

A distributed brain response predicting the facial expression of acute nociceptive pain

Reviewed Preprint

v2 • August 29, 2024

Revised by authors

Reviewed Preprint

v1 • October 4, 2023

Marie-Eve Picard , Miriam Kunz, Jen-I Chen, Michel-Pierre Coll, Étienne Vachon-Pressseau, Tor D Wager, Pierre Rainville

Department of psychology, Université de Montréal, Montreal, Quebec, Canada • Centre de recherche de l'institut universitaire de gériatrie de Montréal, Montreal, Quebec, Canada • Department of medical psychology and sociology, Medical faculty, University of Augsburg, Augsburg, Germany • School of psychology, Université Laval, Quebec, Quebec, Canada • Faculty of Dentistry, McGill University, Montreal, Quebec, Canada • Department of Anesthesia, McGill University, Montreal, Quebec, Canada • Alan Edwards Centre for Research on Pain, McGill University, Montreal, Quebec, Canada • Department of psychological and brain sciences, Dartmouth College, Hanover, New Hampshire, United States • Stomatology department, Faculté de médecine dentaire, Université de Montréal, Montreal, Quebec, Canada

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

Abstract

Pain is a private experience observable through various verbal and non-verbal behavioural manifestations, each of which may relate to different pain-related functions. Despite the importance of understanding the cerebral mechanisms underlying those manifestations, there is currently limited knowledge on the neural correlates of the facial expression of pain. In this functional magnetic resonance imaging (fMRI) study, noxious heat stimulation was applied in healthy volunteers and we tested if previously published brain signatures of pain were sensitive to pain expression. We then applied a multivariate pattern analysis to the fMRI data to predict the facial expression of pain. Results revealed the inability of previously developed pain neurosignatures to predict the facial expression of pain. We thus propose a Facial Expression of Pain Signature (FEPS) conveying distinctive information about the brain response to nociceptive stimulations with minimal or no overlap with other pain-relevant brain signatures associated with nociception, pain ratings, thermal pain aversiveness, or pain valuation. The FEPS may provide a distinctive functional characterization of the distributed cerebral response to nociceptive pain associated with the socio-communicative role of non-verbal pain expression. This underscores the complexity of pain phenomenology by reinforcing the view that neurosignatures conceived as biomarkers must be interpreted in relation to the specific pain manifestation(s) predicted and their underlying function(s). Future studies should explore other pain-relevant manifestations and assess the specificity of the FEPS against simulated pain expressions and other types of aversive or emotional states.

eLife assessment

Picard et al. propose a Facial Expression Pain Signature (FEPS) derived from functional magnetic resonance imaging (fMRI) data to predict facial expressions associated with painful heat stimulation. This **important** work advances our understanding of the brain mechanisms associated with facial expressions of pain. It provides **solid** evidence that facial expressions of pain contain information that is complementary to other pain-related brain processes. The work will be of broad interest to researchers from varied fields ranging from neurosciences to psychology and affective sciences.

<https://doi.org/10.7554/eLife.87962.2.sa3>

Introduction

Theories of pain communication highlight the diversity in pain manifestations, which occur through multiple channels: verbal reports, vocal complaints, change of postures, and facial expressions. From an evolutionary perspective, several manifestations of pain appear to be preserved across vertebrate phyla and reflect various functional roles to preserve the integrity of the organism (Sneddon 2019 [↗](#)). Withdrawal behaviour allows the individual to move away from the noxious source, while facial expressions in social species convey information about the presence of a potential threat and an appeal for assistance (Hadjistavropoulos et al. 2011 [↗](#)). Functionally distinct manifestations imply at least partly segregated neurophysiological processing (Sliwa et al. 2022 [↗](#)). Previous functional magnetic resonance imaging (fMRI) studies investigating the neural correlates of acute pain have suggested that spontaneous or induced fluctuations in pain facial expression partly reflect changes in activity within the cortical targets of the spino-thalamo-cortical pathways. Kunz and colleagues (2011) [↗](#) found a positive association between facial responses to pain and the activity in the posterior insula, the primary somatosensory area, and the anterior cingulate cortex. These fluctuations are, however, independent from changes in stimulus intensity and are inversely related to activity in prefrontal regions (Kunz et al. 2011 [↗](#); Vachon-Preseu et al. 2016 [↗](#); Kunz, Chen, and Rainville 2020 [↗](#)). This suggests that pain facial expression may reflect the integration of activity across distributed brain networks processing ascending nociceptive signals, determining action policy, and gating efferent facial motor outputs (see Kunz et al. 2011 [↗](#) and Kunz, Chen, and Rainville 2020 [↗](#) for further discussion).

The interest in developing pain neuromarkers has led researchers to use multivariate pattern analysis to investigate the distributed brain mechanisms underlying the experience of pain evoked by acute nociceptive stimuli. However, functional magnetic resonance imaging (fMRI) studies have revealed not one but several brain signatures of acute experimental pain that may reflect the diversity and complexity of pain-related function. The Neurological Pain Signature (NPS; Wager et al. 2013 [↗](#)) was developed to predict changes in pain induced by variations in stimulus intensity and captured by subjective reports, reflecting primarily the cerebral contributions to acute nociceptive pain (Krishnan et al. 2016 [↗](#); Wager et al. 2013 [↗](#)). To account for spontaneous fluctuations in the perception of pain intensity, the Stimulus-Independent Intensity of Pain Signature (SIIPS-1) was trained on noxious thermal trials after statistically removing the effects of the stimulus intensity and the NPS response (Woo et al. 2017 [↗](#)). More recently, the affective dimension of pain has received more attention, resulting in a multivariate pattern predictive of negative affect ratings to thermal pain, referred here as the Thermal Pain Aversive Signature (TPAS; Ćeko et al. 2022 [↗](#)). Finally, a signature was elaborated to characterise the neuronal

representations associated with the valuation of pain (PVP) in the context of a decision task involving a cost-benefit analysis of future pain against a monetary reward (Coll et al. 2022). Taken together, those signatures have contributed to improve our understanding of the neurobiological mechanisms of pain, as reflected in self-report or explicit decision making.

Facial expression has been used as a reliable behavioural measure of pain across different mammal species (Costa et al. 2014; Craig 1992; Evangelista et al. 2019; Langford et al. 2010; Sotocina et al. 2011), but few studies have investigated the brain mechanisms associated with the spontaneous non-verbal communication of pain in human (Kunz et al. 2011; Kunz, Chen, and Rainville 2020). As an automatic behavioral manifestation, pain facial expression is considered to be an indicator of activity in nociceptive systems, and to reflect perceptual and affective-evaluative processes. Here, we assessed the association between pain facial expression and the available pain-relevant brain signatures and we applied multivariate analysis with machine learning techniques to develop a predictive brain activation model of the facial responses to pain.

Results and Discussion

The Facial Action Coding System (FACS; Ekman and Friesen 1978) was used to quantify the facial expression of pain in healthy participants while brain responses evoked by brief moderately painful heat stimulation were recorded using fMRI. For each trial, the intensity and the frequency of pain-related action units were scored and combined into a FACS composite score (Methods). The association with the NPS, the SIIPS-1, the PVP, and the TPAS was assessed across the whole brain using the correlation between the FACS scores and the dot product computed between each signature and the activation maps for each individual trial.

Pain facial expression was not significantly associated with NPS expression ($\text{pearson-}r = .06$; $p = .20$; 95% CI = $[-.03, .14]$), TPAS expression ($\text{pearson-}r = .05$; $p = .26$; 95% CI = $[-.04, .14]$), PVP expression ($\text{pearson-}r = .02$; $p = .67$) and SIIPS-1 expression ($\text{pearson-}r = .07$; $p = .10$; 95% CI = $[-.01, .16]$). These low values indicate that the available pain-relevant brain signatures show poor sensitivity to the facial expression of pain. This motivated the development of a new multivariate brain pattern to predict pain expression.

We used a multivariate approach at the voxel level across the whole brain to develop the FEPS (see Methods). A LASSO principal component regression was applied to predict the FACS composite scores from the trial-by-trial fMRI activation maps. The FEPS was able to predict the FACS composite scores with a performance significantly above chance level (averaged cross-validated prediction across 10 folds: $\text{pearson-}r = .54 \pm .10$ (95% CI = $[.39, .64]$); $R^2 = .22 \pm .08$ (95% CI = $[-.09, .33]$); RMSE = 0.99 ± 0.08 (95% CI = $[0.88, 1.10]$); $p < .001$ compared to a permuted distribution; **Figure 1**-AB; also see the scatterplot and the ideal regression line across all data points in **Figure S3-A**). These results indicate that we were able to develop a multivariate brain pattern accounting for some variance in the facial responses related to pain.

The distributed pattern of activity predicting pain expression was projected back on the brain to examine the spatial distribution of higher weights contributing to the prediction (**Figure 1C**). Positive weight clusters were found in the primary motor cortex (M1; bilateral), the frontal pole, the right posterior parietal cortex and the dorsal part of the parietal operculum, adjacent to the secondary somatosensory cortex (S2) (**Table S1**). These regions are respectively associated with motor control, reward and affective value, attentional processes, and nociceptive pain processing (Price 2000; Rushworth et al. 2011; Shackman et al. 2011). Regions showing negative weights included the dorso-lateral PFC (dlPFC), the ventro-lateral PFC (vlPFC), the mid cingulate cortex (MCC), the subgenual ACC, the ventral part of the parietal operculum, the precuneus, and the vmPFC (**Table S2**). Negative weights imply that increased activity in those regions is

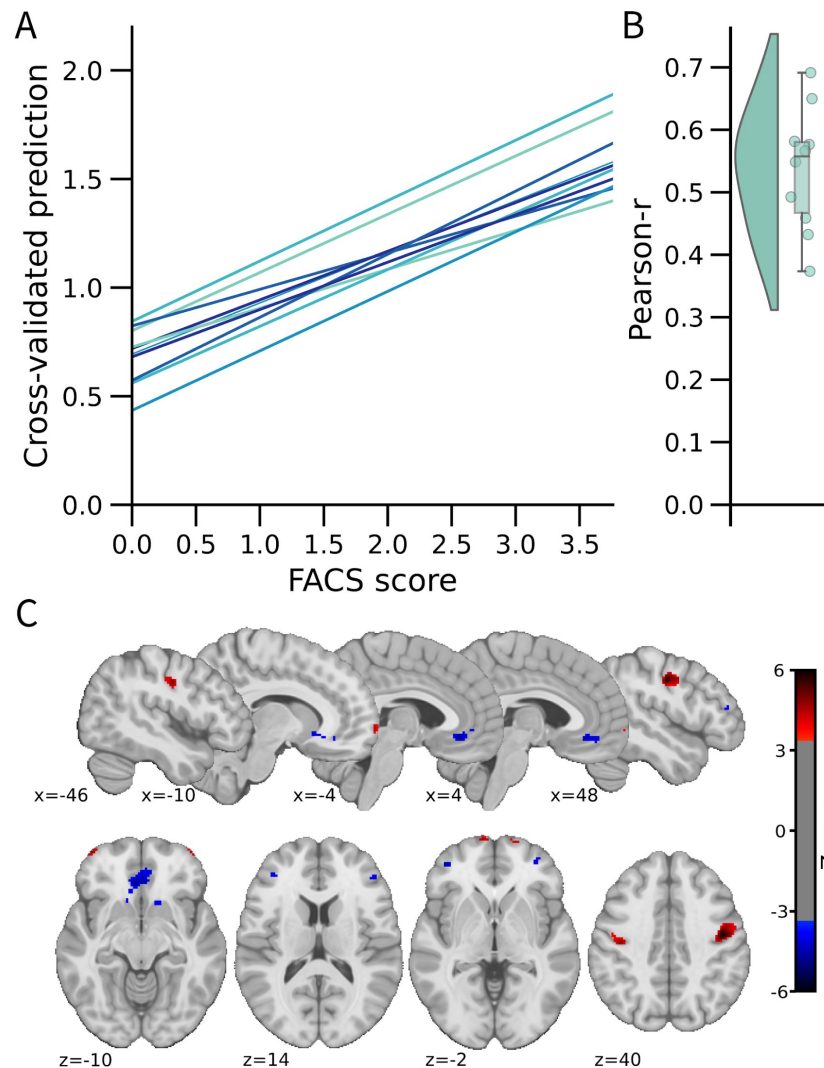


Figure 1

FEPS: a brain signature of the facial expression of pain.

