Robust effects of predictability across experimental contexts: Evidence from event-related potentials

Wenjia Zhang [a,b,1], Wing-Yee Chow [c,1], Bo Liang [a], Suiping Wang [a,b,*]

[a] Center for the Study of Applied Psychology and School of Psychology, South China Normal University, Guangzhou, China
[b] Guangdong Provincial Key Laboratory of Mental Health and Cognitive Science, South China Normal University, Guangzhou, China
[c] Division of Psychology and Language Sciences, University College London, London, UK

ABSTRACT

Many experiments have shown that comprehenders can generate predictions about upcoming inputs on the fly, but relatively little is known about whether and how comprehenders' sensitivity to predictability may be modulated by the experimental context. The present study used event-related potentials (ERPs) in two experiments to ask whether changing the overall predictive validity of the stimuli will affect comprehenders' brain responses to predictable as well as unpredictable words by manipulating the filler sentences, which made up 50% of the stimuli in each experiment. Contrary to the prediction that predictable words should be processed more easily and elicit a smaller N400 response in a more prediction-encouraging experimental context, we found that participants' N400 response to predictable as well as unpredictable words was smaller when the overall predictive validity of the stimuli was low (that is, when the filler items were incongruous compared to when they were predictable). Further, even though the use of different filler sentences did modulate comprehenders' ERP and behavioural responses, it did not modulate the effect of target word predictability on participants' ERP responses at all. We take the present findings to suggest that healthy young adults' brain responses are inherently sensitive to the predictability of the incoming linguistic stimuli, and that this robust sensitivity can be observed regardless of the make-up of the experimental stimuli.

1. Introduction

Over the past two decades, psycholinguistic research has gathered much evidence that comprehenders can generate predictions about upcoming language on the fly (Kamide, 2008; Kuperberg and Jaeger, 2016; Kutas and Federmeier, 2010). Studies using different experimental techniques have commonly found that comprehenders are immediately sensitive to the predictability of incoming linguistic stimuli during real time comprehension. For example, studies that examined the quantitative relationship between predictability and reading times have demonstrated that the two are inversely correlated over the whole range of predictability (Boston et al., 2008; Demberg and Keller, 2008; Kliegl et al., 2006; Rayner and Well, 1996; Smith and Levy, 2013).

Studies that measured comprehenders' event-related potentials (ERPs) during reading and listening comprehension have linked two key ERP components to predictive processing, namely, the N400, which is a negative-going ERP component that peaks at around 400 ms post-stimulus onset, and the late frontal positivity, which is typically observed between 600 and 900 ms.

On one hand, the amplitude of the N400 has been shown to vary inversely with a word's predictability (e.g., Guntkr et al., 1997; Kutas and Hillyard, 1984; Michael et al., 2006; Wlotko and Federmeier, 2012), such that the N400 response elicited by a word is reduced when that word is more predictable. A word's predictability is commonly operationalised as the proportion of trials in which speakers continue the sentence context with that word in an untimed sentence fragment completion task (also known as cloze probability; Taylor, 1953). Under the view that the size of the N400 to a word indexes the ease of accessing the semantic representations associated with that word in long-term memory (Federmeier and Kutas, 1999; Lau et al., 2013; Lau et al., 2014), the N400's sensitivity to predictability has been taken to reflect facilitated lexical semantic access for more predictable words (Federmeier and Kutas, 1999; Lau et al., 2013, 2014). Meanwhile, under the "semantic integration" view of the N400, such sensitivity to

---

predictability can be taken to show that the meaning of more predictable words can be integrated into the context more easily. ,[23].

On the other hand, the late frontal positivity has been linked to violations of predictions. This ERP effect was first reported by Kutas (1993), who found that unpredictable but congruent endings elicited a larger left frontal post-N400 positivity than predictable endings in highly constraining sentence frames. This positivity has since been documented in a growing number of studies (Delong et al., 2011; Federmeier et al., 2007; Kutas, 1993; Wlotko and Federmeier, 2012) and can be dissociated from another late positive component (LPC or the P600), which has a posterior distribution and is elicited by words that are anomalous (e.g., DeLong et al., 2014; Van Petten and Luka, 2012). The late frontal positivity has been taken to index the processing cost of prediction violations (Federmeier, 2007) and inhibition of predicted words (Kutas, 1993).

In the present study we aim to extend previous findings and ask to what extent comprehenders' brain response to predictable and unpredictable words during sentence comprehension may be modulated by the experimental context. Previous studies have shown that the sensitivity of some ERP components associated with language processing may be modulated by the makeup of the experimental stimuli (e.g., Brown et al., 2000; Hahne and Friederici, 1999; Holcomb, 1988; Lau et al., 2013). For example, Hahne and Friederici (1999) examined the effect of phrase structure violations in German and varied the proportion of sentences containing phrase structure violations (20% vs. 80% violation). They found that, while an early left anterior negativity was elicited and equally pronounced under both proportion conditions, the P600 effect observed in the low proportion condition was completely reversed in the high proportion condition. Such sensitivity to the experimental context has been taken to suggest that processes that underlie the ERP component in question are under strategic control.

In the study of prediction during language comprehension, recent findings from visual world eye-tracking and self-paced reading paradigms have also been taken to propose that comprehenders' sensitivity to predictability may be modulated by the extent to which prediction is "encouraged" by the experimental context (e.g., Brothers et al., 2017; Huettig and Guerra, 2019). However, to our knowledge, no previous ERP studies have examined whether and how comprehenders' brain responses' sensitivity to predictability may be modulated by the makeup of the experimental stimuli during reading comprehension.

In order to examine the extent to which comprehenders' sensitivity to predictability may be under strategic control and may be magnified (or dampened) depending on the experimental context, in the present study we asked whether the two ERP components that have been associated with predictive processing (the N400 and the late potential positivity) are sensitive to the proportion of trials in which a sentence ends with a highly predictable word in a given stimulus block (also known as predictive validity). We examined how the overall predictive validity of the stimuli may impact comprehenders' ERP response to more or less predictable words during sentence comprehension in two experiments. In Experiment 1 we focused on the facilitative effect of successful prediction by examining comprehenders' ERP responses to experimental sentences that ended with a highly predictable word; in Experiment 2 we focused on the inhibitory effect of prediction errors by ending the same sentences with an unpredictable word (see Table 1). In both experiments we manipulated the overall predictive validity of the stimuli across different stimulus blocks by intermixing the experimental sentences

Table 1
Sample experimental and filler sentences in Experiments 1 and 2.

