Social Cognitive and Affective Neuroscience, 2019, 1119–1129

doi: 10.1093/scan/nsz089
Advance Access Publication Date: 10 December 2019
Original article

Aberrant reward prediction error during Pavlovian appetitive learning in alexithymia

Francesca Starita, Mattia Pietrelli, Caterina Bertini, and Giuseppe di Pellegrino

Department of Psychology, Center for Studies and Research in Cognitive Neuroscience, University of Bologna, 40126 Bologna (BO), Italy

Correspondence should be addressed to Giuseppe di Pellegrino, Center for Studies and Research in Cognitive Neuroscience, Department of Psychology, University of Bologna Viale Europa, 980 47521 Cesena (FC), Italy. E-mail: g.dipellegrino@unibo.it

Abstract

Extensive literature shows that alexithymia, a subclinical trait defined by difficulties in identifying and describing feelings, is characterized by multifaceted impairments in processing emotional stimuli. Nevertheless, its underlying mechanisms remain elusive. Here, we hypothesize that alexithymia may be characterized by an alteration in learning the emotional value of encountered stimuli and test this by assessing differences between individuals with low (LA) and high (HA) levels of alexithymia in the computation of reward prediction errors (RPEs) during Pavlovian appetitive conditioning. As a marker of RPE, the amplitude of the feedback-related negativity (FRN) event-related potential was assessed while participants were presented with two conditioned stimuli (CS) associated with expected or unexpected feedback, indicating delivery of reward or no-reward. No-reward (*vs* reward) feedback elicited the FRN both in LA and HA. However, unexpected (*vs* expected) feedback enhanced the FRN in LA but not in HA, indicating impaired computation of RPE in HA. Thus, although HA show preserved sensitivity to rewards, they cannot use this response to update the value of CS that predict them. This impairment may hinder the construction of internal representations of emotional stimuli, leaving individuals with alexithymia unable to effectively recognize, respond and regulate their response to emotional stimuli.

Key words: alexithymia; prediction error; reinforcement learning; feedback-related negativity; Pavlovian conditioning

Alexithymia is a subclinical trait defined by difficulties in identifying and describing feelings, is a style of thinking devoid of introspection and affective thinking (Sifneos, 1973; Taylor *et al.*, 1991) and affects about 10% of the general population (Taylor *et al.*, 1991). Individuals with alexithymia have multifaceted impairments in processing emotional stimuli, which range from the recognition of emotional stimuli, to the response to them, to the regulation of such response and its appropriate use to guide decision-making (for a comprehensive review see Luminet *et al.*, 2018; Teixeira *et al.*, 2018). Crucially, despite this evidence, the basic mechanisms that may underlie such difficulties remain poorly understood. We have previously argued that alexithymia may be conceptualized as an impairment in updating the value of encountered stimuli, and we showed that individuals with alexithymia have reduced learning of the aversive value of conditioned stimuli (CS) during Pavlovian threat conditioning, despite preserved response to unconditioned stimuli (US) (Starita *et al.*, 2016). Therefore, individuals with alexithymia appear able to respond to stimuli that are biologically prepared to trigger an emotional response. Nevertheless, they appear unable to exploit this response to actively shape the internal representation of aversive stimuli to encompass, alongside stimuli that unconditionally elicit an emotional response, those that co-occur with them. Here, we extend this investigation and ask whether such difficulty is present also when individuals with alexithymia have to learn the value of appetitive stimuli.

Received: 29 October 2018; **Revised:** 2 September 2019; **Accepted:** 30 September 2019

Reinforcement learning theories argue that reward prediction errors (RPEs; i.e. the difference between expected and experienced reward; Schultz, 1998) drive learning about the value of stimuli in the environment (Daw and Tobler, 2014). When a neutral stimulus co-occurs with an unexpected reward, the RPE is computed by updating the affective value of the neutral stimulus, which acquires an appetitive connotation. Electroencephalographic studies suggest that the feedback-related negativity (FRN) event-related potentials (ERPs) may encode an RPE-like signal (Holroyd and Coles, 2002; Nieuwenhuis *et al.*, 2004; Walsh and Anderson, 2012; Sambrook and Goslin, 2015). The FRN is a negative deflection in electrical potential observed at fronto-central electrodes between 200 and 350 ms after feedback presentation and results from subtracting the potential following feedback indicating reward omission from that following reward delivery (Sambrook and Goslin, 2015). Importantly, this component is modulated by expectation, such that it is more negative for unexpected compared to expected reward-related feedback (e.g. Donkers and Boxtel, 2005; Potts *et al.*, 2006; Hajcak *et al.*, 2007; Holroyd and Krigolson, 2007; Holroyd *et al.*, 2011; Walsh and Anderson, 2011, 2013).

Given the above information, we hypothesized alexithymia to be related to altered computation of RPE during Pavlovian appetitive conditioning. To test this, differences in the amplitude of the FRN were assessed in individuals with high (HA) and low (LA) levels of alexithymia in response to expected and unexpected reward-related feedback. Two CS (CS1 and CS2) were presented to participants. In 80% of trials, CS1 was followed by feedback indicating delivery of monetary reward, while CS2 of no-reward, constituting an expected reward-related feedback condition. Importantly, the task also included two conditions that violated such expectations. In the remaining 20% of trials, the stimulus–feedback association was inverted, resulting in an unexpected reward-related feedback. In LA, this condition was hypothesized to lead to the computation of a RPE, manifested as an enhanced FRN in response to the unexpected compared to the expected reward-related feedback. On the contrary, HA was hypothesized to show a deficit in the computation of such RPE, showing reduced modulation of the FRN compared to LA. In addition to electrophysiological measures, we collected verbal reports of subjective liking of the two CS and of the contingency between each CS and reward-related feedback, in order to test whether alexithymia also affects explicit aspects of emotional learning.

Methods

Participants

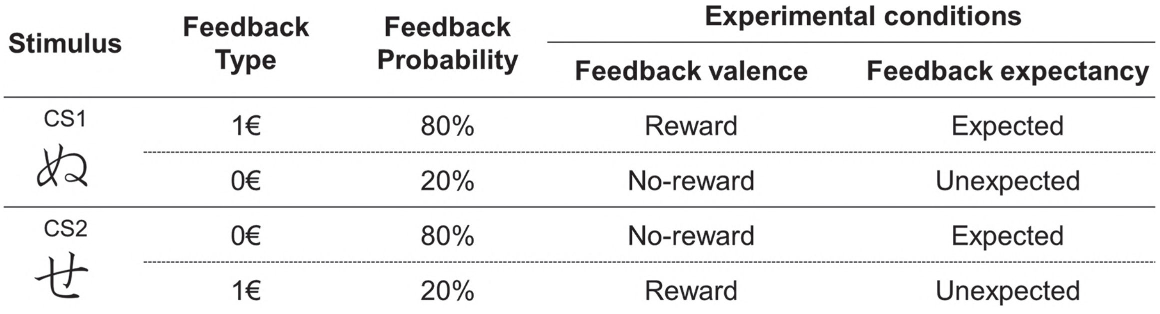
A total of 300 individuals completed the 20-item Toronto Alexithymia Scale (TAS-20; Taylor *et al.*, 2003). Depending on the score, individuals were classified as LA (TAS-20 $\leq$ 36) or HA (TAS-20 $\geq$ 61) (Franz *et al.*, 2004) and were then randomly contacted to participate in the study. Once in the laboratory, the alexithymia module of the structured interview for the diagnostic criteria for psychosomatic research (DCPR; Mangelli *et al.*, 2006) was administered to confirm the TAS-20 classification (LA, DCPR <3; HA, DCPR ≥3). An individual with discordant classification on the two measures did not complete the task ($n = 1$). Due to the high co-occurrence of alexithymia and depression (Li *et al.*, 2015), individuals completed the Beck Depression Inventory (Beck *et al.*, 1961) and did not complete the task if their score was higher than the moderate/severe depression cut-off (i.e. 19, $n = 1$).

Following a power analysis based on effect size reported in a previous electroencephalogram (EEG) study on alexithymia and error-related negativity ($\eta_p^2 = 0.208$) (Maier *et al.*, 2016), a fronto-central component that is also considered a prediction error signal (Gehring *et al.*, 1993; Alexander and Brown, 2011) and a sample of 20 participants per group resulted appropriate for testing the main hypothesis of the study (2 groups: LA and HA × 2 expectancy: expected, unexpected repeated measures ANOVA) with power = 0.85. A total of 43 healthy volunteers were recruited (22 LA and 21 HA). Data from two LA and one HA were removed from analysis due to technical issues with the EEG recording. This left a total of 40 participants in the analysis: 20 LA (6 males; age $M = 21.42$, s.d. = 1.57 years; TAS-20 $M = 31.80$, s.d. = 2.82) and 20 HA (6 males; age $M = 21.97$, s.d. = 2.27 years; TAS-20 $M = 64.00$, s.d. = 4.30). Participants had equivalent educational background and were students at the University of Bologna. The Bioethics Committee of the University of Bologna approved the study, and participant's consent was obtained according to the Declaration of Helsinki.

