Brain Research 1788 (2022) 147942

Contents lists available at ScienceDirect

Brain Research

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

Benefits and costs of predictive processing: How sentential constraint and word expectedness affect memory formation

Gerrit Höltje *, Axel Mecklinger

Experimental Neuropsychology Unit, Department of Psychology, Saarland University, Saarbrücken, Germany

ARTICLE INFO

Keywords:
Event-related potential
Subsequent memory effect
Recognition memory
Predictive processing
Schema
False memory

ABSTRACT

This study investigated how the strength of schema support provided by strongly (SC) and weakly constraining (WC) sentences affects the encoding of expected and unexpected words, and how this is reflected in event-related potentials (ERPs). In a surprise recognition memory test, words studied on the previous day were presented together with new words and lures that were expected but not presented in the study phase. ERPs recorded in the study phase were compared for subsequently remembered and forgotten words. Better memory performance for expected over unexpected words was electrophysiologically supported by a parietal subsequent memory effect (SME) reflecting enhanced item-specific encoding of contextually expected words. SC sentences not only facilitated the semantic integration of sentence-ending words, as reflected in reduced N400 amplitudes, but also enabled the rapid successful encoding of these words into memory, which is evidenced by an SC > WC pattern in memory performance and correlations between pre- and post-stimulus SMEs for SC sentences. In contrast, words processed in WC sentence contexts necessitated sustained elaborative encoding processes as reflected in a late frontal slow wave SME. Expected but not presented words were associated with high rates of false positive memory decisions, indicating that these words remained in a state of high accessibility in memory even one day after the study phase. These mnemonic costs of predictive processing were more pronounced for expected words from SC sentences than from WC sentences and could reflect the lingering of strong semantic predictions which were associated with the pre-updating of sentence representations.

1. Introduction

Learning is most effective when new information can be related to a schema, an associative network structure extracted over multiple experiences (Alba and Hasher, 1983; Bartlett, 1932; Bransford and Johnson, 1972; Hebscher et al., 2019). Schemas can be activated by contextual information and allow the prediction of future events that have previously been associated with similar contexts (Ghosh and Gilboa, 2014). Activated schemas facilitate the encoding of congruent or expected events and enable the formation of elaborated memory representation which are easily accessible at retrieval (Craik and Tulving, 1975; Greve et al., 2019; Staresina et al., 2009). For example, the sentence "She went to the bathroom and cleaned her teeth with" leads to the pre-activation of the sentence-ending word "toothbrush". The schema account predicts that the contextually expected word "toothbrush" is associated with better memory than an incongruent word like "screwdriver" because the semantic elaboration of the expected word "toothbrush" is supported by the schema activated by the sentence context, whereas the incongruent word "screwdriver" is not. However, the word "dental floss" is less expected than "toothbrush" without being incongruent within the sentence context. Words like "dental floss" violate schema-based predictions and elicit expectancy mismatches, the processing of which entails processes of schema accommodation and assimilation (Ghosh and Gilboa, 2014; Gilboa and Marlatte, 2017; Piaget, 1952). Expectancy mismatch-related processing is aimed at reducing future prediction errors and should enhance memory for the eliciting event (Friston, 2010; Greve et al., 2017; Henson and Gagnepain, 2010). Evidence in support of this view comes from research showing that unexpected feedback correcting an erroneous response made with high confidence captures attention and is associated with better memory for the correct response as compared with lower confidence errors (Butterfield and Metcalfe, 2006).

The present study investigated how the strength of schema support provided by a sentence context modulates the encoding of expected words confirming predictions and unexpected words eliciting expectancy mismatches, and how this is reflected in event-related potentials

* Corresponding author.
E-mail address: gerrit.hoeltje@googlemail.com (G. Höltje).

https://doi.org/10.1016/j.brainres.2022.147942
Received 28 June 2021; Received in revised form 7 January 2022; Accepted 6 May 2022
Available online 10 May 2022
0006-8993/

(ERPs). While a considerable number of studies have explored the brain regions mediating schema-based learning (for a review, see Gilboa and Marlatte, 2017), not much is known about the temporal dynamics of the processes involved in schema-based memory encoding. In an illustrative study (Höltje et al., 2019), we recently made use of the high temporal resolution of ERPs and employed a subsequent memory approach in which neural activity recorded during the encoding of events is compared for subsequently remembered and forgotten events (for reviews, see Cohen et al., 2015, or Paller and Wagner, 2002). We found a parietally-distributed subsequent memory effect (SME) starting around 300 ms after the onset of words that were congruent with a given categorical context (e.g., the words "steel" or "zinc" in a context like "a metal", but not for words incongruent in this context (e.g., "bear"). This suggests that with a semantically congruent context successful memory encoding can start as early as 300 ms after the onset of the critical words.

Sentence contexts which provide a wealth of preexisting associative connections enable strong schema-based predictions regarding the sentence-final word and thereby facilitate the processing of expected words. For example, a strongly constraining context like "He locked the door with the " could lead to an increased activation of some type of information for words like "key" which match with those pre-activated by this context (Piai et al., 2016). Word expectedness can be determined using a cloze procedure in which participants are asked to complete a sentence frame with the word they find most fitting. The frequency with which a word is used to complete a sentence context is its cloze probability in that context (Taylor, 1953). In ERP studies, the facilitated processing of contextually predicted words is reflected in N400 amplitudes which are monotonically graded with cloze probabilities and larger (more negative) for unexpected words than for expected words (Kutas and Federmeier, 2011). To address the key question of the present study, namely how the strength of the schema activated by a sentence context affects the encoding of expected and unexpected sentence-ending words, we derived sentence contexts that were either strongly (SC: "In this heat the flower urgently needs more…") or weakly constraining (WC: "Before turning in his bachelor's thesis, Luke makes an appointment with his…") regarding the sentence-ending word from a separate cloze norming study.

In the learning phase of the experiment, participants were presented the aforementioned sentences completed either by highly expected (SC: "water"; WC: "professor") or by unexpected but congruent words (SC: "protection"; WC: "advisor"). One day later, participants returned to the lab for a surprise recognition memory test in which they were asked to discriminate between target words that had been presented as sentence-ending words in the learning phase and unrelated new words. The EEG was recorded during the presentation of the sentences and words in the learning phase and compared for subsequently remembered and forgotten items. This experimental design enabled us to investigate the mnemonic consequences and SMEs associated with confirmed predictions (expected words) and expectancy mismatches (unexpected words). The schema account predicts that expected words should benefit most from the schema support provided by a sentence context, in particular for highly predictive SC sentences, and that these words should therefore be remembered better than less expected words completing WC sentences and unexpected words. On the contrary, a recent study by Rommers and Federmeier (2018b) suggests that high word predictability can induce a top-down verification mode of word processing and as a consequence little attention is paid to expected words, resulting in shallow processing and encoding of these words. Thus, it is possible that memory for expected words is diminished due to the shallow encoding of these words in the learning phase (Craik and Lockhart, 1972; Craik and Tulving, 1975). On the other hand, unexpected words could elicit prediction errors which capture attention and thereby support memory encoding for these words. If large prediction errors boost declarative learning (Greve et al., 2017; Henson and Gagnepain, 2010), then unexpected words should be associated with superior memory. This prediction is consistent with previous studies in which words eliciting prediction errors were associated with superior memory (Corley et al., 2007; Federmeier et al., 2007; Haeuser and Kray, 2021). Notably, in a recent study by Hubbard et al. (2019), neither sentential constraint (strong or weak) nor word expectedness affected recognition memory for sentence-ending words. It is conceivable that in the Hubbard et al. (2019) study the processing of predictable words induced a top-down verification mode which was associated with shallow processing of expected words and, as a consequence, poor memory performance for these words. In the present study, a one second delay between the presentation of the sentence context and the sentence-ending target word was introduced in order to preclude the induction of a top-down verification mode and to boost processes related to the prediction of the upcoming target word. ERP subsequent memory effects elicited by expected and unexpected words completing sentences that provided either strong or weak contextual support were compared to examine encoding mechanisms associated with these words. If, as predicted by the schema account, predictive sentence contexts support the encoding of expected words, then the encoding of these words should at the electrophysiological level be supported by a parietal SME that has previously been shown to support the encoding of schema-congruent words (Höltje et al., 2019). Alternatively, if unexpected words elicit expectancy mismatches which boost memory encoding, then these words should be associated with larger SMEs than expected ones.

By definition, predictions of upcoming words should emerge and be detectable at the neurophysiological level before predictable words are encountered. In support of this view, recent ERP studies on language comprehension found a frontally-distributed sustained negative potential preceding the onset of sentence-ending words that was more pronounced for strongly than for weakly constraining sentences. It has been proposed that this component reflects processes involved in the generation of semantic predictions and can serve as a neurophysiological index of meaning expectancy (Grisoni et al., 2017; León-Cabrera et al., 2017, 2019). Similar sustained anterior negativities elicited during the processing of sentences with long-distance dependencies between elements have recently been proposed to reflect the maintenance of discourse information in working memory (Cruz Heredia et al., 2021). We therefore extended the ERP analysis to the time interval preceding the onset of sentence-ending words and also explored whether neural activity preceding the onset of target words in the learning phase predicted subsequent memory in a similar way as neural activity following the target word.

Besides the exploration of schema effects on memory for sentence-ending words, a second aim of the present study was to investigate the fate of expected but never actually seen words in memory. Unexpected words disconfirming a strong semantic prediction induced by a sentence context elicit a late frontal positivity (LFP) which could reflect the suppression of the previously predicted word (Federmeier et al., 2007; Höltje et al., 2019; Ness and Meltzer-Asscher, 2018a). However, recent studies indicate that previously expected words even though they were not presented can remain in a state of increased accessibility in memory and give raise to false memory decisions (Hubbard et al., 2019; Rich and Harris, 2021; Rommers and Federmeier, 2018a). In the study by Hubbard et al. (2019), participants read strongly and weakly constraining sentences with expected or unexpected sentence-ending words. Subsequently, recognition memory was probed for studied sentence-ending words, words that were expected but not presented (expected lures), and new words. Expected lures, namely words which were highly predictable but not presented as the sentence-ending word, were associated with more false positive memory decisions than new words. The results of the Hubbard et al. (2019) study suggest that predictive processing initiated by a single sentence context can create mnemonic costs when predicted words fail to arrive. The facilitated processing of predicted words could be due to a pre-activation of these words in memory. If predicted but not presented words remain in a state of increased activation, then these words could be associated with increased processing fluency and false positive memory decisions at retrieval. The present study aimed to extend the findings obtained by Hubbard et al. (2019) by testing whether false memories for expected but not presented words are modulated by context strength. If stronger schema contexts allow a stronger pre-activation of the predicted words, then expected but not presented words from SC sentence contexts should be falsely recognized more often than expected words from WC sentence contexts.

2. Results

2.1. Behavioral results

As evidenced by a high proportion of correct responses to comprehension questions in the study phase (M = 0.93, SEM = 0.01), participants complied with the instructions and paid attention to the content of the sentences. In the test phase, Pr scores (M = 0.20, SEM = 0.01) were significantly larger than zero, $t(34)$ = 17.82, p <.001, d = 3.01, indicating that participants were well able to discriminate between studied target words and lures. Mean hit rates and false alarm rates in each condition are given in Table 1. As confirmed by t-tests, mean hit rates were above chance level (i.e., > 0.50) in the SC-EXP, SC-UNEXP, and WC-EXP conditions (all p-values < 0.01), but not in the WC-UNEXP condition, $t(34)$ = 0.92, p =.37.

