The Interaction of Context Constraints and Predictive Validity during Sentence Reading

René Terporten[1], Eleanor Huizeling[2], Karin Heidlmayr[3], Peter Hagoort[2,4], and Anne Kösem[5]

Abstract

■ Words are not processed in isolation; instead, they are commonly embedded in phrases and sentences. The sentential context influences the perception and processing of a word. However, how this is achieved by brain processes and whether predictive mechanisms underlie this process remain a debated topic. Here, we employed an experimental paradigm in which we orthogonalized sentence context constraints and predictive validity, which was defined as the ratio of congruent to incongruent sentence endings within the experiment. While recording electroencephalography, participants read sentences with three levels of sentential context constraints (high, medium, and low). Participants were also separated into two groups that differed in their ratio of valid congruent to incongruent target words that could be predicted from the sentential context. For both groups, we investigated modulations of alpha power before, and N400 amplitude modulations after target word onset. The results reveal that the N400 amplitude gradually decreased with higher context constraints and cloze probability. In contrast, alpha power was not significantly affected by context constraint. Neither the N400 nor alpha power were significantly affected by changes in predictive validity. ■

INTRODUCTION

In daily language use, words are not processed in isolation, but are embedded in phrases or sentences. It is known that sentences create a context that can bias the perception and processing of a word. For instance, contextual information processed during sentence reading is known to facilitate the processing of new linguistic input (Kuperberg & Jaeger, 2016). Although this phenomenon is well documented (Freunberger & Roehm, 2017; Ito, Corley, Pickering, Martin, & Nieuwland, 2016; Frank, Otten, Galli, & Vigliocco, 2015; Van Petten & Luka, 2012; Kutas & Federmeier, 2011; DeLong, Urbach, & Kutas, 2005), the mechanisms at the neurobiological origins of processing sentential linguistic information are still debated (Nieuwland et al., 2020; Huettig & Guerra, 2019; Huettig, 2015). On the one hand, the effects of context constraints could occur incrementally via integration mechanisms, which consist in integrating the (bottom–up) activated word meaning with its context upon its presentation (Huettig, 2015; Lau, Holcomb, & Kuperberg, 2012; Bar, 2007; Gerrig & McKoon, 1998). Conversely, the processing of contextual information could result from neurobiological mechanisms that support linguistic prediction (Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007). On the basis of the contextual information, the brain could build predictions about certain linguistic features of the incoming words before the arrival of the sensory evidence.

Brain oscillatory responses in the alpha (8–12 Hz) frequency range have been linked to linguistic predictive mechanisms before the occurrence of a target word (Piai, Rommers, & Knight, 2018; Rommers, Dickson, Norton, Wlotko, & Federmeier, 2017; Wang, Hagoort, & Jensen, 2017; Lam, Schoffelen, Uddén, Hultén, & Hagoort, 2016). Alpha desynchronization before word presentation was shown to be greater for highly predictable words compared with unpredictable words, as determined by the prior sentential context constraints (Piai et al., 2018; Rommers et al., 2017; Wang et al., 2017; Bastiaansen & Hagoort, 2015; Willems, Oostenveld, & Hagoort, 2008), although the direct link between alpha power and linguistic predictability has been challenged in a recent report from our laboratory (Terporten, Schoffelen, Dai, Hagoort, & Kösem, 2019). Terporten and colleagues (2019) used varying degrees of sentential constraints to influence linguistic predictability. Alpha power before target word occurrence was found to be modulated by context constraint. However, against initial expectations, it was not monotonically related to the strength of sentential context constraints. Instead, alpha power before the sentence-final word was lower for sentences with medium context constraints as compared with sentences with high or low

[1]Radboud University, Nijmegen, The Netherlands, [2]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, [3]Université Paris-Nanterre, [4]Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands, [5]Centre for Research in Neuroscience Lyon, France

https://doi.org/10.1162/jocn_a_02082

context constraints. We argued, based on these results, that alpha power was not a suitable marker of the amount of predictability of a word ending, where we would have expected a gradual change in alpha power as a function of context constraint (Terporten et al., 2019). However, given that alpha power was sensitive to the sentence context constraints, we did not fully exclude the hypothesis that alpha oscillations could be linked to mechanisms involved in linguistic predictions. For example, it could be that alpha relates to the amount of competition between possible continuations. Specifically, it could be that the number of potential lexical candidates as continuation is low under both high context constraints and low context constraints (as few items are competing in high context constraints, and that prediction would be deemed too difficult under low context constraints). In contrast, medium context constraints could present an intermediate level, where participants are still attempting to predict but under more challenging constraints (Terporten et al., 2019).

To further test to what degree alpha oscillations are involved in predictive mechanisms during language processing, we here experimentally dissociated the effect of sentential context from the effect of the linguistic predictive validity. Predictive validity refers to the relevance and reliability of predictive mechanisms during sentence processing (Brothers, Dave, Hoversten, Traxler, & Swaab, 2019; Lau et al., 2012). Manipulating predictive validity during reading can either be done explicitly, by asking participants to actively predict, or not, the sentence ending (Brothers, Swaab, & Traxler, 2017), or implicitly, by manipulating the relevance of sentential contextual information in the prediction of sentence endings or word-pairs associations (with the hypothesis that individuals flexibly adapt to the global predictability of the larger linguistic context by no longer predicting when the contextual information is deemed unreliable; Brothers et al., 2019; Delaney-Busch, Morgan, Lau, & Kuperberg, 2019; Lau et al., 2012; Brown, Hagoort, & Chwilla, 2000; Holcomb, 1988). Previous EEG studies have shown that the amplitude of N400 responses and post-N400-positivities (PNPs), which have been shown to be sensitive to semantic plausibility and predictability (Kuperberg, Brothers, & Wlotko, 2020; DeLong, Quante, & Kutas, 2014; Van Petten & Luka, 2012), are dependent on predictive validity. Bigger N400 and PNP amplitudes were observed when participants were explicitly asked to predict the last word of the sentence, compared with when they were asked to understand the sentence (Brothers et al., 2017). Manipulating predictive validity by changing the proportion of predictable relative to unpredictable word-pairs or sentences endings within the experimental design, the modulation of the N400 by context was found to be more pronounced when the implicit predictive validity was high, that is, when the experimental block contained a large number of highly predictable word-pairs (Delaney-Busch et al., 2019; Lau et al., 2012) or when the speaker was uttering sentences with highly predictable endings (Brothers et al., 2019). Here, our main aim was to test whether alpha oscillations before target presentation would also be sensitive to the predictive validity of context.

The current study therefore investigated whether alpha oscillations before target presentation, as well as N400/PNP responses after target presentation, were sensitive to the predictive validity of sentential context. Participants passively read sentences with either a high (HC), medium (MC), or low (LC) context constraint. Predictive validity was manipulated as the ratio of congruent to incongruent sentence endings within the set of sentences presented to the participant. Specifically, participants were split into two groups: The high-predictive validity group received 80% congruent target words to the sentence context, the low-predictive validity group received mainly 80% incongruent target words with regard to sentential context (see Table 1 for stimuli examples). We measured the influence of sentence context constraint and language predictive validity on three distinct neural

Table 1. Example Dutch Sentences in Each Condition with Their English Translation

Condition	*Stimulus Sentence (Congruent/Incongruent Ending)*
HC	(NL) Morgen geeft de *priester* een toespraak in een (kerk/bal).
	(EN) Tomorrow the *priest* will give a speech in the (church/ball).
MC	(NL) Morgen geeft de *weduwe* een toespraak in een (kerk/bal).
	(EN) Tomorrow the *widow* will give a speech in the (church/ball).
LC	(NL) Morgen geeft de *enthousiasteling* een toespraak in een (kerk/bal).
	(EN) Tomorrow the *enthusiast* will give a speech in the (church/ball).

The context constraining conditions—HC, MC, and LC constraints—were manipulated by changing one context constraining word (in *italic*). The sentence ending was either congruent or incongruent with the context (within parentheses).

markers, the N400, the PNP, and alpha (8–12 Hz) power, recorded with EEG. We investigated how N400/PNP amplitudes at (sentence-final) target word onset and pretarget word alpha power were modulated by context constraints and by the predictive validity of the target words. In line with the literature, we expected the N400/PNP amplitudes to be gradually influenced by sentence context constraints and cloze probability, with higher constraints/cloze probability resulting in a reduction in amplitude (with a less negative N400 and less positive PNP). Aiming to replicate our earlier study (Terporten et al., 2019), we expected pretarget alpha power to be modulated by sentence context constraints and to be nonmonotonically linked to the predictability of the target word. Expecting an association between alpha power and the brain's employment of predictive mechanisms during the encoding of sentential context, we hypothesized that the modulation of alpha power before target word onset would be observed in the high-predictive validity group and not in the low-predictive validity group.

