Electrophysiological Correlates and Predictive Power of ERP Components for False Memories

Hugo Marte-Santana, Laura V. Sánchez-Vincitore, & Daniel Cubilla-Bonnetier.

Neurocognition and Psychophysiology Laboratory, Universidad Iberoamericana (UNIBE), Santo Domingo, Dominican Republic.

Corresponding Author: Hugo Marte-Santana (h.marte@unibe.edu.do).

Keywords: memory, false memories, ERP, P300, N400, LPC, fuzzy-trace theory.

Abstract

This study investigates the neural correlates of false memory using event-related potentials (ERPs) in a Deese-Roediger-McDermott (DRM) paradigm. The primary aim of this study was to assess the predictive utility of three ERP components (P3, FN4, LPC) in distinguishing between true and false memories. Additionally, we examined the influence of semantic and orthographic similarity, as well as emotional valence (positive, neutral, negative), on memory accuracy and reaction times. EEG data were collected from 33 participants using a 128-channel system, analyzed through repeated measures ANOVA and logistic regression. Results indicated significantly heightened P3 and LPC amplitudes for false memories as well as more negative peak amplitudes for FN4 component. No significant ERP amplitude differences were found between semantically and orthographically related critical items. Logistic regression models demonstrated promising predictive power for P3 and FN4 components in distinguishing false from true memories, although not reaching forensic standards of specificity. These findings suggest that while certain ERP components show potential in identifying false memories, further refinement and inclusion of additional neural markers are necessary for robust forensic applications. Future directions include expanding the ERP model with additional components and exploring advanced EEG analyses such as time-frequency and spectral density to enhance predictive accuracy and deepen our understanding of false memory mechanisms.

1. Introduction

Memory is not a repository of immutable information (Armaly & Enders, 2023). In fact, information processed by memory can undergo alterations and be stored incompletely (Arce et al., 2023; Stanek et al., 2024). The missing elements can be filled in with self-generated information that is contextually consistent with the original information (Arce et al., 2023), altering the memories in their content, location, temporality, or involved persons (Greenspan & Loftus, 2020; Otgaar et al., 2022; Wixted et al., 2021). These altered memories are what we know as false memories (Armaly & Enders, 2023; Nichols & Loftus, 2019; Spets et al., 2021; Steffens & Mecklenbräuker, 2007).

A theoretical approach that could describe the phenomena of false memories is the fuzzy-trace theory (Brainerd & Reyna, 1990). This theory proposes that memory relies on the storage of two types of representations: verbatim-traces and gist-traces. The former are literal content about the details of the processed event or information, and the latter are more abstract and associative general representations of said information. This theory suggests that, as time passes, memory becomes more dependent on gist-traces than verbatim-traces, and that by combining the two types of traces, individuals fill in the missing literal information with contextually coherent but inaccurate information, producing a false memory.

There are recognized elements that act as mediators in the generation of these false memories, factors that seem to facilitate their occurrence. This includes the coherence of the false memory with the context of the remembered situation for images or events (Arce et al., 2023; Bamatraf et al., 2016), or the semantic field (concepts related by their meaning) and formal similarity (phonological or orthographic) when it comes to linguistic material (Alakbarova et al., 2021; Chadwick et al., 2016; Gatti et al., 2023; Griffin & Schnyer, 2020). Coane et al. (2021) observed how formal similarities between words generated an increase in erroneous recognitions. Griffin & Schnyer (2020) obtained similar results, observing a higher number of false alarms in new words with orthographic and semantic similarities.

Another mediating factor in the generation of false memories is emotional arousal, which has been linked to modulations in information storage in memory (Bland et al., 2016; Brainerd et al., 2008; Hamann, 2009; Knott et al., 2022; Li et al., 2021; Zimmerman & Kelley, 2010). Emotional arousal refers to the positive (victory, happiness), negative (war, defeat), or neutral (table, chair) content of items. Negative cognitive biases have shown effects in enhancing the accuracy of memory retrieval (Ellis et al., 1984; Kensinger, 2007; Kensinger & Corkin, 2003). Similarly, it has been observed that negative emotional content could facilitate information storage in memory (Kensinger, 2007; Xie & Zhang, 2017). Emotional content or arousal has also been shown to increase the generation of false memories (Bookbinder & Brainerd, 2016; Griffin & Schnyer, 2020; Yüvrük & Kapucu, 2022). Griffin & Schnyer (2020) observed that the negative emotional arousal of distractors—compared to neutral arousal—tended to increase the number of false positives when distractors shared semantic context with the words to be remembered, but tended to decrease it when the words (false positive and item to be remembered) shared an orthographic relationship. These results support that orthographically related stimuli provide a more reliable method to test the effect of emotional arousal on false memory generation.

Whether the memory generated is true or false, in both cases, individuals believe they remember the event accurately (Greenspan & Loftus, 2020). Thus, it could be argued that the underlying neural mechanisms are similar in both cases. However, true and false memories appear to exhibit differentiated physiological patterns, specifically in brain electrical activity (Favre et al., 2020; Li et al., 2021; Volz et al., 2019). These patterns can be identified using electroencephalographic (EEG) measurements, employing event-related potential (ERP) components: true and false memories are associated with distinct underlying electrophysiological patterns (Chen et al., 2012; Volz et al., 2019; Wang et al., 2023), suggesting the complexity of false memory generation mechanisms.

Event-related potentials (ERPs) are positive or negative electrical deflections in brain activity resulting from perceptual or cognitive processes. These electrical deflections occur at specific times (latency), with perceptual potentials occurring earlier (lower latency) than cognitive ones (Luck & Kappenman, 2013).

