The N400 ERP component reflects an error-based implicit learning signal during language comprehension

Alice Hodapp | Milena Rabovsky

Department of Psychology, University of Potsdam, Potsdam, Germany

Correspondence
Alice Hodapp, Department of Psychology, University of Potsdam, 14476 Potsdam, Germany. Email: alice.hodapp@uni-potsdam.com

Funding information
German Research Foundation, Grant/Award Number: RA 2715/2-1

Edited by: Dr. John Foxe

Abstract
The functional significance of the N400 evoked-response component is still actively debated. An increasing amount of theoretical and computational modelling work is built on the interpretation of the N400 as a prediction error. In neural network modelling work, it was proposed that the N400 component can be interpreted as the change in a probabilistic representation of meaning that drives the continuous adaptation of an internal model of the statistics of the environment. These results imply that increased N400 amplitudes should correspond to greater adaptation, which can be measured via implicit memory. To investigate this model derived hypothesis, the current study manipulated expectancy in a sentence reading task to influence N400 amplitudes and subsequently presented the previously expected vs. unexpected words in a perceptual identification task to measure implicit memory. As predicted, reaction times in the perceptual identification task were significantly faster for previously unexpected words that induced larger N400 amplitudes in the previous sentence reading task. Additionally, it could be demonstrated that this adaptation seems to specifically depend on the process underlying N400 amplitudes, as participants with larger N400 differences during sentence reading also exhibited a larger implicit memory benefit in the perceptual identification task. These findings support the interpretation of the N400 as an implicit learning signal driving adaptation in language processing.

KEYWORDS
adaptation, EEG, expectancy, implicit learning, prediction error

1 | INTRODUCTION

1.1 | Error-driven learning

Neural computations have been suggested to be predictive (Elman, 1990; Friston, 2005; McClelland, 1994; Rao & Ballard, 1999; Schultz et al., 1997). Based on internal models of the environment, these implicit predictions...
allow for error-based learning: When the input deviates from expectations, prediction errors occur and allow for an adaptation of the current internal model to make better predictions in the future (Dell & Chang, 2014; Henson & Gagnepain, 2010; St. John & McClelland, 1990). This error-driven implicit learning could be the basis of our ability to adjust to the statistical regularities in our environment and ever-changing input. Prediction errors driving adaptation are central to both reinforcement learning (Sutton & Barto, 1998) and predictive coding theories (Friston, 2005; Rao & Ballard, 1999). Furthermore, various event-related brain potentials (ERPs) in human electroencephalography (e.g., feedback related negativity: Chase et al., 2010; Cohen & Ranganath, 2007; Walsh & Anderson, 2012; error related negativity: Holroyd & Coles, 2002; mismatch negativity: Garrido et al., 2009; Wacongne et al., 2012) have been suggested to reflect a prediction-error based update of internal models. In the domain of language comprehension, prediction errors have not only been linked to oscillatory activity (Lewis & Bastaiaansen, 2015) but also the N400 ERP component (Bornkessel-Schlesewsky & Schlesewsky, 2019; Fitz & Chang, 2019; Kuperberg et al., 2020; Rabovsky et al., 2018; Rabovsky & McRae, 2014).

1.2 Models of the N400 ERP component

The N400 is a centro-parietally distributed negative-going ERP component that peaks around 400 ms after presentation of a potentially meaningful stimulus and is attenuated by an item’s predictability within a given context (Kutas & Hillyard, 1980). The N400 amplitude is negatively correlated with a word’s cloze probability (Kutas & Hillyard, 1984), operationalized as the proportion of participants who complete a sentence fragment with the respective word in an independent norming task. While the N400 is reduced for predictable words in a graded manner and is smaller for sentence continuations that semantically overlap with the expected continuation (Federmeier & Kutas, 1999), it is not influenced by sentence constraint; that is, N400 amplitudes for unexpected words do not differ between situations in which a word is unexpected because the sentence context did not trigger any specific expectations, and situations in which a word is unexpected because a specific word was predicted that differs from the encountered word (Federmeier et al., 2007; Kutas & Hillyard, 1984).

Despite the large number of studies on the N400, its functional significance is still actively debated (for a review, see Kutas & Federmeier, 2011). For many years, the N400 literature has often focused on the debate whether the ERP reflected integration, that is, the effort to integrate an already retrieved word into the sentence context (e.g., Brown & Hagoort, 1993; Hagoort et al., 2009) or lexical access, that is, the effort to retrieve a word from memory (e.g., Laszlo & Federmeier, 2011; Lau et al., 2008). On the other hand, some recent theoretical developments have been characterized by a focus on prediction and specifically prediction errors. The prediction (or preactivation) perspective has often been linked to the lexical access perspective (i.e., lexical access is facilitated due to preactivation of predicted words in memory) as opposed to the integration perspective, but in principle the question of whether the N400 reflects a prediction error is independent of the question of whether the relevant process operates at the word or the sentence level (see, e.g., Rabovsky, 2020, for discussion).

Both verbally descriptive theoretical accounts (Bornkessel-Schlesewsky & Schlesewsky, 2019; Kuperberg et al., 2020) and computational models of the N400 (Fitz & Chang, 2019; Rabovsky et al., 2018; Rabovsky & McRae, 2014) have been based on this prediction error perspective.

