Affective neural signatures do not distinguish women with emotion dysregulation from healthy controls:
A mega-analysis across three task-based fMRI studies

Sicorello, M.[1], Herzog, J.[1], Wager, T.D.[2], Ende G.[3], Müller-Engelmann, M.[4], Herpertz, S.C.[5] Bohus M.[6], Schmahl, C.[1], Paret, C.[1,7*], Niedtfeld, I.[1*]

* Equally contributing authors

Affiliations

1) Department of Psychosomatic Medicine and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim/Heidelberg University, Mannheim, Germany
2) Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA
3) Department Neuroimaging, Central Institute of Mental Health Mannheim, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany
4) Department of Clinical Psychology and Intervention, Institute of Psychology, Goethe-University Frankfurt, Germany
5) Department of General Psychiatry, Center for Psychosocial Medicine, Heidelberg University
6) Institute of Psychiatric and Psychosomatic Psychotherapy, Central Institute of Mental Health Mannheim, Medical Faculty Mannheim, Heidelberg University, Germany
7) Sagol Center for Brain Function, Wohl Institute for Advanced Imaging, Tel-Aviv Sourasky Medical Center and Sagol School of Neuroscience, School of Psychological Sciences and Faculty of Medicine, Tel-Aviv University, Israel

Abstract

Pathophysiological models are urgently needed for personalized treatments of mental disorders. However, most potential neural markers for psychopathology are limited by low interpretability, prohibiting reverse inference from brain measures to clinical symptoms and traits. Neural signatures—i.e. multivariate brain-patterns trained to be both sensitive *and* specific to a construct of interest—might alleviate this problem, but are rarely applied to mental disorders. We tested whether previously developed neural signatures for negative affect and discrete emotions distinguish between healthy individuals and those with mental disorders characterized by emotion dysregulation, i.e. Borderline Personality Disorder (BPD) and complex Post-traumatic Stress Disorder (cPTSD). In three different fMRI studies, a total sample of 192 women (49 BPD, 62 cPTSD, 81 healthy controls) were shown pictures of scenes with negative or neutral content. Based on pathophysiological models, we hypothesized higher negative and lower positive reactivity of neural emotion signatures in participants with emotion dysregulation. The expression of neural signatures differed strongly between neutral and negative pictures (average Cohen's d = 1.17). Nevertheless, a mega-analysis on individual participant data showed no differences in the reactivity of neural signatures between participants with and without emotion dysregulation. Confidence intervals ruled out even small effect sizes in the hypothesized direction and were further supported by Bayes factors. Overall, these results support the validity of neural signatures for emotional states during fMRI tasks, but raise important questions concerning their link to individual differences in emotion dysregulation.

Keywords: neuroimaging, emotion regulation, borderline personality disorder, post-traumatic stress disorder, neural signature, meta-analysis

1. Introduction

1.1 Background

About 30% of the global population are estimated to suffer from a mental disorder during their lifetime, accompanied by significant human and societal costs (Steel et al., 2014; Whiteford et al., 2013). As for most physical maladies, biological explanations have a long history in this realm (Barondes, 1990). In the last 20 years, functional neuroimaging in particular has become a fundamental research strategy to improve our understanding of mental disorders. Most commonly, clinical researchers, practitioners, and patients are interested in features of the brain to infer clinical traits on a psychological level. For such *reverse inference*, neurobiological features must be both sensitive and specific, i.e. highly predictive of the psychological concept of interest, but not other distinct concepts (Poldrack, 2011). Unfortunately, with few exceptions, classic neural measures like average regional activity are not task-specific (Yarkoni et al., 2011) and have low test-retest reliability (Elliott et al., 2020), precluding reverse inference from brain activity to complex psychological constructs.

Neural signatures have been proposed as a solution to this problem (Woo et al., 2017). They can be defined as statistical models, which predict a psychological concept from brain data with great precision, but also distinguish it from similar but meaningfully different concepts (Kragel et al., 2018). For example, a machine learning-based multivariate neural signature of physical pain can be highly predictive of self-reported pain ratings, but distinguishes it from the concept of socio-emotional 'pain' following social rejection and vice versa (Woo et al., 2014). Hence, neural signatures ensure interpretability regarding psychological states above other brain-based approaches. Moreover, they might remedy the very low test-retest reliability of non-pattern brain measures (Gianaros et al., 2020; Kragel et al., 2020) as well as increase statistical power by limiting the number of statistical comparisons to a single neural indicator for the process of interest. Despite these advantages, validated neural signatures have rarely been applied to explain individual differences, particularly regarding clinical research questions on mental disorders.

Some mental disorders such as Borderline Personality Disorder (BPD) and complex Post-traumatic Stress Disorder (cPTSD) are characterized by pervasive emotion dysregulation, comprising increased emotional reactivity and deficits in emotion regulation (American Psychiatric Association, 2013; Brewin et al., 2017; Carpenter & Trull, 2013; Linehan, 1993). For the reactivity component, dominant pathophysiological models posit that presumably emotion-generating brain regions are hyperactive in response to negative (or even neutral)

stimuli (Brendel et al., 2005; Sicorello & Schmahl, 2020; Swartz et al., 2015). Especially for the amygdala, there is compelling evidence of hyperactivity in these disorders (Bryant et al., 2020; Schulze et al., 2019). Still, amygdala hyperactivity does not warrant reverse inference to heightened emotional reactivity, as it is not specific to negative emotions, but rather involved in a large spectrum of both valence-independent emotional and non-emotional processes (Cunningham & Brosch, 2012; Lindquist et al., 2016; Ousdal et al., 2008; Sander et al., 2003; Todorov, 2012; Wager et al., 2015). Hence, there is still no clear evidence demonstrating emotional hyperreactivity on a brain basis in these disorders.

Several neural signatures of emotions have been developed which are suitable to address this issue, which draw from sparse distributed information across the brain. The picture induced negative emotion signature (PINES; Chang et al., 2015) predicted one-item self-ratings of negative affect following negative pictures with a product-moment correlation above .90, outperforming single resting-state networks and regions, demonstrated dissociability from neural patterns of physical pain, and maintained its cross-validated accuracy in a hold-out sample. Complementary to this pattern for global negative affect, Kragel and LaBar (2015) developed seven patterns which distinguish discrete video-induced emotions from each other at an accuracy close to 40% (chance is ≈14%), including the emotions of fear, anger, sadness, surprise, amusement, contentment and a neutral reference state. Classification accuracy was also above chance when tested on music clips, supporting cross-modal validity. Moreover, in a large resting state fMRI sample of young healthy university students, spontaneous activity of the sadness pattern was associated with an epidemiological depression scale, while the fear pattern was associated with trait anxiety (Kragel et al., 2016). This study provides first evidence that individual differences in the expression of neural emotion networks might map on traits related to the differential experience of emotions on a self-report level.

Expanding this approach to a clinical setting, we tested herein whether the activity of these previously developed neural signatures for general negative affect (i.e. PINES; Chang et al., 2015) and discrete emotions (Kragel & LaBar, 2015) in response to pictures of negative (versus neutral) scenes distinguished women with emotion dysregulation from healthy controls. Negative scenes are among the most common stimuli to study negative emotional reactivity in mental disorders (McDermott et al., 2018). Analyses were conducted across three datasets, each including a clinical group characterized by emotion dysregulation (2 BPD, 1 cPTSD), aggregating results with a mega-analytic approach based on individual participant data.

First, we tested whether neural signatures were differentially expressed in the two experimental conditions. When viewing negative pictures, we expected the pattern expression

of negative affect (PINES signature) as well as fear, anger, and sadness (discrete emotion signatures) to be increased (hypothesis 1). Second, for the main research question, common models of the disorders predict heightened reactivity of negative emotions. Here, this translates to increased reactivity of the patterns for negative affect as well as fear, anger, and sadness in participants with emotion dysregulation (hypothesis 2).

