Neural Basis of the Effect of EMDR Therapy

Study Protocol and Statistical Analysis Plan

26/11/2023

Participants

Using the G*Power 3.1.9.7 software (Faul et al., 2009), an a priori power analysis was conducted to determine the required sample size. The analysis was based on a mixed-design ANOVA framework, consistent with the primary hypothesis testing the interaction between Group (experimental vs. control) and Time (pre- vs. post-stimulation). As no previous ERP studies had specifically examined the effects of Butterfly Tapping on LPP modulation, the expected effect size was set to a conventional medium value (Cohen's f=0.25), following Cohen's guidelines (Cohen, 1988). The significance level (α) was set at 0.05, and statistical power (1−β) was set to 0.95. Under these assumptions, the estimated minimal sample size was 36 participants. To account for potential data loss due to EEG artifacts or technical issues, a total of 38 university students and staff were initially recruited.

This power estimation represents a conservative approximation of the effective statistical sensitivity of the study, as it does not explicitly model features known to enhance power in ERP research, including the repeated-measures structure, the use of differential waveforms (negative minus neutral), and the averaging of a large number of trials across spatially defined electrode clusters. From a sensitivity perspective, the final sample size can therefore be considered sufficient to detect Group × Time interaction effects of small-to-medium magnitude, which is consistent with effect sizes typically reported in studies examining LPP modulation during emotion regulation and affective interventions. Consequently, the study was adequately powered to address its primary neurophysiological hypotheses.

Participants were healthy, aged between 18 and 40 years (mean age = 25.1 years; SD = 3.0) (55% of females) with normal or corrected-to-normal vision. Inclusion criteria required full right-handedness as self-reported (Edinburgh Handedness Inventory; Oldfield, 1971), no history of neurological or psychiatric disorders, and no current use of psychoactive medications. All participants were naïve to the purpose of the study. Data was collected between July 2024 and February 2025. Participants were pseudo-randomly assigned to the experimental (Exp) or control (Con) group, which were balanced for sex and age.

Self-Report Measures

To verify the comparability of the two groups, participants completed the trait sub-scale of the State-Trait Anxiety Inventory (STAI-Y2; Spielberger et al., 1971) at the beginning of the experimental session. This 20-item self-report questionnaire evaluates stable personality-related tendencies to experience anxiety across time and situations. Independent samples t-tests confirmed no significant differences between groups ($t < 1$).

To capture transient emotional responses associated with the experimental manipulation, participants completed two tests, the State subscale of the STAI (STAI-Y1; Spielberger et al., 1971) and the Visual Analogue Scale (VAS; Gift, 1989), immediately after the Tasks described below. Specifically, the STAI-Y1 consists of 20 items assessing how anxious the participant feels "at the moment", rated on a 4-point Likert scale ranging from 1 ("not at all") to 4 ("very much so"). It provides a reliable index of momentary anxiety and is sensitive to situational changes, making it well-suited to evaluating the

effects of affective processing interventions. Furthermore, the VAS was used to quantify subjective emotional discomfort related to the negative emotional content of the task. Participants were asked to indicate the intensity of discomfort experienced while viewing the negative images, with endpoints labeled “no discomfort” (0) and “extreme discomfort” (100). The VAS is a well-validated tool for capturing subtle affective changes in a fast and intuitive manner and is particularly useful in protocols involving repeated measures.

2.3. Stimuli

A total of 240 images were selected from the Nencki Affective Picture System (NAPS; Marchewka et al., 2014), a standardized database widely used for emotion research. Specifically, this database includes realistic photographs categorized by five content types: people, faces, animals, objects, and landscapes, each rated along three affective dimensions: valence (from 1 = very negative to 9 = very positive), arousal (from 1 = very calm to 9 = highly arousing), and approach/avoidance motivation (from 1 = strong tendency to avoid to 9 = strong tendency to approach).