1.3 The N400 and adaptation

Here, we focus on predictions derived from a neural network model of sentence comprehension, the Sentence Gestalt model (St. John & McClelland, 1990), which was used to model N400 amplitudes as the update in a probabilistic representation of sentence meaning in response to new cues, that is, incoming words (Rabovsky et al., 2018). In the Sentence Gestalt model, the amplitude of the model’s N400 correlate corresponds to the magnitude of change in the model’s hidden Sentence Gestalt layer activation induced by each incoming word. Because at any given point in sentence comprehension, the model’s Sentence Gestalt layer activation represents the model’s probabilistic prediction of sentence meaning, the change in this activation induced by the new incoming word corresponds to the prediction error contained in the previous representation. In neural network models, processing and learning are intertwined, as input continuously influences the models’ internal weights. Interestingly, in the Sentence Gestalt Model, the update in the hidden layer activation (i.e., the model’s internal prediction error and N400 correlate) has been used as the basis to adapt the model’s connection weights (Rabovsky et al., 2018), which implies that the model’s N400 correlate reflects the error signal used as a basis for model adaptation. Adaptation is meant here as a gradual and continuous implicit learning process that updates the model’s internal probability estimates in reaction to a specific
The facilitated processing of stimuli is characteristic for repetition priming paradigms and constitutes a well-established measure of implicit learning (Cofer, 1967). Repetition is usually accompanied by reductions in brain activity (Henson, 2003), which has also been observed specifically for the N400 in word lists (Rugg, 1985) and sentences (Besson et al., 1992; Van Petten & Kutas, 1991). The finding of an interaction between expectancy (or cloze probability) and repetition priming can offer some insights into a possible relationship between N400 amplitudes and implicit learning. In these studies, incongruent sentences lead to a larger N400 amplitude at first presentation compared to a congruent sentence but also to a larger reduction in N400 amplitude when the sentence is repeated after a delay (Besson et al., 1992). A similar pattern emerges when expectancy is manipulated via contextual constraint. Words that were unexpected due to a weakly constraining sentence (i.e., a sentence that did not trigger any specific predictions), showed a larger repetition effect when encountered again in a different sentence, compared with expected endings in highly constraining sentences (Rommers & Federmeier, 2018a). The expectancy and repetition interaction is in line with the suggested increased adaptation entailed by larger N400 amplitudes and was successfully simulated in the Sentence Gestalt model by Rabovsky et al. (2018), where the signal that was used as the basis to drive learning in the network corresponded exactly to the model’s N400 correlate.

On a single word level, the model would predict that weights are adjusted so that the word is more likely to be predicted in the future (equivalent to an increased base rate frequency estimate or an increased prior in a Bayesian account; see, e.g., Bornkessel-Schlesewsky & Schlesewsky, 2019; Delaney-Busch et al., 2019; Kuperberg, 2016; Rabovsky & McRae, 2014). This continuous error-driven adaptation is proposed to reflect a lifelong learning process, with the aim to capture the statistical regularities in the linguistic environment with as little error as possible to generate accurate predictions and optimize performance. If a comprehender encounters an unexpected word, it will elicit a large N400 amplitude. If the word was unexpected because of, for example, a low internal frequency representation, this estimate is updated (corrected) so that the comprehender will be less surprised by the word the next time it is encountered. The word will also likely reappear in the near future, as long as the context stays the same (e.g., the topic of a conversation or text that is read). The necessary updating process is proposed to be based on the error reflected in N400 amplitudes. This explains the typical repetition paradigm findings as reflecting a reduced prediction error (i.e., N400) upon second presentation and after model adaptation, since the model now has an adjusted internal estimate. According to the model predictions, the described N400-based adaptation process should also influence behaviour in implicit memory tasks: As N400 amplitudes are suggested to reflect model adaptation via prediction errors, a large N400 in response to a word should be reflected in a larger update of internally represented probability (and implicit anticipation) than if the same word was processed with a small prediction error (smaller N400 amplitude and smaller update of internally represented probability). This process is illustrated in Figure 1. Note that this description of the assumed relationship between N400 amplitude and implicit memory task performance presupposes that the respective word had the same probability distribution prior to model adaptation, as this then results in a higher internal baseline probability after processing the word in a high N400 context. To investigate this process, an experimental manipulation is necessary, which makes it possible that the same words are processed with different N400 amplitudes.

An experimental manipulation of N400 amplitude can be achieved by making use of the component’s sensitivity to a sentence’s cloze probability (Kutas & Hillyard, 1984). If a word is encountered in a low-cloze context (referred to as unexpected condition from now on) it will be processed with a larger N400 compared to the same word in a high-cloze setting (expected condition). This study therefore employed a sentence reading task that contained both contextually expected and unexpected items with the assignment of words to expectancy condition fully counterbalanced across participants. The reading task will serve as a manipulation of N400 amplitudes—while keeping words constant—for the critical subsequent implicit memory task. For our manipulation to work, we expect to replicate the classic N400 effect, with larger N400 amplitudes in response to unexpected words than expected words in response to our stimuli.
1.4 N400 and behavioural implicit memory performance

When an item is encountered more than once, it does not only influence brain activity (lower activity at second presentation; Henson, 2003) but also behavioural measures: Performance in response to such a repeated item has repeatedly been demonstrated to be improved in a variety of tests such as lexical decision, naming, and perceptual identification tasks (e.g., Feustel et al., 1983; Scarborough et al., 1977; Stark & McClelland, 2000). This facilitated processing (e.g., percentage of correct responses or reaction times) due to prior experience emerges without explicit memory or conscious recollection of the respective item. Instead, this benefit reflects implicit memory, which is defined as the influence of prior exposure in the absence of (or independent of) explicit memory (Graf & Schacter, 1985; Schacter & Graf, 1986a; Schacter & Graf, 1986b).