Previously, we observed that naturalistic everyday life stressors are associated not only with higher negative affect, but also lower positive affect (Sicorello, Dieckmann, et al., 2020). Therefore, we included additional analyses on neural signatures for positive emotions as well. We predicted the pattern expression of amusement and contentment to be decreased in the negative condition. We predicted stronger deactivation of these patterns in the emotion dysregulation groups. For the surprise pattern, we expected a higher expression in the negative condition, but had no directional between-group hypothesis. Last, the neutral pattern indicates the presence (or absence) of any discrete emotional state. As the paradigm is designed to elicit negative emotions, we expected neutral states to be decreased in the negative condition and more strongly so in the emotion dysregulation group.

2. Methods and Materials

2.1 Samples and Procedure

Three studies comprising a total of 192 women were included in the analyses of which 111 had a diagnosis of BPD or cPTSD. All participants were presented negative and neutral pictures during fMRI.

Study 1 comprised 57 women (29 with BPD, 28 healthy controls) who participated in a randomized controlled trial on BPD psychotherapy (German Clinical Trials Register: DRKS00000778). Only results from cross-sectional data collected before the intervention are reported here. Participants completed an fMRI experiment with three event-related runs, all with the same structure and number of trials. Each run involved a negative and a neutral condition presented after a “view” instruction. Either negative pictures or pictures of objects where shown, respectively. The experiment also involved regulate-conditions that were not analyzed here, where participants had to regulate their emotional response. Pictures were presented for 6s. Longitudinal results on therapy-effects in this sample have been published previously (Niedtfeld et al., 2017; Schmitt et al., 2016; Winter et al., 2017).

Study 2 comprised 40 women (20 with BPD, 20 healthy controls), who completed three runs of a picture viewing task with different designs: block-design (one picture per block, 18s),

mixed-design (three pictures per block, 6s each), and event-related design (6s per picture). Participants viewed negative pictures (negative condition) and scrambled images (neutral condition). Data on the healthy group have been published previously (Paret et al., 2014).

Study 3 comprised 95 women (62 with cPTSD, 33 healthy controls), who were recruited from a larger randomized controlled psychotherapeutic trial (German Clinical Trials Register: DRKS00005578), and therapy-effects were recently published (Bohus et al., 2020). Only results from cross-sectional data collected before the intervention are reported here. Additionally to the DSM-5 criteria for PTSD, participants met at least three out of nine DSM-IV criteria for BPD, including criterion six for emotional instability. Negative pictures and neutral pictures were presented as distractors within a Sternberg working memory task for 1.5s and entered the analysis as negative condition and neutral condition, respectively. Neutral pictures matched with the negative pictures for complexity and content were used in the neutral baseline condition. FMRI data from 34 women of the cPTSD group have been published previously to test a different hypothesis against a trauma-exposed healthy control group (Sicorello, Thome, et al., 2020). The trauma-exposed control group was not included in the analyses here.

Comprehensive descriptions of sample characteristics, designs, procedures, scanning parameters, and preprocessing for all three studies can be found in the supplemental material.

2.2 Pattern expression

We downloaded the pattern-masks of each neural signature (PINES and the seven discrete emotion signatures) from the CANlab github repository: https://github.com/canlab. These pattern masks are freely available and consist of a brain image with a regression weight for each brain voxel. Pattern expression was calculated as the dot product between the pattern mask and an image containing beta weights from the first-level analysis for the respective regressor of interest (negative or neutral condition), separately for each picture condition, run, and participant. For the PINES, pattern expression reflects the predicted negative affect rating. For discrete emotions, pattern expression is a continuous indicator to what degree a given emotion category is more likely than the remaining categories. Notably, expression values cannot be directly compared between studies, as their scale depends on scanning parameters, scanner-specific gain and signal characteristics, and analysis choices. Expression values can, however, be compared across task conditions and participants if these values can be assumed to be constant across participants. As an index of reactivity, pattern expression during the neutral condition was subtracted from pattern expression during the negative condition.

As an indicator of internal consistency, we calculated the reliability for the pattern responses as Cronbach's alpha between experimental runs when more than one run was available (studies 1 and 2). All runs occurred in the same fMRI session. For study 1, pattern responses had a mean reliability of $\alpha = .58$, ranging from $\alpha = .48$ for anger to $\alpha = .66$ for the PINES and fear. As could be expected from previous reports (Gianaros et al., 2020; Kragel et al., 2020), the reliability was higher for pattern expression than for the mean response in an amygdala-hippocampal region-of-interest (ROI; $\alpha = .14$), which was defined from the thresholded mask of a previous functional meta-analysis on emotion processing in BPD, (Schulze et al., 2019; https://identifiers.org/neurovault.collection:3751). For study 2, pattern responses had a mean reliability of $\alpha = .64$, ranging from $\alpha = .56$ for amused to $\alpha = .72$ for fear. Again, reliability of the amygdala-hippocampal ROI was substantially lower at $\alpha = .31$. The correlation between pattern expressions in the event-related design and the two block designs was lower than between the two block designs, but not in a range indicating conclusive differences, given the sample size: r(event-related, block) = .26, r(event-related, mixed-block) = .37, r(block, mixed-block) = .57.

2.3 Statistical Analyses

2.3.1 Negative versus neutral condition

To test whether the expression of neural signatures differed between the negative and the neutral condition in studies 1-3, reflecting pattern reactivity, one-sample *t*-tests were conducted on the difference scores. Cohen's *d* was calculated as the mean difference score divided by the standard deviation of difference scores. The three runs of study 1 were averaged for this analysis, as the runs showed good compatibility in terms of sufficient internal consistency and only small differences in mean effects. Runs of study 2 were analyzed separately, to allow the inspection of design-dependent effects and as the three runs had large differences in mean activations, due to the different stimulus presentation parameters.

The corresponding within-person mega-analysis was conducted using a two-level multilevel analysis framework, with difference scores nested within participants (because of the multiple runs in studies 1 and 2). The difference score Δ_{ijk} of run i within participant j of study k was regressed on a fixed intercept γ_{000}, including random intercepts for study-participants ζ_{0jk}. as well as a residual term ε_{ijk}: $\Delta_{ijk} = \gamma_{000} + \zeta_{0jk} + \varepsilon_{ijk}$. Due to the low number of studies, the study-wise random intercept ζ_{00k} was not included. Moreover, Δ_{ijk} was scaled on the run-specific standard deviation $SD_{i\bullet k}$. With this scaling, γ_{000} is in the metric of the Cohen's *d* used for single study analyses and on a compatible scale between studies and runs, regardless

of e.g. design effects. All frequentist multilevel analyses were conducted using the lmer function of the lme4 package in R version 4.0.3 (Bates et al., 2015) and restricted maximum likelihood estimation.

2.3.2 Group effects

For single studies, differences between the clinical and the healthy groups were tested with two-sample *t*-tests for unequal variances and pattern reactivity (Δ) as the dependent variable. Cohen's *d* was calculated as the difference in group means divided by the pooled standard deviation.

The mega-analysis was specified as $\Delta_{ijk} = \gamma_{100}(\text{group}) + \zeta_{0jk} + \varepsilon_{ijk}$, where γ_{100} represents the fixed effect of group. As for within-analysis, the corresponding random effect for group $\zeta_{10k}(\text{group})$ was not included due to the low number of studies. The group variable was recoded within runs, so that all intercepts (and their variance) are zero. Therefore, the fixed intercept γ_{000} and its variance between studies ζ_{00k} can be omitted from the model. For balanced group sizes (study 2), this can be achieved by coding groups as -0.5 and 0.5, with the regression weight representing the mean difference between groups. For unbalanced group sizes (studies 1 and 3), weighted effect coding was used (Grotenhuis et al., 2017). Moreover, Δ_{ijk} was standardized within runs by subtracting the run-specific mean and dividing by the run-specific pooled standard deviation. With this standardization, γ_{100} is in the metric of Cohen's *d*, as used for single study analyses.

