Title Page

Cardiac interoceptive learning is modulated by emotional valence perceived from facial expressions

Running Title

Valence modulates interoceptive learning

Authors and Primary Affiliations:

Amanda C. Marshall[*1], Antje Gentsch[1], Lena Schröder[1], Simone Schütz-Bosbach[1]

[1] Department of Psychology, General and Experimental Psychology Unit, Ludwig-Maximilians University, D-80802 Munich, Germany

Corresponding Author Address:

Amanda Marshall

Ludwig-Maximilians-University Munich, General and Experimental Psychology

Leopoldstr. 13, D-80802 Munich, Germany

Phone: +49 (0) 89 / 2180 72525

Fax: + 49 (0) 89 / 2180 5211

Email: amanda.marshall@psy.lmu.de

Abstract

Interoception refers to the processing of homeostatic bodily signals. Research demonstrates that interoceptive markers can be modulated via exteroceptive stimuli and suggests that the emotional content of this information may produce distinct interoceptive outcomes. Here, we explored the impact of differently valenced exteroceptive information on the processing of interoceptive signals. Participants completed a repetition-suppression paradigm viewing repeating or alternating faces. In experiment 1, faces wore either angry or pained expressions to explore the interoceptive response to different types of negative stimuli in the observer. In experiment 2, expressions were happy or sad to compare interoceptive processing of positive and negative information. We measured the heartbeat evoked potential (HEP) and visual evoked potentials (VEP) as a respective marker of intero- and exteroceptive processing. We observed increased HEP amplitude to repeated sad and pained faces coupled with reduced HEP and VEP amplitude to repeated angry faces. No effects were observed for positive faces. However, we found a significant correlation between suppression of the HEP and VEP to repeating angry faces. Results highlight an effect of emotional expression on interoception and suggest an attentional trade-off between internal and external processing domains as a potential account of this phenomenon.

Keywords: attention; emotional valence; heartbeat evoked potential; interoception

Total: 4969 words

Introduction

We know from personal experience that different external situations elicit distinct feeling states. For example, while public speaking might produce a feeling of nervous excitement, a sad piece of music may evoke a feeling of melancholy. In both instances, the body responds to the situation with a palpable change of homeostatic signals; public speaking is associated with a significant increase of cardiovascular activity (Al'Absi et al., 1997; Kothgassner et al., 2016) while sad music reduces both skin conductance and heart rate (White & Rickard, 2015; Garrido, 2017). Internal bodily awareness arising from the processing of such autonomic signals is referred to as interoception (Garfinkel et al., 2015). Surprisingly, interoception has largely been treated as a closed system in scientific terms and studies have only recently begun addressing the interplay between exteroceptive and interoceptive domains. Work in this regard has linked interoceptive processing to visual perception. Enhanced cortical processing of the heartbeat signal has been shown to predict conscious perception of a visual stimulus (Park et al., 2014), while presenting a visual cue in tune with participants' heartbeat makes it harder to detect (Salomon et al., 2016). We recently explored the reverse relationship, namely whether exteroceptive material affected interoceptive processing. For this, we focused on the Heartbeat Evoked Potential (HEP) as an established marker of interoceptive processing (Pollatos, Kirsch & Schandry, 2005; Couto et al., 2015). The HEP component manifests with a fronto-central distribution around 200 to 400 ms after the R-wave in the electrocardiogram (Schandry & Montoya, 1996; Park et al., 2014). Increased HEP amplitude has been shown to coincide with higher heartrate detection accuracy (Pollatos & Schandry, 2004; Terhaar et al., 2012). Furthermore, source modelling of the component has linked its origin to the right insula and anterior cingulate cortex (Pollatos, Kirsch & Schandry, 2005; Park et al., 2017), brain structures implicated in the processing of

interoceptive signals (Park et al., 2014). In a previous study, we modulated the HEP by presenting neutral and angry facial expressions across trials in which they were either repeated or alternated. Examining the simultaneously recorded ECG signal, we found that neutral and angry expressions elicited distinctly different patterns of cardiac activity. Furthermore, presenting the same facial expression twice led to a significant change of HEP amplitude (Marshall et al., 2017). We interpreted this amplitude change as a reflection of top-down interoceptive learning which was enhanced by trials in which the same cardiac pattern was repeated. Crucially, our findings provided an indication that this interoceptive process may distinguish between differentially valenced stimuli. Our results showed that HEP modulation changed as a function of the presented facial expression; while neutral repetitions produced a significant increase of HEP amplitude, repeating negative expressions reduced HEP expression. Furthermore, HEP suppression to repeated negative faces correlated with suppression of a simultaneously recorded visual evoked potential linked to the processing of facial stimuli. This association between visual evoked potentials (VEP) and the HEP suggests a link between exteroceptive and interoceptive stimulus processing. Our work lends support to theories suggesting top-down mechanisms for interoceptive processing (Seth et al., 2012) and indicates that interoception is a dynamic process which processes and responds to exteroceptive stimulation. Moreover, it provides a promising indication that different types of exteroceptive information are treated differently by the interoceptive system and suggests a relationship between the processing of extero- and interoceptive stimulus material. Such findings would be the first to capture an underlying neural mechanism reflecting the distinct interoceptive states we experience in response to different environmental situations. In addition, they would approximate state of the art knowledge of exteroceptive stimulus processing for which the impact of valence has been

established and extensively studied. For example, Batty and Taylor (2003) explored the expression of visual evoked potentials to neutral or emotive facial expressions and observed that negative expressions elicited greater VEP activation compared to both positive and neutral ones. Similarly, Palomba, Angrilli and Mini (1997) explored heart rate and VEP responses to pleasant, unpleasant and neutral stimuli and reported that both pleasant and unpleasant pictures elicited a different cardiac pattern and greater VEP amplitude compared to neutral stimuli. Relative to neutral events, emotionally salient stimuli have also been shown to produce greater repetition suppression effects across visual areas measured by fMRI (Ishai et al., 2004) as well as VEPs captured via MEG (Ishai et al., 2006).

