bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

N400: Event-Related Potentials in an Associative Word Pair Learning Paradigm

Maryam Farshad[a*], Yuri G. Pavlov[a,b], Boris Kotchoubey[a]

[a]Institute of Medical Psychology and Behavioral Neurobiology, University of Tübingen, Germany

[b]Department of Psychology, Ural Federal University, Ekaterinburg, Russian Federation

*maryam.farshad@student.uni-tuebingen.de

pavlovug@gmail.com

boris.kotchoubey@uni-tuebingen.de

Abstract

The study investigated the effect of unintentional learning of semantically unrelated word pairs on event-related brain potentials. Two experiments were conducted, in whose acquisition phase participants listened to five word pairs of semantically unrelated words, each pair being repeated twenty times. In the test phase of Experiment I, these “old” pairs were presented mixed with “new” pairs containing other words. In the test phase of Experiment II a third condition was added in which the first word in a pair was one of the words presented during acquisition but the second word was new. In both experiments the second word in new word pairs elicited a large negative deflection identified as an N400, whose amplitude in Experiment II did not depend on whether the first word was old or new. The result is primarily caused by building new associations between words, while the repetition effect played a relatively minor role.

Keywords: Associative Word Pair Learning, N400, ERP, Old Pairs, New Pairs

1. Introduction

Event-related potentials (ERPs) provide a technique of extracting information in the order of milliseconds on the time course of cognitive processes. The averaged ERP waveforms consist of a sequence of positive and negative voltage deflections (Luck, 2005). Particularly, a negative deflection peaking around 400 ms post-stimulus was labeled N400 (Kutas & Hillyard, 1980). More precisely, it is negative-going at particular scalp locations relative to a specific reference derivation (e.g., posterior sites relative to recordings behind the ears), referred to a 100- or 200-ms pre-stimulus baseline (Kutas & Federmeier, 2011). It is a characteristic part of the normal

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

brain response to words and other meaningful (or potentially meaningful) stimuli, but is typically suppressed whenever the word is strongly prepared by its context (e.g., "I drink my coffee with cream and *sugar*").

In contrast to the N400 component, the N400 *effect* is defined as a *difference* between the responses to context-appropriate (with a suppressed N400) and context-inappropriate (with a manifested N400) stimuli. This effect is pronounced in response to violations of semantic expectancies. For example, if sentences are presented one word at a time on a video monitor (or acoustically, word-by word in a loudspeaker), a large N400 is elicited by the last word of the sentence when this word violates the expectancy created during the sentence ("incongruent ending" e.g., "I drink my coffee with cream and *dog*"), as compared with the word that agrees with the expectancy ("congruent ending", see above) (Balass, Nelson and Perfetti,2010; Kutas & Federmeier, 2011).

In addition to the above depicted sentence paradigm with congruent versus incongruent endings, similar results were obtained in experiments with word pairs. The N400's amplitude is typically larger for words that are semantically unrelated to the preceding word than for words that are semantically related (see Kutas & Hillyard, 1984). Its amplitude is inversely related to the degree of semantic expectancy and the modulation of the N400 amplitude is referred to as the semantic priming effect (Brualla, Romero, Serrano, & Valdizán, 1998). The generation of this expectancy requires access to the semantic aspects of the prime (Chwilla, Brown, & Hagoort, 1995).

Borovsky, Elman and Kutas (2012) examined the impact of contextual constraint on word learning. They set out ERP experiments as participants performed a lexical decision task in which known and newly unknown learned words served as primes for *known words* that were either semantically related or unrelated to prime words. Although based on previous findings, the N400 component is sensitive to acquisition of word meanings after multiple exposures, the authors addressed a fundamental process by which adults learned novel words after only a single exposure and the effect of old and new presentation of words in constraint sentence. A reduction of the N400 amplitude to semantically related target words, was observed after known primes and after novel unknown primes in a strongly constraining context, but not after novel unknown primes in a free context. The N400 effect after novel words is driven by an association between the novel word's meaning and its relationship to other words in the lexicon.

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Cruse et al. (2014) applied the N400 word pair paradigm to test preserved linguistic function in non-communicative patients. In the first experiment, they used strongly semantically similar word pairs. For the second experiment strongly associated word pairs were selected. N400 effect was obtained in both experiments being significantly larger in the second experiment than in the first one.

Ortu, Allan and Donaldson (2013) attempted to separate semantic association from semantic relatedness. They employed three classes of word-pairs including pairs with moderate association, high association and without association. The degree of semantic relationship did not differ between the three classes. Both moderately and highly associated word pairs elicited a suppressed N400 as compared to unrelated word pairs. Also Rhodes and Donaldson (2008) manipulated the presence of associative and semantic relationships. They asked the participants to read and remember visually presented word pairs that shared either an associative relationship, a semantic relationship, both associative and semantic relationship, or no relationship whatsoever. The data revealed a strong N400 effect for both associated and associated+semantically related word pairs compared with unrelated pairs. Like Ortu et al's (2013) study, these findings suggested that the association between words, rather than the semantic relationship, is critical for the processing of meaning.

As Perea and Rosa (2002) declared, semantic complexity is commonly studied by investigating words pairs' association like sheet-PAPER, health-DISEASE and lightning-THUNDER. Different kinds of semantic relationships are sought; some of them describe a part-whole relation between lexical items as PART(*X*, *Y*), where *X* is part of *Y*. For instance, the compound *door knob* contains the part–whole relation: PART (*knob*, *door*) (Girju, Badulescu, & Moldovan, 2006).

In contrast to semantic relationship, an association between two words is due to their spatial and/or temporal co-occurrence in the real world and/or in language which is an important driving force in learning and forming semantic representations (McRae, Khalkhali & Hare, 2012). Ortu et al., (2013) emphasized that the N400 effect is sensitive to how often two stimuli were presented contiguously within an individual's past experience. Because of their co-occurrence, the term word-pair describes any variable based on concepts that are paired with other concepts and the two words occur together in similar context. An example of such co-occurring concepts is the pair rock and roll. (Lancia, 2007; Buchanan, Holmes, Teasley & Hutchison, 2013). Although a rock can roll, their main association are related to the music

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

style. Thus, although meanings of two words are not necessarily related to each other, one of them can strongly reminds us of the other one.

However, some semantically related word pairs do not have shared association. Examples are semantic-only antonyms (ORDER-*CHAOS*) or category-coordinate pairs (IRON-*COPPER*). In a study by Perea et al. (2002) priming for pairs that are associatively and semantically related appears to have a faster rise time than for semantic-only pairs (Perea & Rosa, 2002). From this perspective, the N400 effect does not reflect the transient activation of semantic-only knowledge. Rather, the N400 effect can also reflect the acquisition of meaning via associations formed and maintained by the proximity of distinct elements within one's experience (Ortu et al., 2013).

Balass et al. (2010) investigated the ERP differences in word learning as a result of the type of word experience. Their experiment included a learning phase and a testing phase. During the former, participants were presented with a total 105 rare English words. The testing phase included the presentation of 105 word pairs, in thirty-five of which the first word was one of the words presented in the learning phase. The first word in other 35 pairs was a new (unlearnt) rare word, and the first word in the resting 35 pairs was a new but familiar low-frequency word. Participants were asked to respond whether the two words were semantically related or not. P600 amplitudes to the first word in high-skilled readers were larger for learnt than for unlearnt words. ERP differences to the second word were observed in the N400 time window. The amplitudes were more positive to both learnt rare words und unlearnt familiar words than to unlearnt rare words. The suppression of the N400 to semantically related words was observed only for pairs in which the first word was either learnt or unlearnt but familiar. This effect of relatedness was not observed for word pairs containing rare unlearnt words. Thus, the P600 and N400 results indicated that word experience affected recognition and meaning retrieval processes.

An alternative approach regards similar phenomena as an effect of words repetition, i.e., as based not on associations but on the mere repetition. According to Bentin and Peled (1990) in tasks such as lexical decision and semantic classification, ERPs varied as a function of the frequency of their previous repetition. The ERP repetition effect takes the form of a widely distributed, long-latency, positive-going shift in the ERPs for repeated items, relative to the ERPs elicited by the first presentation (Rugg, Doyle, & Holdstock, 1994).

More recently, Li, Guo and Jiang (2008) compared repeated learning effects during encoding and retrieval on pairs of phrases. Their experiment consisted of three parts: encoding, memory

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

interference, and memory retrieval. During the initial encoding learning part, participants were told to judge whether the target phrase contained a noun or a verb. After the first presentation, half of the previously seen phrases were presented again. The participants were asked again to distinguish between nouns and verbs. During the memory interference part, a three-digit number was presented and participants were asked to continually subtract three from this number and provide verbal responses. In the memory retrieval part, each encoding stage was followed by one of two retrieval tests: a lexical decision test (word versus non-word judgments) or a recognition test (new versus studied judgments). The results showed that during encoding, repetition made the target task of judging noun/verb easier and faster. During retrieval, repeated words were responded to faster than non-repeated words in both lexical decision and recognition tests. The ERP effects of repetition occurred during both encoding and retrieval at similar scalp locations but started earlier and lasted longer during encoding than retrieval.

Importantly, the data on the ERP repetition effect (Rugg et al., 1994) allow to make a similar prediction as the data on semantic effects. Because learnt words have been repeated, their ERPs should contain a larger positivity. The difference is, however, that repetition would result in more positive ERPs to learnt words than to unlearnt words.

Although studies have revealed large positive-going ERP response to repeated words, little is known about the effects of repetition on *semantically unrelated word pairs*. Since word association suppresses the N400, we asked whether the same effect can be attained by originally unrelated and non-associated words learnt together in the course of unintentional learning. Thus we compared ERPs to unrelated word pairs that have been recently unconsciously learnt ("old" pairs) and to other unrelated word pairs that were presented for the first time ("new pairs").

Therefore, the aim of our study was to answer the questions, (i) whether unintentional learning of semantically not associated word pairs would result in N400 suppression similar to that in semantically associated pairs, and (ii) if yes, whether this effect is to be explained by increased familiarity of repeatedly presented words or is based on building new associations between the members of a pair. If we expect results similar to those in Balass et al. (2010), consistent repetition of unrelated word pairs may make them acting like related word pairs, that is, the N400 to the second word in a pair would be reduced. Thus, in the first experiment we expected that the N400 to the second word in a pair to be smaller in learnt word pairs than in new word pairs. To clarify the latter issue in the second experiment, we introduced a novel element in the paradigm: in addition to old and new word pairs, we had pairs where the old (well-learnt) prime was combined with a new target.