2.3.3 Bayes factors

Bayes factors were calculated for all models to quantify the relative evidence of the H_0 over the H_1 (e.g. effect = 0 versus effect ≠ 0), using the low information cauchy prior with a scale factor of 0.707, which is the default of the R package used here and was previously suggested for psychological applications (Wagenmakers et al., 2018). Bayes factors are a ratio between p(Data|H_1) and p(Data|H_0), with values above 3 (or below 1/3) often used as a minimum cutoff for claims of evidence in favor of one hypothesis over the other, although continuous interpretations are recommended as well (Jarosz & Wiley, 2014). BF_{10} denotes evidence for the H_1, divided by the evidence for H_0; BF_{01} denotes evidence for the H_0, divided by the evidence for H_1. BF_{10} equals $1/BF_{01}$ and vice versa.

To compute Bayes factor for tests in singles studies, the function ttestBF() of the Bayes factor package was used in R (Morey & Rouder, 2018). For mega-analyses, the multilevel models were refitted using the *brms* package (Bürkner, 2017), comparing models with (H_1) and models without (H_0) the effect of interest using the function bayes_factor().

In accordance with our hypotheses stated in the introduction, all Bayes factors reflected directional one-sided tests, except for the between-group effect of surprise. This was achieved by modelling a half-cauchy for the H_1 in the hypothesized direction. We argue this is appropriate here, as the Bayes factor should reflect evidence for/against the alternative hypothesis of interest, e.g. neural expression of fear is higher when viewing pictures with negative content (and neither zero *nor* lower).

2.3.4. Reproducible Analyses

Data and annotated R scripts to reproduce the main analyses can be found on: https://github.com/MaurizioSicorello/MVPAemoDys_Analyses.git.

Demographic information used for sample description and fMRI images are not openly provided. Requests for primary data should be addressed directly to the corresponding author.

3. Results

3.1 Comparison between negative and neutral pictures

In line with our hypothesis, both mega-analyses and single-study analyses indicated that neural signatures of negative affect and negative emotions were expressed more strongly while viewing negative pictures, except for the sadness pattern (figure 1, table 1). Likewise, neural signatures of positive emotions were expressed more strongly in the neutral conditions. Effect sizes were overall large, ranging between $d = 0.81$ (anger) and $d = 2.07$ (PINES/negative affect). Only the signature for sadness had a small effect of $d = -0.23$, which went in the opposite direction than expected, i.e. sadness was expressed more strongly in the neutral condition. The null hypothesis that the condition effect for sadness is zero or negative was 60 times more likely than the hypothesized positive effect, i.e. increased neural expression in the negative condition. Study 2 indicated that mixed-block design elicited the largest effects and the event-related design the smallest effects, as has been previously reported for the mass univariate ROI approach (Paret et al., 2014).

In the original validation study, the PINES distinguished the highest and the lowest negative affect ratings at an accuracy of 93.5% (Chang et al., 2015). A logistic regression of picture condition on PINES expression revealed mostly lower but compatible accuracies, with the highest accuracy in the mixed-block design of study 2 and the lowest accuracy in study 3, which had the shortest stimulus presentation duration: Study 1 = 82% [74.21%, 88.94%]; Study $2_{\text{Event-Related}}$ = 71.25% [60.05%, 80.82%]; Study 2_{Block} = 87.50% [78.21%, 93.84%]: Study 2_{Mixed} = 97.50% [91.26%, 99.70%]; Study 3 = 64% [56.41%, 70.52%] (95% confidence intervals in

brackets calculated based on Clopper & Pearson (1934).

In sum, these results overall support our first hypothesis that neural signatures of emotions are differentially expressed when viewing negative and neutral pictures in the hypothesized directions, except for the sadness pattern. The estimated effect sizes were very large, but also appeared to depend on design aspects of the studies.

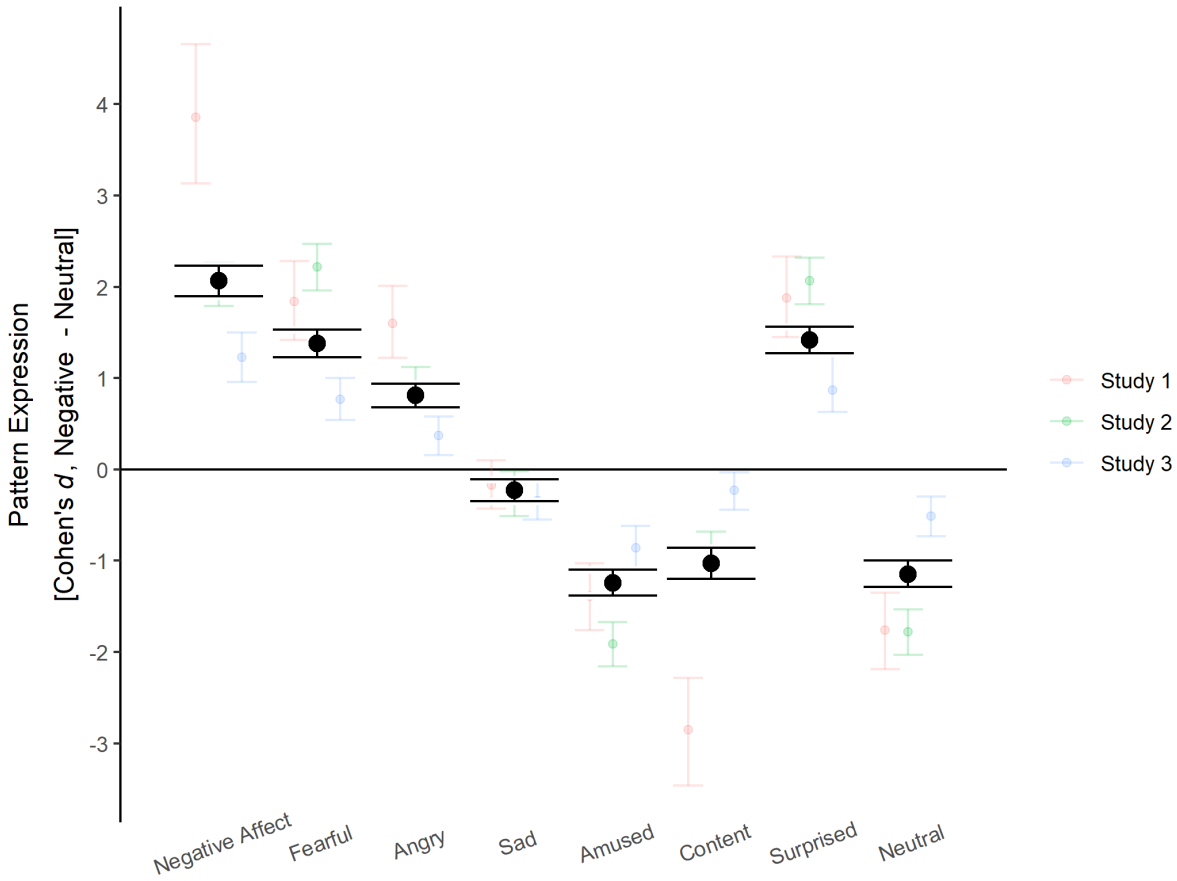

Figure 1. Differences in the expression of neural emotion signatures between the negative and the neutral condition. Error bars show 95% confidence intervals.