Among the components associated with memory processes is the P300 (Amin et al., 2015; Steiner et al., 2013), a positive deflection occurring around 300 milliseconds after stimulus presentation, also related to advanced stages of cognitive processing (Gongora et al., 2021; Sara et al., 1994). This component has shown variations in amplitude as a result of processing in working memory tasks (Harwood et al., 2022; Käthner et al., 2014; Simpson & Rafferty, 2021), as well as an increase in P3 amplitudes for false memories compared to correct rejections (Volz et al., 2019).

Another component associated with false memories is N400 in frontal montages (FN400). This is a negative deflection occurring 400 ms after a stimulus presentation, traditionally associated with language processing, semantic incongruities, and the familiarity of semantic content (Hodapp & Rabovsky, 2021; Kuipers & Thierry, 2011; Kutas & Federmeier, 2011; Morett et al., 2021). It has also shown changes in its amplitudes during the occurrence of false memories (Chen et al., 2012; Griffin & Schnyer, 2020). Griffin & Schnyer (2020) observed greater negativity amplitudes of this component for false memories compared to correct identifications.

Similarly, the late-positive component (LPC), a positive deflection occurring after 600 ms, has been linked to the retrieval process of information from memory (Voss & Paller, 2007). This is another component related to the formation of false memories, as an increase in its amplitude has been observed during true memories compared to false positives and false memories (Griffin & Schnyer, 2020; Kiat & Belli, 2017). An increase in the amplitudes of this component has also been associated with heightened cognitive demand across various contexts (Deeny et al., 2014; MacNamara et al., 2019; MacNamara & Proudfit, 2014).

False memories have an impact on subjects' subjective perception of reality (Armaly & Enders, 2023; Greenberg, 2004), their behavior (Murphy et al., 2019), and consequently, their interaction with others. Moreover, the occurrence of false memories can affect the functioning of certain institutions, such as the justice system, as it may inadvertently affect the accuracy of specific testimonies in legal proceedings (Greenspan & Loftus, 2020;

Wixted et al., 2021). Hence, there is significant relevance in researching the neural mechanisms underlying these alterations in memory content.

As mentioned previously, evidence shows differences between true and false memories in some of these components (Chen et al., 2012; Favre et al., 2020; Griffin & Schnyer, 2020; Volz et al., 2019; Voss & Paller, 2007). However, there is a lack of studies proposing an analysis of these components as a complex of signals that could determine whether a memory is true or false, with the consequent applications this would allow. Therefore, the main purpose of the present study was to determine if it is possible to construct predictive models reliable enough to discriminate false memories from true ones using a complex of ERP components (P3, FN4, and LPC) as predictor variables. Additionally, this study had two additional objectives. Firstly, in the behavioral domain, it aimed to confirm the existence of factors influencing the generation of true and false memories, such as the semantic or orthographic proximity of the critical stimuli used, or their positive, neutral, or negative emotional content. Secondly, in the electrophysiological domain, it aimed to confirm that there are differences in the amplitude of the mentioned components between true and false memories.

In the behavioral domain, we hypothesize that more false memories will be triggered by critical items (items similar to those in a word list but not presented) of semantic rather than orthographic type (based on findings by Ballou & Sommers, 2008). Additionally, we expect that the negative emotional valence of critical items will also correlate with a higher number of false memories (Bland et al., 2016; Brainerd et al., 2008; Davis & Loftus, 2007). In the electrophysiological domain, we hypothesize that there will be differences in the amplitude of the mentioned ERP components depending on the type of memory (increased P3 amplitudes for false memories, decreased FN4 negativity, increased LPC amplitudes), based on accumulated evidence (Chen et al., 2012; Favre et al., 2020; Griffin & Schnyer, 2020; Volz et al., 2019; Voss & Paller, 2007). Lastly, regarding the production of predictive models using a complex of ERP components, no preliminary hypotheses were formulated, as this is an exploratory study without precedents in the literature.

2. Method

2.1. Participants

The present study employed a within-subject design. Sample size calculation was conducted using GPower 3.1 (Faul et al., 2009) for a repeated measures ANOVA with an effect size of 0.30, alpha of 0.05, power of 0.95, 1 group, and 3 measurements, resulting in a required sample size of 31 participants.

For this study, 57 young adult volunteers participated, all university students (39 female; mean age = 21.3 years; SD = ±1.72). Behavioral performance data (accuracy and reaction times) were collected from these participants. Individuals with reading or language disorders, or diagnosed with psychological, psychiatric, neurological, or neuropsychological conditions, were excluded from the study beforehand.

A total of 24 participants were excluded from the EEG analysis. This high exclusion rate resulted from the dependency of brain data on the participants' behavioral responses, specifically the identification of "false memories." To remain in the study group, participants needed to generate a sufficient number of false memories. Consequently, participants were excluded if they had insufficient segments for target items (< 10%) (n=2), semantic items (n=7), orthographic items (n=8), and distractor items (n=5). Additionally, participants with more than 10 bad channels were excluded (n=2).This resulted in a final sample of 33 participants (23 female; mean age = 21.1 years; SD = ±1.99) with normal or corrected vision included in the analysis.

2.2. Task Design

We used an adaptation of the Deese-Roediger-McDermott (DRM) word retention and recognition paradigm in our study. In this paradigm, participants are presented with a list of words to remember (study list) (Figure 1). Following a distractor task of successive subtractions, they are presented with a list where they must identify the words that were presented in the study list (identification list).

Figure. 1. Experimental Design Outline.

The identification lists contained, in addition to words from the study list (target items), words that were semantically and orthographically similar to those in the study list (critical items), and words unrelated to the study list (distractor items). False positives—critical items or distractors recognized in the study list—were considered false memories. Similarly, target and critical items could have different emotional content (positive, negative, or neutral).