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2. Material and methods

2.1. Participants

Twenty-two subjects participated in this study. The group comprised thirteen females and nine males with a mean age=25.23 (range=19-33, standard deviation: 12.45) recruited from the university of Tubingen student population. None of them reported visual, auditory or neurological deficits. According to the Edinburgh Handedness Scale (Oldfield, 1971), seventeen participants were right handed, four left handed and one participant was ambidextrous. They were either native German speakers or came to Germany before the age of 6 and started their primary school here. Informed consent was obtained from all subjects included in the study. They received expense allowance of 8 €/h for their participation.

2.2. Procedure and stimulus material

We carried out two learning experiments, each entailing two phases. The result of learning in the first (acquisition) phase should be manifested in the second (test) phase.

Experiment 1 contained an acquisition phase, in which five word pairs were presented 20 times each, ten times in one order (e.g., cat/table) and ten times in the reverse order (e.g., table/cat). In the test phase of Experiment 1 two hundred word pairs were presented including 100 pairs that had been presented in the acquisition phase ("old" pairs), and 100 completely new pairs. The acquisition phase of Experiment 2 was identical to the acquisition phase of Experiment 1, but the five employed word pairs did not repeat any word from Experiment 1. The test phase included 300 word pairs subdivided into three categories: 100 "old" pairs and 100 new pairs were analogous to those in Experiment 1. An additional category involved 100 word pairs, whose first word was the same as one of the "old" words, learnt in the acquisition phase. However, the second word in the pair was new. These pairs were designated as "old-new pairs". There was no significant difference between "old" and "new" words in terms of word frequency, the average length in ms and the average number of syllables.

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The two above-described EEG experiments were preceded by a pilot paper-and-pencil experiment with fifteen students of medicine and psychology (nine females, all native German speakers). They were presented with about 1000 different word pairs and were required to judge the association strength within each pair using the scale from 0 (no association at all) to 8 (very strong association). Moreover, they were instructed that they should regard word pairs such as man/woman and cat/mouse as "8". After this, only word pairs whose association strength was never assessed above 3 were employed in the two EEG experiments.

In both experiments word pairs were presented in a pseudorandom order, and one old pair was never presented twice in a row. Furthermore, in the test phases not more than four pairs of the same category (i.e., old, new, old-new) were presented in immediate succession. The interval between the offset of the first word and the onset of the second word within each pair was 100 ms. Due to the variable length of the words, the onset-to-onset interval varied from 500 to 1100 ms. The second word was followed by a motor response (see below), and the next word pair started 1.8 s after the motor response. In each experiment the test phase started immediately after the end of the acquisition phase without a break. There was a break of about 30 minutes between the two experiments, during which participants were busy with different kinds of tasks.

Participants were asked to compare the number of syllables in the two words of each pair. If both words had the equal number of syllables, participants had to press one of the keyboard buttons (left or right Ctrl buttons on a standard keyboard), and if they had different number of syllables, the other button. The side of the response was counterbalanced among the participants. The task was irrelevant with respect to the aim of the experiments. It was used just to increase attention to the words. Because the mean length of experiment 1 was around 19 minutes and the experiment 2 was about 27 minutes, we were concerned that participants' attention would decline if no overt response is required.

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2.3. EEG recording and ERP analysis

The EEG was recorded from 64 channels online referenced to Cz. The EEG was acquired with 1000 Hz sampling rate with low-pass 280 Hz filter. The active electrodes (EasyCap) were placed using the 10-10 percent system. BrainVision Analyzer Version 2.1.2.327 was used for preprocessing of EEG. First the data were downsampled off-line to 256 Hz and then re-referenced to a common average reference. After that, offline low pass 30 Hz and 0.1 Hz high-pass filters were applied. Bad channels were replaced using spherical interpolation. An Independent Component Analysis (ICA) was employed to separate and remove activity related to ocular artifacts. Finally, the data were again re-referenced to average mastoids to warrant the comparison with the published results, most of which have been reported with mastoid reference. Three different categories of epochs have been defined, including (old epochs, new epochs and old-new epochs). Each epoch started -500 ms before a stimulus and lasted for 1500 ms after it (i.e., the total of 2000 ms). The baseline was defined as the average amplitude between -200 ms and 0 ms. All these processes were carried out in every segment in every experiment separately. Based on our experiment design, we had seven segment categories in each of the 22 subjects. Finally, ERPs were averaged for first and second words of all categories separately and the average and standard deviation were computed.

The mean ERP amplitude was measured in the time windows of 250-550 ms (N400), 550-750 ms (Late 1 window) and 750-1000 ms (Late 2 window). The statistical analysis was conducted using SPSS version 23 (IBM corp.) using Analysis of variances (ANOVAs) within the General Linear Model. Specifically, two ANOVAs were employed. The first ANOVA over midline electrodes included the factors Site (5 levels: Fz, FCz, Cz, CPz, Pz) and Condition. The factor Condition included three levels in Experiment 1 (acquisition phase; old pairs in the test phase; new pairs in the test phase) and four levels in Experiment 2 (acquisition phase; old pairs, new

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

pairs, and old-new pairs in the test phase), see Figures 1, 2, 4, 8, 9 for traces of the ERP signals at midline electrodes.

The second ANOVA over lateral electrodes involved the factors Condition (identical to the first ANOVA), Region of Interest (ROI: 4 levels) and Side (2 levels, i.e. left versus right). For this sake we built average values for eight ROIs (4 ROIs on each side) including: left anterior ROI: F1, F3, FC1, FC3, FC5, FT7, left central ROI: C1, C3, C5, T7, left temporo-parietal ROI: CP1, CP3, CP5, TP7, P5, P7, left posterior ROI: P1, P3, PO3, PO7, O1, right anterior ROI: F2, F4, FC2, FC4, FC6, FT8, right central ROI: C2, C4, C6, T8, right temporo-parietal ROI: CP2, CP4, CP6, TP8, P6, P8 and right posterior ROI: P2, P4, PO4, PO8, O2.

The significance level was set at .05, and for all effects with more than two levels degrees of freedom were corrected for non-sphericity using Greenhouse-Geisser epsilon. Below, corrected degrees of freedom will be reported.

3. Results

3.1. Experiment I, First word

3.1.1. N400 (350-550 ms time window)

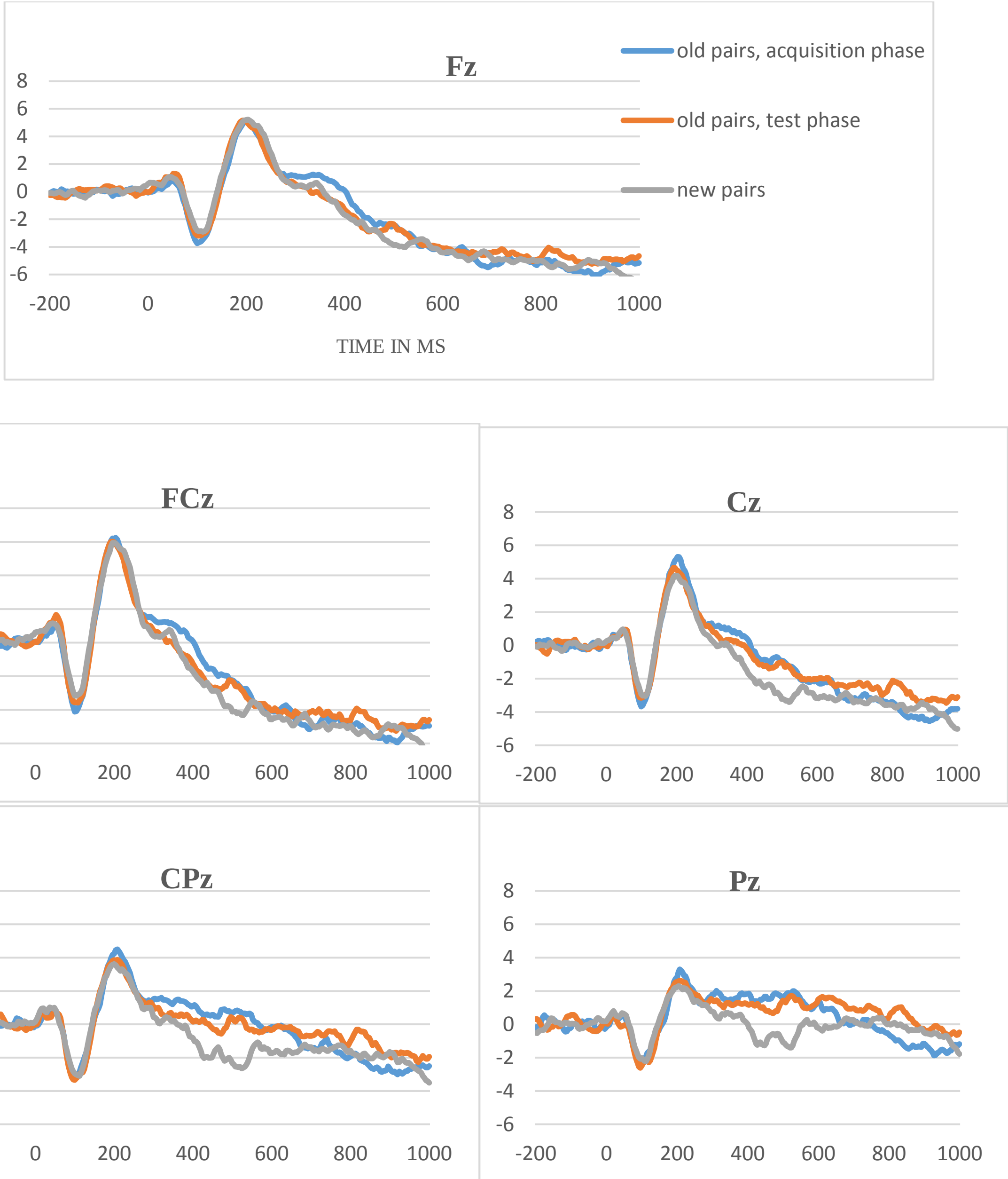
3.1.1.1. midline electrodes

As ERP waveforms in Figure 1 displays, a significant main effect of Condition ($F(1, 30)$ = 6.24, p = .010, η^2 =.22) resulted from the N400 amplitude for old pairs in the test phase (mean 0.60 µV) being significantly larger than in the acquisition phase (mean 0.49 µV, p = .044), and for new pairs (mean 2.26 µV) it was larger than for old pairs in the acquisition phase (p = .009). A significant main effect of Site in the midline electrodes ($F(1, 27)$ = 12.44, p = .001, η^2 =.37) resulted from the fact that the amplitude at Fz was larger than at CPz (p = .006) and Pz (p = .001). These two changes led to a significant interaction between Condition and Site ($F(3, 69)$ = 3.44, p = .018, η^2 =.14), which is shown in Figure 2. The analysis revealed that although the N400 was largest at Fz and FCz, between-condition differences at these sites did not reach significance (all p > .05). At Cz, the N400 for new pairs was larger than in the acquisition phase (p = .008), and at Pz it was larger for new pairs than for old pairs in both