3.2 Comparison between clinical groups and healthy controls

Most mega-analytic group effects were very small (all $|d| \leq 0.17$; figure 2, table 2). Contrary to hypothesis 2—i.e. higher neural pattern reactivity of negative emotions in participants with emotion dysregulation compared to healthy controls—the former actually showed lower reactivity of neural signatures for negative affect, fear, and anger. The upper confidence limit for these three emotions did not include values higher than $d = 0.12$ and Bayes factors favored the null hypothesis of equal or smaller neural signature reactivity in the emotion dysregulation groups. While the emotion dysregulation group did show the expected tendency

of higher expression for sadness, the effect was very small (d = 0.06), confidence intervals covered zero and had an upper limit at a small effect size of d = 0.31, and the Bayes factor favored the null (BF_{01} = 9.54). Moreover, the condition-wise analyses indicated this emotion signature might not be a valid measure given the stimulus material. Group effects for neutral states, amusement, contentment, and surprise did not differ considerably from zero. These results were supported by Bayes factors, except for surprise, whose Bayes factor was relatively inconclusive (BF_{01} = 2.15). On a single study-basis, this pattern was overall present in studies 1 and 2. The descriptive effect directions in study 3 were more compatible with the theoretical predictions, albeit with miniscule effect sizes and inconclusive Bayes factors. These results were stable when a binary indicator for psychotropic medication was included as a covariate (figure S1).

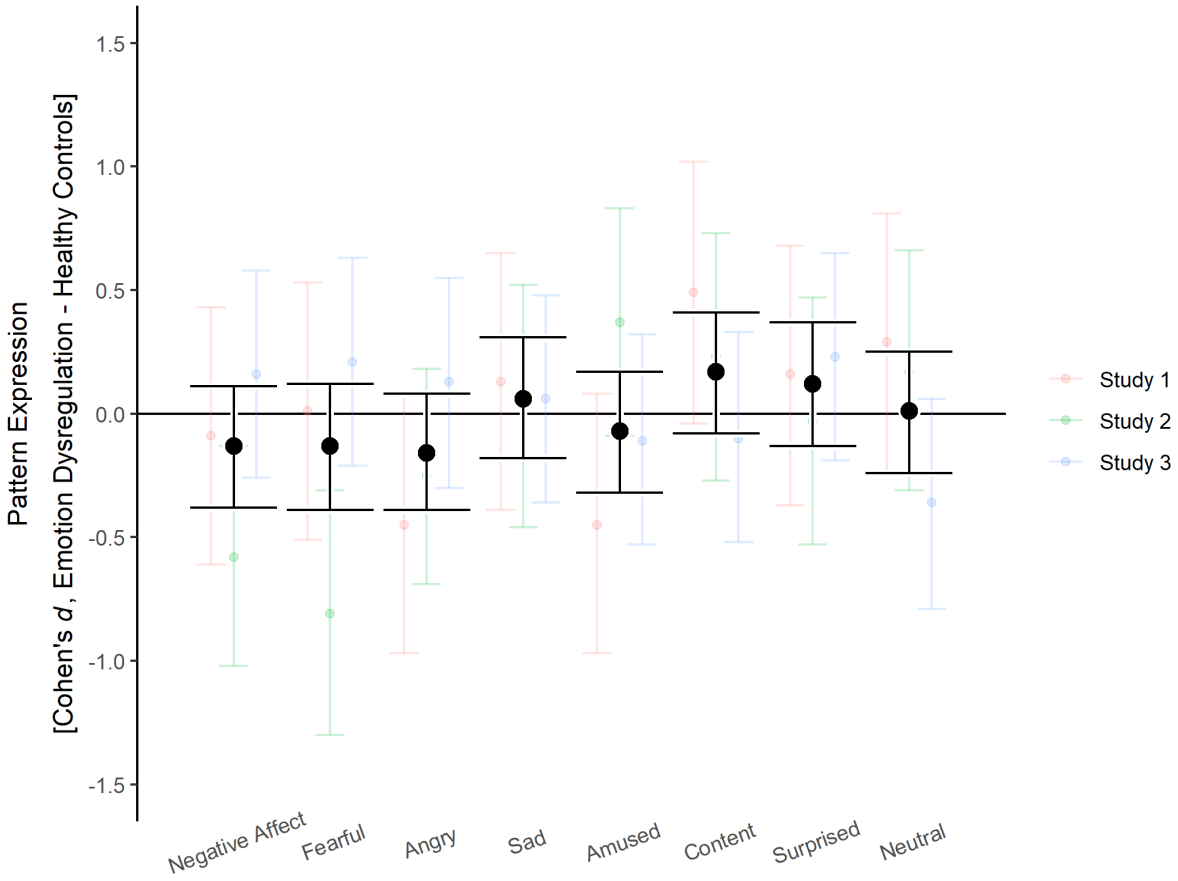

Figure 2. Group differences in the reactivity of neural emotion signatures between participants with emotion dysregulation and healthy controls. Error bars show 95% confidence intervals.

3.3 Exploratory Analyses: Group effects on the neutral baseline

There is some evidence that people with emotion dysregulation have a higher propensity to interpret neutral stimuli as negative (Daros et al., 2013; Mitchell et al., 2014), accompanied by heightened amygdala responses (Donegan et al., 2003; Lischke et al., 2017; Niedtfeld et al., 2010). As this might diminish group differences in the negative-neutral contrast, we repeated the between-group analyses of section 3.2 with activation in the neutral condition as the dependent variable, instead of the difference between negative and neutral conditions.

In these analyses, all confidence intervals contained zero by a considerable margin (figure S2). Still, even statistically non-significant group effects on the neutral baseline might diminish group effects on the difference scores used to indicate neural reactivity. In the neutral condition, participants with emotion dysregulation had slightly increased responses for the fear pattern (d = 0.17, 95% CI = [-0.08, 0.43]) and decreased responses for the contentment pattern (d = -0.10, 95% CI = [-0.35, 0.15]). Hence, the hypothesized effects for these two patterns might be diminished by group differences in response to the neutral condition. All other effects were in the opposite direction of what would be expected if an increased responsiveness to neutral stimuli accounts for the null effects reported in section 3.2 (e.g. participants with emotion dysregulation had a lower expression of the PINES signature and a higher expression of the neutral signature).

To follow up on the potential attenuation effect for fear and contentment, we repeated the mega-analytic procedure on pattern expression in the negative condition against the implicit baseline (figure S3). The estimates for the fear and contentment patterns were almost perfectly zero, although confidence intervals of the fear pattern still included small to moderate effect sizes (fear: d = 0.00, 95% CI = [-0.26, 0.26]; contentment: d = 0.01, 95% CI = [-0.24, 0.26]).

Coincidentally, we observed that the confidence interval of the effect of lower negative affect in the emotion dysregulation group vs. the healthy control group no longer contained zero (d = -0.32, 95% CI = [-0.57, -0.07]), which differs from the results for the negative-neutral contrast.

4. Discussion

To translate neurobiological models of mental disorders into the clinical language of traits and symptoms, neural markers have to be both sensitive *and* specific to the psychological concept of interest. This is rarely the case for properties of discrete anatomical brain regions

like the amygdala, which nonetheless has been frequently used as an indicator of negative emotional processes in affect-related disorders, while it is also involved in a broad set of psychological phenomena other than emotions. Here, we used machine learning-based multivariate neural signatures for emotional states to test whether people with emotion dysregulation show signs of hyperreactive neuro-emotional systems. This assumption of leading psychopathological models was assessed in three independent studies from our lab, investigating participants diagnosed with either BPD or cPTSD and healthy controls.

Neural signatures of negative affect (Chang et al., 2015) and discrete emotions (Kragel & LaBar, 2015) showed strong differential expression between the negative and the neutral condition in the expected directions (hypothesis 1), supporting their validity and accuracy, even when transferred to a different lab, experimental design, and population than the initial validation studies. Effect sizes were very large and supported by very large Bayes factors in each of the three studies. Moreover, study 2 indicates that effect sizes might be partly related to stimulus presentation parameters such as exposure time. Notably, the effect observed with the sadness signature was in the opposite direction than expected (neutral > negative condition). As the stimuli were chosen based on valence and arousal ratings, it is possible that sadness-inducing pictures were underrepresented or that sadness is harder to induce with briefly presented pictures.