The task consisted of 24 study lists (12 words per list, totaling 288 items) and 24 identification lists (24 words per list, totaling 576 items), amounting to 864 items in total. The emotional valence of the items was balanced across the study lists (4 items of each valence per list, 96 per emotional valence). The distribution of items in the identification lists was pseudo-randomized but balanced across all lists, ensuring an equal number of items of each type and valence (Table 1). Items were presented in Spanish, the participants' native language.

Table 1. Distribution of items in the identification list by type and valence.

Type	Negative	Positive	Neutral	total
Target	48	48	48	144
Orthographic	48	48	48	144
Semantic	48	48	48	144
Distractors	0	0	144	144
total	144	144	288	576

2.3. Instruments

The experimental design programming and collection of behavioral data were conducted using E-prime® (v.2.0; E-Prime Psychology Software Tools Inc., Pittsburgh, USA) on a Dell® OptiPlex 7020 computer (Intel Core i3, Windows 7 Professional, 4.00GB RAM, 64-bit). Stimuli were presented on a Dell® SE2219H monitor (21.5"; 60Hz). Participant responses were recorded using a Chronos™ response box (Psychology Software Tools; PST-100430).

For continuous electroencephalographic data collection, the study utilized the Geodesic EEG System (GES; series 400, Electrical Geodesics, Inc. (EGI®)) with 128 channels and Netstation Acquisition 4.5.7 software, sampling at 1000 Hz with impedance maintained below 5kΩ.

The statistical analysis was conducted using JASP software (v. 0.16.4; JASP Team, 2024).

2.4. Data processing

The EEG data processing was conducted using Net Station Tools 4.5.7 software. A bandpass filter was applied at 0.1 to 30 Hz (Passband Gain: 99.9% (0.1 dB); Stopband Gain: 1.0% (40.0 dB); Rolloff: 2.00 Hz). Subsequently, it was segmented into windows from 250 milliseconds before stimulus presentation to 1500 milliseconds after stimulus presentation. Blinks and vertical eye movements were identified by inspecting the data collected by electrodes above and below the eyes (channels 8, 126, 25, 127). Horizontal eye movements were recorded using channels to the left and right of the eyes (channels 125 and 128).

Artifacts from bad channels were automatically identified using a criterion where a voltage peak greater than 200μV was considered. To detect blinks, the peak had to exceed 140μV above the relevant channels. Regarding eye movements, the peak had to exceed 55μV. Any channel with 20% or more of its segments marked as noisy was classified as a bad channel. Segments were targeted for removal when they contained more than 10 noisy channels. Channels flagged as bad were eliminated from the dataset and interpolated using spherical spline modeling ($\bar{x}$ = 0.91; SD =1.82) (see Perrin, Pernier, Bertrand, & Echallier, 1989), leveraging data obtained from neighboring sensors. On average, 0.52 channels were substituted per recording (SD = 1.45). Noisy segments marked as bad were excluded from the analysis ($\bar{x}$ = 189.7; SD = 59.24; 21.95%). Remaining usable trials were counted and compared for each condition.

To mitigate any potential bias stemming from a particular reference channel during recording, the data underwent offline re-referencing to the average reference. Subsequently, the mean waveforms underwent baseline correction, utilizing the 200ms pre-stimulus period, to counteract drift and minimize the impact of background noise.

Signal to noise ratio (SNR) was calculated using the following formula:

$$\mathbf{SNR = S / N}$$

Where S is the peak amplitude of the ERP component of interest and N was the standard deviation of the single-trial EEG data within the same time window. SNR for P3 was 5.99, FN4 was 6.03 and for LPC was 5.98.

To analyze the proposed components, the data were segmented into regions of interest (ROIs). The electrodes that comprise each ROI were determined based on a principal component analysis.

For the P3 component (250 to 550ms) ROI: Parietal: Pz (l = 0.834; u = 0.304), Poz (l = 0.881; u = 0.224), PO3 (l = 0.890; u = 0.207), PO4 (l = 0.957; u = 0.083), P1 (l = 0.873; u = 0.238), P2 (l = 0.868; u = 0.246); FN4 (300 to 550ms) ROI: Right Frontal: F2 (l = 0.802; u = 0.357), Fp2 (l = 0.879; u = 0.227), AF4 (l = 0.912; u = 0.169), AF8 (l = 0.743; u = 0.447), E8 (l = 0.883; u = 0.221); LPC (550 to 800ms) ROI: Parietal: Pz (l = 0.837; u = 0.299), Poz (l = 0.889; u = 0.209), PO3 (l = 0.921; u = 0.152), PO4 (l = 0.923; u = 0.132), P1 (l = 0.838; u = 0.297), P2 (l = 0.838; u = 0.297)).

Time windows for ERPs components were extracted from Volz et al. (2019).

2.5. Procedure

The study has received approval from the Ethics Committee of Universidad Iberoamericana (CEI2023-12). Participants received all study information and signed informed consent forms before the EEG electrode application. They were then led to a soundproof chamber and instructed to take a seat facing the computer screen for task presentation.

The electrode cap was connected to the signal amplifier, and the participant was instructed on how to proceed with the task. At this point, the experimenter left the soundproof chamber, informing the participant to begin executing the experiment.

2.6. Statistical Analysis

For behavioral analyses, a repeated measures ANOVA was conducted, including stimulus type (target, semantic, orthographic, and distractor) and emotional valence of the stimulus (positive, negative, and neutral). Given that distractors only had a neutral emotional valence, they were excluded from the ANOVA that included this variable.

To determine differences in the electrical activity of individual components, a repeated measures ANOVA was used. For the components P3, FN4, and LPC, a 4x3 ANOVA was conducted, including stimulus type and emotional valence (positive vs. negative vs. neutral). Correct recognitions were used for target items, while false alarms were used for critical items.

When sphericity assumptions were not met, corrections were applied using the method of greatest observed power (Greenhouse-Geisser or Huynd-Feldt). Post-hoc comparisons were conducted using the Bonferroni multiple comparisons test.