Experiment 1 (Predictable target word)	Experiment 2 (Unpredictable target word)
张律师 赢了 这场 官司。	张律师 赢了 这场 球赛。
Lawyer Zhang won this-CL <u>lawsuit</u>.	Lawyer Zhang won this-CL <u>ballgame</u>
"Lawyer Zhang won this <u>lawsuit</u>."	"Lawyer Zhang won this <u>ballgame</u>

Filler sentences (for both experiments)	
Predictable filler	数学家 成功地 解开了 难题。
	Mathematician successfully solved <u>problem</u>.
	"The mathematician successfully solved the <u>problem</u>."
Incongruous filler	舞台上 明星们 正在 演唱 法国。
	On-stage celebrities are-now singing <u>France</u>.
	"The celebrities are singing <u>France</u> on the stage."

with one of two types of filler sentences (predictable or incongruous sentences; see Table 1) with a 1:1 target-to-filler ratio, resulting in stimulus blocks with a high (predictable filler blocks) or low predictive validity (incongruous filler blocks).

The primary goals of the present study were to examine if and how either of the two ERP components associated with prediction may be modulated when the experimental context contained a higher or lower proportion of sentences in which the final word is highly predictable. We entertained two ways in which changing predictive validity may modulate comprehenders' sensitivity to predictability. Firstly, an experimental context in which all stimuli are highly predictable (as in the predictable filler blocks in Experiment 1) may encourage predictive processing and make predictable words even easier to process. Meanwhile, an experimental context in which all sentences end with an unpredictable word (as in the incongruous filler blocks in Experiment 2) may weaken comprehenders' commitment to their predictions and they may experience less processing disruption upon encountering an unpredictable word. We examined these possibilities in Experiments 1 and 2 respectively.

In addition, we also combined the data from Experiments 1 and 2 to examine whether and how the effect of predictability (that is, the difference between comprehenders' response to predictable and unpredictable target words) may differ between the predictable fillers and incongruous filler blocks. However, as was the case in many previous studies that examined the effect of predictability, the predictable target words in the present stimuli were also more strongly semantically related to words in the sentence context than the unpredictable target words. For example, in the sample stimuli presented in Table 1, the predictable target word "lawsuit" is likely to be primed by the preceding words in the sentence than the unpredictable target word "ballgame". As such, effects of predictability may also be attributed to differences in the extent to which predictable vs. unpredictable target word may have been primed by preceding words in the sentence context. Therefore, we remain agnostic about the mechanisms that underlie comprehenders' sensitivity to a word's predictability and will refer to effects of predictability without committing to specific prediction and/or priming mechanisms.

2. Experiment 1

In this experiment, we examined participants' ERP responses to highly predictable words during sentence comprehension in stimulus blocks with a high or low predictive validity. In the predictable filler blocks, all of the stimuli (i.e., fillers as well as experimental items) ended with a highly predictable word (high predictive validity). Meanwhile, in the incongruous filler blocks, all of the fillers ended with a semantically incongruous words, such that only half of the stimuli ended with a predictable word (low predictive validity).

Based on the view that the N400 response to a word is reduced when that word's semantic representation can be accessed more easily, if an experimental context in which all stimuli are highly predictable

---

[2] The present study is not aimed to address this debate about the functional significance of the N400. For a more detailed discussion of the competing views of the N400 see Lau et al. (2008).

[3] The present study is mainly concerned with the effect of predictability (operationalised as cloze probability). However, it should also be noted that the N400's sensitivity cannot be fully captured by cloze probability alone (e.g., Chow et al., 2016; Federmeiers and Kutas, 1999; Nieuwland et al., in press).

(predictable filler blocks) can encourage predictive processing and make predictable words even easier to process, then the predictable target words in the experimental items should elicit a smaller N400 response in the predictable filler blocks than in the incongruous filler blocks.

2.1. Method

2.1.1. Participants

Eighteen students (7 males; mean age = 21.1) from South China Normal University participated in the present study. All participants were native speakers of Mandarin Chinese, right-handed, had normal or corrected-to-normal vision, and had no reading disabilities or neurological disorders. The present study was approved by the Psychology Research Ethics Committee in South China Normal University. All participants provided written informed consent prior to the experiment and were paid for their participation. Data from two additional participants were excluded due to excessive artefacts (>30%) in their ERP data.

2.1.2. Materials

The materials consisted of 80 experimental sentences and 80 filler sentences (40 predictable fillers + 40 incongruous fillers) with comparable length and syntactic complexity. All sentences were segmented into word-like units for stimulus presentation in the ERP experiment; the average number of segments did not differ between the experimental and filler sentences (experimental sentences: 5.65 ($SD = 0.86$); predictable filler sentences: 5.68 ($SD = 0.83$); incongruous filler sentences: 5.48 ($SD = 0.88$); $F < 1$).

All the experimental and filler sentences were normed for cloze probability and plausibility. In an offline sentence-completion task, we asked 30 native Mandarin Chinese speakers who did not participate in the ERP experiment to provide the most likely continuation for each sentence frame. We defined contextual constraint as the cloze probability of the most likely completion for a given sentence frame. All experimental and filler sentences had a context constraint greater than 0.63 (average = 0.82). We then selected the sentence-final words based on participants' sentence completion data. For the experimental sentences and the predictable filler sentences, the sentence-final word was the most likely completion; for the incongruous filler sentences, the sentence-final word was semantically incongruous and always had zero cloze probability (see Table 1).

Further, we obtained semantic plausibility ratings for all sentences (including those used in Experiment 2). An additional 27 participants who also did not participate in the ERP experiment were asked to rate the plausibility of each sentence on a 5-point scale (ranging from 1 = 'not acceptable at all' to 5 = 'fully acceptable'). The average rating for the experimental sentences, predictable filler sentences, and incongruous filler sentences were 4.82 ($SD = 0.19$), 4.81 ($SD = 0.23$) and 1.37 ($SD = 0.42$) respectively.

The sentences were distributed evenly into 4 stimulus blocks (2 predictable filler blocks and 2 incongruous filler blocks) of 40 sentences each. In a predictable filler block, 20 experimental sentences were intermixed and presented together with 20 predictable filler sentences; in an incongruous filler block, 20 experimental sentences were intermixed and presented together with 20 incongruous filler sentences. As a result, all of the sentences ended with a predictable (high cloze probability) word in the predictable filler blocks, and only half of the sentences ended with a predictable word while the other half ended with an unpredictable and semantically congruous word in the incongruous filler blocks. The same type of stimulus blocks were always presented together, and we counterbalanced the order of the blocks across participants such that half of the participants saw the predictable filler blocks first while the other half saw the incongruous filler blocks first. Further, the assignment of the experimental sentences into blocks was also counterbalanced such that a given experimental sentence was presented in a predictable filler block to half of the participants and in an incongruous filler block to the other half.

2.1.3. Procedure

We used the software package E-Prime (Psychology Software Tools, Pittsburgh, PA) for stimulus presentation and behavioural response collection. Sentences were presented in black fonts against a light grey background. Each trial began with the presentation of a fixation cross for 400 ms, followed by a 400 ms blank screen. Subsequently a sentence was presented one word at a time, with each word appearing on the screen for 400 ms and an inter-stimuli interval (ISI) of 200 ms. The final word appeared with a period to mark the end of the sentence. After a blank screen of 1200 ms, participants were prompted to rate the semantic plausibility of the sentence on a 7-point scale (ranging from 1 = completely unacceptable to 7 = fully acceptable). They were asked to respond as quickly as possible by pressing the corresponding number keys on a keyboard. Participants were asked to avoid eye movements during the presentation of the sentences.