Experimental task and procedure

The Pavlovian appetitive conditioning task consisted in the presentation of two CS and followed by a feedback indicating the delivery of reward or no-reward. The CS was a 3 cm white square with a Japanese hiragana on it, reward feedback consisted in the writing '1€' and no-reward feedback in the writing '0€'. In order to manipulate reward expectations, the percentage of reward delivery was varied for each CS. CS1 was followed by the delivery of reward in 80% of trials (i.e. expected reward condition) and no-reward in 20% of trials (i.e. unexpected no-reward condition). Contrarily, CS2 was followed by the delivery of no-reward in 80% of trials (i.e. expected no-reward condition) and reward in 20% of trials (i.e. unexpected reward condition; Figure 1). The task included eight blocks of 108 randomized trials (expected reward, 40 trials; expected no-reward, 40 trials; unexpected reward, 10 trials; unexpected no-reward, 10 trials; catch, 8 trials). Each trial consisted in the presentation of a fixation cross in the center of the screen (500 ms), followed by the presentation of CS to the right or left of the fixation cross (1500 ms), followed by the feedback (1000 ms) and followed by a jittered inter trial interval (ITI; 1000–1500 ms) during which a blank screen appeared (Figure 2). Catch trials consisted in the presentation of a scrambled CS followed by the ITI. After four blocks, hiragana was changed so that participants had to learn new stimulus–feedback associations. Stimuli appeared on a 17 inch colored monitor (60 Hz refresh rate), at a viewing distance of 80 cm. A PC running E-prime 2.0 (Psychology Software Tools, Pittsburgh, PA) controlled the stimulus presentation.

Participants were instructed that on each trial, a stimulus would appeared on the left or right of the screen followed by the feedback '1€' or '0€' that indicated the reward for that trial. Their task was to pay attention to the stimulus–feedback association because they would earn part of the reward at the end of the task. Importantly, no information was provided regarding which stimulus would be associated with which feedback, and participants had to learn this from experience. Additionally, as soon as the stimulus appeared, participants had to press one of the two keys on the keyboard, corresponding to the side of stimulus presentation. They were informed that neither response speed nor accuracy would influence feedback appearance, in order to eliminate the possibility of participants attributing administration of reward to their actions. Finally, participants were told that sometimes a scrambled picture (i.e. catch trial) would appear, and they should not press any key. Because reward delivery was



Stimulus	Feedback Type	Feedback Probability	Experimental conditions	
			Feedback valence	Feedback expectancy
CS1 ぬ	1€	80%	Reward	Expected
	0€	20%	No-reward	Unexpected
CS2 せ	0€	80%	No-reward	Expected
	1€	20%	Reward	Unexpected

Fig. 1. Illustration of experimental conditions with examples CS1 and CS2 stimuli.

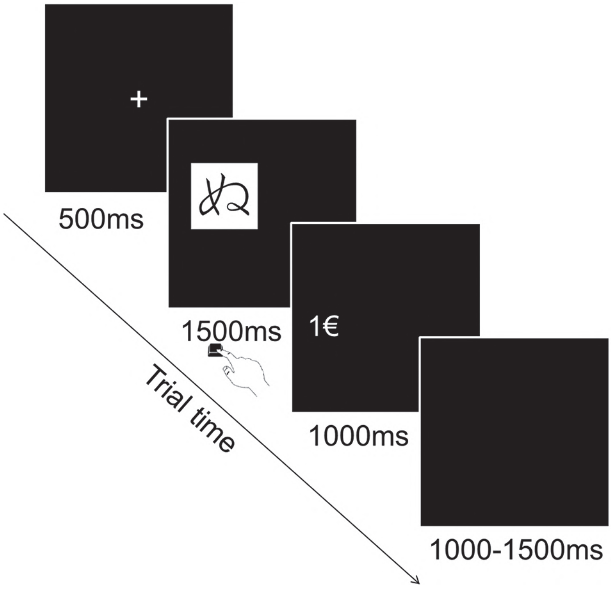


Fig. 2. Illustration of experimental trial.

not related to participants' keypress at the presentation of the CS, these catch trials were introduced to reduce the possibility that participants pressed keys without paying attention to the actual CS on the screen. Participants could take a break to rest at the end of each block. At the end of the fourth and eighth block, participants reported the number of different CS seen and described them. They also rated how much they liked them and reported what they noticed about the stimulus–feedback association (see Dependent measures below). At the end of the session, participants received 10€ as a reward.

EEG recording and pre-processing

The EEG was recorded with Ag/AgCl electrodes (Fast n Easy Electrodes, Easycap, Herrsching, Germany) from 59 electrode sites (Fp1, Fp2, AF3, AF4, AF7, AF8, F1, F2, F3, F4, F7, F8, FC1, FC2, FC3, FC4, FC5, FC6, FT7, FT8, C1, C2, C3, C4, C5, C6, T7, T8, CP1, CP2, CP3, CP4, CP5, CP6, TP7, TP8, P1, P2, P3, P4, P5, P6, P7, P8, PO3, PO4, PO7, PO8, O1, O2, FPz, AFz, Fz, FCz, Cz, CPz, Pz, POz and Oz) and from the right mastoid. The reference electrode was placed on the left mastoid and the ground electrode on the right cheek. Signal impedance was maintained below 5 KΩ. The electro-oculogram (EOG) was recorded from above and below the left eye and from the outer canthi of both eyes. The EEG and EOG were recorded with a band-pass filter of 0.01–100 Hz and a slope of 12 dB/oct, amplified by a BrainAmp DC amplifier (Brain Products, Gilching, Germany) and digitized at a sampling rate of 1000 Hz.

The EEG data were pre-processed using EEGLAB toolbox, version 14.1.0 (Delorme and Makeig, 2004) and custom routines written in MATLAB R2016b (The MathWorks, Natick, MA). The ERP data were re-referenced offline to the linked mastoid (Luck, 2014) and subjected to low- and high-pass filtering with a cut-off at 30 and 1 Hz, respectively (EEGLAB function pop_eegfiltnew). The transition band was 7.5 Hz for the low-pass filter (−6 dB/octave; 441 pts) and 1 Hz for the high-pass filter (−6 dB/octave; 3301 pts). Stimulus-locked epochs from −200 to 2500 ms relative to the appearance of the CS were extracted from the continuous EEG, in order to extract the EEG activity during the entire presentation of both CS and reward-related feedback. Epochs were baseline corrected using the average voltage during the 200 ms pre-CS window. Then, epochs whose voltage exceeded 400 μV were excluded, in order to remove epochs with large peaks. In addition, epochs whose voltage deviated more than 5 s.d. from the mean of the joint probability distribution were excluded, in order to remove trials with improbable data. The number of epochs left was as follows: LA: $M_{\text{rew\_exp}} = 303.20$ (min = 225), $M_{\text{rew\_unexp}} = 75.75$ (min = 55), $M_{\text{no-rew\_exp}} = 301.95$ (min = 202), $M_{\text{no-rew\_unexp}} = 75.25$ (min = 55); HA: $M_{\text{rew\_exp}} = 309.95$ (min = 272), $M_{\text{rew\_unexp}} = 77.05$ (min = 67), $M_{\text{no-rew\_exp}} = 309.00$ (min = 278), $M_{\text{no-rew\_unexp}} = 77.45$ (min = 66). The remaining number of epochs did not differ between groups or feedback valence (all $P \geq 0.246$) but differed as a function of feedback expectancy [$F(1,38) = 6907.049$, $P < 0.001$] with trials with unexpected feedback being less than trials with expected feedback ($M_{\text{exp}} = 304.39$, $M_{\text{unexp}} = 76.16$). Note that feedback expectancy did not interact with the factors group or feedback valence (all $P \geq 0.118$). Importantly, in the analyses that included feedback expectancy as a factor, to avoid a possible influence of differing numbers of trial on the average ERP results, the trial number of the experimental conditions was matched through data re-sampling as described in the dependent measures below. Finally, to correct the remaining artifacts, the data were subjected to a temporal independent component (IC) analysis (Jutten and Herault, 1991; Makeig et al., 1996) using the infomax algorithm (Bell and Sejnowski, 1995). The resulting component matrix was visually inspected for ICs-representing stereotyped artifact activity for horizontal (saccades) and vertical (blinks) eye movements, which were manually rejected.