A) Relationship between the actual and the predicted FACS composite scores for each cross-validation fold. B) Distribution of the Pearson's r scores across the cross-validation folds. C) Predictive weight map of pain expression thresholded at FDR corrected $q < .05$ using bootstrap tests performed with 5000 samples (with replacement). The thresholded map is shown for visualization and interpretation purposes only, although the prediction was made using voxel weights across the whole brain. MNI coordinates of the clusters with the related z-score can be found in [Tables S1](#) and [S2](#). The colour bar represents the z-scored regression weights reflecting the positive and negative association with the magnitude of the FACS composite score of pain expression.

associated with decreased facial response, consistent with a role in the inhibition of pain expression (Kunz et al. 2011 [↗](#)). The contribution of the dlPFC and the vlPFC to the model's prediction aligns with the role of these regions in inhibitory control and cognitive regulation, respectively (Goldin et al. 2008 [↗](#)).

A supplementary analysis was conducted to evaluate whether the activity pattern in the primary motor cortex (M1) alone could be sufficient for predicting facial expressions. The choice of this particular region was informed by prior research indicating that M1 presented the strongest correlation with the facial expression of pain (Kunz et al., 2011 [↗](#)). If the facial expression of pain primarily reflected a motor component without providing substantial insights into the pain experience, then the activity of the motor cortex alone should have been equally effective as the activity of the whole brain in predicting the FACS scores. This M1 model did lead to a significant prediction of pain facial expression (see **Figure S2** [↗](#)), but the whole brain model was significantly better ($t(532) = 2.73, p = .003$).

These results are consistent with the distributed nature of brain activity associated with the production and regulation pain facial expression reflecting in part the ascending nociceptive response and the ensuing affective processes, as well as top-down socio-affective regulation underlying the implementation of learned display rules (Karmann et al. 2016 [↗](#)).

This study was not designed to assess the specificity of the FEPS against other aversive states but the warm stimulation condition allowed us to compare the FEPS response to a painful vs a non-painful thermal stimulation. The FEPS expression score was computed on the activation maps of the warm trials and compared to the pain trials, with and without pain facial expression (**Figure S3-B** [↗](#)). The results indicated higher expression scores in the painful condition than in the warm condition (Pain - Warm contrast: $EMM = 0.24 \pm 0.02, t(1034) = 10.33, p < .0001, 95\% CI = [0.20, 0.29]$, Cohen's $d = 0.63$). Given the intra-individual and inter-individual variability in facial responses to pain, we repeated this comparison, stratifying the trials within the pain condition based on FACS scores: null FACS scores (FACS=0) and non-null FACS scores (FACS>0). The FEPS scores were larger in the pain condition where facial responses were displayed, compared to both the pain condition without facial expression and the warm condition (Pain_{FACS>0} - Pain_{FACS=0} contrast: $EMM = 0.58 \pm 0.03, t(1065) = 17.19, p < .0001, 95\% CI = [0.50, 0.66]$, Cohen's $d = 1.69$; Pain_{FACS=0} - Warm contrast: $EMM = -0.06 \pm 0.03, t(1057) = -2.24, p = .07, 95\% CI = [-0.13, 0.003]$, Cohen's $d = -0.18$; Pain_{FACS>0} - Warm contrast: $EMM = 0.52 \pm 0.03, t(1055) = 19.64, p < .0001, 95\% CI = [0.46, 0.58]$, Cohen's $d = 1.52$). Note that the comparison of Pain_{FACS>0} vs Pain_{FACS=0} is redundant with the regression approach used as the primary analysis model (**Figure 1** [↗](#)) and should not be considered as additional evidence. The observation that the Pain_{FACS=0} trials did not differ significantly from the Warm trials and that both conditions showed a mean score close to 0 (**Figure S3-B** [↗](#)) indicate that the FEPS does not respond to innocuous thermal stimuli and only responds to noxious heat when a facial expression is produced.

Several regions identified in the FEPS have also been reported in other pain-related brain signatures. Regions predictive of pain facial expression and pain intensity (NPS and SIIPS-1) include S2, the vmPFC, and the precuneus. The vlPFC is a region that does not receive direct spino-thalamo-cortical nociceptive afferents, and was reported both in the FEPS and in the SIIPS-1. Overlap between the FEPS and the PVP (Pain Value Pattern) include regions associated with reward and affect (i.e. OFC). Finally, the primary motor cortex, and S2 were also reported as contributing regions in the TPAS. The weak positive association between both the SIIPS-1 and the facial expression in addition with the spatial comparison showing some common regions across these pain-relevant signatures suggest possible shared features with the FEPS.

We computed the cosine similarity between the FEPS and other pain-related brain signatures to further examine the shared and specific representations between those predictive patterns (Methods). Cosine similarity ranging from 0.00 to 0.10 was found between the FEPS and the other

pain-related brain signatures reflecting the overall low similarity between the signatures at the whole-brain level (**Figure 2A**). The highest similarity value with the FEPS were found for the SIIPS-1, consistent with the notion that the facial expression of pain may reflect, at least partly, changes in brain responses associated with spontaneous fluctuations in pain experience captured by pain ratings. Similarity with the FEPS was further assessed across different cortical networks (**Figure 2B**). The significant positive similarity with the SIIPS-1 at the frontoparietal level and in the default-mode network may suggest common mechanisms in self-representation, prediction and emotional regulation of the pain experience that would be reflected in both facial expression and subjective reports (Pan et al. 2018). Recruitment of the frontoparietal network may also be involved in the conscious representation of the pain context, making nociceptive information available for integration into decision-making processes (Coll et al. 2022; Bastuji et al. 2016; Cul, Baillet, and Dehaene 2007; Zheng et al. 2020). The similarity between the FEPS and the SIIPS-1 in the somatomotor network indicates potential overlaps between the sensory aspect of the pain experience captured by the facial expression and the pain intensity ratings. This is consistent with our previous report showing that changes in pain facial expression by the cognitive modulation of perceived pain intensity is correlated to changes in the nociceptive response of the somatosensory cortex (Kunz, Chen & Rainville, 2020). Finally, the convergent similarities in the limbic network with the PVP is consistent with a key role of affective pain processing influencing facial expression, and perceived pain value (Garcia-Larrea and Peyron 2013; Roy et al. 2009).

In research and clinical practice, verbal reports of perceived pain intensity are considered to be the gold standard for measuring pain. Other measures that are often weakly correlated with those subjective reports, like facial expressions of pain, are often considered a less valid metric of the experience of pain even though they provide important complementary information on pain-related processes (Hadjistavropoulos et al. 2011). The Facial Expression of Pain Signature (FEPS) was able to predict the magnitude of the facial expression of pain above chance level. Regions that contribute to the prediction include motor and pain-related areas associated with both sensory and affective processing of pain. Although it shares, to some extent, similar representations with other pain-relevant signatures within various cerebral networks, the FEPS is distinctive from these other signatures. Results of this study provide unique evidence of the complementary information provided by facial expression on pain-related brain processes. Future studies must provide a more comprehensive account of diverse pain manifestations and their related function to better capture the pain phenomenon in its entirety.

Limitations

We recognize that our study has several limitations. First, given our limited sample size, further research will be necessary to verify the generalizability of the FEPS across other samples, but also across diverse experimental conditions (e.g. electrical, mechanical, and chemical pain) and populations (e.g. young vs old, chronic pain). Conducting future generalizability studies is crucial to ensure that the FEPS is a valid signature and is not only a result of model overfitting.

Even though the model developed from the entire brain activity could predict pain facial expression scores (FACS scores) beyond chance levels, it is important to highlight its inability to accurately predict the higher facial expression scores. This observation may be explained by the positive asymmetry in the distribution of facial expression scores, despite the log transformation applied to mitigate the observed skewness in the behavioral data. It is possible that the brain signature of pain facial expression might not adequately capture the heterogeneity of facial expressiveness in the population, especially for highly expressive individuals (Kunz et al., 2008). To address this limitation in the generalizability of the model, it could be re-trained with more observations associated with high FACS scores, thereby improving the extraction of predictive

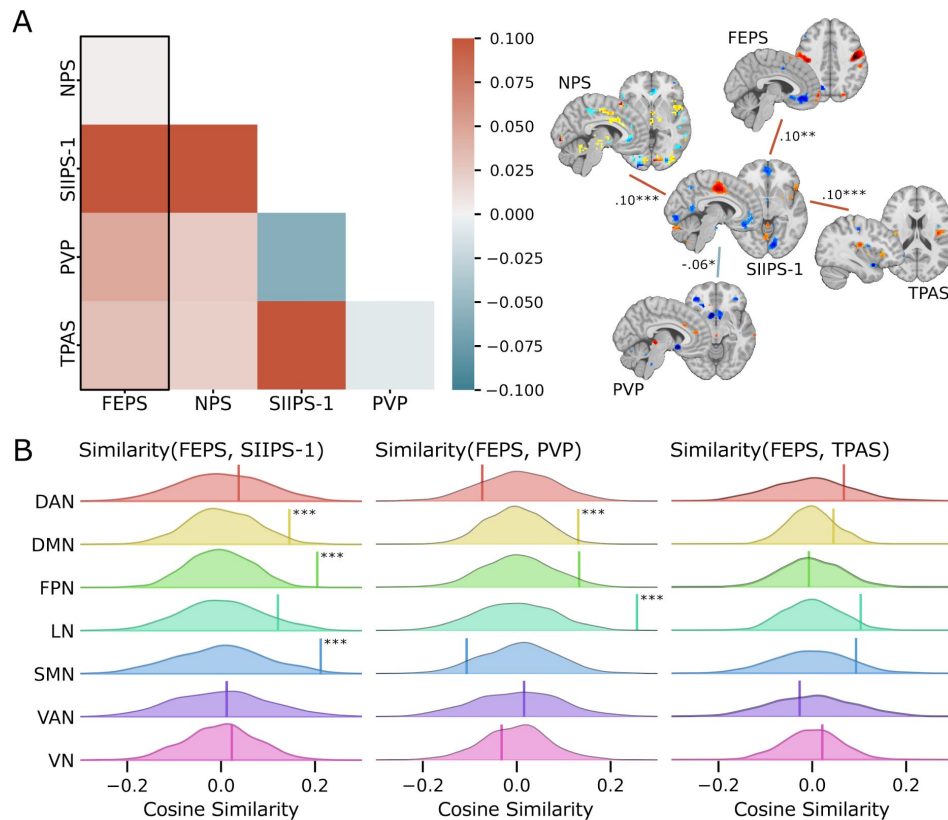


Figure 2

Spatial similarity between the FEPS and other pain-related signatures.

A) Pattern similarity between the FEPS and other pain-related brain signatures using the weights of the full brain patterns. Pattern similarities were computed at the voxel level using the cosine similarity; a value of 1 reflects proportional patterns; a value of 0 reflects orthogonal patterns; a value of -1 reflects patterns of opposite direction. The left panel shows the similarity matrix, and the right panel shows only the significant similarities between the pain-related signatures (* $p < .05$; ** $p < .01$; *** $p < .001$). B) Deconstructing the pattern similarity with regards to seven cortical networks as defined in the Yeo atlas²⁴: Visual Network (VN); Somatomotor Network (SMN); Dorsal Attention Network (DAN); Ventral Attention Network (VAN); Limbic Network (LN); Frontoparietal Network (FPN); Default Mode Network (DMN). Null distributions computed using permutation tests are shown, and the actual similarity values are represented by the vertical bar. Significant similarity values were found in the FPN (similarity = .20; $p = .002$), the SMN (similarity = .21; $p = .02$), and the DMN (similarity = .15; $p = .04$) for the SIIPS-1, in the LN (similarity = .26; $p = .001$), and DMN (similarity = .13; $p = .03$) for the PVP.

features associated with a greater level of facial expressivity. It is also possible that non-linear models might provide a better prediction and/or that the higher pain expressiveness might engage additional brain mechanisms not captured here.

The whole-brain model also overestimated the lowest facial responses with the intercept of the regression lines being systematically greater than 0 (see **Figure 1A** [↗](#)). This means that even when no facial expression was detected, there was still some brain activation matching the FEPS. Again, this may reflect a limitation of the linear model applied here. It is also possible that the FACS method might miss subtle movements of the face and/or that the FEPS captures meaningful variability in the pain-related brain activity below the threshold of facial expression.

It is also essential to assess the specificity of the FEPS in future studies. This involves examining whether the FEPS responds specifically to facial expressions of pain rather than broadly reflecting any facial movements including simulated facial expressions of pain and facial expression of emotions. Despite the similarities between facial expressions of pain and those associated with negative emotions, it is possible to behaviorally distinguish them. This suggests the potential to identify distinct brain patterns predictive of different facial responses ([Simon et al. 2008](#) [↗](#)). However, to our knowledge, there are no available brain imaging datasets to assess the specificity of the FEPS in that context.

