**Emotion Regulation Task**

**Procedure**

Participants completed an emotion regulation task directly before (T0) and after (T1) the neurofeedback training as well as after 6 weeks (T2). They were instructed either to view negative and neutral pictures without modifying their emotions or to down-regulate their feelings toward negative pictures. Furthermore, participants were instructed not to turn away their gaze or to close their eyes, nor to focus exclusively on non-emotional parts of the picture. The participants’ eyes were tracked by a camera system (SMI BeGaze, Teltow, Germany) to encourage subjects to stay with the task but data were not analyzed. Trial order was pseudorandomized and counterbalanced with no more than two consecutive conditions of the same type. 10 % of trials did not include a startle probe (2 trials per condition). Before the emotion regulation task began, 10 startle probes were presented consecutively without any picture stimuli. In total, the paradigm consisted of 36 trials (12 trials per condition) and lasted 20 min.

Each trial began with a 2000-ms presentation of an instructional cue (view, down-regulate), followed by a fixation cross displayed for 1,000 ms. Next, a neutral or negative picture was presented for 7,500 ms. A startle probe (50 ms, 95dB white noise burst) was presented through headphones at 7s into the regulation phase. Probe time was jittered between 1000s and 3000s. Self-assessment Manikins (SAM Ratings; Bradley & Lang, 1994) were presented after presentation of each picture. Participants rated on a 1-9 Likert scale how positive/negative and aroused/calm they felt at that moment. Lower scores on the valence scale indicate that they felt more positive; lower scores on the arousal scale indicate that they felt calmer. By pressing buttons on a keyboard, subjects moved a bar from left to right to select SAMs corresponding to their subjective valence and arousal. The initial bar position was random and the final position of the bar at the end of the rating was logged. Valence and arousal rating scales were displayed consecutively for 5,000 ms each. Intertrial intervals were jittered between 3,500 and 5,500 ms.

**Picture Stimuli**

Stimuli were taken from the standardized picture series (Lang, Bradley, & Cuthbert, 2008; Marchewka, Żurawski, Jednoróg, & Grabowska, 2014) and were presented with the Presentation software (Neurobehavioral Systems, Berkeley, CA) in semi-randomized order with restriction of no more than two consecutive trials from the same condition, and no more than three consecutive trials with negative pictures. For each assessment (T0, T1 and T2) sets of 24 negative pictures (valence: M = 2.36, SD = .68; arousal: M = 6.86, SD = .23) and 12 neutral pictures (valence: M = 5.21, SD = .59; arousal: M = 2.57, SD = .26) were created (normative ratings based on representative samples (Lang et al., 2008; Marchewka et al., 2014)). Arousal and valence ratings differed significantly between the sets (both ps <.001). The pictures were further divided into two sets (balanced for content, valence, and arousal), which resulted in three conditions depending on instruction and picture type: View neutral pictures (LookNeu), view negative pictures (LookNeg) and down-regulate emotions while viewing negative pictures with reappraisal (RegNeg). Assignment of negative picture sets to LookNeg and RegNeg condition was alternated between subjects.

**Measures**

**Emotion-modulated Startle**

The eye blink was measured by electromyogram (EMG). Two Ag-AgCl electrodes were placed on the orbicularis oculi muscle below the left eye, and a ground electrocardiogram electrode was attached on the lower rip bow1. The raw EMG signal was sampled at 1000 Hz, and the gain was amplified by 2000. High-pass (50 Hz) and low-pass (500 Hz) filters were applied to the data with AcqKnowledge software (BIOPAC Systems; Goleta, CA). EMG data were integrated over 10 samples and analyzed offline with Clip, a C++based, semi-automated program (Kinzig, Schulz, Curio, & Schächinger, 2008). Startle response was defined as the difference between peak (20–120 ms after stimulus onset) and baseline (20 ms prior to stimulus onset) signal. Trials including movement artifacts, excessive baseline activity (exceeding 2 standard deviations [SD] above baseline mean), or non-responses (peak < four SD above baseline mean) were excluded (mean % = 9.56 [SD = 6.52]) of all trials across participants). Startle data from one participant was excluded because of excessive noise (more than 30% missing). Finally, amplitudes were z-standardized within participants and transformed to T-scores with mean = 50 and SD = 10. Responses were averaged across participants for each condition.