The current study used a perceptual identification (or perceptual fluency) paradigm adapted from Stark and McClelland’s (2000) work on repetition priming to assess implicit memory. The paradigm is implemented as a target word of increasing clarity with time (see Figure 2). Perceptual identification tests have been suggested as a reliable measure of implicit memory (Buchner & Wippich, 2000), but in their traditional implementation as near-threshold stimuli identification task offer only binary correct/incorrect responses. In contrast, the task implemented in this study allowed the recording of continuous reaction time data (for details, see methods section). The framework that is built upon here suggests—in addition to the general benefit induced by prior presentation—better performance for items that have been previously encountered in the unexpected (more negative N400) condition compared to expected condition, as the processing with a larger prediction error should lead to increased adaptation and implicit memory.

1.5 The role of positive ERP components

To demonstrate the specificity of the hypothesized relationship between N400 amplitudes and adaptation, we were also interested in examining possible relationships between post N400 positivities and adaptation. The N400 is not the only language related ERP component that was shown to be sensitive to predictability, but also post-N400 positivities (posterior and frontal) that have, amongst other things, been interpreted as reflecting the costs of prediction violations (Van Petten & Luka, 2012). The late posterior positivity/P600 response is observed in response to mostly syntactic and some semantic incongruities.
(anomalous words that lead to an incongruent interpretation) and commonly interpreted as being related to reanalysis and additional processing (e.g., Brothers et al., 2020; Kuperberg et al., 2020; Van Petten & Luka, 2012) and top-down allocation of attention via the coeruleus-norepinephrine system (Nieuwenhuis et al., 2005; Yu & Dayan, 2005). A frontally distributed positive component has been suggested to be elicited when an input disconfirms predictions but is nevertheless plausible (Federmeier et al., 2007; Federmeier et al., 2010; Quante et al., 2018; Van Petten & Luka, 2012) or when the context needs to be reanalysed (Wlotko et al., 2010) and (in contrast to the N400) is sensitive to specific lexical predictions rather than graded expectations at the level of meaning (Thornhill & Van Petten, 2012). Both positive ERP components mentioned have been suggested to reflect a conflict between predicted and bottom-up input (late frontal positivity: Delong et al., 2011; late posterior positivity: Kuperberg et al., 2020; Meerendonk et al., 2009), which implies that effects of our experimental manipulation on these positive components could be possible. Additionally, the late positivities have previously been suggested to represent learning (or adaptation) of the current statistical environment (Kuperberg et al., 2020). Since the current study focuses on investigating the relation between N400 amplitude and model adaptation, the hypotheses will not address the effect of positive ERP components. However, possible consequences of the experimental manipulation on positive ERP components will be compared to N400 effects to make the framework put forward here more precise and demonstrate the specificity of the presumed relation between N400 amplitudes and implicit memory formation.

1.6 | The current study

The present study experimentally manipulated N400 amplitudes by varying expectancy (cloze probability) in a sentence reading task and presenting the critical target words in a subsequent behavioural perceptual identification task measuring implicit memory to test the hypothesis that the N400 reflects an implicit learning signal based on prediction errors. If the N400 manipulation during sentence reading is successful, we expect to see the classic N400 effect: more negative amplitudes for the unexpected condition than the expected condition. Critically, we expect target words from this previously unexpected condition to be recognized faster than words from the previously expected condition in the subsequent behavioural implicit memory task.

2 | MATERIAL AND METHODS

2.1 | Participants

The study was preregistered on the Open Science Framework (https://osf.io/wg8nt/) and deviations are stated as such. The experiment is part of a research project whose protocols were approved by the Ethics Committee of the German Psychological Association (DGPs; MR102018). All participants gave written informed consent before the experimental session in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Thirty-three naive volunteers (12 male) participated in the experiment. Participants were compensated by either course credits or money (15€/h). Age ranged from 18 to 35 years (M = 23.88; SD = 5.60). All were
native speakers of German and right-handed (as assessed via the Edinburgh Handedness Inventory). All had normal or corrected-to-normal vision and none reported a history of neurological or psychiatric disorders.

2.2 | Stimuli and procedure

The experiment consisted of a sentence reading task with the conditions expected and unexpected, followed by a short working memory task (n-back task; see below for details) to prevent possible explicit memory effects, and finished with an implicit memory task, which included the target words from the previous sentence reading task with the conditions previously (during sentence reading) expected, previously unexpected and not previously seen. The experiment was therefore divided into three tasks and all stimulation was controlled by running the Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007) in MATLAB R2020a (MathWorks Inc. Natick, MA, USA). Overall, the experiment lasted approximately 1 h.

2.3 | Sentence reading task

The stimuli consisted of 120 German sentences (see Table 1 for example sentences and the OSF for the complete list of sentences: https://osf.io/wg8nt/) that were created in pairs that ended with the same target noun. The sentence pairs only differed in a minimal aspect, enough to change the expected ending of the sentence but keeping the baseline of words preceding the target word as similar as possible. For the unexpected condition, target words had a low cloze probability but were overall possible. Cloze probability was operationalized as the proportion of participants (n = 25) who used the critical word as sentence continuation in an online cloze test (on www.prolific.co). Nouns with the highest cloze probability in the unexpected sentences were never used as a target word in another sentence, as to prevent downstream consequences of expected but not seen words (Hubbard et al., 2019; Rommers & Federmeier, 2018b). The sentence pairs were divided across three counterbalanced lists, so that each target word was presented to an individual participant in either its expected or unexpected version, or not presented at all (serving as a not seen control condition for the implicit memory task later). Each of the lists contained 40 expected, 40 unexpected and 50 filler sentences. Neither conditions nor lists differed in their lexical characteristics (See Table 2). Sentences were divided into part A (first half of the sentence list) and part B (second half), which contained half of each stimulus condition, respectively. The order of part A and B was counterbalanced across participants and was kept the same for the implicit memory task to keep the distance between the presentation of a word during sentence reading and its repetition during the implicit memory task within a roughly comparable range. With this procedure, we aimed to avoid that, for example, the target item of the last sentence would by chance appear as the first implicit memory item, which would increase implicit memory effects, or that, for example, the target item of the very first sentence would by chance appear as the very last implicit memory item, which would decrease implicit memory effects, thus adding random noise to our measures of implicit memory. Sentence order was randomized within participants and it was additionally ensured that no more than two unexpected sentences would be presented successively.