Due to our limited sample size, we were not able to analyze each pain-related facial action unit separately or to explore different combinations. This could be valuable, especially considering the reported differences in the association between diverse pain-related action units and the sensory and affective components of pain ([Kunz et al., 2012](#); [Kunz, Chen & Rainville, 2020](#) [↗](#)).

Despite the limitations of this study, the evidence provided by our findings highlight the importance of facial expression as a complementary source of information to assess central nociceptive processes and acute pain. These results should be regarded as a benchmark for future research on non-verbal pain-related manifestations and may provide a foundation for the assessment of brain mechanisms underlying non-verbal communication across domains.

Methods

Participants

Secondary analyses of brain imaging data acquired in 34 healthy participants were performed in this study (18 women, 16 men; age mean $\pm$ SD = 23.4 $\pm$ 2.5 years) ([Kunz et al. 2011](#) [↗](#)). No participants reported having a history of neurological or psychiatric disease, nor chronic pain. Participants reported not using alcohol nor analgesic for at least 24h before the experimental session. All participants provided written informed consent and received monetary compensation for their participation. The information about the video recording of the face was indicated in the consent form but this was not emphasized and it was not mentioned at the time of data acquisition. All procedures were approved by the ethics committee of the Centre de recherche de l'institut universitaire de g riatrie de Montr al.

Study design

Pre-experimental session

Participants were submitted to a pre-experimental session to assess the range of thermal pain sensitivity using a magnitude estimation procedure. All participants included in this study had normal thermal pain sensitivity (see [Kunz et al., 2008](#) for more details regarding the procedure;

Duncan et al., 1992). The degree of facial expressiveness was also evaluated (low/non expressive (n=13): facial responses in 20% or less of painful trials; expressive (n=21): facial responses in more than 20% of painful trials; Kunz et al., 2011). Our sample was representative of the interindividual variability in facial expressivity of pain as reported in the literature (Kunz et al., 2006, 2008) and data from all participants were used in the present study.

Thermal stimuli

Thermal stimulations were induced using a Medoc TSA-2001 thermosensory stimulator with a MRI-compatible 3 x 3 cm² contact probe (Medoc), positioned at the level of the lower left leg. Thermal stimuli lasted 9 seconds (2s ramp up, 5s plateau at targeted temperature, 2s ramp down) and were followed by an interstimulus interval of 18-25 seconds. The experiment was programmed using E-prime software (Psychology Software Tools, Pittsburgh, PA). A baseline temperature of 38°C was applied between stimuli for all participants. The target temperatures were determined individually before the MRI scans to induce a warm non-painful sensation in the control condition, and a moderate to strong self-reported pain intensity (targeting 75-80/100 on the pain scale, where 50/100 corresponds to pain threshold; temperature (mean±SD) = 47.8±0.90 °C). Participants were not aware that the warm and painful temperatures remained constant across trials. The order of the control condition and experimental condition was pseudorandomized. There were 8 trials for each experimental condition per run, for a total of 16 trials per condition and a total of 544 trials per condition across all participants (34 participants x 16 pain trials). After each stimulus, participants rated the warm or pain sensation by moving a cursor with an MRI-compatible response key on a computerized visual analog scale.

Facial expression

The facial expression of the participants was recorded using an MRI-compatible camera (MRC Systems) mounted onto the head coil of the MRI scanner. To be able to quantify facial expressions that occurred during the stimulation, a signal was sent from the stimulator to the sound card to automatically mark the onset of each stimulus on the video recording. Two certified FACS-coders evaluated the video recordings to rate the frequency, and intensity (on a 5-point scale) of pain-related action units (AUs; AU4, AU6-7, AU9-10, and AU43) for each trial (see Kunz et al. 2011 for details about the AUs selection). In the frequency count, an AU was considered as occurring either for the first time or if it was already present, an intensity increase of 2 points was recorded as a new occurrence. Additionally, if there was a distinct interruption followed by a reappearance of the AU for a given time window, it was also added to the frequency count. From the frequency and intensity scores, a composite score (FACS composite score) was computed by taking the product between the mean AU frequency and mean AU intensity values, reflecting pain expression for each trial. A logarithmic transformation was applied in order to normalize the FACS composite scores (Transformed scores = log(FACS scores + 1); skewness = 0.75, kurtosis = -0.84). The transformed FACS composite scores during the painful trials were used as the predictive variable. All results were reported based on the log transformed scores. To examine whether the facial expression could be confounded with the pain reports, we predicted the trial-by-trial FACS composite scores from the pain ratings using a mixed effect model with the participants as a random effect, and allowing random slopes. Pain ratings were not associated to facial responses ($R^2 = .008$, $R^2 = .74$, $\beta = 0.10 \pm 0.07$, 95% CI = [-0.01; 0.21], $t(32.57) = 1.82$, $p = .07$, **Table S3**). Even if the likelihood and the strength of facial expression of pain generally increase with pain ratings in response to the increase of stimulus intensity, this result is not surprising in the present context where the stimulus intensity is held constant, and spontaneous fluctuations in both facial expression and subjective ratings are observed (see Kunz et al. (2011) and Kunz, Karos, and Vervoort (2018) for a discussion on those results).

To test for a possible habituation or sensitization effect on the facial expressivity throughout the experiment, a mixed effect model was conducted with the trials, the runs, and the interaction between the trials and the runs considered as fixed effects, and the participants as random effect.

No evidence of such habituation or sensitization on the log transformed FACS scores were found in the results ($R^2 = .00$; $R^2 = .71$; trial: $\beta = -0.002 \pm 0.02$, 95% CI = [-0.03; 0.03], $t(496.01) = -0.12$, $p = .90$; run: $\beta = 0.02 \pm 0.22$, 95% CI = [-0.22; 0.25], $t(496.16) = 0.15$, $p = .88$; trial×run: $\beta = 0.001 \pm 0.02$, 95% CI = [-0.05; 0.05], $t(496.05) = 0.05$, $p = .96$; **Figure S1-A** and **Table S4**). The same analysis conducted on the pain intensity ratings led to the same conclusions ($R^2 = .01$; $R^2 = .49$; Trials: $\beta = 0.01 \pm 0.21$, 95% CI = [-0.40; 0.42], $t(496.13) = 0.04$, $p = .97$; Runs: $\beta = 1.98 \pm 1.47$, 95% CI = [-0.91; 4.87], $t(496.47) = 1.35$, $p = .18$; Trials×Runs: $\beta = -0.03 \pm 0.29$, 95% CI = [-0.61; 0.54], $t(496.21) = -0.12$, $p = .91$; **Figure S1-B** and **Table S5**).

Anatomical and functional acquisition

MRI images were acquired using a 12-channel head coil 3T Siemens TRIO scanner. T1-weighted structural data were collected using a MP-RAGE sequence (TR = 2300ms, TE = 2.91ms, flip angle = 9°, FOV = 256mm, matrix size = 240×256, voxel size = 1×1×1.2mm, 170 whole-brain volumes). Functional data were acquired using an EPI T2*-weighted sequence (TR = 3000ms, TE = 30ms, flip angle = 90°, FOV = 220×220mm², voxel size = 3.44×3.44×3.40mm, 40 interleaved axial slices).

Preprocessing of fMRI data

The fMRI data were preprocessed using SPM8 (Statistical Parametric Mapping, Version 8; Wellcome Department of Imaging Neuroscience, London, United Kingdom) on MATLAB 7.4 (The MathWorks Inc, Natick, Massachusetts, United States). Preprocessing steps included a slice-timing correction, a correction for head movements, and co-registration between functional and anatomical images for each participant. Functional images were normalized into the MNI space. A spatial smoothing procedure (6mm FWHM gaussian kernel) and a high pass filter (128s) were also applied.

BOLD signal was modelled using a canonical hemodynamic response function. First level analyses were computed using the GLM to obtain a pain activation map for each trial using SPM8. Additionally, the six movements parameters and averaged signals from the white matter and the cerebrospinal fluid were included as nuisance regressors. Eleven trials were discarded due to excessive movements in the painful condition, and eight trials were excluded in the warm condition, resulting in a total of 533 and 536 activation maps for the pain and the warm conditions, respectively. The activation maps of individual painful trials were used to develop the Facial Expression of Pain Signature (FEPS).

Analyses

Association between the facial expression of pain and pain-related brain signatures

The dot product between the Neurologic Pain Signature (NPS; Wager et al. 2013), the Stimulus Intensity Independent Pain Signature-1 (SIIPS-1; Woo et al. 2017), the Predictive Value of Pain (PVP; Coll et al. 2022), the Thermal Pain Aversive Signature (TPAS; Čeko et al. 2022), and the trial-by-trial activation maps was computed to derive a measure of similarity (pattern expression) between the unthresholded signatures and the activation maps. These scalar values were then correlated with the FACS composite scores to assess the association between the facial expression of pain and the NPS, the SIIPS-1, the PVP, and the TPAS, separately. Pearson-r correlation coefficients, and p-values are reported.

Multivariate pattern analysis

We applied a Least Absolute Shrinkage and Selection Operator Principal Component Regression (LASSO-PCR) with a 10-fold cross-validation procedure for multivariate pattern analysis (Wager et al. 2011) using scikit-learn implementation (Pedregosa et al. 2011). The algorithm was trained on ~70% of the data and tested on the remaining ~30%, and the LASSO alpha hyperparameter was set at 1.0. The analyses were performed using the trial-by-trial activation maps as input to the model and the participants as a grouping factor (i.e. data from a given participant could only be either in the training set or the testing set for a given cross-validation fold). This procedure was used to predict FACS composite scores from activation maps. The performance of each regression model was evaluated using Pearson's correlation coefficient (*pearson-r*), coefficient of determination (R^2 ; computed using scikit-learn), and root mean square error (RMSE). The scikit-learn's implementation of the coefficient of determination was used, allowing negative values if the model showed worse performance compared to an unfitted model (i.e. a horizontal line). The averaged performance metrics across folds are reported for each analysis. To test if the models performed significantly above chance, permutation tests were computed using 5000 iterations, leading to a p-value corresponding to the probability that the R^2 between the observed and predicted FACS scores would be obtained by chance. A bootstrap resampling procedure was also performed to evaluate the stability of the voxel contribution to the model performance, and to derive confidence intervals for the performance metrics. This procedure consists of randomly sampling 5000 times the dataset with replacement. The resulting samples contain the same number of observations as the dataset. The LASSO-PCR procedure as described above is then applied on each sample. Z-scores and p-values are calculated on the overall regression coefficients.

This analysis procedure was first applied at the whole brain level. It was repeated using a spatial mask including only the precentral region bilaterally. This mask was derived from the Oxford-Harvard Cortical Atlas (Caviness et al. 1996). This secondary analysis was conducted to verify if the pattern of activity within the primary motor cortex (M1) might be sufficient to predict facial expression. The performance between the model based on the primary motor cortex activity and the whole-brain model was compared using a corrected resampled t-test (Nadeau and Bengio 1999). To ensure that the whole brain model prediction of the facial responses was not confounded with the pain ratings, we predicted the facial composite scores from the FEPS pattern expression scores (i.e., dot product between the trial-by-trial activation maps and the unthresholded FEPS signature), and included the trial-by-trial pain ratings using a mixed effect model including the participants as a random effect, and allowing the slopes to vary. Variance in the facial composite scores was significantly explained by the FEPS pattern expression scores ($R^2_{\text{GLMM(m)}} = .40$; $R^2_{\text{GLMM(c)}} = .69$; FEPS scores: $\beta = 0.62 \pm 0.04$, 95% CI = [0.54; 0.70], $t(431.79) = 14.8$, $p < .001$; pain ratings: $\beta = 0.02 \pm 0.05$, 95% CI = [-0.07; 0.11], $t(30.34) = 0.48$, $p = .68$; FEPS scores $\times$ pain ratings: $\beta = -0.09 \pm 0.04$, 95% CI = [-0.17; -0.01], $t(194.89) = -2.25$, $p = .03$; see **Table S6**).

These results confirm the prediction of the facial response by the FEPS scores even when pain ratings are included as a predictor in the model. However, a small but significant negative interaction between the FEPS scores and the pain ratings was found. This possible moderator effect indicates that, for a constant stimulus, the positive slopes between the FEPS scores and the facial responses is slightly reduced when pain ratings are higher. This may reflect a saturation effect across the two output channels (i.e., verbal reports and facial expression).