Most importantly, the neural signatures did not differentiate between participants with and without emotion dysregulation, speaking against the main hypothesis of the present study (hypothesis 2). Except for the sadness and amusement signatures, all effects went in the opposite direction from the theoretical predictions, i.e. smaller negative emotional reactivity and positive emotional reactivity in the emotion dysregulation group vs. the healthy group. The corresponding confidence intervals ruled out even small effect sizes in the expected direction, below $|d| = 0.20$ and Bayes factors favored the null hypothesis for all signatures, except for surprise, which was inconclusive. Similar patterns emerged for separate analyses on studies 1 and 2, while the results in study 3 were less conclusive in terms of Bayes factors. These results could not be explained by a heightened response to the neutral condition in those with BPD and cPTSD, which has been observed previously for amygdala reactivity.

These findings are incompatible with the dominant pathological model of BPD and provide evidence against either the theoretical, experimental, or neurobiological assumptions of the present study, which we discuss below. Either way, important implications arise for future research. To discuss these potential explanations of the reported results, we mainly draw from the BPD literature, as the cPTSD literature is still relatively limited and the BPD criterion

for emotional instability was the cardinal criterion for inclusion in study 3.

Showing participants pictures of scenes with negative content is among the most common tasks to experimentally investigate heightened emotional reactivity in mental disorders and affect-related traits. This approach rests on the implicit assumptions that (1) the clinical phenomenon of heightened emotional reactivity is not fully accounted for by more negative environments, a lower threshold for emotional responses, or difficulties in emotion regulation, (2) emotional reactivity can be observed outside of its naturalistic daily life context, and (3) the emotion-inducing effect of experimental stimuli is not limited to stimuli personalized according to thematic relevance. If correct, these assumptions naturally lead to the conclusion that people with emotion dysregulation must have generally hyperresponsive emotion generating biological systems, whose exploration could aid the understanding and treatment of such disorders. Further, our aim to investigate these biological systems with neural signatures was based on the assumption that (4) neural signatures represent the best available neural markers for such systems, due to their high accuracy for emotional states.

Apart from qualitative clinical impression of therapeutic practitioners, there is empirical evidence for increased reactivity to discrete naturalistic everyday life stressors in BPD (Glaser et al., 2008; Hepp et al., 2018). Notably, such studies cannot easily distinguish precisely which aspects of emotion processing are aberrant, due to their relatively low temporal resolution (assumption 1). Experimental settings offer higher control and better temporal resolution, but suffer from limited ecological validity, as stressors are presented outside of their natural context (assumption 2). A recent meta-analytic review found that the literature is surprisingly inconclusive concerning experimentally induced emotional reactions in BPD (Bortolla et al., 2020). While they did find moderate experimental group effects on affective self-ratings in their meta-analysis, many studies did not include a pre-measurement, potentially confounding tonic negative emotions and emotional reactivity, or only had pre- and post-task ratings, which might capture other processes than stimulus-contingent real-time responses. Moreover, peripheral-physiological effects were negligibly small and/or statistically not significant. Interestingly, there was no statistically significant difference in effect sizes dependent on whether stimuli were thematically related to BPD (assumption 3).

Taken together, it is possible that typical laboratory designs, as used in our studies, are not well-suited to probe individual differences in emotional reactivity which generalize to everyday life or that clinical subgroups with opposing phenomenology cancel each other's effects. Alternatively, it is possible that the neural signatures do not capture the psychological concept of interest well (assumption 4). If the concepts of interest are emotions as they are

measured by self-reports, this seems unlikely for the PINES, as it correlated with self-reports above $r = .90$ in both the training and the hold-out sample, which employed a design similar to ours. Still, it is possible that when asked for their mood directly after seeing a negative picture, participants partly rate the picture content, rather than exclusively their emotions, which could have impeded the construct validity of the PINES. Nevertheless, this argument does not hold for the discrete emotion signatures, which distinguish emotion categories and were associated with trait depressiveness and anxiety in a well-powered resting-state study.

Another neurobiological explanation of the null results might be the presence of stable physiological between-person noise (e.g. cerebrovasculature or hematocrit levels; D'Esposito et al., 2003; Yang et al., 2014). A recent meta-analysis demonstrated that test-retest reliability of resting state fMRI diminishes considerably after artefact correction, indicating the presence of such stable between-person noise (Noble et al., 2019). The neural signatures used here have been developed to explain variance without explicit differentiation of the within- or between-person level and their high accuracy might be preferentially due to variance within individuals. Notably, while machine learning-based approaches have been increasingly used to differentiate between clinical groups based on fMRI data (Gao et al., 2018; Woo et al., 2017), these approaches do not necessarily lead to interpretable neural markers, as groups might differ on many confounded dimensions.

4.1 Limitations and future directions

The mega-analyses did not include random slopes for studies, as the low number of studies does not allow a sensible estimate of between-study variance. Hence, the generalizability to other experimental investigations is limited and a wider range of effect sizes should be expected (Yarkoni, 2020). This limitation on generalizability is especially important, as studies included only female participants, due to potential gender-differences in symptom presentation (Sansone & Sansone, 2011). Study 2 indicated that stimulus presentation parameters might be one important influence on effect size differences, at least for within-person effects.

Another limitation to consider is the reliability of fMRI-based neural markers (Elliott et al., 2020). Testing the internal consistency for multi-run studies 1 and 2 indicated that reliability was considerably higher for neural signatures than for an amygdala-hippocampal cluster from a BPD meta-analysis, but still lower than desirable, ranging from $\alpha = .48$ to $\alpha = .72$. These estimates could be used in future studies to correct expected effect sizes for unreliability in power analyses.

As in most BPD studies which used fMRI designs with negative scenes, there were no affective self-ratings directly following pictures. Such ratings would be necessary to closely replicate the core assumption of the neural emotion signatures, that is, they predict momentary subjective affect ratings by means of BOLD responses to affective stimuli across different populations. More research is urgently needed to confirm the strict validity of neural signatures in clinical populations. Post-session valence ratings of negative pictures did not differ considerably between participants, as has been previously reported (Koenigsberg et al., 2009; Schulze et al., 2011), but are not necessarily a valid surrogate of *momentary* affect, immediately following negative trials. While these tasks have been frequently used, there has been to our knowledge no thorough psychometric validation to ensure their usefulness for research on individual differences on the psychological end. Therefore, we suggest a systematic assessment of their test-retest reliability and validity in terms of associations with clinically relevant traits, independent of neuroimaging techniques. As stated above, it is unclear whether valence ratings following the session should continue to replace self-ratings of affect immediately following image-exposure.

4.2 Conclusion

Neural signatures of emotions appear to be valid and transferable tools to investigate within-person relationships, but their utility to understand individual differences remains unclear. Contrary to theoretical expectations, we did not find differences between people with and without emotion dysregulation. We offer to share our analysis pipelines with other research groups to reanalyze existing datasets. This could be done efficiently and lead to a more comprehensive picture of the relationship between neural signatures and emotion-related traits. Apart from neurobiological approaches, more research is needed concerning the psychometric properties and ecological validity of typical experimental tasks used to probe affective traits.

Acknowledgements

Thanks are due to Rosemarie Kluetsch, Steffen Hoesterey, Michael Rieß and Claudia Stief for support in data collection, Dorina Winter and Ruth Schmitt for study setup and data processing, to Madita Stirner, Katharina Brunner, and Elena Buck for compiling questionnaire data and to the KFO256-Central Project team for providing excellent organizational support and clinical diagnostics. Study 1 was funded by the German Research Foundation (grant no. SCHM 1526/8-2; HE2660/7-2). Study 2 was supported by the German Research Foundation (grant no. KFO 256, EN 361/13-2). Study 3 was financed by the German Ministry of Education and Research (BMBF) RELEASE 01KR1303A.