Hit rates were submitted to a two (Constraint: Strong, weak) by two (Expectedness: Expected, Unexpected) by-participant ANOVA that yielded a significant main effect of Constraint, $F(1,34) = 7.53$, p <.05, η_p^2 = 0.18, reflecting better memory for target words following strongly constraining sentence frames (M = 0.59, SEM = 0.02) as compared to words completing weakly constraining sentence frames (M = 0.55, SEM = 0.02). Furthermore, a significant main effect of Expectedness was obtained, $F(1,34) = 31.84$, p <.001, η_p^2 = 0.48, reflecting better memory for expected words (M = 0.60, SEM = 0.02) than for unexpected ones (M = 0.54, SEM = 0.02). The Constraint by Expectedness interaction did not reach significance, $F(1,34)$ = 1.77, p =.19, η_p^2 = 0.05.

False alarm rates for words that were predicted but not actually seen during the study phase (SC and WC lures) and for new words were analyzed in a one-way ANOVA including the factor Item Status. The effect of Item Status was significant, $F(2,68) = 67.32$, p <.001, η_p^2 = 0.66. Subsidiary t-tests revealed that expected lures were associated with more false positive memory decisions than new words (SC lures: M = 0.48, SEM = 0.02; WC lures: M = 0.44, SEM = 0.02; New lures: M = 0.29, SEM = 0.02; SC vs. New: $t(34)$ = 10.50, p <.001, d = 1.48; WC vs. new: $t(34)$ = 8.58, p <.001, d = 1.17), and that SC lures were associated with higher false alarm rates than WC lures, $t(34)$ = 2.41, p <.05, d = 0.28.

2.2. ERP results

A summary of the results obtained in the analysis of ERP mean amplitudes in each time window is provided in Table 2.

2.2.1. Pre-stimulus ERPs

As evident from pre-stimulus ERP waveforms time-locked to the offset of sentence contexts in the study phase, shown in Fig. 1a, SC and WC sentences elicited a negative slow wave preceding the onset of target words. ERP mean amplitudes measured at frontal and frontocentral electrodes were analyzed in a series of ANOVAs, each contrasting mean amplitudes in two adjacent time windows, to test whether effects of Constraint and Memory differed between the four consecutive time windows (−800 to −600 ms, −600 to −400 ms, −400 to −200 ms, −200 ms to target word onset). In case no significant interaction involving Time window was found, mean amplitudes were averaged across the two time windows included in the ANOVA and subsequently contrasted with activity in the next adjacent time window (for a similar approach, see Bridger et al., 2014, or Höltje et al., 2019). Following this procedure, mean amplitudes were collapsed across the first three time windows (no significant interactions involving Time window, all p-values >0.20). Comparing mean amplitudes in the first three time windows (−800 to −200 ms) with those in the fourth and final stimulus-preceding time window (−200 to word onset) in a three-way repeated-measures ANOVA including the factors Constraint, Memory, and Time

Table 1
Mean proportions and standard deviations of "old" responses to targets (hit rates) and lures (false alarm rates) in the memory test.

Condition	"Old" responses
SC-EXP target	0.60 (0.12)
SC-UNEXP target	0.56 (0.12)
WC-EXP target	0.59 (0.12)
WC-UNEXP target	0.52 (0.13)
SC lure	0.48 (0.14)
WC lure	0.44 (0.14)
New lure	0.29 (0.11)

Table 2
Significant effects obtained in the analysis of ERP mean amplitudes in each time window.

Time window	Significant effects
−200 to −800 ms	**2(Constraint: SC, WC) × 2 (Memory: Hits, Misses) ANOVA:**ME Memory (Hits < Misses)
180–250 ms	**2(Constraint: SC, WC) × 2(Expectedness: EXP, UNEXP) × 2 (Memory: Hits, Misses) ANOVA:**ME Memory (Hits > Misses), ME Constraint, IE Constraint × Expectedness (SC-EXP > SC-UNEXP, WC-EXP = WC-UNEXP)
300–500 ms	**2(Antpos: Anterior, Posterior) × 2(Constraint: SC, WC) × 2 (Expectedness: EXP, UNEXP) × 2 (Memory: Hits, Misses) ANOVA:** ME Constraint, ME Expectedness, ME Memory, IE Antpos × Memory, IE Expectedness × Memory, IE Antpos × Expectedness × Memory **Anterior 2(Constraint: SC, WC) × 2(Expectedness: EXP, UNEXP) × 2 (Memory: Hits, Misses) ANOVA:**ME Constraint (SC > WC), ME Memory (Hits > Misses) **Posterior 2(Constraint: SC, WC) × 2(Expectedness: EXP, UNEXP) × 2 (Memory: Hits, Misses) ANOVA:**ME Constraint (SC > WC), ME Expectedness, ME Memory, IE Expectedness × Memory (SME EXP > UNEXP > 0)
500–700 ms	**2(Antpos: Anterior, Posterior) × 2(Constraint: SC, WC) × 2 (Expectedness: EXP, UNEXP) × 2 (Memory: Hits, Misses) ANOVA:** ME Memory, IE Constraint × Expectedness, IE Antpos × Expectedness × Memory **Anterior 2(Constraint: SC, WC) × 2(Expectedness: EXP, UNEXP) × 2 (Memory: Hits, Misses) ANOVA:**ME Memory (Hits > Misses) **Posterior 2(Constraint: SC, WC) × 2(Expectedness: EXP, UNEXP) × 2 (Memory: Hits, Misses) ANOVA:**ME Memory (Hits > Misses), IE Constraint × Expectedness (WC-EXP > WC-UNEXP, SC-EXP = SC-UNEXP)
700–1200 ms	**2(Antpos: Anterior, Posterior) × 2(Constraint: SC, WC) × 2 (Expectedness: EXP, UNEXP) × 2 (Memory: Hits, Misses) ANOVA:** ME Memory, ME Expectedness, IE Antpos × Expectedness, IE Constraint × Memory, IE Antpos × Constraint × Memory **Anterior 2(Constraint: SC, WC) × 2(Expectedness: EXP, UNEXP) × 2 (Memory: Hits, Misses) ANOVA:** ME Expectedness, ME Memory, IE Constraint × Expectedness (SC-UNEXP > SC-EXP, WC-UNEXP = WC-EXP), IE Constraint × Memory (SCM WC > SC = 0) **Posterior 2(Constraint: SC, WC) × 2(Expectedness: EXP, UNEXP) × 2 (Memory: Hits, Misses) ANOVA:**ME Memory (Hits > Misses)

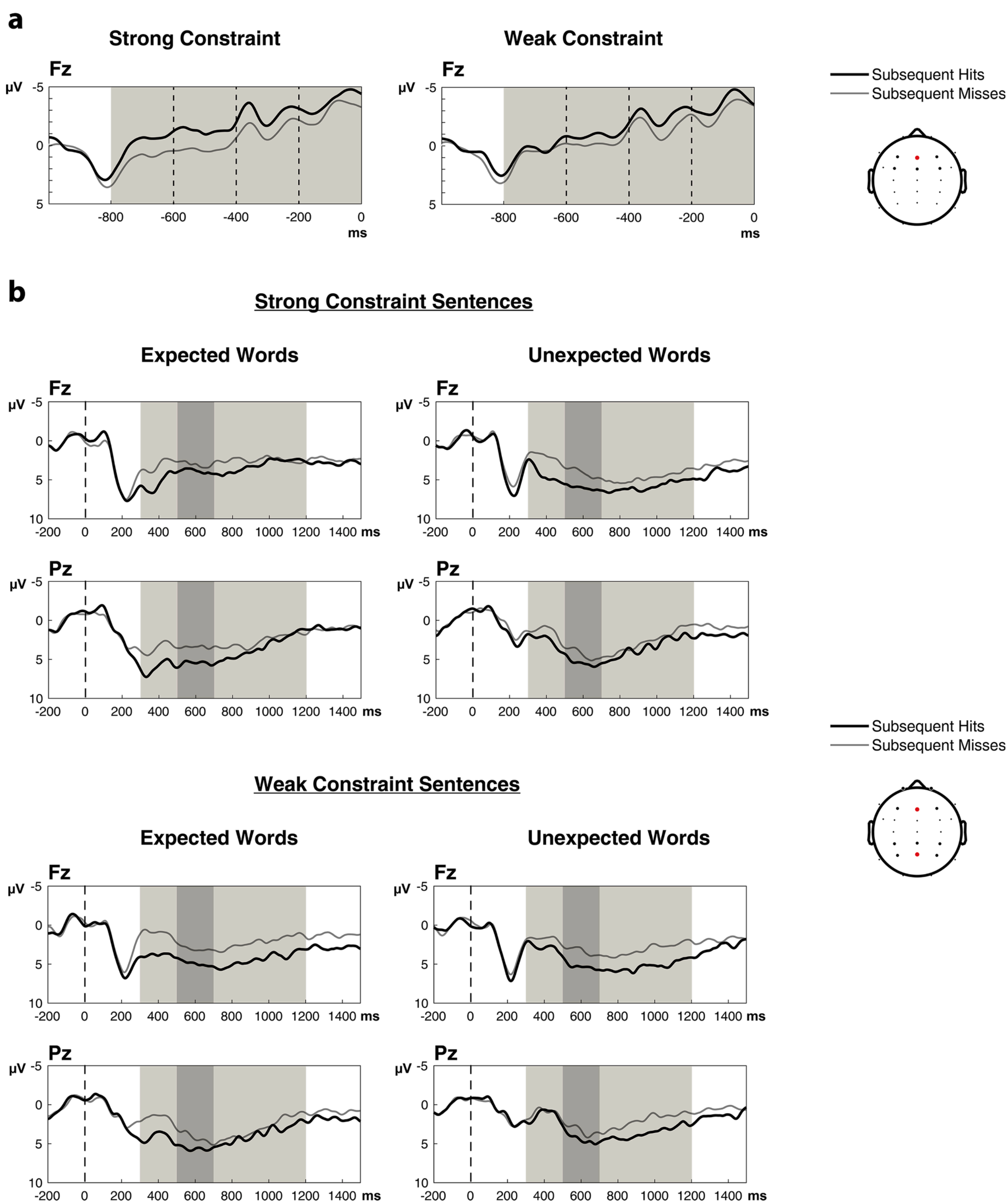

Fig. 1. ERP waveforms elicited at representative scalp electrodes by the onset of target words in the study phase. Shaded areas indicate the 300–500 ms, 500–700 ms, and 700–1200 ms time windows in which mean amplitudes were analyzed.

Window yielded a significant Time Window by Memory interaction, $F(1,28) = 4.55$, $p < .05$, $\eta_p^2 = 0.14$. Consequently, mean amplitudes in the −800 to −200 ms and −200 to word onset time windows were analyzed in two separate ANOVAs including the factors Constraint and Memory. Mean amplitudes in the −800 to −200 ms time window did not differ as a function of Constraint, $F < 1$, but were associated with a significant main effect of Memory, $F(1,28) = 15.27$, $p < .001$, $\eta_p^2 = 0.36$, reflecting more negative mean amplitudes for hits ($M = -1.48$, $SEM = 0.31$ μV) than for misses ($M = -0.55$, $SEM = 0.25$ μV). No Constraint by Memory interaction was obtained, $F < 1$. In the analysis of mean amplitudes during the 200 ms preceding target word onset, the main effect of Memory did not reach significance, and neither did the main effect of Constraint or the Constraint by Memory interaction, Fs < 1.