Participants completed 10 practice trials prior to the first predictable filler block and the first incongruous filler block. The make-up of the practice trials matched that of the stimuli block that was to follow immediately, such that the practice sentences prior to a predictable filler block always ended with a predictable word, while half of the practice sentences preceding an incongruous filler block ended with a semantically incongruous word.

In addition, in order to further separate the predictable filler blocks and the incongruous filler blocks and to reduce potential carryover effects, participants were asked to perform a simple flanker task (Kelly et al., 2008) upon completing the first two blocks. In each trial participants were presented with a visual array (e.g., "»»>") and were asked to press one of two buttons to indicate the direction of the middle arrow. Three types of trials were included: (i) congruent, where all arrows pointed in the same direction (<<<<</>>>>), (ii) incongruent, where the middle arrow pointed in a different direction than its neighbouring arrows (<<><</>>>>), and (iii) neutral, where dashes were placed around the middle arrow (–<–/–>–). This task comprised of 60 trials and lasted about 3 min. An average experimental session lasted around 40 min in total.

2.1.4. EEG recording and data analysis

Participants were tested individually in a sound-attenuating, electrically shielded booth. We used the Brain Products system for the electroencephalograph (EEG) and electrooculogram (EOG) recording and EEGLAB for data analysis (Delorme and Makeig, 2004). EEG was recorded with 30 scalp electrodes (10–20 System) and EOG was recorded from electrodes placed below and above the left eye and at the outer canthus of each eye. EEG signals were referenced online to the left mastoid and re-referenced offline to the average of the two mastoid electrodes. The AFz electrode on the cap served as ground. Impedance was kept below 5 kΩ for all scalp electrodes. The EEG and EOG signals were sampled at 1000 Hz online and filtered digitally with a 0.02–30 Hz band pass filter offline. Epochs were computed for the 1000 ms after the onset of the target word relative to a 200 ms pre-stimulus baseline. Epochs with ocular and movement artefacts exceeding ±80 μV were rejected. This affected 14% of experimental trials.

Statistical analyses on average voltage amplitudes were conducted separately for two time windows: 300–500 ms for the N400 and 600–900 ms for the late frontal positivity. Within each time window we conducted a repeated measures ANOVA which included three within-subject factors: filler type (predictable fillers, incongruous fillers), anteriority (anterior, central, posterior), and laterality (left, midline, right). Crossing the factors of anteriority and laterality yielded 9 areas of interest (AOIs) spanning 18 scalp electrodes: left-anterior: F3, FC3; midline-anterior: Fz, FCz; right-anterior: F4, FC4; left-central: C3, CP3; midline-central: Cz, CPz; right-central: C4, CP4; left-posterior: P3, O1; midline-posterior: Pz, Oz; right-posterior: P4, O2 (see Fig. 1). The Greenhouse-Geisser correction was applied in cases where the sphericity assumption was violated (Greenhouse and Geisser, 1959).
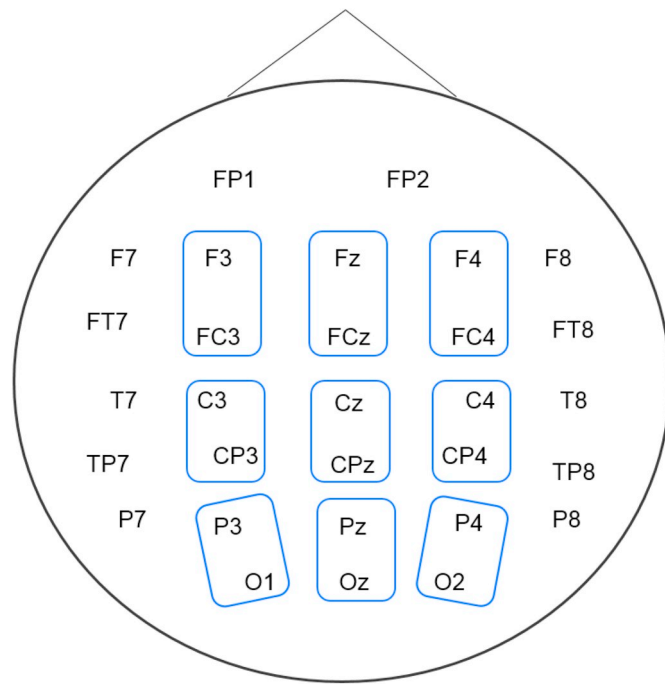
Fig. 1. Thirty-channel montage and the 9 AOIs used for statistical analysis in Experiments 1 and 2.

2.2. Results

2.2.1. Behavioural Results

Participants' average plausibility ratings and response times are presented in Table 2. Paired sample t-tests comparing participants' responses to the experimental sentences in the predictable and incongruous filler conditions revealed no significant difference in either measure ($|ts| < 1$).

2.2.2. ERP Results

Fig. 2 shows the grand average ERPs at the predictable target words and the topographic distribution of the effects of filler type in the 300–500 ms and the 600–900 ms intervals. Results of the statistical analyses are presented in Table 3.

In the 300–500 ms time window, the $2 \times 3 \times 3$ repeated measures ANOVA showed a marginally significant main effect of filler type ($F(1, 17) = 4.32$, $p < 0.1$, $\eta_p^2 = 0.20$) (predictable filler blocks: mean = 1.65, SE = 0.48; incongruous filler blocks: mean = 2.35, SE = 0.42) and interaction between filler type and laterality ($F(2, 34) = 2.72$, $p < 0.1$, $\eta_p^2 = 0.14$). Follow-up comparisons within each level of laterality showed that the N400 response was significantly smaller (i.e., less negative) in the incongruous filler blocks than in the predictable filler blocks in left and midline regions ($ps < 0.05$), but not in the right region ($p > 0.1$).

In the 600–900 ms time window, there was no significant effect of filler type ($F < 1$) (predictable filler blocks: mean = 2.50, SE = 0.48;

Table 2
Average plausibility ratings and response times (and standard errors) in Experiment 1.

	Plausibility rating (out of 7)	Response times (ms)
Predictable experimental sentences		
Predictable filler blocks	6.41 (0.12)	717 (54)
Incongruous filler blocks	6.46 (0.10)	727 (55)
Filler sentences		
Predictable fillers	6.46 (0.12)	749 (59)
Incongruous fillers	1.42 (0.09)	718 (73)

incongruous filler blocks: mean = 2.35, SE = 2.22). Even though visual inspection of the data suggested that participants' ERPs were numerically more positive in the incongruous filler blocks than in the predictable filler blocks in the right-posterior region of the scalp, there was no statistically significant interaction between filler type and either topographic factor ($ps > 0.1$).