1We preregistered 42 participants, but due to the COVID-19 pandemic data acquisition had to stop and we decided to analyse the data of the 33 participants for which we obtained data prior to the shutdown.

| Condition | Sentence |
|-----------|----------|
| Expected  | 1 a. Der Zauberer holte ein Kaninchen aus seinem Hut |
| Expected  | 1 b. Der Tierwärter holte ein Kaninchen aus seinem Hut |
| Expected  | 2 a. In Bayern angekommen, aß der Tourist eine frisch gebackene Brezel |
| Expected  | 2 b. In Italien angekommen, aß der Tourist eine frisch gebackene Brezel |
| Expected  | 3 a. Elias wollte Mia eine Postkarte schreiben und fragte nach ihrer Adresse |
| Expected  | 3 b. Elias wollte Mia anrufen und fragte nach ihrer Adresse |
| Expected  | 4 a. Nach einer langen Regenzeit versprach der Wetterbericht nun endlich wieder Sonne |
| Expected  | 4 b. Nach einer langen Dürrezeit versprach der Wetterbericht nun endlich wieder Sonne |

Note: The expected or unexpected target words are underlined.
In the sentence reading task, participants were instructed to attentively read each sentence for comprehension. Each trial started with a fixation cross until the participant pressed a button, after which a 300-ms black screen was presented. Then each word appeared for 250 ms at the centre of the screen with an interstimulus interval of 300 ms. Word presentation time was adjusted for words with more than 12 letters (20 ms per additional letter). The sentence final word appeared for 300 ms (not adjusted to word length to keep presentation time stable) and was followed by an 800 ms interstimulus interval before the fixation cross reappeared. The longer overall display time of word plus interstimulus interval for the target words was chosen in order to reduce noise during target word presentation for the duration of the entire time segment for ERP analysis, and prevent, for example, blinking of participants upon presentation of the fixation cross during the time segment of the late positivities. Participants were unaware of the upcoming implicit memory task while reading the sentences. During the sentence reading task, participant’s EEG was recorded to analyse the effect of the manipulation on the critical word’s N400 amplitude. If N400 amplitudes differed between the two experimental conditions, their effect on implicit memory can be analysed.

2.4 | Implicit memory task

Participants’ implicit memory for the target words was then assessed via a behavioural perceptual identification paradigm adapted from Stark and McClelland’s (2000) work on repetition priming. Each participant was presented with all 120 target words that were used across lists in the sentence reading task. Again, the order of the stimuli was randomized within participant; however, the overall order of part A and B stimuli (as compared with the sentence reading task) stayed the same, to control for strong timing effects (see more detailed explanation in the last section). Participants started each trial with a button press. After a 300-ms interval with a blank screen, a mask (15 #s, two more than the number of letters in the longest target word) was presented for 14 frames immediately followed by the first target word presentation for one frame (1 frame = 16.67 ms). This procedure continued, until the participants pressed a button to indicate that they had recognized the word. With each round, the mask duration was decreased by one frame, whereas the stimulus presentation was increased by one (see illustration in Figure 2). From the participants’ perspective, this set up induced the perception of the word flashing on the screen and becoming clearer with each repetition, which allowed the recording of reaction times corresponding to a word’s perceptual fluency. They were instructed to identify the word as early on and as accurately as possible. After button press, a cue was presented on the screen, prompting participants to type in their response, that is, which word they had recognized. Participants received feedback for incorrect and too slow trials (after 25 repetitions the word became fully visible, and the trial aborted). Only correct trials were analysed.

2.5 | N-back task

After the sentence task, participants completed four blocks of a n-back task (increasing from $n = 1$ to $n = 4$) with numbers to minimize any potential explicit memory effects of sentence reading. Numbers were presented at the centre of the screen for 500 ms with an interstimulus interval of 2000 ms, and the task lasted approximately 10 mins. In the n-back task participants are instructed to respond when a stimulus matches the one that appeared $n$ items before, which increases working memory load with each block (Kirchner, 1958). The n-back task was not designed to be analysed (no practice rounds, too short) but was screened to make sure participants were trying to solve the task (no random button presses, at least some correct responses).
2.6 | Data acquisition and analysis

2.6.1 | EEG recording

During sentence reading, participant’s electroencephalogram (EEG) was recorded continuously from 64 active Ag/AgCl electrodes positioned according to the extension of the international 10–20 system. The EEG was referenced online to the left mastoid and electrode impedances were kept below 5 kΩ. The data were acquired at a sampling rate of 1000 Hz and amplified (BrainVision BrainAmp amplifier with a bandpass filter of 0.016–0.250 Hz and a time constant of 10 s).