Author contribution

Conflicts of interest

C. Schmahl and C. Paret have served as consultants to Boehringer Ingelheim Pharma.

5. References

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders, 5th ed* (5th ed.). American Psychiatric Association.

Barondes, S. H. (1990). The biological approach to psychiatry: history and prospects. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *10*(6), 1707–1710. https://doi.org/10.1523/JNEUROSCI.10-06-01707.1990

Bates, D., Maechler, M., & Bolker, B. (2015). Walker., S. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bohus, M., Kleindienst, N., & Hahn, C. (2020). Dialectical behavior therapy for posttraumatic stress disorder (DBT-PTSD) compared with cognitive processing therapy (CPT) in complex presentations of PTSD …. *JAMA: The Journal of the American Medical Association*. https://doi.org/10.1001/jamapsychiatry.2020.2148

Bortolla, R., Cavicchioli, M., Fossati, A., & Maffei, C. (2020). Emotional reactivity in borderline personality disorder: Theoretical considerations based on a meta-analytic review of laboratory studies. *Journal of Personality Disorders*, *34*(1), 64–87. https://doi.org/10.1521/pedi_2018_32_382

Brendel, G. R., Stern, E., & Silbersweig, D. A. (2005). Defining the neurocircuitry of borderline personality disorder: Functional neuroimaging approaches. *Development and Psychopathology*, *17*(4), 1197–1206. https://doi.org/10.1017/S095457940505056X

Brewin, C. R., Cloitre, M., Hyland, P., Shevlin, M., Maercker, A., Bryant, R. A., Humayun, A., Jones, L. M., Kagee, A., Rousseau, C., Somasundaram, D., Suzuki, Y., Wessely, S., van Ommeren, M., & Reed, G. M. (2017). A review of current evidence regarding the ICD-11 proposals for diagnosing PTSD and complex PTSD. *Clinical Psychology Review*, *58*, 1–15. https://doi.org/10.1016/j.cpr.2017.09.001

Bryant, R. A., Felmingham, K. L., Malhi, G., Andrew, E., & Korgaonkar, M. S. (2020). The distinctive neural circuitry of complex posttraumatic stress disorder during threat processing. *Psychological Medicine*, 1–8. https://doi.org/10.1017/S0033291719003921

Bürkner, P.-C. (2017). brms: An R package for bayesian multilevel models using stan. *Journal of Statistical Software, Articles*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Carpenter, R. W., & Trull, T. J. (2013). Components of emotion dysregulation in borderline personality disorder: a review. *Current Psychiatry Reports*, *15*(1), 335. https://doi.org/10.1007/s11920-012-0335-2

Chang, L. J., Gianaros, P. J., Manuck, S. B., Krishnan, A., & Wager, T. D. (2015). A sensitive and specific neural signature for picture-induced negative affect. *PLoS Biology*, *13*(6), 1–28. https://doi.org/10.1371/journal.pbio.1002180

Clopper, C. J., & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, *26*(4), 404–413. https://doi.org/10.2307/2331986

Cunningham, W. A., & Brosch, T. (2012). Motivational salience: Amygdala tuning from traits, needs, values, and goals. *Current Directions in Psychological Science*, *21*(1), 54–59. https://doi.org/10.1177/0963721411430832

Daros, A. R., Zakzanis, K. K., & Ruocco, A. C. (2013). Facial emotion recognition in borderline personality disorder. *Psychological Medicine*, *43*(9), 1953–1963. https://doi.org/10.1017/S0033291712002607

D'Esposito, M., Deouell, L. Y., & Gazzaley, A. (2003). Alterations in the BOLD fMRI signal with ageing and disease: a challenge for neuroimaging. *Nature Reviews. Neuroscience*, *4*(11), 863–872. https://doi.org/10.1038/nrn1246

Donegan, N. H., Sanislow, C. A., Blumberg, H. P., Fulbright, R. K., Lacadie, C., Skudlarski, P., Gore, J. C., Olson, I. R., McGlashan, T. H., & Wexler, B. E. (2003). Amygdala

hyperreactivity in borderline personality disorder: Implications for emotional dysregulation. *Biological Psychiatry*, *54*(11), 1284–1293. https://doi.org/10.1016/S0006-3223(03)00636-X

Elliott, M. L., Knodt, A. R., Ireland, D., Morris, M. L., Poulton, R., Ramrakha, S., Sison, M. L., Moffitt, T. E., Caspi, A., & Hariri, A. R. (2020). What is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis. *Psychological Science*, *31*(7), 792–806. https://doi.org/10.1177/0956797620916786

Gao, S., Calhoun, V. D., & Sui, J. (2018). Machine learning in major depression: From classification to treatment outcome prediction. *CNS Neuroscience & Therapeutics*, *24*(11), 1037–1052. https://doi.org/10.1111/cns.13048

Gianaros, P. J., Kraynak, T. E., Kuan, D. C.-H., Gross, J. J., McRae, K., Hariri, A. R., Manuck, S. B., Rasero, J., & Verstynen, T. D. (2020). Affective brain patterns as multivariate neural correlates of cardiovascular disease risk. *Social Cognitive and Affective Neuroscience*, *March*, 1–12. https://doi.org/10.1093/scan/nsaa050

Glaser, J.-P., Van Os, J., Mengelers, R., & Myin-Germeys, I. (2008). A momentary assessment study of the reputed emotional phenotype associated with borderline personality disorder. *Psychological Medicine*, *38*(9), 1231–1239. https://doi.org/10.1017/S0033291707002322

Grotenhuis, M. te, M., Pelzer, B., Eisinga, R., Nieuwenhuis, R., Schmidt-Catran, A., & Konig, R. (2017). When size matters: advantages of weighted effect coding in observational studies. *International Journal of Public Health*, *62*(1), 163-167. https://doi.org/10.1007/s00038-016-0901-1

Hepp, J., Lane, S. P., Wycoff, A. M., Carpenter, R. W., & Trull, T. J. (2018). Interpersonal stressors and negative affect in individuals with borderline personality disorder and community adults in daily life: A replication and extension. *Journal of Abnormal*

Psychology, *127*(2), 183–189. https://doi.org/10.1037/abn0000318

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting bayes factors. *The Journal of Problem Solving*, *7*(1), 1–9. https://doi.org/10.7771/1932-6246.1167

Koenigsberg, H. W., Fan, J., Ochsner, K. N., Liu, X., Guise, K. G., Pizzarello, S., Dorantes, C., Guerreri, S., Tecuta, L., Goodman, M., New, A., & Siever, L. J. (2009). Neural correlates of the use of psychological distancing to regulate responses to negative social cues: A study of patients with borderline personality disorder. *Biological Psychiatry*, *66*(9), 854–863. https://doi.org/10.1016/j.biopsych.2009.06.010

Kragel, P. A., Han, X., Kraynak, T., Gianaros, P., & Wager, T. (2020). fMRI can be highly reliable, but it depends on what you measure (v2). *PsyArXiv*. https://doi.org/10.31234/osf.io/9eaxk

Kragel, P. A., Knodt, A. R., Hariri, A. R., & LaBar, K. S. (2016). Decoding spontaneous emotional states in the human brain. *PLoS Biology*, *14*(9), 1–19. https://doi.org/10.1371/journal.pbio.2000106

Kragel, P. A., Koban, L., Barrett, L. F., & Wager, T. D. (2018). Representation, pattern information, and brain signatures: From neurons to neuroimaging. *Neuron*, *99*(2), 257–273. https://doi.org/10.1016/j.neuron.2018.06.009

Kragel, P. A., & LaBar, K. S. (2015). Multivariate neural biomarkers of emotional states are categorically distinct. *Social Cognitive and Affective Neuroscience*, *10*(11), 1437–1448. https://doi.org/10.1093/scan/nsv032

Lindquist, K. A., Satpute, A. B., Wager, T. D., Weber, J., & Barrett, L. F. (2016). The brain basis of positive and negative affect: Evidence from a meta-analysis of the human neuroimaging literature. *Cerebral Cortex* , *26*(5), 1910–1922. https://doi.org/10.1093/cercor/bhv001