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

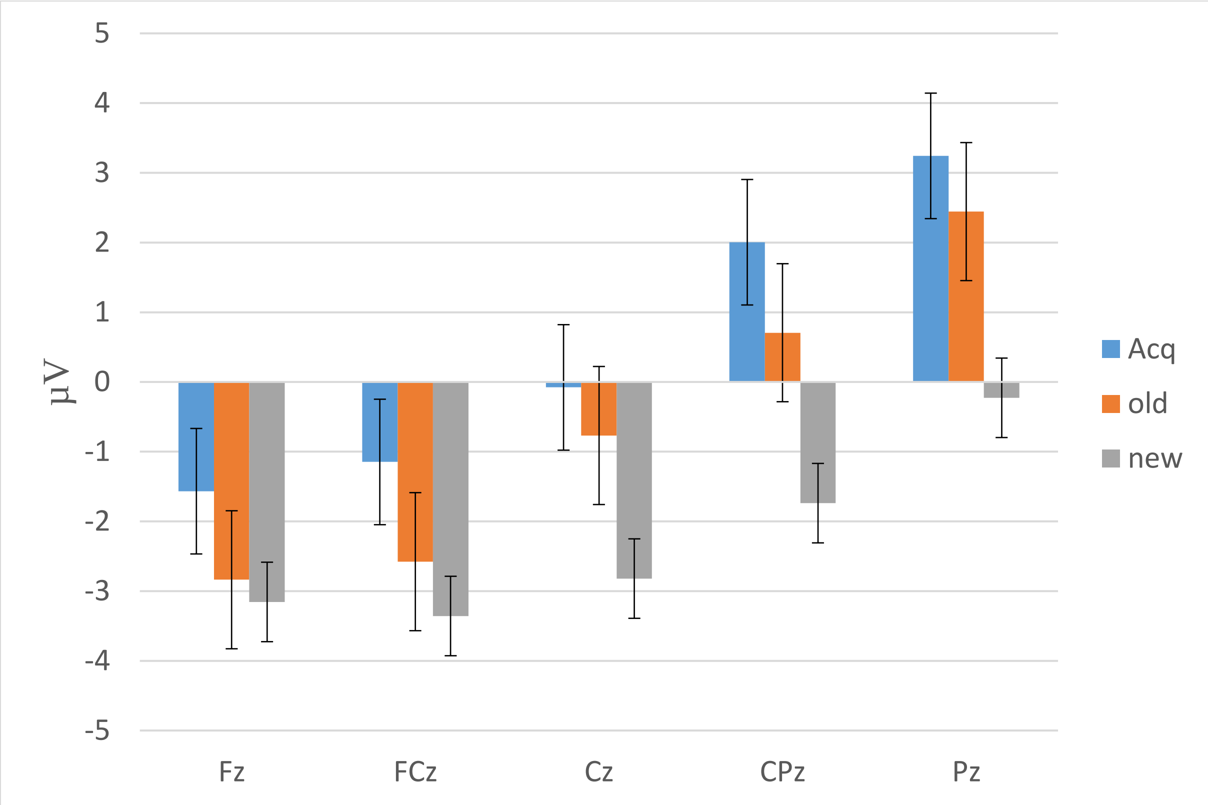
acquisition (p = .001) and test phase (p = .002). Finally, at CPz all three conditions differed significantly from each other: the N400 for old pairs was larger in the test phase than in the acquisition phase (p = .047), and for new pairs it was larger than for old pairs in both acquisition phase (p = .002) and test phase (p = .008).

Figure 1 ERP grand average waveforms at midline electrodes (Fz, FCz, Cz, CPz and Pz) to the 1st word in the 1st experiment, averaged with a 200 ms baseline. Negativity is plotted downwards. Blue lines: Condition 1 (old pairs in the acquisition phase), orange lines: Condition 2 (old pairs in the test phase), grey lines: Condition 3 (new pairs).



bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2 Mean N400 values in all Conditions at midline electrodes. Larger amplitudes are plotted downwards because N400 is a negative wave.



3.1.1.2. lateral electrodes

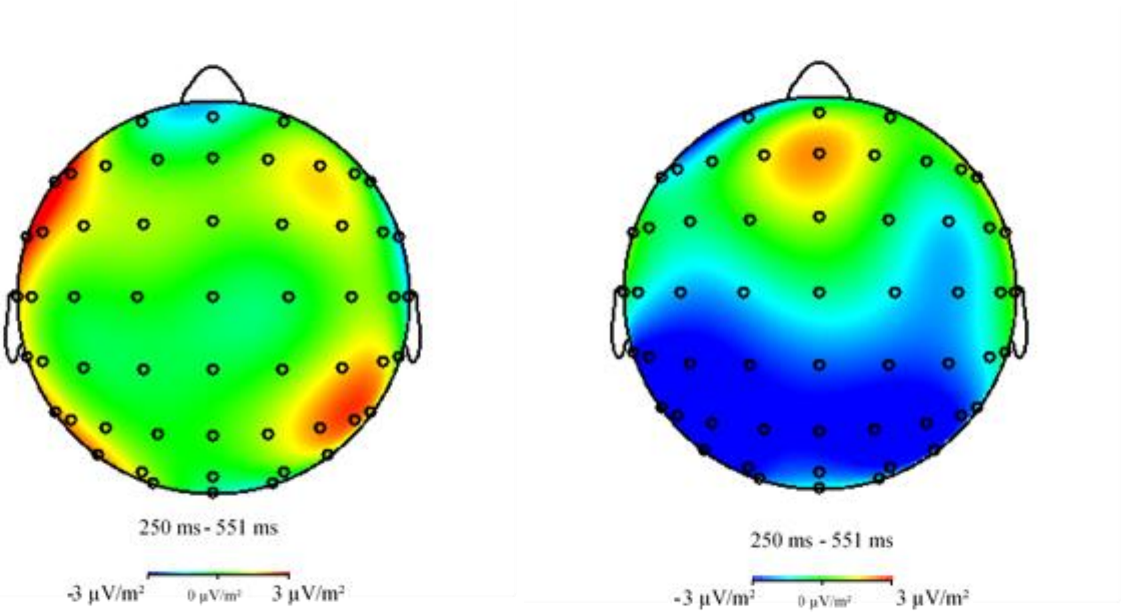
As can be seen in Figure 3, like at the midline, the main effect of Condition ($F(1,30) = 7.06$, $p = .002$, $\eta^2 = .25$) indicated that the N400 for new pairs in the test phase (mean 1.13 µV) was significantly ($p = .006$) larger than the amplitude in the acquisition phase (mean .96 µV), and also larger than the amplitude for old pairs in the test phase (mean .10 µV, $p = .044$). The amplitude for old pairs in the test phase was significantly larger than in the acquisition phase ($p = .030$). In addition, a significant main effect of ROI ($F(1,28) = 5.48$, $p = .002$, $\eta^2 = .20$) indicated that the amplitude over temporo-parietal ROIs was larger than the amplitude over central ($p = .019$) and posterior ROIs ($p = .023$). In addition, the same result indicated that the amplitude over anterior ROIs was larger than the amplitude over central ($p = .020$) and posterior ROIs ($p = .040$).

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

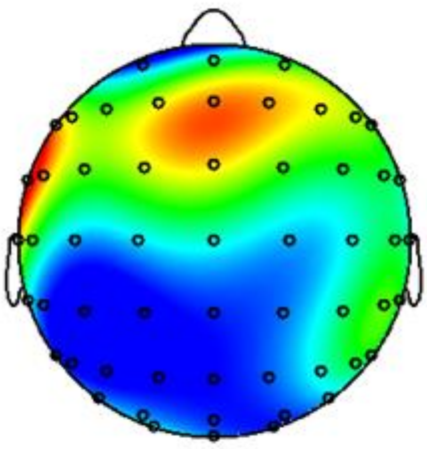
A significant main effect of Side was found (F(1,21) = 5.03, p = .036, η^2 = .19) resulted from the amplitude on the left side being larger than on the right side.

In addition, its interaction with Condition was significant (F(1, 28) = 3.96, p = .027, η^2 = .15), which indicated that in response to old stimuli the amplitude on the left side was larger than on the right side in both acquisition phase (F(1, 21) = 5.08, p = .035, η^2 = .19) and test phase (F(1, 21) = 7.51, p = .012, η^2 = .26). However, the side difference was not significant for the N400 to new stimuli (F(1, 21) = 1.57, p = .223, η^2 = .07).

Figure 3 Difference brain maps for the N400 time window to the 1st word in the 1st experiment. Top left: difference between old pairs in the acquisition phase and the same pairs in the test phase. Top right: difference between new pairs and old pairs in the acquisition phase. Bottom: difference between new pairs and old pairs in the test phase.



bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



3.2. Experiment I, Second word

3.2.1. N400

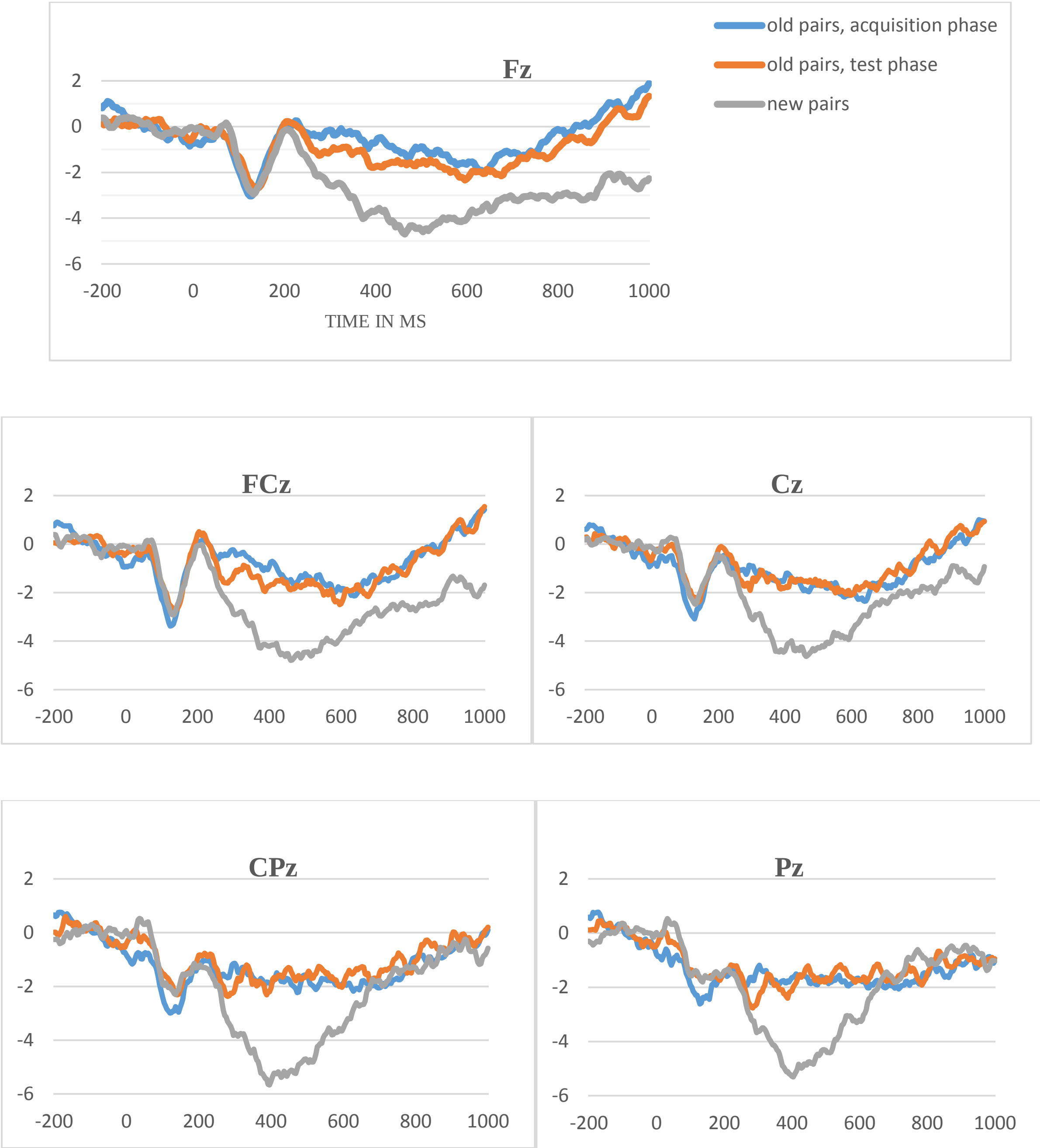
3.2.1.1. midline electrodes

As shown in Figure 4, the ERP waveforms at midline electrodes displays a significant main effect of Condition ($F(1, 39) = 15.11$, $p < .001$, $\eta^2 = .41$) resulting from the amplitude for new pairs being significantly ($p < .001$; mean 8.03 µV) larger than the amplitude for old pairs in the test phase (mean 3.26 µV) and in the acquisition phase (2.68 µV). In contrast, the difference between old pairs in the acquisition phase and in the test phase was not significant ($p = .538$).

Besides this, a significant main effect of Site ($F(1, 37) = 5.32$, $p = .012$, $\eta^2 = .20$) indicated that the amplitude at CPz was larger than at Fz, FCz and Cz (all $p < .05$) and the amplitude at Pz was larger than at Fz, $p = .028$.

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 4 ERP grand average waveforms at midline electrodes (Fz, FCz, Cz, CPz and Pz) to the 2nd word in the 1st experiment, averaged with a 200 ms baseline



3.2.1.2. lateral electrodes

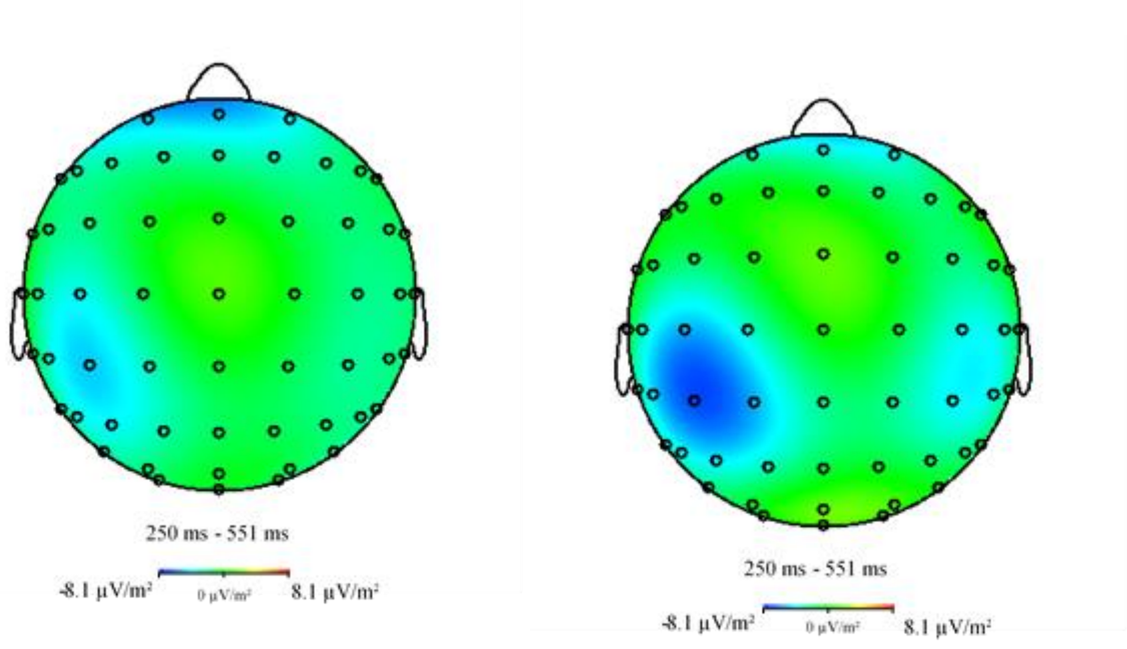
As can be seen in Figure 5, a significant main effect of Condition ($F(1, 39) = 19.07$, $p < .001$, $\eta^2 = .47$) indicated that the amplitude for new pairs in the test phase (mean 6.20 µV) was

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

significantly larger than the amplitude for old pairs (mean 2.23 µV) and in the acquisition phase (mean 1.80 µV) (both $p < .001$). On the other hand, although the amplitude for old pairs in the test phase was larger than in the acquisition phase, the corresponding difference was not significant ($p > .5$).

The amplitude over anterior ROIs was significantly ($p < .001$) larger than over central, temporo-parietal and posterior ROIs, which gave rise to a significant main effect of ROI ($F(1, 33) = 5.36$, $p = .014$, $\eta^2 = .20$).

Figure 5 Difference brain maps for the N400 time window to the 2nd word in the 1st experiment. Left: difference between new pairs and old pairs in the acquisition phase. Right: difference between new pairs and old pairs in the test phase.



3.2.2. Late 1 (550-750 ms time window)

There were no significant effects of Condition at both midline and lateral electrodes.

3.2.3. Late 2 (750-1000 ms time window)

3.2.3.1. midline electrodes

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The amplitude for new pairs in the test phase (mean -3.01 µV) was more negative than the amplitude for old pairs (mean -.20 µV) in the test phase and in the acquisition phase (mean -.29 µV) creating a significant main effect of Condition (df = 1.92 / 40.33, F = 4.62, p =.017, η^2 = .18).

3.2.3.2. lateral electrodes

Like at the midline, the amplitude for new pairs in the test phase (mean -1.95 µV) was more negative than old pairs in the acquisition phase (mean -.19 µV, p = .023) giving rise to a significant main effect of Condition (df = 1.91 / 40.29, F = 3.49, p = .041, η^2 = .14).

3.3. Experiment II: First word

3.3.1. N400

3.3.1.1. midline electrodes

ERP waveforms at midline electrodes (Fz, FCz, Cz, CPz and Pz) to the first word in second experiment. As can be seen in the Figure, a significant main effect of Site ($F(1, 27) = 15.48$, $p < .001$, $\eta^2 = .42$) was due to the fact that the amplitude at Fz and FCz was significantly larger than at Cz ($p = .040$), CPz ($p = .006$) and Pz ($p = .001$). Neither the main effects of Condition nor its interactions were significant.

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 8 ERP grand average waveforms at midline electrodes (Fz, FCz, Cz, CPz and Pz) to the 1st word in the 2nd experiment, averaged with a 200 ms baseline. Blue lines: Condition 1 (old pairs in the acquisition phase), orange lines: Condition 2 (old pairs new pairs in the test phase), grey lines: Condition 3 (new pairs) and yellow lines: condition 4 (oldnew pairs).

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3.3.1.2. lateral electrodes

Like at the midline electrodes, the main effect of Condition and its interactions with other factors were not significant. However, the amplitude over temporo-parietal ROIs was larger than over anterior (p = .013), central (p = .002) and posterior ROIs (p = .013); also the amplitude over both anterior and posterior ROIs were significantly larger than over central ROIs, which led to a significant main effect of ROI ($F(1, 29) = 7.12$, p =.006, $\eta^2 = .26$). In addition, the main effect of Side ($F(1, 20) = 4.62$, $p = .044$, $\eta^2 = .18$) indicated that the amplitude on the left side was larger than on the right side.