2.2.2. *Post-stimulus ERPs*

ERPs elicited by subsequently remembered and forgotten words are depicted in Fig. 1b. Fig. 2 shows the waveforms as a function of Constraint and Expectedness. Mean amplitudes in three consecutive time windows (300–500 ms, 500–700 ms, and 700–1200 ms) were analyzed in three separate two (Antpos: Anterior, posterior) by two (Constraint: SC, WC) by two (Expectedness: EXP, UNEXP) by two (Memory: Hits, Misses) ANOVAs. Only effects involving one of the three experimental conditions are reported.

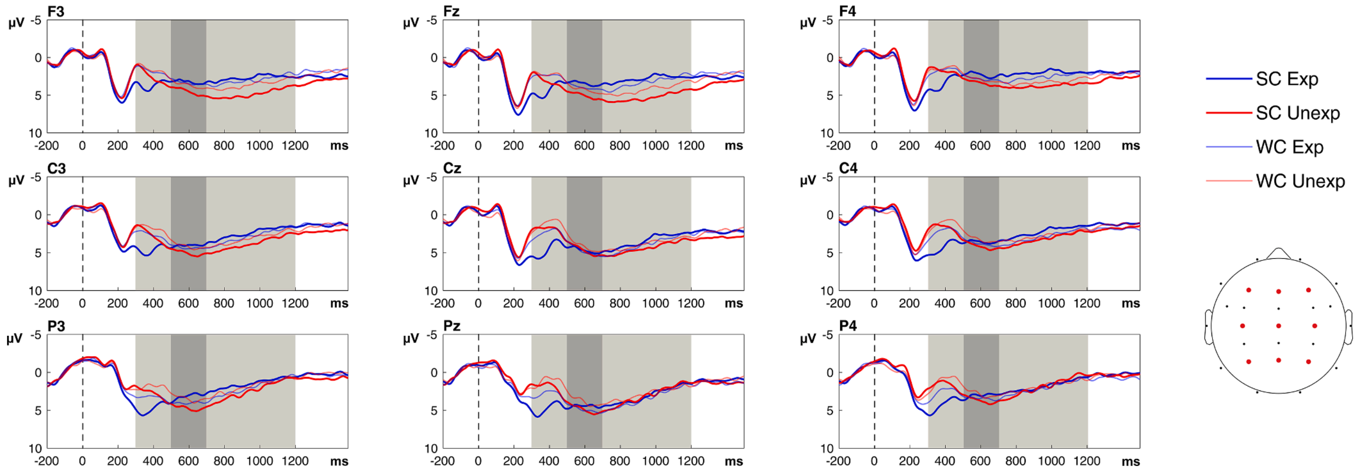

Fig. 2. Pre-stimulus subsequent memory effects (a) at an electrode representative for the anterior electrode cluster (Fz), separate for strong and weak constraint sentences. Shaded areas indicate the four consecutive time windows of 200 ms length each in which mean amplitudes were analyzed. Post-stimulus SMEs (b) at two electrodes representative for the anterior and posterior electrode clusters (Fz and Pz), separate for expected and unexpected words completing strong and weak constraint sentences. Shaded areas indicate the 300–500 ms, 500–700 ms, and 700–1200 ms time windows in which mean amplitudes were analyzed.

2.2.2.1. 300–500 ms time window. The analysis of mean amplitudes in the 300–500 ms time window, depicted in Fig. 1b and Fig. 2, yielded significant main effects of Constraint, $F(1,28) = 16.92$, $p < .001$, $\eta_p^2 = 0.38$, Expectedness, $F(1,28) = 18.61$, $p < .001$, $\eta_p^2 = 0.40$, and Memory, $F(1,28) = 34.99$, $p < .001$, $\eta_p^2 = 0.56$, qualified by significant interactions between Antpos and Expectedness, $F(1,28) = 20.95$, $p < .001$, $\eta_p^2 = 0.43$, Expectedness and Memory, $F(1,28) = 4.53$, $p < .05$, $\eta_p^2 = 0.14$, and a triple interaction between Antpos, Expectedness, and Memory, $F(1,28) = 4.95$, $p < .05$, $\eta_p^2 = 0.15$. To follow up the significant interactions involving Antpos, mean amplitudes in the anterior and posterior electrode clusters were analyzed in two separate ANOVAs including the factors Constraint, Expectedness, and Memory.

At anterior electrodes, mean amplitudes differed as a function of Constraint, $F(1,28) = 10.09$, $p < .01$, $\eta_p^2 = 0.27$, reflecting more positive amplitudes elicited by SC sentences ($M = 2.81$, $SEM = 0.59$ µV) than by WC ones ($M = 1.71$, $SEM = 0.48$ µV), and Memory, $F(1,28) = 34.42$, $p < .001$, $\eta_p^2 = 0.55$, indicating more positive amplitudes for hits ($M = 3.13$, $SEM = 0.54$ µV) than for misses ($M = 1.41$, $SEM = 0.52$ µV). The main effect of Expectedness did not reach significance, $F(1,28) = 4.00$, $p = .06$, $\eta_p^2 = 0.13$, and neither did the Constraint by Expectedness interaction, $F(1,28) = 2.66$, $p = .11$, $\eta_p^2 = 0.09$, or any other effect, all *p*-values > 0.22.

At posterior electrodes, as for the anterior electrodes, a main effect of Constraint was obtained, $F(1,28) = 20.61$, $p < .001$, $\eta_p^2 = 0.42$, indicating more positive amplitudes elicited by SC sentences ($M = 3.41$, $SEM = 0.60$ µV) than by WC ones ($M = 2.14$, $SEM = 0.57$ µV). Main effects of Expectedness, $F(1,28) = 37.12$, $p < .001$, $\eta_p^2 = 0.57$, and Memory, $F(1,28) = 27.32$, $p < .001$, $\eta_p^2 = 0.49$, were qualified by a significant interaction between these two factors, $F(1,28) = 7.03$, $p < .001$, $\eta_p^2 = 0.20$. To disentangle the Memory by Expectedness interaction, subsequent memory effects (SMEs) were calculated as the difference between subsequent hits and misses separately for expected and unexpected words. These effects differed from zero (EXP: $M = 2.42$, $SEM = 0.47$ µV, $t(28) = 5.18$, $p < .001$, $d = 0.96$; UNEXP: $M = 0.94$, $SEM = 0.38$ µV, $t(28) = 2.49$, $p < .05$, $d = 0.46$), and a paired *t*-test revealed that expected words were associated with a larger SME than unexpected ones, $t(28) = 2.65$, $p < .05$, $d = 0.64$.

To summarize, between 300 and 500 ms mean amplitudes at posterior electrodes were modulated by Expectedness, as it is typically found for the N400 (Kutas and Federmeier, 2011), and Constraint. Even though N400 effects and simultaneously occurring SMEs can be functionally dissociated because N400 amplitudes were larger (i.e., more negative) for words completing WC sentences than for words completing SC sentences, whereas SMEs did not differ between SC and WC sentences but were modulated by Expectedness (i.e., larger SMEs for expected than for unexpected words) it is possible that an interaction between the two effects emerged at the between-subjects level. For example, individuals showing strong facilitated semantic processing as reflected in a large N400 expectancy effect could also show strong SMEs reflecting contextually supported memory encoding, and vice versa. If facilitated semantic processing supports item-specific memory encoding, then N400 expectancy effects and parietal SMEs should correlate positively (i.e., larger N400 effects should be associated with larger SMEs). In total, four correlations were computed, namely between global N400 effects (i.e., the difference in amplitudes between expected and unexpected words, collapsed across Constraint and Memory) and SMEs in each of the four experimental conditions at posterior electrodes. One dataset had to be excluded from this analysis due to the amplitude of the SME elicited by SC-UNEXP words which deviated by more than three standard deviations from the sample mean. As evident from Fig. 3 (left side), N400 effects and SMEs elicited by WC-UNEXP words were negatively correlated, $r(28) = -0.52$, $p < .01$. WC-EXP words were associated with a positive and smaller correlation that did not exceed the Bonferroni-corrected threshold of significance, $r(28) = 0.40$, $p = .04$. Words completing SC sentences were not associated with significant correlations, SC-EXP: $r(28) = 0.15$, $p = .46$; SC-UNEXP: $r(28) = -0.32$, $p = .10$. To summarize, no positive correlations between N400 expectancy effects and parietal SMEs were obtained, suggesting that individual differences in facilitated semantic processing and successful memory formation were not positively associated. Rather, N400 effects and parietal SMEs elicited by WC-UNEXP words correlated negatively, which suggests that in this condition, strong item-specific encoding was associated with poor (semantic) expectancy processing.

Both pre- and post-stimulus neural activity predicted subsequent memory, but it is unclear whether these two types of SMEs reflect processes that work in concert to support successful memory formation, or whether they reflect independent processes. To further explore the relationship between pre- and post-stimulus SMEs we computed four correlations between pre-stimulus SMEs elicited by SC and WC sentences and post-stimulus parietal SMEs in the four experimental conditions. Pre-stimulus SMEs were calculated based on mean amplitudes between 800 and 200 ms before target word onset at anterior electrodes. One dataset had to be excluded from this analysis due to the amplitude of the SME elicited by SC-UNEXP words which deviated by more than three standard deviations from the sample mean. Notably, as the polarity of pre- and post-stimulus was reversed (i.e., more negative amplitudes in

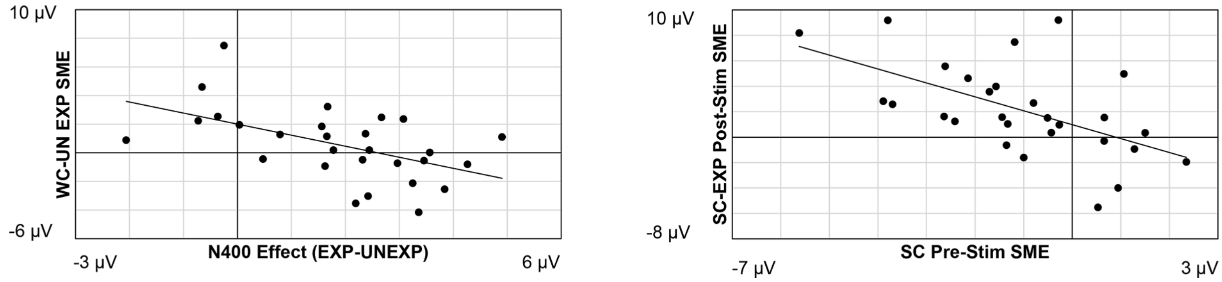

Fig. 3. Correlation between N400 effects, measured as the differences in amplitudes between expected and unexpected words, and parietal SMEs, measured as the differences in amplitudes between subsequent hits and misses, elicited by WC-UNEXP words (left side). N400 effects and parietal SMEs were calculated based on mean amplitudes in the 300–500 ms time window at posterior electrodes. Correlation between pre-stimulus SMEs elicited by SC sentences and parietal SMEs elicited by SC-EXP words (right side). Pre-stimulus SMEs were calculated based on mean amplitudes between 800 and 200 ms before target word onset at anterior electrodes.

the pre-stimulus interval predicted successful remembering, and vice versa in the post-stimulus interval), negative correlations would indicate a positive association between large pre- and post-stimulus SMEs. As evident from Fig. 3 (right side), pre-stimulus SMEs elicited by SC sentences were strongly and negatively correlated with post-stimulus SMEs elicited by SC-EXP words, $r(28) = -0.56$, $p < .01$. The correlation with post-stimulus SMEs for SC-UNEXP words was not significant, $r(28) = 0.32$, $p = .10$, and no significant correlations were found between pre-stimulus SMEs elicited by WC sentences and post-stimulus SMEs elicited by target words completing these sentences (WC-EXP: $r(28) = 0.28$, $p = .15$; WC-UNEXP: $r(28) = -0.25$, $p = .21$).