METHODS

Participants

In total, 70 participants were invited from the participant pool of the Max Planck Institute for Psycholinguistics, Nijmegen. All participants gave their informed written consent in accordance with the Declaration of Helsinki and the local ethics committee (Commissie Mensgebonden Onderzoek region Arnhem-Nijmegen). All participants were Dutch native speakers, were right-handed, had normal or corrected-to-normal vision, and did not suffer from any neurological impairment or dyslexia. After completion of the experiment, the participants received €18. One participant did not finish the experiment and was excluded such that 69 participants (mean age = 25 years, range = 19–41 years; 20 men) were included for the analyses.

Stimulus Material

The stimulus set used in this study consisted of 203 critical sentence triplets from Terporten and colleagues (2019). Although the original stimulus set contained only congruent sentence endings, an additional set of sentences was created with incongruent (semantically implausible) sentence endings (see Table 1 for examples). This approach resulted in two sets of sentence triplets: a congruent and an incongruent set of sentence triplets. Manipulations in predictive validity were achieved by changing the ratio of congruent to incongruent sentence endings for the final experimental set of sentence triplets. Each sentence for both groups belonged to either a HC, MC, or LC constraining condition. The degree of constraint for a given sentence was manipulated by changing one word, the *context constraining word*. This word was always at the same position within a sentence with regard to a triplet (Table 1). Across the conditions, these context-constraining words were matched with regard to word length, $F(2, 606) = 0.78$, $p = .457$, with a mean (SE) of HC: 7.12 (2.26); MC: 7.1 (2.54); LC: 7.37 (2.61), and word frequency, $F(2, 584) = 1.98$, $p = .138$, with mean (SE) of HC: 2.4 (0.78); MC: 2.56 (0.87); LC: 2.5 (0.84), based on the Dutch SUBTLEX-NL database (Keuleers, Brysbaert, & New, 2010). The degree of context constraints was measured in Terporten and colleagues (2019) by using a sentence completion task in which participants had to fill in a missing target word based on a preceding sentence context. The degree of context constraint per sentence was evaluated by calculating the percentage of participants that finished a sentence with the same word. A percentage of context constraint of 50% means that half of the participants filled in the sentence ending with the same word. The degree of context constraints differed significantly between constraining conditions, $F(2, 606) = 442.84$, $p < .001$. HC sentences showed the strongest degree of context constraints, mean (SE) = 77% (17.74); followed by MC, mean (SE) = 50% (18.67); and LC, mean (SE) = 28% (11.97). As in Terporten and colleagues (2019), for all sentences in the congruent condition, the most answered word from the HC sentences was chosen as the ending target word for all sentences within the context constraint triplet (see Table 1 for examples). Once the sentences were constructed, we computed the cloze probability of the sentence endings, that is, the percentage of participants that responded with the selected target word as the sentence ending. The cloze probabilities of the congruent target words differed significantly between conditions, $F(2, 606) = 468.16$, $p < .001$, with HC showing the highest cloze probability, mean (SE) = 77% (17.74); followed by MC, mean (SE) = 42% (25.94); and LC, mean (SE) = 15% (15.82). Measures of context constraints highly correlated with measures of cloze probability for *congruent* target words ($r = .93$, $p < .001$). In addition to the congruent stimulus set, a stimulus set was created with 203 sentences ending with *incongruent* target words. The congruent and the incongruent stimulus set differed significantly from each other on pretested ratings of plausibility, $F(1, 1312) = 4772.23$, $p < .001$. The incongruent target words did not occur in the pretest of the congruent stimulus set, and therefore, all have a cloze probability of 0%. Congruent and incongruent target words were matched on word length, $t(404) = -1.12$, $p = .264$, with a mean (SE) of congruent: 5.79 (0.14); incongruent: 6.0 (0.13), and word frequency, $t(404) = 1.29$, $p = .199$, with a mean (SE) of congruent: 3.07 (0.05); incongruent: 2.98 (0.04); based on the Dutch SUBTLEX-NL database (Keuleers et al., 2010).

A practice stimulus set was also created, including a selection of 50 sentences in total, split in congruent and incongruent sentences from Wang and colleagues (2017). Half of the sentences were defined as HC, whereas the other half was defined as LC for each congruency condition separately (see Wang et al., 2017, for details). For the EEG experiment, six counterbalanced lists were created. Three of these lists contained 80% of congruent target words, whereas the other three lists contained 80% of incongruent target words. The practice stimulus set was thought to bias participants' expectation of the predictive validity of the context constraints, toward the respective proportion of (in)congruent target words in the critical stimulus set. For all lists, the three levels of context constraints were randomly distributed across the set.

Experimental Procedure

Participants were comfortably seated in front of a screen in a dimly illuminated room. They were instructed to rest their right arm on the table in front of them and to access a button box with their right hand. At 70 cm and with a 25°–35° viewing angle, a screen was located to which all stimuli were projected. Written stimuli were shown in black, on a gray background.

Participants were instructed to silently read a word-by-word display of sentences on the screen and to focus on the content of each sentence. It was explained that sometimes (after 25% of the sentences; participants were not informed about the precise percentage) a question would be prompted about the content of the previously displayed sentence. For instance, the question "Did the person have something on his head?" could appear after the presentation of the sentence "The old king wears on his head a remarkable crown." The participants were required to answer this question with "yes" or "no" by button press. The answer possibilities (yes/no) were displayed randomly on the left or right side of the screen, and a left or right button had to be pressed accordingly. The occurrence of these questions throughout the experiment was at random intervals. The goal of these questions was to ensure that the participant kept processing the content of the sentence material throughout the experimental session.

A trial began with the presentation of a fixation cross in the middle of the screen for 500 msec. This was followed by a blank screen for a jittered interval of 500–1200 msec. The sentences were presented word-by-word. Each word was displayed for 200 msec, followed by a blank screen of 800 msec. The ISI of 1000 msec was chosen to avoid the influence of the evoked response from stimulus onset onto pretarget alpha activity. Another blank screen occurred for 2000 msec (Figure 1) after (sentence final) target word offset. In 25% of the cases, the catch question was displayed, with the full question centered on the screen and the yes–no answers randomly split to the left or right side.

Participants were presented with 50 practice sentences from Wang and colleagues (2017) at the start of the experiment to prime the statistics of the experimental predictive validity. This was followed by 203 critical sentences (203 trials). The 203 sentences from the main experiment were presented in a random order that were predetermined in

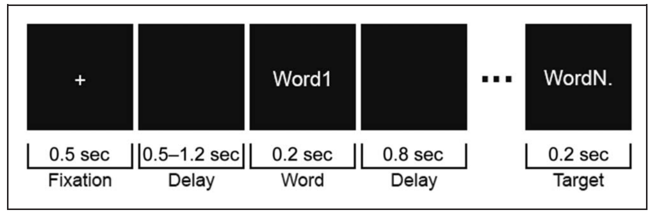

Figure 1. A schematic display of a trial procedure showing the duration of each screen. A trial began with the display of a fixation period, followed by a blank screen. Subsequently, the sentence was visually displayed by a word-by-word presentation, up to the final word as indexed by the period. Between words, a black screen served as delay before a subsequent word was shown.

six different lists. The lists were constructed so that one sentence per triplet was selected for each participant and that they were counterbalanced with regard to context constraint. Three of the lists were composed of sentences that mostly had congruent endings (high-predictive validity list) or incongruent endings (low predictive validity list). The list was chosen at random for each participant, which resulted in half of the participants having a high-predictive validity list, whereas the other half had to the low predictive validity list. The practice set had the same proportion of congruent sentences that the main experiment, that is, the practice set had mostly congruent (resp. incongruent) endings when the main experimental list had mostly congruent (resp. incongruent endings). Trial presentation was divided into four blocks, separated by self-paced breaks in between. In total, the experimental procedure took 60 min.