Response of the FEPS to pain and warm

To test if the FEPS was more activated during the painful condition compared to the warm conditions, the FEPS expression scores were further computed on the warm trials. A linear mixed model was performed to examine the relation between the FEPS scores and the experimental conditions (Warm and Pain), considering those conditions as a fixed categorical effect and the participants as a random effect. Contrasts on the estimated marginal means (least-squares means)

were conducted to assess the statistical significance of the difference between the FEPS expression scores on the warm trials and the painful trials. Given that there are also trials where the FACS scores were equal to zero in the painful condition, the same analyses were repeated, this time separating the FEPS expression scores based on whether FACS scores were greater or equal to 0 in the painful condition. The estimated marginal means (EMM) contrasts between experimental conditions were reported with the associated *t*-value and *p*-value (corrected for multiple comparisons using Tukey method).

Spatial similarity across the FEPS and pain-related brain signatures

Similarity between the FEPS and other pain-related brain signatures was assessed using the cosine similarity computed across all voxels ($\text{Similarity}(X, Y) = \frac{X \cdot Y}{\|X\| \|Y\|}$). This metric was computed on the unthresholded NPS, SIIPS-1, PVP, and TPAS maps. To further explore the cortical similarities between the FEPS and other pain-related brain signatures, we also computed the cosine similarity across different cortical networks (i.e. visual, somatomotor, dorsal attention, ventral attention, limbic, frontoparietal, and default mode)(Yeo et al. 2011 [\[1\]](#)). Permutation tests ($n = 1000$) using generative null models preserving the spatial autocorrelation of a target brain map were used to assess the significance of the similarity between the brain signatures (Burt et al. 2020 [\[2\]](#)). *P*-values were calculated from the generated null distributions as the fraction of permuted samples where the absolute cosine similarity was equal or larger than the cosine similarity obtained on the original signatures.

Acknowledgements

We wish to thank André Cyr and Carolyn Hurst from the Unité de Neuroimagerie Fonctionnelle for their help with the data acquisition, and Andréanne Proulx, François Paugman and Pierre Bellec for their insights on the analyses as part of BrainHack School. We also want to acknowledge Compute Canada for the computational resources allocated to conduct the analyses. M.-E.P. was supported by scholarships from the Fonds de Recherche du Québec-Nature et Technologies and from the Institut de Valorisation des Données. This study was funded by a discovery grant from the Natural Sciences and Engineering Research (RGPIN-2018-06799) to P.R. and the NIH (R01 DA035484) to T.D.W.

Credit statement

Marie-Eve Picard: Conceptualization, Software, Formal analysis, Writing-Original draft, Visualization. **Miriam Kunz:** Conceptualization, Methodology and data acquisition, Data curation. **Jen-I Chen:** Methodology and data acquisition, Data curation, Formal analysis. **Michel-Pierre Coll:** Software, Validation, Visualization. **Étienne Vachon-Pressseau:** Conceptualization, Methodology, Validation. **Tor D. Wager:** Conceptualization, Validation, Funding acquisition. **Pierre Rainville:** Conceptualization, Methodology and data acquisition, Writing-Original draft, Funding acquisition. All authors reviewed and approved the final version.

Data Availability

Participants did not provide informed consent for the sharing of individual data but unthresholded group maps are available upon request. Unthresholded and thresholded FEPS patterns are available on Neurovault (<https://neurovault.org/collections/13924/> [\[3\]](#)). The SIIPS-1, the PVP and the TPAS are available on Github (https://github.com/canlab/Neuroimaging_Pattern_Masks [\[4\]](#)). The NPS is available upon request from T.D.W.

Code Availability

All custom Python scripts used to produce the analyses are available at https://github.com/me-picard/picard_feps_2022 .

Supplementary Material

Figure S1

Behavioral scores across trials.

A) Fluctuations in the log transformed FACS composite scores across trials in the two runs (shaded areas represent the standard deviation). There was no significant effect of trials and runs (p 's > .8; see results). B) Distribution of the log transformed FACS composite scores (each point represents one trial). C) Fluctuations in the pain intensity ratings across trials between runs. There was no significant effect of trials and runs (p 's > .15; see results). D) Distribution of the pain intensity ratings.

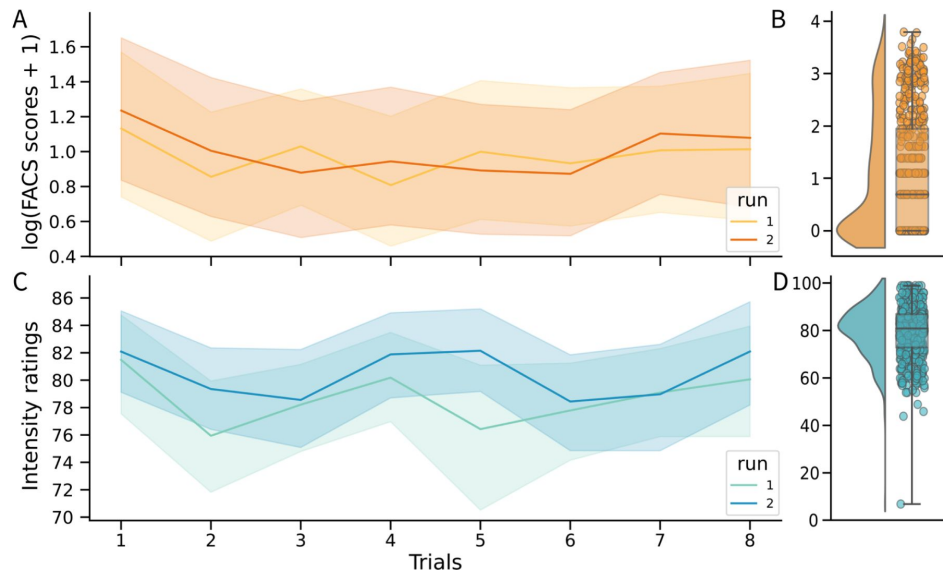


Figure S2

Predictive performance of the M1-based model.

A) Relationship between the actual and the predicted FACS composite scores for each cross-validation fold ($k=10$) using only the activity from the primary motor cortex as defined by the Oxford-Harvard Cortical Atlas (Caviness et al. 1996). B) Distribution of the Pearson-r scores across the 10 cross-validation folds ($\text{Pearson-}r = .28 \pm .14$ (95% CI = [.17; .49]); $R^2 = -.02 \pm .12$ (95% CI = [-.41; .16]); $\text{RMSE} = 1.13 \pm 0.11$ (95% CI = [0.99; 1.23]); $p = .01$).

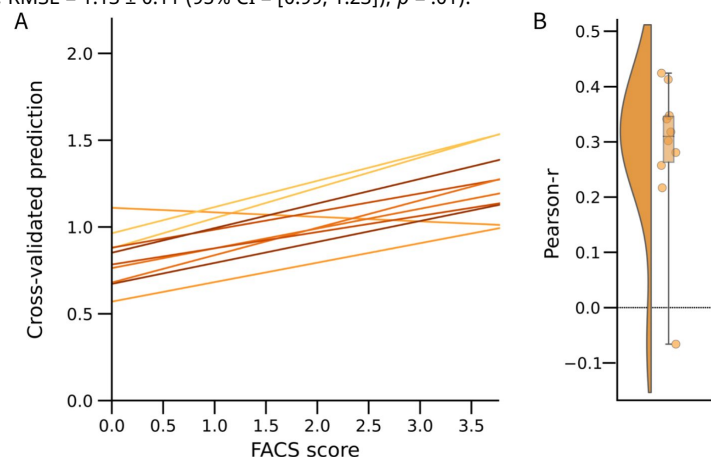


Figure S3

FEPS pattern expression and pain facial expression.

A) FEPS pattern expression scores significantly predict the log transformed FACS scores (see [Figure 1](#) in main text for the results of the 10-fold cross-validation). Each point represents a trial in the pain condition. B) Distribution of the FEPS pattern expression scores in the warm (nonpainful) condition and in the pain condition, separating the trials with and without pain facial expression. The FEPS pattern expression was computed using the dot product between the FEPS weights and the trial-by-trial activation maps. The whiskers on the box plots extend to either 1.5 times the interquartile range or to the farthest point within that range. Statistical results are reported in the main text.

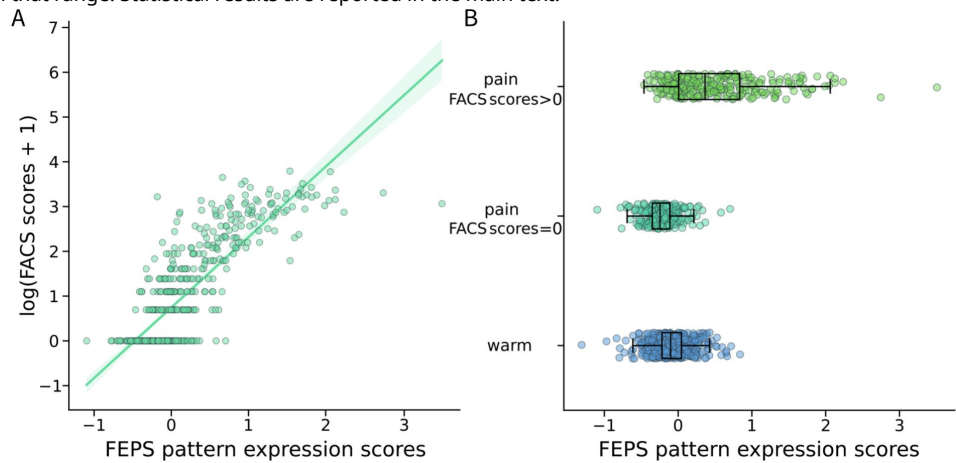


Table S1.

Peak regions with positive weights contributing to the prediction of the facial expression scores.

Regions	x	y	z	Clusters size (mm ³)	Peaks z-score
Frontal pole (bilateral)	-10	70	-2	1928	4.43
Lateral inferior frontal gyrus (bilateral)	-42	60	-8	336	4.78
Precentral gyrus (M1; right)	44	-14	38	3640	5.96
Precentral gyrus (M1; left)	-46	-10	36	2472	4.90
Parietal operculum (right)	48	-26	34	256	3.58
Inferior parietal lobule (right)	56	-28	42	56	3.20
Posterior cingulate gyrus (right)	20	-40	46	304	3.90
Occipitotemporal/ Inferior temporal gyrus (left)	-48	-80	8	168	3.96

The main clusters are shown in Figure 1

Table S2.

Peak regions with negative weights contributing to the prediction of the facial expression scores.

Regions	x	y	z	Clusters size (mm ³)	Peaks z-score
vlPFC (right)	36	52	-2	1128	-4.09
vlPFC (left)	-40	50	-2	1288	-4.71
dIPFC (right)	42	46	20	1136	-4.06
dIPFC (left)	-38	40	14	688	-4.04
vmPFC/subgenual ACC (bilateral)	-4	34	-10	5560	-4.80
aMCC	0	20	30	24	-3.13
Ventral striatum	14	18	-10	536	-3.94
Superior frontal gyrus	24	10	58	168	-3.43
Temporal operculum (right)	52	8	-6	152	-4.00
Middle frontal gyrus (premotor area; left)	-32	0	58	160	-3.71
MCC (right)	4	4	44	16	-3.27
Post central (S1)/inferior parietal lobule (right)	50	-36	50	88	-3.37
Fusiform gyrus (right)	28	-60	-14	56	-3.45
Precuneus (left)	-14	-66	42	544	-4.08
Precuneus (right)	12	-66	54	48	-3.31

The main clusters are shown in Figure 1

Table S3.

Mixed-effects models results for the effect of pain ratings on the logarithmic transformed FACS scores.

<i>Predictors</i>	Log(FACS scores)		
	<i>Estimates</i>	<i>CI</i>	<i>p</i>
(Intercept)	0.99	0.66 – 1.33	<.001
Pain ratings	0.08	-0.01 – 0.21	.069
Random Effects			
σ^2	0.36		
τ_{00} Subject	0.95		
τ_{11} Subject.Rating	0.05		
ρ_{01} Subject	0.64		
ICC	0.74		
N _{Subject}	34		
Observations	533		
Marginal R ² / Conditional R ²	0.008 / 0.737		

Table S4.

Mixed-effects models results for the effect of runs and trials on the logarithmic transformed FACS scores

Log(FACS scores)			
<i>Predictors</i>	<i>Estimates</i>	<i>CI</i>	<i>p</i>
(Intercept)	1.00	0.63 – 1.37	<.001
Trial	-0.00	-0.03 – 0.03	.902
Run	0.02	-0.22 – 0.25	.884
Trial × Run	0.00	-0.05 – 0.05	.962
Random Effects			
σ^2	0.39		
τ_{00} Subject	0.95		
ICC	0.71		
N _{Subject}	34		
Observations	533		
Marginal R ² / Conditional R ²	0.000 / 0.710		

Table S5.