2.2.2.2. 500–700 ms time window. In the analysis of mean amplitudes in the 500–700 ms time window, depicted in Fig. 1b and Fig. 2, the main effect of Memory, $F(1,28) = 38.56$, $p < .001$, $\eta_p^2 = 0.58$, the Constraint by Expectedness interaction, $F(1,28) = 6.20$, $p < .05$, $\eta_p^2 = 0.18$, and the triple interaction between Antpos, Expectedness, and Memory, $F(1,28) = 4.59$, $p < .05$, $\eta_p^2 = 0.14$, reached significance. To further explore the significant interaction involving Antpos, mean amplitudes at anterior and posterior electrodes were analyzed in two separate ANOVAs including the factors Constraint, Expectedness, and Memory.

Mean amplitudes at anterior electrodes differed as a function of Memory, $F(1,28) = 28.95$, $p < .001$, $\eta_p^2 = 0.51$, reflecting more positive amplitudes for subsequent hits ($M = 4.22$, $SEM = 0.72$ μV) than for misses ($M = 2.80$, $SEM = 0.65$ μV). No further effects reached significance, all p-values >0.20.

At posterior electrodes, a significant main effect of Memory, $F(1,28) = 27.14$, $p < .001$, $\eta_p^2 = 0.49$, indicated more positive amplitudes for hits ($M = 4.88$, $SEM = 0.72$ μV) than for misses ($M = 3.35$, $SEM = 0.56$ μV). Further, the Constraint by Expectedness interaction was significant, $F(1,28) = 28.58$, $p < .001$, $\eta_p^2 = 0.27$. As revealed by subsidiary t-tests, the significant Congruency by Expectedness interaction indicates that WC-UNEXP words were associated with more negative amplitudes than WC-EXP ones (WC-EXP: $M = 4.33$, $SEM = 0.69$ μV; WC-UNEXP: $M = 3.63$, $SEM = 0.66$ μV; $t(28) = 3.21$, $p < .01$, $d = 0.19$), whereas SC-EXP and -UNEXP words did not differ in the 500 to 700 ms time interval (SC-EXP: $M = 3.89$, $SEM = 0.56$ μV; SC-UNEXP: $M = 4.59$, $SEM = 0.76$ μV; $t(28) = 1.58$, $p = .13$, $d = 0.18$).

To summarize, words completing SC sentences elicited strong N400 expectancy effects at posterior electrodes in the 300–500 ms time window, but these effects did not extend into the adjacent 500–700 ms time window. In contrast, words completing WC sentences were associated with numerically weaker N400 effects in the 300–500 ms time window, but these effects were longer-lasting and extended into the 500–700 ms

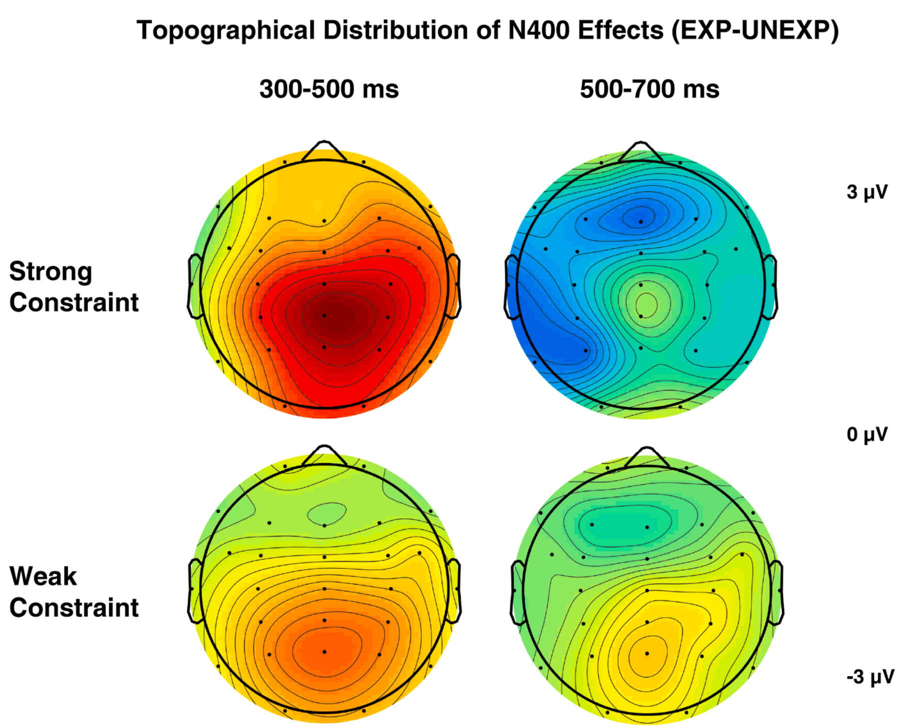

Fig. 4. Scalp topographies for N400 expectancy effects in the 300–500 ms and 500–700 ms time windows.

time window, which is also evident from Fig. 4 showing the topographic distribution of the EXP-UNEXP differences in both conditions and time windows.

2.2.2.3. 700–1200 ms time window. The analysis of mean amplitudes in this late time window, depicted in Fig. 1b and Fig. 2, yielded significant main effects of Expectedness, $F(1,28) = 5.14$, $p < .05$, $\eta_p^2 = 0.16$, and Memory, $F(1,28) = 31.62$, $p < .001$, $\eta_p^2 = 0.53$, qualified by significant Antpos by Expectedness, $F(1,28) = 17.73$, $p < .001$, $\eta_p^2 = 0.39$, and Constraint by Memory interactions, $F(1,28) = 7.22$, $p < .05$, $\eta_p^2 = 0.21$, and a triple interaction involving Antpos, Constraint, and Memory, $F(1,28) = 7.60$, $p < .05$, $\eta_p^2 = 0.21$. To follow up the significant interactions involving Antpos, mean amplitudes in the anterior and posterior electrode clusters were analyzed in two separate ANOVAs including the factors Constraint, Expectedness, and Memory.

In the analysis of mean amplitudes at anterior electrodes, significant main effects of Expectedness, $F(1,28) = 13.33$, $p < .01$, $\eta_p^2 = 0.32$, and Memory, $F(1,28) = 24.75$, $p < .001$, $\eta_p^2 = 0.47$, were obtained, qualified by significant Constraint by Expectedness, $F(1,28) = 5.76$, $p < .05$, $\eta_p^2 = 0.17$, and Constraint by Memory interactions, $F(1,28) = 10.90$, $p < .01$, $\eta_p^2 = 0.28$. Follow-up *t*-tests revealed that SC-UNEXP words elicited more positive amplitudes than SC-EXP ones (SC-EXP: $M = 2.70$, $SEM = 0.62$ μV; SC-UNEXP: $M = 4.47$, $SEM = 0.61$ μV; $t(28) = 3.46$, $p < .01$, $d = 0.53$), whereas the difference in mean amplitudes between WC-EXP and -UNEXP words was only marginally significant (WC-EXP: $M = 3.06$, $SEM = 0.62$ μV; WC-UNEXP: $M = 3.60$, $SEM = 0.53$ μV; $t(28) = 2.04$, $p = .05$, $d = 0.17$). To follow up the significant Constraint by Memory interaction, SME were calculated as differences in amplitudes between subsequently remembered and forgotten words completing strongly and weakly constraining sentences. Words completing WC sentences elicited an SME larger than zero ($M = 1.96$, $SEM = 0.32$ μV, $t(28) = 6.11$, $p < .001$, $d = 1.14$), whereas those completing SC sentences did not ($M = 0.62$, $SEM = 0.34$ μV, $t(28) = 1.85$, $p = .08$, $d = 0.34$). A paired *t*-test revealed that the SME elicited by words completing WC sentences was larger than the one associated with words completing SC sentences, $t(28) = 3.30$, $p < .01$, $d = 0.75$.

The analysis of mean amplitudes at posterior electrodes yielded a significant main effect of Memory, $F(1,28) = 21.62$, $p < .001$, $\eta_p^2 = 0.44$, reflecting more positive amplitudes for subsequent hits ($M = 3.19$, $SEM = 0.55$ μV) than for misses ($M = 2.10$, $SEM = 0.47$ μV). No further effects reached significance, all *p*-values > 0.21.

To summarize, as evidenced by a significant Constraint by Expectedness interaction, unexpected words completing SC sentences elicited a late frontal positivity between 700 and 1200 ms after word onset. In this late time window, SMEs at anterior electrodes were modulated by sentential constraint, i.e., they were larger for WC sentences than for SC sentences.

2.2.2.4. Post hoc analyses in the 180–250 ms time window. A visual inspection of the ERP waveforms elicited by the onset of words, depicted in Fig. 1b and Fig. 2, suggests that systematic differences between the experimental conditions were already present in an early time window preceding the effects of interest in the 300 – 500 ms interval. To further explore these unexpected effects, mean amplitudes in this early time window at electrode Fz were analyzed in an ANOVA including the factors Constraint, Expectedness, and Memory. The main effect of Memory was significant, $F(1,28) = 7.94$, $p < .01$, $\eta_p^2 = 0.22$, and reflected more positive amplitudes for subsequent hits ($M = 6.89$, $SEM = 0.63$ μV) than for misses ($M = 6.18$, $SEM = 0.56$ μV). Even though a visual inspection of Fig. 2b suggests that SMEs were larger for unexpected than for expected words, this turned out not to be the case, as there was no significant interaction involving Memory, all *p*-values > 0.29. However, the main effect of Constraint, $F(1,28) = 4.74$, $p < .05$, $\eta_p^2 = 0.15$, was qualified by a significant Constraint by Expectedness interaction, $F(1,28) = 8.10$, $p < .01$, $\eta_p^2 = 0.22$. Subsidiary *t*-tests revealed that expected words were associated with more positive amplitudes than unexpected words when completing SC sentences (SC-EXP: $M = 7.35$, $SEM = 0.69$ μV; SC-UNEXP: $M = 6.27$, $SEM = 0.55$ μV; $t(28) = 2.44$, $p < .05$, $d = 0.31$), but not when completing WC sentences (WC-EXP: $M = 6.13$, $SEM = 0.60$ μV; WC-UNEXP: $M = 6.38$, $SEM = 0.68$ μV; $t(28) = -0.66$, $p = .52$, $d = -0.07$).

3. Discussion

3.1. Memory effects of sentential constraint and word expectedness

This study investigated how the strength of schema support provided by a sentence context affects the encoding of expected and unexpected words, and how this is reflected in neural activity preceding and following the onset of sentence-ending words. We hypothesized that predictive sentential contexts should facilitate the processing of expected words in particular. Furthermore, predictive sentence contexts should activate schemas which enable the formation of more stable and elaborated memory traces for expected words and enhance memory for these words. Consistent with this hypothesis, expected words were associated with higher hit rates in the recognition memory test one day later as compared with unexpected ones irrespective of sentential constraint. It is conceivable that, even though all sentence-ending words in the learning phase were congruent with the preceding sentences, expected words were perceived as being more congruent than unexpected ones, with the consequence being that expected words benefited more from schema support provided by the sentences than unexpected words did. Thus, word expectedness could in fact be considered as reflecting variations in schema congruency (for a similar approach, see a recent study by Quent et al. (2021), in which images of objects which were presented in expected or unexpected locations were used as manipulations of schema congruency and incongruency). Apparently, as sentential constraint did not modulate the effect of expectedness on memory, even weakly constraining sentences provided strong enough contexts to boost the encoding of expected words into memory and to make them more readily assessable in the subsequent memory test. We also found that words completing strongly constraining (SC) sentences in the learning phase were remembered better than words completing WC sentences. This finding suggests that highly predictive sentences did indeed facilitate the encoding of congruent words even when they were of low expectedness.