Dependent measures

FRN. First, to test the role of feedback valence in eliciting the FRN, we calculated the FRN using a difference wave approach (Walsh and Anderson, 2011a, 2012, 2013), separately grouping

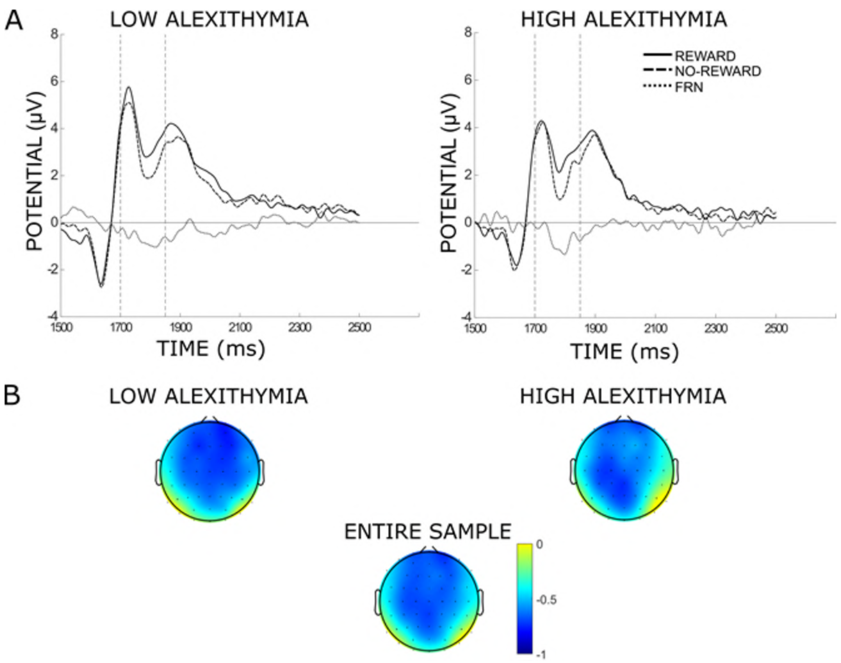


Fig. 3. (A) Grand average waveforms of low and high alexithymia group averaged between electrodes Fz and FCz for reward, no-reward conditions and FRN. A total of 1500 ms on the *x*-axis indicates time point of feedback appearance during the trial; vertical dashed lines indicate time interval for analysis. (B) Scalp topographies of the mean voltage difference between no-reward and reward feedback in the interval 200–350 ms following feedback presentation for the entire sample, the low and high alexithymia group.

early blocks (i.e. 1-2-5-6) together and late blocks (i.e. 3-4-7-8) together, enabling us to test whether the FRN amplitude changed as experiment progressed. Scalp topographies for the entire sample of the mean voltage difference between no-reward and reward feedback in the interval 200–350 ms following feedback presentation showed maximum difference at fronto-central electrodes (Figure 3B). FRN was calculated as the mean amplitude difference between no-reward and reward feedback averaged between electrodes Fz and FCz (Figure 3A), where FRN has been previously reported (for a meta-analysis see Sambrook and Goslin, 2015)).

Then, to test the role of feedback expectancy in modulating the FRN and in eliciting an RPE, we calculated the FRN separately for the expected and unexpected conditions, again by separately grouping early blocks (i.e. 1-2-5-6) together and late blocks (i.e. 3-4-7-8) together. Because the number of trials for the unexpected condition was significantly smaller than that for the expected condition (see the EEG recording and pre-processing above), the trial number was matched through data re-sampling. For each participant, we identified the condition with the smallest number of trials and the corresponding number of trials was randomly drawn from each of the remaining conditions for 1000 iterations. Then, an average ERP was calculated for each iteration separately for each condition (Garofalo *et al.*, 2016). The resulting components were then averaged over all iterations separately for each condition, producing four ERPs per participant used for data analysis. Scalp topographies for the entire sample of the mean voltage difference between no-reward and reward feedback for the expected and unexpected conditions in the interval 200–350 ms following feedback showed maximum difference in activation at fronto-central electrodes (Figure 4B). The FRN (i.e. difference between no-reward and reward feedback) for expected and unexpected conditions was calculated averaged between electrodes Fz and FCz (Figure 4A).

P300. To verify whether or not alexithymia specifically affected the FRN, P300 in response to the feedback was isolated for each condition, separately grouping early blocks (i.e. 1-2-5-6) together and late blocks (i.e. 3-4-7-8) together. The P300 is a positive deflection in electrical potential observed at posterior centro-parietal electrodes peaking 300–600 ms after stimulus presentation, which has been interpreted as a marker of attentional allocation (Polich, 2007). For example, the P300 is elicited in response to motivationally salient stimuli, such as target stimuli in attention tasks, emotional pictures and gain or loss of money (Nieuwenhuis *et al.*, 2005; Polich, 2007; Olofsson *et al.*, 2008). Here, the P300 in response to feedback was defined as the mean voltage in the interval 300–450 ms following feedback appearance averaged between electrodes Cz and CPz, where scalp topographies showed maximum activation.

Accuracy and response time to CS. The percentage of accurate responses and the average response time for accurate responses for each participant and for each CS were calculated to test differences in response to the CS, separately grouping early blocks (i.e. 1-2-5-6) together and late blocks (i.e. 3-4-7-8) together.

CS–feedback contingency awareness. At the end of the fourth and eighth block of the task, to evaluate understanding of the task and explicit learning of CS–feedback association, three open questions were asked to participants. The first asked to report how many stimuli they saw and the second to briefly describe them. All participants reported the correct number of stimuli and were able to describe them. In addition, the third question asked to report what participants noticed about the stimulus–feedback association, for each stimulus separately. For CS1, a score of 1 was given for correctly reporting that it was mostly associated with 1€; for CS2, a score of 1 was given for correctly reporting that it was mostly associated with 0€. A score of 0.5

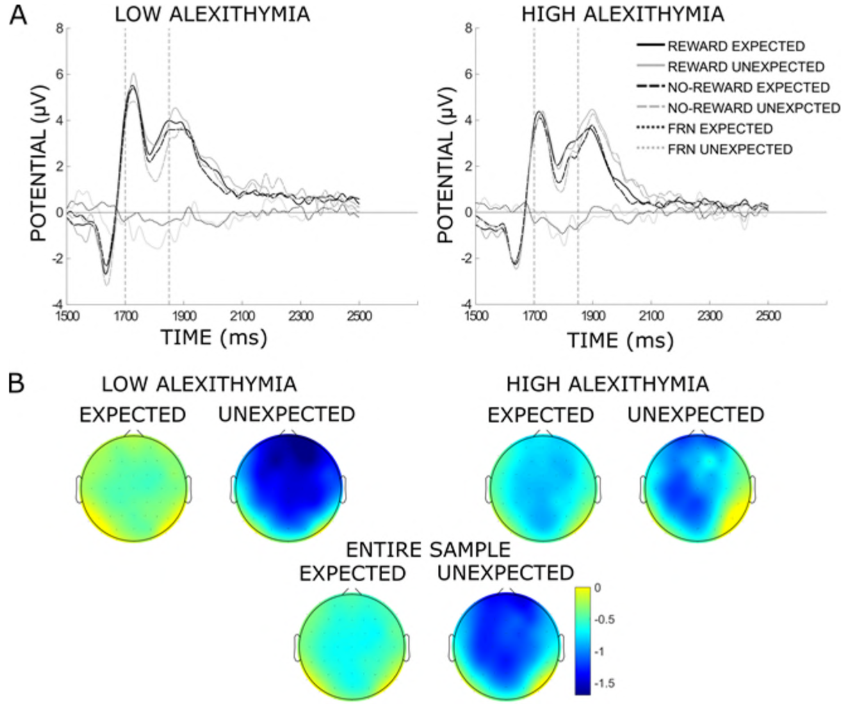


Fig. 4. (A) Grand average waveforms of low and high alexithymia group averaged between electrodes Fz and FCz for reward, no-reward and FRN in the expected and unexpected conditions. A total of 1500 ms on the *x*-axis indicates time point of feedback appearance during the trial; vertical dashed lines indicate time interval for analysis. (B) Scalp topographies of the mean voltage difference between reward and no-reward feedback for the expected and unexpected conditions in the interval 200–350 ms following feedback presentation for the entire sample, the low and high alexithymia group.

was given for reporting that the stimulus was associated with reward and no-reward but failing to identify which feedback was more probable for that stimulus. All other responses were given a score of 0. For each participant, for each CS, we calculated the median of the contingency scores collected at the end of the fourth and eighth block and run the analyses using these values. Note that repeating statistical analyses on the electrophysiological and self-report liking data, excluding participants who had a score below 1, did not significantly affect the results.

Subjective report of liking. At the end of the fourth and eighth block of the task, to assess subjective value rating of the two CS following conditioning, participants reported how much they liked each CS. They made a mark with a pen on a visual analogue scale made of a straight horizontal continuous line of 17.5 cm length with two anchors at its left and right edges, respectively, 0 (not at all) and 100 (extremely). For each participant, a liking score was calculated by measuring with a ruler the distance (cm) on the line between the not at all anchor and the participant's mark, converting it to a percentage value providing a range of scores from 0 to 100. Then, for each participant, for each CS, we calculated the median of the liking scores collected at the end of the fourth and eighth block, which was then used for statistical analyses. Data for one LA participant were missing because the software failed to record the responses.