Linehan, M. M. (1993). *Cognitive-behavioral Treatment of Borderline Personality Disorder*. Guilford Press. https://books.google.de/books?id=UZim3OAPwe8C

Lischke, A., Herpertz, S. C., Berger, C., Domes, G., & Gamer, M. (2017). Divergent effects of oxytocin on (para-)limbic reactivity to emotional and neutral scenes in females with and without borderline personality disorder. *Social Cognitive and Affective Neuroscience*, *12*(11), 1783–1792. https://doi.org/10.1093/scan/nsx107

McDermott, T. J., Kirlic, N., & Aupperle, R. L. (2018). Roadmap for optimizing the clinical utility of emotional stress paradigms in human neuroimaging research. *Neurobiology of Stress*, *8*, 134–146. https://doi.org/10.1016/j.ynstr.2018.05.001

Mitchell, A. E., Dickens, G. L., & Picchioni, M. M. (2014). Facial emotion processing in borderline personality disorder: A systematic review and meta-analysis. *Neuropsychology Review*, *24*(4), 166–184. https://doi.org/10.1007/s11065-014-9254-9

Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of Bayes Factors for Common Designs. R package version 0.9.12-4.2*. https://cran.r-project.org/package=BayesFactor

Niedtfeld, I., Schmitt, R., Winter, D., Bohus, M., Schmahl, C., & Herpertz, S. C. (2017). Pain-mediated affect regulation is reduced after dialectical behavior therapy in borderline personality disorder: a longitudinal fMRI study. *Social Cognitive and Affective Neuroscience*, *12*(5), 739–747. https://doi.org/10.1093/scan/nsw183

Niedtfeld, I., Schulze, L., Kirsch, P., Herpertz, S. C., Bohus, M., & Schmahl, C. (2010). Affect regulation and pain in borderline personality disorder: A possible link to the understanding of self-injury. *Biological Psychiatry*, *68*(4), 383–391. https://doi.org/10.1016/j.biopsych.2010.04.015

Noble, S., Scheinost, D., & Constable, R. T. (2019). A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. *NeuroImage*, *203*,

116157. https://doi.org/10.1016/j.neuroimage.2019.116157

Ousdal, O. T., Jensen, J., Server, A., Hariri, A. R., Nakstad, P. H., & Andreassen, O. A. (2008). The human amygdala is involved in general behavioral relevance detection: Evidence from an event-related functional magnetic resonance imaging Go-NoGo task. *Neuroscience*, *156*(3), 450–455. https://doi.org/10.1016/j.neuroscience.2008.07.066

Paret, C., Kluetsch, R., Ruf, M., Demirakca, T., Kalisch, R., Schmahl, C., & Ende, G. (2014). Transient and sustained BOLD signal time courses affect the detection of emotion-related brain activation in fMRI. *NeuroImage*, *103*, 522–532. https://doi.org/10.1016/j.neuroimage.2014.08.054

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. *Neuron*, *72*(5), 692–697. https://doi.org/10.1016/j.neuron.2011.11.001

Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: An evolved system for relevance detection. *Reviews in the Neurosciences*, *14*(4), 303–316. https://doi.org/10.1515/REVNEURO.2003.14.4.303

Sansone, R. A., & Sansone, L. A. (2011). Gender patterns in borderline personality disorder. *Innovations in Clinical Neuroscience*, *8*(5), 16. https://www.ncbi.nlm.nih.gov/pmc/articles/pmc3115767/

Schmitt, R., Winter, D., Niedtfeld, I., Herpertz, S. C., & Schmahl, C. (2016). Effects of psychotherapy on neuronal correlates of reappraisal in female patients with borderline personality disorder. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*, *1*(6), 548–557. https://doi.org/10.1016/j.bpsc.2016.07.003

Schulze, L., Domes, G., Krüger, A., Berger, C., Fleischer, M., Prehn, K., Schmahl, C., Grossmann, A., Hauenstein, K., & Herpertz, S. C. (2011). Neuronal correlates of cognitive reappraisal in borderline patients with affective instability. *Biological*

Psychiatry, *69*(6), 564–573. https://doi.org/10.1016/j.biopsych.2010.10.025

Schulze, L., Schulze, A., Renneberg, B., Schmahl, C., & Niedtfeld, I. (2019). Neural correlates of affective disturbances: A comparative meta-analysis of negative affect processing in borderline personality disorder, major depressive disorder, and posttraumatic stress disorder. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*, *4*(3), 220–232. https://doi.org/10.1016/j.bpsc.2018.11.004

Sicorello, M., Dieckmann, L., Moser, D., Lux, V., Luhmann, M., Neubauer, A. B., Schlotz, W., & Kumsta, R. (2020). Highs and lows: Genetic susceptibility to daily events. *PloS One*, *15*(8), e0237001. https://doi.org/10.1371/journal.pone.0237001

Sicorello, M., & Schmahl, C. (2020). Emotion dysregulation in borderline personality disorder: A fronto-limbic imbalance? *Current Opinion in Psychology*, *37*, 114–120. https://doi.org/10.1016/j.copsyc.2020.12.002

Sicorello, M., Thome, J., Herzog, J., & Schmahl, C. (2020). Differential effects of early adversity and PTSD on amygdala reactivity: The role of developmental timing. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. https://doi.org/10.1016/j.bpsc.2020.10.009

Steel, Z., Marnane, C., Iranpour, C., Chey, T., Jackson, J. W., Patel, V., & Silove, D. (2014). The global prevalence of common mental disorders: A systematic review and meta-analysis 1980-2013. *International Journal of Epidemiology*, *43*(2), 476–493. https://doi.org/10.1093/ije/dyu038

Swartz, J. R., Knodt, A. R., Radtke, S. R., & Hariri, A. R. (2015). A neural biomarker of psychological vulnerability to future life stress. *Neuron*, *85*(3), 505–511. https://doi.org/10.1016/j.neuron.2014.12.055

Todorov, A. (2012). The role of the amygdala in face perception and evaluation. *Motivation and Emotion*, *36*(1), 16–26. https://doi.org/10.1007/s11031-011-9238-5

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E.-J., van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., … Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*(1), 58–76. https://doi.org/10.3758/s13423-017-1323-7

Wager, T., Kang, T. D., Johnson, J. D., Nichols, T. D., Satpute, T. E., & Barrett, A. B. (2015). A bayesian model of category-specific emotional brain responses. *PLoS Computational Biology*, *11*(4), 1004066. https://doi.org/10.1371/journal.pcbi.1004066

Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., Charlson, F. J., Norman, R. E., Flaxman, A. D., Johns, N., Burstein, R., Murray, C. J. L., & Vos, T. (2013). Global burden of disease attributable to mental and substance use disorders: Findings from the Global Burden of Disease Study 2010. *The Lancet*, *382*(9904), 1575–1586. https://doi.org/10.1016/S0140-6736(13)61611-6

Winter, D., Niedtfeld, I., Schmitt, R., Bohus, M., Schmahl, C., & Herpertz, S. C. (2017). Neural correlates of distraction in borderline personality disorder before and after dialectical behavior therapy. *European Archives of Psychiatry and Clinical Neuroscience*, *267*(1), 51–62. https://doi.org/10.1007/s00406-016-0689-2

Woo, C. W., Chang, L. J., Lindquist, M. A., & Wager, T. D. (2017). Building better biomarkers: Brain models in translational neuroimaging. *Nature Neuroscience*, *20*(3), 365–377. https://doi.org/10.1038/nn.4478

Woo, C. W., Koban, L., Kross, E., Lindquist, M. A., Banich, M. T., Ruzic, L., Andrews-Hanna, J. R., & Wager, T. D. (2014). Separate neural representations for physical pain and social rejection. *Nature Communications*, *5*, 5380. https://doi.org/10.1038/ncomms6380