Mixed-effects models results for the effect of runs and trials on the pain ratings

Pain ratings			
<i>Predictors</i>	<i>Estimates</i>	<i>CI</i>	<i>p</i>
(Intercept)	78.58	75.30 – 81.85	<.001
Trial	0.01	-0.40 – 0.42	.965
Run	1.98	-0.91 – 4.87	.179
Trial × Run	-0.03	-0.61 – 0.54	.907
Random Effects			
σ^2	59.34		
τ_{00} Subject	56.91		
ICC	0.49		
N _{Subject}	34		
Observations	533		
Marginal R ² / Conditional R ²	0.007 / 0.493		

Log(FACS scores)			
<i>Predictors</i>	<i>Estimates</i>	<i>CI</i>	<i>p</i>
(Intercept)	1.00	0.82 – 1.17	<.001
FEPS expression scores	0.62	0.54 – 0.70	<.001
Pain ratings	0.02	-0.07 – 0.11	.630
FEPS expression scores × Pain ratings	-0.09	-0.17 – -0.01	.025
Random Effects			
σ^2	0.29		
τ_{00} Subject	0.25		
τ_{11} Subject.Rating	0.03		
ρ_{01} Subject	0.56		
ICC	0.49		
N _{Subject}	34		
Observations	533		
Marginal R ² / Conditional R ²	0.396 / 0.690		

Table S6.

Mixed-effects models results for the effect of the FEPS expression scores and the pain ratings on the logarithmic transformed FACS scores

References

- Bastuji H      , Frot Maud, Perchet Caroline, Magnin Michel, Garcia-Larrea Luis (2016) **Pain Networks from the inside: Spatiotemporal Analysis of Brain Responses Leading from Nociception to Conscious Perception** *Human Brain Mapping* **37**:4301–15 <https://doi.org/10.1002/hbm.23310>
- Burt Joshua B., Helmer Markus, Shinn Maxwell, Anticevic Alan, Murray John D. (2020) **Generative Modeling of Brain Maps with Spatial Autocorrelation** *NeuroImage* **220** <https://doi.org/10.1016/j.neuroimage.2020.117038>
- Caviness V. S., Meyer J., Makris N., Kennedy D. N. (1996) **MRI-Based Topographic Parcellation of Human Neocortex: An Anatomically Specified Method with Estimate of Reliability** *Journal of Cognitive Neuroscience* **8**:566–87 <https://doi.org/10.1162/jocn.1996.8.6.566>
-      Marta, Kragel Philip A., Woo Choong-Wan, L         Marina, Wager Tor D. (2022) **Common and Stimulus-Type-Specific Brain Representations of Negative Affect** *Nature Neuroscience* **25**:760–70 <https://doi.org/10.1038/s41593-022-01082-w>
- Coll Michel-Pierre, Slimani Hocine, Woo Choong-Wan, Wager Tor D., Rainville Pierre, Vachon-Pressseau   tienne, Roy Mathieu (2022) **The Neural Signature of the Decision Value of Future Pain** *Proceedings of the National Academy of Sciences* **119** <https://doi.org/10.1073/pnas.2119931119>
- Costa Emanuela Dalla, Minero Michela, Lebelt Dirk, Stucke Diana, Canali Elisabetta, Leach Matthew C. (2014) **Development of the Horse Grimace Scale (HGS) as a Pain Assessment Tool in Horses Undergoing Routine Castration** *PLOS ONE* **9** <https://doi.org/10.1371/journal.pone.0092281>
- Craig Kenneth D (1992) **The Facial Expression of Pain Better than a Thousand Words?** *APS Journal* **1**:153–62 [https://doi.org/10.1016/1058-9139\(92\)90001-S](https://doi.org/10.1016/1058-9139(92)90001-S)
- Cul Antoine Del, Baillet Sylvain, Dehaene Stanislas (2007) **Brain Dynamics Underlying the Nonlinear Threshold for Access to Consciousness** *PLOS Biology* **5** <https://doi.org/10.1371/journal.pbio.0050260>
- Ekman Paul, Friesen Wallace V. (1978) **Facial Action Coding System**
- Evangelista Marina C., Watanabe Ryota, Leung Vivian S. Y., Monteiro Beatriz P., O'Toole Elizabeth, Pang Daniel S. J., Steagall Paulo V. (2019) **Facial Expressions of Pain in Cats: The Development and Validation of a Feline Grimace Scale** *Scientific Reports* **9** <https://doi.org/10.1038/s41598-019-55693-8>
- Garcia-Larrea Luis, Peyron Roland (2013) **“Pain Matrices and Neuropathic Pain Matrices: A Review.”** *PAIN Biennial Review of Pain* **154**:S29–43 <https://doi.org/10.1016/j.pain.2013.09.001>
- Goldin Philippe R., McRae Kateri, Ramel Wiveka, Gross James J. (2008) **“The Neural Bases of Emotion Regulation: Reappraisal and Suppression of Negative Emotion.”** *Biological Psychiatry, Stress Anxiety, and Post-Traumatic Stress Disorder* **63**:577–86 <https://doi.org/10.1016/j.biopsych.2007.05.031>

Hadjistavropoulos Thomas, Craig Kenneth D., Duck Steve, Cano Annmarie, Goubert Liesbet, Jackson Philip L., Mogil Jeffrey S., et al. (2011) **A Biopsychosocial Formulation of Pain Communication** *Psychological Bulletin* **137**:910–39 <https://doi.org/10.1037/a0023876>

Karmann Anna Julia, Maihöfner Christian, Lautenbacher Stefan, Sperling Wolfgang, Kornhuber Johannes, Kunz Miriam (2016) **The Role of Prefrontal Inhibition in Regulating Facial Expressions of Pain: A Repetitive Transcranial Magnetic Stimulation Study** *The Journal of Pain* **17**:383–91 <https://doi.org/10.1016/j.jpain.2015.12.002>

Anjali Krishnan, Woo Choong-Wan, Chang Luke J, Ruzic Luka, Xiaosi Gu, López-Solà Marina, Jackson Philip L, Pujol Jesús, Fan Jin, Wager Tor D (2016) **“Somatic and Vicarious Pain Are Represented by Dissociable Multivariate Brain Patterns.”** *eLife* **5** <https://doi.org/10.7554/eLife.15166>

Kunz Miriam, Stefan Lautenbacher Jen-I. Chen, Vachon-Presseau Etienne, Rainville Pierre (2011) **Cerebral Regulation of Facial Expressions of Pain** *Journal of Neuroscience* **31**:8730–38 <https://doi.org/10.1523/JNEUROSCI.0217-11.2011>

Kunz Miriam, Chen Jen-I, Rainville Pierre (2020) **Keeping an Eye on Pain Expression in Primary Somatosensory Cortex** *NeuroImage* **217** <https://doi.org/10.1016/j.neuroimage.2020.116885>

Kunz Miriam, Karos Kai, Vervoort Tine, Vervoort Tine, Karos Kai, Trost Zina, Prkachin Kenneth M. (2018) **When, How, and Why Do We Express Pain?** *Social and Interpersonal Dynamics in Pain: We Don't Suffer Alone* :101–19 https://doi.org/10.1007/978-3-319-78340-6_6

Langford Dale J., Bailey Andrea L., Chanda Mona Lisa, Clarke Sarah E., Drummond Tanya E., Echols Stephanie, Glick Sarah, et al. (2010) **Coding of Facial Expressions of Pain in the Laboratory Mouse** *Nature Methods* **7**:447–49 <https://doi.org/10.1038/nmeth.1455>

Claude Nadeau, Bengio Yoshua (1999) **“Inference for the Generalization Error.”** *Advances in Neural Information Processing Systems*.

Pan Junhao, Zhan Liying, Hu ChuanLin, Yang Junkai, Wang Cong, Li Gu Shengqi, Zhong, et al. (2018) **“Emotion Regulation and Complex Brain Networks: Association Between Expressive Suppression and Efficiency in the Fronto-Parietal Network and Default-Mode Network.”** *Frontiers in Human Neuroscience* **12**

Pedregosa Fabian, Varoquaux Gaël, Gramfort Alexandre, Michel Vincent, Thirion Bertrand, Grisel Olivier, Blondel Mathieu, et al. (2011) **Scikit-Learn: Machine Learning in Python** *Journal of Machine Learning Research* **12**:2825–30

Price Donald D (2000) **Psychological and Neural Mechanisms of the Affective Dimension of Pain** *Science* **288**:1769–72 <https://doi.org/10.1126/science.288.5472.1769>

Roy Mathieu, Mathieu Piche Jen-I, Chen Isabelle Peretz, Rainville Pierre (2009) **“Cerebral and Spinal Modulation of Pain by Emotions.”** *Nature Precedings February* <https://doi.org/10.1038/npre.2009.2885.1>

Rushworth Matthew F. S., Noonan MaryAnn P., Boorman Erie D., Walton Mark E., Behrens Timothy E. (2011) **Frontal Cortex and Reward-Guided Learning and Decision-Making** *Neuron* **70**:1054–69 <https://doi.org/10.1016/j.neuron.2011.05.014>

- Shackman Alexander J., Salomons Tim V., Slagter Heleen A., Fox Andrew S., Winter Jameel J., Davidson Richard J. (2011) **The Integration of Negative Affect, Pain and Cognitive Control in the Cingulate Cortex** *Nature Reviews Neuroscience* **12**:154–67 <https://doi.org/10.1038/nrn2994>
- Simon Daniela, Craig Kenneth D., Gosselin Frederic, Belin Pascal, Rainville Pierre (2008) **Recognition and Discrimination of Prototypical Dynamic Expressions of Pain and Emotions** *PAIN®* **135**:55–64 <https://doi.org/10.1016/j.pain.2007.05.008>
- Sliwa Julia, Mallet Marion, Christiaens Maëlle, Takahashi Daniel Y. (2022) **Neural Basis of Multi-Sensory Communication in Primates** *Ethology Ecology & Evolution* **34**:322–43 <https://doi.org/10.1080/03949370.2021.2024266>
- Sneddon Lynne U (2019) **Evolution of Nociception and Pain: Evidence from Fish Models** *Philosophical Transactions of the Royal Society B: Biological Sciences* **374** <https://doi.org/10.1098/rstb.2019.0290>
- Sotocina Susana G, Sorge Robert E, Zaloum Austin, Tuttle Alexander H, Martin Loren J, Wieskopf Jeffrey S, Mapplebeck Josiane CS, et al. (2011) **The Rat Grimace Scale: A Partially Automated Method for Quantifying Pain in the Laboratory Rat via Facial Expressions** *Molecular Pain* **7**:1744–8069 <https://doi.org/10.1186/1744-8069-7-55>
- Vachon-Pressseau Etienne, Roy Mathieu, Woo Choong-Wan, Kunz Miriam, Martel Marc-Olivier, Sullivan Michael J., Jackson Philip L., Wager Tor D., Rainville Pierre (2016) **Multiple Faces of Pain: Effects of Chronic Pain on the Brain Regulation of Facial Expression** *Pain* **157**:1819–30 <https://doi.org/10.1097/j.pain.0000000000000587>
- Wager Tor D., Atlas Lauren Y., Leotti Lauren A., Rilling James K. (2011) **Predicting Individual Differences in Placebo Analgesia: Contributions of Brain Activity during Anticipation and Pain Experience** *Journal of Neuroscience* **31**:439–52 <https://doi.org/10.1523/JNEUROSCI.3420-10.2011>
- Wager Tor D., Atlas Lauren Y., Lindquist Martin A., Roy Mathieu, Woo Choong-Wan, Kross Ethan (2013) **An fMRI-Based Neurologic Signature of Physical Pain** *New England Journal of Medicine* **368**:1388–97 <https://doi.org/10.1056/NEJMoa1204471>
- Woo Choong-Wan, Schmidt Liane, Krishnan Anjali, Jepma Marieke, Roy Mathieu, Lindquist Martin A., Atlas Lauren Y., Wager Tor D. (2017) **Quantifying Cerebral Contributions to Pain beyond Nociception** *Nature Communications* **8** <https://doi.org/10.1038/ncomms14211>
- Yeo Thomas B. T., Krienen Fenna M., Sepulcre Jorge, Sabuncu Mert R., Lashkari Danial, Hollinshead Marisa, Roffman Joshua L., et al. (2011) **The Organization of the Human Cerebral Cortex Estimated by Intrinsic Functional Connectivity** *Journal of Neurophysiology* **106**:1125–65 <https://doi.org/10.1152/jn.00338.2011>
- Zheng Weihao, Woo Choong-Wan, Yao Zhijun, Goldstein Pavel, Atlas Lauren Y, Roy Mathieu, Schmidt Liane, et al. (2020) **Pain-Evoked Reorganization in Functional Brain Networks** *Cerebral Cortex* **30**:2804–22 <https://doi.org/10.1093/cercor/bhz276>

Editors

Reviewing Editor

José Biurrun Manresa

National Scientific and Technical Research Council (CONICET), National University of Entre Ríos (UNER), Oro Verde, Argentina

Senior Editor

Christian Büchel

University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Reviewer #2 (Public Review):

Summary.