In the present study, the images used were divided into two emotional categories (negative and neutral), and each category was presented across two separate tasks (Task 1 and Task 2). Each task included 60 images of the same emotional valence, resulting in 60 negative images in Task 1, 60 negative images in Task 2, 60 neutral images in Task 1, and 60 neutral images in Task 2. Importantly, the image sets used in Task 1 and Task 2 were not repeated across time points. Specifically, each participant viewed one set at T0 and the other at T1, with the set assignment randomized across participants. This design allowed us to evaluate the generalizability of any observed effects to novel, previously unseen stimuli. All images were selected based on normative valence ratings, approach–avoidance motivation, and arousal, ensuring internal consistency within emotional cate-gories (see Table 1).

Table 1. Mean ratings of valence, avoidance/approach motivation (Av/Ap), and arousal of the 240 images included in the tasks.

Valence	Av/Ap	Arousal	
Negative – Task 1	2.57	2.86	6.86
Negative – Task 2	2.66	2.87	6.73
Neutral – Task 1	5.53	5.49	4.67
Neutral – Task 2	5.54	5.44	4.72

2.4. Task

Participants were comfortably seated in a dimly lit room, approximately 100 cm from a 32” 16:9 computer monitor. Visual stimuli were presented using Presentation Software (Neurobehavioral Systems, Albany, CA, USA). A small yellow fixation point (0.15×0.15° of visual angle) on a grey background remained visible at the center of the screen throughout the experiment to maintain gaze stability and minimize ocular artifacts. The experimental task was a variant of a simple detection task, where the participant had to respond as soon as possible to the visual stimuli described above using a response key with their right index finger, regardless of their emotional content.

In each trial, images were presented in full-screen mode (subtending 35.4 x 31.7°) at the center of the display for 350 ms, then the fixation point remained for a randomized inter-stimulus interval ranging from 1200 to 3000 ms. The task was structured into 12 consecutive runs, alternating in emotional valence: 6 runs of negative images and 6 runs of neutral images. Each run contained 60 images of the

same valence. The order of runs (starting with neutral or negative) was counterbalanced across participants to control for sequence effects. Each run lasted about 2 min and 30 sec, resulting in a recording session duration of approximately 35 min, including breaks. For each valence, 360 trials were collected, for a total of 720 trials per participant. This task has been executed during EEG recordings before and after the stimulation session described below.

Stimulation

Following the first task session, participants executed a 15-minute self-administered tactile stimulation session, which differed depending on group assignment. The Exp group performed the BT (Artigas & Jarero, 2014). The posture followed the traditional butterfly hug configuration (Figure 1): participants were seated on a chair and were instructed to cross their hands and rest their fingertips gently over the suprasternal notch, with thumbs crossed and hands positioned just below the clavicles. From this position, they had to alternate bilateral tapping with their fingertips while keeping their eyes closed and mentally focusing on the negative images previously viewed during the task to maintain affective engagement across the session and ensure continuity between the pre- and post-stimulation EEG recordings. This instruction was intended to prevent disengagement from the emotional context during the stimulation interval. The tapping rhythm was modelled on the original Butterfly Hug procedure (Artigas & Jarero, 2014), and was kept consistent across participants, approximating a rate of one tap per second for standardization purposes in the experimental setting. On the other hand, in the Con group, the posture and instructions were identical. However, instead of rhythmic alternate tapping, participants rested their hands in place without movement (static stimulation). The posture-matched static control condition was selected to provide a conservative baseline by maintaining identical posture, environmental setting, and task structure across conditions. This control was designed to minimize confounds related to body position, visual input, and general task demands. Importantly, the present control condition was not in-tended to function as an active motor or attentional control, but rather as a conservative comparison allowing assessment of whether observed changes in affective neural responsiveness exceeded those attributable to time-on-task or repeated exposure alone.