3.4. Experiment II, Second word

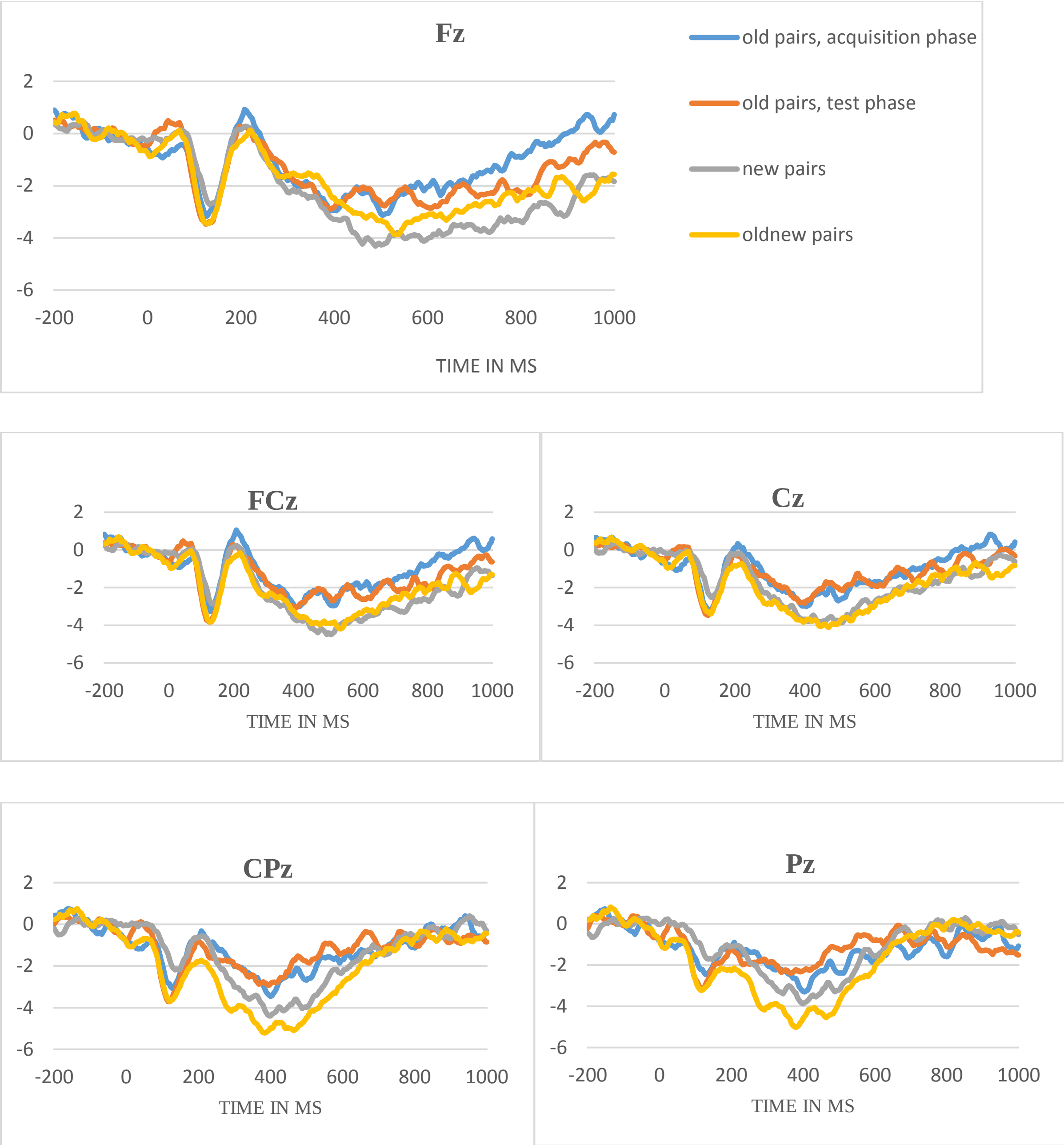
3.4.1. N400 effects (350-550 ms time window)

3.4.1.1. midline electrodes

As the ERP waveforms at midline electrodes (Fz, FCz, Cz, CPz and Pz) to the second word in the second experiment in Figure 9 display, a significant main effect of Condition ($F(2, 43) = 7.88$, $p < .001$, $\eta^2 = .28$) resulted from the fact that the amplitude for old-new pairs (mean 7.31 μV) was significantly larger than for old pairs in the test phase (mean 3.94 μV, $p < .001$) and in the acquisition phase (mean 4.23 μV, $p = .01$). Similarly, the amplitude for new pairs (mean 6.68 μV) was significantly larger than for old pairs in the test phase ($p = .002$) and in the acquisition phase ($p = .017$). While the amplitude for old pairs appeared smaller in the test phase than in the acquisition phase, this difference was not significant ($p = .769$). Further, the difference between the amplitudes for old-new pairs and for new pairs was not significant either ($p = .259$).

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 9 ERP grand average waveforms at midline electrodes (Fz, FCz, Cz, CPz and Pz) to the 2nd word in the 2nd experiment, averaged with a 200 ms baseline.



3.4.1.2. lateral electrodes

As Figure 10 displays, like at the midline, a significant main effect of Condition ($F(2, 43) = 8.27$, $p < .001$, $\eta^2 = .29$) emerged from the fact that the amplitude for old-new pairs (mean 5.20

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

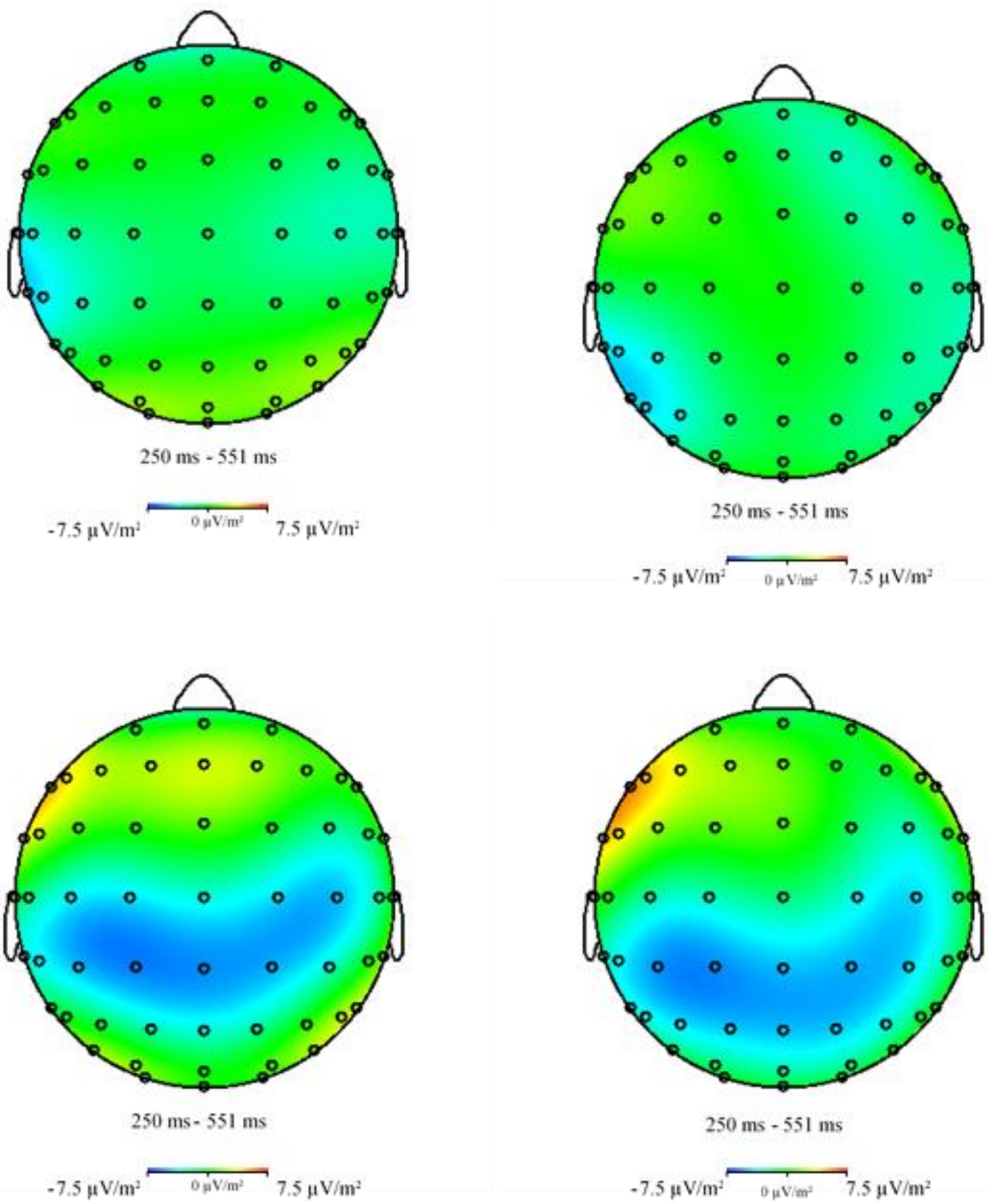
μV) was significantly larger than for old pairs in the test phase (mean 2.67 μV, $p < .001$) and in the acquisition phase (mean 2.96 μV, $p = .007$). The amplitude for new pairs (mean 5.00 μV) was also significantly larger than for old pairs in the test phase ($p = .001$) and in the acquisition phase ($p = .007$). And like at the midline, the difference between old pair in the acquisition phase and in the test phase, and the difference between new pairs and old-new pairs were not significant.

Moreover, the amplitude over anterior ROIs was larger than over central ($p = .048$), temporo-parietal ($p < .001$) and posterior ROIs ($p = .007$) and again, the amplitude over central was larger than over posterior ($p = .034$) ROIs, which led to a significant main effect of ROI ($F(1, 32) = 5.92$, $p = .010$, $\eta^2 = .22$).

Figures 11 and 12 show that the amplitude for new pairs was larger than for old-new pairs on the left side, particularly in the anterior region. In contrast, the amplitude for old-new pairs was larger than for new pairs on the right side, particularly in the central and posterior regions. This resulted in the significant interactions between Condition and ROI ($F(4, 86) = 2.69$, $p = .032$, $\eta^2 = .12$), between Condition and Side ($F(2, 41) = 9.43$, $p < .001$, $\eta^2 = .32$), between ROI and Side ($F(2, 43) = 6.32$, $p = .003$, $\eta^2 = .24$), as well as in a triple interaction of Condition, ROI, and Side ($F(4, 82) = 3.50$, $p = .010$, $\eta^2 = .15$). The apparent differences between the acquisition phase and the same (i.e., old) stimuli in the test phase were not significant in any region.