2.3. Discussion

In Experiment 1, we used predictable target words to examine whether comprehenders' response to predictable words was modulated by the overall predictive validity of the stimuli. We reasoned that, if the high predictive validity of the stimuli in the predictable filler blocks can make predictable target words even easier to process, then comprehenders should show a reduced N400 response to predictable target words in the predictable filler blocks than in the incongruous filler blocks. This prediction was not confirmed by the present results. In fact, we saw the opposite pattern – participants showed a numerically larger, not smaller, N400 response to predictable target words in the predictable filler blocks than in the incongruous filler blocks. This suggests that predictable target words were not processed any more easily when the experimental context has a higher predictive validity.

Therefore, the results thus far do not provide evidence that comprehenders are any more likely to engage in prediction or make stronger predictions when the predictive validity of the experimental context is high. However, since the manipulation of filler type did not have a statistically significant impact on participants' behavioural or ERP responses in the present experiment, at the moment we cannot rule out the possibility that the manipulation of filler type was simply ineffective. We aimed to address this potential concern and to look at another aspect of predictive processing, namely, violations of predictions, in Experiment 2.

3. Experiment 2

This experiment has two primary goals. First, we aimed to extend the findings in Experiment 1 by asking whether an experimental context in which all sentences end with an unpredictable word might weaken comprehenders' commitment to their predictions and, as such, make the processing of an unpredictable word less costly. We adapted the experimental items from Experiment 1 such that they all ended with an unpredictable (zero cloze probability) but congruous word (e.g., "Lawyer Zhang won this <u>ballgame</u>."). As in Experiment 1, all of the filler items in the predictable filler blocks ended with a highly predictable word, and all of the filler items in the incongruous filler blocks ended with a semantically incongruous words. As a result, 50% and 0% of all the stimuli ended with a predictable word in the predictable filler and incongruous filler blocks respectively.

With the linking hypothesis that the late frontal positivity is larger when an unpredictable but congruous word violates stronger predictions, if comprehenders are less likely to predict (or make weaker predictions) when the overall predictive validity of the stimuli is low, then the unpredictable target words should elicit a smaller late frontal positivity in the incongruous filler blocks than in the predictable filler blocks. In addition, comprehenders' judgments of the plausibility of the experimental sentences may also be affected the type of fillers used in a given stimulus block. For instance, while the unpredictable experimental items may be rated poorly when they are presented alongside filler sentences that are highly predictable (predictable filler blocks), the same sentences may be rated more favourably when all the filler sentences are incongruous (incongruous filler blocks).

Further, we aimed to examine whether and how the effect of predictability may be modulated by the predictive validity of the stimuli by combining the data from both experiments. Based on previous findings, we expect a comparison between comprehenders' ERP response to predictable and unpredictable target words to reveal an N400 effect
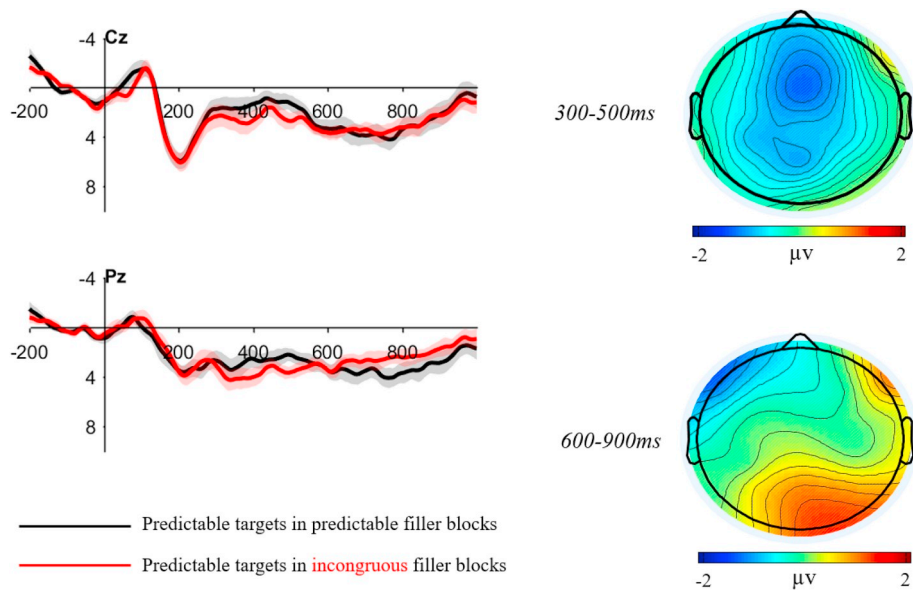
Fig. 2. Grand average ERPs to predictable target words at central electrode CZ and posterior electrode PZ and topographic distribution of ERP effects (predictable filler minus incongruous filler) in the 300–500 ms and 600–900 ms intervals in Experiment 1. A 20 Hz low pass filter were applied to the waveforms for the purpose of illustration.

Table 3
Omnibus ANOVA F-values at the target word in Experiment 1.

	df	300–500 ms	600–900 ms
Filler type	1, 17	4.32°	<1
Filler type × Anteriority	2, 34	1.72	2.37
Filler type × Laterality	2, 34	2.72°	<1
Filler type × Anteriority × Laterality	4, 68	1.78	<1

° $p < 0.1$; * $p < 0.05$; ** $p < 0.01$.

followed by a late frontal positivity. However, if comprehenders' sensitivity to predictability is weakened when the overall predictive validity of the stimuli is low, then the effect of predictability on both the N400 and the late frontal positivity should be smaller in the incongruous filler blocks than in the predictable filler blocks. However, as we mentioned in the Introduction, the present study was not designed to isolate the effect of prediction beyond that of semantic priming. Therefore, we remain agnostic about the specific mechanisms that underlie comprehenders' sensitivity to a word's predictability.

3.1. Methods

3.1.1. Participants

Eighteen participants (6 males; mean age = 21.1) from the same participant pool as Experiment 1 participated in the present experiment. None of them had participated in Experiment 1 or either of the norming studies. All participants gave informed consent and were paid for their participation. Data from two additional participants were excluded due to excessive artefacts (>30%) in their ERP data.

3.1.2. Materials

As in Experiment 1, the materials consisted of 80 experimental sentences and 80 filler sentences (40 predictable fillers + 40 incongruous fillers). The only difference between the stimuli of the present experiment and those used in Experiment 1 was that the predictable sentence-final target word in all of the experimental items was replaced with a word that was unpredictable but semantically congruous. These unpredictable target words had zero cloze probability and the sentences

had an average plausibility rating of 4.08 out of 5 (SD = 0.36; see description of norming procedures in Experiment 1). Further, the unpredictable target words in the present experiment were matched with the predictable target words in Experiment 1 on their number of strokes (predictable: 15.8 vs. unpredictable: 16.3; $t < 1$) and word frequency (predictable: 22 vs. unpredictable: 20 per miillion; $t < 1$; Cai and Brysbaert, 2010). However, as we noted previously, the unpredictable targets words tended to be less semantically related to the words in the sentence context than their predictable counterparts. The same set of predictable and incongruous fillers from Experiment 1 were used in this experiment, such that the stimuli had 50% and 0% predictive validity in the predictable filler and incongruous filler conditions respectively.