The objective of this study was to further our understanding of the brain mechanisms associated with facial expressions of pain. To achieve this, participants' facial expressions and brain activity were recorded while they received noxious heat stimulation. The authors then used a decoding approach to predict facial expressions from functional magnetic resonance imaging (fMRI) data. They found a distinctive brain signature for pain facial expressions (FEPS). This signature had minimal overlap with brain signatures reflecting other components of pain phenomenology, such as signatures reflecting subjective pain intensity or negative effects.

Strength.

The authors used a rigorous approach involving multivariate brain decoding to predict the occurrence and intensity of pain facial expressions during noxious heat stimulation. The analyses are solid and well-conducted. This is an important study of fundamental and clinical relevance.

Weakness.

Despite those major strengths, the main weakness of the study is that the design and analyses do not allow us to know if the FEPS is really specific to pain expressions. Based on the analysis, it is possible to conclude that this brain signature is present when a participant is in a state of pain and displays a facial expression. However, it is possible that it would also be present when a participant experiences (another) negative state and displays (another) facial expression. It will be important, in future work, to investigate the specificity of this brain signature.

<https://doi.org/10.7554/eLife.87962.2.sa2>

Reviewer #3 (Public Review):

In this manuscript, Picard et al. propose a Facial Expression Pain Signature (FEPS) as a distinctive marker of pain processing in the brain. Specifically, they attempt to use functional magnetic resonance imaging (fMRI) data to predict facial expressions associated with painful heat stimulation.

The main strengths of the manuscript are that it is built on an extensive foundation of work from the research group, and that experience can be observed in the analysis of fMRI data and the development of the machine learning model. Additionally, it provides a comparative account of the similarities of the FEPS with other proposed pain signatures. The main

weaknesses of the manuscript are the absence of a proper control condition to assess the specificity of the facial pain expressions, as well as several limitations in the experimental setup.

I believe that the authors partially succeed in their aims, as described in the introduction, which are to assess the association between pain facial expression and existing pain-relevant brain signatures, and to develop a predictive brain activation model of the facial responses to painful thermal stimulation. However, they list several limitations in the study that should be addressed in future research in order to establish whether FEPS truly conveys distinctive information about the brain response to nociceptive stimuli.

<https://doi.org/10.7554/eLife.87962.2.sa1>

Author response:

The following is the authors' response to the original reviews.

Summary of the changes

Changes in the manuscript were made to clarify some ambiguities raised by the reviewers and to improve the report following their recommendations. A summary of the main changes is listed below:

- The title was changed to better reflect the results of this study - Re-training the model on log transformed FACS scores.
- Testing the specificity of the FEPS to facial expression of pain within this experimental setup by comparing it to the activation maps obtained from the Warm stimulation condition.
- Testing for sensitization/habituation of the behavioral measures (FACS scores and pain ratings).
- Adding a section in the discussion to better address the limitations of this study and provide potential directions for future studies.

Other changes target areas where the original manuscript may have been ambiguous or lacked precision. To address these concerns, additional details have been incorporated, and certain terms have been revised to ensure a more precise and transparent presentation of the information.

Public Reviews:

Reviewer #1 (Public Review):

Picard et al. report a novel neural signature of facial expressions of pain. In other words, they provide evidence that a specific set of brain activations, as measured by means of functional magnetic resonance imaging (fMRI), can tell us when someone is expressing pain via a concerted activation of distinctive facial muscles. They demonstrate that this signature provides a better characterization of this pain behaviour when compared with other signatures of pain reported by past research. The Facial Expression of Pain Signature (FEPS) thus enriches this collection and, if further validated, may allow scientists to identify the neural structures subserving important non-verbal pain behaviour. I have, however, some reservations about the strength of the evidence, relating to insufficient characterization of the underlying processes involved.

We are thankful for the summary of our work. We are hopeful that the modifications made in the latest version effectively address these concerns. The changes are outlined in the

summary above, and detailed in the following point-by-point response.

Strengths:

The study relies on a robust machine-learning approach, able to capitalise on the multivariate nature of the fMRI data, an approach pioneered in the field of pain by one of the authors (Dr. Tor Wager). This paper extends Wager's and other colleagues' work attempting to identify specific combinations of brain structures subserving different aspects of the pain experience while examining the extent of similarity/dissimilarity with the other signatures. In doing so, the study provides further methodological insight into fine-grained network characterization that may inspire future work beyond this specific field.

We are thankful for the positive comments.

Weaknesses:

The main weakness concerns the lack of a targeted experimental design aimed to dissect the shared variance explained by activations both specific to facial expressions and to pain reports. In particular, I believe that two elements would have significantly increased the robustness of the findings:

(1) Control conditions for both the facial expressions and the sensory input. An efficient signature should not be predictive of neutral and emotional facial expressions (e.g., disgust) other than pain expressions, as well as it should not be predictive of sensations originating from innocuous warm stimulation or other unpleasant but non-painful stimulation.

We do recognize the lack of specificity testing for the FEPS, especially towards negative emotional facial expressions. This would be relevant to test given the behavioural overlap between the facial expressions of pain and disgust, fear, anger, and sadness (Kunz et al., 2013; Williams, 2003). The experimental design used in this study did not include other negative states. However, we fully support the necessity of collecting data throughout those conditions, and we believe that the present study highlights the importance of such a demonstration. Future research should involve recording facial expressions while exposing participants to stimuli that elicit a range of negative emotions but, to our knowledge, such combination of fMRI and behavioural data is currently unavailable. As raised by the reviewer, this approach would allow us to assess the specificity of the FEPS to the facial expression evoked by pain compared to different affective states. We would like to emphasise that specificity and generalizability testing is a massive amount of work, requiring multiple studies to address comprehensively. A Limitations paragraph addressing this research direction has been added to the Discussion. A conclusion was added to the abstract as follows: "Future studies should explore other pain-relevant manifestations and assess the specificity of the FEPS against other types of aversive or emotional states."

(2) Graded intensity of the sensory stimulation: different intensities of the thermal stimulation would have caused a graded facial expression (from neutral to pain) and graded verbal reports (from no pain to strong pain), thus offering a sensitive characterisation of the signal associated with this condition (and the warm control condition).

However, these conditions are missing from the current design, and therefore we cannot make a strong conclusion about the generalisability of the signature (regardless of whether it can predict better than other signatures - which may/may not suffer from similar or other methodological issues - another potential interesting scientific question!).

The authors seem to work on the assumption that the trials where warm stimulation was delivered are of no use. I beg to disagree. As per my previous comment, warm trials (and associated neutral expressions) could be incorporated into the statistical model to increase the classification sensitivity and precision of the FEPS decoding.

The experience of pain can fluctuate for a fixed intensity or after controlling statistically for the intensity of the stimulation (Woo et al., 2017). Consistent with this, the current study focused on spontaneous facial expression in response to noxious thermal stimuli delivered at a constant intensity that produced moderate to strong pain in every participant. As the reviewer points out, this does not allow us to characterise and compare the stimulus-response function of facial expression and pain ratings. The advantage of the approach adopted is to maximise the number of trials where facial expression is more likely to occur, while ensuring that changes in facial expression and pain ratings are not confounded with changes in stimulus intensity. The manuscript has been revised to clarify that point. However, we do agree that it would be interesting to conduct more studies focusing on facial expression in response to a range of stimulus intensities. This discussion has been added to the Limitations paragraph.

Furthermore, following the reviewer's suggestion, we performed complementary analyses on the warm trials in the proposed revisions. The dot product (FEPS scores) between the FEPS and the activation maps associated with the warm condition was computed. A linear mixed model was conducted to investigate the association between FEPS scores and the experimental condition (warm vs pain). The trials in the pain condition were divided into two conditions: null FACS scores (painful trials with no facial response; FACS scores = 0) and non-null FACS scores (painful trials with a facial response; FACS > 0). The details of this analysis have been added to the manuscript (see Response of the FEPS to pain and warm section in the Methods; lines 427 to 439) as well as the corresponding results (see Results and Discussion; lines 138 to 158). The FEPS scores were larger in the pain condition where a facial response was expressed, compared to both the pain condition without facial expression and the warm condition. These results confirmed the sensitivity of the FEPS to facial expression of pain.

Reviewer #2 (Public Review):

Summary:

The objective of this study was to further our understanding of the brain mechanisms associated with facial expressions of pain. To achieve this, participants' facial expressions and brain activity were recorded while they received noxious heat stimulation. The authors then used a decoding approach to predict facial expressions from functional magnetic resonance imaging (fMRI) data. They found a distinctive brain signature for pain facial expressions. This signature had minimal overlap with brain signatures reflecting other components of pain phenomenology, such as signatures reflecting subjective pain intensity or negative effects.

We appreciate this concise and accurate summary of our study.

Strength:

The manuscript is clearly written. The authors used a rigorous approach involving multivariate brain decoding to predict the occurrence and intensity of pain facial expressions during noxious heat stimulation. The analyses seem solid and well-conducted. I think that this is an important study of fundamental and clinical relevance.

Weaknesses:

Despite those major strengths, I felt that the authors did not sufficiently explain their own interpretation of the significance of the findings. What does it mean, according to them, that the brain signature associated with facial expressions of pain shows a minimal overlap with other pain-related brain signatures?

We express our sincere gratitude for the valuable insights and constructive comments on the strengths and weaknesses of the current study. We thank reviewer 2 for the encouragement to reinforce our interpretation of the significance of the findings, while acknowledging the limitations raised by the three reviewers.

A few questions also arose during my reading.

Question 1: Is the FEPS really specific to pain expressions? Is it possible that the signature includes a facial expression signal that would be shared with facial expressions of other emotions, especially since it involves socio-affective regulation processes? Perhaps this question should be discussed as a limit of the study?

We acknowledge this limitation as outlined in response to Reviewer #1. We have incorporated a Limitations paragraph to provide a more in-depth discussion of this limitation and to explore potential future avenues (lines 225 to 268). Again, please note that the demonstration of specificity is an incremental process that requires a systematic comparison with other conditions where facial expressions are produced without pain. A concluding sentence was added to the abstract to encourage specificity testing in future studies, as indicated above.

Question 2: All AUs are combined together in a composite score for the regression. Given that the authors have other work showing that different AUs may be associated with different components of pain (affective vs. sensory), is it possible that combining all AUs together has decreased the correlation with other pain signatures? Or that the FEPS actually reflects multiple independent signatures?

The question raised is consistent with the work of Kunz, Lautenbacher, LeBlanc and Rainville (2012), and Kunz, Chen and Rainville (2020). In the current study, the pain-relevant action units were combined in order to increase the number of trials where a facial response to pain was expressed, thus enhancing the robustness of our analyses. Given the limited sample size, our current dataset is unfortunately insufficient to perform such analysis as there would not be enough trials to look at the action units separately or in subgroups. While the approach of combining the different AUs has proven to be valid and useful, we recognize the value of investigating potential independent signatures associated with the different AUs within the FEPS, and examining whether those signatures can lead to more similar patterns compared to previously developed pain signatures. This discussion has been included in the Limitations paragraph in the Discussion (lines 225 to 268).

Question 3: Is facial expressivity constant throughout the experiment? Is it possible that the expressivity changes between the beginning and the end of the experiment? For instance, if there is a habituation, or if the participant is less surprised by the pain, or in contrast if they get tired by the end of the experiment and do not inhibit their expression as much as they did at the beginning. If facial expressivity changes, this could perhaps affect the correlation with the pain ratings and/or with the brain signatures; perhaps time (trial number) could be added as one of the variables in the model to address this question.

The concern raised by the reviewer is legitimate. We conducted a mixed-effects model to assess the impact of successive trials and runs on facial expressivity. Results indicate that the

FACS scores did not change significantly throughout the experiment, suggesting no notable effect of habituation or sensitization on the facial expressivity in our study. Details about the analysis and the results have been added to the Facial Expression section in the Methods (lines 335 to 346).