2.6.2 | EEG analysis

EEG data were pre-processed and analysed with MATLAB R2020a using the EEGLab (Delorme & Makeig, 2004) toolbox. Data were re-referenced offline to the average of the left and right mastoids. The EEG was filtered with a 0.1-Hz high-pass filter (two-pass Butterworth with a 12 dB/oct roll-off) and low-pass filtered at 30 Hz (two-pass Butterworth with a 24 dB/oct roll-off). The continuous EEG data were then segmented into epochs ranging from −200 to 1000 ms relative to target word onset. A 200-ms baseline was subtracted. Eye blinks were corrected by means of independent component analysis (Jung et al., 2000; Makeig et al., 1997). Bad channels were identified by their variance (in case it exceeded an absolute z-score of 3) and were interpolated using a spherical spline function (Perrin et al., 1989). All segments with values that exceeded ±75 μV at any channel were automatically excluded. N400 data were analysed for a preregistered centro-parietal region of interest (ROI: CPz, Cz, CP1, CPz, CP2, P1, Pz, and P2) within a 300- to 500-ms time window. The frontal post N400 positivity was analysed in a frontal ROI (Fp1, Fp2, AF7, AF3, AF4, and AF8), and the parietal post N400 positivity (P600) was analysed in a parietal ROI (P3, P1, Pz, P2, P4, PO3, POz, PO4, O1, Oz, and O2) in a window between 600 and 800 ms (as this analysis was not preregistered because it only serves to demonstrate the specificity of our results concerning the N400, ROIs and time windows were based on Kuperberg et al., 2020; see supporting information for an additional analysis based on electrode sagitallity and laterality).

2.6.3 | Statistics

We performed a linear mixed-effects model (LMM) analysis using the package lme4 (Bates et al., 2014) as implemented in R (R Core Team, 2018) to investigate the effect of experimental condition on N400 amplitude and implicit memory, as well as the overall correlation between N400 and implicit memory. Due to the skewed distribution of reaction times, correct responses were log-transformed (not preregistered). For all models predicting reaction times, word frequency was added into each model as additional fixed effect (log transformed and scaled). Following the recommendations of Barr et al. (2013), we tried to fit the maximal random effect structure as justified by the design. In the behavioural analysis (implicit memory task data), random intercepts and slopes could be fitted by subjects and random intercepts by item. To aid convergence, we removed random correlations. For the effect of condition (previously expected, unexpected and not seen), Helmert coding was used. For the EEG analysis (reading task data), the full random effects structure was fitted. The conditions (expected, unexpected) were sum coded (−0.5, 0.5). The significance of fixed effects was determined via likelihood ratio tests, comparing the fit of the model to that of a model with the same random effects structure without the respective fixed effect. All analysis scripts can be found on OSF (https://osf.io/wg8nt/).

3 | RESULTS

3.1 | EEG results from reading task

Figure 3 shows the ERP time-locked to the onset of the target word. In line with previous studies, a clear N400 is visible in a centro-parietal ROI and modulated by expectancy condition (Figure 3a). The amplitude of the N400 in response to unexpected words was more negative than for expected words by 1.32 μV (SE = 0.38, t = −3.51, χ² = 10.48, p = 0.001). The experimental manipulation of N400 amplitudes in the reading task was therefore successful. Additionally, we ran an exploratory analysis (i.e., not preregistered) of post N400 positivities related to unexpected sentence continuations (Kuperberg et al., 2020; Van Petten & Luka, 2012). There was a significant effect of expectancy on the late frontal positivity (Figure 3b; expected vs. unexpected: β = 0.64, SE = 0.30, t = 2.09, χ² = 4.22, p = 0.04) but not on the P600 (parietal positivity; expected vs. unexpected: β = 0.46, SE = 0.27, t = 1.71, χ² = 2.86, p = 0.091).

3.2 | Reaction times from implicit memory task

The arithmetic means of the reaction time data show an 84-ms benefit for previously unexpected compared to not
seen and an 81-ms benefit for previously unexpected compared with expected words (Table 3). Log transformed data are depicted in Figure 4. In a LMM analysis on log-transformed data, the behavioural results showed the classic priming effect (not seen words vs. all previously seen words, regardless of condition) on reaction times ($\beta = 0.02$, $SE = 0.01$, $t = 3.26$, $\chi^2 = 9.15$, $p = 0.0022$) that is well established in literature. Critically, previous expectancy of the words influenced reaction times: Previously unexpected words were recognized faster than previously expected words ($\beta = 0.03$, $SE = 0.01$, $t = 4.62$, $\chi^2 = 16.47$, $p < 0.001$). There was an additional effect of word frequency ($\beta = 0.02$, $SE = 0.01$, $t = 3.72$, $\chi^2 = 16.47$, $p < 0.001$). Error rates did not differ between conditions (Table 3).

|                | Unexpected | Expected | Not seen |
|----------------|------------|----------|----------|
| Reaction time in ms | 1932 (511) | 2013 (549) | 2016 (558) |
| Error rate in %     | 4.7 (4.0%) | 4.2 (3.5%) | 4.9 (4.4%) |

Critical, previous expectancy of the words influenced reaction times: Previously unexpected words were recognized faster than previously expected words ($\beta = -0.03$, $SE = 0.01$, $t = -4.62$, $\chi^2 = 16.47$, $p < 0.001$). There was an additional effect of word frequency ($\beta = -0.02$, $SE = 0.01$, $t = -3.72$, $\chi^2 = 16.47$, $p < 0.001$). Error rates did not differ between conditions (Table 3).