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 10 Difference brain maps for the N400 time window to the 2nd word in the 2nd experiment. Top left: difference between new pairs and old pairs in the acquisition phase. Top right: difference between new pairs and old pairs in the test phase. Bottom left: difference between old/new pairs and old pairs in the acquisition phase. Bottom right: difference between old/new pairs and old pairs in the test phase.



bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 11 Mean N400 values in all Conditions at left lateral electrodes. Larger amplitudes are plotted downwards because N400 is a negative wave

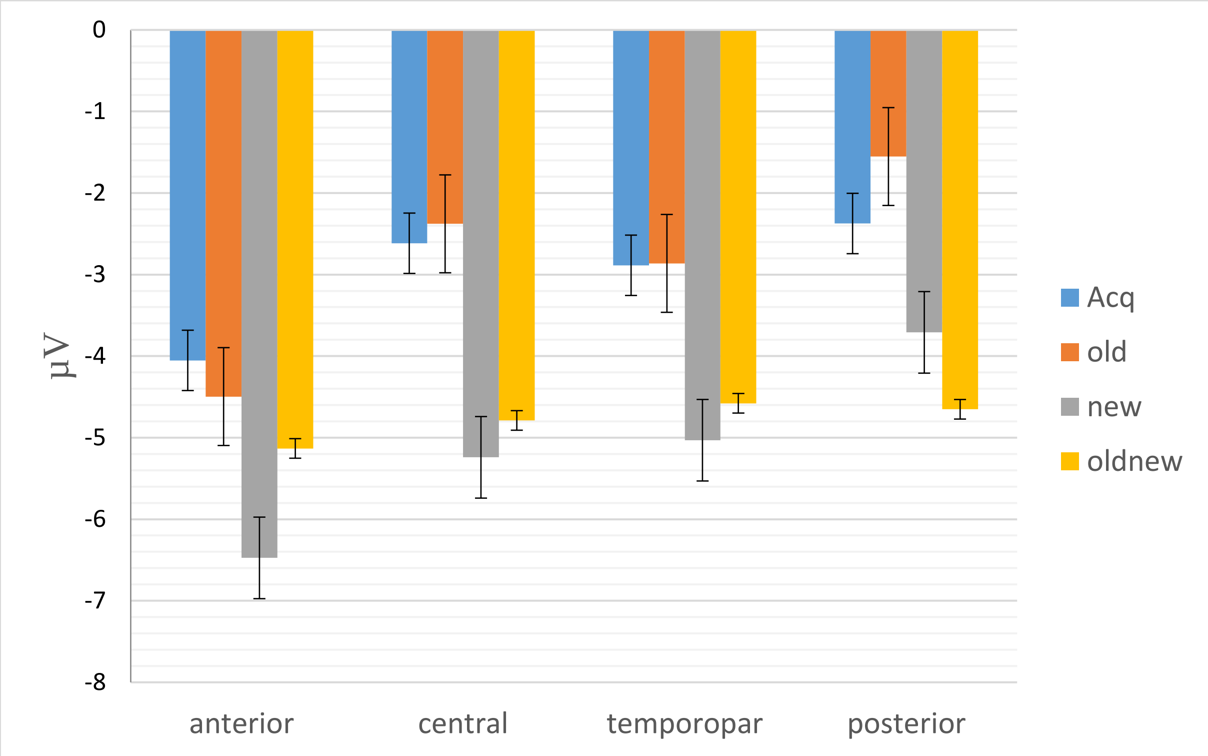
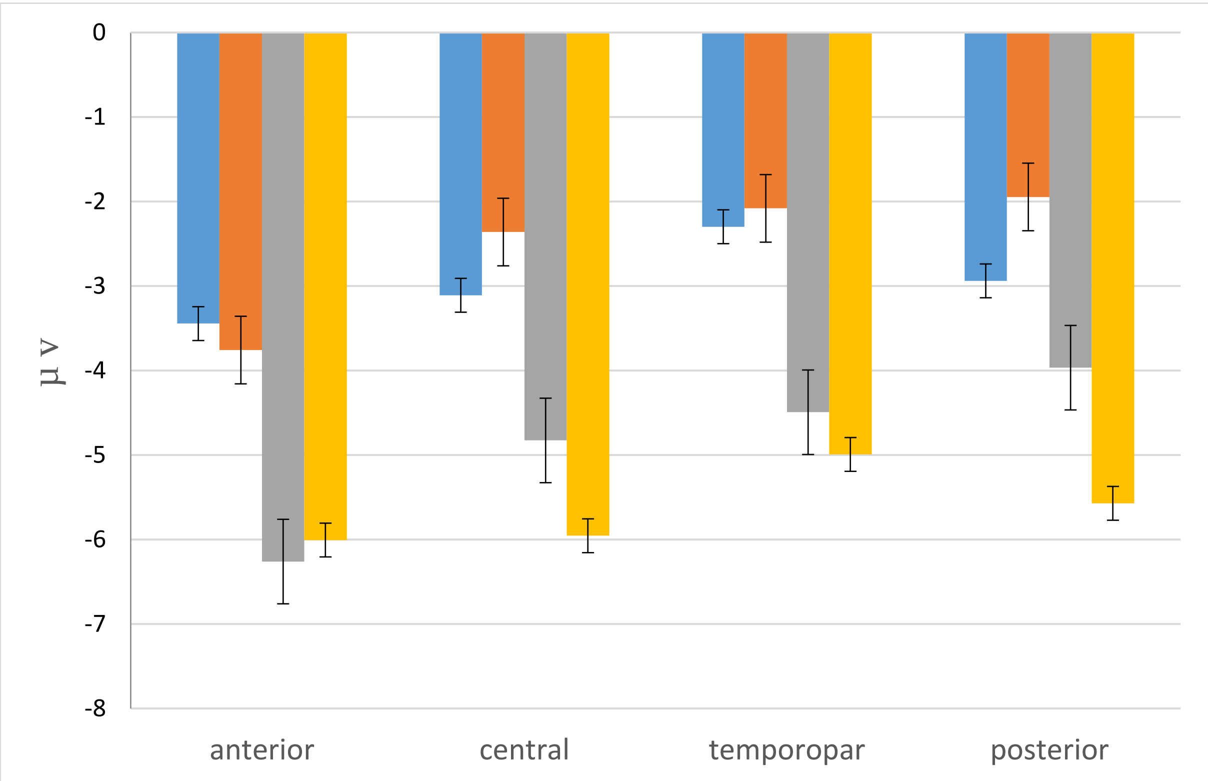


Figure 12 Mean N400 values in all Conditions at right lateral electrodes



bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4. Discussion

4.1. Summary of the data

The aim of the study was to investigate the effect of unintentional learning of semantically unrelated word pairs on semantic ERP components. Particularly, we asked the question whether the N400 effect can be attributed to a repetition effect or to a learnt association between the two unrelated words.

The modulation of ERPs by word repetition was investigated in two experiments, both containing an acquisition phase and a test phase. In both experiments, healthy subjects heard series of semantically unrelated word pairs. A portion of the words were repetitions of previously presented items.

In the test phase of the first experiment, “old” pairs repeated those presented during the acquisition phase, while “new” pairs contained other words that had not been presented before. In the test phase of the second experiment a third condition was added in which the first word in a pair was one of the words presented during acquisition, but the second word was new.

Almost all significant effects of the study were specific for the time window of 350-550 ms, which we identified as the N400. In the *first experiment*, the amplitude for old pairs in the test phase was significantly more negative than to the same stimuli in the acquisition phase. There was also a more negative deflection for new pairs than for old pairs especially at CPz and lateral electrodes.

In response to the second word, in the test phase N400 amplitude for new pairs was significantly more negative than for old pairs, which can be attributed either to the effect of learnt association or that of repetition, or both.

In the *second experiment* the amplitude in the N400 time window to the first word was similar in all four conditions (including old pairs in the acquisition phase and in the test phase, new pairs and old-new pairs in the test phase). Although in some leads (e.g., Pz and CPz) this amplitude appeared to be larger to new than old first words, this effect was very small and neither the main effect of condition nor (as might be expected) a Condition by Site interaction even remotely approached significance.

The amplitude to new second words was significantly larger than for old second words, independently of whether the first word in the pair was old or new. The effect was most pronounced in the central, temporo-parietal and posterior regions. The relatively small N400 amplitude to old words did not differ between the test phase and the acquisition phase.

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To sum up, the main result is that after having been repeated, the previously unrelated word pairs were acting like related word pairs, i.e., the N400 to the second word was strongly reduced. The result is further justified by pointing to the learnt association to the second word for old pairs and the lack thereof for new and old-new pairs.

4.2. Comparable studies

4.2.1. ERP repetition effect

A number of studies have examined the effects of repetition on the ERPs elicited by words in lists. Initially Rugg (1985) compared the effects of repetition to those of associative semantic priming. A technique whereby the exposure to one stimulus influences a response to a subsequent stimulus is called priming. Priming does not necessarily depend on the activation of preexisting memory representations. Instead, a new representation capable of supporting priming can be established at the time of study (Weingarten et al., 2016). In comparison to our result, Rugg (1985) found that both repeated and semantically primed words elicited more positive ERPs than did new or unrelated words in the latency range of 300 to 500 msec post stimulus. This was consistent with the conclusion that both repeated and related words elicit a smaller N400 in lists.

In a following study, Rugg (1987) investigated a combination of repetition and semantic effects. Unlike our study, the experiments included both words and pronounceable non-words. Subjects silently counted occasional non-words against a background of meaningful words, a portion of which were either semantic associates or repetitions of a preceding word. Repeated words were distinguished by waveforms containing an initial transient negative-going deflection called N140. In contrast, semantically primed words showed a relatively small, positive-going modulation around 500 msec. In contrast to the current study, the repetition effect in Rugg (1987) did not appear as a typical N400. Since that experiment included two types of stimuli (words and non-words), onset latency differed between the two studies. One may suppose (although we do not have a direct comparison) that the task to distinguish words from non-words is more difficult, or stronger interfere with semantic processing, than our task to count syllables, and that for this reason word/non-word discrimination took more time. In the light of the above, it seems reasonable to assume that the differences between word and non-word repetition effects on ERPs are attributable to the dissimilar properties of these two types of item, which is completely different from our experiment restricted to one kind of stimulus, unrelated word pairs.

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Van Petten, Kutas, Kluender, Mitchiner and McIsaac (1991) additionally investigated the effects of repetition as a function of word frequency. Whereas repeated low-frequency words elicited a positive ERP component after 500 ms post-stimulus, the predominant effect for repeated high-frequency words was a reduction of N400 amplitude. Although we did not specifically check the effect of word frequency, most words presented in our study belonged to highly to moderately frequent German words. Thus our result is in line with that of Van Petten et al. (1991).

4.2.2. ERP old-new effect

Curran (1999) showed reliable differences in brain electrical activity between old and new stimuli. ERPs were more positive (referred to mastoids) to old than new stimuli from about 400 to 800 ms. He investigated the old/new effect that distinguished between recently presented items and unpresented items and, like Balass et al. (2010), revealed a rather late P600 effect manifested around 500–800 ms after the onset of a word. The P600 was characterized by larger amplitudes for old presented items than for new unpresented items. As mentioned above, this interval is comparable with the late 1 window in the present study, wherein we did not show any significant effect. In contrast, we found a small but significant effect later (around 750-1000 ms) where a positive amplitude for old second words in the first experiment was larger than for new second words.