In this study, we aimed to explore whether differently valenced exteroceptive cues produce a distinct neural signature indexing differentiated interoceptive processing of environmental information. We conducted two experiments in which we explored whether the interoceptive response distinguished between different types of negative stimuli (experiment 1) and between positive and negative information (experiment 2). Across both experiments, participants were shown a repetition-suppression paradigm in which three types of facial expressions were either repeated or alternated (see Marshall et al., 2017). Both experiments included neutral facial expressions to capture the previously reported effect of stimulus repetition on HEP amplitude (Marshall et al., 2017). In experiment 1, we presented participants with angry and painful expressions. Both stimuli signal negative emotions. However, the subjective response in the observer is markedly different. Observing another's pain may elicit an emphatic response. Furthermore, pain results from an internal signal (e.g. conveyed by skin, muscles or an inner organ) and is thus a highly relevant signal for internal processing. Conversely, observing another's anger elicits defensive reactions and may constitute a less relevant cue for the processing of interoceptive states. We thus

hypothesized that painful faces would prime internal processing to a greater extent than angry faces. In experiment 2, we contrasted the interoceptive response to positive and negative stimuli by presenting happy or sad facial expressions. Relative to neutral expressions, both types of stimuli convey qualitative information about internal states. However, similar to painful expressions sadness is closely associated with an introspective focus on bodily states and feelings. Sad faces may thus likewise act as a more relevant trigger for interoceptive processing compared to happy expressions. Based on earlier findings, we predicted a change of HEP and VEP amplitude in repeated relative to alternated trials, as well as enhanced repetition effects for the VEP in response to emotive stimuli. We further predicted an effect of valence on HEP amplitude. For experiment 1, we expected enhanced HEP amplitude for painful expressions and decreased HEP amplitude for angry expressions relative to neutral faces. We further expected to replicate the correlation between the HEP and the VEP response to the face stimuli observed in our earlier study. For experiment 2, we expected elevated HEP amplitude to sad and happy expressions relative to neutral faces. We expected this elevation to be particularly pronounced for sad expressions.

Materials and Methods

Experiment 1

Participants.

Twenty-five participants (11 female, right-handed, mean age: 26.72 ± 4.63 years) with normal or corrected-to-normal vision took part in the study. All provided written informed consent and received payment or student credit for their participation. Procedures were approved by the ethics committee of the Ludwig-Maximilians University Munich in accordance with the Declaration of Helsinki (1991; p.1194). Depressive symptoms were assessed with the Beck Depression Inventory (BDI-II; Beck et al., 1961). Anxiety was

measured using the State-Trait Anxiety Inventory (STAI; Spielberger et al., 1983). Both conditions have been associated with altered interoceptive processing. We thus screened for any individuals scoring above the clinical cut-off point. However, all participants scored within the normal range. Sample size was determined with a power analysis. This indicated we had 80% power to detect the small to medium effect (Cohen's $\delta = 0.42$; $\alpha = 0.05$) of stimulus repetition on HEP amplitude observed in a previous study.

Stimuli.

Materials consisted of 10 actors (5 males/5 females; Caucasian ethnicity) portraying angry, neutral and pained facial expressions (Motreal Pain and Affective Face Clips database; Simon et al., 2008). Pre-validation of this stimulus set led us to exclude 2 actors (1 male/1 female; see supplementary material). The remaining 24 stimuli were used for the experiment.

Procedure.

The experiment began with a standard heartbeat tracking task (Schandry & Montoya, 1996). Participants reported the number of heartbeats they silently counted during three time periods (25, 35, 45 s) presented in random order. We explicitly discouraged guessing the number of heartbeats. Heartbeat tracking score was calculated using the following formula:

$$\frac{1}{3} \sum (1 - [(\textit{recorded heartbeats} - \textit{counted heartbeats}) \div \textit{recorded heartbeats}])$$

The subsequent experiment consisted of a repetition-suppression paradigm (see Fig. 1). After a training session (15 trials), participants completed 18 experimental blocks of 40 trials. We presented 3 types of blocks (6 of each). In block type 1 participants encountered neutral and pained facial expressions. Block type 2 contained neutral and angry faces and

block type 3 presented angry and pained facial expressions. We chose this design to reduce the number of different events within each block.

In each trial, the same face was presented twice for a duration of 500 ms, interspersed with a 500 ms fixation screen. A jittered inter-trial interval (1.5 – 2.5 s) separated each trial. The presented face wore either the same (repetition trials) or different (alternation trials) facial expression across both iterations. Expressions were counter-balanced within and across blocks so that each was presented with the same frequency and equally often in first and second position of the sequence. Participants' monitored the sequence for occasional arrows pointing to the left or right. Arrows were superimposed on the first or second face. Participants responded to their appearance by pressing a left or right button for which they received immediate feedback. These catch trials ensured participants' engagement and occurred on 20 % of all trials (balanced across conditions). They were discarded from later analyses. The experimental session ended after participants filled out the questionnaires.

ECG recording and processing.

The ECG signal was recorded at a sampling rate of 500 Hz from two bipolar electrodes placed below the left clavicle and the left pectoral muscle. ECG data was offline filtered between 1 – 40 Hz. R-peaks were detected using the EEGLAB plugin FMRIB 1.21 (Niazy et al., 2005). For the control analysis (see supplementary materials) we extracted R-peak amplitude, heart period power and interbeat-intervals (between R-peaks) using the open source R-HRV package implemented in R (Rodríguez-Liñares et al., 2008).

EEG recording and processing.

Continuous EEG signals were recorded at a sampling rate of 500 Hz using a 64-channel active electrode system (actiCAP, Brain products GmbH, Gilching, Germany).

Offline EEG data were pre-processed in EEGLAB (EEGLAB 9.0.3, University of San Diego, San Diego, CA) and BrainVision Analyzer (BrainVision Analyzer 2.0, Brain products GmbH). In EEGLAB, the continuous EEG signal was filtered between 0.1 and 40 Hz and re-referenced to a common average reference. Independent component analysis was conducted on the signal to determine stereotypical components reflecting eye movements, blinks and the cardiac field artefact (CFA). These were removed based on the visual inspection of 40 independent components. Components relating to the CFA as well as eye movements and blinks were characterized by a time course (projection matrix) and a scalp map (weighting matrix). We identified ICs relating to the CFA by searching for components displaying the bimodal topography commonly associated with this artefact. In addition, we searched for a frequency peak around 5 Hz and a rhythmically repeating time course. We removed 2 – 4 components per participants (average of 1.72 related to the CFA across participants). For the HEP, data was segmented into 1000 ms periods relative to the onset of the inter-trial interval marker. Within this post-stimulus interval, epochs were further segmented into periods ranging from -100 to 600 ms relative to the R-peak marker as obtained from the continuously recorded ECG. For the VEP, data was segmented from -100 to 500 ms relative to the onset of the second facial stimulus. Participants completed 120 trials per condition. Artefact correction led to an average trial rejection of 13%, leaving an average of 104 epochs per condition (minimum 99). No difference in retained epochs was observed across conditions for either component (all $p_s > .05$). VEPs and HEPs were calculated by averaging across trials for each condition using the -100 ms interval prior to stimulus onset or R-peak marker for baseline correction. A current-source-density (CSD) transformation was applied to HEP epochs to reduce potentially remaining contamination of HEPs by residual CFA overlap.

Statistical analysis.