3.3 | Direct relationship N400 and reaction times

There was no effect of N400 amplitude during sentence reading on reaction times in the perceptual identification task ($\beta = -0.007$, $SE = 0.004$, $t = -1.66$, $\chi^2 = 2.76$, $p = 0.097$) when controlling for the effect of frequency ($\beta = -0.02$, $SE = 0.01$, $t = -3.05$, $\chi^2 = 13.36$, $p = 0.037$). The slight deviance in frequency effect (with respect to the frequency effect in the main analysis of the behavioural results) is due to the analysis only containing trials.
which survived artefact rejection in EEG pre-processing. The missing effect of N400 amplitudes on reaction times can be explained by the intuitive notion that words which are easier to access in semantic memory (such as, e.g., high frequency words) elicit a lower amplitude N400 (Kutas & Federmeier, 2011; Van Berkum, 2009) and are also easier to identify in, for example, a perceptual identification task. Therefore, the predicted relationship between N400 amplitudes and later implicit memory effects can only be observed when strictly controlling for any confounding variables. This can be achieved by keeping words constant between conditions that differ in elicited N400 amplitudes and counterbalancing the assignment of target words to expected versus unexpected conditions across participants in a Latin-square design as implemented in this experiment. Thus, to get a better understanding if N400 amplitude differences (rather than the categories expected/unexpected) had a direct effect on reaction times, an additional exploratory (i.e., not pre-registered) analysis was conducted and revealed a significant positive correlation between participants’ individual N400 amplitude differences (expected minus unexpected) during sentence reading and the respective difference in reaction times in the perceptual identification task with $r = 0.49$ [95% CI: 0.18, 0.72], $p = 0.003$ (Figure 5a). The data were standardized to prevent influences of pure amplitude or reaction time differences across participants. As the expectancy manipulation in the sentence reading task also manipulated a late frontal positivity, an analogous analysis was conducted and found no support for a correlation between late frontal positivity differences and reaction time differences between participants (Figure 5b; $r = -0.054$ [95% CI: -0.39, 0.29], $p = 0.763$).

FIGURE 4 Reaction times (log transformed) in the perceptual identification task across participants and items by condition, that is, unexpected, expected or not seen during the preceding sentence reading task

FIGURE 5 Correlation of within participant (a) N400 differences (b) late frontal positivity differences (both expected minus unexpected) in the reading task and the respective reaction time difference from the subsequent implicit memory task. Error bands indicate the SEM
This study tested the hypothesis that N400 amplitudes reflect an implicit learning signal driving adaptation by experimentally manipulating N400 amplitudes via word expectancy during sentence reading and subsequently presenting the critical words in a perceptual identification task to measure implicit memory.

**4.1 N400 amplitude and implicit memory**

In line with well-established findings in the literature on the N400, the experimental manipulation successfully influenced N400 amplitudes (Figure 3). If these differences in N400 amplitude induced by the manipulation in fact reflect a prediction-error that drives implicit learning, then the words used in the sentence reading task should differ in their repetition benefit in the subsequent implicit memory task. Words that were unexpected in the sentence reading task (and were therefore processed with a larger prediction error, i.e., N400) should have a larger increase in internally represented probability than words previously encountered in the expected condition. As predicted, the manipulation of N400 amplitudes influenced participants’ implicit memory: Previously unexpected words (larger N400 in reading task) were recognized faster than expected words (Figure 4). Our results suggest that word expectancy does not independently influence both N400 amplitudes and adaptation.

The correlation between N400 amplitude differences and subsequent implicit memory (Figure 5a; but not between post N400 positivity differences and implicit memory, Figure 5b) demonstrates that participants with overall larger N400 effects also exhibit larger adaptation effects, supporting the proposed relationship between N400 amplitudes and adaptation. The results are consistent with the theory that the N400 reflects an implicit semantic prediction error that drives model adaptation (Rabovsky et al., 2018; Rabovsky & McRae, 2014), which predicts unexpected words to lead to larger N400 amplitudes and enhanced implicit memory. More generally, the results are relevant to the broader literature regarding the N400 as a prediction-error and predictive coding in language.

The findings reported here are also in line with results from studies that were not explicitly designed to investigate this question. In a study aiming to investigate ERP correlates of memory encoding, the authors observed a relation between a N400-like negativity for single words presented during the study phase and subsequent implicit memory performance as reflected in a stem completion task during test (Schott et al., 2002). Meyer et al. (2007) found a positive correlation between the amplitude of the N400 at encoding and the size of the early old/new ERP effect at test. Further support for the N400 as a learning signal comes from the field of language learning in early childhood: The presence of N400 effects as well as N400 priming effects seem to be predictive of infants’ later language skills (Friedrich & Friederici, 2006, 2010). Even though these results were not originally interpreted this way, they can be naturally explained by understanding the N400 as an implicit semantic prediction error that drives implicit learning.

Not only the N400, but also a late frontal positive component was influenced by our expectancy manipulation. As there was no relationship between amplitude differences and RT differences, it is very unlikely that frontal positivities, rather than N400 amplitudes have driven the observed adaptation effect. The late positivities could however still correspond to higher order error processing, for example, environmental uncertainty (or unexpected surprise, Yu & Dayan, 2005) influencing higher order parameters (Kuperberg et al., 2020), or model switch to other previously learned models (Gershman & Niv, 2012; Qian et al., 2012). Disentangling these interpretations is beyond the scope of this experiment. As noted above, we only analysed post N400 positivities to demonstrate the specificity of the observed relationship between N400 amplitudes and implicit memory formation.