In addition to the late effect, an ERP study showed that the facilitation of the processing of previously learnt words is associated with an attenuation of a negative peak normally occurring around 400 ms post stimulus (e.g., Karayanidis, Andrews, Ward & McConaghy, 1991). According to Kutas and Van Petten (1988) and Fischler and Raney (1991), the difference in ERPs to previously encountered and new words involves the modulation not only of a late positive component, but also an earlier region of the waveform centred on a negative-going deflection peaking around 400 msec. The amplitude of this deflection is attenuated when elicited by old as opposed to new words. Rugg, Pearl, Walker, Roberts and Holdstock (1994) reported that the ERP to learnt words (as compared with unlearnt words) was more positive-going in the latency range of the N400 component, when item repetition occurred over the interval of less than a few minutes.

Thus about a dozen ERP studies have now documented that N400 amplitude can be reduced by word repetition in lists. However, the ERP effects in most similar studies, even if they included

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the N400 time window, extended well beyond this window and did not necessary overlap with the apparent peak of the N400.

These considerations suggest that the effects observed in these word-list experiments may consist of the modulation of more than one ERP component, for example, that repeated (or learnt) words elicit a larger late positive wave as well as a smaller N400. This is in contrast to our experiment, in which the effects were rather limited by N400 interval. It should be noticed, that contrary to the previous studies, we presented words in pairs and not in lists.

4.3. Repetition effect versus learnt association effect

ERPs recorded during the presentation of sentences that are completed with incompatible words show enhanced negativity around 400 ms after presentation of critical items. Similarly, a comparison between semantically related versus unrelated word pairs consistently reveals an N400 to the latter as compared with the former. However, less attention has been directed to this ERP effect for repeated unrelated pair words.

Esper (1973) emphasized the principal difference between two kinds of associations: inner associations are based on semantic or logical relationships, while outer associations are formed by repetition (cit. for Maki, McKinley, & Thompson, 2004). Thus “cat” and “mammal” are innerly associated even though these two words rarely occur in close proximity in most real sentences. The suppressed N400 to the second word in the learnt word pairs in our current study may be attributed to the outer associations in terms of Esper. On the other hand, in the first experiment there was a significantly larger N400 to the first word in new pairs than in old repeated pairs. Therefore, one may suggest that the effect related to the first word is the repetition effect, while the (larger) N400 difference to the second word might be a combination of repetition and association effects.

However, this interpretation is complicated by the fact that in the first experiment new first words were always related to new second words, and therefore, after a new first word, the subject already knew that the whole pair was new. That is, if subjects hear the first word, they already know that the second word is either new or old. Since if the first word is new, the second word is also new, we cannot be sure whether the effect of the first word is because the first word is new or because the whole association is new. As a result, the first new word in the first experiment can be the subject of either repetition effect or learnt association effect, or both.

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The second experiment sheds a light to this issue. In the test phase we presented word pairs consisting of old first words and new second words. This means that when hearing the first "old" word the participant was unable to know whether the pair is old or new. If N400 suppression to the first word in the first experiment is a kind of pure repetition effect, this effect should be independent of the fate of the second word in a pair; therefore, there would be no reason to expect any difference between the responses to the first word in the first and the second experiment. However, we found that the N400 to the first word in the second experiment was virtually absent. A weak effect could albeit be seen in single electrodes (e.g., CPz) but it was not significant (neither as the main effect nor as an interaction).

At the same time, the N400 to the second word in a pair was substantially smaller in learnt word pairs than in new word pairs. This difference in the response to the second word did not depend on whether the first word in the pair was old or new. Together, these facts speak rather against the repetition effect and indicate that this effect, even if present, was very small as compared with the very large and highly significant effect strongly manifested in the response to the second word and, to a much less extent, to the first word in the first experiment.

Another argument against the repetition effect is a small but significant difference between N400 responses to *old first words* in the acquisition and in test phases of the first experiment. Because each word was presented twenty times already in the acquisition phase, we did not expect any difference as compared with the test phase. But if such a difference takes place, on the basis of the repetition effect one might expect a positive ERP shift in the test phase; just to the contrary, the N400 was more negative in the test phase than to the same words in the acquisition phase.

Finally, the length and the shape of the effect obtained in the present study is highly atypical for repetition and old-new effects. The effect manifested itself only as the difference in the amplitude of the N400 to learnt and unlearnt words. As said above, the repetition effect can be manifested in the suppression of the N400 to repeated items, but it always continues beyond N400 window and lasts up to 800 ms post-stimulus. No such prolonged effect was observed in the present data.

4.4. Implicit learning

Large effects of previously presented word pairs were found in experiments to recognition memory. In a paradigmatic study, Stern & Hasselmo (2008) first presented subjects with a series of stimuli to be learned. After a variable delay period, a list of previously presented stimuli

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

mixed with new ones is presented. The above authors assumed that previously presented items gave rise to a range of confidence ratings based on their familiarity. Performance of these tasks requires recognition memory, and in our study, if the first word is new, since participants knew they would not encounter an old word again, further processing of that item for later recognition was not required.

The present data set and the word pairs list studies reviewed earlier converge to indicate that the N400 repetition effect is susceptible to implicit learning factors.

Although in our data only minor effects of repetition were found, the effect cannot be zero just because repetition was an essential part of the experimental design.

Many have used learning as a way to produce automaticity (Logan, 1998; Schneider & Chein, 2003; Wilkins & Rawson, 2011). Automatic processes operate through a relative set of associative connections and require training to develop fully. According to Logan (1990), since repetition and automaticity as a product of learning share common characteristics, they can be accounted for by the same theory, namely the instance theory of automaticity as a component of a larger learning theory. Thus each instance of stimulus presentation is stored in memory for repeated items and adds to the speed of responses to subsequent stimuli, thus explaining the improvement of speed with practice. Because, however, the repetition effect cannot solely explain our result, the instance theory is not enough and should be added with an associative model by which repetition would strengthen associations for words.

4.5. Automatic Spreading Activation

To summarize, the data support the idea that repeated presentation of word pairs leads to establishment of new associations between stimuli. Since these new associations manifest themselves in the suppression of the N400, one may speculate that they work like semantic associations. These effects of word repetition on ERPs were unrelated to a consistent relationship between repetition and response and, in particular, were largely confined to the second word in already familiar word pairs.

To our knowledge, the theoretical basis that the association effect found in the present data can best be explained upon is the Automatic Spreading Activation (*ASA)* model (Collins & Loftus, 1975). In the following, we focused on the issue of whether this model can allow access to newly formed associations between unrelated word pairs.

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

As far as the automaticity assumption is concerned, Koriat (1981) provided evidence that the association effect in the lexical decision task is automatic in the sense that although anticipation of a target is not a necessary condition for associative facilitation to occur, its presence enhances facilitation. This suggestion is in line with the notion that the semantic facilitation effect in the lexical decision task was to be attributed to the spreading-activation process (Schvaneveldt & Meyer, 1973). A concept can be represented as a node in a network, with properties of the concept represented as labeled relational links from the node to other concept nodes (Collins & Loftus, 1975).

According to many authors (e.g., Neely, 1977; Neely & Keefe, 1989; Hill, Strube, Roesch-Ely & Weisbrod, 2002; Daltrozzo, Wioland & Kotchoubey, 2012**)**, the mental lexicon is assumed to be a semantic network with related words in neighboring nodes. ASA occurs because the lexical access to a word activates the corresponding node and this activation spreads to the neighboring nodes corresponding to related words. When a prime is presented, a set of nodes corresponding to related words are pre-activated. The spreading activation lowers the threshold for the processing of all related nodes. Therefore, when a related (i.e., already pre-activated) word is presented, its processing requires *less* additional activation, thus resulting in a *smaller* amplitude of the N400 to this word.

Note that the ASA model presumes the same basic underlying mechanism for N400 suppression to *semantically related* second words and to second words in word pairs that have been recently presented and *learnt in semantically unrelated sets*, as compared with semantically unrelated word pairs *presented for the first time*.

In our experiment we simulated a real life situation, and the idea was to check which components of word processing would change as a consequence of the new word association learnt during the acquisition phase. For instance, during life, HORSE has been associated with other words like COW. It is referred to as semantic priming, and the associations can be driven by this automatic mechanism. According to the ASA model, there is a spreading activation from the HORSE node to the COW node, and the pre-activation inhibits the N400 to COW. Let us assume that, in contrast, there is no association between HORSE and FAST, and thus no activation spreads from the HORSE node to the FAST node, resulting in a large N400 to FAST. When the unrelated word pair is repeated, the brain is in the search of an interpretation of this newly emerging connection (see Kotchoubey, 2006; Schlesewsky & Bornkessel-Schlesewski, 2019). The activation spreads to broader and broader networks of nodes related to the two words (maybe, particularly the first word in a pair). Some of these activated nodes are related to *both*

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

words in the pair and, by virtue of their activation, build a semantic bridge between them. In the above example, repeated presentation of HORSE may activate many neighboring nodes such as LEGS, RIDING, and finally RACE, which should run FAST to win a match. In other words, a relationship that originally is just a learnt (outer, in the sense of Esper, 1973) association, in the course of consistent ASA can become very much like an inner (meaningful) association and work like the latter. In such a case, identical N400 effects cannot be surprising.

This model is supported by the data about the dynamics of ERP in the case of an intervening concept (Chwilla, Kolk & Mulder, 2000). These authors investigated mediated two-step priming (e.g., from LION to STRIPES via TIGER). The N400 two-step priming effect appears to indicate that participants succeed in integrating word pairs which are indirectly related. The main assumption is that activation spreads along the associative pathways from related concepts to their associates. When a lexical decision to the prime as well as the target was required, mediated priming occurred for N400. One possible account of the presence of mediated priming is that attention to the prime was enhanced by the lexical decision response and evoked a large boost in activation, causing the automatic activation process to spread deep in the network. Chwilla et al. (2000) proposed that the mediated priming pattern in the double lexical decision task was most compatible with a spreading activation account.

In the case of two intervening concepts, the facilitation is referred to as three-step priming (e.g., a prime MANE is associated with the target STRIPES through mediators LION and TIGER), which is again consistent with spreading activation models. Since these models assume that activation spreads to all interconnected nodes, it should also spread from directly related concepts to associates of these concepts (Chwilla & Kolk, 2002). These authors indicated that three-step pairs were not semantically related. However, they referred to frequency of co-occurrence as the frequency with which the prime and the target occur together in ordinary language and apply it as a measure of semantic relatedness. The underlying rationale was that concepts that frequently co-occur share many aspects of their contextually based meaning.

One final matter needs to be considered. Throughout this study, we assumed that in our experiment the words were semantically unrelated pairs. If we match our findings to frequency of co-occurrence in which the unrelated word pairs are repeated, the co-occurrence of association may increase.

4.6. Limitations and further possibilities

Although the numerous data of the literature suggest that the effects should be specific for meaningful words, or maybe for meaningful stimuli in general, this supposed specificity is not

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

proven. Perhaps we need at least two additional experiments in which (i) meaningless non-words and (ii) equally complex (acoustically) but not word-like sounds would be presented in pairs like in the current experiments.