Based on theoretical accounts assuming that prediction errors elicited by expectancy mismatches promote learning (Henson and Gagnepain, 2010; see also Kuperberg and Jaeger, 2016, for a similar argument in the field of language comprehension) presumably by capturing attention and the ensuing thorough encoding of the eliciting events (Butterfield and Metcalfe, 2006; Fazio and Marsh, 2009), it was predicted that there should be a memory advantage for unexpected words due to the prediction errors they elicit, and that this effect should depend on the strength of predictions induced by a sentence context. This hypothesis was not confirmed. It is possible that unexpected words occurred too frequently in the learning phase to provoke strong prediction errors that could have affected memory for these words. Reggev et al. (2017) examined recognition memory for words learned in the context of a semantically congruent or incongruent word and varied the proportion of incongruent word pairs at study. In support of the aforementioned view, memory performance for words learned in the context of an incongruent word was higher when the proportion of incongruent word pairs at study was lowered and their distinctiveness increased. Conversely, memory for congruent words was completely unaffected by the proportion manipulation. Future studies should therefore consider manipulating the proportions of expected and unexpected words at study to identify boundary conditions under which schema-congruency and expectancy mismatches enhance memory.

Contrary to our results, a recent study by Hubbard et al. (2019) using a similar approach did not find similar effects of sentential constraint and word expectedness on hit rates in a recognition memory test for sentence-ending words. One important methodological difference between the two studies is that sentences were presented word by word in the Hubbard et al. (2019) study, with a short presentation time of 200 ms per word and 300 ms interstimulus intervals. In contrast, in the present study, all words constituting a sentence context were presented for five seconds and separated from the presentation of the sentence-ending target word (1.5 s) by a one second delay. This experimental procedure was intended to boost processes related to the prediction of the target word. Previous studies have found that the confirmation of predicted word form representations or the more efficient visual feature extraction for highly predictable words is reflected in enhanced frontal P200 amplitudes for expected words processed in highly predictive contexts (Federmeier et al., 2005; Lau et al., 2013). Thus, our finding that expected words completing strongly constraining sentences elicited larger frontal P200 amplitudes than unexpected words and those completing weakly constraining sentences indicates that that the presentation mode employed in the present study indeed contributed to the generation of stronger word form predictions. These strong predictions could have boosted the facilitating effect of predictive processing on the encoding of expected words, resulting in superior memory for these words. Thus, in line with recent studies highlighting the importance of the top-down modulation of predictive processes in language comprehension (Brothers et al., 2017; Kuperberg, 2021; Lau et al., 2013), the divergent results in behavioral memory performance for sentence-ending words between the present study and the Hubbard et al. (2019) study indicate that predictive mechanisms in sentence comprehension can be strongly modulated by processing strategies induced by the experimental procedure employed in a given study.

3.2. Schema effects on memory for expected but not presented words

A second important aim of the present study was to explore the fate of words that were expected but never actually seen during the study phase and presented as lures in the recognition memory test one day later. As suggested by the results obtained by Hubbard et al. (2019), predicted but not presented words remain in a state of increased pre-activation which could lead to an increased fluency associated with the processing of expected lures at test and, as a consequence, an increased rate of false positive memory decisions for these words. If stronger schema-based predictions lead to a stronger pre-activation of predicted words, then expected lures from SC sentence contexts should be associated with a higher rate of false alarms than expected lures from WC sentence contexts. This is exactly what we found: Expected lures were associated with higher false alarm rates than unrelated new words, and expected words from SC sentences were associated with more false alarms than those from WC sentences. This pattern of results shows that schema-based learning and predictive processing is not only associated with mnemonic benefits but can also be detrimental for subsequent memory presumably because predicted but not presented words are pre-activated which in turn leads to increased fluency during the processing of these words at test. The results obtained in the present study confirm and extend those obtained by Hubbard et al. (2019) using a short retention interval of several minutes and indicate that the accessibility of memory representations for words that were expected but not actually seen in the learning phase was increased even one day after the processing of the sentences. Hence, our results provide additional evidence for the view that the representations of lexical predictions linger in memory (Hubbard et al., 2019; Rommers and Federmeier, 2018a). Rich and Harris (2021) further distinguished into an active and a passive version of the lingering activation account. They argued that the increased activation of a predicted word could either be actively maintained in memory, or that it could passively and gradually decay following the disconfirmation of the prediction. To decide between the active and the passive version of the lingering activation account, Rich and Harris (2021) compared reading times for previously predicted words after that prediction had been disconfirmed under near and far distance conditions between the prediction site and the previously predicted target word. Even though far distance sentences differed from near distance sentences only in one additional adverbial preceding the target word (e.g., "It was obvious that John was afraid of creepy crawlers. He screamed when he saw the mouse in the corner but *somewhat surprisingly* he didn't notice the **spider** crawling on the wall", adverbial in italics, target word in bold), shorter reading times for previously predicted words were obtained under near distance conditions only, which suggests that the increased activation of predicted words can decay rapidly during sentence processing. In stark contrast to the results obtained in the recent studies by Rich and Harris (2021) and Lai et al. (2021), which are rather in favor of the passive version of the lingering activation account, our data suggest that the lingering activation of expected but not presented words can affect memory decisions even after retention intervals as long as 24 h. However, it seems unlikely that the heightened activation of expected but not presented words was actively maintained for many hours. It was recently proposed that strong lexical predictions are associated with a pre-updating of the sentence context with the predicted word in working memory (Lau et al., 2013; Ness and Meltzer-Asscher, 2018b). In the present study, the one second delay between sentence contexts and target words at study presumably boosted the prediction of target words and the pre-updating of sentence representations. If a pre-updated sentence representation was not thoroughly revised upon the disconfirmation of the predicted word, then the predicted but not presented word could linger in memory for a longer time without being actively maintained. Our finding that predicted but not presented sentence-ending words were associated with a high frequency of false positive memory decisions one day after the initial processing of the sentences could reflect the lingering of pre-updated sentence representations with the predicted but not presented word. Thus, our results suggest that the downstream consequences of predictive processing depend on the strength of predictions: Strong lexical predictions associated with the pre-updating of the sentence representation are able to affect memory decisions even after long time intervals. On the contrary, weaker predictions may be associated with a short-lived pre-activation of the predicted words as reflected in rather implicit measures of memory such as N400 repetition effects (Rommers and Federmeier, 2018a).

3.3. N400 effects and parietal SMEs

N400 amplitudes between 300 and 500 ms at posterior electrodes were more negative for unexpected vs. expected words and more positive for words completing SC sentences than for those completing WC sentences. Notably, even though the N400 expectedness effect (i.e., more negative amplitudes for unexpected than for expected words) was numerically larger for words completing SC sentences than for those completing WC sentences, the Constraint by Expectedness interaction did not reach significance. Apparently, the difference in cloze probabilities between expected words ($M = 0.29$, $SD = 0.09$) and unexpected words ($M = 0.02$ SD $= 0.02$) completing WC sentences was large enough to elicit a N400 expectedness effect for WC sentences which was reflected in a main effect of Expectedness rather than in a Constraint by Expectedness interaction. Our results suggest that even the schema support provided by our weakly constraining sentence contexts was strong enough to facilitate the processing of expected words. However, another effect for the N400 also suggests that strongly constraining sentence contexts more effectively facilitated the processing of expected words: N400 expectedness effects elicited by words in WC sentences temporally extended into the 500–700 ms time window, whereas for words completing SC sentences expectedness effects were temporally restricted to the 300 to 500 ms time window. This finding could reflect the facilitated and accelerated integration of words processed in a strongly constraining sentence context (for a similar finding, see León-Cabrera et al., 2019).

As predicted, a parietal SME indicative of the encoding of item-specific details (Fabiani et al., 1986; Kamp et al., 2017) was present in the 300–500 ms time window. This result confirms our previous finding that schema congruency boosts the encoding of item-specific details (Höltje et al., 2019). In further support of this view, a recent study conducted in our lab shows that the parietal SME also supports the contextually supported learning of new compound words (Meßmer et al., 2021). In the present study, expected words elicited larger parietal SMEs than unexpected words. Thus, the pattern of parietal SMEs paralleled the behavioral memory advantage for expected words over unexpected words, which suggests that in particular the encoding of expected words was at the electrophysiological level supported by the parietal SME. This pattern of the results highlights the mnemonic benefits associated with the processing of confirmed predictions by showing that predictive sentence contexts facilitate the processing of words confirming predictions, as reflected in N400 amplitudes, and also boost item-specific memory encoding for these words over less expected words.

N400 and parietal subsequent memory effects were present in the same 300–500 ms time window at posterior electrodes, which raises the question whether the facilitated semantic processing reflected in the N400 attenuation also contributes to item-specific memory encoding as reflected in the parietal SME. In the present study, N400 amplitudes were smaller (i.e., more positive) for words completing SC than WC sentences. Thus, N400 effects were dissociable from parietal SME effects which did not differ between words completing either sentence type (SC or WC) but were modulated by word expectedness instead. Furthermore, if facilitated semantic processing as reflected in the N400 contributes to memory encoding, then N400 expectancy effects and parietal SME should correlate positively on the individual subject level. This was not the case in the present study. Surprisingly, however, there was a negative correlation between N400 expectancy effects and SMEs elicited by unexpected words when these words completed weakly constraining sentences. It is conceivable that this significant negative correlation was driven by a subset of participants who showed a reversed expectancy N400 effect for weakly constraining sentences (see Fig. 3, left panel). For this subgroup of participants expectations regarding the sentence-final words in WC sentences deviated from those of most other participants. These participants tended to process unexpected words like expected ones, resulting in a reversed N400 effect and large SMEs for these words. To summarize, our results suggest that the early parietal SME supported the item-specific encoding of expected words that leads to memory representations that are distinctive from other presentations and more likely to be retrieved in a subsequent memory test. This processing takes place in close temporal proximity to the N400 but is independent from the facilitated semantic processing reflected in the N400.

3.4. Late frontal positivity effects and frontal slow wave SMEs

As evidenced by a significant Constraint by Expectedness interaction, ERPs elicited by unexpected words completing SC sentences were associated with a late frontal positivity (LFP) between 700 and 1200 ms that is thought to reflect expectancy mismatch-related processing (Federmeier et al., 2007; Kuperberg et al., 2019; Quante et al., 2018). The late time period in which LFP effects were most pronounced is consistent with the timing of functionally similar frontal slow waves in recent studies (DeLong et al., 2014; DeLong and Kutas, 2020; Höltje et al., 2019). In the present study, in the same 700–1200 ms time window, an SME at frontal electrodes was selectively revealed for words completing WC sentences. Similar late frontal slow wave SMEs are associated with associative episodic memory encoding (Fabiani et al., 1990; Forester et al., 2020; Kamp et al., 2017). It is conceivable that the processing of words in WC sentences without strong contextual support necessitated sustained elaborative encoding processes reflected in the late frontal slow wave SME. These late encoding processes may help to establish context-word associations in particular when contextual support is weak. In contrast, strongly constraining sentences could facilitate the early formation of bindings between words and the preceding contexts without requiring further encoding efforts. This could be the case if preexisting associative connections contained in highly predictive sentence contexts are associated with enhanced semantic elaboration and relational binding operations at encoding that render the resulting memory trace more accessible for subsequent memory tests (Staresina et al., 2009). Thus, even though sustained encoding efforts as reflected in the late frontal slow wave predicted subsequent memory for words processed in weakly predictive contexts, SC sentences and the strong contextual support they provide are presumably more effective in promoting memory formation, which is at least indirectly supported by the higher hit rates in the SC condition.