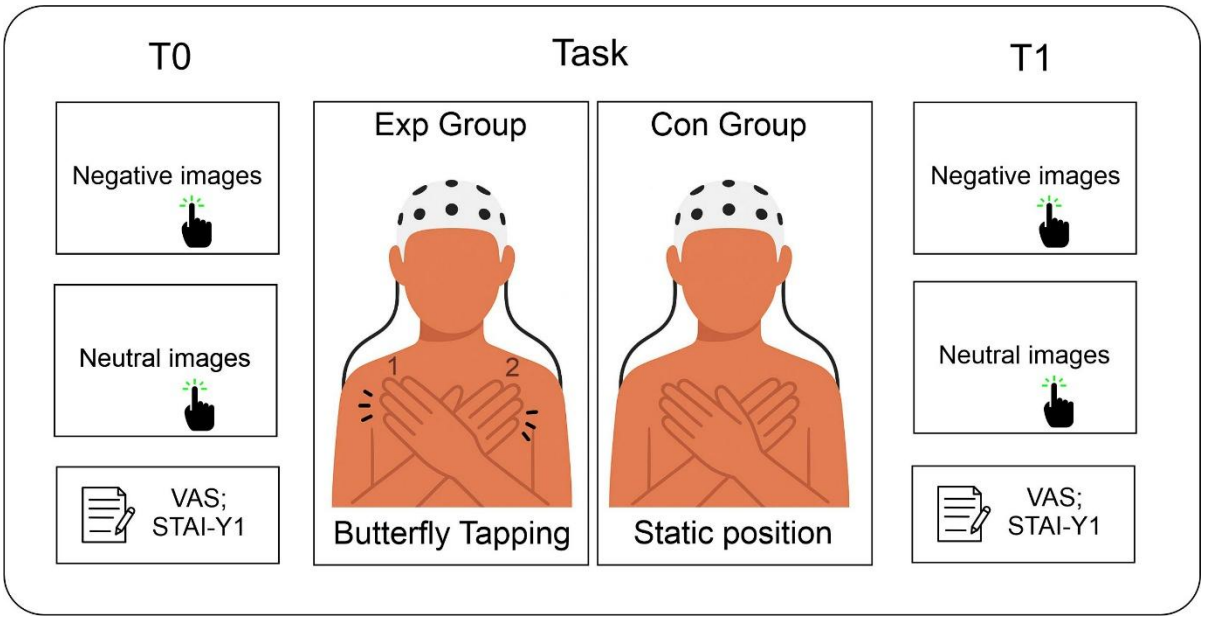

Figure 1. Schematic representation of experimental design. At T0, participants performed the task and the State-Trait Anxiety Inventory – State version (STAI-Y1) and the Visual Analogue Scale (VAS) for discomfort. Participants assigned to the experimental (Exp) group performed a 15-minute self-administered bilateral tapping session (Butterfly Tapping), while subjects of the control (Con) group

adopted the same posture without performing any tapping (static position). At T1, participants repeated the task, followed again by the STAI-Y1 and the VAS. During the task, EEG was continuously recorded.

The session was structured into four blocks of 3 minutes each, interleaved with one-minute pauses. This structure provides sufficient exposure to stimulation while minimizing muscular fatigue and maintaining attention. Before the stimulation, the participants rehearsed the alternating pattern until a consistent and relaxed tempo was achieved. No auditory pacing was used during the actual session to preserve a spontaneous rhythm.

Experimental Procedure

The participants were fitted with the EEG cap and received instructions for the upcoming tasks. The experiment followed a pre-post design with three consecutive phases (Figure 1): Baseline assessment (T0), a 15-minute stimulation session, and post-assessment (T1). At both T0 and T1, participants performed the task while the EEG was recorded and completed the STAI-Y1 and the VAS.