Another unclear issue concerns the role of attention. ASA theory presumes that the effects should be largely independent of attention. In accord with this idea, some N400 effects were even obtained independently of conscious awareness of stimuli (e.g., Kiefer, 2002). Other data indicate, however, that active instruction to respond to stimuli can increase the amplitude of the N400 as compared with more passive conditions (e.g., Erlbeck et al., 2014). In the present study we used an instruction that demanded participants' attention to stimuli. In some contexts, particularly if we think on a clinical application of the paradigm in neurologic patients, the issue of attention may be very important. A replication of the present experiments without an active instruction would help to clarify the issue.

Finally, the theory presumes that the effects mediated by ASA require very brief interstimulus intervals. Our hypothesis can, therefore, be tested when the experiments are replicated with longer intervals between words in pairs.

5. Conclusion

The present findings provide evidence that the acquisition of new word associations results in the suppression of N400 wave in semantically unrelated word pairs, very much similar to the suppression observed in semantically related word pairs. The learning is acquired across multiple trials in the absence of a declarative task. Although our data may also include a repetition effect, this effect, if present, plays a relatively minor role as compared with the effect of associative learning.

The observed effect seems to be similar in most participants, particularly in the (simpler) Experiment 1. However, larger samples are necessary to understand the potential extent of individual differences in the strength of the newly built associations. The present hypothesis, which relates the N400 learnt association effect to automatic processes, needs further support from independent experiments. A simple test of the correctness of the present explanation would be increasing interstimulus interval in word pairs. According to the theory, ASA rapidly decays unless the subject makes effort to maintain it via rehearsal (Neely, 1977). Thus we can predict the increase of the interval between words in pairs would result in the decrement of the observed N400 effect.

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgments

The study was supported by the German Research Society (Deutsche Forschungsgemeinschaft), Grant KO-1753/13.

References

Balass, M., Nelson, J. R., & Perfetti, C. A. (2010). Word learning: An ERP investigation of word experience effects on recognition and word processing. *Contemporary Educational Psychology, 35*(2), 126-140.

Bentin, S., & Peled, B. S. (1990). The contribution of stimulus encoding strategies and decision-related factors to the repetition effect for words: electrophysiological evidence. *Journal of Memory and Cognition, 18*, 359-366.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2019). Towards a neurobiologically plausible model of language-related, negative event-related potentials. *Frontiers in Psychology*, 10, 298.

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. *Lang Learn Dev.*, *8*(3), 278-302.

Brualla, J., Romero, M. F., Serrano, M., & Valdizán, J. R. (1998). Auditory event-related potentials to semantic priming during sleep. *Electroencephalography and Clinical Neurophysiology*, *108*(3), 283-290.

Buchanan, E. M., Holmes, J. L., Teasley, M. L., & Hutchison, K. A. (2013). English semantic word-pair norms and a searchable Web portal for experimental stimulus creation. *Journal of Behavior Research Methods, 45*(3), 746-775.

Chwilla, D. J., Brown, C. M., & Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology*, *32*(3), 274-285.

Chwilla, D. J., & Kolk, H. H. J. (2002). Three-step priming in lexical decision. *Memory and Cognition*, *30*(2), 217-225.

Chwilla, D. J., Kolk, H. H. J., & Mulder, G. (2000). Mediated priming in the lexical decision task: Evidence from event-related potentials and reaction time. *Journal of Memory and Language*, *42*(3), 314-341.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407-428.

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Curran, T. (1999). The electrophysiology of incidental and intentional retrieval: ERP old/new effect in lexical decision and recognition memory. *Neuropsychologia*, *37*, 771-785.

Cruse, D., Beukema, S., Chennu, S., Malins, J. G., Owen A. M. & McRae, K. (2014). The reliability of the N400 in single subjects: Implications for patients with disorders of consciousness. *Neuroimage Clin.*, 9(4), 788-799.

Daltrozzo, J., Wioland, N., & Kotchoubey, B. (2012). The N400 and Late Positive Complex (LPC) effects reflect controlled rather than automatic mechanisms of sentence processing. *Brain Sciences*, *2*(3), 267-297.

Erlbeck, H., Kübler, A., Kotchoubey, B., & Veser, S. (2014). Task instructions modulate the attentional mode affecting the auditory MMN and the semantic N400. *Frontiers in Human Neuroscience*, 8, 654.

Fischler, I., & Raney, G. E. (1991). Language by eye: Behavioral and psychophysiological approaches to reading. In Jennings, J. R., & Coles, M. G. H. (Eds.), *Wiley psychophysiology handbooks. Handbook of cognitive psychophysiology: Central and autonomic nervous system approaches* (pp. 511-574). Oxford, England: John Wiley & Sons.

Girju, R., Badulescu¸ A., & Moldovan, D.I., (2006). Automatic Discovery of Part–Whole Relations. *Computational Linguistics 32*(1), 83-135.

Kiefer, M. (2002). The N400 is modulated by unconsciously perceived masked words: further evidence for an automatic spreading activation account of N400 priming effects. *Cognitive Brain Research*, *13*, 27-39.

Li, Q., Guo, C., & Jiang, Y. (2008). Brain potentials and repetition effects during encoding and retrieval of words. *Neuroreport*, *19*(14), 1365-1368.

Hill, H., Strube, M., Roesch-Ely, D., & Weisbrod, M. (2002). Automatic vs. controlled processes in semantic priming differentiation by event-related potentials. *International Journal of Psychophysiology*, *44*(3), 197-218.

Hoddes, E., Dement, W.C., & Zarcone V. (1973). Quantification of sleepiness: a new approach. *Psychophysiology*, *10*, 431-436.

Karayanidis, F., Andrews S., Ward, Pb. & Mcconaghy, N. (1991). Effects of inter-item lag on word repetition: an event-related potential study. *Psychophysiology*, *28*(3), 307-318.

Koriat, A. (1981). Semantic facilitation in lexical decision as a function of prime-target association. *Memory and Cognition*, *9*(6), 587-598.

Kotchoubey, B. (2006). Event-related potentials, cognition, and behavior: a biological approach. *Neurosci. Biobehav*, *30*(1), 42-65.

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Kutas M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology, 62*, 621-647.

Kutas, M., & Hillyard, S. A. (1980). Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. *Science, 207*(4427), 203-205.

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

Kutas, M., & Van Petten, C. (1988). Event-related brain potential studies in language. *Advances in psychophysiology*, *3*, 139-187.

Lancia, F. (2007). Word co-occurrence and similarity in meaning. In Salvatore, S., & Valsiner, J. (Eds.) *Mind as infinite dimensionality*. Charlotte, NC: Information Age Publishers.

Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology*, *22*(1), 1-35.

Logan, G. D. (1998). Toward an instance theory of automatization. *Psychological Review*, *95*(4), 492-527.

Luck, S. J. (2005). An introduction to the event-related potential technique. Cambridge, Mass: MIT Press.

Maki, W. S., McKinley, L. N., & Thompson, A. G. (2004). Semantic distance norms computed from an electronic dictionary (WordNet). *Behavior Research Methods, Instruments, and Computers*, *36*(3), 421-431.

McRae, K., Khalkhali, S., & Hare, M. (2012). Semantic and associative relations: Examining a tenuous dichotomy. In Reyna, V. F., Chapman, S. B., Dougherty, M. R. & Confrey, J. (Eds.). The Adolescent Brain: Learning, Reasoning, and Decision Making (pp. 39- 66). Washington, DC: APA.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*(3), 226-254.

Neely, J. H., & Keefe, D. E. (1989). Semantic context effects in visual word processing: A hybrid prospective/retrospective processing theory. In Bower, G. H. (Eds.). *The psychology of learning and motivation: Advances in research and theory* (Vol. 24, pp. 207-248). New York: Academic Press.

Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97-113.

bioRxiv preprint doi: https://doi.org/10.1101/741728; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ortu, D., Allan, K., & Donaldson, D. I. (2013**).** Is the N400 effect a neurophysiological index of associative relationships? *Neuropsychologia, 51*(9), 1724-1748.

Perea, M., & Rosa, E. (2002). The effects of associative and semantic priming in the lexical decision task. *Psychological Research, 66*, 180-194.

Rhodes S. M., & Donaldson D. I. (2008). Association and not semantic relationships elicit the N400 effect: electrophysiological evidence from an explicit language comprehension task. *Psychophysiology*, *45*(1), 50-59.

Rugg, M. D. (1985). The effects of semantic priming and word repetition on event-related potentials. *Psychophysiology, 22*(6), 642-647.

Rugg, M. D. (1987). Dissociation of semantic priming, word and non-word repetition effects by event-related potentials. *The Quarterly Journal of Experimental Psychology*, *39*(1), 123-148.

Rugg, M. D., Doyle, M. C., & Holdstock, J. S. (1994). Modulation of event-related brain potentials by word repetition: Effects of local context. *Psychophysiology*, *31*(5), 447-459.

Rugg, M. D., Pearl, S., Walker, P., Roberts, R. C. & Holdstock, J. S. (1994). Word repetition effects on event-related potentials in healthy young and old subjects, and in patients with Alzheimer-type dementia. *Neuropsychologia*, *32*(4), 381-398.

Schneider, W., & Chein, J. M. (2003). Controlled & automatic processing: behavior, theory, and biological mechanisms. *Journal of Cognitive Science*, *27*(3), 525-559.

Schvaneveldt, R. W., & Meyer, D. E. (1973). Retrieval and comparison processes in semantic memory. In Kornblum, S. (Ed.), Attention and performance IV (pp. 395-409). New York: Academic Press.

Stern, C. E., & Hasselmo, M. E. (2008). The Neurobiological Basis of Recognition Memory in *Learning and Memory: A Comprehensive Reference* (Vol. 3, pp. 131-141). Elsevier.

Van Petten, C., Kutas, M., Kluender, R., Mitchiner, M., & McIsaac, H. (1991). Fractionating the Word Repetition Effect with Event-Related Potentials. *Journal of Cognitive Neuroscience*, *3*(2), 131-150.

Weingarten, E., Chen, Q., McAdams, M., Yi, J., Hepler, J., & Albarracín, D. (2016). From primed concepts to action: A meta-analysis of the behavioral effects of incidentally presented words. *Psychological Bulletin*, *142*(5), 472-497.

Wilkins, N. J., & Rawson, K. A. (2011). Controlling retrieval during practice: Implications for memory-based theories of automaticity. *Journal of Memory and Language*, *65*(2), 208-221.