The possibility of predicting the presence of true or false memories through the P3, FN4, and LPC components was tested using a logistic regression model.

3. Results

Addressing the behavioral objective regarding factors influencing false memory generation, in the ANOVA on participants' execution accuracy, we observed a main effect of stimulus type - target, semantic, orthographic, and distractor ($F (1.467, 56) = 261.046$; $p < 0.001$; $\eta2 = 0.823$), as well as emotional valence - positive, negative, and neutral ($F (2, 56) = 50.303$; $p < 0.001$; $\eta2 = 0.026$). Items with orthographic similarity ($M = 0.92$; $SD = 0.06$), semantic similarity ($M = 0.89$; $SD = 0.08$), and distractors ($M = 0.97$; $SD = 0.04$) were better identified (correct rejections) than target stimuli ($M = 0.56$; $SD = 0.15$) (incorrect rejections). Significant differences were observed in accuracy rates among target items, critical items, and distractors, but not between items with semantic and orthographic similarity (see Table 2). Similarly, stimuli with neutral emotional valence ($M = 0.79$; $SD = 0.095$) and negative emotional valence ($M = 0.76$; $SD = 0.119$) were better identified than those with positive emotional valence ($M = 0.71$; $SD = 0.141$). Significant differences were observed between stimuli with negative and positive valence (MD = 0.042; SE = 0.007; t = 5.682; p < .001; d = 0.343), as well as between stimuli with neutral valence and those with positive (MD = 0.074; SE = 0.007; t = 9.999; p < .001; d = 0.603) and negative valence (MD = 0.032; SE = 0.007; t = 4.317; p < .001; d = 0.260).

Table 2. Post hoc comparisons for accuracy by stimulus type.

stimulus type		MD	SE	t	Cohen's d	pbonf
orthographic	target	0.357	0.016	21.951	3.926	< .001***
	semantic	0.029	0.016	1.782	0.319	0.459
	distractor	-0.054	0.016	-3.302	-0.591	0.007**
target	semantic	-0.328	0.016	-20.169	-3.607	< .001***
	distractor	-0.411	0.016	-25.253	-4.516	< .001***
semantic	distractor	-0.083	0.016	-5.084	-0.909	< .001***

P-value adjusted for comparing a family of 6. * 0.05. **0.01, *** <0.001.

Regarding reaction times, a similar pattern was observed with a main effect found for stimulus type (F (1.144, 56) = 7.381; p = 0.006; η2 = 0.116), as well as for emotional valence of stimuli (F (2, 56) = 22.231; p < 0.001; η2 = 0.063). In this case, distractor items were recognized the fastest (M = 1000.45; SD = 308.99), followed by semantic similarity items (M = 1090.7; SD = 369.28), orthographic similarity items (M = 1118.98; SD = 345.88), and lastly target items (M = 1221.80; SD = 622.88). Significant differences were observed in reaction times between target items and semantic similarity items (MD = 131.095; SE = 47.457; t = 2.762; p = 0.038; d = 0.305) and distractor items (MD = 221.342; SE = 47.457; t = 4.664; p < .001; d = 0.515). However, no statistically significant differences were observed between target items and orthographic similarity items (p > 0.05), nor between semantic and orthographic similarity items (p > 0.05), as well as between distractor items and critical items (semantic and orthographic similarity) (p > 0.05). Similarly, items with neutral emotional valence were recognized faster (M = 1143.83; SD = 446.02), followed by items with negative emotional valence (M = 1255.19; SD = 401.97), and lastly items with positive emotional valence (M = 1282.46; SD = 497.70). Significant differences were found between items with neutral emotional valence and items with negative emotional valence (MD = 111.364; SE = 22.029; t = 5.055; p < .001; d = 0.244), as well as between items with neutral emotional valence and items with positive emotional valence (MD = -138.631; SE = 22.029; t = -6.293; p < .001; d = -0.303).

However, no significant differences were found between items with positive and negative emotional valence.

Responding to the objective in the electrophysiological domain, for the P3 component, we observed a main effect of stimulus type (F (2, 32) = 18.347; $p < 0.001$; η2 = 0.364) (Figure 2). However, we did not find significant differences in emotional valence of the stimuli ($p > 0.05$). Specifically, we observed lower amplitudes for target stimuli (M = 3.359; SD = 2.674) and distractor stimuli (M = 1.401; SD = 0.238) compared to stimuli with orthographic similarity (M = 5.450; SD = 3.112) and semantic similarity (M = 6.358; SD = 3.882). Significant differences were observed in P3 component amplitudes between target stimuli and stimuli with orthographic similarity (MD = -2.09; SE = 0.732; t = -2.856; p = 0.032; d = -0.719) and semantic similarity (MD = -2.999; SE = 0.732; t = -4.097; $p < 0.001$; d = -1.032), as well as between distractor stimuli and stimuli with orthographic similarity (MD = -4.048; SE = 0.732; t = -5.531; $p < .001$; d = -1.393) and semantic similarity (MD = -4.957; SE = 0.732; t = -6.722; $p < .001$; d = -1.706). No significant differences were observed in P3 component amplitudes between stimuli with orthographic similarity and semantic similarity ($p > 0.05$), nor between target stimuli and distractor stimuli ($p > 0.05$).

Figure. 2. Top: Event-related potentials at central scalp locations with the comparison of the grand averages for false alarms (red line) and correct rejections (blue line) Bottom: scalp maps comparing Correct Rejections (left) and False Alarms (right).