EEG Recording and ERP Analysis

During the task, continuous EEG was recorded using BrainProducts GmbH (Gilching, Germany) equipment, including Recorder 1.2 software and three BrainAmp amplifiers. Two of the amplifiers were connected to a 64-channel ActiCap active electrode system, mounted according to the 10-10 International System. EEG data were sampled at 250 Hz and band-pass filtered offline using a second-order zero-phase Butterworth filter (0.01 – 60 Hz) with an additional 50 Hz notch filter. All signals were referenced to the average of M1 and M2 electrodes. Bipolar electrooculographic (EOG) activity was recorded using a third BrainAmp ExG amplifier, with electrodes placed at the outer canthi of left and right eyes (horizontal EOG), and above and below the left eye (vertical EOG). Electrode impedances were kept below 5 kΩ. On average, 6.7 ± 3.4 electrodes were interpolated in the Con group and 6.1 ± 4.2 in the Exp group. Ocular artifacts (blinks and movements) were removed using the Gratton and Coles algorithm (Gratton et al., 1983), and residual noise was excluded via semi-automatic artifact rejection, discarding epochs with amplitudes exceeding ± 70 μV. After artifact rejection, an average of 8.1% of trials were excluded from the Con group and 6.7% from the Exp group. To measure post-stimulus activity, 1200 epochs were extracted time-locked to stimulus onset, ranging from 200 ms before to 1000 ms after it. The 200 ms pre-stimulus interval served as the baseline. For each subject, ERP waveforms were averaged separately for negative and neutral stimuli. ERP components were identified using the "collapsed localizer" approach (Luck & Gaspelin, 2017), averaging across all groups and conditions to identify the electrodes to include in statistical analysis. Furthermore, to define analysis intervals, the Global Field Power (GFP) was computed. Specifically, the GFP is a reference-free measure that summarizes the overall distribution of electrical activity across the scalp into a single waveform, providing a measure of the global strength of the ERP signal at each time point. It is computed as the root mean square of the voltage values recorded at all electrodes, simultaneously, across the scalp. The main GFP peaks were used to identify the most prominent ERP components, and the time windows were determined by including the intervals in which the GFP amplitude exceeded 80% of the local maximum. Similarly, corresponding electrodes showing at least 80% of the maximum amplitude within each interval were grouped into spatial pools. Two separate GFP waveforms were generated: one based on the grand average of all post-stimulus activity, and another based on the differential waveforms obtained by subtracting neutral from negative trials. While the first waveform displayed multiple peaks, including those corresponding to early and late visual components, the second one highlighted two time-windows with strong affective modulation: Frontal Positivity (188–228 ms) and central LPP (cLPP; 456–552 ms). For statistical analysis, we focused on the components extracted by the difference waves, capturing the emotion-related

modulation of interest. The electrodes included in the analysis were selected according to the scalp distribution of maximal amplitude in this window, averaging all sites whose activity exceeded 80% of the GFP peak. This procedure identified two electrode pools: a frontal cluster including F1, Fz, F2, FC1, FCz, and FC2, and a centro-parietal cluster comprising Cz, C1, C2, CP1, CPz, and CP2, both consistent with the canonical distribution of the LPP component (Hajcak et al., 2010). Mean amplitudes in this time window and pool were extracted for each condition and used in subsequent statistical comparisons.

Statistical Analysis

Prior to statistical testing, the normality assumption for all variables was assessed using the Shapiro–Wilk test. None of the distributions deviated significantly from normality. The assumption of homogeneity of variance was tested using Levene's test, which confirmed that the data met the criterion for homoscedasticity. The behavioral and psychological measures were analyzed using a 2×2×2 ANOVA, with Group (Exp vs. Con) as a between-subjects factor and Time (T0 vs. T1) and Valence (Negative vs. Neutral) as a within-subjects factor. Furthermore, a 2×2 ANOVA was conducted on the mean amplitudes of the Frontal Positivity and cLPP difference waves, with Group and Time as factors. Effect sizes were reported using partial eta squared (ηp^2). Where significant main effects or interactions were found, Bonferroni-corrected post-hoc tests were applied. All tests adopted an α threshold of 0.05. Analyses were performed using Statsoft Statistica 12.0 (StatSoft Inc., Tulsa, OK, USA).