We employed a permutation-based approach to determine the morphology (latency and topography) of event-related components. Our hypothesis concerned the way heartbeat and visual evoked potentials responded to differently valenced stimuli. Therefore, we created a new set of difference values by subtracting alternation from repetition trials (Sel, Azevedo & Tsakiris, 2017). Based on the findings of Marshall and colleagues (2017) we then compared angry against neutral conditions. We proceeded to identify neural phenomena that varied with this effect of valence and calculated associated point-estimate statistics (F-values) across the time window ranging from -100 to 600 ms from the onset of the R-peak marker in the inter-trial-interval for the HEP and -100 to 500 ms from the onset of the second facial stimulus for the VEP. Time windows were averaged into 100 ms windows (i.e. 100 ms $\sim$ 200 ms etc.) prior to the permutation analysis. We then permuted the dataset by shuffling across conditions and subjects and re-computing the statistics 1000 times, providing a null distribution corresponding to each time point. Across each permutation, the maximum F-value was logged, providing a distribution of maximal values obtained under H_0. We then compared original point-estimates to this distribution, choosing the values that fell into or above the 95^{th} percentile as significant candidates for subsequent analysis. We determined ERP topography in the same manner. For this analysis all 64 electrodes were treated as a distinct variable.

We used the electrodes and time windows identified via this bottom-up approach for the subsequent main analysis. For this, we averaged across all electrodes exhibiting an effect of valence for the time window in which this effect reached statistical significance. For each condition, we hereby created a single variable which reflected the amplitude of HEP and VEP across spatial and temporal points exhibiting a statistically robust effect of valence. We

chose an analysis of variance approach for the main analysis which allowed us to move beyond the binary comparison afforded by the permutation test. We therefore submitted difference values (repetition – alternation) to a one-way (valence: angry vs. painful vs. neutral) repeated measures ANOVA. Bonferroni-corrected paired t-tests were used for follow-up comparisons. In addition to the standard frequentist approach, we reported Bayes Factors for all primary analyses. Bayesian Repeated Measures ANOVAs were conducted in JASP (Love et al., 2016) adopting the default prior settings (fixed effects = 0.5; random effects = 1; covariates = 0.354; auto sampling).

Results

Behavioural and questionnaire data.

Participants responded accurately to 65.68 ± 14.1% of catch trials. Paired t-tests comparing reaction times on response trials (mean = 433 ± 64 ms) did not differ between experimental manipulations (all $p_s > 0.44$; $BF_{10} = 0.87$). Questionnaire results (BDI = 6.7 ± 6.4; STAI trait = 40.72 ± 12.5; STAI state = 39.3 ± 12.1) correspond to previously reported student samples (Paulus & Stein, 2010). However, we observed relatively low scores on the heartbeat tracking task (0.48 ± 0.29). We attributed this to the stringent instructions we gave to our participants in which we discouraged participants to estimate heartbeats or name beats they had not explicitly felt.

Visual evoked response.

The permutation test returned an effect of valence between 100 – 200 ms after the onset of the facial stimulus. This effect manifested over right parietal-occipital electrodes (Pz, POz, P2, PO4; Cohen's $\delta > 0.2$). Using the average of this electrode pool for the analysis of variance calculation returned a main effect of valence $F_{2,48} = 5.35$, $p = .008$, $\eta^2 = 0.19$ ($BF_{10} = 32.88$). Bonferroni corrected paired-tests of this effect revealed a significant

difference between VEP expression to angry and neutral faces (t_{24} = 2.51, p = .029; BF_{10} = 21.03; *Mean change score* = 0.91) and between angry and pained faces (t_{24} = 3.61, p = .001; BF_{10} = 26.57; *Mean change score* = 1.45). No difference was observed between neutral and pained expressions (t_{24} = 1.04 p = .46; BF_{10} = 0.13; *Mean change score* = 0.54). Results thus demonstrate that VEP expression to repeated angry faces differs significantly to repetition effects observed for both neutral and pained expressions. While repeated angry faces produced suppression of the VEP, repeated neutral and pained faces produce a slight elevation of VEP amplitude (see Figure 2).

Heartbeat evoked potential.

The permutation test found an effect of valence between 200 – 300 ms after the R-wave peak over frontal-central electrodes (F1, Fz, F2, FC1, FCz, FC2; $\delta > 0.4$) which corresponds to previous observations of the HEP component (Schandry & Montoya, 1996; Marshall et al., 2017). The analysis of variance calculation of this effect observed a main effect of valence $F_{1,48}$ = 7.91, p = .001, η^2 = 0.25 (BF_{10} = 27.63). Paired t-tests used to follow up this effect once again found a significant difference between angry and neutral faces (t_{24} = 2.43, p = .029; BF_{10} = 13.47; *Mean change score* = 1.66) as well as angry and pained faces (t_{24} = 4.62, p < .001; BF_{10} = 38.66; *Mean change score* = 2.73). No difference emerged between neutral and pained faces (t_{24} = 1.33, p = .24; BF_{10} = 0.57; *Mean change score* = 1.07). Repetition effects for the heartbeat evoked potential also differ significantly from angry to pained and neutral stimuli. Repetition of angry faces produces strong suppression of HEP amplitude which significantly differs from the repetition enhancement of the HEP occurring in response to neutral and painful faces *(*see Figure 3*)*. Furthermore, results revealed a significant correlation between repetition suppression of the early VEP and subsequent HEP to angry repeated faces ($\boldsymbol{\rho}$ = 0.39, p = 0.018; *BF_{10} = 61.12;* see Figure 4). No

significant correlations were observed between VEP and HEP expression for either neutral (ρ = 0.16, p = 0.45; $BF_{10} = 0.004$) or pained stimulus repetitions (ρ = 0.04, p = 0.86; $BF_{10} = 0.007$). This finding suggests an association between the neural response indexing the processing of intero- and exteroceptive information specifically for angry facial expressions. We also explored whether HEP effects correlated with explicit interoceptive accuracy (scores on the heartbeat detection task). We found no evidence for this when correlating scores with HEP amplitude collapsed across all conditions (ρ = -0.17, p = 0.39; $BF_{10} = 0.09$). This corresponds to previous work (Park et al., 2014) and suggests that HEP modulation reflects a transient state of interoceptive processing rather than persistent interoceptive accuracy.

Control analysis.

We conducted extensive control analysis for both experiments (see supplementary material). For both experiments, these analyses demonstrated that facial expressions elicited different patterns of cardiac activity. Our paradigm rests on the assumption that repeating trials iterate highly similar cardiac patterns evoked by presenting the same facial expression twice. Relative to alternating trials in which it remains unpredictable, the heartbeat signal can thus be approximated and processed more efficiently. For this to hold true, both types of expressions must elicit a different cardiac signal. Findings thus fulfil the baseline requirement of our paradigm. However, cardiac differences did not persist for the later time window in which the HEP was measured. Finally, a surrogate R-peak analysis demonstrated that the HEP was locked to processing the heartbeat signal rather than other changes in the EEG.

Experiment 2

In experiment 2 we compared the interoceptive response to positive and negative exteroceptive stimuli by presenting happy and sad expressions alongside neutral faces within the same repetition-suppression framework used for experiment 1.

Participants.

We recruited twenty-five participants (15 females; all right-handed, mean age: 25.6 ± 4.96 years) with normal or corrected-to-normal vision. A previous power analysis indicated we had 80% power to detect the small to medium effect (Cohen's $\delta = 0.44$; $\alpha = 0.05$) of stimulus repetition on HEP discovered in experiment 1.