3.1.3. Procedure

The experimental procedures were identical to those in Experiment 1.

3.1.4. EEG recording

The EEG recording procedures were identical to those in Experiment 1.

3.1.5. Data analysis

The procedures for behavioural and ERP data analysis were identical to those in Experiment 1. A total of 16% of experimental trials were excluded from data analysis due to artefacts in the EEG data.

In addition, we analysed the data of both experiments together to examine the effect of predictability and its potential interaction with filler type. Plausibility ratings and response times were analysed with a 2 × 2 mixed-model ANOVA; ERP data in each time window (300–500 ms and 600–900 ms) were analysed with a 2 (Predictability) × 2 (Filler type) × 3 (Anteriority) × 3 (Laterality) mixed-model ANOVA. Predictability was treated as a between-subjects factor while filler type, anteriority and laterality were treated as within-subjects factors.

3.2. Results

3.2.1. Behavioural Results

Participants' average plausibility ratings and response times are presented in Table 4. Paired sample t-tests comparing participants' responses to the experimental sentences in the predictable and incongruous filler blocks revealed no significant difference in plausibility ratings ($t(17) = -1.61$, $p > 0.1$), but that response times were longer in the predictable filler blocks than in the incongruous filler blocks ($t(17) = 2.40$, $p < 0.05$).

3.2.2. ERP results

Fig. 3 shows the grand average ERPs at the unpredictable target words and the topographic distribution of the effects of filler type in the 300–500 ms and the 600–900 ms intervals. Results of the statistical analyses are presented in Table 5.

In the 300–500 ms time window, repeated measures ANOVA showed a marginally significant main effect of filler type ($F(1, 17) = 3.67$, $p < 0.10$, $\eta_p^2 = 0.18$) (predictable filler blocks: mean = 0.18, SE = 0.58; incongruous filler blocks: mean = 0.95, SE = 0.65) and no interaction between filler type and either topographic factor ($ps > 0.2$). As in Experiment 1, the N400 was once again smaller in the incongruous filler blocks than in the predictable filler blocks but the difference failed to reach statistical significance. There were no significant effects involving filler type ($Fs < 1$) in the 600–900 ms time window (predictable filler blocks: mean = 3.10, SE = 0.41; incongruous filler blocks, mean = 3.21, SE = 0.50).

3.2.3. Combined Analysis (Experiments 1 and 2)

Behavioural Results. Fig. 4 shows the average plausibility ratings and RTs of experimental sentences in both experiments. We found a significant main effect of predictability in both measures (plausibility ratings: $F(1, 34) = 111.07$, $p < 0.001$, $\eta\eta_p^2 = 0.77$; response times: $F(1, 34) = 19.44$, $p < 0.001$, $\eta_p^2 = 0.36$), showing that participants responded more quickly to predictable sentences than unpredictable sentences, and that their plausibility ratings were also higher for predictable sentences than for unpredictable sentences. No other significant effects were found ($ps > 0.05$).

ERP Results. Fig. 5 shows the effects of predictability (unpredictable minus predictable) in the predictable and incongruous filler blocks. Results of the statistical analyses are presented in Table 6[4].

Table 4
Average plausibility ratings and response times (and standard errors) in Experiment 2.

	Plausibility rating (out of 7)	Response times (ms)
Unpredictable experimental sentences		
Predictable filler blocks	4.45 (0.15)	1181 (90)
Incongruous filler blocks	4.73 (0.18)	1065 (70)
Filler sentences		
Predictable fillers	6.49 (0.10)	718 (79)
Incongruous fillers	1.23 (0.04)	537 (59)

---

[4] Visual inspection of the ERP waveforms suggested that comprehenders' ERPs were more negative for unpredictable words than predictable words even prior to the N400 time window. We conducted additional statistical analysis in the 50–150 ms and 150–250 ms time windows and found a statistically significant effect in the 50–150 ms time window ($F(1, 34) = 4.37$, $p < 0.05$) and a marginally significant effect in the 150–250 ms time window ($F(1, 34) = 3.75$, $p < 0.1$).

In the 300–500 ms time window, the omnibus ANOVA revealed significant main effects of predictability ($F(1, 34) = 4.60$, $p < 0.05$, $\eta_p^2 = 0.12$) and filler type ($F(1, 34) = 7.98$, $p < 0.01$, $\eta_p^2 = 0.19$) and a marginally significant interaction between filler type and laterality ($F(2, 68) = 2.79$, $p < 0.10$, $\eta\eta_p^2 = 0.08$). Follow-up comparisons within each level of laterality revealed that the effect of filler type was significant in left and midline regions ($ps < 0.01$), but only marginally significant in the right region ($p < 0.10$). These results showed that the N400 response was smaller for predictable target words than unpredictable target words, and that it was smaller in the incongruous filler blocks than in the predictable filler blocks. Crucially, there were no significant interactions between predictability and filler type ($ps > 0.30$), which shows that the effect of predictability did not differ between the predictable and incongruous filler blocks.

In the 600–900 ms time window, there were no significant effects involving either predictability or filler type ($ps > 0.10$) in the omnibus ANOVA, despite the fact that unpredictable targets seemed to elicit a larger late frontal positivity than predictable targets in the ERP waveforms (Fig. 5; predictable targets: mean = 2.36, SE = 0.45; unpredictable targets: mean = 3.15, SE = 0.46). We suspect that its failure to reach statistical significance may be due to the fact that predictability was manipulated between experiments.

3.3. Discussion

In Experiment 2, we extended the findings in Experiment 1 and asked whether if comprehenders are less likely to predict (or make weaker predictions) when the overall predictive validity of the stimuli is low. We examined participants' ERP responses to *unpredictable* target words while keeping the same predictable vs. incongruous filler manipulation. We found that participants' N400 response to the unpredictable target words was numerically smaller in the incongruous filler blocks than in the predictable filler blocks, replicating our findings with predictable target words in Experiment 1. Meanwhile, participants' late frontal positivity to the unpredictable target words did not differ at all between the incongruous and predictable filler blocks. Further, we found that participants were slower to respond to the experimental sentences in the plausibility judgment task when the filler sentences were highly predictable compared to when they were incongruous, suggesting that the filler type manipulation did have an impact on participants' processing of the experimental sentences.