Results

Normality of EEG data was tested and parametric analyses were conducted to test the hypotheses. Post-hoc comparisons were conducted using the Newman–Keuls test.

Table 1. Results of the one-sample *t*-tests comparing the mean amplitude of FRN against zero in low and high alexithymia group

	Early blocks	Late blocks
LA	$M = -0.67$ $t(19) = -3.30$, $P = 0.004$	$M = -0.74$ $t(19) = -3.76$, $P = 0.001$
HA	$M = -0.52$ $t(19) = -2.80$, $P = 0.011$	$M = -0.67$ $t(19) = -2.69$, $P = 0.014$

FRN

First, we tested whether reward-related feedback elicited an FRN (i.e. difference between no-reward and reward) separately in LA and HA. One-sample *t*-tests comparing the mean amplitude of FRN, grouped for early (i.e. 1-2-5-6) and late (i.e. 3-4-7-8) blocks separately, against zero indicated that reward-related feedback elicited an FRN at all tested electrode sites both in LA and HA (Table 1).

Next, we assessed whether the mean amplitude of the FRN differed between LA and HA. The 2 × 2 mixed-design ANOVA (block, early and late; group, HA and LA) indicated no significant difference between groups in the amplitude of FRN ($P = 0.600$), no significant difference between blocks ($P = 0.613$) and no interaction between the two ($P = 0.857$). Therefore, the amplitude of the FRN elicited by the presentation of reward-related feedback did not differ significantly between LA and HA.

Next, in order to test the differences between groups in the computation of prediction error, we compared the mean amplitude of the FRN in response to expected and unexpected reward-related feedback for LA and HA. The 2 × 2 × 2 mixed-design ANOVA (block, early and late; expectancy, expected and unexpected; group, LA and HA) revealed a main effect of expectancy ($F(1,38) = 4.19$, $P = 0.047$, $\eta_p^2 = 0.10$), which was

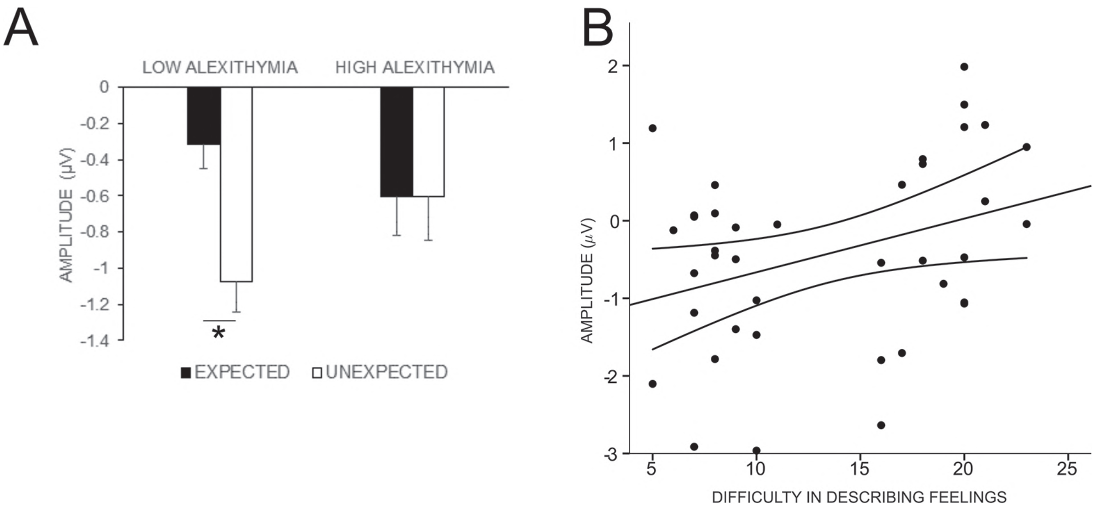


Fig. 5. (A) FRN mean amplitude for expected and unexpected feedback as a function of alexithymia group. The low, but not high, alexithymia group showed enhancement of the FRN in response to unexpected (*vs* expected) reward-related feedback. Error bars represent standard errors. $^*P < 0.05$. (B) Scatterplot showing that difficulty in describing feelings predicted prediction error magnitude (i.e. amplitude difference between FRN unexpected and FRN expected). Dashed lines represent 95% confidence interval. $P < 0.05$.

qualified by an expectancy by group interaction ($F(1,38) = 4.13$, $P = 0.049$, $\eta_p{}^2 = 0.10$). Crucially, LA showed a more negative FRN for unexpected than expected reward-related feedback ($M_{\text{unexp}} = -1.07$, $M_{\text{exp}} = -0.31$, $P = 0.031$). On the contrary, HA showed no significant difference in the amplitude of the FRN between unexpected and expected reward-related feedback ($M_{\text{unexp}} = -0.60$, $M_{\text{exp}} = -0.60$, $P = 0.992$; Figure 5A). No other main effects or interactions were significant (all $P \geq 0.173$). Therefore, in LA, but not HA, unexpected reward-related feedback elicited an RPE.

Finally, we ran a stepwise multiple regression in order to understand which of the three components of alexithymia influenced the current results on RPE. Participants' scores on the TAS-20 subscales [i.e. difficulty in identifying feelings (DIF), difficulty in describing feelings (DDF) and externally oriented thinking (EOT)] were the independent variables, and the amplitude difference between FRN for unexpected and expected feedback, averaged between early and late blocks, was the dependent variable. Only the score on the difficulty in describing feelings subscale made a significant contribution to the regression ($R^2 = 0.12$, $F(1,37) = 4.88$, $P = 0.033$; DDF: $\beta = 0.34$, $t(37) = 2.21$, $P = 0.033$; DIF: $P = 0.549$; EOT: $P = 0.883$, Figure 5B). Thus, the more participants had difficulties in describing their feelings, the smaller their prediction error.

P300

To verify that alexithymia specifically affected RPE, differences between groups in the P300 in response to the feedback were tested. The 2 × 2 × 2 × 2 mixed-design ANOVA (block, early and late; feedback expectancy, expected and unexpected; feedback valence, reward and no-reward; and group, LA and HA) revealed the main effect of valence ($F(1,38) = 23.22$, $P < 0.001$, $\eta_p{}^2 = 0.38$) indicating that the P300 was more positive for reward than no-reward feedback ($M_{\text{reward}} = 4.13$, $M_{\text{no-reward}} = 3.54$). In addition, there was a marginally significant valence by expectancy interaction ($F(1,38) = 3.34$, $P = 0.075$, $\eta_p{}^2 = .08$). No other main effects or interactions were significant (all $P \geq 0.130$). These results suggest that participants devoted more attention to the reward than no-reward feedback, but that groups did not differ significantly in the amount of attentional resources devoted to feedback; thus, alexithymia did not affect the computation of the P300 significantly.

Accuracy and response time to the CS

The 2 × 2 × 2 mixed-design ANOVA (block, early and late; stimulus, CS1 and CS2; and group, LA and HA) on response times showed a significant main effect of block ($F(1,38) = 5.11$, $P = 0.029$, $\eta_p{}^2 = .12$), indicating that participants were faster in the first than the second group of blocks ($M_{\text{early}} = 436.94$ ms, s.d.$_{\text{early}} = 83.89$ ms, $M_{\text{late}} = 423.19$ ms, s.d.$_{\text{late}} = 72.64$ ms). No other effects were significant (all $P \geq 0.110$; LA: CS1: $M_{\text{early}} = 417.92$ ms, s.d.$_{\text{early}} = 57.25$ ms, $M_{\text{late}} = 405.61$ ms, s.d.$_{\text{late}} = 47.29$ ms, CS2: $M_{\text{early}} = 413.64$ ms, s.d.$_{\text{early}} = 58.39$ ms, $M_{\text{late}} = 405.69$ ms, s.d.$_{\text{late}} = 48.12$ ms; HA: CS1: $M_{\text{early}} = 458.00$ ms, s.d.$_{\text{early}} = 100.11$ ms, $M_{\text{late}} = 439.78$ ms, s.d.$_{\text{late}}$85.67 ms, CS2: $M_{\text{early}} = 458.20$ ms, s.d.$_{\text{early}} = 103.24$ ms, $M_{\text{late}} = 441.66$ ms, s.d.$_{\text{late}} = 93.56$ ms). The 2 × 2 × 2 mixed-design ANOVA (block, early and late; stimulus, CS1 and CS2; and group, LA and HA) on percentage accuracy showed no significant main effect or interaction (all $P \geq 0.236$; LA: CS1: $M_{\text{early}} = 93.45\%$, s.d.$_{\text{early}} = 8.71\%$, $M_{3\text{–}4\text{–}7\text{–}8} = 92.40\%$, s.d.$_{\text{late}} = 11.57\%$, CS2: $M_{\text{early}} = 92.82\%$, s.d.$_{\text{early}} = 11.89\%$, $M_{\text{late}} = 92.82\%$, s.d.$_{\text{late}} = 12.22\%$; HA: CS1: $M_{\text{early}} = 93.85\%$, s.d.$_{\text{early}} = 6.05\%$, $M_{\text{late}} = 94.50\%$, s.d.$_{\text{late}} = 1.97\%$, CS2: $M_{\text{early}} = 94.20\%$, s.d.$_{\text{early}} = 5.35\%$, $M_{\text{late}} = 95.12\%$, s.d.$_{\text{late}} = 1.93\%$).

