Supplementary Information

for the manuscript "*Pain management decisions in emergency hospitals are predicted by brain activity during empathy and error monitoring*" by C. Corradi-Dell'Acqua, M. Foester, G. Sharvit, L. Trueb, E. Foucault, Y. Fournier, P. Vuilleumier & O. Hugli

Supplementary Methods

Detailed description of the neuroimaging tasks

ED nurses were engaged in a neuroimaging study involving three experimental paradigms testing empathy, risk taking, and error monitoring. The experiment took place in the Brain and Behaviour Laboratory (http://bbl.unige.ch/) of the University of Geneva. All participants met MRI safety requirements (no metallic objects in the body, no familial history of epilepsy, etc.), and were placed supine in the scanner with the head fixated by firm foam pads. Stimuli were presented on an LCD projector using either E-Prime 2.0 (Psychology Software Tools, Inc.) or Cogent 2000 (Wellcome Dept., London, UK), and were observed through a mirror mounted on the MRI headcoil. Key-presses were recorded on an MRI-compatible bimanual response button box. The paradigms employed were the following.

Empathy for pain. In this task, 120 colour pictures were presented depicting hands in either painful or non-painful situations. These pictures were sorted in four categories of 30 images each. *Painful* images described hands in pain, as visible by wounds/marks on the skin and by the display of an object (scalpel, syringe, etc.) acting on the skin surface. *Control* images were neutral stimuli matched with the previous category for hand laterality (right/left), orientation, and associated visual content (presence of objects), but purged from any painful/arousing feature. *Arousing* (and *ArousingControl*) images described hands in emotionally aversive (and matched neutral controls), but painless situations (hands holding knifes/guns, hands with handcuffs). Each of these 120 stimuli was presented for 2500 ms, followed by an inter-trial interval that ranged from 2500 to 4100 ms (mean = 3300 ms) with

incremental steps of 320 ms. Participants were asked to perform a handedness task, i.e., press one key if the stimulus depicted a right hand but another key if the stimulus was a left hand. The 120 stimuli were presented in a randomized order together with 30 null-events, in which an empty screen replaced the stimuli. This task was built using E-Prime 2.0 (Psychology Software Tools, Inc.) and lasted about 15 minutes.

Balloon Analogue Risk Task (BART). Nurses had to press a key repeatedly in order to inflate a virtual balloon as much as possible, and stop just before it exploded. If they stopped before explosion, they received a monetary gain proportional to the volume of air pumped; however they got nothing if the balloon exploded. Each trial started with an empty balloon placed on a tip of an inflator. The balloon could then be inflated for a maximum of 11 times through key-press, each of which was associated with an increasing probability of explosion (from 0% to 100%), but also an increased monetary reward (from 0.1 to 4.3 CHF). In this context, participants' choice to inflate the balloon led to two possible feedbacks: a larger balloon together with the graphical display of the current monetary gain (e.g., "+ 0.6 CHF"), or a "negative" feedback in which a picture of the balloon explosion was displayed together with the text "you have lost". At each step, participants were free to discontinue the inflation, which led to a "positive" feedback ("you have won"), and the money amount gained during this trial was added to their overall earning.

The experimental session comprised 28 independent trials, each separated by an inter-trial interval ranging between 2000 and 4000 ms. Within each trial, different inflations were separated by an interval ranging between 1500 and 2500 ms. During that time, the inflator was coloured in red, to signal participants to withhold any response until it got green. Win/loss

feedbacks lasted 2000 ms. Once every six trials, participants were displayed the question “to which extent this task makes you feel anxious?” together with a visual analog scale that ranged from “not anxious at all” to “extremely anxious”. Participants had 5 seconds to slide a marker on the bar until it reached the position corresponding to their judgment. The overall experimental session never exceeded 15 minutes. This task was built using Cogent 2000 (Wellcome Dept., London, UK).

Social Harm Avoidance Monitoring Experiment (SHAME)***.*** The nurse inside the MRI scanner took turns with a colleague outside the scanner in performing and observing a dot-counting task. The colleague was one of the other nurses of the experimental group, who was matched with the participant based on availability (with no further constraint in terms of seniority or interpersonal closeness). The experiment was organized in 14 blocks, 7 in which the subject in the scanner performed the task (whilst the subject outside the scanner observed the same display), alternating with 7 in which the task was performed by the subject outside. Each block comprised 5 trials, each starting with the simultaneous presentation of two clusters of white dots on a grey background separated by a vertical line (duration 500 ms, with inter-stimulus interval ranging from 1.5–3.5 s). The participants in charge of playing indicated which side contained the largest amount of dots. The trial difficulty was adjusted on-line throughout the whole experiment (at participant unawareness) to ensure a comparable amount of correct and incorrect trials for each participant[1].