Additionally, when we analysed the data from both experiments together, we found that (i) participants' N400 response to both predictable and unpredictable target words was reduced in the incongruous filler blocks then in the predictable filler blocks, (ii) predictable target words elicited a much smaller N400 response than unpredictable target words, and unpredictable target words elicited a larger (albeit statistically insignificant) late frontal positivity than predictable words, and crucially, (iii) the effect of predictability on comprehenders' ERP response was not modulated by filler type at all, such that both the N400 effect and the late frontal positivity did not differ between the predictable and incongruous filler blocks.

Taken together, the present results showed that the overall predictive validity of the stimuli did have an impact on participants' processing of the experimental sentences, but it did not modulate the sensitivity of their ERP responses to predictability.

4. General discussion

The present study examined the effect of experimental context on comprehenders' sensitivity to predictability in two ERP experiments. We manipulated the overall predictive validity of the experimental stimuli in both experiments by presenting the target items alongside the same number of filler items that were either highly predictable or incongruous.
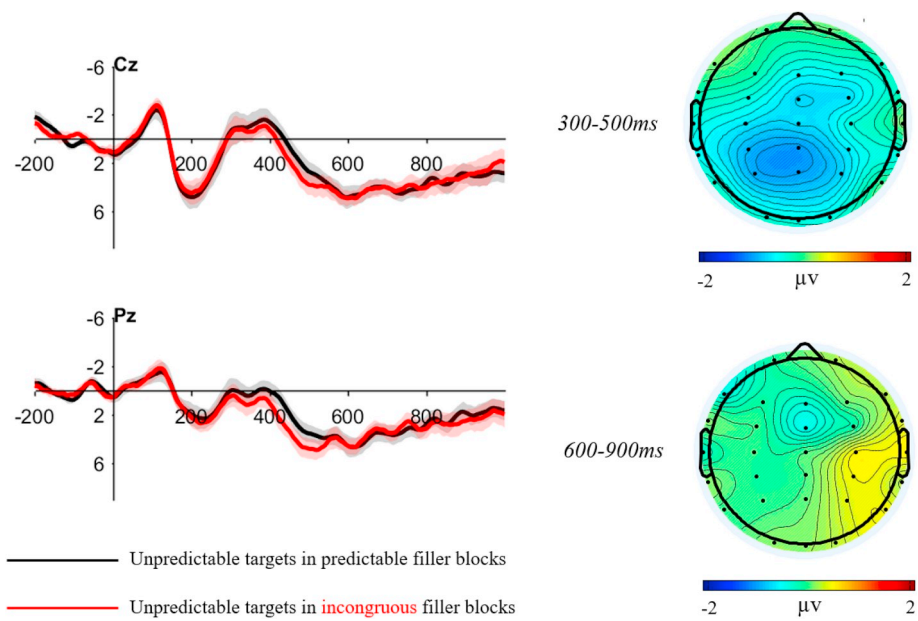
Fig. 3. Grand average ERPs to unpredictable target words at central electrode CZ and posterior electrode PZ and topographic distribution of ERP effects (predictable filler minus incongruous filler) in the 300–500 ms and 600–900 ms intervals in Experiment 2. A 20 Hz low pass filter were applied to the waveforms for the purpose of illustration.

Table 5
Omnibus ANOVA F-values at the target word in Experiment 2.

	df	300–500 ms	600–900 ms
Filler type	1, 17	3.67°	<1
Filler type × Anteriority	2, 34	<1	<1
Filler type × Laterality	2, 34	<1	<1
Filler type × Anteriority × Laterality	4, 68	1.55	<1

° $p < 0.1$; * $p < 0.05$; ** $p < 0.01$.

In the first experiment, we asked whether an experimental context in which all stimuli are highly predictable can facilitate the processing of predictable words by examining participants' ERP responses to predictable target words with a focus on the N400 component; in the second experiment we asked whether an experimental context in which all stimuli end with an unpredictable word can make unpredictable words

less difficult to process by examining participants' ERP responses to unpredictable but congruous target words with a focus on the late frontal positivity. Contrary to our expectations, we found that predictable words elicited a numerically smaller N400 response in incongruous filler blocks than in predictable filler blocks, while unpredictable words elicited an identical late frontal positivity in both incongruous and predictable filler blocks. These results suggested that neither the facilitative effect of prediction nor the inhibitory effect of prediction violations was modulated by the overall predictive validity of the stimuli.

Further, when we analysed the data from both experiments together, we found that the effect of target word predictability on comprehenders' ERP response was not modulated by the overall predictive validity of the stimuli at all. That is, the predictability manipulation elicited the same N400 effect and (non-statistically significant) late frontal positivity regardless of whether the experimental items were presented alongside predictable or incongruous fillers. Many previous studies have reported
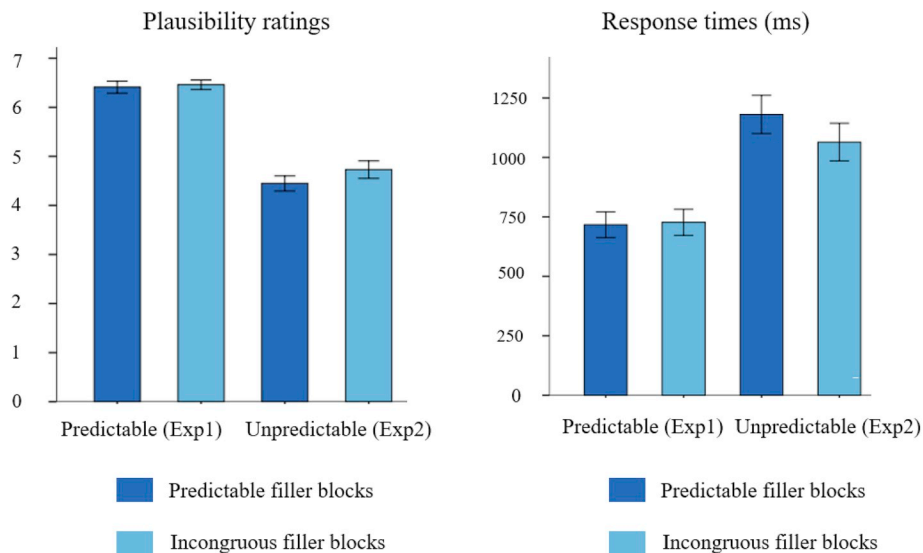