When analyzing the amplitudes of the FN4 component, we observed a main effect of stimulus type (F (1.859, 32) = 17.350; $p < 0.001$; η2 = 0.352) (Figure 3). Items with the most negative amplitudes were those with semantic similarity (M = -10.065; SD = 6.429), followed by items with orthographic similarity (M = -8.556; SD = 4.446), target items (M = -5.290; SD = 3.531), and lastly distractor items (M = -2.957; SD = 1.962). In post hoc analyses for stimulus type, we found significant differences in the amplitudes of this component between target items and items with orthographic similarity (MD = 3.266; SE = 1.087; t = 3.004; p = 0.022; d = 0.742) and semantic similarity (MD = 4.775; SE = 1.087; t = 4.392; $p < 0.001$; d = 1.085), as well as between distractor items and items with orthographic similarity (MD = 5.599; SE = 1.087; t = 5.149; $p < 0.001$; d = 1.237) and semantic similarity (MD = 7.108; SE = 1.087; t = 6.537; $p < 0.001$; d = 1.616). However, no significant differences were observed between items with orthographic similarity and semantic similarity ($p > 0.05$), nor between target items and distractor items ($p > 0.05$).

Figure. 3. Top: Event-related potentials at right frontal scalp locations with the comparison of the grand averages for false alarms (Blue line) and correct rejections (red line). Bottom: scalp maps comparing correct Rejections (left) and False Alarms (right).

Regarding the LPC component, we observed a main effect of item type ($F (2, 32) = 13.197$; $p < 0.001$; $\eta 2 = 0.292$) (Figure 2). For this component, we found the highest amplitudes in items with semantic similarity ($M = 6.105$; $SD = 4.256$), followed by items with orthographic similarity ($M = 5.635$; $SD = 3.835$), while target items ($M = 3.107$; $SD = 2.587$) and distractor items ($M = 1.764$; $SD = 1.238$) showed lower amplitudes. Significant differences were observed when comparing the amplitudes of target items with those of items with orthographic similarity ($MD = -2.528$; $SE = 0.804$; $t = -3.142$; $p = 0.013$; $d = -0.789$) and semantic similarity ($MD = -2.998$; $SE = 0.804$; $t = -3.727$; $p = 0.002$; $d = -0.936$), as well as between distractor items and items with orthographic similarity ($MD = -3.871$; $SE = 0.804$; $t = -4.813$; $p < 0.001$; $d = -1.208$) and semantic similarity ($MD = -4.341$; $SE = 0.804$; $t = -5.397$; $p < 0.001$; $d = -1.355$). However, no significant differences were found when comparing items with orthographic similarity and semantic similarity ($p > 0.05$). Additionally, there were no significant differences in the amplitudes of this component between target items and distractor items ($p > 0.05$).

We conducted a logistic regression analysis to evaluate the impact of the variables LPC, P3, and FN4 on true/false memories. As shown in Table 3, the final model, which excluded the LPC variable, was significant ($\chi^2 = 17.750$, $df = 294$, Nagelkerke's $R^2 = .25$, $p < .001$). The coefficient for the variable P3 was positive and significant (B standardized = .99, Wald $\chi^2 = 21.373$, $p < .001$), indicating a strong positive association with the dependent variable. Similarly, the coefficient for FN4 was negative and significant, though less pronounced (B standardized = -.75, Wald $\chi^2 = 15.108$, $p < .001$), suggesting a weaker negative association with the dependent variable.

Table 3. Coefficients of logistic regression model (predicting false memories by P3, FN4 and LPC)

Parameter*	B estimate	SE	B standardized	Odds Ratio	Wald statistic	*p*
(Intercept)	-1.62	0.369	1.02	0.198	19.343	<.001
P3	0.28	0.060	0.99	1.322	21.373	<.001
FN4	-0.14	0.037	-0.75	0.865	15.108	<.001

*The LPC component was excluded from the final model

The obtained model demonstrated an accuracy of .72, a sensitivity of .86, but a specificity of .45, with a precision of .76 (Table 4).

Table 4. Confusion matrix of logistic regression model

	Predicted		
Observed	Targets	Critical	% Correct
Targets	45	54	45.45
Critical	28	170	85.86
Overall % Correct			72.39

4. Discussion

The main objective of this study was to determine the feasibility of constructing predictive models to discriminate between false and true memories using a complex of ERP components (P3, FN4, and LPC) as predictor variables. Additionally, we aimed to confirm the facilitation effect on false memory generation from semantic and orthographic proximity, as well as from positive, neutral, or negative emotional load. In the electrophysiological domain, our goal was to validate the presence of amplitude differences in the P3, FN4, and LPC components between true and false memories.

In our results, we did not observe significant differences between critical items (orthographically and semantically related) in the generation of false memories, unlike the findings of Griffin & Schnyer (2020), who reported more false memories generated by semantically related stimuli compared to orthographically related ones. This discrepancy between our study and Griffin & Schnyer's (2020) findings regarding critical items (semantic vs. orthographic) may be attributed to the fact that our study used items in Spanish, a transparent language where words are pronounced as they are written, unlike Griffin & Schnyer's study which used stimuli in English, a language with opaque orthography (irregularities in grapheme-to-phoneme correspondences). Similarly, in comparing semantically and orthographically related items, we also did not observe significant differences in reaction times between them. Moreover, we found that critical items and distractors were recognized (correctly rejected) better than target items (incorrectly rejected).

In our behavioral results, we also observed that items with positive emotional valence resulted in the highest number of false alarms. In contrast, Yüvrük & Kapucu (2022) reported an opposite effect, noting a higher incidence of false alarms for items with negative emotional valence. These authors attribute their findings to potential effects of increased strength in associative networks of items and biases in participant response patterns. The difference in our results could stem from the fact that lists in the study by Yüvrük & Kapucu (2022) were presented in blocks of emotional content, whereas our lists contained all three types of emotional valence. This effect may be supported by evidence suggesting that content with negative emotional loading facilitates information storage in memory (Kensinger, 2007; Xie & Zhang, 2017).