Data Acquisition

The participants' EEG was recorded online. A custom actiCAP 64-electrode montage (Brain Products) was used, with 58 equidistant electrodes mounted in the cap. Four electrodes measured EOG, with two HEOG electrodes placed next to the left and right eyes. VEOG was measured by placing an electrode above and below the participant's left eye. The reference electrode was placed on the left mastoid. The ground electrode was placed on the forehead, above the nasion. Data were filtered online with a high-pass filter of 0.02 Hz and a low-pass filter of 500 Hz.

Data Preprocessing

All data were analyzed using the MATLAB 2016a open source toolbox Fieldtrip (Oostenveld, Fries, Maris, & Schoffelen, 2011). Data were segmented 1.5 sec before and after the onset of the target word for each trial, which included the blank 800-msec period before the target word presentation. The segmented data were low-pass filtered at 150 Hz and high-pass filtered at 0.1 Hz and rereferenced to the average of the left and right mastoids. The 50-Hz line-noise component was removed using a discrete

Fourier transform filter. Remaining strong line noise and muscle artifacts were identified first by visual inspection of amplitude variance over trials, and the corresponding trials were removed. Second, artifacts related to eye movements were removed by means of an independent component analysis (fastICA; Hyvärinen & Oja, 2000), followed by back projection. Bad channels were repaired by replacing them with the plain average of the nearest neighbors. Third, the resulting data were again visually inspected on a trial-by-trial basis and trials with remaining artifacts were removed. From this procedure and for both groups, 6% of trials were excluded on average from further analysis.

ERP Analysis

ERPs were investigated to observe N400 and PNP amplitude modulations after target word onset as a function of Constraint (HC, MC, LC), Congruency (congruent, incongruent) and Predictive Validity (high, low). Per condition, preprocessed epochs were low-pass filtered at 35 Hz. Baseline correction was performed using a time window of −300 to 0 msec relative to target word onset. The N400 component was calculated by averaging amplitudes from 250 to 600 msec following target word onset. A time window of 250–600 msec was selected for the N400 analysis, rather than the typical 300–500 msec, to remain consistent with Terporten and colleagues (2019). Both a 250- to 600-msec and 300- to 500-msec time window yielded a similar pattern of results. The PNP was calculated by averaging amplitudes from 600 to 1000 msec (Kuperberg et al., 2020) following target word onset. Cluster-based permutation statistics (Maris & Oostenveld, 2007) were used to identify a cluster of channels that resulted from a significant difference in N400/PNP amplitude between levels of the factor congruency, irrespective of the factors predictive validity and constraint, so as to select the sensors that were most responsive to semantic N400/PNP congruency effects. Resulting clusters included 12 and four sensors selected for the N400 and PNP analyses, respectively. For subsequent statistical analyses, the average amplitude over time and over the channels belonging to this cluster was extracted per participant and per trial separately for the N400 and PNP relevant time windows. All statistical analyses on the extracted data were performed in R software (R Core Team, 2019) Version 4.1.2 by fitting a linear mixed-effects model, using *lmer* from *lme4* (Bates, Mächler, Bolker, & Walker, 2015), to the interaction of the factors, Constraint (within-subject factor): HC, MC, LC; Congruency (within-subject factor): congruent, incongruent; Predictive Validity (between-subjects factor): high, low, with participant × congruency random intercepts and slopes, and a treatment contrast coding scheme to compare the HC condition with each other level of constraint. A model including additional random slopes for constraint did not converge and a model comparison procedure revealed no significant difference in model fit between models with and without random slopes for constraint ($\chi^2 = 3.09$, $p > .10$). We therefore report the model with by-subject random slopes for congruency only. The estimates of the model were interpreted using R's Type II *anova* function and *glht* from the package *multcomp* for further analysis of pairwise comparisons. Correction for multiple comparisons was performed using the Tukey method (Tukey, 1949).

Time–Frequency Analysis

Time–frequency analysis was performed on a time window of −1500 to 1500 msec relative to target word onset. Power was estimated for a frequency range of 2–40 Hz, using a fixed 500-msec sliding Hanning window in time steps of 50 msec, and frequency steps of 2 Hz. No baseline correction was performed on the time–frequency data. Cluster-based permutation statistics (Maris & Oostenveld, 2007) were performed across sensors, with alpha power averaged between −540- and 0-msec time window relative to target word onset. The relevant time window was selected based on our previous findings of the effects of sentence context constraints on alpha (8–12 Hz) power (Terporten et al., 2019). The cluster-based permutations statistics identified a cluster of eight channels that showed an effect of the factor Constraint, irrespective of predictive validity in alpha (8–12 Hz) power (cluster statistic: *F* test of factor Constraint, statistical threshold of $p = 5\%$). This spatial cluster was used as ROI for further analyses. To identify which frequencies were modulated by constraint, a cluster-based permutation analysis was performed across frequencies, during the −540- to 0-msec time window relative to target word onset and within the ROI, to evaluate the effect of the factor Constraint across a broad frequency spectrum (2–30 Hz). For visualization only, alpha (8–12 Hz) power modulation over time was plotted by selecting the average power for the cluster-specific channels within a time window of −1.0 to 1.0 sec relative to target word onset.

For subsequent statistical analyses, the average power over the respective channel cluster from −540 to 0 msec relative to target word onset was extracted per participant and per trial within the ROI. A linear mixed-effects model was fitted to the interaction of the factors, Constraint (within-subject factor): HC, MC, LC; Predictive Validity (between-subjects factor): high, low, with random intercepts for participant. Please note that the factor Congruency was not included in the analysis of alpha power modulations, because we investigated alpha activity before the arrival of the target word. A model containing additional by-subject random slopes for Constraint did not converge. A model comparison revealed that including random slopes for Constraint did not improve the model fit ($\chi^2 = 1.76$, $p > .10$), and so we report the model with random intercepts only. We applied a

Table 2. Model Summary of Comprehension Question Accuracy

Random Effects				
Groups	*Name*	*Variance*	*SD*	
Subject	Intercept	2.78	1.67	
Residual		0.24	0.49	
Fixed Effects	*Estimate*	*SE*	*z Value*	*Pr(>|t|)*
Intercept	3.26	0.35	9.31	< .001
Congruency IC	−1.04	0.31	−3.32	.001
Constraint low	−0.30	0.48	−0.64	.524
Constraint med	−0.64	0.47	−1.37	.172
CongruencyIC:Constraint low	1.32	0.43	3.04	.002
CongruencyIC:Constraint med	1.00	0.42	2.37	.018

df = degrees of freedom; Estimate = beta coefficient; *SD* = standard deviation; *SE* = standard error; IC = incongruent, med = medium constraint.

treatment contrast coding scheme to compare the HC condition with each other level of constraint. The estimates of the model were interpreted using R's *anova* (Type II) function.

RESULTS

Behavioral Performance

Participants' accuracy on the comprehension questions confirmed that they were paying attention to the content of the sentences, with a mean performance of 91% (*SD* = 5.61%), 84% (*SD* = 6.97%), and 89% (*SD* = 6.39) for the HC, MC, and LC sentences, respectively, for the high-predictive validity group. The means of the low-predictive validity group were 83% (*SD* = 7.32%), 84% (*SD* = 3.52%), and 91% (*SD* = 4.01) for the HC, MC, and LC sentences, respectively. Generalized linear mixed-effects models were performed to assess the effect of context constraint and predictive validity on comprehension accuracy. Models containing by-subject and by-item random slopes for constraint or only by-subject random slopes for constraint did not converge, and a model comparison procedure revealed that adding these random slopes did not improve the model fit ($\chi^2 = 3.43$, $p > .10$ and $\chi^2 = 1.05$, $p > .10$, respectively). We therefore reported the models with random intercepts for subject and item only. The model revealed that accuracy was significantly higher for the high- compared with low-predictive validity group ($p < .001$) but did not differ across levels of constraint. A significant interaction was found between the factors Constraint and Predictive Validity (treatment contrast coding scheme against HC: LC $p = .002$, MC $p = .012$). The direction of the slopes from HC to LC and HC to MC reversed and were more positive for the low- compared with high-predictive validity group (see model summary in Table 2).

N400 Amplitude Modulation after Target Word Onset

Cluster permutation statistics were used to identify which channels should be selected for the analysis of the interaction effects based on the factor Congruency, so as to select the channels that are maximally responsive to the contextual modulations of the N400. Collapsed over Predictive Validity and Constraint, the N400 amplitude was significantly modulated by congruency in a cluster over central-posterior electrodes ($p = .003$; Figure 2). To investigate the Constraint × Congruency × Predictive Validity interaction, we analyzed the effect of these factors on N400 amplitude in the identified cluster in a linear mixed-effects model. The modulation of the N400 amplitude as a function of constraint and congruency are presented for each predictive validity group in Figure 3.