Fig. 4. Average plausibility ratings and response times for experimental sentences in Experiments 1 (predictable) and 2 (unpredictable).
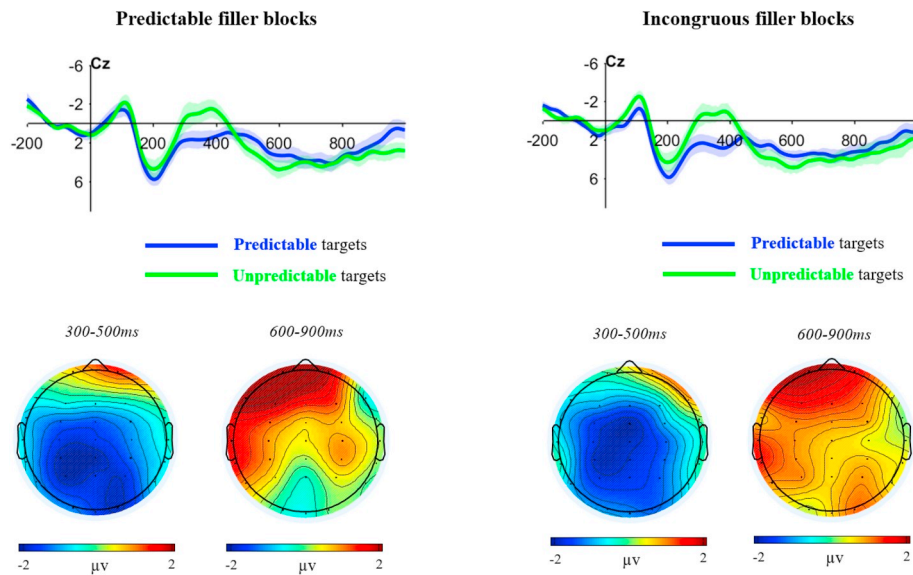
Fig. 5. Grand average ERPs to the predictable and unpredictable target words at Cz in the predictable filler blocks (left) and the incongruous filler blocks (right). The scalp maps show the topographic distribution of the effect of predictability (unpredictable minus predictable) in the 300–500 ms and 600–900 ms intervals.

Table 6
Omnibus ANOVA F-values at the target word in the combined analysis.

	df	300–500 ms	600–900 ms
Predictability	1, 34	4.60*	1.75
Filler type	1, 34	7.98**	<1
Predictability × Filler type	1, 34	<1	<1
Predictability × Anteriority	2, 68	<1	1.60
Predictability × Laterality	2, 68	<1	<1
Filler type × Anteriority	2, 68	<1	2.30
Filler type × Laterality	2, 68	2.79°	<1
Predictability × Filler type × Anteriority	2, 68	1.40	<1
Predictability × Filler type × Laterality	2, 68	<1	<1
Predictability × Filler type × Anteriority × Laterality	4, 136	<1	<1

° $p < 0.1$; * $p < 0.05$; ** $p < 0.01$.

a reduced N400 response to predictable words and a larger late frontal positivity to unpredictable but semantically congruous words (for a review see Van Petten and Luka, 2012). What is new in the present results is that these widely reported ERP effects of predictability seemed completely unaffected by the overall predictive validity of the stimuli.

Note, however, it was not the case that comprehenders were simply completely insensitive to the filler type manipulation. In both experiments we found that participants' N400 response to the target words was smaller in the incongruous filler blocks than in the predictable filler blocks, while in Experiment 2 we found that the filler type manipulation also affected how quickly participants were able to respond to the experimental sentences in a plausibility judgment task. Set against these observations, the present findings suggested that even though the overall predictive validity of the stimuli did impact comprehenders' processing of the stimuli, its impact did not differ between more or less predictable words.

4.1. Effects of filler type

One unexpected finding from the present study is that comprehenders' N400 response to both predictable and unpredictable target words was reduced when the filler items were incongruous (compared to when the fillers were predictable). For the predictable target words (i.e., Experiment 1), one may argue that participants were more 'surprised' to see a predictable sentence-final word in the incongruous filler blocks (where only 50% of the sentences ended with a predictable word) than

in the predictable filler blocks (where 100% of the sentences ended with a highly predicable word). An ERP component known as P3b is known to vary as a function of stimulus probability (Polich, 2007). Since the P3b is a positive-going ERP component which may occur at a similar time and with a similar scalp distribution as the N400, one might argue that the "reduced N400" may in fact be an increased P3b response, indexing the reduced probability of a predictable target word in the incongruous filler blocks relative to the predictable filler blocks.

However, this account would not be able to accommodate the results for the unpredictable target words in Experiment 2. This is because, even though the unpredictable words should have been more 'surprising' in the predictable filler condition (where 50% as opposed to 100% of the sentences ended with an unpredictable word), the N400 response was still smaller (or the P3b was still larger) in the incongruous filler blocks than in the predictable filler blocks. Therefore, at the moment we do not have a complete account of the effect of filler type in the present study, and we hope that these data may stimulate further research to investigate this issue.

4.2. Relating the present findings to Brothers et al. (2017)

At first glance, the present ERP findings may appear at odds with the recent findings in a self-paced reading study by Brothers et al. (2017). In that study, they manipulated the predictive validity of their stimuli by presenting predictable and unpredictable experimental sentences along with filler sentences that were either (i) all predictable, (ii) 50% predictable sand 50% unpredictable, or (iii) all unpredictable. They found that, contrary to the effect of lexical frequency which was preserved across all levels of predictive validity, the effect of predictability on participants' reading times was greatly attenuated when most of the stimuli were unpredictable.

However, we believe the apparent differences in the results may be better understood by considering the differences between the tasks and dependent measures used in the present study and Brothers et al. (2017).

One key difference between the two studies the control participants had (or did not have) over the rate at which the linguistic stimuli were presented. Brothers et al. (2017) used a self-paced reading paradigm, in which participants read sentences one word at a time and can control when to proceed from one word to the next by pressing a button. As such, participants were free to adjust their reading speed throughout the experiment. Meanwhile, in the present study, stimuli were presented

at the same speed across all conditions, which means that participants had no control over the rate at which words were presented on the computer screen.[5]

Another important difference between these studies lies in their dependent measures. In Brothers et al. (2017), participants had to press a button to proceed from one word to the next and their reading times (or more precisely, the time it took them to press a button upon seeing a word) were taken as a measure of processing cost, such that longer reading times were taken to reflect higher processing costs. Meanwhile, in the present study we measured participants' ERP brain responses directly from their scalp as they read, bypassing all overt behaviours (e. g., button presses). Following previous research, we took the amplitude of their N400 response and late frontal positivity to make inferences about underlying cognitive processes (e.g., reduced N400 responses are taken to reflect facilitated long-term semantic access).

We believe that the present study and Brothers et al. (2017) can in fact complement each other in at least two ways. First, by providing a richer measure of real-time processing, ERPs may be able to reveal smaller changes in cognitive processes that are difficult to observe in reading time measures, or changes that occur at different stages of processing that are not distinguishable using reading times alone. Further, although the self-paced reading paradigm allows comprehenders to read at a more natural pace, it cannot tell us whether predictability has an effect on comprehenders' processing beyond their reading speed. For instance, Brothers et al.'s (2017) observation that predictability had almost no effect of comprehenders' reading times may be taken to show that comprehenders can adjust their reading speed (or reading strategy) depending on the overall predictive validity of the stimuli, such that they may read more slowly and rely more heavily on the bottom-up information when most of the stimuli had an unpredictable ending. Meanwhile, presenting stimuli at the same rate across conditions has allowed us to examine the effect of predictability beyond reading speed. In fact, the present results suggest that, when reading speed is held constant across conditions, predictability has the same effect on comprehenders' brain responses (both the N400 and the late frontal positivity) regardless of the predictive validity of the stimuli.