Regarding the observed amplitudes of ERP components, our results reveal higher amplitudes of the P300 component for false memories. Volz et al. (2019) observed a similar pattern in the amplitude behavior of this component. This may be attributed to differentiated memory mechanisms between true and false memories. Given that this component has been linked to working memory processes (Amin et al., 2015; Steiner et al., 2013) and cognitive processing (Gongora et al., 2021; Sara et al., 1994), it could be hypothesized that false memories involve higher cognitive demand mechanisms, possibly due to the predominance of gist-traces, which would require more elaborate cognitive processing of information and the establishment of more complex chains of associations.

For the FN4 component, we observed that false memories generate higher negative peaks compared to both target items and distractors, corroborating findings by Griffin & Schnyer (2020). These results could be attributed to the novelty response experienced in correctly rejecting novel items and the familiarity of target items (Hodapp & Rabovsky, 2021).

Regarding the LPC component, we observed higher amplitudes for false memories compared to correct identifications of target items. This contrasts with findings from other authors such as Griffin & Schnyer (2020) and Kiat & Belli (2017). Griffin & Schnyer attribute this difference to the confidence with which participants respond. However, this component is sensitive to cognitive demand, and item recognition may require fewer cognitive resources because it relies on verbatim-trace compared to false alarms, which also involve fuzzy-trace mechanisms. Another possible explanation for these discrepancies is that Griffin & Schnyer (2020) worked with a population diagnosed with depression, which could lead to attenuated neural activity, and Kiat & Belli (2017) used a misinformation paradigm that may involve additional neural circuits.

For none of the ERP components were there any observed differences between the types of critical items (orthographic and semantic similarity). Similarly, no differences were observed in the amplitudes between target items and distractors.

The results of the logistic regression to predict the presence of false memory based on the amplitude of P300, FN4, and LPC components, while promising due to the significance of the first two components, did not demonstrate sufficient specificity for their unequivocal use as forensic evidence. This could be due to the small sample size, as small sample sizes can lead to reduced statistical power.

However, we believe that there is an opportunity to explore significant differences in other components not covered in this study, to incorporate them into the regression equation and enhance their predictive power.

5. Conclusions

We have observed significant differences in the amplitudes recorded in the ERP components P3, FN4, and LPC in false memories induced by a DRM paradigm. No differences were observed between semantically and orthographically related items in the amplitudes of these components or in the behavioral data.

The main limitation of this study was the loss of participants, which could have been mitigated by adding a greater number of items to the identification lists.

In future research, in addition to continuing the line of investigation related to predicting false memories by incorporating new ERP components into the model, we propose exploring other types of electroencephalographic data analysis such as time-frequency analysis or spectral density analysis.

6. Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

7. Author Contributions: CRediT

HM contributed substantially to the conception of the work, data collection, analysis and interpretation, and drafted the manuscript. DC contributed substantially to the conception of the work, analyzed and interpreted the data, and revised the manuscript critically for important intellectual content. LS contributed substantially to the conception of the work, data collection, and revised the manuscript critically for important intellectual content. All authors provided approval for publication of the content and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

8. Acknowledgments

The authors express their gratitude to Eduardo Reynoso, María Fernanda Polanco, Hilce Fortuna and Rogert Espaillat for their assistance with the literature review and data collection. We also express our gratitude to Andrea Guaido for her assistance with the graphical abstract.

9. Data availability statement

The datasets generated for this study can be found in the Open Science Framework as "Electrophysiological Correlates and Predictive Power of an ERP Component Complex for False Memories.", https://osf.io/jksef/ (DOI 10.17605/OSF.IO/JKSEF).

References

Alakbarova, D., Hicks, J. L., & Ball, B. H. (2021). The influence of semantic context on false memories. *Memory & Cognition*, *49*(8), 1555-1567. https://doi.org/10.3758/s13421-021-01182-1

Amin, H. U., Malik, A. S., Kamel, N., Chooi, W.-T., & Hussain, M. (2015). P300 correlates with learning & memory abilities and fluid intelligence. *Journal of NeuroEngineering and Rehabilitation*, *12*(1), 87. https://doi.org/10.1186/s12984-015-0077-6

Arce, R., Selaya, A., Sanmarco, J., & Fariña, F. (2023). Implanting rich autobiographical false memories: Meta–analysis for forensic practice and judicial judgment making. *International Journal of Clinical and Health Psychology*, *23*(4). Scopus. https://doi.org/10.1016/j.ijchp.2023.100386

Armaly, M. T., & Enders, A. M. (2023). Filling in the Gaps: False Memories and Partisan Bias. *Political Psychology*, *44*(2), 281-299. Scopus. https://doi.org/10.1111/pops.12841

Ballou, M. R., & Sommers, M. S. (2008). Similar phenomena, different mechanisms: Semantic and phonological false memories are produced by independent mechanisms. Memory & cognition, 36(8), 1450-1459.

Bamatraf, S., Hussain, M., Aboalsamh, H., Qazi, E.-U.-H., Malik, A. S., Amin, H. U., Mathkour, H., Muhammad, G., & Imran, H. M. (2016). A System for True and False Memory Prediction Based on 2D and 3D Educational Contents and EEG Brain Signals. *Computational Intelligence and Neuroscience*, *2016*, 8491046. https://doi.org/10.1155/2016/8491046

Bland, C. E., Howe, M. L., & Knott, L. (2016). Discrete emotion-congruent false memories in the DRM paradigm. *Emotion*, *16*(5), 611-619. https://doi.org/10.1037/emo0000153

Bookbinder, S. H., & Brainerd, C. J. (2016). Emotion and false memory: The context–content paradox. *Psychological Bulletin*, *142*(12), 1315-1351. https://doi.org/10.1037/bul0000077

Brainerd, C. J., & Reyna, V. F. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review*, *10*(1), 3-47. https://doi.org/10.1016/0273-2297(90)90003-M