3.5. Stimulus-preceding SMEs

A remaining question in our discussion is whether there was neurophysiological evidence for predictive processing and successful memory formation even before the onset of the critical words. ERPs recorded during the 1000 ms interval between the offset of sentence contexts and the onset of sentence-ending words in the study phase elicited a sustained negative potential at anterior electrodes that has been hypothesized to reflect processes associated with semantic predictions (Grisoni et al., 2017; León-Cabrera et al., 2017, 2019). León-Cabrera et al. (2019, 2017) compared ERPs elicited by SC and WC sentence contexts (e.g., SC: “The goalkeeper managed to catch the”; WC: “As a present she gave her son a”) with ERPs elicited by non-semantic sentence contexts that were semantically meaningless but grammatically plausible (NS, e.g., “Helade algoroa seujohi nua”[1]) and found that amplitudes of the sustained anterior negativity followed an SC < WC < NS pattern. Consistent with the studies by León-Cabrera et al. (2019, 2017), in the present study both SC and WC sentences elicited a frontal slow wave. This could be the case because not only SC but also WC sentences may have enabled semantic predictions to some extent. Notably, WC sentences in the present study had a relatively high cloze probability of 29% as compared to 6.1% in the León-Cabrera et al. (2019) study This supports the view that even weakly to moderately predictive sentence contexts enabled predictive processing in the present study but less so in the León-Cabrera study.

Between 800 and 200 ms preceding the onset of sentence-ending words, more negative frontal slow wave amplitudes predicted subsequent memory for upcoming words in both constraint conditions. This finding is reminiscent of a similar pre-stimulus SME in a study by Otten et al. (2006) and could reflect the engagement of working memory (WM) control processes. Support for this view comes from studies showing that frontally-distributed negative slow waves are associated with higher order WM control processes (Bosch et al., 2001; Ruchkin et al., 2003). Similar sustained anterior negativities elicited during the processing of sentences with long-distance dependencies between elements have recently been proposed to reflect the maintenance of discourse information in WM (Cruz Heredia et al., 2021). In a similar vein, the stimulus-preceding slow wave observed in the present study could reflect the maintenance of contextually based predictions in WM during the one second delay between the sentence context and the target word. If the computation and maintenance of semantic predictions in WM facilitates the semantic integration of upcoming words and supports the rapid encoding of words that confirm predictions, then large pre-stimulus SMEs should be associated with large post-stimulus SMEs elicited by words confirming strong predictions. This is exactly what we found: Pre-stimulus SMEs elicited by SC sentences correlated positively with parietal SMEs elicited by SC-EXP words, whereas correlations were smaller and nonsignificant for unexpected words and WC sentences. To

[1] Please note that semantically meaningful sentence contexts were originally presented in Spanish in the studies by León-Cabrera et al. (2019, 2017). Only the example for the NS condition is provided in its original version here.

summarize, we found that a sustained negative potential occurring prior to the presentation of sentence-ending words predicted subsequent memory in particular for words that confirmed strong schema-based predictions. These results provide strong evidence in support of the view that sentence comprehension not only involves the active prediction of upcoming words (Altmann and Mirković, 2009; Kutas et al., 2011; Levy, 2008; Pickering and Garrod, 2013) but also that neural activity reflecting pre-stimulus predictive processing does contribute to the successful encoding of these words.

3.6. Early frontal SMEs and P200 effects

As revealed by a post hoc analysis, ERPs elicited by the onset of sentence-ending words in the study phase were associated with a frontally-distributed positive deflection between 180 and 250 ms that predicted subsequent memory. This finding, however preliminary and requiring replication, suggests that post-stimulus neural activity predictive of subsequent memory can emerge considerably earlier than between 300 and 400 ms, as found in previous studies (Höltje et al., 2019; Kamp et al., 2017; Otten and Donchin, 2000). As evidenced by a Constraint by Expectedness interaction, mean amplitudes in the 180–250 ms time window were more positive for expected words completing SC sentences than for unexpected words and those completing WC sentences. This finding is consistent with previous studies reporting enhanced frontal P200 amplitudes for expected words processed in highly predictive contexts and could reflect the confirmation of the predicted orthographic representation of a word or the more efficient visual feature extraction for highly predictable words (Federmeier et al., 2005; Lau et al., 2013).

4. Method

4.1. Participants

Thirty-six young adults participated in the experiment. All participants were German native speakers, right-handed as confirmed by the Edinburgh Handedness Inventory (Oldfield, 1971), had normal or corrected-to-normal vision, and no self-reported neurological or psychiatric conditions. The experimental procedures were approved by the ethics board of the Faculty of Human and Business Sciences at Saarland University. Participants gave their informed consent before the experiment and received money (10 € per hour) or course credit as a compensation for their participation. Data from one participant could not be analyzed because the second session of the experiment was canceled. Thus, behavioral analyses are based on data from 35 participants (28 female). Their age ranged between 19 and 32 years, with a median age of 24 years. Due to the exclusion criteria for ERP data, these analyses are based on a lesser number of data sets (see Section 4.4).

4.2. Stimuli

Two hundred and forty sentence frames were used in the main ERP experiment, half of which were strongly constraining (SC) with regard to the sentence-ending word, as evidenced by cloze probabilities determined in a separate norming study (see below). The other half of the sentence frames were weakly constraining (WC), meaning that these sentence frames were not associated with the strong expectation of a specific sentence-ending word. In the main ERP experiment, half of the sentence frames were completed by expected target words associated with a high cloze probability, and the other half by unexpected target words with a cloze probability close to zero, resulting in four experimental conditions: SC sentence frames with expected target words (SC-EXP), SC sentence frames with unexpected target words (SC-UNEXP), WC sentence frames with expected target words (WC-EXP), and WC sentence frames with unexpected target words (WC-UNEXP). For examples of the stimuli, see Table 3.

Table 3
Examples of the sentences and words that were derived from the cloze norming study and used in the main ERP experiment. Please note that minor adaptations were made to the stimuli for the translation from German into English.

	Sentence Frame	Expected	Unexpected
SC	Because Jens has to get up early the next day, he soon goes to (the)	bed	kids' room
WC	When Benjamin arrives at home in the evening, his flatmate surprises him with a	dinner	song
SC	Caro got back her exam and was very proud of her	grade	handwriting
WC	Esther sits down at the table and writes a/an	list	advertisement
SC	Even though Deborah is very good at mental arithmetic, she uses a	calculator	pencil
WC	Modern society is based on the greek	democracy	teaching
SC	When Klara needs a break from studying, she gets up for a	walk	jump
WC	Even though his parents were against it, Oliver decided to become a	butcher	cook

Sentence frames and target words for the main ERP experiment were derived from an independent cloze study with 42 participants (31 female) whose age ranged between 18 and 20 years (*Md* = 21), and who did not participate in the main experiment. Four hundred sentences were generated and divided into two lists of 200 sentences each. Participants were presented the sentences from one list without the sentence-ending words and asked to complete each sentence frame with the word they would generally most expect to finish the sentence. Participants were also asked to provide a second sentence-ending word in order to obtain a larger number of less expected completions (see Federmeier et al, 2007, for a similar procedure). The number of ratings per sentence ranged between 8 and 22 (*Md* = 20) for best completions and between 10 and 22 (*Md* = 20) for "next best" completions.

From the resulting database 120 SC and 120 WC sentences were selected for which the best completions had mean cloze probabilities of 0.83 (*SD* = 0.13, range 0.60–1) and 0.29 (*SD* = 0.09, range 0.09–0.45), respectively, and divided into two lists of 60 SC and 60 WC sentences each that were matched for length (number of words including the sentence-ending word) both within and across lists (list 1: $M_{SC} = 9.65$, $SD_{SC} = 1.71$; $M_{WC} = 9.88$, $SD_{WC} = 1.81$; list 2: $M_{SC} = 9.57$, $SD_{SC} = 1.92$; $M_{WC} = 9.70$, $SD_{WC} = 1.78$; all p-values >0.35). In the main ERP experiment, sentence frames were either completed by their best completions or by less expected words with cloze probabilities close to zero (use as best completions: $M_{SC} = 0.01$, $SD_{SC} = 0.02$, range 0–0.10; $M_{WC} = 0.02$, $SD_{WC} = 0.02$, range 0–0.06; use as "next best" completions: $M_{SC} = 0.06$, $SD_{SC} = 0.04$, range 0–0.25; $M_{WC} = 0.04$, $SD_{WC} = 0.03$, range 0–0.16). These target words were singular nouns matched for word length (number of letters) and word frequency as measured by normalized lemma frequencies retrieved from the dlexDB database (Heister et al., 2011), see Table 4 for details. One hundred and twenty additional singular nouns that matched the target words in word length ($M_{NEW} = 6.91$, $SD = 2.19$, all p-values >0.51) and frequency ($M_{NEW} = 49.16$, $SD = 83.34$, all p-values >0.47) were retrieved from the dlexDB database and presented as new words in the test phase of the experiment.

4.3. Procedure

The main ERP experiment consisted of a study (40 min) and a test phase (50 min) separated by 24 h. Please note that, even though EEGs

Table 4
Mean (standard deviation) word lengths and frequencies for target words.

	SC-EXP	SC-UNEXP	WC-EXP	WC-UNEXP
Length	6.92 (2.76)	6.85 (2.59)	6.84 (2.43)	6.83 (2.78)
Frequency	49.83 (101.33)	44.47 (79.75)	47.39 (100.03)	46.62 (70.10)

were recorded in both sessions, only the study phase data are reported here. The preparation of the EEG recording took about 45 min each. Thereafter, participants were seated in front of a 19” computer screen with a resolution of 1280 × 1024 pixels in an electrically shielded and sound-attenuated booth. Experimental tasks were presented using the E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA) and participants used a keyboard for their responses. All list and key assignments were balanced across participants.

4.3.1. Study phase

Participants were presented the sentence frames from one of the two sentence lists completed by expected target words and those from the second sentence list completed by unexpected target words. Thus, participants processed each of the 240 sentence frames once and 60 trials in each of the four experimental conditions, plus eight practice trials. The 240 study trials were divided into six blocks of 40 trials each, separated by self-paced breaks. Participants were instructed to read the sentence frames and words carefully and in one third of all trials were asked to answer a yes/no comprehension question referring to the sentence frame. Trials were presented in pseudorandomized order to make sure that, firstly, no more than three trials of the same experimental condition were presented in direct succession and, secondly, no more than three successive trials contained a comprehension question.

Each trial was initiated by a fixation cross (500 ms), followed by a sentence frame (5000 ms) and a blank display (500 ms). Another fixation cross (500 ms) was shown before the target word (1500 ms) appeared, followed by a blank display (500 ms) and, in one third of all trials, a comprehension question (self-paced, max. 5000 ms) to which participants could respond by pressing the “c”- and “n”- keys of the keyboard. Trials were separated by an inter-trial interval jittered between 1500 and 2000 ms.

To verify that participants processed the sentence frames and words as instructed and paid attention to the content of the sentences, the proportion of correct responses to comprehension questions was calculated and analyzed.

4.3.2. Test phase

The 240 target words from the study phase were presented together with an equal number of new words in a surprise recognition memory test. Importantly, new words consisted of 120 unrelated new words and 120 words that had been expected but not actually seen in the study phase. That is, for each of the 120 sentence frames that had been completed by an unexpected word in the study phase, the more expected but not seen word was presented as an expected lure in the test phase. Old and new words were presented in pseudorandomized order, so that not more than three adjacent target or lure items were presented in direct succession. The 480 test trials were divided into six blocks of 80 trials each and separated by self-paced breaks. In the beginning of each trial, a fixation cross (500 ms) was presented, followed by a word (1000 ms). Participants were instructed to decide for each word whether it was old or new using a six-step confidence scale („sure old“, „probably old“, „maybe old“, „maybe new“, „probably new“, „sure new“). After the presentation of the word, a blank screen appeared for 1000 ms. Then, the question „Old or New?“ appeared, together with a depiction of the rating scale. The old/new decision could be given as soon as the word was presented. After the participants’ response, a blank screen was shown jittered between 1500 and 2000 ms before the next trial started.