Critically, each response was followed by a thermal stimulation given to the arm of the participant outside the scanner (3 s of raise time, 2 s of plateau, 3 s for returning to baseline). Correct responses were always followed by a painless temperature (38°C), whereas incorrect

responses had 50% of probability to be associated with a painless or painful temperature (set by subject-specific pain threshold; average 48°C ± 1.45; see below for details). The thermal stimulus was associated with a visual feedback (5 s) informing about the performance in the task and the painfulness of the temperature, and was followed by inter-trial-interval of variable duration (from 2-9s). The thermal stimulation was delivered through a MSA Thermotest thermode (SOMEDIC Sales AB, Sweden). This task was built using Cogent 2000 and lasted about 12 minutes.

Pain Thresholding Procedure

In line with previous studies, individual temperatures were determined through a double random staircase (DRS) algorithm[2 3]. Our DRS procedure selected a given temperature on each experimental trial according to the previous response of the participant in a pain unpleasantness rating scale. Trials rated as more unpleasant than the given cut-off (corresponding to 8 out of 10 on a visual analogic scale) led to a subsequent lower temperature in the next trial; whereas trials rated as less unpleasant than the given cut-off led to a subsequent higher temperature. This resulted in a sequence of temperatures that rapidly ascended towards, and subsequently converged around, a subjective pain unpleasantness threshold, which was in turn calculated as the average value of the first four temperatures leading to a direction change in the sequence. In order to avoid participants anticipating a systematic relationship between their rating and the subsequent temperature, two independent staircases were presented randomly. Initial thermal stimulations for the two staircases were 41°C and 43°C. Within each staircase, stimulus temperatures increased or

decreased with steps of 3°C, while smaller changes (1°C) occurred following direction flips in the sequence. None of our subjects was stimulated at temperature larger than 52°C.

Post-Scanning quantitative debrief session

The nurses taking part to the neuroimaging study were subjected to a post-scanning debriefing session of about 50 minutes, which comprehended standardized tests as well as a set of custom-based items. In particular, participants underwent the *inclusion of other in the self* (IOS) scale[4] to assess to which extent the felt close with the "colleague" engaged in the SHAME outside the MRI scanner. Furthermore, participants also rated the degree to which they felt particular emotions (pain, fear, shame, guilt, sadness, and anger) when they were engaged in the SHAME in the MRI scanner (but not when they were performing the task as confederates outside the scanner). All ratings were carried out on a Liker scale ranging from 1 to 5, with the exception of the rating of pain which was carried out on a verbal numeric rating scale ranging from 1 to 10. Participants were asked to rate their subjective experience associated with any kind of error event: i.e., they did a mistake or when they observed their colleague making a mistake, either with a painful or painless outcome. See Koban et al.[1] for more details about this rating session.

Finally, participants were asked to rate each of the 120 stimuli used in the Empathy for Pain paradigms in terms of familiarity ("*how much is the content described in this picture familiar to you?*"), emotional intensity ("*how intense is the emotion triggered by this image?*"), emotional valence (*"does this image elicit positive or negative emotions?"*), and pain ("*how intense is the pain felt by the hand depicted on this image?*"). The rating session was carried

using E-Prime 2.0, and divided in four blocks, one for each question, during which all 120 stimuli were rated on a Likert scale rating from 1 to 10 (with the exception of the emotional valence rating, in which a Likert scale ranging from -4 to +4 was used). To avoid habituation biases due to the presentation of the same stimuli four times, the order of the blocks and order of the stimuli within each block was randomized across participants. See Corradi-Dell'Acqua et al.[5] for more details about the subjective rating session.