A linear mixed-effects model (model summary in Table 3), with by-subject random slopes for congruency and a treatment contrast coding scheme, revealed significant effects of both Congruency, $F(2, 71.5) = 96.02$, $p < .001$, and Constraint, $F(2, 13109.3) = 19.33$, $p < .001$ (please note that in this study, constraint covaries with cloze probability). The N400 amplitude was significantly greater (more negative) in the LC compared with the HC condition ($p < .001$) and LC vs MC condition ($p < .001$). There was no significant difference between the HC and MC conditions ($p > .10$). The N400 amplitude was additionally significantly greater (more negative) in

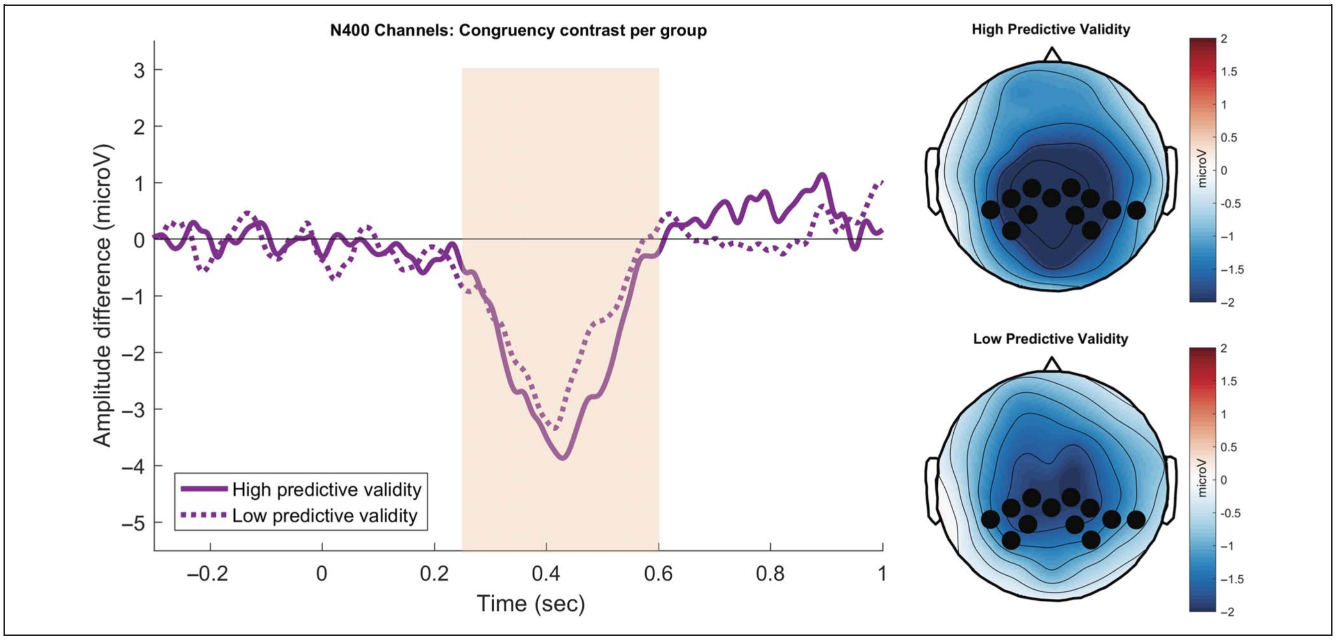

Figure 2. N400 congruency contrast effect between predictive validity groups, averaged over constraints. (Left) The N400 effect for the time window 250–600 msec for each group, as averaged over sentence context constraints. (Right) The topography of the N400 effect and the sensor selection based on the top 20% of *t* values from the cluster-based permutation statistics.

the incongruent compared with congruent the condition ($p < .001$). Although the low predictive validity group displayed overall more positive N400 amplitudes compared with the high-predictive validity group (Figure 3), the main effect of Predictive Validity did not reach significance, $F(1, 71.1) = 3.97$, $p = .050$). There was a significant Constraint × Congruency interaction, $F(2, 13125.9) = 3.47$, $p = .031$), which stemmed from a greater effect of Constraint (LC vs. HC) in the congruent relative to incongruent condition ($p = .007$). There was no significant interaction between Congruency × Predictive Validity, $F(2, 71.5) = 1.75$, $p > .10$, or Constraint × Predictive Validity, $F(2, 13125.6) = 0.43$, $p > .10$, and no significant three-way interaction between Constraint × Congruency × Predictive Validity, $F(2, 13125.9) = 0.86$, $p > .10$.

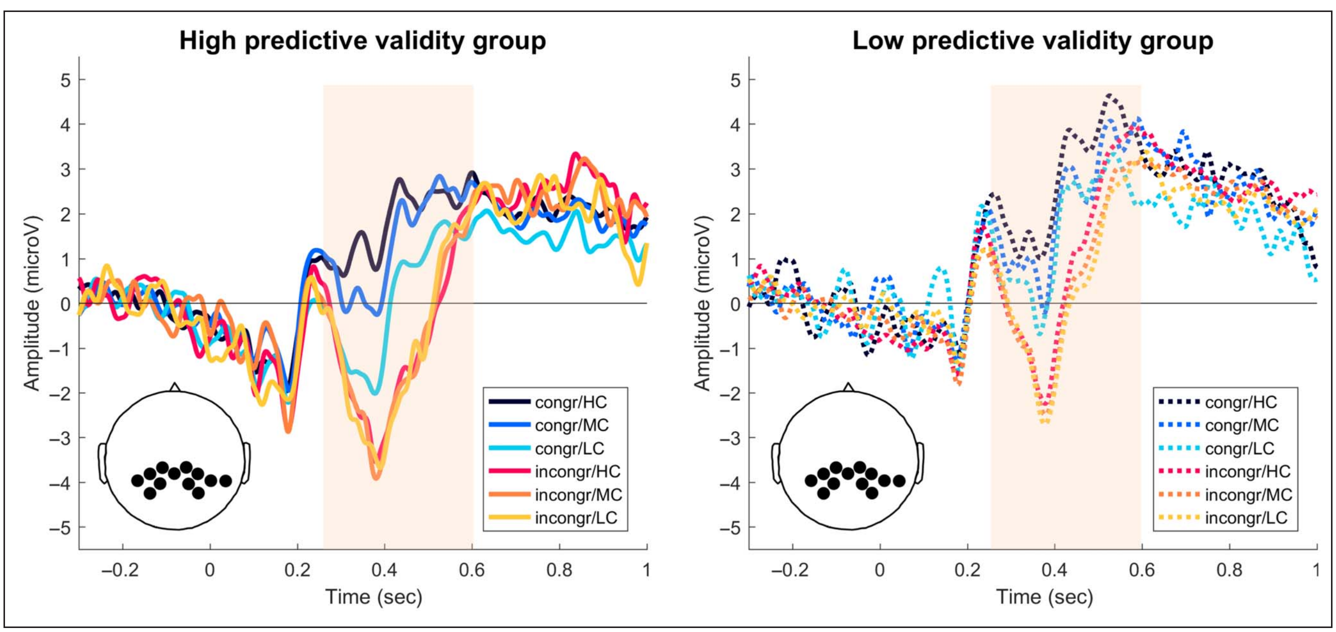

Figure 3. N400 amplitude modulations split by predictive validity. Per group, incongruent target words (red colors) resulted in a stronger N400 amplitude than congruent target words (blue colors), within a time window of 250 msec and 600 msec after target word onset. The effect of context constraints was only observed for congruent target words, for both groups. No significant effect of Predictive Validity was observed.