Brainerd, C. J., Stein, L. M., Silveira, R. A., Rohenkohl, G., & Reyna, V. F. (2008). How Does Negative Emotion Cause False Memories? *Psychological Science*, *19*(9), 919-925. https://doi.org/10.1111/j.1467-9280.2008.02177.x

Chadwick, M. J., Anjum, R. S., Kumaran, D., Schacter, D. L., Spiers, H. J., & Hassabis, D. (2016). Semantic representations in the temporal pole predict false memories. *Proceedings of the National Academy of Sciences*, *113*(36), 10180-10185. https://doi.org/10.1073/pnas.1610686113

Chen, H., Voss, J., & Guo, C. (2012). Event-related brain potentials that distinguish false memory for events that occurred only seconds in the past. *Behavioral and brain functions: BBF*, *8*, 36. https://doi.org/10.1186/1744-9081-8-36

Coane, J. H., McBride, D. M., Huff, M. J., Chang, K., Marsh, E. M., & Smith, K. A. (2021). Manipulations of List Type in the DRM Paradigm: A Review of How Structural and Conceptual Similarity Affect False Memory. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.668550

Davis, D., & Loftus, E. F. (2007). Internal and external sources of misinformation in adult witness memory. En *The handbook of eyewitness psychology, Vol I: Memory for events* (pp. 195-237). Lawrence Erlbaum Associates Publishers.

Deeny, S., Chicoine, C., Hargrove, L., Parrish, T., & Jayaraman, A. (2014). A simple ERP method for quantitative analysis of cognitive workload in myoelectric prosthesis control and human-machine interaction. *PLoS ONE*, *9*(11). Scopus. https://doi.org/10.1371/journal.pone.0112091

Ellis, H. C., Thomas, R. L., & Rodriguez, I. A. (1984). Emotional mood states and memory: Elaborative encoding, semantics processing, and cognitive effort. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(3), 470-482. https://doi.org/10.1037/0278-7393.10.3.470

Favre, G., Horat, S. K., François R, Herrmann, Isabelle, Gothuey, Marco CG, Merlo, & Pascal, Missonnier. (2020). *Neurophysiological Signature of Memory Performance during the DRM Task*. https://doi.org/10.23937/2690-3172/1710007

Gatti, D., Marelli, M., Mazzoni, G., Vecchi, T., & Rinaldi, L. (2023). Hands-on false memories: A combined study with distributional semantics and mouse-tracking. *Psychological Research*, *87*(4), 1129-1142. Scopus. https://doi.org/10.1007/s00426-022-01710-x

Gongora, M., Nicoliche, E., Magalhães, J., Vicente, R., Teixeira, S., Bastos, V. H., Bittencourt, J., Cagy, M., Basile, L. F., Budde, H., Velasques, B., & Ribeiro, P. (2021). Event-related potential (P300): The effects of levetiracetam in cognitive performance. *Neurological Sciences*, *42*(6), 2309-2316. Scopus. https://doi.org/10.1007/s10072-020-04786-8

Greenberg, D. L. (2004). President Bush's false [flashbulb] memory of 9/11/01. *Applied Cognitive Psychology*, *18*(3), 363-370. https://doi.org/10.1002/acp.1016

Greenspan, R. L., & Loftus, E. F. (2020). Eyewitness confidence malleability: Misinformation as post-identification feedback. *Law and Human Behavior*, *44*(3), 194-208. Scopus. https://doi.org/10.1037/lhb0000369

Griffin, N. R., & Schnyer, D. M. (2020). Memory distortion for orthographically associated words in individuals with depressive symptoms. *Cognition*, *203*, 104330. https://doi.org/10.1016/j.cognition.2020.104330

Hamann, S. (2009). Towards Understanding Emotion's Effects on Memory. *Emotion Review*, *1*(2), 114-115. https://doi.org/10.1177/1754073908100433

Harwood, V., Kleinman, D., Puggioni, G., & Baron, A. (2022). The P300 event related potential predicts phonological working memory skills in school-aged children. *Frontiers in Psychology*, *13*, 918046. https://doi.org/10.3389/fpsyg.2022.918046

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

Käthner, I., Wriessnegger, S. C., Müller-Putz, G. R., Kübler, A., & Halder, S. (2014). Effects of mental workload and fatigue on the P300, alpha and theta band power during operation of an ERP (P300) brain-computer interface. *Biological Psychology*, *102*(1), 118-129. Scopus. https://doi.org/10.1016/j.biopsycho.2014.07.014

Kensinger, E. A. (2007). Negative Emotion Enhances Memory Accuracy: Behavioral and Neuroimaging Evidence. *Current Directions in Psychological Science*, *16*(4), 213-218. https://doi.org/10.1111/j.1467-8721.2007.00506.x

Kensinger, E. A., & Corkin, S. (2003). Effect of Negative Emotional Content on Working Memory and Long-Term Memory. *Emotion*, *3*(4), 378-393. https://doi.org/10.1037/1528-3542.3.4.378

Kiat, J. E., & Belli, R. F. (2017). An exploratory high-density EEG investigation of the misinformation effect: Attentional and recollective differences between true and false perceptual memories. *Neurobiology of Learning and Memory*, *141*, 199-208. https://doi.org/10.1016/j.nlm.2017.04.007

Knott, L., Wilkinson, S., Hellenthal, M., Shah, D., & Howe, M. L. (2022). Generative processing and emotional false memories: A generation "cost" for negative false memory formation but only after delay. *Cognition and Emotion*. https://www.tandfonline.com/doi/abs/10.1080/02699931.2022.2128063

Kuipers, J. R., & Thierry, G. (2011). N400 amplitude reduction correlates with an increase in pupil size. *Frontiers in Human Neuroscience*, *JUNE*. Scopus. https://doi.org/10.3389/fnhum.2011.00061

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*(1), 621-647. https://doi.org/10.1146/annurev.psych.093008.131123

Li, Y., Ni, Z., He, R., Zhang, J., Zhang, Z., Yang, S., & Yin, N. (2021). The influence of positive emotion and negative emotion on false memory based on EEG signal analysis. *Neuroscience Letters*, *764*, 136203. https://doi.org/10.1016/j.neulet.2021.136203

Luck, S. J., & Kappenman, E. S. (Eds.). (2013). The Oxford handbook of event-related potential components. Oxford university press.