To assess memory performance, Pr scores (Snodgrass and Corwin, 1988) were calculated as the difference between the proportions of correct and incorrect „old“-decisions (hits and false alarms). For this purpose, the corresponding three steps of the confidence scale (sure, probably, maybe) were collapsed into „old“- and „new“-decisions. Condition-specific analyses were performed on hit and false alarm rates.

4.4. EEG recording and processing

The EEG was recorded from 28 Ag/AgCl scalp electrodes embedded in an elastic cap with positions according to the 10–20 electrode system (Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC3, FCz, FC4, FC6, T7, C3, Cz, C4, T8, CP3, CPz, CP4, P7, P3, Pz, P4, P8, O1, O2, and A2). The vertical and horizontal EOG was recorded from four electrodes placed above and below the right eye and at the canthi of the left and right eyes. The electrodes were on-line referenced to a left mastoid electrode (A1), and AFz was used as a ground electrode. The EEG was amplified with a BrainAmp DC amplifier (Brain Products GmbH, Gilching, Germany) from 0.016 to 250 Hz and digitized at 500 Hz.

For off-line processing of the EEG data, the EEGLAB (Delorme and Makeig, 2004) and ERPLAB (Lopez-Calderon and Luck, 2014) toolboxes for MATLAB (The MathWorks Inc., Natick, MA) were used. Electrodes were re-referenced to the average of the left and right mastoid electrodes. The data were bandpass-filtered at 0.1–30 Hz using a second order Butterworth filter. A Parks-McClellan Notch filter was applied to the data to eliminate line noise at the frequency of 50 Hz. Pre-stimulus segments of 1000 ms length were extracted from the offset of the sentence frame to the onset of the target word and baseline-corrected based on activity during the first 100 ms of the segments. Post-stimulus segments were extracted from 200 ms before onset of the target word to 1500 ms thereafter and baseline-corrected based on activity during the 200 ms before target word onset. Independent component analysis (ICA) was applied to the segmented data to correct for ocular artifacts. Components associated with ocular artifacts were identified and rejected manually based on their activations and topographies. Segments containing artifacts were rejected using the following criteria: A minimal and maximal allowed total amplitude of ±100 μV, a maximal difference of values of 100 μV during intervals of 200 ms (window steps of 100 ms), and a maximal allowed voltage step of 30 μV/ms. On average, 6.34% of all segments were rejected.

4.5. ERPs

Post-stimulus ERPs were averaged for every combination of the factors Constraint (strong, weak), Expectedness (expected, unexpected), and Memory (hits, misses). Old words judged as “old” or “new” in the test phase were counted as hits and misses, respectively. For this purpose, the corresponding steps of the confidence scale were collapsed into “old”- and “new”-decisions. Hits were calculated based on “probably old” and “sure old” responses that can be assumed to reflect memory, whereas “maybe old” responses, which are likely to include decisions based on guessing, were discarded. Six data sets had to be excluded from the post-stimulus ERP analysis because there were not enough artifact-free trials (<7) to calculate reliable ERPs in one of the conditions (for recent SME studies using a similar criterion for trial selection, see Höltje et al., 2019; Kamp et al., 2017, 2018). The means and ranges of trial numbers per condition and participant were as following for post-stimulus ERPs: $M = 22$, range 7–40 (SC-EXP hits), $M = 20$, range 9–34 (SC-EXP misses), $M = 19$, range 7–33 (SC-UNEXP hits), $M = 23$, range 10–35 (SC-UNEXP misses), $M = 21$, range 8–41 (WC-EXP hits), $M = 21$, range 10–29 (WC-EXP misses), $M = 17$, range 7–32 (WC-UNEXP hits), $M = 24$, range 12–40 (WC-UNEXP misses). Pre-stimulus ERPs were averaged for every combination of the factors Constraint and Memory. For the sake of comparability of results, the pre-stimulus ERP analysis was based on the same 29 data sets as the analysis of the post-stimulus ERP data. Thus, all ERP analyses are based on data from $N = 29$ participants. The means and ranges of trial numbers per condition and participant were as following for pre-stimulus ERPs: $M = 42$, range 18–72 (SC hits), $M = 44$, range 24–63 (SC misses), $M = 40$, range 18–73 (WC hits), $M = 46$, range 23–69 (WC misses). Grand average waveforms were low-pass filtered at 12 Hz for illustration purposes.

Pre-stimulus ERP mean amplitudes were measured in four successive time windows of 200 ms length each, ranging from −800 ms to the onset of the target word. The electrode montage consisted of six fronto-central electrodes (F3, Fz, F4, FC3, FCz, FC4) at which pre-stimulus potentials associated with semantic predictions are largest in prior studies on pre-stimulus ERP activity (Grisoni et al., 2017; León-Cabrera et al., 2017, 2019).

Post-stimulus ERP mean amplitudes were measured in three consecutive time windows, including the 300–500 ms time window in which N400 effects are typically largest (Kutas and Federmeier, 2011) and SMEs emerge (Höltje et al., 2019; Kamp et al., 2017; Otten and Donchin, 2000), the adjacent 500–700 ms time window, and the 700–1200 ms time window in which late frontal positivity effects were expected to be largest (DeLong et al., 2011; Höltje et al., 2019; Quante et al., 2018). In order to capture both frontally- and parietally-distributed SMEs (Höltje et al., 2019; Kamp et al., 2017), N400 effects that are most pronounced at posterior electrodes (Kutas and Federmeier, 2011), and activity related to the late frontal positivity which is usually largest over anterior recording sites (Federmeier et al., 2007; Höltje et al., 2019; Kuperberg et al., 2019), the electrode montage consisted of 11 electrodes that cover anterior and posterior brain regions, divided into two electrode clusters (anterior: Fp1, Fp2, F3, Fz, F4; posterior: CP3, CPz, CP4, P3, Pz, P4).

4.6. Statistical analyses

All statistical analyses were conducted using IBM SPSS software. Behavioral and electrophysiological measures were analyzed using repeated-measures ANOVAs and dependent *t*-tests. Greenhouse-Geisser corrected degrees of freedom and *p*-values are reported whenever the assumption of sphericity was violated. Significant effects were decomposed using lower level ANOVAs and dependent *t*-tests. As measures of effect sizes, partial eta squared (η_p^2) are reported for ANOVA results. For independent *t*-tests, Cohen's *d* was calculated. For dependent *t*-tests, *d* was calculated according to Dunlap et al. (1996), taking into account the correlations between measurements. Error margins in graphs represent 95% confidence intervals based on the mean square error of the depicted effect (Jarmasz and Hollands, 2009).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Ann-Kathrin Brill, Isabelle Strack, Katja Haller, Leon Christidis, Annalisa Lipp, Tim Rüterbories and Sandra Glaser for assistance in creating the stimulus material and in data collection.

This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 232722074 – SFB 1102, project A6.

References:

Alba, J.W., Hasher, L., 1983. Is memory schematic? Psychol. Bull. 93 (2), 203–231. https://doi.org/10.1037/0033-2909.93.2.203.

Altmann, G.T.M., Mirković, J., 2009. Incrementality and prediction in human sentence processing. Cogn. Sci. 33 (4), 583–609. https://doi.org/10.1111/j.1551-6709.2009.01022.x.

Bartlett, F.C., 1932. Remembering: A Study in Experimental and Social Psychology. Cambridge University Press.

Bosch, V., Mecklinger, A., Friederici, A.D., 2001. Slow cortical potentials during retention of object, spatial, and verbal information. Cogn. Brain Res. 10 (3), 219–237. https://doi.org/10.1016/S0926-6410(00)00040-9.

Bransford, J.D., Johnson, M.K., 1972. Contextual prerequisites for understanding: Some investigations of comprehension and recall. J. Verb. Learn. Verb. Behav. 11 (6), 717–726. https://doi.org/10.1016/S0022-5371(72)80006-9.

Bridger, E.K., Bader, R., Mecklinger, A., 2014. More ways than one: ERPs reveal multiple familiarity signals in the word frequency mirror effect. Neuropsychologia 57 (1), 179–190. https://doi.org/10.1016/j.neuropsychologia.2014.03.007.

Brothers, T., Swaab, T.Y., Traxler, M.J., 2017. Goals and strategies influence lexical prediction during sentence comprehension. J. Mem. Lang. 93, 203–216. https://doi.org/10.1016/j.jml.2016.10.002.

Butterfield, B., Metcalfe, J., 2006. The correction of errors committed with high confidence. Metacogn. Learn. 1 (1), 69–84. https://doi.org/10.1007/s11409-006-6894-z.

Cohen, N., Pell, L., Edelson, M.G., Ben-Yakov, A., Pine, A., Dudai, Y., 2015. Peri-encoding predictors of memory encoding and consolidation. Neurosci. Biobehav. Rev. 50, 128–142. https://doi.org/10.1016/j.neubiorev.2014.11.002.

Corley, M., MacGregor, L.J., Donaldson, D.I., 2007. It's the way that you, er, say it: Hesitations in speech affect language comprehension. Cognition 105 (3), 658–668. https://doi.org/10.1016/j.cognition.2006.10.010.

Craik, F.I.M., Lockhart, R.S., 1972. Levels of processing: A framework for memory research. J. Verb. Learn. Verb. Behav. 11 (6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X.

Craik, F.I.M., Tulving, E., 1975. Depth of processing and the retention of words in episodic memory. J. Exp. Psychol. Gen. 104 (3), 268–294.

Cruz Heredia, A. A. L., Dickerson, B., Lau, E. (2021). Towards understanding sustained neural activity across syntactic dependencies. *Neurobiol. Lang.*, 1–22. https://doi.org/10.1162/nol_a_00050.

DeLong, K.A., Kutas, M., 2020. Comprehending surprising sentences: sensitivity of post-N400 positivities to contextual congruity and semantic relatedness. Lang. Cogn. Neurosci. 35 (8), 1044–1063. https://doi.org/10.1080/23273798.2019.1708960.

DeLong, K.A., Quante, L., Kutas, M., 2014. Predictability, plausibility, and two late ERP positivities during written sentence comprehension. Neuropsychologia 61 (1), 150–162. https://doi.org/10.1016/j.neuropsychologia.2014.06.016.

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. https://doi.org/10.1111/j.1469-8986.2011.01199.x.

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. https://doi.org/10.1016/j.jneumeth.2003.10.009.

Dunlap, W.P., Cortina, J.M., Vaslow, J.B., Burke, M.J., 1996. Meta-analysis of experiments with matched groups or repeated measures designs. Psychol. Methods 1 (2), 170–177. https://doi.org/10.1037/1082-989X.1.2.170.

Fabiani, M., Karis, D., Donchin, E., 1986. P300 and recall in an incidental memory paradigm. Psychophysiology 23 (3), 298–308. https://doi.org/10.1111/j.1469-8986.1986.tb00636.x.

Fabiani, M., Karis, D., Donchin, E., 1990. Effects of mnemonic strategy manipulation in a Von Restorff paradigm. Electroencephalogr. Clin. Neurophysiol. 75 (2), 22–35. https://doi.org/10.1016/0013-4694(90)90149-E.