Table 3. Model Summary of N400 Modulation

Random Effects				
Groups	*Name*	*Var*	*SD*	*Corr*
Subject	Intercept	7.95	2.82	
Congruency[T.incongruent]		0.58	0.76	0.01
Residual		59.70	7.73	

Fixed Effects	*Est*	*SE*	*df*	*t*	*Pr(>|t|)*
Intercept	1.79	0.51	75.70	3.51	.001
Constraint[T.low]	−1.65	0.26	13104.17	−6.40	< .001
Constraint[T.med]	−0.42	0.26	13104.24	−1.62	.105
Congruency[T.incongruent]	−2.78	0.43	452.51	−6.50	< .001
Group[T.inc]	1.03	0.79	107.91	1.29	.199
Constraint[T.low]:Congruency[T.incongruent]	1.56	0.58	13122.90	2.70	.007
Constraint[T.med]:Congruency[T.incongruent]	0.41	0.57	13128.75	0.72	.470
Constraint[T.low]:Group[T.inc]	0.56	0.58	13125.30	0.95	.341
Constraint[T.med]:Group[T.inc]	−0.18	0.58	13128.44	−0.31	.757
Congruency[T.incongruent]:Group[T.inc]	0.93	0.61	454.00	1.54	.126
Constraint[T.low]:Congruency[T.incongruent]:Group[T.inc]	−1.03	0.82	13124.12	−1.26	.208
Constraint[T.med]:Congruency[T.incongruent]:Group[T.inc]	−0.26	0.82	13128.57	−0.31	.754

df = degrees of freedom; Est = beta coefficient; *SD* = standard deviation; *SE* = standard error; IC = incongruent; med = medium constraint; Var = variance.

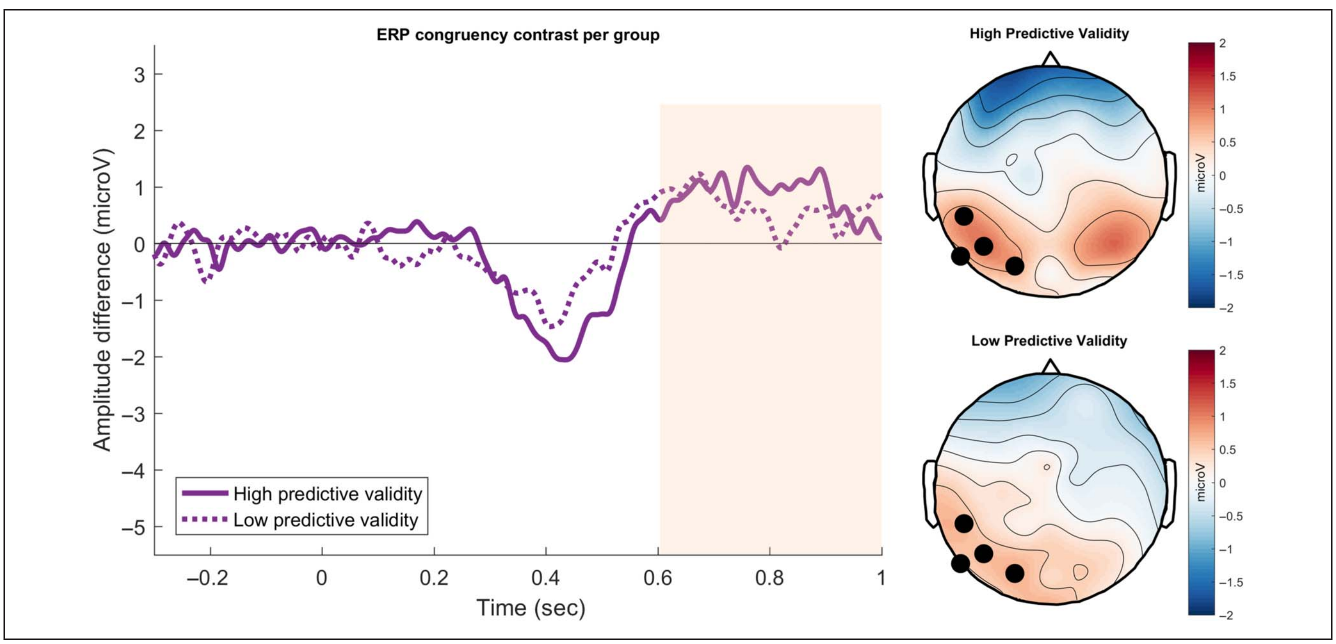

Figure 4. PNP congruency effect comparison between predictive validity groups, averaged over constraints. (Left) The PNP congruency effect for the time window 600–1000 msec as averaged over sentence context constraints. (Right) The topography of the PNP congruency effect and the sensor selection based on the corresponding cluster from the cluster-based permutation statistics.

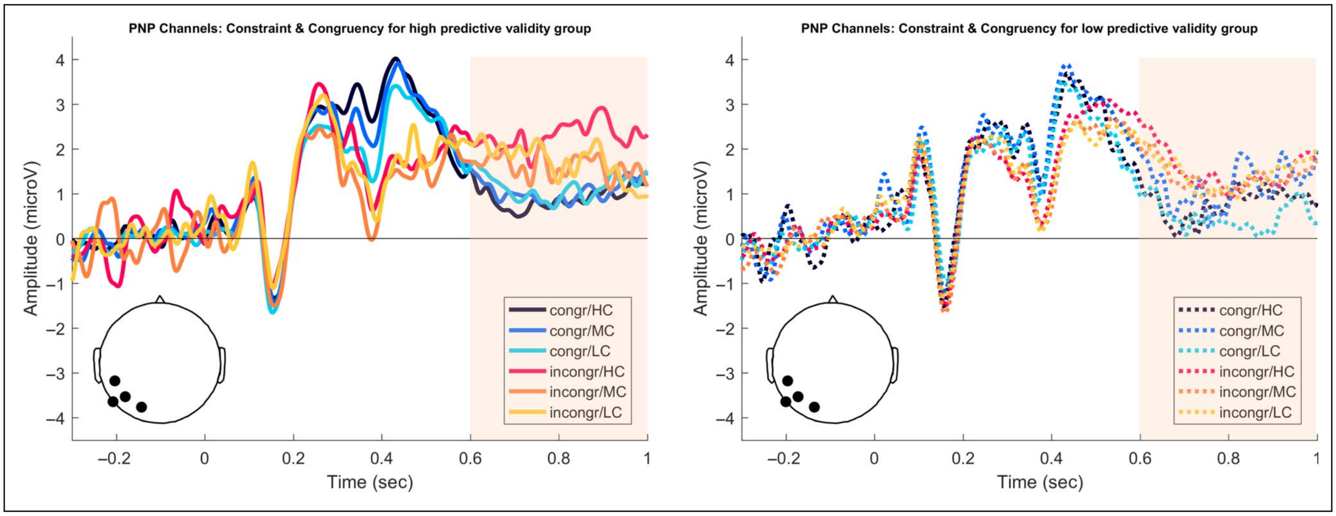

Figure 5. PNP amplitude modulations split by predictive validity. Per predictive validity group, incongruent target words (red colors) resulted in a more positive PNP amplitude than congruent target words (blue colors), within a time window of 600 and 1000 msec after target word onset. The effect of context constraints was not observed for either group. The groups did not differ with respect to constraint or congruency.

PNP Amplitude Modulation after Target Word Onset

We investigated the effect of predictive validity on the PNP amplitude 600–1000 msec relative to target word onset. Cluster permutation statistics were again used to identify which channels should be selected for analysis of the interaction effects based on the congruency contrast. Collapsed over predictive validity and constraint, the modulation of PNP amplitude by congruency in a cluster over left

Table 4. Model Summary of PNP Modulation

Random Effects					
Groups	*Name*	*Var*	*SD*	*Corr*	
Subject	Intercept	0.95	0.97		
Congruency[T.incongruent]		0.31	0.55	−0.10	
Residual		38.60	6.21		
Fixed Effects	*Est*	*SE*	*df*	*t*	*Pr(>|t|)*
Intercept	0.90	0.22	97.47	4.07	< .001
Constraint[T.low]	0.25	0.21	13095.92	1.18	.238
Constraint[T.med]	0.15	0.21	13096.09	0.71	.477
Congruency[T.incongruent]	1.34	0.34	513.62	3.94	< .001
Group[T.inc]	−0.14	0.41	270.49	−0.34	.735
Constraint[T.low]:Congruency[T.incongruent]	−0.77	0.46	13113.88	−1.66	.098
Constraint[T.med]:Congruency[T.incongruent]	−0.82	0.46	13120.19	−1.77	.077
Constraint[T.low]:Group[T.inc]	−0.48	0.47	13115.44	−1.03	.303
Constraint[T.med]:Group[T.inc]	0.27	0.47	13118.66	0.58	.565
Congruency[T.incongruent]:Group[T.inc]	−0.57	0.49	515.40	−1.17	.243
Constraint[T.low]:Congruency[T.incongruent]:Group[T.inc]	0.90	0.66	13114.80	1.37	.172
Constraint[T.med]:Congruency[T.incongruent]:Group[T.inc]	0.23	0.66	13119.65	0.35	.727

df = degrees of freedom; Est = beta coefficient; *SD* = standard deviation; *SE* = standard error; IC = incongruent; med = medium constraint; Var = variance.

