that our family tree is not a real tree.
74	Pred	Katie put the flowers in an expensive vase that we got for Christmas.
74	Unpred	In March, the engineer bought an expensive vase for his sister.
75	Pred	After inhaling smoke from the fire she needed fresh air so she walked far away.
75	Unpred	The most important factor in their restaurant plan is having fresh air circulate the patio.
76	Pred	Sherry had to read lips because she was deaf and had been since birth.
76	Unpred	The energetic Pitbull in the park was deaf so he could not hear his owner.
77	Pred	When babies are hungry they may often cry until they are fed.
77	Unpred	I found that my dog would often cry when I left for work.
78	Pred	After raking the yard Pat jumped into the pile of leaves and quickly regretted it.
78	Unpred	The increased difficulty on the trail is a result of leaves falling from the surrounding trees.

79	Pred	She married just for money and not for love so I think it will end in divorce.
79	Unpred	They now know that if it wasn't for love the protest would have failed.
80	Pred	I could not remember his name for the life of me.
80	Unpred	The diligent waiter saw his name on the business cards that were left on the table.
81	Pred	To pay for the car, Al simply wrote a check to the dealership.
81	Unpred	The other day, Miriam told me she found a check for $100.
82	Pred	Bradley prefers cats over dogs but his girlfriend still convinced him to get a dog.
82	Unpred	The six birds flew over dogs that were sleeping in the park.
83	Pred	They raised pigs on their farm which had been in the family for generations.
83	Unpred	We decided that it's their farm and we shouldn't tell them how to run it.
84	Pred	John felt sorry, but it was not his fault that she missed her bus.
84	Unpred	None of us knew that it was his fault that we got in trouble.
85	Pred	The lecture should last about one hour if everything goes smoothly.
85	Unpred	I never really understood why one hour goes by so quickly when I'm having fun.
86	Pred	The wealthy child attended a private school far away from his home.
86	Unpred	The movie theater always had private school brochures in the lobby.
87	Pred	The knight readied for battle and drew his sword with a vengeance in his eyes.
87	Unpred	Stacey was astounded that in his room his sword was hanging right above his bed.
88	Pred	John swept the floor with a broom while he was at work.
88	Unpred	After my sister had discovered a broom in the backyard she was convinced witches exist.
89	Pred	The exit was marked by a large sign hanging from the ceiling.
89	Unpred	This is the last time a large sign will convince me to stop for fast food.
90	Pred	The grass was tall because Tim didn't mow the lawn last week.

90	Unpred	Her husband was especially excited about the lawn because he could now host barbecues.

BIBLIOGRAPHY

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.

Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67, 1-48.

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K. M., & Robbins, K. A. (2015). The PREP pipeline: standardized preprocessing for large-scale EEG analysis. Frontiers in neuroinformatics, 9, 16.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. Psychological review, 108, 204.

Delorme, A. & Makeig, S. (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics Journal of Neuroscience Methods 134:9-21

Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., & Kliegl, R. (2011). Coregistration of eye movements and EEG in natural reading: analyses and review. Journal of Experimental Psychology: General, 140, 552.

Frisson, S., Harvey, D. R., & Staub, A. (2017). No prediction error cost in reading: Evidence from eye movements. Journal of Memory and Language, 95, 200-214.

Grainger, J., & Holcomb, P. J. (2009). Watching the word go by: On the time-course of component processes in visual word recognition. Language and linguistics compass, 3, 128-156.

Hagoort, P., Baggio, G., & Willems, R. M. (2009). Semantic unification. In The cognitive neurosciences, 4th ed. (pp. 819-836). MIT press.

Henderson, J. M., Luke, S. G., Schmidt, J., & Richards, J. E. (2013). Co-registration of eye movements and event-related potentials in connected-text paragraph reading. Frontiers in systems neuroscience, 7, 28.

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. Psychophysiology, 30, 47-61.

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. Language and cognitive processes, 5(4), 281-312.

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. Journal of Memory and Language, 86, 157-171.

Kong, L., Zhang, B., Zhang, J. X., & Kang, C. (2012). P200 can be modulated by orthography alone in reading Chinese words. Neuroscience letters, 529(2), 161-165.