Yang, Z., Craddock, R. C., & Milham, M. P. (2014). Impact of hematocrit on measurements of the intrinsic brain. *Frontiers in Neuroscience*, *8*, 452. https://doi.org/10.3389/fnins.2014.00452

Yarkoni, T. (2020). The generalizability crisis. *The Behavioral and Brain Sciences*, 1–37. https://doi.org/10.1017/S0140525X20001685

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, *8*(8), 665–670. https://doi.org/10.1038/nmeth.1635

Table 1

Differences in neural pattern expression between negative and neutral condition

	Negative emotions				Positive emotions		Other emotions	
	Negative Affect	Fear	Anger	Sadness	Amusement	Contentment	Surprise	Neutral
Study 1	3.86 [3.13, 4.66] BF_{10} > 100	1.84 [1.42, 2.28] BF_{10} > 100	1.60 [1.22, 2.01] BF_{10} > 100	-0.17 [-0.43, 0.1] BF_{10} = 0.07	-1.39 [-1.76, -1.03] BF_{10} > 100	-2.85 [-3.46, -2.28] BF_{10} > 100	1.88 [1.45, 2.33] BF_{10} > 100	-1.76 [-2.19, -1.35] BF_{10} > 100
Study 2								
Event-related	1.45 [1.01, 1.91] BF_{10} > 100	1.26 [0.85, 1.69] BF_{10} > 100	0.49 [0.16, 0.83] BF_{10} = 20.57	-0.16 [-0.48, 0.16] BF_{10} = 0.09	-0.62 [-0.97, -0.28] BF_{10} > 100	-0.80 [-1.17, -0.44] BF_{10} > 100	1.10 [0.71, 1.51] BF_{10} > 100	-1.21 [-1.63, -0.80] BF_{10} > 100
Block	1.87 [1.36, 2.41] BF_{10} > 100	2.49 [1.88, 3.16] BF_{10} > 100	0.96 [0.59, 1.35] BF_{10} > 100	-0.32 [-0.65, 0.00] BF_{10} = 0.06	-2.28 [-2.91, -1.71] BF_{10} > 100	-0.82 [-1.19, -0.46] BF_{10} > 100	1.99 [1.47, 2.56] BF_{10} > 100	-1.47 [-1.94, -1.03] BF_{10} > 100
Mixed-Block	2.76 [2.10, 3.48] BF_{10} > 100	2.90 [2.21, 3.65] BF_{10} > 100	1.23 [0.82, 1.65] BF_{10} > 100	-0.32 [-0.65, 0.00] BF_{10} = 0.06	-2.84 [-3.58, -2.16] BF_{10} > 100	-1.19 [-1.61, -0.78] BF_{10} > 100	3.11 [2.38, 3.91] BF_{10} > 100	-2.66 [-3.36, -2.02] BF_{10} > 100
Study 3	1.23 [0.96, 1.5] BF_{10} > 100	0.77 [0.54, 1.00] BF_{10} > 100	0.37 [0.16, 0.58] BF_{10} = 79.45	-0.34 [-0.55, -0.13] BF_{10} = 0.03	-0.86 [-1.10, -0.62] BF_{10} > 100	-0.23 [-0.44, -0.03] 2.59	0.87 [0.63, 1.30] BF_{10} > 100	-0.51 [-0.73, -0.30] BF_{10} > 100
Mega-Analysis	2.07 [1.90, 2.23] BF_{10} > 100	1.38 [1.23, 1.53] BF_{10} > 100	0.81 [0.68, 0.94] BF_{10} > 100	-0.23 [-0.35, -0.11] BF_{10} = 0.02	-1.24 [-1.38, -1.10] BF_{10} > 100	-1.03 [-1.20, -0.86] BF_{10} > 100	1.42 [1.27, 1.56] BF_{10} > 100	-1.15 [-1.29, -1.00] BF_{10} > 100

Note. Estimates are Cohen's *d*. Numbers in brackets are 95% confidence intervals. BF_{10} = Bayes factor of the alternative hypothesis over the null hypothesis.

Table 2

Differences in neural pattern reactivity (negative – neutral condition) between emotion dysregulation and healthy control group

	Negative emotions				Positive emotions		Other emotions	
	Negative Affect	Fear	Anger	Sadness	Amusement	Contentment	Surprise	Neutral
Study 1	-0.09 [-0.61, 0.43] BF_{01} = 4.76	0.01 [-0.51, 0.53] BF_{01} = 3.57	-0.45 [-0.97, 0.08] BF_{01} = 9.09	0.13 [-0.39, 0.65] BF_{01} = 2.50	-0.45 [-0.97, 0.08] BF_{01} = 0.06	0.49 [-0.04, 1.02] BF_{01} = 10.0	0.16 [-0.37, 0.68] BF_{01} = 3.23	0.29 [-0.24, 0.81] BF_{01} = 7.14
Study 2								
Event-related	-0.37 [-0.99, 0.26] BF_{01} = 6.25	-0.36 [-0.98, 0.27] BF_{01} = 6.25	-0.20 [-0.82, 0.42] BF_{01} = 4.76	-0.39 [-1.01, 0.24] BF_{01} = 6.25	-0.08 [-0.70, 0.54] BF_{01} = 2.70	0.32 [-0.31, 0.94] BF_{01} = 5.88	0.15 [-0.47, 0.77] BF_{01} = 2.94	0.36 [-0.27, 0.98] BF_{01} = 6.25
Block	-0.99 [-1.64, -0.33] BF_{01} = 11.11	-0.95 [-1.60, -0.29] BF_{01} = 11.11	-0.34 [-0.96, 0.28] BF_{01} = 5.88	0.30 [-0.32, 0.93] BF_{01} = 1.41	0.68 [0.04, 1.31] BF_{01} = 9.09	0.20 [-0.43, 0.82] BF_{01} = 4.76	-0.23 [-0.85, 0.39] BF_{01} = 2.63	0.05 [-0.57, 0.67] BF_{01} = 3.57
Mixed-Block	-0.37 [-0.99, 0.26] BF_{01} = 6.25	-1.11 [-1.77, -0.44] BF_{01} = 12.5	-0.21 [-0.83, 0.41] BF_{01} = 5.00	0.18 [-0.44, 0.80] BF_{01} = 2.04	0.51 [-0.12, 1.14] BF_{01} = 7.14	0.17 [-0.45, 0.79] BF_{01} = 4.55	-0.01 [-0.63, 0.61] BF_{01} = 3.23	0.12 [-0.50, 0.74] BF_{01} = 4.17
Study 3	0.16 [-0.26, 0.58] BF_{01} = 2.27	0.21 [-0.21, 0.63] BF_{01} = 1.79	0.13 [-0.3, 0.55] BF_{01} = 2.70	0.06 [-0.36, 0.48] BF_{01} = 3.57	-0.11 [-0.53, 0.32] BF_{01} = 2.94	-0.10 [-0.52, 0.33] BF_{01} = 3.12	0.23 [-0.19, 0.65] BF_{01} = 2.70	-0.36 [-0.79, 0.06] BF_{01} = 0.69
Mega-Analysis	-0.13 [-0.38, 0.11] BF_{01} = 11.56	-0.13 [-0.39, 0.12] BF_{01} = 7.28	-0.16 [-0.39, 0.08] BF_{01} = 53.01	0.06 [-0.18, 0.31] BF_{01} = 9.54	-0.07 [-0.32, 0.17] BF_{01} = 4.79	0.17 [-0.08, 0.41] BF_{01} = 9.68	0.12 [-0.13, 0.37] BF_{01} = 2.15	0.01 [-0.24, 0.25] BF_{01} = 21.70

Note. Estimates are Cohen's *d*. Numbers in brackets are 95% confidence intervals. BF_{01} = Bayes factor of the null hypothesis over the alternative hypothesis.