MacNamara, A., Jackson, T. B., Fitzgerald, J. M., Hajcak, G., & Phan, K. L. (2019). Working Memory Load and Negative Picture Processing: Neural and Behavioral Associations With Panic, Social Anxiety, and Positive Affect. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *4*(2), 151-159. Scopus. https://doi.org/10.1016/j.bpsc.2018.04.005

MacNamara, A., & Proudfit, G. H. (2014). Cognitive load and emotional processing in generalized anxiety disorder: Electrocortical evidence for increased distractibility. *Journal of Abnormal Psychology*, *123*(3), 557-565. Scopus. https://doi.org/10.1037/a0036997

Morett, L. M., Landi, N., Irwin, J., & McPartland, J. C. (2021). *N400 amplitude, latency, and variability reflect temporal integration of beat gesture and pitch accent during language processing*. https://doi.org/10.1016/j.brainres.2020.147059

Murphy, G., Loftus, E. F., Grady, R. H., Levine, L. J., & Greene, C. M. (2019). False Memories for Fake News During Ireland's Abortion Referendum. *Psychological Science*, *30*(10), 1449-1459. https://doi.org/10.1177/0956797619864887

Nichols, R. M., & Loftus, E. F. (2019). Who is susceptible in three false memory tasks? *Memory*, *27*(7), 962-984. Scopus. https://doi.org/10.1080/09658211.2019.1611862

Otgaar, H., Howe, M. L., & Patihis, L. (2022). What science tells us about false and repressed memories. *Memory*. https://www.tandfonline.com/doi/abs/10.1080/09658211.2020.1870699

Sara, G., Gordon, E., Kraiuhin, C., Coyle, S., Howson, A., & Meares, R. (1994). The P300 ERP component: An index of cognitive dysfunction in depression? *Journal of Affective Disorders*, *31*(1), 29-38. https://doi.org/10.1016/0165-0327(94)90124-4

Simpson, T. G., & Rafferty, K. (2021). Investigating the P300 Response as a Marker of Working Memory in Virtual Training Environments. *IEEE Transactions on Human Machine Systems*. https://doi.org/10.1109/THMS.2021.3064815

Spets, D. S., Karanian, J. M., & Slotnick, S. D. (2021). False memories activate distinct brain regions in females and males. *Neuroimage: Reports*, *1*(4), 100043. https://doi.org/10.1016/j.ynirp.2021.100043

Stanek, M. L., Boaz, K. M., Cordes, C. N., Niese, T. D., Long, K. E., Risner, M. S., Blasco, J. G., Suzelis, K. N., Siereveld, K. M., Rorabaugh, B. R., & Zoladz, P. R. (2024). Social evaluative stress enhances central detail memory, reduces false memory, and results in intrusive memories that last for days. *Neurobiology of Learning and Memory*, *209*, 107906. https://doi.org/10.1016/j.nlm.2024.107906

Steffens, M. C., & Mecklenbräuker, S. (2007). False memories: Phenomena, theories, and implications. *Zeitschrift für Psychologie/Journal of Psychology*, *215*(1), 12-24. https://doi.org/10.1027/0044-3409.215.1.12

Steiner, G. Z., Barry, R. J., & Gonsalvez, C. J. (2013). Can working memory predict target-to-target interval effects in the P300? *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *89*(3), 399-408. https://doi.org/10.1016/j.ijpsycho.2013.07.011

Volz, K., Stark, R., Vaitl, D., & Ambach, W. (2019). Event-related potentials differ between true and false memories in the misinformation paradigm. *International Journal of Psychophysiology*, *135*, 95-105. https://doi.org/10.1016/j.ijpsycho.2018.12.002

Voss, J. L., & Paller, K. A. (2007). Neural correlates of conceptual implicit memory and their contamination of putative neural correlates of explicit memory. *Learning & Memory*, *14*(4), 259-267. https://doi.org/10.1101/lm.529807

Wang, C., Li, Y., Wang, L., Liu, S., & Yang, S. (2023). A study of EEG non-stationarity on inducing false memory in different emotional states. *Neuroscience Letters*, *809*, 137306. https://doi.org/10.1016/j.neulet.2023.137306

Wixted, J. T., Wells, G. L., Loftus, E. F., & Garrett, B. L. (2021). Test a Witness's Memory of a Suspect Only Once. *Psychological Science in the Public Interest*, *22*(1_suppl), 1S-18S. Scopus. https://doi.org/10.1177/15291006211026259

Xie, W., & Zhang, W. (2017). Negative emotion enhances mnemonic precision and subjective feelings of remembering in visual long-term memory. *Cognition*, *166*, 73-83. https://doi.org/10.1016/j.cognition.2017.05.025

Yüvrük, E., & Kapucu, A. (2022). False (or biased) memory: Emotion and working memory capacity effects in the DRM paradigm. *Memory & Cognition*, *50*(7), 1443-1463. https://doi.org/10.3758/s13421-022-01298-y

Zimmerman, C. A., & Kelley, C. M. (2010). "I'll remember this!" Effects of emotionality on memory predictions versus memory performance. *Journal of Memory and Language*, *62*(3), 240-253. https://doi.org/10.1016/j.jml.2009.11.004