CS–feedback contingency awareness

To assess within-group differences on CS–feedback contingency awareness (Figure 6A) for each group, we conducted the Wilcoxon Matched Paired Test contrasting contingency scores for CS1 and CS2. These tests showed no significant within-group difference in CS–feedback contingency awareness either for LA ($P = 1.000$) or HA ($P = 0.592$). This indicates that, in either group, there was no significant difference between CS in explicit understanding of CS–feedback association.

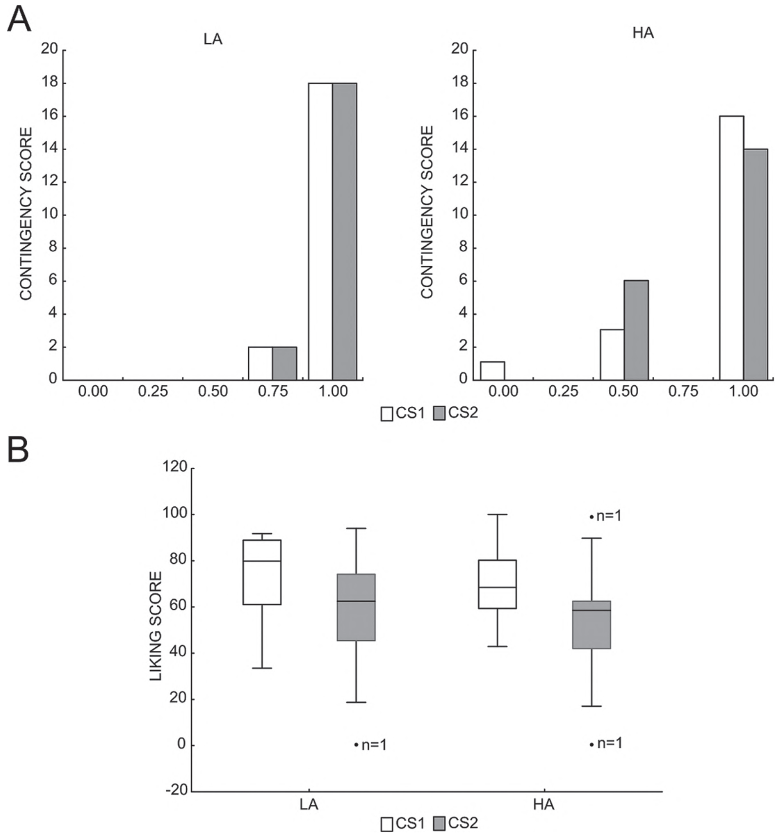


Fig. 6. (A) Histogram plots showing the frequency (in counts) of the CS–feedback contingency awareness scores of the low (LA) and high alexithymia groups (HA) for CS1 and CS2. (B) Box and whisker plots of the CS liking scores of LA and HA for CS1 and CS2. The horizontal line within the box indicates the median, boundaries of the box indicate the 25th and 75th percentile and the whiskers indicate the non-outliers minimum and maximum. The dot indicates outliers.

In addition, to assess between-group differences on CS–feedback contingency awareness, for each CS, we conducted the Mann–Whitney U-test contrasting median contingency scores of LA and HA participants. There were no differences between groups neither in the test conducted on scores for CS1 ($P = 0.525$) nor on scores for the CS2 ($P = 0.218$). This indicates that groups did not differ significantly in their understanding of CS–feedback associations.

Subjective report of liking

To assess within-group differences on subjective reports of CS liking (Figure 6B), for each group, we conducted the Wilcoxon Matched Paired Test contrasting liking scores for CS1 and CS2. For LA, results showed a significant difference in liking ($T = 41.00$, $Z = 2.17$, $P = 0.030$), such that LA liked CS1 (80% rewarding; Mdn = 79.83) more than CS2 (20% rewarding; Mdn = 62.50). For HA, results showed a marginally significant difference in liking between the two CS ($T = 56.00$, $Z = 1.83$, $P = 0.067$), indicating a tendency to like CS1 (80% rewarding; Mdn = 68.46) more than CS2 (20% rewarding; Mdn = 58.52).

In addition, to assess between-group differences on CS liking, for each CS, we conducted the Mann–Whitney U-test contrasting median liking scores of LA and HA participants. Results showed no difference between groups neither in the test conducted on scores for CS1 ($P = 0.133$) nor on scores for CS2 ($P = 0.384$). This indicates that groups did not differ significantly in their subjective reports of CS liking.

Discussion

Here, we tested the hypothesis that alexithymia is related to altered learning of the value of CS during Pavlovian appetitive conditioning. To this end, electroencephalography was recorded to assess differences in the computation of RPE, i.e. the amplitude of the FRN, in response to expected and unexpected reward-related feedback in individuals with high (HA) and low (LA) levels of alexithymia. We found that both in LA and HA, the FRN was elicited in response to the valence of reward-related feedback (i.e. omitted *vs* delivered reward), and its amplitude did not differ significantly between groups. Crucially, we found that unexpected (*vs* expected) reward-related feedback enhanced FRN in LA but not in HA, confirming impaired computation of RPE in HA. Additionally, the size of RPE decreased with increasing difficulty in describing feelings as measured on the TAS-20 (Taylor *et al.*, 2003). This result was not associated with a global reduction in brain activity or in feedback processing, but to a specific impairment in the computation of RPE, because additional analyses on the amplitude of the P300 component in response to feedback showed no difference between groups. Finally, on a behavioral level, groups showed comparable understanding of explicit CS–feedback contingencies and liked the CS mostly predicting reward more than that mostly predicting no-reward, although for HA this effect was only marginally significant. This possibly suggests that alexithymia may affect implicit but not explicit aspects of emotional learning, in keeping with Starita *et al.* (2016). Nevertheless, the marginally significant effect calls for further investigation on this issue.

The results from LA are in line with the previous literature that investigated the FRN and P300 in response to reward-related feedback. They support the theory that FRN encodes an RPE-like signal (Holroyd and Coles, 2002; Nieuwenhuis *et al.*, 2004; Walsh and Anderson, 2012; Sambrook and Goslin, 2015). An FRN was elicited in response to feedback valence (i.e. omitted *vs* delivered reward) and was more negative when feedback was unexpected than expected, due to more negative ERP following unexpected (*vs* expected) no-reward, and more positive ERP following unexpected (*vs* expected) reward (Figure 4, as in e.g. Donkers and van Boxtel, 2005; Potts *et al.*, 2006; Hajcak *et al.*, 2007; Holroyd and Krigolson, 2007; Holroyd *et al.*, 2011; Walsh and Anderson, 2011, 2013). Furthermore, the P300 in response to feedback was more positive for feedback indicating reward than no-reward (as in e.g. Hajcak *et al.*, 2005; Hajcak *et al.*, 2007; Bellebaum and Daum, 2008). Therefore, LA were not only sensitive to feedback valence, showing differential electrophysiological response to the delivery of reward as opposed to no-reward, but they also used such response to shape the value representation of the CS. This was evidenced by the computation of the RPE in response to feedback that violated acquired expectations about the CS–feedback contingency.