**4.2 Unpredicted information or prediction violation?**

The Sentence Gestalt Model from which the hypotheses were derived implements implicit prediction error as the amount of unpredicted semantic information, also known as Bayesian surprise (distance between prior and posterior distributions due to new information; Doya et al., 2007; Itti & Baldi, 2009), and not in the sense of a violated prediction per se. In this framework, predicted semantic activation reduces N400 amplitude, rather than a violation triggering an all-or-nothing response. This notion of a graded prediction-error that encodes unpredicted information is not only in line with predictive coding accounts (where prediction errors are “explained away” by predictability) but also recent discussions based on typical N400 findings (Kuperberg, 2016; Kuperberg et al., 2020; Rabovsky & McClelland, 2020; Rabovsky & McRae, 2014). Sentences can differ in their constraint (amount of preceding information that can be used to predict the next word, defined
as the cloze probability of the most used sentence completion), however as discussed in the introduction unexpected words in highly constraining sentences (i.e., a prediction violation) lead to similar N400 amplitudes as words in a low constraining sentence where no predictions could be violated (Federmeier et al., 2007; Kutas & Hillyard, 1984). In both cases—prediction violation in a highly constraining sentence or no prediction in a low constraining sentence—the incoming word delivers a similar amount of new semantic information that was not predicted. This new information is therefore encoded within the model’s prediction error and can be used to update the model. Interestingly, unexpected words in a highly constrained context (no prediction violation) not only elicit similar N400 amplitudes as words in a weakly constrained sentence but they also do not differ in magnitude of the repetition effect (Lai et al., 2021). This finding is in line with the view that it is not the expectation violation per se but rather the process underlying N400 amplitudes (i.e., unpredicted semantic information) that is critical for internal model adaptation and implicit memory. The distinction between defining prediction errors as encoding the amount of unpredicted information and the notion of a prediction error as a prediction violation is important, as it helps to reconcile findings that may seem conflicting (in the sense that one might conclude from them that the magnitude of repetition effects does not depend on prediction errors) but where the apparent conflict is due to the definition of prediction error as prediction violation rather than as the amount of unpredicted information (e.g., Lai et al., 2021). The account of the N400 as reflecting the amount of unpredicted semantic information contrasts with the reported prerequisites for the emergence of the late frontal positivity, which seems to be sensitive to the disconfirmation of specific lexical predictions and a constraining context (even for comparisons with similar N400 amplitudes, e.g., Federmeier et al., 2007; Thornhill & Van Petten, 2012). This points to exiting directions for future research, dissociating the different implementations of prediction errors and their downstream consequences for adaptation and memory.

4.3 Implicit memory formation and learning

The framework of a prediction-error dependent adaptation process that is put forward here describes a continuous lifelong learning process. We see the acquisition of a skill such as language learning as a process of continuous error-driven adaptation to the statistical regularities in the linguistic environment. Initially a baby’s expectations and priors will be very unspecific and broad, leading to large prediction errors (in the sense of unpredicted information) and thus strong adaptation. On the other hand, prediction errors in adults with roughly stable linguistic environments will often be much smaller and their internal representation of probabilities in the linguistic environment will presumably oscillate around the true probability distribution of their linguistic environment, with some overrepresentation based on recent exposure. While it is interesting to note that typical N400 developments across the lifespan (decrease in N400 effects with age: e.g., Atchley et al., 2006; Benau et al., 2011; Holcomb et al., 1992, Kutas & Iragui, 1998) can be captured by the Sentence Gestalt model (Rabovsky et al., 2018), the process thought to underly N400 amplitudes is not specific to language learning in children but is thought to continuously and gradually adapt the language and meaning system of adults to current statistical regularities in their environments.

Linking the N400 to implicit memory is in line with a study on amnesia that demonstrated preserved N400 repetition effects in patients with explicit memory impairment but relatively intact implicit memory (Olichney et al., 2000). It is also important to note that the experiment reported did not investigate novel word learning or explicit learning. While the learning of novel words (Borovsky et al., 2012; Gambi et al., 2021) seems to be influenced by predictability manipulations, sentence constraint (and therefore prediction violation) seems to play a central role (Borovsky et al., 2012). The underlying process therefore likely differs from the proposed implicit adaptation described here. Positive ERP components that are sensitive specifically to these prediction violations might reflect this (later) more conscious process (Kuperberg et al., 2020; Rabovsky et al., 2018).

5 CONCLUSION

The current study manipulated N400 amplitudes via word expectancy in a sentence reading task. This manipulation influenced the repetition priming benefit in a subsequent implicit memory task. Combined, the results suggest that larger N400 amplitudes reflect an implicit learning signal leading to increased adaptation, which entails higher represented probability and thus better accessibility of the respective word in successive tasks. Therefore, the current study found experimental support for a prediction derived from computational modelling work (Rabovsky et al., 2018; Rabovsky & McRae, 2014), namely, that the N400 ERP component reflects an error-based implicit learning signal during language comprehension.
ACKNOWLEDGEMENTS
Funding was provided by an Emmy Noether grant from the German Research Foundation (grant RA 2715/2-1) to Milena Rabovsky. We thank Antonia Heinrich for help with stimulus preparation.

CONFLICT OF INTEREST
The authors declare no conflict of interests.

AUTHOR CONTRIBUTIONS
A.H. and M.R. designed the study. A.H. collected the data. A.H. analysed the data. A.H. and M.R. wrote the manuscript.

PEER REVIEW
The peer review history for this article is available at https://publons.com/publon/10.1111/ejn.15462.

DATA AVAILABILITY STATEMENT
Grand averaged EEG files, data, and scripts needed for statistical analysis and all figures presented can be found on OSF (https://osf.io/wg8nt/). Due to the General Data Protection Regulation (GDPR), raw EEG data are only available from the corresponding author upon request.