Imaging processing

Data Acquisition. Functional images were acquired using a 3T whole-body scanner (Trio TIM, Siemens) with a 32-channel head coil. We used a multiplex sequence[6], with TR = 650 ms, TE = 30 ms, flip angle = 50°, 36 interleaved slices, 64 x 64 in-slice resolution, 3 x 3 x 3 mm voxel size, and 3.9 mm slice spacing. The multiband accelerator factor was 4, and parallel acquisition techniques (PAT) was not used. A fieldmap was also estimated through the acquisition of 2 functional images with a different echo times (short TE = 5.19 ms; long TE = 7.65). Finally, a structural image was acquired using a T1 weighted 3D sequence (MPRAGE, TR = 1900 ms, TI = 900 ms, TE = 2.27 ms, flip angle = 9°, PAT factor = 2, 192 sagittal slices, 1 x 1 x 1 mm voxel size, 256 x 256 in-slice resolution.

Preprocessing. Statistical analysis was performed using the SPM12 software (http://www.fil.ion.ucl.ac.uk/spm/). For each subject, all functional images were realigned and unwrapped using a field map image, to account for geometric distortions due to magnetic field inhomogeneity. Subsequently the functional images were normalized to a template based on 152 brains from the Montreal Neurological Institute (MNI) with a voxel-size resolution of 3 x 3 x

3 mm, using a deformation field estimated on a coregistered structural image. Finally, the normalized images were smoothed by convolution with an 8 mm full-width at half-maximum Gaussian kernel.

General Linear Model. Preprocessed images from each task were analysed using the General Linear Model (GLM) framework implemented in SPM, consistently with previous studies using the same paradigms[1 5 7 8]. For the Empathy for Pain task, trial time onsets from each of the four conditions were modelled with a delta function. Additionally, for each condition we also included an additional vector in which participants Response Times were modulated parametrically[5]. For the BART, we modelled with a delta function all inflation events (in which participants were prompted a decision), with the probability of explosion fed as additional parametric regressor. Furthermore, we also modelled positive (win) and negative (loose) feedbacks as separate regressors[7 8]. For the SHAME we modelled, separately for each player, all trials in which participants were prompted with a judgment, as well as all kinds of feedback (correct, incorrect painless, incorrect painful) with separate delta function[1].

For all tasks, we accounted for putative habituation effects in neural responses of each condition by using the time-modulation option implemented in SPM, which creates a regressor in which the block/trial order is modulated parametrically. Furthermore, each regressor was convolved with a canonical hemodynamic response function and associated with its first order temporal derivative. To account for movement-related variance, we included six differential movement parameters as covariates of no interest. Low-frequency signal drifts were filtered using a cutoff period of 128 sec. Serial correlations in the neural signal were accounted through exponential covariance structures, as implemented in the 'FAST' option of SPM12. Global

scaling was applied, with each fMRI value rescaled to a percentage value of the average whole-brain signal for that scan.

Functional contrasts, testing differential parameter estimates images associated with one experimental condition *vs.* the other were then fed in a second level, one-sample t-test using random-effect analysis. Similarly, parameter estimates of conditions of interest were fed to univariate linear regressions, using one of the three clinical measures of interest as predictors. Effects were considered significant if exceeded $p < 0.05$, family-wise correction for multiple comparisons at the cluster level (with an underlying height threshold of $p < 0.001$, uncorrected). In addition, we report also effects surviving $p < 0.05$ small volume corrected for masks of interests, defined through by previous studies in which independent lay populations were engaged in the same studies implemented here[1 5 8].

Region of interest masks. For each task, we identified an inclusive mask, comprehending only those coordinates of theoretical interest, as obtained by reanalysing data from independent researches employing the same paradigms[1 5 8], under similar preprocessing and modelling settings of the current study. The only exceptions were related to those datasets in which a field map image was not available, for which no unwrapping was applied during the preprocessing stage. Consequently, in these cases, the deformation field for the normalization was estimated directly from the functional images (instead from a coregistered structural volume), to minimize the impact of geometric distortions related the magnetic field inhomogeneity[9]. Furthermore, as all these previous datasets were acquired with long repetition time (> 2 sec), serial autocorrelations were accounted with standard first-order autoregressive AR(1) model (as opposed to the FAST option for rapid sequences).