posterior electrodes did not reach significance ($p = .076$; Figure 4). We still used this cluster as a ROI for the linear mixed-effects model (Figure 5). The model with by-subject random slopes for Congruency and a treatment contrast coding scheme (model summary in Table 4) revealed a significant effect of Congruency, $F(2, 77.2) = 22.91$, $p < .001$, with a higher amplitude in the incongruent relative to congruent condition. Yet, the significant effect of Congruency is inflated by the selection of sensors that are a priori responsive to the congruency effect. There was no significant main effect of Constraint, $F(2, 13100.2) = 0.03$, $p > .10$, or Predictive Validity, $F(2, 79.2) = 1.25$, $p > .10$, and no significant interaction between Constraint and Congruency, $F(1, 13116.8) = 2.29$, $p > .10$; Congruency and Predictive Validity, $F(1, 77.2) = 0.41$, $p > .10$; or Constraint and Predictive Validity, $F(1, 13116.5) = 0.99$, $p > .10$, and no significant three-way interaction between Constraint, Congruency, and Predictive Validity, $F(2, 13116.8) = 1.01$, $p > .10$.

Alpha Power Modulations before Target Word Onset

Alpha (8–12 Hz) power modulation before target word onset was investigated to study the influence of context constraint and predictive validity on brain states before the occurrence of the target word. We expected alpha power to be modulated by context constraints (Terporten

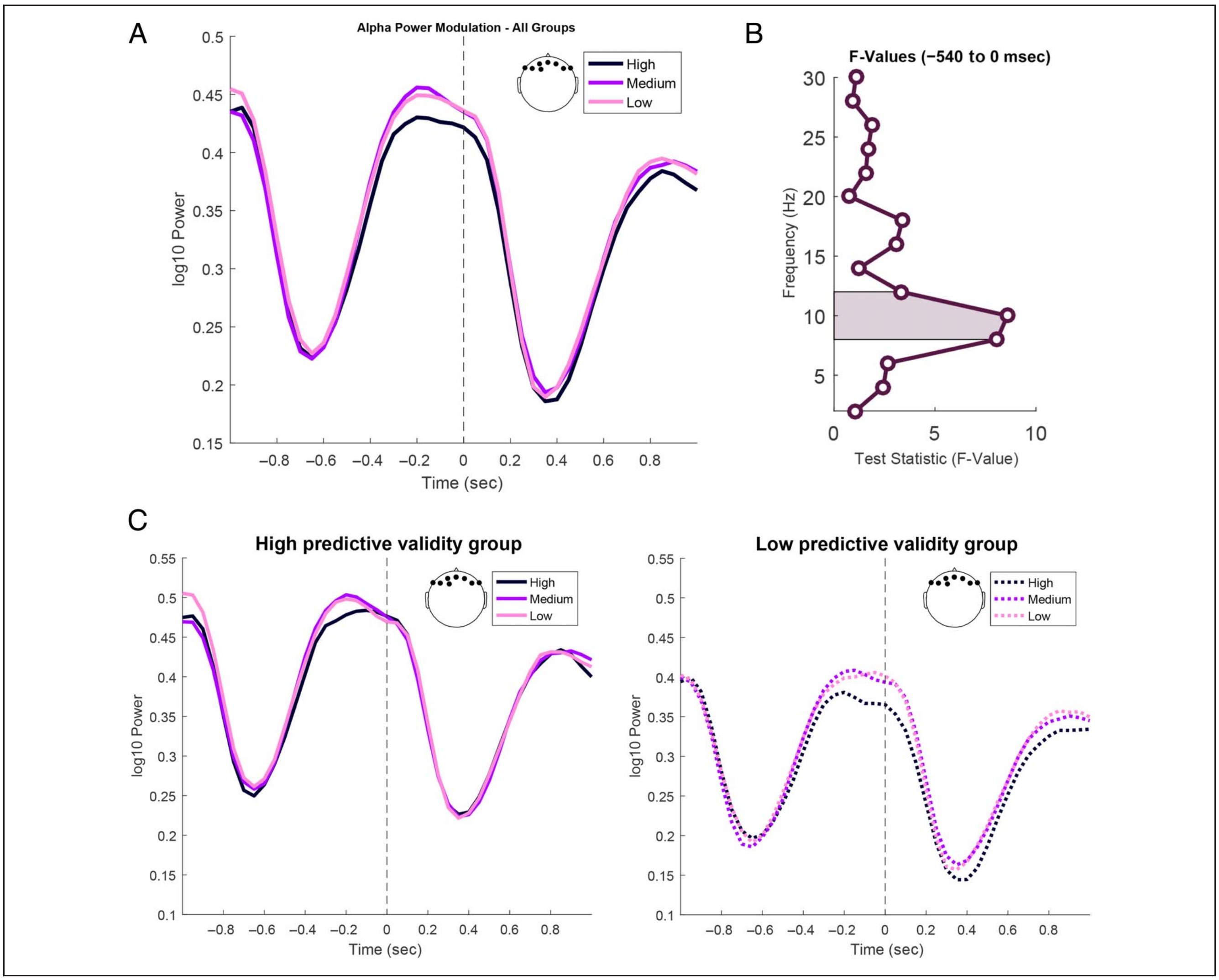

Figure 6. Alpha power modulations as a function of context constraints and power modulations across a broad frequency spectrum. (A) Alpha power modulations as a function of context constraints averaged across predictive validity groups. Pretarget word (−540 msec to 0 msec) alpha power is modulated by sentence context constraints. HC contexts induce a stronger alpha power decrease than MC or LC sentence contexts. This effect is most pronounced over frontal electrodes. (B) Power modulations across a broad frequency spectrum as a function of sentence context constraints, irrespective of predictive validity. The shaded part marks the alpha frequency band of interest (8–12 Hz). The power spectrum displays a peak in the effect of constraint around the alpha (8–12 Hz) frequency band, suggesting that alpha as compared with other frequency bands serves as a cognitive marker that is specific to variations in sentence context constraint. (C) Alpha power modulations as a function of sentence context constraints, split by predictive validity. The pretarget (−540 msec to 0 msec) word alpha power modulations suggest an interaction between Predictive Validity and Context Constraints, as the effect of constraints appears to be stronger for the incongruent group. This interaction however does not reach significance.

Table 5. Model Summary of Alpha Power Modulation

Random Effects					
Groups	*Name*	*Var*	*SD*		
Subject	Intercept	0.13	0.36		
Residual		0.07	0.27		
Fixed Effects	*Est*	*SE*	*df*	*t*	*Pr(>|t|)*
Intercept	0.43	0.06	70.85	7.12	< .001
Constraint[T.low]	0.01	0.01	13210.00	1.26	.209
Constraint[T.med]	0.01	0.01	13210.00	1.70	.089
Group[T.inc]	−0.11	0.09	70.88	−1.24	.220
Constraint[T.low]:Group[T.inc]	0.01	0.01	13210.00	0.99	.323
Constraint[T.med]:Group[T.inc]	0.01	0.01	13210.00	0.64	.523

df = degrees of freedom; Est = beta coefficient; *SD* = standard deviation; *SE* = standard error; IC = incongruent; med = medium constraint; Var = variance.

et al., 2019; Rommers et al., 2017; Wang et al., 2017; Piai, Roelofs, & Maris, 2014), but did not expect a monotonic relationship between context constraint and alpha power based on previous results (Terporten et al., 2019). We further expected to observe this effect to be strongest over frontoparietal sensors as in our previous study.