Our main result is that, contrary to LA, HA did not exhibit an RPE in response to unexpected reward-related feedback, as evidenced by lack of modulation of FRN by feedback expectancy. This impairment in the computation of RPE in HA was not related to a general impairment in the computation of the FRN. There were no differences between HA and LA in the amplitude of the FRN elicited by reward-related feedback valence. Thus, electrophysiological data suggest that the sensitivity to rewards of HA appears comparable to LA. Nevertheless, HA are unable to use this response to update the value of stimuli in the environment that predict the occurrence of such rewards. This is in line with our previous study on Pavlovian threat conditioning, in which HA showed preserved psychophysiological response to an aversive US but had reduced response to a CS that predicted it (Starita *et al.*, 2016), and on instrumental learning, in which HA showed decreased learning of the value of aversively motivated actions (Starita and di Pellegrino, 2018). Taken together, the results of these studies suggest that alexithymia is characterized by a general alteration of associative emotional learning, although the precise mechanisms (e.g. explicit and/or implicit) contributing to this alteration remain to be clearly identified. Individuals with alexithymia appear able to respond to stimuli that are biologically prepared to trigger an emotional response. Nevertheless, they appear unable to use such response to construct an internal representation of emotional stimuli that include, alongside US, those that predict them. As a consequence, HA may be at the mercy of emotional stimuli. Internal representations are predictive models that enable individuals to anticipate the emotional future, so that organisms can prepare to respond to coming emotional stimuli, rather than simply react to them once they have occurred (Öhman and Mineka, 2001; McNally and Westbrook, 2006). These predictive representations are not only crucial for effective recognition, response and response regulation to the emotional stimuli *per se* but also for anticipating the consequences of these stimuli enabling optimal decision-making (Bubic *et al.*, 2010). HA are impaired in all these aspects. They have impairments in the identification of the emotional stimuli (Grynberg *et al.*, 2012; Ihme *et al.*, 2014a,b; Starita *et al.*, 2018), the physiological (Franz *et al.*, 2003; Neumann *et al.*, 2004; Pollatos *et al.*, 2008; Bermond *et al.*, 2010) and behavioral response to these stimuli (Sonnby-Borgström, 2009; Scarpazza *et al.*, 2014, 2015, 2018), the regulation of such response (Swart *et al.*, 2009; Pollatos and Gramann, 2012) and its use to guide decision-making (Ferguson *et al.*, 2009; Patil and Silani, 2014a,b; Scarpazza *et al.*, 2017). Therefore, the impaired construction of internal representations of emotional stimuli in alexithymia may represent a mechanism underlying the difficulties in processing emotional stimuli.

There is evidence that the amplitude of FRN may be related to activity in the dopaminergic system, which is crucially involved in emotional learning and in encoding RPEs (Schultz, 1998, 2016). This component originates from activity in the anterior cingulate cortex (ACC; Holroyd *et al.*, 2004; Warren *et al.*, 2015), which may be related to phasic changes in dopamine activity resulting from dopaminergic projections from midbrain structures to this area (Nieuwenhuis *et al.*, 2004; Walsh and Anderson, 2012; Sambrook and Goslin, 2015). Therefore, the differences observed in HA in the amplitude of FRN may be related to differences in the dopaminergic system. In addition, differences in activity in the ACC have been shown in alexithymia (van der Velde *et al.*, 2014), possibly suggesting an overlapping neural mechanism for the difficulty in emotional learning and the broader difficulties in processing emotional stimuli.

The lack of modulation of FRN by feedback expectancy but the preserved enhancement of P300 suggest that alexithymia is associated with a specific impairment in the computation of RPE, rather than a global impairment in feedback processing. Also, it highlights the subclinical nature of alexithymia, which may affect only specific aspects of emotional processing. For example, in the context of error processing, electrophysiological evidence showed that alexithymia affects rapid and automatic error monitoring in an emotional (*vs* neutral) task context at the time of erroneous responses, as measured by the error-related negativity, but not later-emerging error awareness, as measured by the error positivity (Maier *et al.*, 2016). Additionally, when examining visual processing of emotional body postures, alexithymia was found to impair early but not later visual processing, as evidenced by lack of modulation of the N190, but preserved modulation of early posterior negativity in response to fearful body postures (Borhani *et al.*, 2016).

The present study has three limitations that we wish to discuss. First, the results show no significant effect of early *vs* late blocks on our dependent variables, suggesting that learning did not vary with increasing exposure to CS–feedback contingencies. Possibly, the effects we see may happen after only a few trials. This may be imputed mainly to the structure of the experiment, which included only two CS, each followed by the expected feedback in 80% of trials, hence making the learning of CS–feedback contingencies straightforward. Additionally, the ERP analysis that requires averaging serval trials together to have reliable data has the disadvantage of limiting the investigation of effects occurring early in the experiment. Future work could use a learning task structured such that it can enable to track the emergence of learning over time and how this is affected by alexithymia. Second, the US used in the experiment is a secondary reinforcer, rather than a primary one, such as food. Although evidence shows that secondary reinforcers activate neural structures overlapping with those activated by primary reinforcers, lead to behavioral learning of contingencies and shape human behavior (Breite and Rosen, 1999; Delgado *et al.*, 2000, 2003, 2005, 2006; Elliott *et al.*, 2000; Delgado, 2007; Knutson *et al.*, 2001a,b), future work should investigate whether or not results change when using a primary reinforcer as US. Third, although we cannot exclude the possibility that response to the CS1/CS2 may influence the feedback ERP, we designed our experimental task following the past FRN literature, where it is common practice to present a visual stimulus, have participants make a response and then provide visual feedback regarding the outcome of the trial. In fact, the time interval between participants' response and feedback can vary significantly between experiments, with intervals ranging from 400 ms to 1 s (Gehring and Willoughby, 2002; Yeung et al., 2005; Hajcak et al., 2006; Holroyd and Krigolson, 2007; Holroyd et al., 2008; Walsh and Anderson, 2011b; Sambrook and Goslin, 2016; Di Gregorio et al., 2019). In our experiment, given average participants' response time of 430.06 ms (s.d. = 78.52 ms), the time between response and feedback was ~1 s. Interestingly, the FRN literature has also shown that an FRN is reliably observed not only when feedback is contingent upon participants' response but also when it is not or when no response at all is made (Yeung et al., 2005). Thus, future studies could replicate the experiment without having participants make any response.

In conclusion, despite preserved sensitivity to rewards, alexithymia is related to impaired computation of RPEs during Pavlovian appetitive conditioning. Thus, although individuals with alexithymia are able to respond to stimuli that are biologically prepared to elicit an emotional response, they are unable to use this response to update the value of stimuli in the environment that predict them. This alters the construction of an internal representation of emotional stimuli that includes, alongside UCS, those that are associated with them. However, these internal representations are crucial for anticipating the emotional future, enabling effective recognition, response and response regulation to emotional stimuli. Therefore, altered construction of internal representations in individuals with alexithymia may leave them at the mercy of emotional stimuli, possibly representing a mechanism underlying their difficulties in emotion processing.

Funding

This work was supported by RFO Grant from the University of Bologna awarded to Giuseppe di Pellegrino.

Conflict of interest

None declared.

Acknowledgements

The authors would like to thank Maria Chiara Caruso, Angela Galanti, Agata Monni and Antonella Nero for their assistance with data collection; Francesco di Gregorio, Sara Garofalo, Martin Maier and Marco Zanon for their support in data analysis; and Elisabetta Làdavas for her comments on an earlier version of the manuscript. The authors would also like to acknowledge the two anonymous reviewers and the Editor Prof. Annett Schirmer for thoroughly reviewing the paper.

References

Alexander, W.H., Brown, J.W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nature Neuroscience*, **14**(10), 1338–44 https://doi.org/10.1038/nn.2921.

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, **4**, 561–71.

Bell, A.J., Sejnowski, T.J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, **7**(6), 1129–59 http://www.ncbi.nlm.nih.gov/pubmed/7584893.

Bellebaum, C., Daum, I. (2008). Learning-related changes in reward expectancy are reflected in the feedback-related negativity. *European Journal of Neuroscience*, **27**(7), 1823–35 https://doi.org/10.1111/j.1460-9568.2008.06138.x.

Bermond, B., Bierman, D.J., Cladder, M.A., Moormann, P.P., Vorst, H.C.M. (2010). The cognitive and affective alexithymia dimensions in the regulation of sympathetic responses. *International Journal of Psychophysiology*, **75**(3), 227–33 https://doi.org/10.1016/j.ijpsycho.2009.11.004.

Borhani, K., Borgomaneri, S., Làdavas, E., Bertini, C. (2016). The effect of alexithymia on early visual processing of emotional body postures. *Biological Psychology*, **115**, 1–8 https://doi.org/10.1016/j.biopsycho.2015.12.010.

Breiter, H.C., Rosen, B.R. (1999). Functional Magnetic Resonance Imaging of Brain Reward Circuitry in the Human. *Annals of the New York Academy of Sciences*, **877**(1), 523–547. https://doi.org/10.1111/j.1749-6632.1999.tb09287.x.

Bubic, A., von Cramon, D. Y., and Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience, 4(March), 1–15.* https://doi.org/10.3389/fnhum.2010.00025

Daw, N.D., Tobler, P.N. (2014). Chapter 15—value learning through reinforcement: the basics of dopamine and reinforcement learning. In: Glimcher, P.W., Fehr, E., editors. *Neuroeconomics*, 2nd edn, San Diego: Academic Press, pp. 283–98.

Delgado, M.R. (2007). Reward-related responses in the human striatum. *Annals of the New York Academy of Sciences*, **1104**, 70–88 https://doi.org/10.1196/annals.1390.002.