ORCID
Alice Hodapp https://orcid.org/0000-0002-7886-0049
Milena Rabovsky https://orcid.org/0000-0001-7729-1027

REFERENCES
Atchley, R. A., Rice, M. L., Betz, S. K., Kwasny, K. M., Sereno, J. A., & Jongman, A. (2006). A comparison of semantic and syntactic event related potentials generated by children and adults. Brain and Language, 99(3), 236–246. https://doi.org/10.1016/j.bandl.2005.08.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language, 68(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. ArXiv:1406.5823

Benau, E. M., Morris, J., & Couperus, J. W. (2011). Semantic processing in children and adults: Incongruity and the N400. Journal of Psycholinguistic Research, 40(3), 225–239. https://doi.org/10.1007/s10573-011-0167-1

Besson, M., Kutas, M., & Petten, C. V. (1992). An event-related potential (ERP) analysis of semantic congruity and repetition effects in sentences. Journal of Cognitive Neuroscience, 4(2), 132–149. https://doi.org/10.1162/jocn.1992.4.2.132

Bornske1-Schlesewsky, I., & Schlesewsky, M. (2019). Toward a neurobiologically plausible model of language-related, negative event-related potentials. Frontiers in Psychology, 10, 298. https://doi.org/10.3389/fpsyg.2019.00298

Borovsky, A., Elman, J. L., & Kutas, M. (2012). Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. Language Learning and Development, 8(3), 278–302. https://doi.org/10.1080/15475441.2011.614893

Brainard, D. H. (1997). The psychophysics toolbox. Spatial Vision, 10(4), 433–436. https://doi.org/10.1163/156856897X000357

Brothers, T., Wlotko, E. W., Warnke, L., & Kuperberg, G. R. (2020). Going the extra mile: Effects of discourse context on two late positivities during language comprehension. Neurobiology of Language, 1(1), 135–160. https://doi.org/10.1016/nol_a_00006

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. Journal of Cognitive Neuroscience, 5(1), 34–44. https://doi.org/10.1162/jocn.1993.5.1.34

Buchner, A., & Wippich, W. (2000). On the reliability of implicit and explicit memory measures. Cognitive Psychology, 40(3), 227–259. https://doi.org/10.1006/cogp.1999.0731

Chase, H. W., Swainson, R., Durham, L., Benham, L., & Cools, R. (2010). Feedback-related negativity codes prediction error but not behavioral adjustment during probabilistic reversal learning. Journal of Cognitive Neuroscience, 23(4), 936–946. https://doi.org/10.1162/jocn.2010.21456

Cofer, C. N. (1967). Conditions for the use of verbal associations. Psychological Bulletin, 68(1), 1–12. https://doi.org/10.1037/h0024713

Cohen, M. X., & Ranganath, C. (2007). Reinforcement learning signals predict future decisions. Journal of Neuroscience, 27(2), 371–378. https://doi.org/10.1523/JNEUROSCI.4421-06.2007

Coltheart, M., Davelaar, E. J., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), Attention and performance VI. Proceedings of the sixth international symposium on attention and performance, Stockholm, Sweden, July 28–August 1, 1975 (pp. 535–555). Hillsdale: Lawrence Erlbaum Associates.

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

Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. Philosophical Transactions of the Royal Society: B: Biological Sciences, 369(1634), 20120394. https://doi.org/10.1098/rstb.2012.0394

Delong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentences. 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. Journal of Neuroscience Methods, 134(1), 9–21. https://doi.org/10.1016/j.nineup.2003.10.009

Doya, K., Ishii, S., Pouget, A., & Rao, R. P. N. (2007). Bayesian brain: Probabilistic approaches to neural coding. MIT Press.

Elman, J. L. (1990). Finding structure in time. Cognitive Science, 14(2), 179–211. https://doi.org/10.1207/s15516709cog1402_1

Federmeyer, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. Journal
Hagoort, P., Baggio, G., & Willems, R. M. (2009). Semantic unification. In M. S. Gazzaniga (Ed.), The cognitive neurosciences (pp. 819–836). MIT Press.

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

Henson, R. N. A. (2003). Neuroimaging studies of priming. Progress in Neurobiology, 70(1), 53–81. https://doi.org/10.1016/S0301-0082(03)00086-8

Holcomb, J. P., Coffey, S. A., & Neville, H. J. (1992). Visual and auditory sentence processing: A developmental analysis using event-related brain potentials. Developmental Neuropsychology, 8(2–3), 203–241. https://doi.org/10.1080/8756569209540525

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. Psychological Review, 109(4), 679–709. https://doi.org/10.1037/0033-295X.109.4.679
Wacongne, C., Changeux, J.-P., & Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *Journal of Neuroscience, 32*(11), 3665–3678. https://doi.org/10.1523/JNEUROSCI.5003-11.2012

Walsh, M. M., & Anderson, J. R. (2012). Learning from experience: Event-related potential correlates of reward processing, neural adaptation, and behavioral choice. *Neuroscience & Biobehavioral Reviews, 36*(8), 1870–1884. https://doi.org/10.1016/j.neubiorev.2012.05.008

Wlotko, E. W., Lee, C.-L., & Federmeier, K. D. (2010). Language of the aging brain: Event-related potential studies of comprehension in older adults. *Lang & Ling Compass, 4*(8), 623–638. https://doi.org/10.1111/j.1749-818X.2010.00224.x

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron, 46*(4), 681–692. https://doi.org/10.1016/j.neuron.2005.04.026

**SUPPORTING INFORMATION**
Additional supporting information may be found in the online version of the article at the publisher’s website.

**How to cite this article:** Hodapp, A., & Rabovsky, M. (2021). The N400 ERP component reflects an error-based implicit learning signal during language comprehension. *European Journal of Neuroscience, 54*(9), 7125–7140. [https://doi.org/10.1111/ejn.15462]