For the Empathy for Pain task, we reanalysed the data from Corradi-Dell'Acqua et al.[5]. As this previous study employed larger database of pictures than ours, we considered as conditions of interest those parameters estimated on the same sub-selection of images which were used for the present study (the remaining images were modelled as separate conditions of no interest), which were then used to identify brain regions associated with the contrast *Painful – Control* (Table S1). For the BART, we reanalysed the study of Schonberg et al.[8] (freely available at https://openneuro.org/datasets/ds000001/). This study included an additional control condition characterized by the inflation of balloon without risk of money loss[8], which was modelled in the first level as separate regressor of no interest. In keeping with results reported by the original study[8], we took into consideration regions implicated in the contrast *monetary loss – implicit baseline* (in this specific dataset, no differential effect between *monetary loss – win* were reported; see Table S2). Finally, for the SHAME, we reanalysed the data from Koban et al.[1] and selected features implicated in one's erroneous performance with painful outcome, compared with others' erroneous trials with painful outcome (*One's – Others' painful errors*; see Table S3). In all these mask, we followed previous studies implementing similar multivariate regression on whole brain data[10 11], and excluded the coordinates in occipital cortex that may be driven by distinctive visual features rather than on the information of interest.

Multivariate Regression

We used Least Absolute Shrinkage and Selection Operator (LASSO)[10–13] and Random Forest (RF) regression[14] to identify distributed patterns of activity across brain that could be predictive of nurses' professional behaviour in Emergency Department.

Feature Selection. For each task, we identified an inclusive mask, comprehending only those coordinates of theoretical interest, as obtained from independent datasets in which lay participants were engaged in similar paradigms than those used here[1 5 8] (see above).

LASSO. For each of these independently-defined masks, we extracted the neural activation associated with corresponding tasks in the present study. The resulting data matrixes (e.g., for the Empathy for Pain Task, 33 nurses x 1077 voxels) were then fed into a LASSO routine (*lasso* function from Matlab R2013b) to identify which components were jointly predictive of nurses' behaviour in their clinical practice (as recorded during the preceding 15 months). To optimize the modelling, but at the same time insure its generalizability to new data, the LASSO regression was conducted in two nested 10-fold cross-validation loops. The first (inner) was aimed at optimizing regularization hyper-parameter λ. The second (outer) was aimed at predicting professional behaviour of a portions of nurse by a model optimized (also in its hyper-parameters) on out-of-sample nurses.

Random Forest Regression. The same data matrixes were also fed to the regression routines implemented in the Matlab-based RF toolbox[14 15]. This analysis involves the implementation of decision trees, which perform recursive partitioning of the neural (feature) space to lead to a non-linear predictive model. As decision tree-based models are susceptible to small perturbations in the dataset, the variance in estimated prediction function was reduced by the RF algorithm through 1000 bootstrap resampling of the original dataset, each of which provides its own contribution (or vote) to the final prediction[14]. Generalizability of the regression was conducted through a 10-fold cross-validation loop, in which the model optimized in a portion of nurses was tested on out-of-sample nurses.

Permutation Analysis. The proficiency of the LASSO and RF procedures was assessed by estimating the mean squared error (MSE), reflecting the deviation between nurses' actual behaviour and the behaviour estimated from the brain activity. This value was considered significant if lower than the 5^{th} percentile of the distribution of 1000 MSEs obtained by rerunning the whole procedure on permuted datasets.

Supplementary Results

Empathy for Pain Task

Tables S4-5 report all behavioural (Table S4) and neuroimaging (Table S5) results associated with this task. Behavioural effects include on-line accuracy and response time, but also post-experimental rating sessions. Consistent with a previous study using the same paradigm[5], negative images (both painful and painless arousing) were associated with longer response times, greater arousal, and lower accuracy, valence and familiarity, with respect to their tailored controls (paired *t*-test: $|t| \geq 1.94$, p (1-tailed) ≤ 0.030; Wilcoxon sign-rank test: $|Z| \geq 2.38$, $p \leq 0.017$). In addition, painful images were associated with higher ratings of harm/pain, than both their controls and painless arousing images ($|t| \geq 7.84$, $p < 0.001$; $|Z| \geq 4.25$, $p < 0.001$). However, unlike in our previous study on lay population, painful images were rated by emergency nurses as more familiar, less negatively-valenced, and less arousing than painless arousing images (the same for their corresponding controls – $|t| \geq 2.40$, $p \leq 0.027$; $|Z| \geq 2.24$, $p < 0.025$).

We then tested whether the behavioural responses to painful images could be predictive of the three clinical indexes of interest. For this purpose we took both online (median response times and accuracy) and offline (post-scanning ratings) measures for the condition of interest (displayed in Table S4), and subjected them to massive univariate linear regression, which led to no significant effects ($|r| \leq 0.28$, *n.s.*). When feeding all six measures to multivariate regression using the same routines used for the analysis of brain data, we found that a reliable prediction of the treatment application could be obtained using RF decision trees (see Table S11). None of the other two indexes could be reliably predicted.