Therefore, the results from both of these studies taken together may suggest that (i) comprehenders may rely more heavily on bottom-up information and as such read more slowly when overall predictive validity is low, but (ii) when the uptake of bottom-up information is held constant, processing of predictable words in the brain is facilitated relative to unpredictable words regardless of the stimuli's overall predictive validity. One prediction that emerges from this proposal is that the present results should replicate provided that the stimulus presentation rate is held constant between conditions (e.g., if the stimuli are presented auditorily instead of visually). Future research will be needed to investigate this.

4.3. Conclusion

The present study found that both the N400 and the late frontal positivity displayed the same sensitivity to a word's predictability regardless of the overall predictive validity of the experimental context. This extends previous findings by showing that even though the makeup of the stimuli can affect how comprehenders process language in real time, it did not make them more (or less) sensitive to a word's predictability. We propose that while comprehenders may read more slowly when overall predictive validity is low, they routinely generate predictions about upcoming language and the facilitative effects of prediction can be observed even when the experimental context does not encourage prediction.

---

[5] This is comparable to natural listening comprehension where listeners also have little control over the stimuli presentation rate.

References

Boston, M.F., Hale, J., Kliegl, R., Patil, U., Vasishth, S., 2008. Parsing costs as predictors of reading difficulty: an evaluation using the Potsdam Sentence Corpus. J. Eye Movement Res. 2 (1).

Brothers, T., Swaab, T.Y., Traxler, M.J., 2017. Goals and strategies influence lexical prediction during sentence comprehension. J. Mem. Lang. 93, 203–216.

Brown, C.M., Hagoort, P., Chwilla, D.J., 2000. An event-related brain potential analysis of visual word priming effects. Brain Lang. 72 (2), 158–190.

Cai, Q., Brysbaert, M., 2010. SUBTLEX-ch: Chinese word and character frequencies based on film subtitles. PLoS One 5 (6), e10729.

Chow, W.Y., Smith, C., Lau, E., Phillips, C., 2016. A "bag-of-arguments" mechanism for initial verb predictions. Lang. Cognit. Neurosci. (5), 1–20.

DeLong, K.A., Quante, L., Kutas, M., 2014. Predictability, plausibility, and two late ERP positivities during written sentence comprehension. Neuropsychologia 61, 150–162.

Delong, K.A., Urbach, T.P., Groppe, D.M., Kutas, M., 2011. Overlapping dual ERP responses to low cloze probability sentence continuations. Psychophysiology 48 (9), 1203–1207.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134 (1), 9–21.

Demberg, V., Keller, F., 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. Cognition 109 (2), 193–210.

Federmeier, K.D., 2007. Thinking ahead: the role and roots of prediction in language comprehension. Psychophysiology 44 (4), 491–505.

Federmeier, K.D., Kutas, M., 1999. A rose by any other name: long-term memory structure and sentence processing. J. Mem. Lang. 41 (4), 469–495.

Federmeier, K.D., Wlotko, E.W., De Ochoa-Dewald, E., Kutas, M., 2007. Multiple effects of sentential constraint on word processing. Brain Res. 1146, 75–84.

Greenhouse, S.W., Geisser, S., 1959. On methods in the analysis of profile data. Psychometrika 24 (2), 95–112.

Guntkr, T.C., Stowe, L.A., Mulder, G., 1997. When syntax meets semantics. Psychophysiology 34 (6), 660–676.

Hahne, A., Friederici, A.D., 1999. Electrophysiological evidence for two steps in syntactic analysis. Early automatic and late controlled processes. J. Cogn. Neurosci. 11 (2), 194–205.

Holcomb, P.J., 1988. Automatic and attentional processing: an event-related brain potential analysis of semantic priming. Brain Lang. 35 (1), 66–85.

Huettig, F., Guerra, E., 2019. Effects of speech rate, preview time of visual context, and participant instructions reveal strong limits on prediction in language processing. Brain Res. 1706, 196–208.

Kamide, Y., 2008. Anticipatory processes in sentence processing. Lang. Linguist. Compass 2 (4), 647–670.

Kelly, A.M.C., Uddin, L.Q., Biswal, B.B., Castellanos, F.X., Milham, M.P., 2008. Competition between functional brain networks mediates behavioral variability. Neuroimage 39 (1), 527–537.

Kliegl, R., Nuthmann, A., Engbert, R., 2006. Tracking the mind during reading: the influence of past, present, and future words on fixation durations. J. Exp. Psychol. Gen. 135 (1), 12–35.

Kuperberg, G.R., Jaeger, T.F., 2016. What do we mean by prediction in language comprehension? Lang. Cognit. Neurosci. 31 (1), 32–59.

Kutas, M., 1993. In the company of other words: electrophysiological evidence for single-word and sentence context effects. Lang. Cogn. Process. 8 (4), 533–572.

Kutas, M., Federmeier, K.D., 2010. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annu. Rev. Psychol. 62, 621–647.

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

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

Lau, E.F., Phillips, C., Poeppel, D., 2008. A cortical network for semantics:(de) constructing the N400. Nat. Rev. Neurosci. 9 (12), 920–933.

Lau, E.F., Weber, K., Gramfort, A., Hamalainen, M.S., Kuperberg, G.R., 2014. Spatiotemporal signatures of lexical-semantic prediction. Cerebr. Cortex 26 (4), 1377–1387.

Michael, D., Reinhold, K., Markus, H., Jacobs, A.M., 2006. Frequency and predictability effects on event-related potentials during reading. Brain Res. 1084 (1), 89–103.

Nieuwland, M.S., Barr, D.J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D.I., Wolfsthurn, S.V.G.Z., 2018. Dissociable Effects of Prediction and Integration during Language Comprehension: Evidence from a Large-Scale Study Using Brain Potentials. bioRxiv (in press).

Polich, J., 2007. Updating P300: an integrative theory of P3a and P3b. Clin. Neurophysiol. 118 (10), 2128–2148.

Rayner, K., Well, A.D., 1996. Effects of contextual constraint on eye movements in reading: a further examination. Psychon. Bull. Rev. 3 (4), 504–509.

Smith, N.J., Levy, R., 2013. The effect of word predictability on reading time is logarithmic. Cognition 128 (3), 302–319.

Taylor, W.L., 1953. Cloze procedure": a new tool for measuring readability. Journal. Q. 30 (4), 415–433.

Van Petten, C., Luka, B.J., 2012. Prediction during language comprehension: benefits, costs, and ERP components. Int. J. Psychophysiol. 83 (2), 176–190.

Wlotko, E.W., Federmeier, K.D., 2012. So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. Neuroimage 62 (1), 356–366.