Cluster permutation analysis was first used to identify which channels should be selected for further analysis and then to identify which frequencies were most prominently modulated by constraint. A comparison of the conditions of constraint, irrespective of predictive validity, revealed a cluster over bilateral frontal electrode sites (Figure 6A) that did not reach significance (p = .085). Despite the cluster not being significant, we investigated the interaction between constraint and predictive validity within the sensors of the identified cluster. We first performed spectral power analysis within the identified cluster, ranging from 2 to 30 Hz over a time window of −540 to 0 msec before target word presentation, selected based on previous results (Terporten et al., 2019). As in our previous study, we show that the effect of context constraint was most prominently observed in the alpha range (Figure 6B), confirming that the effect of sentence context constraint was most prominently observed in the alpha frequency range relative to other frequency bands. We then performed a linear mixed-effects model analysis, with random intercepts for participants, to analyze the interaction between Constraint and Predictive Validity (model summary in Table 5). The model revealed a significant effect of constraint, $F(2, 13214) = 5.61$, $p = .004$, on alpha power, but this effect should be taken with caution as we selected a priori the sensors that were most responsive to context constraint. Alpha power was lowest for the HC, followed by the LC and MC conditions (see Figure 6). No significant effect of Predictive Validity, $F(1, 70) = 1.36$, $p > .10$, and no Constraint × Predictive Validity interaction, $F(2, 13214) = 0.50$, $p > .10$, were observed.

DISCUSSION

The current study addressed the issue of whether the processing of previous semantic context could be affected by the predictive validity of contextual information, for sentences with three degrees of contextual constraints. The validity of these predictions was manipulated group-wise, by changing the proportion of sentence final (target) words that were congruent to the previously established sentence context. Pretarget alpha oscillatory power and post-target N400/ PNP amplitude modulations were investigated as functional markers for the interaction between context constraint and predictive validity.

N400 amplitude was modulated by both the congruency of the target word with its preceding context, as well as the amount of sentential context constraints (correlated with the cloze probability of the target word). However, the N400 did not significantly differ between the predictive-validity groups. For both groups, incongruent target words resulted in a stronger (more negative) N400 amplitude than congruent target words, which is in line with classic N400 findings (Kutas & Federmeier, 2011). A graded difference in N400 amplitude as a function of sentence context constraints was only found for congruent target words, which also confirms our expectations (Kutas & Federmeier, 2011) and replicates earlier investigations of this stimulus material (Terporten et al., 2019). The effects of context constraints and target word congruency on N400 amplitude were not significantly affected by predictive validity.

The robustness of the observed N400 effects across predictive validity groups speaks against a top–down modulation of linguistic processing as a function of predictive validity and stands in contrast to previous evidence (Brothers et al., 2017, 2019; Delaney-Busch et al., 2019; Lau et al., 2012). Brothers and colleagues (2019) showed an effect of predictive validity on the processing of final words only for highly constrained sentential contexts; hence, sentences endings were either highly or poorly predictable. In contrast, we also presented sentences with medium- and low-context constraints, and orthogonalized the effect of context constraint and predictive validity. Therefore, it is possible that our manipulation of predictive validity was less strong than in Brothers and colleagues (2019), because two thirds of our congruent stimulus set had medium to low predictable sentence endings. In addition, differences in experimental design could have led to differences in how explicit the manipulation of predictive validity is to the participants. In Brothers and colleagues (2017), variations in predictive validity were done explicitly contrary to our current approach, by explicitly instructing participants to either predict or not the sentence endings during the experiment. This could potentially mean that if not otherwise explicitly instructed, participants do not automatically engage in predictive processing during sentence comprehension. The current results are additionally inconsistent with findings from previous semantic priming paradigms (Delaney-Busch et al., 2019; Lau et al., 2012), which demonstrated that the proportion of valid predictions modulated the N400 amplitude for highly predictable word pairs. We speculate that the process underlying linguistic predictions created in semantic priming paradigms, which would rely on semantic associations (Brothers et al., 2017; Boudewyn, Gordon, Long, Polse, & Swaab, 2012; Lau et al., 2012; Kuperberg, Paczynski, & Ditman, 2010), might differ from predictions processes recruited during full sentences processing.

Next to effects of semantic congruency, context constraints and predictive validity on the N400 time window, we also explored their potential effects on a later time window. PNP have also been shown to be sensitive for contextual constraints (Brothers et al., 2017; Van Petten & Luka, 2012; DeLong, Urbach, Groppe, & Kutas, 2011; Federmeier et al., 2007). The PNP could be linked to reanalysis of problematic semantic input that relates to the previous context (Van Petten & Luka, 2012; Kuperberg, 2007; Kolk, Chwilla, van Herten, & Oor, 2003). In a literature overview, Van Petten and Luka (2012) note that the PNP can be influenced by semantic congruency and constraint. Their influence however is expressed by different topographies, with semantic (in)congruencies affecting the PNP over parietal electrode sites, and semantic constraints predominantly affecting PNPs over frontal electrode sites (Thornhill & Van Petten, 2012; DeLong et al., 2011; Federmeier et al., 2007). We did not find statistical evidence for an overall effect of congruency on the PNP, irrespective of context constraints and predictive validity. Yet, we were able to identify a trend in the data, showing an unspecific late positivity over left lateralized posterior electrodes. For these posterior electrodes, we were unable to observe effects of constraint and predictive validity. A limitation in our study is that the effect of context constraints was confounded with cloze probability for both post-targets neural signatures (N400 and PNP). Knowing that N400 is only weakly sensitive to constraint (Kuperberg et al., 2020; Van Petten & Luka, 2012; Kutas & Hillyard, 1984), it is possible that the effects of constraints on the N400 are rather reflecting cloze probability effects. The interplay between context constraint and cloze probability may have also created the spatial overlap of distinct frontal and posterior PNP components making them difficult to dissociate in the current design (Brouwer & Crocker, 2017).

Cloze probability could not affect alpha power modulations as they were measured before the arrival of the cloze; therefore, alpha modulations could only be interpreted as an effect of context constraints (and predictive validity). Yet, contrary to our expectations, alpha power was not significantly modulated as a function of context constraints, nor as a function of predictive validity. The main effect of sentential context constraints was only found for frontal electrode sites, for the time window −540 to 0 msec relative to target word onset as predefined from Terporten and colleagues (2019). Within this ROI, stronger alpha power decrease was observed for HC as compared with MC or LC. Although the stronger power decrease for high constraints compared with lower constraints is in line with earlier findings (Piai et al., 2018; Wang et al., 2017), the current results do not replicate our previous work that used a fine-grained constraint modulation (Terporten et al., 2019). On the basis of Terporten and colleagues (2019), we expected the alpha power decrease to be strongest for MC, followed by the other conditions; however, we observed that alpha desynchronization was strongest for HC instead for MC. This again speaks against a direct or linear relation between pretarget alpha activity and target word predictability in line with Terporten and colleagues (2019). If the modulations in pretarget alpha power reflected processes underlying linguistic prediction, we would have further expected that alpha power would be modulated by the predictive validity of the context. This is not what we observed: Alpha power was not significantly affected by the predictive validity of sentences. The investigation of alpha oscillations was based on Terporten and colleagues (2019). We also explore the effect of context constraint on a wider frequency range, including, for example, the beta (16–20 Hz) and theta (4–8 Hz) frequency bands, as beta (Wang et al., 2017; Lam et al., 2016; Lewis & Bastiaansen, 2015) as well as the theta (Rommers et al., 2017; Molinaro, Barraza, & Carreiras, 2013) frequency bands have been linked to linguistic prediction before. Yet, as in Terporten and colleagues (2019), effects of context constraints were most prominently observed in the alpha frequency band.

In conclusion, the current study shows that predictive validity did not alter the processing of sentential context constraints and cloze probability as measured by pretarget alpha power and post-target N400/PNP amplitudes. These results do not give conclusive insight on a link between N400/PNP and alpha power and the predictability of a target word based on global contextual information.

Corresponding author: Anne Kosem, Centre for Research in Neuroscience Lyon, Lyon Neuroscience Research Center (CRNL), INSERM U1028, 95 bd Pinel, 69500, Bron, France, or via e-mail: anne.kosem@inserm.fr.

Data Availability Statement

Data and code related to this study can be requested to the Max Planck Institute repository via the following access link: https://archive.mpi.nl/mpi/islandora/search/Terporten?type=dismax.

Funding Information

Agence Nationale de la Recherche (https://dx.doi.org/10.13039/501100001665), grant number: ANR-21-CE37-0003. Nederlandse Organisatie voor Wetenschappelijk Onderzoek (https://dx.doi.org/10.13039/501100003246), grant number: Spinoza award.

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance. The authors of this paper report its proportions of citations by gender category to be: M/M = .432; W/M = .159; M/W = .159; W/W = .25.