Balloon Analog Risk Task (BART)

Tables S6-7 report all behavioural (Table S6) and neuroimaging (Table S7) results associated with this task. Behavioural data refer to the number of inflation in win trials, median response times, the amount of money gained in the overall session, and the median subjective rating. For each of these four measures, a linear relationship with the clinical indexes of interest was tested with no significant results ($|r| \leq 0.30$, *n.s.*). Furthermore, when feeding all four behavioural measures to multivariate regression using the same LASSO and RF approaches employed for the analysis of brain signal, we found no reliable prediction (see Table S11).

Social Harm Avoidance Monitoring Experiment (SHAME)

Tables S8-9 report all the behavioural effects associated with the SHAME. Post-scanning ratings obtained from all four kinds of errors suggest that nurses felt greater empathic pain, but also greatest sadness, when observing an error with painful (vs. painless) outcome. Instead, greater shame, guilt and anger were reported when nurses committed themselves an error (regardless of its painful/painless outcome) relative to when they observed the confederate in the scanner committing an error.

We then tested whether the behavioural responses to SHAME could be predictive of the three clinical indexes of interest. For this purpose we took the ratings of pain, guilt, shame, fear, sadness and anger (acquired in the post-scanning session) associated with One's Painful Errors (see Table S9). Furthermore we also took into consideration three online measures from when the subject was playing the task in the MRI (see Table S8): the median response times, accuracy, and median trial difficulty (as difficulty was automatically adapted according to

participants' performance, the median trial difficulty throughout the experimental session is an indirect measure of how challenging the task was for each subject). For each of these nine measures, a linear relationship with the clinical indexes of interest was tested with no significant results ($|r| \leq 0.31$, *n.s.*). When feeding all 9 measures to multivariate regression using the same routines used for the analysis of brain data, we found that a reliable prediction of the Documentation Rate could be obtained using RF decision trees (see Table S11). None of the other two indexes could be reliably predicted.

Finally, Table S10 reports the regions involved when participants observed harmful consequences of their own errors, relative to the condition in which pain was self-caused by the colleague outside the scanner. We assessed whether these responses could be influenced by the personal/professional relationship between the pair of nurses engaged in the task. Personal closeness was assessed by the IOS questionnaire[4] as implemented in the post-scanning debrief, whereas professional closeness was assessed by calculating the absolute difference in age and years of experience between the two nurses engaged in the task (no difference reflects stronger similarity in professional status than a large difference). We then run a univariate linear regression analysis, in which the neural responses to one's painful errors were fitted against each of these three measures (IOS, age difference, experience differences). No significant effects of personal/professional closeness were found.

Supplementary References

1. Koban L, Corradi-Dell'Acqua C, Vuilleumier P. Integration of error agency and representation of others' pain in the anterior insula. *J Cogn Neurosci* 2013; **25**: 258–72

2. Corradi-Dell'Acqua C, Tusche A, Vuilleumier P, Singer T. Cross-modal representations of first-hand and vicarious pain, disgust and fairness in insular and cingulate cortex. *Nat Commun* 2016; **7**: 10904

3. Sharvit G, Corradi-Dell'Acqua C, Vuilleumier P. Modality-specific effects of aversive expectancy in anterior insula and medial prefrontal cortex. *Pain* 2018; **159**: 1529–1542

4. Aron A, Aron EN, Smollan D. Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *J Pers Soc Psychol* 1992; **63**: 596–612

5. Corradi-Dell'Acqua C, Hofstetter C, Vuilleumier P. Felt and Seen Pain Evoke the Same Local Patterns of Cortical Activity in Insular and Cingulate Cortex. *J Neurosci* 2011; **31**: 17996–8006

6. Feinberg DA, Moeller S, Smith SM, et al. Multiplexed Echo Planar Imaging for Sub-Second Whole Brain FMRI and Fast Diffusion Imaging. *PLOS ONE* 2010; **5**: e15710

7. Rao H, Korczykowski M, Pluta J, Hoang A, Detre JA. Neural correlates of voluntary and involuntary risk taking in the human brain: an fMRI Study of the Balloon Analog Risk Task (BART). *NeuroImage* 2008; **42**: 902–10