Reviewer #3 (Public Review):

In this manuscript, Picard et al. propose a Facial Expression Pain Signature (FEPS) as a distinctive marker of pain processing in the brain. Specifically, they attempt to use functional magnetic resonance imaging (fMRI) data to predict facial expressions associated with painful heat stimulation. The main strengths of the manuscript are that it is built on an extensive foundation of work from the research group, and that experience can be observed in the analysis of fMRI data and the development of the machine learning model. Additionally, it provides a comparative account of the similarities of the FEPS with other proposed pain signatures. The main weaknesses of the manuscript are the absence of a proper control condition to assess the specificity of the facial pain expressions, a few relevant omissions in the methodology regarding the original analysis of the data and its purpose, and a biased interpretation of the results.

I believe that the authors partially succeed in their aims, as described in the introduction, which are to assess the association between pain facial expression and existing pain-relevant brain signatures, and to develop a predictive brain activation model of the facial responses to painful thermal stimulation. However, I believe that there is a clear difference between those aims and the claim of the title, and that the interpretation of the results needs to be more rigorous.

We wish to express our appreciation for the insightful and constructive critique provided. The limitation pertaining to the absence of specificity testing had been addressed in response to Reviewer #1, and it has been incorporated into the manuscript (lines 251 to 258).

The commentary made by Reviewer #3 has drawn our attention to a critical concern, namely the potential misalignment between the study findings and our original title. Consequently, we have changed the title to “A distributed brain response predicting the facial expression of acute nociceptive pain”. We also revised the interpretation of the results in the discussion section and we have added a section on limitations.

Recommendations for the Authors:

Reviewer #1 (Recommendations For The Authors):

I hope the following comments will be useful to improve the manuscript.

Abstract

I felt the abstract could be more clear in terms of experimental or scientific questions, hypotheses/expectations, and findings. I also feel the abstract should briefly support the conclusive claim ("is better than...": how better? Or according to what criterion? This may be more relevant than the final conclusive general sentence that does not specifically address the significance of the findings).

The abstract was revised to reinforce the functional perspective adopted to interpret brain activity produced by noxious stimuli and predicting various pain-relevant manifestations. We also mention explicitly the other pain-relevant signatures against which the FEPS is compared in this report, and we added a concluding sentence highlighting the importance of assessing the specificity of the FEPS in future studies.

I would postpone the discussion around pain signature and anticipate the one about the brain mechanisms of facial expressions of pain. This will allow you to reinforce the logical flow of rationale, literature gap/question, why the problem is important, and study aims. Only then go for a review of relevant literature on signatures before providing a more specific final paragraph about the study-specific questions, expectations, and implementation. At the moment this is limited to a single very descriptive short paragraph at the end of the intro.

The introduction was structured to guide the readers through a comprehensive understanding of different pain neurosignatures. The introduction aimed to establish a robust rationale for the subsequent analyses detailed in the results section. Indeed, the presentation of that literature ensured that the discussion around pain signatures is contextualised within a broader continuous framework. We acknowledge the reviewer's comment on the limited description of the brain mechanisms of facial expression of pain. However, this was addressed in several previous reports of our laboratory (Kunz et al. 2011; Vachon-Pressseau et al. 2016; Kunz, Chen, and Rainville 2020). We have added some more details about the brain mechanisms of facial expression, and highlighted those references in the first paragraph of the introduction.

Methods and Results

(1) Was there any indication of power based on the previous work or the other signature papers? If yes, how that would inform the present analysis?

The NPS was trained on 20 participants that experienced 12 trials at each of four different intensities. The assessment of the effect sizes was performed on the Neurological Pain Signature in [Han et al. \(2022\)](#). That study revealed a moderate effect size for predicting between-subject pain reports, and a large one for predicting within-subject pain reports. We trained our model on 34 participants that underwent 16 trials. We expected our results to show a smaller effect size as the current experimental design only allowed us to examine spontaneous changes in the facial expression, as noted in the comments made by Reviewer #1. However, the best way to calculate the unbiased effect size of the results presented in the current study would be to test the unchanged model on new independent datasets (see [Reddan, Lindquist, and Wager, 2017](#)). Unfortunately, such datasets do not currently exist.

(2) I would clarify to the reader what is meant by normal range of thermal pain and why is this relevant. Also, I did not find data about this assessment nor about the assessment of facial expressiveness (or reference to where it can be found).

We changed this formulation to “All participants included in this study had normal thermal pain sensitivity” and we added a few references. By targeting a healthy population with normal thermal pain sensitivity, our study sought to identify a predictive brain pattern related to facial expression evoked by typical responses to pain that could eventually be generalised to other individuals from the same population. Details about the assessment of facial expressiveness have been added in the appropriate section in the Methods.

(3) That pain ratings are only weakly associated with facial responses is, in its own right, an interesting finding, as a naïve reader would expect the two to be highly positively correlated. I'd suggest discussing this aspect (in reference to previous research) as it is interesting on both theoretical and empirical grounds.

The likelihood and the strength of pain facial expression generally increase with pain ratings in response to acute noxious stimuli of increasing physical intensities, thereby leading to a positive association between the two responses that is driven by the stimulus. However, the

poor correlation or the dissociation between facial pain expression and pain rating is a very well known phenomenon that can be demonstrated easily using experimental methods where the stimulus intensity is held constant and spontaneous fluctuations are observed in both facial expression and pain ratings. This result was not discussed in the current manuscript as it was already addressed in the work of Kunz et al. (2011) and Kunz, Karos and Vervoot (2018). We added the references to these studies in the revised manuscript (lines 330 to 334).

(4) It may be worth having CIs throughout the whole set of analyses.

Thanks for the suggestions, this was an oversight. The confidence intervals have been added in the manuscript where applicable.

(5) I would clarify if there are two measures of the brain signature: dot-product and activation map. Relatedly, I cannot find where the authors explained what "FEPS pattern expression scores". Can the authors please clarify?

The clarification has been added in the manuscript (lines 413 to 414).

(6) There seems to be the assumption that the relationship between pain-relevant brain signatures and facial expressions of pain would be parametric and linear. However, this might not hold true. Did the authors test these assumptions?

We indeed decided to use a linear regression technique (i.e. LASSO regression) to model the association between the brain activity and the facial expression of pain. The algorithm choice was mainly based on the simplicity and the interpretability of that approach, and our limited number of observations. The choice was also coherent with previous studies in the domain (e.g. Wager et al., 2011; Wager et al., 2013; Krishnan et al. 2016; Woo et al., 2017). Using a linear model, we were able to predict above chance level the facial expression evoked by pain using the fMRI activation. However, it is legitimate to think that more complex non linear models can better capture the brain patterns predictive of that behavioural manifestation of pain.

(7) Did the authors assess whether the FACS were better to be transformed/normalised? More generally, I would report any data assessment/transformation that has not been reported.

Thank you for this highly relevant suggestion. FACS scores were indeed not normally distributed and the analyses were conducted again to predict the log transformed FACS scores. This transformation was effective to normalize the distribution (skewness = 0.75, kurtosis = -0.84). The predictive model was confirmed on transformed data.

(8) Page 12: I am not clear on whether all the signatures are included in the same model (like a multiple regression) or if separate regressions are calculated per signature. The authors seem to imply that several regressions have been computed (possibly one per comparison with each signature?).

The correlation between the FACS scores and the pain-related signatures was computed separately for each signature. This information has been clarified.

(9) MVPA: See my main comment about warm trials and experimental/statistical design. For example, the LASSO regression model for the pain trials could be compared with a model using warm trials besides (or instead of) the unfitted model. Otherwise, add the

warm trials as another predictor or within the subject level in a dummy fixed factor comprising pain and warm trials.

The inclusion of warm trials in the model training would be inconsistent with the goal of the main analysis to predict the facial expression of pain when a noxious pain stimulus is presented. Secondary analyses were conducted to compare the response of the FEPS to the warm trials compared to noxious pain trials. The dot product between the FEPS and the activation maps (FEPS scores) associated with the warm condition was computed. A linear mixed model was conducted to investigate the association between FEPS scores and the experimental condition (warm vs pain). Additional contrasts compared the warm trials with the pain trials with and without pain facial expression. The details of this analysis have been added to the manuscript (see Response of the FEPS to pain and warm in the Methods) as well as the corresponding results (see Results and Discussion).

(10) I would clarify for the reader why the separate M1 analysis has been run. Although obvious, I feel the reader would benefit from the specific hypothesis about this control analysis being spelled out together with the other statistical hypotheses within the statistical design in a more streamlined manner.

We extended the discussion on the rationale of that analysis and its interpretation taking into account the most recent results using the log transformed FACS scores (lines 125 to 133).

(11) The mixed model aimed to assess the relationship between pain ratings FEPS scores and facial scores is a crucial finding. I believe it speaks to the importance of a more complete design, which I already highlighted. I have a couple of technical questions: did the authors assess random slopes too? And, what was the strategy used to determine the random effects structure?

The linear mixed model considered the participants as a random effect, with random intercepts, considering the grouping structure in our data (i.e., each participant completed multiple trials). The reported results in the original manuscript were considering fixed slopes. However, following the reviewer's comment, we re-computed the mixed linear models allowing the slopes to vary according to the intensity ratings. The results were changed in the manuscript to represent the output of those models.

(12) The text from lines 63 to 67 could go in the methods.

We decided to include those lines within the Result and Discussion section to give the reader more specification about the FACS scores, as this term is subsequently referenced in the following part of the Results and Discussion section. We are concerned that putting this information only in the Methods section would disrupt the reading.

Reviewer #2 (Recommendations For The Authors):

p. 4-5. When you report the positive weight clusters, you follow up with a sentence specifying which cognitive processes those brain regions are typically associated with. However, when you report the negative weight clusters, you do not specify the cognitive processes typically associated with those brain areas. I think that providing that information would be helpful to the readers.

Thanks for noticing this omission. The information has been added in the most recent version of the manuscript (lines 119 to 121).

p. 9. You specify that the degree of expressiveness of participants was evaluated. How did you evaluate expressiveness? Did you use this variable in your analyses? Were participants excluded based on their degree of expressiveness?

Details about the assessment of facial expressiveness have been added in the appropriate section in the Methods (lines 285 to 289).

p. 10. You explain that two certified FACS-coders evaluated the video recordings to rate the frequency of AUs. Could you please provide more details about the frequency measure? I think that there are different ways in which this could have been done. For instance, were the videos decomposed into frames, and then the frequency measured by summing the number of frames in which the AU occurred? Or was it "expression-based", so one occurrence of an AU (frequency of 1) would correspond to the whole period between its activation onset and offset? Both ways have pros and cons. For example, if the frequency represents the number of frames, then it controls for the total duration of the AU activation within a trial (pro); but if there were multiple activations/deactivations of the AU within one trial, this will not be controlled for (con). And vice-versa with the second way of calculating frequency.

Details about the frequency scores have been added to the manuscript (lines 315 to 319).

p. 11. When you explained how you calculated the association between the facial expression of pain and pain-related brain signatures, I felt that there was some information missing. Did you use the thresholded maps (available in the published articles), or did you somehow have access to the complete, voxel-by-voxel, raw regression coefficient maps?

The unthresholded maps were used. The information has been clarified in the latest version of the manuscript, as well as the details about the availability of the maps (see Data Availability section at the end of the manuscript).

Reviewer #3 (Recommendations For The Authors):

Format

The authors will notice that many observations about the manuscript are related to missing information and a lack of graphical representations. I believe the topic and the content of the manuscript are too complex to condense into a short report.

Title

The claim of the title is simply not substantiated by the content of the manuscript. Demonstrating that the FEPS is a distinctive (i.e., specific) marker of pain processing requires a substantially different experimental design, with more rigorous controls and a broader set of painful stimulations. The manuscript would benefit from a more accurate title.

We agree that the title could better align with our findings. We modified the title accordingly : "A distributed brain response predicting the facial expression of acute nociceptive pain".

Abstract

I find it puzzling that the authors claim that there is limited knowledge of the neural correlates of facial expression of pain given what they describe in the first paragraph of

the introduction. Besides, they propose to reanalyze a dataset that has been extensively described in Kunz et al. (2011), which is unlikely to provide any new significant information.

