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

Alice Hodapp (University of Potsdam)
Milena Rabovsky (University of Potsdam)
alice.hodapp@uni-potsdam.de

The functional significance of the N400 is still actively debated. Recent verbal theories (Bornkessel-Schlesewsky & Schlesewsky, 2019; Kuperberg et al., 2019) and computational models of the N400 (Fitz & Chang, 2019; Rabovsky et al., 2018; Rabovsky & McRae, 2014) have increasingly focused on a prediction error perspective. One interpretation of these prediction errors proposes that unpredicted input elicits a prediction error which is used as a basis to update the internal model to make better predictions in the future. This process is thought to underlie a lifelong learning process continuously adjusting internal estimates to the statistics of the (current) environment. Interpreting the N400 as a prediction error driving internal model adaptation can explain N400 reduction in repetition paradigms (as observed e.g., by Besson et al., 1992) as reflecting smaller prediction errors following adjusted probability estimates. Interestingly, these reductions are larger for greater initial N400 amplitudes, in line with the proposed internal update based on the N400. Repetition priming has repeatedly demonstrated increased behavioral performance by previous exposure to stimuli in implicit memory tasks. If larger prediction errors (i.e., N400) lead to greater adaptation, this should be reflected in behavioral implicit memory measures.

Here, we investigated this prediction by experimentally manipulating the N400 via target word predictability (cloze probability) in a sentence reading task, while recording participants' EEG. The same target words appeared in the expected (high-cloze) and unexpected (low-cloze) condition in a latin-square design. After a short break to avoid explicit memory effects, participants were presented the critical words from the reading task in an implicit memory task (perceptual identification task: e.g., Stark & McClelland, 2000). We analyzed our data using linear mixed effects models with random by participants and by items intercepts and slopes and included word frequency as an additional fixed effect.

As expected, the manipulation of cloze probability in the sentence reading task influenced N400 amplitude in response to the critical word. Previously unexpected words did not only elicit larger N400 amplitudes (Figure 1A; $\hat{\beta}$ = 1.32 µV, *SE* = .38, $\chi 2$ = 10.48, *p* = .001) than expected words but were also recognized faster in the subsequent implicit memory task (Figure 1B: $\hat{\beta}$ = -0.03, *SE* = 0.01, χ^2 = 16.47, *p* < .001). Participants' N400 differences in the expected minus unexpected condition during sentence reading and the respective reaction time differences in the subsequent implicit memory task (both standardized) were analyzed and demonstrated that participants with greater N400 amplitude difference also showed a greater implicit memory benefit for previously unexpected words (Figure 2: $\hat{r}$ = 0.49 [95 % CI: 0.18, 0.72], *p* = .003). As an exploratory analysis we additionally investigated post N400 positivities related to unexpected sentence continuations (Kuperberg et al., 2019; Van Petten & Luka, 2012) and while there was a significant late frontal positive ERP component in our EEG data ($\hat{\beta}$ = 0.64, *SE* = 0.30, χ^2 = 4.22, *p* = .04), no support for a correlation between its amplitude and implicit memory differences was found ($\hat{r}$ = -0.054 [95 % CI: -0.39, 0.29], *p* = .763). We conclude from the results that experimentally manipulated N400 amplitudes lead to increased implicit memory formation. This is in line with the interpretation of the N400 as a prediction error that drives adaptation and learning. The correlation between within-participant differences makes it unlikely that expectancy independently manipulated N400 amplitude and implicit memory. Rather our findings support a prediction derived from computational modelling work, namely that the N400 ERP component reflects a learning signal during language comprehension. This finding is of relevance for predictive processing theories in language.