Korinth, S. P., Sommer, W., & Breznitz, Z. (2012). Does silent reading speed in normal adult readers depend on early visual processes? Evidence from event-related brain potentials. Brain and language, 120(1), 15-26.

Kramer, A. F., & Donchin, E. (1987). Brain potentials as indices of orthographic and phonological interaction during word matching. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13(1), 76.

Kretzschmar, F., Schlesewsky, M., & Staub, A. (2015). Dissociating word frequency and predictability effects in reading: Evidence from coregistration of eye movements and EEG. Journal of Experimental Psychology: Learning, Memory, and Cognition, 41, 1648.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension?. Language, cognition and neuroscience, 31(1), 32-59.

Kutas, M. (1987). Event-related brain potentials (ERPs) elicited during rapid serial visual presentation of congruous and incongruous sentences. Electroencephalography and Clinical Neurophysiology, 40(Suppl), 406-411.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. Science, 207(4427), 203-205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. Nature, 307(5947), 161-163.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. Trends in cognitive sciences, 4(12), 463-470.

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., & Iragui, V. (1998). The N400 in a semantic categorization task across 6 decades. Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section, 108(5), 456-471.

Kuznetsova A, Brockhoff PB, Christensen RHB (2017). "lmerTest Package: Tests in Linear Mixed Effects Models." Journal of Statistical Software, 82(13), 1-26. doi: 10.18637/jss.v082.i13 (URL: https://doi.org/10.18637/jss.v082.i13).

Laszlo, S., & Federmeier, K. D. (2011). The N400 as a snapshot of interactive processing: Evidence from regression analyses of orthographic neighbor and lexical associate effects. Psychophysiology, 48, 176-186.

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. Journal of cognitive neuroscience, 25(3), 484-502.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. Nature Reviews Neuroscience, 9, 920.

Li, X., Zhang, Y., Xia, J., & Swaab, T. Y. (2017). Internal mechanisms underlying anticipatory language processing: Evidence from event-related-potentials and neural oscillations. Neuropsychologia, 102, 70-81.

Liceralde, V. R. & Gordon, P. (2018). Consequences of power transforms in linear mixed-effects models of chronometric data. https://psyarxiv.com/f73mh/

Luck, S. J. (2005). Ten simple rules for designing ERP experiments. Event-related potentials: A methods handbook, 262083337.

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). Psychophysiology, 54(1), 146-157.

Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: state of the art and emerging mechanisms. European Journal of Neuroscience, 48(7), 2609-2621.

Nieuwland, M. S., & Kuperberg, G. R. (2008). When the truth is not too hard to handle: An event-related potential study on the pragmatics of negation. Psychological Science, 19(12), 1213-1218.

Nieuwland, M. S., & Van Berkum, J. J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. Journal of cognitive neuroscience, 18(7), 1098-1111.

Otten, M., & Van Berkum, J. J. (2008). Discourse-based word anticipation during language processing: Prediction or priming?. Discourse Processes, 45(6), 464-496.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The EZ Reader model of eye-movement control in reading: Comparisons to other models. Behavioral and brain sciences, 26(4), 445.

Reingold, E. M., & Rayner, K. (2006). Examining the word identification stages hypothesized by the EZ Reader model. Psychological Science, 17(9), 742-746.

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.

Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. Language and Linguistics Compass, 9, 311-327.

Staub, A. (2020). Do Effects of Visual Contrast and Font Difficulty on Readers' Eye Movements Interact With Effects of Word Frequency or Predictability?. Journal of Experimental Psychology: Human Perception and Performance.

Staub, A., & Goddard, K. (2019). The role of preview validity in predictability and frequency effects on eye movements in reading. Journal of Experimental Psychology: Learning, Memory, and Cognition, 45, 110.

Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. Journal of Memory and Language, 82, 1-17.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. Acta psychologica, 30(0), 276-315.

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. Journalism Bulletin, 30, 415-433.

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. Memory & cognition, 18(4), 380-393.

Veldre, A., & Andrews, S. (2018). Beyond cloze probability: Parafoveal processing of semantic and syntactic information during reading. Journal of Memory and Language, 100, 1-17.

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

White, S. J., & Staub, A. (2012). The distribution of fixation durations during reading: effects of stimulus quality. Journal of Experimental Psychology: Human Perception and Performance, 38, 603.