Stimuli.

Materials for the paradigm came from the Radboud Faces Database (Langner et al., 2010). The initial stimulus set consisted of 39 young adults (18 females, 18 males; Caucasian ethnicity) modelling a happy, sad or neutral expression. Based on a pre-validation of the stimulus set (see supplementary material) we chose 8 actors (4 female/4 male) for the subsequent experiment.

Procedure.

We followed the same procedure reported for experiment 1. For the repetition-suppression paradigm participants viewed happy, sad and neutral faces presented in the same block and trial structure reported for experiment 1.

EEG processing and statistical analysis.

We employed the same EEG/ECG set up, data pre-processing and statistical approach to determine HEP latencies and topographies reported for experiment 1. For experiment 2, our permutation test compared the difference scores (repetition – alternation) between sad and neutral conditions. The average of statistically valid time points and electrodes was

subsequently entered into an analysis of variance calculation to compare all three levels of our valence variable (repetition – alternation trials for happy, sad and neutral conditions).

Results

Behavioural and questionnaire data.

Participants responded accurately to 68.3 ± 14% of catch trials (mean reaction time 459 ± 49 ms). Paired t-tests observed no differences in reaction times across the different experimental conditions (all $p_s > 0.34$; $BF_{10} = 0.77$). Participants mean heartbeat perception score (0.52 ± 0.27), as well as their scores on the STAI (state: 35.96 ± 7.3; trait: 38.44 ± 8.8) and BDI (5.2 ± 5.0) compared to values obtained in experiment 1.

Visual evoked response.

Results of the permutation test returned no effects of valence across any electrode pools or latencies (all p_s > than α .05 cut-off value; see Figure 5).

Heartbeat evoked potential.

Similar to experiment 1, the permutation test revealed an effect of valence between 200 – 300 ms after the R-wave peak over frontal-central electrodes (F1, Fz, F2, FC1, FCz, FC2). The analysis of variance calculation across all three levels of the effect likewise found a significant effect of valence $F_{2,48} = 7.68$, $p = .001$, $\eta^2 = 0.24$ ($BF_{10} = 22.78$). Bonferroni corrected follow-up t-tests found a significant difference between sad and neutral facial expressions ($t_{24} = 2.42$, $p = .034$, $BF_{10} = 19.83$; *Mean change score* = 1.14) and between sad and happy expressions ($t_{24} = 3.92$, $p < .001$, $BF_{10} = 25.29$; *Mean change score* = 2.26). No significant difference emerged between happy and neutral facial expressions ($t_{24} = 1.41$, $p = .22$; $BF_{10} = 0.60$; *Mean change score* = 1.12). Results demonstrate that repetition effects for sad facial expressions significantly differ from those to neutral and happy faces. Repeated sad faces produce a strong elevation of HEP amplitude which significantly differs from the

marginal HEP elevation produced by repeated neutral and happy faces (see Figure 6). The correlation between absolute HEP amplitude and heartbeat tracking score failed to reach significance ($\boldsymbol{\rho}$ = -0.12, p = 0.5; $BF_{10} = 0.93$).

Discussion

In this study, we explored the effects of emotional valence on interoceptive processing. In experiment 1, we presented angry and painful facial expressions to explore the impact of different adverse contexts. In experiment 2, participants viewed happy and sad faces to test the difference between positive and negative contexts. We observed significant enhancement of the HEP for repeating painful and sad faces. Although we found no significant difference between pained and neutral faces in experiment 1, this effect was less pronounced for repeated neutral and happy faces. In addition, we found a significant reduction of HEP and VEP amplitude to repeated angry faces. Furthermore, we found a significant correlation between repetition suppression of the HEP and VEP in response to repeated angry faces. Results extend our earlier work showing HEP modulation by repeating exteroceptive events (Marshall et al., 2017). Here, we interpreted HEP modulation as a marker of top-down interoceptive learning. Our simultaneously recorded ECG signal demonstrated that different emotional expressions elicited distinct patterns of cardiac activity. While alternation trials therefore produced different heartbeat signals across the first and second stimulus presentation within a trial, repetition trials (presenting the same facial expression twice) repeated highly similar cardiac patterns. We thus suggest that enhancement of the HEP in repeated trials reflects the construction of interoceptive templates enabled by the repeating heartbeat signal. This interpretation rests on findings within the exteroceptive domain where repetition enhancement of sensory potentials is commonly characterized as a reflection of internal templates subsequently used to make

predictions about upcoming stimuli (Henson, Shallice & Dolan, 2000; Turk-Browne, Leber & Chun, 2007; Müller et al., 2012). Interpreting HEP modulation in this framework corresponds to reports linking higher HEP amplitude to enhanced cardiac processing (Pollatos & Schandry, 2004; Terhaar et al., 2012) and interoceptive cardiac learning (Canales-Johnson et al., 2015). Our current dataset highlights an effect of emotional valence on this interoceptive learning process. Results hereby approximate findings in the exteroceptive domain (Batty & Taylor, 2003; Ishai et al., 2006) as well as studies reporting a modulating impact of valence on other internal, pre-reflective forms of bodily self-awareness such as agency (Gentsch et al., 2015; Yoshi & Haggard, 2017).

For visual evoked potentials, an effect of valence manifested only for repeated angry expressions. However, reduced VEP amplitude for this stimulus corresponds to previous reports suggesting repetition-suppression of early visual components as an indication of more efficient visual processing (Recasens et al., 2014), particularly in response to angry faces known to capture and hold attention (Koster et al., 2004). Crucially, we replicated the significant correlation between interoceptive and exteroceptive measures in response to repeated angry faces which suggests that reduced interoceptive processing coincided with a more efficient exteroceptive response to this stimulus (Marshall et al., 2017).

This association suggests attentional focus as a potential mechanism underlying the effect of valence on interoceptive processing, particularly during observation of angry facial expressions. Attention is an established modulator of repetition effects, known to produce greater repetition suppression of attended relative to unattended visual stimuli (Summerfield & Egner, 2009; Escera et al., 2003; Opitz et al., 2002). Attention may therefore similarly modulate interoceptive learning for emotive stimuli. In this respect, painful and sad contexts may prime interoceptive focus as they arise from internal signals and suggest

introspective states. Conversely, angry faces may direct attention to the external environment by suggesting emotional states directed outwards (e.g. at an opponent or a frustrating situation). Further, sad or painful contexts may generate an empathic response which induces a strong interoceptive focus via bodily resonance mechanisms (Lamm, Decety & Singer, 2011). These may counteract exteroceptive attentional modulation. Repetition enhancement of the HEP for painful and sad faces may thus be a function of increased attention to homeostatic signals, while repetition suppression of the HEP for angry faces may result from priming increased attention to the exteroceptive domain. Attentional allocation may thus facilitate an interoceptive response and learning for sad and painful stimuli while impeding interoceptive learning for angry faces. However, this interpretation rests on the assumption that suppression of the HEP and VEP towards angry faces are two distinct processes (i.e. suppression signifies reduced interoceptive learning for the HEP while suggesting more efficient perceptual processing for the VEP). In support of this hypothesis, suppression of perceptual components is commonly interpreted as more efficient neural processing (Wiggs & Martin, 1998). Furthermore, VEP suppression occurs in an early time window commonly attributed to low-level perceptual processes (Recasens et al., 2014) while HEP suppression occurs at a later time which is generally associated with higher-order enhancement effects. We would further argue that it is not parsimonious to offer two conflicting accounts of HEP amplitude. Thus, a more cogent explanation is achieved by framing both types of HEP expression in terms of interoceptive learning which can either be facilitated or reduced by differently valenced exteroceptive cues.