Fazio, L.K., Marsh, E.J., 2009. Surprising feedback improves later memory. Psychon. Bull. Rev. 16 (1), 88–92. https://doi.org/10.3758/PBR.16.1.88.

Federmeier, K.D., Mai, H., Kutas, M., 2005. Both sides get the point: Hemispheric sensitivities to sentential constraint. Mem. Cogn. 33 (5), 871–886. https://doi.org/10.3758/BF03193082.

Federmeier, K.D., Wlotko, E.W., De Ochoa-Dewald, E., Kutas, M., 2007. Multiple effects of sentential constraint on word processing. Brain Res. 1146 (1), 75–84. https://doi.org/10.1016/j.brainres.2006.06.101.

Forester, G., Halbeisen, G., Walther, E., Kamp, S.-M., 2020. Frontal ERP slow waves during memory encoding are associated with affective attitude formation. Int. J. Psychophysiol. 158 (May), 389–399. https://doi.org/10.1016/j.ijpsycho.2020.11.003.

Friston, K., 2010. The free-energy principle: A unified brain theory? Nat. Rev. Neurosci. 11 (2), 127–138. https://doi.org/10.1038/nrn2787.

Ghosh, V.E., Gilboa, A., 2014. What is a memory schema? A historical perspective on current neuroscience literature. Neuropsychologia 53 (1), 104–114. https://doi.org/10.1016/j.neuropsychologia.2013.11.010.

Gilboa, A., Marlatte, H., 2017. Neurobiology of schemas and schema-mediated memory. Trends Cogn. Sci. 21 (8), 618–631. https://doi.org/10.1016/j.tics.2017.04.013.

Greve, A., Cooper, E., Kaula, A., Anderson, M.C., Henson, R.N., 2017. Does prediction error drive one-shot declarative learning? J. Mem. Lang. 94, 149–165. https://doi.org/10.1016/j.jml.2016.11.001.

Greve, A., Cooper, E., Tibon, R., Henson, R.N., 2019. Knowledge is power: Prior knowledge aids memory for both congruent and incongruent events, but in different ways. J. Exp. Psychol. Gen. 148 (2), 325–341. https://doi.org/10.1037/xge0000498.

Grisoni, L., Miller, T.M.C., Pulvermüller, F., 2017. Neural correlates of semantic prediction and resolution in sentence processing. J. Neurosci. 37 (18), 4848–4858. https://doi.org/10.1523/JNEUROSCI.2800-16.2017.

Haeuser, K.I., Kray, J., 2021. Effects of prediction error on episodic memory retrieval: evidence from sentence reading and word recognition. Lang. Cogn. Neurosci. 1–17. https://doi.org/10.1080/23273798.2021.1924387.

Hebscher, M., Wing, E., Ryan, J., Gilboa, A., 2019. Rapid cortical plasticity supports long-term memory formation. Trends Cogn. Sci. 23 (12), 989–1002. https://doi.org/10.1016/j.tics.2019.09.009.

Heister, J., Würzner, K.M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., Kliegl, R., 2011. dlexDB - Eine lexikalische Datenbank für die psychologische und linguistische Forschung. Psychologische Rundschau 62 (1), 10–20. https://doi.org/10.1026/0033-3042/a000029.

Henson, R.N., Gagnepain, P., 2010. Predictive, interactive multiple memory systems. Hippocampus 20 (11), 1315–1326. https://doi.org/10.1002/hipo.20857.

Höltje, G., Lubahn, B., Mecklinger, A., 2019. The congruent, the incongruent, and the unexpected: Event-related potentials unveil the processes involved in schematic encoding. Neuropsychologia 131, 285–293. https://doi.org/10.1016/j.neuropsychologia.2019.05.013.

Hubbard, R.J., Rommers, J., Jacobs, C.L., Federmeier, K.D., 2019. Downstream behavioral and electrophysiological consequences of word prediction on recognition memory. Front. Hum. Neurosci. 13 (291) https://doi.org/10.3389/fnhum.2019.00291.

Jarmasz, J., Hollands, J.G., 2009. Confidence intervals in repeated-measures designs: The number of observations principle. Canadian J. Exp. Psychol./Revue Canadienne de Psychologie Experimentale 63 (2), 124–138. https://doi.org/10.1037/a0014164.

Kamp, S.-M., Bader, R., Mecklinger, A., 2017. ERP subsequent memory effects differ between inter-item and unitization encoding tasks. Front. Hum. Neurosci. 11, 30. https://doi.org/10.3389/fnhum.2017.00030.

Kamp, S.-M., Bader, R., Mecklinger, A., 2018. Unitization of word pairs in young and older adults: Encoding mechanisms and retrieval outcomes. Psychol. Aging 33 (3), 497–511. https://doi.org/10.1037/pag0000256.

Kuperberg, G.R., 2021. Tea with milk? A hierarchical generative framework of sequential event comprehension. Top. Cogn. Sci. 13 (1), 256–298. https://doi.org/10.1111/tops.v13.110.1111/tops.12518.

Kuperberg, G.R., Brothers, T., Wlotko, E.W., 2019. A tale of two positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. J. Cognit. Neurosci. 32 (1), 12–35. https://doi.org/10.1162/jocn_a_01465.

Kuperberg, G.R., Jaeger, T.F., 2016. What do we mean by prediction in language comprehension? Lang. Cogn. Neurosci. 31 (1), 32–59. https://doi.org/10.1080/23273798.2015.1102299.

Kutas, M., DeLong, K.A., Smith, N.J., 2011. A look around at what lies ahead: prediction and predictability in language processing. In: Bar, M. (Ed.), Predictions in the Brain: Using Our Past to Generate a Future. Oxford University Press, pp. 190–207. https://doi.org/10.1093/acprof:oso/9780195395518.003.0065.

Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annu. Rev. Psychol. 62 (1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123.

Lai, M.K., Rommers, J., Federmeier, K.D., 2021. The fate of the unexpected: Consequences of misprediction assessed using ERP repetition effects. Brain Res. 1757, 147290. https://doi.org/10.1016/j.brainres.2021.147290.

Lau, E.F., Holcomb, P.J., Kuperberg, G.R., 2013. Dissociating N400 effect of prediction from association in single word contexts. J. Cognit. Neurosci. 484–502 https://doi.org/10.1162/jocn_a_00328.

León-Cabrera, P., Flores, A., Rodríguez-Fornells, A., Morís, J., 2019. Ahead of time: Early sentence slow cortical modulations associated to semantic prediction. NeuroImage 189 (January), 192–201. https://doi.org/10.1016/j.neuroimage.2019.01.005.

León-Cabrera, P., Rodríguez-Fornells, A., Morís, J., 2017. Electrophysiological correlates of semantic anticipation during speech comprehension. Neuropsychologia 99, 326–334. https://doi.org/10.1016/j.neuropsychologia.2017.02.026.

Levy, R., 2008. Expectation-based syntactic comprehension. Cognition 106 (3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006.

Lopez-Calderon, J., Luck, S.J., 2014. ERPLAB: an open-source toolbox for the analysis of event-related potentials. Front. Hum. Neurosci. 8 (April), 213. https://doi.org/10.3389/fnhum.2014.00213.

Meßmer, J.A., Bader, R., Mecklinger, A., 2021. The more you know: Schema-congruency supports associative encoding of novel compound words. Evidence from event-related potentials. Brain Cogn. 155 (October), 105813 https://doi.org/10.1016/j.bandc.2021.105813.

Ness, T., Meltzer-Asscher, A., 2018a. Lexical inhibition due to failed prediction: Behavioral evidence and ERP correlates. J. Exp. Psychol. Learn. Mem. Cogn. 44 (8), 1269–1285. https://doi.org/10.1037/xlm0000525.

Ness, T., Meltzer-Asscher, A., 2018b. Predictive pre-updating and working memory capacity: Evidence from event-related potentials. J. Cognit. Neurosci. 30 (12), 1916–1938. https://doi.org/10.1162/jocn_a_01322.

Oldfield, R.C., 1971. The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 9 (1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4.

Otten, L.J., Donchin, E., 2000. Relationship between P300 amplitude and subsequent recall for distinctive events: Dependence on type of distinctiveness attribute. Psychophysiology 37 (5), 644–661. https://doi.org/10.1017/S004857720098171X.

Otten, L.J., Quayle, A.H., Akram, S., Ditewig, T.A., Rugg, M.D., 2006. Brain activity before an event predicts later recollection. Nat. Neurosci. 9 (4), 489–491. https://doi.org/10.1038/nn1663.

Paller, K.A., Wagner, A.D., 2002. Observing the transformation of experience into memory. Trends Cogn. Sci. 6 (2), 93–102. https://doi.org/10.1016/S1364-6613(00)01845-3.

Piaget, J. (Ed.), 1952. The Origins of Intelligence in Children. W W Norton & Co, New York.

Piai, V., Anderson, K.L., Lin, J.J., Dewar, C., Parvizi, J., Dronkers, N.F., Knight, R.T., 2016. Direct brain recordings reveal hippocampal rhythm underpinnings of language processing. Proc. Natl. Acad. Sci. 113 (40), 11366–11371. https://doi.org/10.1073/pnas.1603312113.

Pickering, M.J., Garrod, S., 2013. An integrated theory of language production and comprehension. Behav. Brain Sci. 36 (4), 329–347. https://doi.org/10.1017/S0140525X12001495.

Quante, L., Bölte, J., Zwitserlood, P. (2018). Dissociating predictability, plausibility and possibility of sentence continuations in reading: evidence from late-positivity ERPs. *PeerJ*, *6*, e5717. https://doi.org/10.7717/peerj.5717.

Quent, J. A., Greve, A., Henson, R. N. (2021). *Shape of U: The relationship between object-location memory and expectedness*. https://doi.org/https://doi.org/10.31234/osf.io/xq37j.

Reggev, N., Sharoni, R., Maril, A., 2017. Distinctiveness benefits novelty (and not familiarity), but only up to a limit: the prior knowledge perspective. Cogn. Sci. 42 (1), 103–128. https://doi.org/10.1111/cogs.12498.

Rich, S., Harris, J.A., 2021. Unexpected guests: When disconfirmed predictions linger. Proceedings of the Annual Meeting of the Cognitive Science Society https://escholarship.org/uc/item/7q0000c8.

Rommers, J., Federmeier, K.D., 2018a. Lingering expectations: A pseudo-repetition effect for words previously expected but not presented. NeuroImage 183, 263–272. https://doi.org/10.1016/j.neuroimage.2018.08.023.

Rommers, J., Federmeier, K.D., 2018b. Predictability's aftermath: Downstream consequences of word predictability as revealed by repetition effects. Cortex 101, 16–30. https://doi.org/10.1016/j.cortex.2017.12.018.

Ruchkin, D.S., Grafman, J., Cameron, K., Berndt, R.S., 2003. Working memory retention systems: A state of activated long-term memory. Behav. Brain Sci. 26 (6), 709–728. https://doi.org/10.1017/S0140525X03000165.

Snodgrass, J.G., Corwin, J., 1988. Pragmatics of measuring recognition memory: Applications to dementia and amnesia. J. Exp. Psychol. Gen. 117 (1), 34–50. https://doi.org/10.1037/0096-3445.117.1.34.

Staresina, B.P., Gray, J.C., Davachi, L., 2009. Event congruency enhances episodic memory encoding through semantic elaboration and relational binding. Cereb. Cortex 19 (5), 1198–1207. https://doi.org/10.1093/cercor/bhn165.

Taylor, W.L., 1953. "Cloze procedure": A new tool for measuring readability. Journalism Q. 30 (4), 415–433. https://doi.org/10.1177/107769905303000401.