Delgado, M.R., Nystrom, L.E., Fissell, C., Noll, D.C., Fiez, J.A. (2000). Tracking the Hemodynamic Responses to Reward and Punishment in the Striatum. *Journal of Neurophysiology*, **84**(6), 3072–3077. https://doi.org/10.1152/jn.2000.84.6.3072.

Delgado, M.R., Locke, H.M., Stenger, V.A., Fiez, J.A. (2003). Dorsal striatum responses to reward and punishment: Effects of valence and magnitude manipulations. *Cognitive, Affective, & Behavioral Neuroscience*, **3**(1), 27–38. https://doi.org/10.3758/CABN.3.1.27.

Delgado, M.R., Miller, M.M., Inati, S., Phelps, E.A. (2005). An fMRI study of reward-related probability learning. *NeuroImage*. **24**(3), 862–873. https://doi.org/10.1016/j.neuroimage.2004.10.002.

Delgado, M.R., Labouliere, C.D., Phelps, E.A. (2006). Fear of losing money? Aversive conditioning with secondary reinforcers. *Social Cognitive and Affective Neuroscience*, **1**(3), 250–9 https://doi.org/10.1093/scan/nsl025.

Delorme, A., Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, **134**(1), 9–21 https://doi.org/10.1016/J.JNEUMETH.2003.10.009.

Di, F., Ernst, B., Steinhauser, M. (2019). Differential effects of instructed and objective feedback reliability on feedback-related brain activity. *Psychophysiology*, e13399 https://doi.org/10.1111/psyp.13399.

Donkers, F.C.L., van Boxtel, G.J.M. (2005). Mediofrontal negativities to averted gains and losses in the slot-machine task. *Journal of Psychophysiology*, **19**(4), 256–62 https://doi.org/10.1027/0269-8803.19.4.256.

Elliott, R., Friston, K.J., Dolan, R.J. (2000). Dissociable neural responses in human reward systems. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, **20**(16), 6159–65 https://doi.org/10.1523/jneurosci.4537-03.2004.

Ferguson, E., Bibby, P.A., Rosamond, S., O'Grady, C., Parcell, A., Amos, C., O'Carroll, R. (2009). Alexithymia, cumulative feedback, and differential response patterns on the Iowa gambling task. *Journal of Personality*, **77**(3), 883–902 https://doi.org/10.1111/j.1467-6494.2009.00568.x.

Franz, M., Schaefer, R., Schneider, C. (2003). Psychophysiological response patterns of high and low alexithymics under mental and emotional load conditions. *Journal of Psychophysiology*, **17**(4), 203–13 https://doi.org/10.1027/0269-8803.17.4.203.

Franz, M., Schaefer, R., Schneider, C., Sitte, W., Bachor, J. (2004). Visual event-related potentials in subjects with alexithymia: modified processing of emotional aversive information? *American Journal of Psychiatry*, **161**(4), 728–35 https://doi.org/10.1176/appi.ajp.161.4.728.

Garofalo, S., Timmermann, C., Battaglia, S., Maier, M.E., di Pellegrino, G. (2016). Mediofrontal negativity signals unexpected timing of salient outcomes. 1–10 https://doi.org/10.1162/jocn_a_01074.

Gehring, W.J., Willoughby, A.R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science*, **295**(5563), 2279–82 https://doi.org/10.1126/science.1066893.

Gehring, W.J., Goss, B., Coles, M.G.H., Meyer, D.E., Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science*, **4**(6), 385–90 https://doi.org/10.1111/j.1467-9280.1993.tb00586.x.

Grynberg, D., Chang, B., Corneille, O., Maurage, P., Vermeulen, N., Berthoz, S., Luminet, O. (2012). Alexithymia and the processing of emotional facial expressions (EFEs): systematic review, unanswered questions and further perspectives. *PLoS One*, **7**(8), e42429 https://doi.org/10.1371/journal.pone.0042429.

Hajcak, G., Holroyd, C.B., Moser, J.S., Simons, R.F. (2005). Brain potentials associated with expected and unexpected good and bad outcomes. *Psychophysiology*, **42**(2), 161–70 https://doi.org/10.1111/j.1469-8986.2005.00278.x.

Hajcak, G., Moser, J.S., Holroyd, C.B., Simons, R.F. (2006). The feedback-related negativity reflects the binary evaluation of good versus bad outcomes. *Biological Psychology*, **71**(2), 148–54 https://doi.org/10.1016/j.biopsycho.2005.04.001.

Hajcak, G., Moser, J.S., Holroyd, C.B., Simons, R.F. (2007). It's worse than you thought: the feedback negativity and violations of reward prediction in gambling tasks. *Psychophysiology*, **44**(6), 905–12 https://doi.org/10.1111/j.1469-8986.2007.00567.x.

Holroyd, C.B., Coles, M.G.H. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, **109**(4), 679–709 https://doi.org/10.1037/0033-295X.109.4.679.

Holroyd, C.B., Krigolson, O.E. (2007). Reward prediction error signals associated with a modified time estimation task. *Psychophysiology*, **44**(6), 913–7 https://doi.org/10.1111/j.1469-8986.2007.00561.x.

Holroyd, C.B., Nieuwenhuis, S., Yeung, N., Nystrom, L., Mars, R.B., Coles, M.G.H., Cohen, J.D. (2004). Dorsal anterior cingulate cortex shows fMRI response to internal and external error signals. *Nature Neuroscience*, **7**(5), nn1238), https://doi.org/10.1038/nn1238.

Holroyd, C.B., Pakzad-Vaezi, K.L., Krigolson, O.E. (2008). The feedback correct-related positivity: sensitivity of the event-related brain potential to unexpected positive feedback. *Psychophysiology*, **45**(5), 688–97 https://doi.org/10.1111/j.1469-8986.2008.00668.x.

Holroyd, C.B., Krigolson, O.E., Lee, S. (2011). Reward positivity elicited by predictive cues. *Neuroreport*, **22**(5), 249–52 https://doi.org/10.1097/WNR.0b013e328345441d.

Ihme, K., Sacher, J., Lichev, V., Rosenberg, N., Kugel, H., Rufer, M., Suslow, T. (2014a). Alexithymia and the labeling of facial emotions: response slowing and increased motor and somatosensory processing. *BMC Neuroscience*, **15**, 40 https://doi.org/10.1186/1471-2202-15-40.

Ihme, K., Sacher, J., Lichev, V., Rosenberg, N., Kugel, H., Rufer, M., Suslow, T. (2014b). Alexithymic features and the labeling of brief emotional facial expressions—an fMRI study. *Neuropsychologia*, **64**, 289–99 https://doi.org/10.1016/j.neuropsychologia.2014.09.044.

Jutten, C., Herault, J. (1991). Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture. *Signal Processing*, **24**(1), 1–10 https://doi.org/10.1016/0165-1684(91)90079-X.

Knutson, B., Adams, C.M., Fong, G.W., Hommer, D. (2001a). Anticipation of Increasing Monetary Reward Selectively Recruits Nucleus Accumbens. *The Journal of Neuroscience*. 21(16), RC159–RC159. https://doi.org/10.1523/JNEUROSCI.21-16-j0002.2001

Knutson, B., Fong, G.W., Adams, C.M., Varner, J.L., Hommer, D. (2001b). Dissociation of reward anticipation and outcome with event-related fMRI. *Neuroreport*, 12(17), 3683–3687 https://doi.org/10.1097/00001756-200112040-00016

Li, S., Zhang, B., Guo, Y., Zhang, J. (2015). The association between alexithymia as assessed by the 20-item Toronto alexithymia scale and depression: a meta-analysis. *Psychiatry Research*, **227**(1), 1–9 https://doi.org/10.1016/j.psychres.2015.02.006.

Luck, S.J. (2014). *An Introduction to the Event-Related Potential Technique, second edition*, The MIT Press (2nd ed.), Cambridge MA. https://doi.org/10.1118/1.4736938

Luminet, O., Bagby, R.M., Taylor, G.J. (2018). *Alexithymia: Advances in Research, Theory, and Clinical Practice*, Cambridge UK: Cambridge University Press, Retrieved from: https://play.google.com/books?id=AvxsDwAAQBAJ.

Maier, M.E., Scarpazza, C., Starita, F., Filogamo, R., Làdavas, E. (2016). Error monitoring is related to processing internal affective states. *Cognitive, Affective, & Behavioral Neuroscience*, **16**(6), 1050–62 https://doi.org/10.3758/s13415-016-0452-1.

Makeig, S., Bell, A.J., Jung, T.-P., Sejnowski, T.J. (1996). Independent component analysis of electroencephalographic data. In: Touretzky, D., Mozer, M., Hasselmo, M., editors, Cambridge, MA: MIT Press, pp. 145–51 http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.36.4803&rep=rep1&type=pdf.

Mangelli, L., Semprini, F., Sirri, L., Fava, G.A., Sonino, N. (2006). Use of the diagnostic criteria for psychosomatic research (DCPR) in a community sample. *Psychosomatics*, **47**(2), 143–6 https://doi.org/10.1176/appi.psy.47.2.143.