We observed similar HEP expressions to painful and sad stimuli. This corresponds to past work highlighting a close relationship between the experience of both states (Bingel et al., 2006; Tracey & Mantyn, 2007). For example, Yoshino and colleagues (2010) reported

that sadness increased participants' sensitivity to a subsequent painful stimulus. The association between both states is further demonstrated by the terminology commonly used to describe sadness which is often referred to as an altered bodily condition related to pain (i.e. a broken heart). Theoretical accounts suggest this is a result of sadness evolving onto a pre-existing pain system which means both experiences share neural and computational mechanisms primarily involving the amygdala and the anterior cingulate cortex (Eisenberger & Libermann, 2004; Wager et al., 2004). The observed similarity between interoceptive cortical processing evoked by sad and painful stimuli thus corresponds to the shared processing architecture suggested by past accounts.

Finally, we wish to highlight some study limitations and future directions which could extend this work. Our study did not measure participants' empathy levels. Neither did it assess the propensity of presented faces to evoke empathetic responses. Given the potential contribution of empathetic mechanisms to our observed effects, future work would benefit from including such measures. Relatedly, we are unable to ascertain whether the observed findings apply only to emotions observed from others or whether they generalize to one's own, self-experienced emotions. Future work could thus explore HEP expression in response to inducing different types of emotions in participants. Finally, we found no effect of positive valence on interoceptive processing. An explanation could be that happy faces, like neutral expressions do not elicit a strong interoceptive focus. Future work in this domain would thus benefit from further investigation of HEP amplitude to positive emotions of different intensities and types to test whether the impact of valence on interoception extends beyond adverse contexts.

In conclusion, we report an effect of emotional expression on interoceptive processing and suggest a potential mechanism underlying the effect in the form of

attentional weighting between intero- and exteroceptive domains. Results hereby emphasize the interaction between intero- and exteroceptive sensory processing and are to the authors' knowledge the first to capture a neural proxy corresponding to the distinct interoceptive states we experience in response to different environmental situations.

Acknowledgements

This work was supported by Deutsche Forschungsgemeinschaft [SCHU 2471/5-1 to S.B.].

Figure Legends

Figure 1. Examples of the emotive facial stimuli used across both experiments as well as the time course of the paradigm. Faces were presented in pairs wearing either the same (repeated trials) or different (alternating trials) facial expressions across both iterations. In 20% of trials a red arrow (</>) appeared superimposed on the first or second face. In these catch trials, participants were required to press a button corresponding to the arrow's direction as quickly as possible.

Figure 2. Left: Amplitude of VEP difference scores (repetition – alternation trials) in response to viewing neutral, angry or pained facial expressions in Experiment 1. Positive waveforms indicate higher VEP amplitudes in repeated relative to alternated trials while negative waveforms indicate repetition suppression of the VEP in repeated relative to alternated trials. Right: Box plots highlighting the significant suppression of VEP amplitude to repeating angry faces (whiskers represent standard deviations).

Figure 3. Left: Amplitude of HEP difference scores (repetition – alternation trials) in response to pained, angry and neutral facial expressions in Experiment 1. Positive waveforms highlight higher amplitude in repeated relative to alternated trials. Negative waveforms highlight

reduced amplitude in repeated relative to alternated trials. Right: Box plots showing the interaction between the three levels of valence: HEP suppression to repeated angry expressions significantly differs from HEP elevation to neutral and pained faces (whiskers represent standard deviation).

Figure 4. Correlation between HEP and VEP repetition suppression (across 120 trials for each participant) in response to angry repeated faces. Reduced VEP amplitude in response to the second, repeated face (at 1100 ms of the trial sequence) significantly correlated with subsequent HEP amplitude suppression (at 1700 ms of the trial sequence).

Figure 5. Left: Amplitude of VEP difference scores (repetition – alternation trials) in response to viewing sad, happy and neutral facial expressions in Experiment 2. Positive waveforms highlight higher amplitude in repeated relative to alternated trials. Negative waveforms highlight reduced amplitude in repeated relative to alternated trials. Right: Box plots showing no difference in VEP repetition effects between the three levels of valence (whiskers represent standard deviation).

Figure 6. Left: Amplitude of HEP difference scores (repetition – alternation trials) in response to viewing sad, happy and neutral facial expressions in Experiment 2. Positive waveforms highlight higher amplitude in repeated relative to alternated trials. Negative waveforms highlight reduced amplitude in repeated relative to alternated trials. Right: Box plots highlighting the interactions between the three levels of valence: HEP elevation in repeated trials with sad expressions differs significantly from HEP elevation in neutral and happy repetition trials (whiskers represent standard deviation).

References

Al'Absi, M., Bongard, S., Buchanan, T., Pincomb, G. A., Licinio, J., & Lovallo, W. R. (1997). Cardiovascular and neuroendocrine adjustment to public speaking and mental arithmetic stressors. Psychophysiology, 34(3), 266-275.

Batty, M., & Taylor, M. J. (2003). Early processing of the six basic facial emotional expressions. *Cognitive Brain Research*, *17*(3), 613-620.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., Erbaugh, J. An inventory for measuring depression. Archives of general psychiatry. 4(6), 561-571 (1961).

Bingel, U., Lorenz, J., Schoell, E., Weiller, C., & Büchel, C. (2006). Mechanisms of placebo analgesia: rACC recruitment of a subcortical antinociceptive network. *Pain*, *120*(1-2), 8-15.

Canales-Johnson, A., Silva, C., Huepe, D., Rivera-Rei, Á., Noreika, V., Garcia, M. D. C., ... & Couto, B. (2015). Auditory feedback differentially modulates behavioral and neural markers of objective and subjective performance when tapping to your heartbeat. Cerebral Cortex, 25(11), 4490-4503.

Couto, B., Adolfi, F., Velasquez, M., Mesow, M., Feinstein, J., Canales-Johnson, A., ... & Manes, F. (2015). Heart evoked potential triggers brain responses to natural affective scenes: a preliminary study. Autonomic Neuroscience, 193, 132-137.

Escera, C., Yago, E., Corral, M. J., Corbera, S., & Nuñez, M. I. (2003). Attention capture by auditory significant stimuli: semantic analysis follows attention switching. European Journal of Neuroscience, 18(8), 2408-2412.