REFERENCES

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*, 280–289. https://doi.org/10.1016/j.tics.2007.05.005, PubMed: 17548232

Bastiaansen, M., & Hagoort, P. (2015). Frequency-based segregation of syntactic and semantic unification during online sentence level language comprehension. *Journal of Cognitive Neuroscience*, *27*, 2095–2107. https://doi.org/10.1162/jocn_a_00829, PubMed: 26042498

Boudewyn, M. A., Gordon, P. C., Long, D., Polse, L., & Swaab, T. Y. (2012). Does discourse congruence influence spoken language comprehension before lexical association? Evidence from event-related potentials. *Language and Cognitive Processes*, *27*, 698–733. https://doi.org/10.1080/01690965.2011.577980, PubMed: 23002319

Brothers, T., Dave, S., Hoversten, L. J., Traxler, M. J., & Swaab, T. Y. (2019). Flexible predictions during listening comprehension: Speaker reliability affects anticipatory processes. *Neuropsychologia*, *135*, 107225. https://doi.org/10.1016/j.neuropsychologia.2019.107225, PubMed: 31605686

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2017). Goals and strategies influence lexical prediction during sentence comprehension. *Journal of Memory and Language*, *93*, 203–216. https://doi.org/10.1016/j.jml.2016.10.002

Brouwer, H., & Crocker, M. W. (2017). On the proper treatment of the N400 and P600 in language comprehension. *Frontiers in Psychology*, *8*, 1327. https://doi.org/10.3389/fpsyg.2017.01327, PubMed: 28824506

Brown, C. M., Hagoort, P., & Chwilla, D. J. (2000). An event-related brain potential analysis of visual word priming effects. *Brain and Language*, *72*, 158–190. https://doi.org/10.1006/brln.1999.2284, PubMed: 10722786

Delaney-Busch, N., Morgan, E., Lau, E., & Kuperberg, G. R. (2019). Neural evidence for Bayesian trial-by-trial adaptation on the N400 during semantic priming. *Cognition*, *187*, 10–20. https://doi.org/10.1016/j.cognition.2019.01.001, PubMed: 30797099

DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, *61*, 150–162. https://doi.org/10.1016/j.neuropsychologia.2014.06.016, PubMed: 24953958

DeLong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology*, *48*, 1203–1207. https://doi.org/10.1111/j.1469-8986.2011.01199.x, PubMed: 21457275

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*, 1117–1121. https://doi.org/10.1038/nn1504, PubMed: 16007080

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75–84. https://doi.org/10.1016/j.brainres.2006.06.101, PubMed: 16901469

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, *140*, 1–11. https://doi.org/10.1016/j.bandl.2014.10.006, PubMed: 25461915

Freunberger, D., & Roehm, D. (2017). The costs of being certain: Brain potential evidence for linguistic preactivation in sentence processing. *Psychophysiology*, *54*, 824–832. https://doi.org/10.1111/psyp.12848, PubMed: 28240780

Gerrig, R. J., & McKoon, G. (1998). The readiness is all: The functionality of memory-based text processing. *Discourse Processes*, *26*, 67–86. https://doi.org/10.1080/01638539809545039

Holcomb, P. J. (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain and Language*, *35*, 66–85. https://doi.org/10.1016/0093-934X(88)90101-0, PubMed: 3179703

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135. https://doi.org/10.1016/j.brainres.2015.02.014, PubMed: 25708148

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 Research*, *1706*, 196–208. https://doi.org/10.1016/j.brainres.2018.11.013, PubMed: 30439351

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks*, *13*, 411–430. https://doi.org/10.1016/S0893-6080(00)00026-5, PubMed: 10946390

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. https://doi.org/10.1016/j.jml.2015.10.007

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*, 643–650. https://doi.org/10.3758/BRM.42.3.643, PubMed: 20805586

Kolk, H. H. J., Chwilla, D. J., van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, *85*, 1–36. https://doi.org/10.1016/S0093-934X(02)00548-5, PubMed: 12681346

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23–49. https://doi.org/10.1016/j.brainres.2006.12.063, PubMed: 17400197

Kuperberg, G. R., Brothers, T., & Wlotko, E. W. (2020). A tale of two positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience*, *32*, 12–35. https://doi.org/10.1162/jocn_a_01465, PubMed: 31479347

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*, 32–59. https://doi.org/10.1080/23273798.2015.1102299, PubMed: 27135040

Kuperberg, G. R., Paczynski, M., & Ditman, T. (2010). Establishing causal coherence across sentences: An ERP study. *Journal of Cognitive Neuroscience*, *23*, 1230–1246. https://doi.org/10.1162/jocn.2010.21452, PubMed: 20175676

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. https://doi.org/10.1146/annurev.psych.093008.131123, PubMed: 20809790

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161–163. https://doi.org/10.1038/307161a0, PubMed: 6690995

Lam, N. H. L., Schoffelen, J.-M., Uddén, J., Hultén, A., & Hagoort, P. (2016). Neural activity during sentence processing as reflected in theta, alpha, beta, and gamma oscillations. *Neuroimage*, *142*, 43–54. https://doi.org/10.1016/j.neuroimage.2016.03.007, PubMed: 26970187

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2012). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience*, *25*, 484–502. https://doi.org/10.1162/jocn_a_00328, PubMed: 23163410

Lewis, A. G., & Bastiaansen, M. (2015). A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex*, *68*, 155–168. https://doi.org/10.1016/j.cortex.2015.02.014, PubMed: 25840879

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024, PubMed: 17517438

Molinaro, N., Barraza, P., & Carreiras, M. (2013). Long-range neural synchronization supports fast and efficient reading: EEG correlates of processing expected words in sentences. *Neuroimage*, *72*, 120–132. https://doi.org/10.1016/j.neuroimage.2013.01.031, PubMed: 23357072

Nieuwland, M., Barr, D., Bartolozzi, F., Busch-Moreno, S., Donaldson, D., Ferguson, H. J., et al. (2020). Dissociable effects of prediction and integration during language comprehension: Evidence from a large-scale study using brain potentials. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *375*, 20180522. https://doi.org/10.1098/rstb.2018.0522, PubMed: 31840593

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869. https://doi.org/10.1155/2011/156869, PubMed: 21253357

Piai, V., Roelofs, A., & Maris, E. (2014). Oscillatory brain responses in spoken word production reflect lexical frequency and sentential constraint. *Neuropsychologia*, *53*, 146–156. https://doi.org/10.1016/j.neuropsychologia.2013.11.014, PubMed: 24291513

Piai, V., Rommers, J., & Knight, R. T. (2018). Lesion evidence for a critical role of left posterior but not frontal areas in alpha–beta power decreases during context-driven word production. *European Journal of Neuroscience*, *48*, 2622–2629. https://doi.org/10.1111/ejn.13695, PubMed: 28887896

R Core Team. (2019). *R: A language and environment for statistical computing*. https://www.R-project.org/

Rommers, J., Dickson, D. S., Norton, J. J. S., Wlotko, E. W., & Federmeier, K. D. (2017). Alpha and theta band dynamics related to sentential constraint and word expectancy. *Language, Cognition and Neuroscience*, *32*, 576–589. https://doi.org/10.1080/23273798.2016.1183799, PubMed: 28761896

Terporten, R., Schoffelen, J.-M., Dai, B., Hagoort, P., & Kösem, A. (2019). The relation between alpha/beta oscillations and the encoding of sentence induced contextual information. *Scientific Reports*, *9*, 1–12. https://doi.org/10.1038/s41598-019-56600-x, PubMed: 31882830

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, *83*, 382–392. https://doi.org/10.1016/j.ijpsycho.2011.12.007, PubMed: 22226800

Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, *5*, 99–114. https://doi.org/10.2307/3001913, PubMed: 18151955

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*, 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015, PubMed: 22019481

Wang, L., Hagoort, P., & Jensen, O. (2017). Language prediction is reflected by coupling between frontal gamma and posterior alpha oscillations. *Journal of Cognitive Neuroscience*, *30*, 432–447. https://doi.org/10.1162/jocn_a_01190, PubMed: 28949823

Willems, R. M., Oostenveld, R., & Hagoort, P. (2008). Early decreases in alpha and gamma band power distinguish linguistic from visual information during spoken sentence comprehension. *Brain Research*, *1219*, 78–90. https://doi.org/10.1016/j.brainres.2008.04.065, PubMed: 18538306