McNally, G.P., Westbrook, R.F. (2006). Predicting danger: the nature, consequences, and neural mechanisms of predictive fear learning. *Learning & Memory*, **13**(3), 245–53 https://doi.org/10.1101/lm.196606.

Neumann, S.A., Sollers, J.J., Thayer, J.F., Waldstein, S.R. (2004). Alexithymia predicts attenuated autonomic reactivity, but prolonged recovery to anger recall in young women. *International Journal of Psychophysiology*, **53**(3), 183–95 https://doi.org/10.1016/j.ijpsycho.2004.03.008.

Nieuwenhuis, S., Holroyd, C.B., Mol, N., Coles, M.G. (2004). Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. *Neuroscience & Biobehavioral Reviews*, **28**(4), 441–8 https://doi.org/10.1016/J.NEUBIOREV.2004.05.003.

Nieuwenhuis, S., Aston-Jones, G., Cohen, J.D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin*, **131**(4), 510–32 https://doi.org/10.1037/0033-2909.131.4.510.

Öhman, A., Mineka, S. (2001). Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychological Review*, **108**(3), 483–522 https://doi.org/10.1037/0033-295X.108.3.483.

Olofsson, J.K., Nordin, S., Sequeira, H., Polich, J. (2008). Affective picture processing: an integrative review of ERP findings. *Biological Psychology*, **77**(3), 247–65 https://doi.org/10.1016/j.biopsycho.2007.11.006.

Patil, I., Silani, G. (2014a). Alexithymia increases moral acceptability of accidental harms. *Journal of Cognitive Psychology*, **26**(5), 597–614 https://doi.org/10.1080/20445911.2014.929137.

Patil, I., Silani, G. (2014b). Reduced empathic concern leads to utilitarian moral judgments in trait alexithymia. *Frontiers in Psychology*, **5**(501), 1–12 https://doi.org/10.3389/fpsyg.2014.00501.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, **118**(10), 2128–48 http://linkinghub.elsevier.com/retrieve/pii/S1388245707001897.

Pollatos, O., Gramann, K. (2012). Attenuated modulation of brain activity accompanies emotion regulation deficits in alexithymia. *Psychophysiology*, **49**(5), 651–8 https://doi.org/10.1111/j.1469-8986.2011.01348.x.

Pollatos, O., Schubö, A., Herbert, B.M., Matthias, E., Schandry, R. (2008). Deficits in early emotional reactivity in alexithymia. *Psychophysiology*, **45**(5), 839–46 https://doi.org/10.1111/j.1469-8986.2008.00674.x.

Potts, G.F., Martin, L.E., Burton, P., Montague, P.R. (2006). When things are better or worse than expected: the medial frontal cortex and the allocation of processing resources. *Journal of Cognitive Neuroscience*, **18**(7), 1112–9 https://doi.org/10.1162/jocn.2006.18.7.1112.

Sambrook, T.D., Goslin, J. (2015). A neural reward prediction error revealed by a meta-analysis of ERPs using great grand averages. *Psychological Bulletin*, **141**(1), 213–35 https://doi.org/10.1037/bul0000006.

Sambrook, T.D., Goslin, J. (2016). Principal components analysis of reward prediction errors in a reinforcement learning task. *NeuroImage*, **124**, 276–86 https://doi.org/10.1016/j.neuroimage.2015.07.032.

Scarpazza, C., di Pellegrino, G., Làdavas, E. (2014). Emotional modulation of touch in alexithymia. *Emotion*, **14**(3), 602–10 https://doi.org/10.1037/a0035888.

Scarpazza, C., Làdavas, E., di Pellegrino, G. (2015). Dissociation between emotional remapping of fear and disgust in alexithymia. *PLoS One*, **10**(10), e0140229 https://doi.org/10.1371/journal.pone.0140229.

Scarpazza, C., Sellitto, M., di Pellegrino, G. (2017). Now or not-now? The influence of alexithymia on intertemporal decision-making. *Brain and Cognition*, **114**, 20–8 https://doi.org/10.1016/j.bandc.2017.03.001.

Scarpazza, C., Làdavas, E., Cattaneo, L. (2018). Invisible side of emotions: somato-motor responses to affective facial displays in alexithymia. *Experimental Brain Research*, **236**(1), 195–206 https://doi.org/10.1007/s00221-017-5118-x.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, **80**(1), 1–27.

Schultz, W. (2016). Dopamine reward prediction error coding. *Dialogues in Clinical Neuroscience*, **18**(1), 23–32 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4826767/.

Sifneos, P.E. (1973). The prevalence of “alexithymic” characteristics in psychosomatic patients. *Psychotherapy and Psychosomatics*, **22**(2), 255–62.

Sonnby-Borgström, M. (2009). Alexithymia as related to facial imitation, mentalization, empathy, and internal working models-of-self and -others. *Neuropsychoanalysis*, **11**(1), 111–28 https://doi.org/10.1080/15294145.2009.10773602.

Starita, F., di Pellegrino, G. (2018). Alexithymia and the Reduced Ability to Represent the Value of Aversively Motivated Actions. *Frontiers in Psychology*, **9**(2587), 1–11. https://doi.org/10.3389/fpsyg.2018.02587.

Starita, F., Ladavas, E., di Pellegrino, G. (2016). Reduced anticipation of negative emotional events in alexithymia. *Scientific Reports*, **6**, 27664 https://doi.org/10.1038/srep27664.

Starita, F., Borhani, K., Bertini, C., Scarpazza, C. (2018). Alexithymia is related to the need for more emotional intensity to identify static fearful facial expressions. *Frontiers in Psychology*, **9**, https://doi.org/10.3389/fpsyg.2018.00929.

Swart, M., Kortekaas, R., Aleman, A. (2009). Dealing with feelings: characterization of trait alexithymia on emotion regulation strategies and cognitive-emotional processing. *PLoS One*, **4**(6), e5751 https://doi.org/10.1371/journal.pone.0005751.

Taylor, G.J., Michael Bagby, R., Parker, J.D.A. (1991). The alexithymia construct: a potential paradigm for psychosomatic medicine. *Psychosomatics*, **32**(2), 153–64 https://doi.org/10.1016/S0033-3182(91)72086-0.

Taylor, G.J., Bagby, R.M., Parker, J.D. (2003). The 20-item Toronto alexithymia scale. *Journal of Psychosomatic Research*, **55**(3), 277–83 https://doi.org/10.1016/S0022-3999(02)00601-3.

Teixeira, R.J., Bermond, B., Moormann, P.P. (2018). *Current Developments in Alexithymia—A Cognitive and Affective Deficit*, Hauppauge, New York: Nova Science Publishers Incorporated, Retrieved from: https://www.researchgate.net/publication/325952621_Current_developments_in_alexithymia_A_cognitive_and_affective_deficit.

van der Velde, J., Gromann, P., Swart, M., Wiersma, D., Haan, L., de Bruggeman, R., Aleman, A. (2014). Alexithymia influences brain activation during emotion perception but not regulation. *Social Cognitive and Affective Neuroscience*, nsu056 https://doi.org/10.1093/scan/nsu056.

Walsh, M.M., Anderson, J.R. (2011a). Learning from delayed feedback: neural responses in temporal credit assignment. *Cognitive, Affective, & Behavioral Neuroscience*, **11**(2), 131–43 https://doi.org/10.3758/s13415-011-0027-0.

Walsh, M.M., Anderson, J.R. (2011b). Modulation of the feedback-related negativity by instruction and experience. *Proceedings of the National Academy of Sciences of the United States of America*, **108**(47), 19048–53 https://doi.org/10.1073/pnas.1117189108.

Walsh, M.M., Anderson, J.R. (2012). Learning from experience: event-related potential correlates of reward processing, neural adaptation, and behavioral choice. *Neuroscience & Biobehavioral Reviews*, **36**(8), 1870–84 https://doi.org/https://doi.org/10.1016/j.neubiorev.2012.05.008.

Walsh, M.M., Anderson, J.R. (2013). Electrophysiological responses to feedback during the application of abstract rules. *Journal of Cognitive Neuroscience*, **25**(11), 1986–2002 https://doi.org/10.1162/jocn_a_00454.

Warren, C.M., Hyman, J.M., Seamans, J.K., Holroyd, C.B. (2015). Feedback-related negativity observed in rodent anterior cingulate cortex. *Journal of Physiology, Paris*, **109**(1), 87–94 https://doi.org/10.1016/j.jphysparis.2014.08.008.

Yeung, N., Holroyd, C.B., Cohen, J.D. (2005). ERP correlates of feedback and reward processing in the presence and absence of response choice. *Cerebral Cortex*, **15**(5), 535–44 https://doi.org/10.1093/cercor/bhh153.