Garrido, S. (2017). Physiological Effects of Sad Music. In Why Are We Attracted to Sad Music? (pp. 51-66). Springer International Publishing.

Garfinkel, S. N., Seth, A. K., Barrett, A. B., Suzuki, K., & Critchley, H. D. (2015). Knowing your own heart: distinguishing interoceptive accuracy from interoceptive awareness. Biological psychology, 104, 65-74.

Gentsch, A., Weiss, C., Spengler, S., Synofzik, M., & Schütz-Bosbach, S. (2015). Doing good or bad: How interactions between action and emotion expectations shape the sense of agency. Social neuroscience, 10(4), 418-430.

Henson, R., Shallice, T., & Dolan, R. (2000). Neuroimaging evidence for dissociable forms of repetition priming. Science, 287(5456), 1269-1272.

Ishai, A., Pessoa, L., Bikle, P. C., & Ungerleider, L. G. (2004). Repetition suppression of faces is modulated by emotion. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(26), 9827-9832.

Ishai, A., Bikle, P. C., & Ungerleider, L. G. (2006). Temporal dynamics of face repetition suppression. *Brain Research Bulletin*, *70*(4-6), 289-295.

Koster, E. H., Crombez, G., Van Damme, S., Verschuere, B., & De Houwer, J. (2004). Does imminent threat capture and hold attention?. Emotion, 4(3), 312.

Kothgassner, O. D., Felnhofer, A., Hlavacs, H., Beutl, L., Palme, R., Kryspin-Exner, I., & Glenk, L. M. (2016). Salivary cortisol and cardiovascular reactivity to a public speaking task in a virtual and real-life environment. Computers in Human Behavior, 62, 124-135.

Lamm, C., Decety, J., & Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage*, *54*(3), 2492-2502.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H.J., Hawk, S.T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. Cognition & Emotion, 24(8), 1377—1388.

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., Wagenmakers, E. J. JASP (Version 0.8)[computer software]. Amsterdam, the Netherlands: Jasp project (2016).

Marshall, A. C., Gentsch, A., Jelinčić, V., & Schütz-Bosbach, S. (2017). Exteroceptive expectations modulate interoceptive processing: repetition-suppression effects for visual and heartbeat evoked potentials. Scientific reports, 7(1), 16525.

Müller, N. G., Strumpf, H., Scholz, M., Baier, B., & Melloni, L. (2012). Repetition suppression versus enhancement—it's quantity that matters. Cerebral cortex, 23(2), 315-322.

Niazy, R. K., Beckmann, C. F., Iannetti, G. D., Brady, J. M., & Smith, S. M. (2005). Removal of FMRI environment artifacts from EEG data using optimal basis sets. Neuroimage, 28(3), 720-737.

Opitz, B., Rinne, T., Mecklinger, A., Von Cramon, D. Y., & Schröger, E. (2002). Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. Neuroimage, 15(1), 167-174.

Palomba, D., Angrilli, A., & Mini, A. (1997). Visual evoked potentials, heart rate responses and memory to emotional pictorial stimuli. *International journal of psychophysiology*, *27*(1), 55-67.

Park, H. D., Correia, S., Ducorps, A., & Tallon-Baudry, C. (2014). Spontaneous fluctuations in neural responses to heartbeats predict visual detection. Nature neuroscience, 17(4), 612-618.

Park, H. D., Bernasconi, F., Salomon, R., Tallon-Baudry, C., Spinelli, L., Seeck, M., ... & Blanke, O. (2017). Neural sources and underlying mechanisms of neural responses to heartbeats, and their role in bodily self-consciousness: An intracranial EEG study. *Cerebral Cortex*, 1-14.

Paulus, M. P., & Stein, M. B. (2010). Interoception in anxiety and depression. Brain structure and Function, 214(5-6), 451-463.

Pollatos, O., & Schandry, R. (2004). Accuracy of heartbeat perception is reflected in the amplitude of the heartbeat-evoked brain potential. Psychophysiology, 41(3), 476-482.

Pollatos, O., Kirsch, W. & Schandry, R. (2005). Brain structures involved in interoceptive awareness and cardioafferent signal processing: a dipole source localization study. Human brain mapping. 26(1), 54-64.

Recasens, M., Leung, S., Grimm, S., Nowak, R., & Escera, C. (2015). Repetition suppression and repetition enhancement underlie auditory memory-trace formation in the human brain: An MEG study. Neuroimage, 108, 75-86.

Rodríguez-Liñares, L., Vila, X., Mendez, A., Lado, M., & Olivieri, D. (2008). RHRV: An R-based software package for heart rate variability analysis of ECG recordings. In 3rd Iberian Conference in Systems and Information Technologies (CISTI 2008),pp. 565-574.

Salomon, R., Ronchi, R., Dönz, J., Bello-Ruiz, J., Herbelin, B., Martet, R., ... & Blanke, O. (2016). The insula mediates access to awareness of visual stimuli presented synchronously to the heartbeat. Journal of Neuroscience, 36(18), 5115-5127.

Schandry, R., & Montoya, P. (1996). Event-related brain potentials and the processing of cardiac activity. Biological psychology, 42(1), 75-85.

Sel, A., Azevedo, R. T., & Tsakiris, M. (2017). Heartfelt self: cardio-visual integration affects self-face recognition and interoceptive cortical processing. Cerebral Cortex, 1-12.

Seth, A. K., Suzuki, K., & Critchley, H. D. (2012). An interoceptive predictive coding model of conscious presence. Frontiers in psychology, 2, 395.

Simon, D., Craig, K. D., Gosselin, F., Belin, P., & Rainville, P. (2008). Recognition and discrimination of prototypical dynamic expressions of pain and emotions. PAIN, 135(1), 55-64.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., Jacobs, G. A. Manual for the state-trait anxiety inventory (Palo Alto, CA, Consulting Psychologists Press, 1983).

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. Trends in cognitive sciences, 13(9), 403-409.

Terhaar, J., Viola, F. C., Bär, K. J. & Debener, S. (2012). Heartbeat evoked potentials mirror altered body perception in depressed patients. Clinical Neurophysiology. 123(10), 1950-1957.

Tracey, I., & Mantyh, P. W. (2007). The cerebral signature for pain perception and its modulation. *Neuron*, *55*(3), 377-391.

Turk-Browne, N. B., Yi, D. J., Leber, A. B., & Chun, M. M. (2006). Visual quality determines the direction of neural repetition effects. Cerebral Cortex, 17(2), 425-433.

Wager, T. D., Rilling, J. K., Smith, E. E., Sokolik, A., Casey, K. L., Davidson, R. J., ... & Cohen, J. D. (2004). Placebo-induced changes in FMRI in the anticipation and experience of pain. *Science*, *303*(5661), 1162-1167.

White, E. L., & Rickard, N. S. (2016). Emotion response and regulation to “happy” and “sad” music stimuli: Partial synchronization of subjective and physiological responses. Musicae Scientiae, 20(1), 11-25.

Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. Current opinion in neurobiology, 8(2), 227-233.

Yoshie, M., & Haggard, P. (2017). Effects of emotional valence on sense of agency require a predictive model. Scientific Reports, 7(1), 8733.

Yoshino, A., Okamoto, Y., Onoda, K., Yoshimura, S., Kunisato, Y., Demoto, Y., ... & Yamawaki, S. (2010). Sadness enhances the experience of pain via neural activation in the anterior cingulate cortex and amygdala: an fMRI study. *Neuroimage*, *50*(3), 1194-1201.

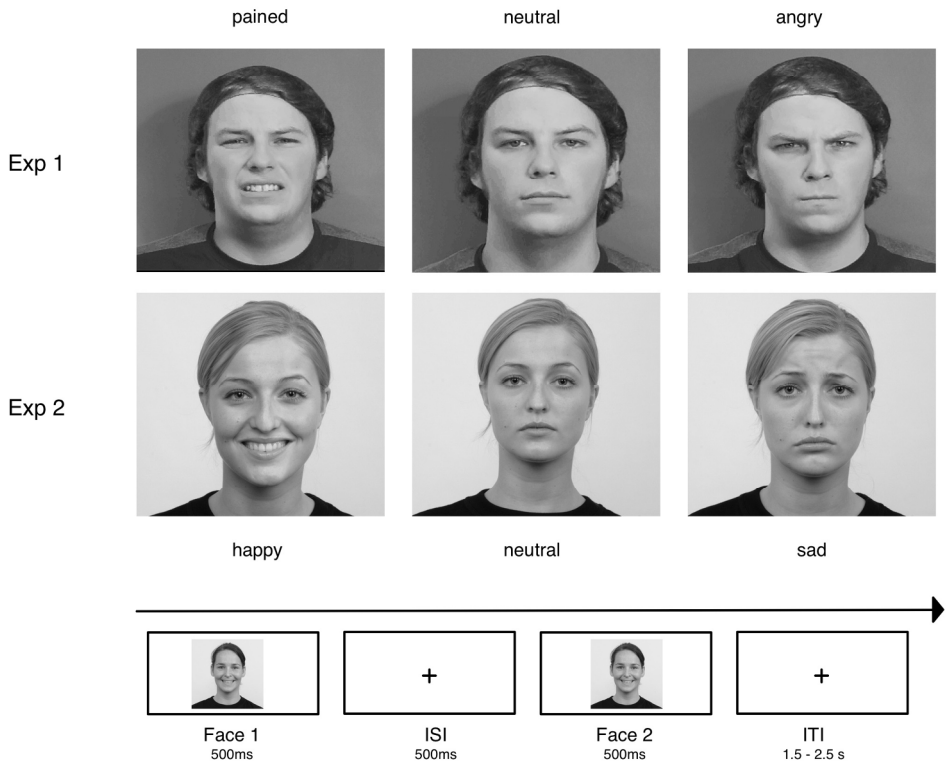

Figure 1. Examples of the emotive facial stimuli used across both experiments as well as the time course of the paradigm. Faces were presented in pairs wearing either the same (repeated trials) or different (alternating trials) facial expressions across both iterations. In 20% of trials a red arrow (</>) appeared superimposed on the first or second face. In these catch trials, participants were required to press a button corresponding to the arrow's direction as quickly as possible.

757x623mm (72 x 72 DPI)

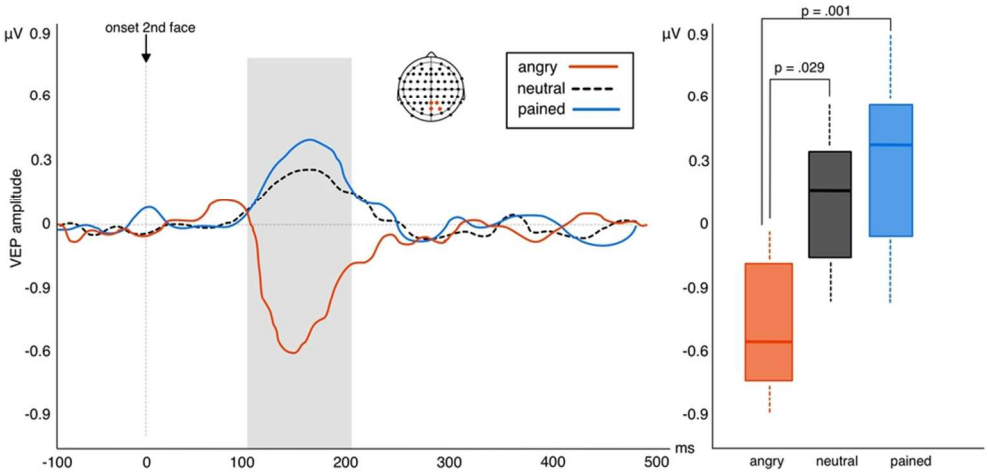

Figure 2. Left: Amplitude of VEP difference scores (repetition – alternation trials) in response to viewing neutral, angry or pained facial expressions in Experiment 1. Positive waveforms indicate higher VEP amplitudes in repeated relative to alternated trials while negative waveforms indicate repetition suppression of the VEP in repeated relative to alternated trials. Right: Box plots highlighting the significant suppression of VEP amplitude to repeating angry faces (whiskers represent standard deviations).

84x42mm (300 x 300 DPI)

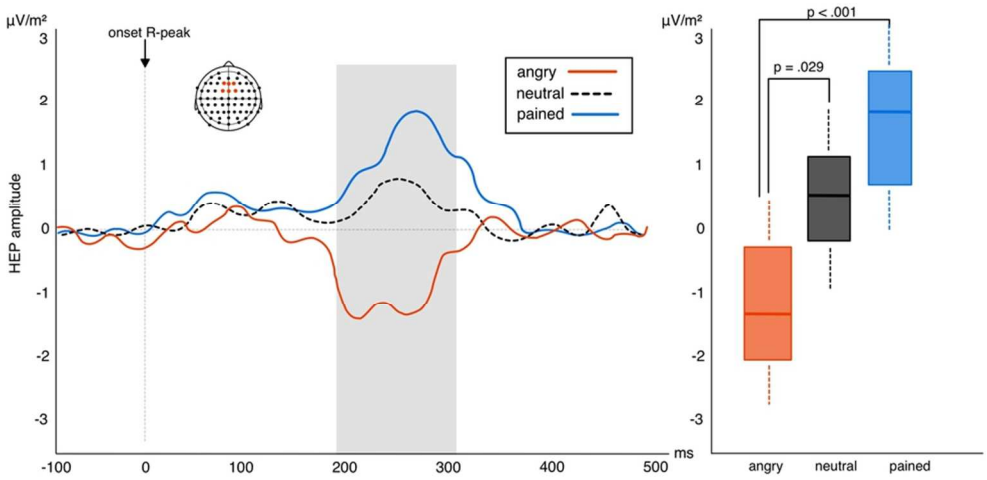

Figure 3. Left: Amplitude of HEP difference scores (repetition – alternation trials) in response to pained, angry and neutral facial expressions in Experiment 1. Positive waveforms highlight higher amplitude in repeated relative to alternated trials. Negative waveforms highlight reduced amplitude in repeated relative to alternated trials. Right: Box plots showing the interaction between the three levels of valence: HEP suppression to repeated angry expressions significantly differs from HEP elevation to neutral and pained faces (whiskers represent standard deviation).