We respectfully disagree with that comment. We considered that three articles (i.e., Kunz et al., 2011; Vachon-presseau et al., 2016; Kunz, Chen and Rainville, 2020) on the topic do constitute limited knowledge, especially if we compare it to the very large body of literature on the neural correlates associated with pain ratings. Except for these three studies, all the other citations pertain to behavioral studies on facial expression of pain, and do not examine the brain activity related to it. Furthermore, we believe that the complementary nature of the analyses performed in Kunz et al. (2011) and in this manuscript offers new insights into our understanding of facial expression in the context of pain. Indeed, the multivariate approach used in this study addresses some limitations present in Kunz et al. (2011) univariate analyses, mainly that it provides a quantifiable way to compare the similarity between different predictive patterns (Reddan and Wager, 2017). We submit that the assessment of the FEPS against several other pain-relevant signatures provides new and important information.

Furthermore, the abstract does not clearly state the aim, and the first line of the results does not match what the authors claim in the preceding line. The take-home message (last sentence) introduces the concept of a biomarker, which, as stated before, cannot be validated with the current data/experimental design. To put it in plain words, a given facial expression (or a composite score derived from a combination of expressions) cannot be a specific biomarker for pain, because a person can always mimic the same expression without feeling pain. Whether a given facial expression can be predicted from brain activity is a different issue, and whether that prediction can differentiate between painful and non-painful origins of the facial expression is another different issue. Unfortunately, neither of those issues can be tested with the current data/experimental design. The abstract would improve if the authors would circumscribe to what they actually tested, which is accurately described in the last sentence of the Introduction.

The abstract was revised accordingly. The term ‘biomarker’ was used in accordance with preceding studies in the field (see Reddan and Wager, 2017; Lee et al., 2021). Please note that we applied the same reasoning to fluctuations in pain expression as previous studies have applied to pain ratings. Of course, we can not dismiss the possibility of someone mimicking facial expressions. Similar reasoning applies to subjective reports, as individuals can intentionally overestimate their pain experience conveyed through verbal reports. This is another case of specificity testing that cannot be addressed in the present study (see new conclusion of the abstract and discussion of limitations). The challenge of pain assessment is a classical problem within both the scientific and the clinical literature. Here, we suggest that the consideration of multiple manifestations of pain is necessary to address this challenge and will provide a more comprehensive portrait of pain-related brain function.

Introduction

I believe that the Introduction would benefit from a strict definition of what is a marker/biomarker/neuromarkers (all those terms are used in the manuscript) and what are its desirable features (validity, reliability, specificity, etc.). I also believe that the Introduction (and the rest of the text) would benefit from a critical assessment of the term "signature". The Introduction describes four existing "signatures", all of them differing in the experimental condition in which acute nociceptive pain is studied, and proposes a fifth one. Keeping with the analogy, I'm wondering whether they should be called (pain) "signatures" if there is a different one for each experimental acute pain

condition, and they are so dissimilar between them when they are tested on the same condition (this dataset).

The last part of that comment raises fundamental methodological potential limitations that should be addressed in more depth in another article. That point goes beyond the scope of a research article. Regarding the stability aspect of the signatures, most of the signatures have not been studied extensively. It is thus difficult to currently assess their reliability. However, Han et al. (2022) showed high within-individual test-retest reliability for the NPS across eight different studies. Given that pain is a multidimensional experience, it is not surprising to find different patterns of activation predictive of different aspects or dimensions of the pain experience (see Čeko et al., 2022 for a similar discussion applied to negative affect).

The authors state that "As an automatic behavioral manifestation, pain facial expression might be an indicator of activity in nociceptive systems, perceptual and evaluative processes, or general negative affect." Doesn't it reflect all three of them? (and instead of or?) Why "might"?

The original sentence has been modified as follows: "As an automatic behavioral manifestation, pain facial expression is considered to be an indicator of activity in nociceptive systems, and to reflect perceptual and affective-evaluative processes" (lines 65 to 67).

Methods

The pain scale should be described. Kunz et al. used a 0-100 scale, where 50 was the pain threshold. This is crucial to interpret the 75-80/100 score for the painful thermal intensity.

The description of the pain scale has been added to the manuscript (lines 299 to 300).

Ratings for warm and painful temperatures should be reported (ideally plotted with individual-trial/subject data). In the same line of reasoning, FACS scores should be reported as well (ideally plotted with individual-trial/subject data). It would be interesting to explore the across-trial variability of pain ratings and FACS scores. That is, do people keep giving the same ratings and making the same facial expression after 16 trials? How much variability is between trials and between subjects?

The point raised in that comment was already addressed in response to a comment made by Reviewer #1 (also see the new Figures S2 and S4; see also lines 335 to 346).

How come only painful trials are analyzed? What if the FEPS signature was the same for warm and painful stimulation, thus reflecting the settings (fMRI experiment, stimulation, etc.) rather than the brain response to the stimuli?

The point raised in that comment was already addressed in response to a comment made by Reviewer #1. There was no pain expression in the warm trials and the FEPS shows no response to warm trials. This is now illustrated in the new Figure S4B (see also lines 138 to 158).

The authors propose to predict the trial-by-trial FACS composite score from the pain ratings using a LMM. However, it is interesting that they aim for an almost constant within- and between-subject pain score (75-80/100) as stated in the Methods. This should theoretically render the linear model invalid since its first (and main) assumption would be that FACS should vary linearly with the pain score. Even if patients were not aware

that the temperatures were constant across trials, the variation in pain scores should be explained by random noise for a constant stimulation intensity.

Reviewer #3 raises an important point that we need to clarify. Contrary to the expectation that FACS responses should be strongly correlated to pain ratings, we posited that these response channels depend at least in part on separate brain networks that may be differentially sensitive to a variety of modulatory mechanisms (attention, emotion, expectancy, motor priming, social context, etc.). This implies that part of the variance in FACS is independent from pain ratings. We, therefore, consider what Reviewer #3 refers to as random noise to be relevant and meaningful fluctuations reflecting endogenous processes influencing one's experience of pain and differentially affecting various output responses.

I noticed that fMRI data was analyzed with SPM5 in the original paper (Kunz et al., 2011) and with SPM8 in this manuscript. Was fMRI data re-processed for this manuscript? Were there any differences between the original analysis and this one that might induce changes in the interpretation of results?

The data were indeed re-processed using SPM8, which was the most recent version available when we started the analyses reported here. We used trial-by-trial activation maps for MVPA, which differs from what was used in the previous study (contrast maps at the level of the conditions, not the trials). We have no reason to believe that the different versions will change the message of this manuscript since those versions do not differ significantly in terms of the fMRI preprocessing pipeline (see SPM8 release notes; <https://www.fil.ion.ucl.ac.uk/spm/software/spm8/>). Furthermore, the aim of this present study is not to compare the different analysis parameters implemented in SPM5 vs SPM8.

What is the rationale for including PVP in the comparison among signatures? The experimental settings in which it was devised are distant from those described here.

The inclusion of the PVP was aimed at enhancing our comparative analysis with the FEPS, as we sought to investigate the potential functional meaning of the FEPS. The PVP was developed to capture the aversive value of pain, a dimension that is conceptually proximal to the interpretation of the facial expression as a manifestation of the affective response to nociceptive pain.

The LASSO-PCR approach is, in my opinion, not a procedure for (brain) decoding in this context. It is accurately described in the section title as a method for multivariate pattern analysis, or as a variable selection and regularization method for a prediction model. Here, brain activity in specific areas related to pain processing can hardly be described as "encoded", and the method just helps select those activations relevant for explaining a certain outcome (in this case, facial expressions).

We understand the point made by reviewer #3. The term brain decoding was changed for multivariate pattern analysis in the latest version of the manuscript.

Details are missing with regards to the dataset split into training, validation, and testing.

Details about the training and testing procedure were added in the manuscript (lines 383 to 385).

This might just be ignorance from me, so I apologize in advance, but what are "contrast" fMRI images? They are mentioned three times in the text but not really described. Are they the "Pain > Warm" contrasts from the original paper?

We apologize for any confusion caused by the use of the term “contrast images” which suggests a direct comparison between two experimental conditions. We have replaced “contrast images” with “activation maps” to provide a more accurate description of the nature of the data used in the multivariate pattern analysis (lines 388 to 389).

In the "Facial expression" section, the authors run an LMM to test the association between pain ratings (response variable) and facial responses (explanatory variable). If I understand correctly, in the "Multivariate pattern analysis" section they test the association between facial composite scores (response variable) and pain ratings (explanatory variable), but they obtain different results.

The analyses were recomputed on the log transformed data, as mentioned previously in the response to reviewers 1-2. The first model (in the “Facial expression” section) used the log transformed FACS scores as a dependent variable, the pain ratings as the fixed effect, and the participants as the random effect. The results of that analysis suggested that the transformed facial expression scores were not significantly associated with the pain ratings ($p = .07$). The second model uses both the FEPS pattern expression scores and pain ratings as fixed effects to predict facial responses. This analysis showed the significant contribution of the FEPS to the prediction of FACS scores ($p < .001$) and no significant effect of the pain ratings. However, a significant interaction was found ($p = .03$) suggesting that the prediction of the pain facial expression by the FEPS may vary with pain ratings (i.e. moderator effect). Those results have been clarified in the “Multivariate pattern analysis” section in the Methods (lines 416 to 426).

In this same section, what are "FEPS pattern expression scores"? They are used three times in the text, but I could not find their description.

The FEPS pattern expression scores correspond to the dot product between the trial-by-trial activation maps and the unthresholded FEPS signature. This information has been added to the manuscript (lines 413 to 414).

It would not be far-fetched to hypothesize that FACS scores could be predicted using solely activity from the motor cortex. The authors attempted to do this, but only with information from M1. Why did they not use the entire motor cortex, or better, regions of the motor cortex directly linked with the AUs described in the manuscript?

The selection of the primary motor area (M1) was based on the results found in Kunz et al. (2011). In this study, M1 showed the strongest correlation with facial expression of pain. There are numerous possibilities of combinations of multiple brain regions considering a variety of criteria based on distributed networks involved in motor, affective, or pain-related processes. We limited our exploration to the region with the strongest hypothesis due to practical feasibility concerns.

Results and Discussion

As a general recommendation, results should present individual data whenever possible. For example, the association between signatures and facial expression should be plotted using scatterplots.

We have added figures showing individual data when it was applicable (Figure S2; Figure S4).

The authors state that the LASSO-PCR model accounts for the facial responses to pain. I believe this is an overstatement, considering:

- A Pearson's r of 0.49 is usually considered low/weak correlation (moderate at best). In the same line, an R^2 of 0.17 means that only 17% of the variance is explained by the model.

More nuanced interpretation of the results has been added to the discussion. A section has been added to highlight the limitations of the study.

- Figure 1 needs to display individual subject data and the ideal regression line.

The model was trained using a k-fold cross-validation procedure. The regression lines thus represent the model's prediction for each one of the 10 folds (i.e. each fold is trained and tested on a different subset of the data). A scatter plot including the ideal regression line computed across all trials and subjects was added in supplementary material to illustrate the relation between the FACS scores and the FEPS pattern expression scores (Figure S4).

- Looking at Figure 1, it is clear that the model has an intercept different from zero. This means that when the FACS score was zero (i.e., volunteers did not make any distinguishable facial expression), the model predicted a score larger than zero. This is not discussed in the manuscript, and in simple terms, it means that there are brain activation patterns when no discernible facial expression is being made by the volunteers. In the original paper by Kunz et al., two groups of subjects were categorized, and one of them was a facially low- or non-expressive group ($n=13$). This fact is not even mentioned in the manuscript.

The categorization in the previous report (Kunz et al., 2012) was based on a pre-experimental session. All subjects were included in the current analysis. This is now indicated in the Methods (lines 287 to 289).

- On the other end of the range in Figure 1, differences between the FACS scores near the maximum range (40) are underestimated by 23 to 33 points! I guess that the RMSE is smaller (6-7 points), because many FACS scores are concentrated on the low end of the scale.

This is a very interesting comment. A section discussing the limits of the model to predict the lower and higher FACS scores has been added in the manuscript (lines 232 to 250).

It is of course acceptable to interpret the low similarity between signatures as a sign that each signature describes a different mechanism related to pain processing. However, I believe that a complete discussion should contemplate other competing hypotheses. Considering that all signatures were developed using a similar painful thermal stimulation protocol, it is reasonable to expect larger similarities between signatures. The fact that they are so dissimilar could be a reflection of model overfit, i.e., all these signatures are just fitted to these particular experimental protocols and data, and do not generalize to brain mechanisms of pain processing.

We appreciate the pertinent observation. We have included a limitations section in which we discussed, among other considerations, the possible overfitting of models and the necessity of pursuing generalizability studies (lines 225 to 268).

<https://doi.org/10.7554/eLife.87962.2.sa0>