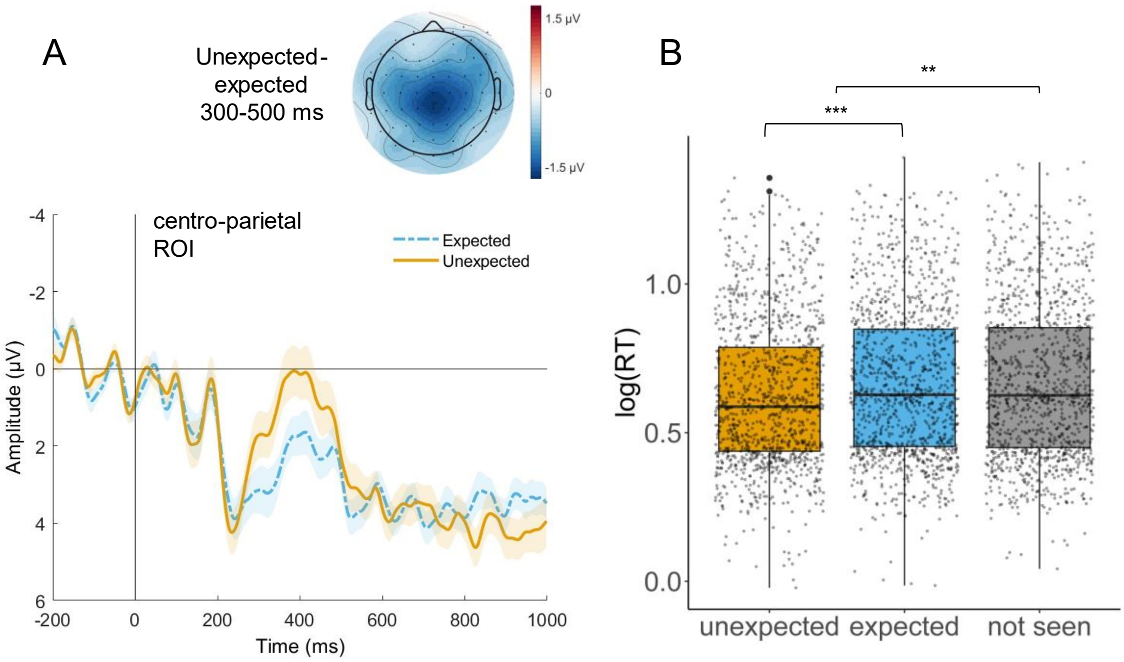

Figure 1. (A) Grand-average waveforms (n = 33) at centro-parietal ROI and topography for the 300-500 ms time window. Negative values are plotted upwards. Error bands indicate the SEM. (B) Reaction times (log transformed) in the implicit memory task (perceptual identification) across participants and items by condition, i.e., unexpected, expected or not seen during the preceding sentence reading task.

References

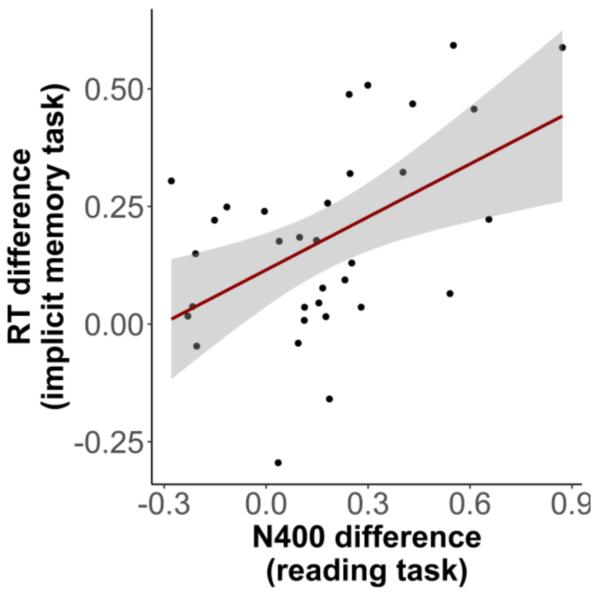

Figure 2. Correlation of within participant N400 differences (expected minus unexpected) in the reading task and the respective reaction time difference from the subsequent implicit memory task. Error bands indicate the SEM.

Besson, M., Kutas, M., & Petten, C. V. (1992). Journal of Cognitive Neuroscience.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2019). Frontiers in Psychology.

Fitz, H., & Chang, F. (2019). Cognitive Psychology.

Kuperberg, G. R., Brothers, T., & Wlotko, E. W. (2019). Journal of Cognitive Neuroscience.

Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2018). Nature Human Behaviour.

Rabovsky, M., & McRae, K. (2014). Cognition.

Stark, C., & McClelland, J. (2000). Journal of Experimental Psychology: LMC

St. John, M. F., & McClelland, J. L. (1990). Artificial Intelligence.

Van Petten, C., & Luka, B. J. (2012). International Journal of Psychophysiology.