84x42mm (300 x 300 DPI)

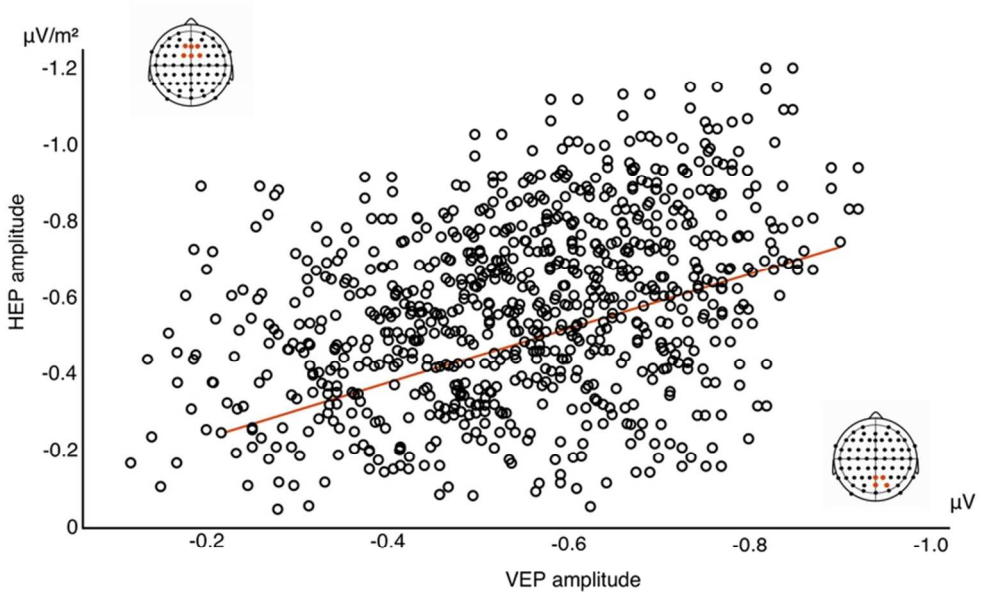

Figure 4. Correlation between HEP and VEP repetition suppression (across 120 trials for each participant) in response to angry repeated faces. Reduced VEP amplitude in response to the second, repeated face (at 1100 ms of the trial sequence) significantly correlated with subsequent HEP amplitude suppression (at 1700 ms of the trial sequence).

73x47mm (300 x 300 DPI)

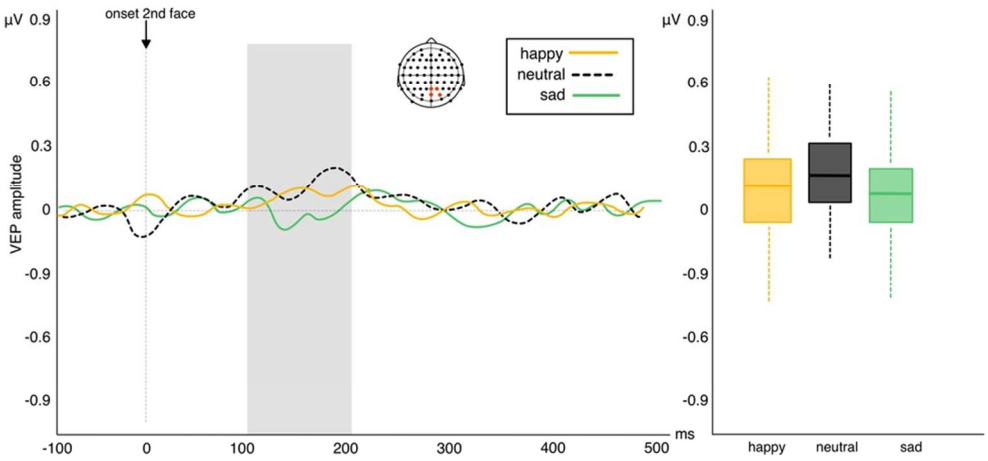

Figure 5. Left: Amplitude of VEP difference scores (repetition – alternation trials) in response to viewing sad, happy and neutral facial expressions in Experiment 2. Positive waveforms highlight higher amplitude in repeated relative to alternated trials. Negative waveforms highlight reduced amplitude in repeated relative to alternated trials. Right: Box plots showing no difference in VEP repetition effects between the three levels of valence (whiskers represent standard deviation).

84x42mm (300 x 300 DPI)

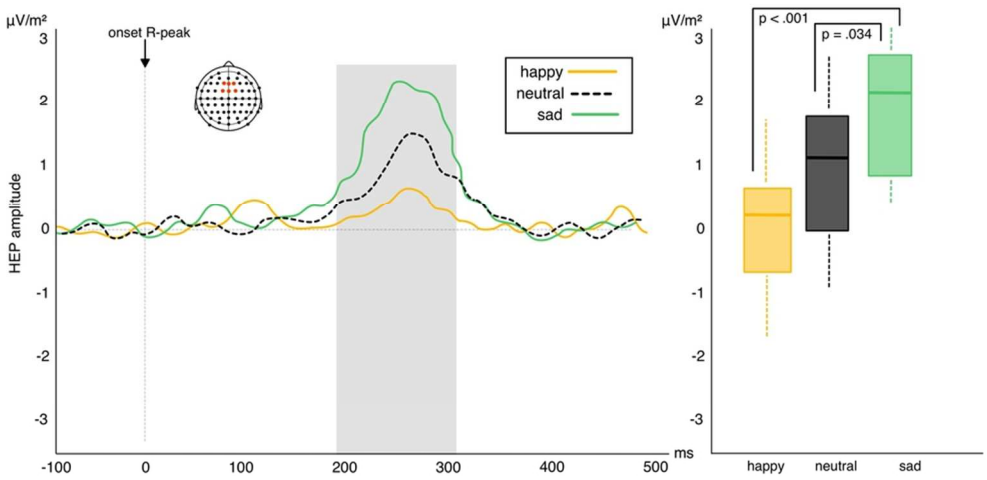

Figure 6. Left: Amplitude of HEP difference scores (repetition – alternation trials) in response to viewing sad, happy and neutral facial expressions in Experiment 2. Positive waveforms highlight higher amplitude in repeated relative to alternated trials. Negative waveforms highlight reduced amplitude in repeated relative to alternated trials. Right: Box plots highlighting the interactions between the three levels of valence: HEP elevation in repeated trials with sad expressions differs significantly from HEP elevation in neutral and happy repetition trials (whiskers represent standard deviation).

84x42mm (300 x 300 DPI)