8. Schonberg T, Fox CR, Mumford JA, Congdon E, Trepel C, Poldrack RA. Decreasing ventromedial prefrontal cortex activity during sequential risk-taking: an FMRI investigation of the balloon analog risk task. *Front Neurosci* 2012; **6**: 80

9. Calhoun VD, Wager TD, Krishnan A, et al. The impact of T1 versus EPI spatial normalization templates for fMRI data analyses. *Hum Brain Mapp* 2017; **38**: 5331–42

10. Chang LJ, Gianaros PJ, Manuck SB, Krishnan A, Wager TD. A Sensitive and Specific Neural Signature for Picture-Induced Negative Affect. *PLoS Biol* 2015; **13**: e1002180

11. Krishnan A, Woo C-W, Chang LJ, et al. Somatic and vicarious pain are represented by dissociable multivariate brain patterns. *eLife* 2016; **5**: e15166

12. Wager TD, Atlas LY, Lindquist MA, Roy M, Woo C-W, Kross E. An fMRI-Based Neurologic Signature of Physical Pain. *N Engl J Med* 2013; **368**: 1388–97

13. Zunhammer M, Geis S, Busch V, Eichhammer P, Greenlee MW. Pain modulation by intranasal oxytocin and emotional picture viewing — a randomized double-blind fMRI study. *Sci Rep* 2016; **6**: 31606

14. Breiman L. Random Forests. *Mach Learn* 2001; **45**: 5–32

15. Jaiantilal A. Classification and regression by randomforest-matlab [Internet]. 2009. Available from: https://code.google.com/archive/p/randomforest-matlab/

16. Friston KJ, Worsley KJ, Frackowiak RSJ, Mazziotta JC, Evans AC. Assessing the significance of focal activations using their spatial extent. *Hum Brain Mapp* 1993; **1**: 210–20

17. Chumbley JR, Friston KJ. False discovery rate revisited: FDR and topological inference using Gaussian random fields. *NeuroImage* 2009; **44**: 62–70

Supplementary Tables

Table S1

Region of Interest mask for the Empathy for Pain Task. Regions included in both univariate and multivariate analysis defined from independent data[5]. The table lists the regions displaying differential activity for the contrast *Painful – Control*.

	SIDE	*Coordinates*			$T_{(27)}$	*Cluster size*
		X	*y*	*z*		
Painful – PainfulControl						
Anterior Insula	R	39	29	-4	3.93	66*
Inferior Frontal Gyrus	R	42	38	5	6.21	
Middle/Posterior Insula	R	42	5	-7	6.85	260***
Amygdala	R	21	-4	-16	7.25	
Anterior Insula	L	-33	23	-1	5.97	
Middle/Posterior Insula	L	-36	-4	8	4.26	292***
Amygdala	L	-21	-4	-19	5.29	
Supramarginal	R	60	-22	38	8.56	301***
Postcentral Gyrus	R	63	-16	29	7.08	
Supramarginal	L	-54	-25	35	7.07	132***
Precentral Gyrus	R	45	8	26	5.82	137***

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$ family-wise corrected for the whole brain

Table S2

Region of Interest mask for the Balloon Analog Risk Task. Regions included in both univariate and multivariate analysis defined from independent data[8]. The table lists the regions displaying differential activity for the contrast *monetary loss – implicit baseline*.

	SIDE	***Coordinates***			***$T_{(15)}$***	***Cluster size***
		x	***Y***	***z***		
Monetary Loss – Implicit Baseline						
Anterior Insula	R	42	23	-4	7.74	184***
Ventral Insula	R	30	14	-19	4.74	
Anterior Insula	L	-39	17	-1	5.30	156**
Ventral Insula	L	-33	14	-16	6.20	
Pre-central Gyrus	R	45	11	32	6.54	167**
Middle Frontal Gyrus	R	42	23	47	5.62	
Inferior Frontal Sulcus	R	45	35	23	4.98	67*
Thalamus/Midbrain	M	-5	-31	37	5.77	85*

*$p < 0.05$ family-wise corrected for multiple comparisons at the cluster level

Table S3

Region of Interest mask for SHAME. Regions included in both univariate and multivariate analysis defined from independent data[1]. The table lists the regions displaying differential activity for the contrast *One's – Others' Painful Errors*. All clusters are displayed with a height threshold corresponding to $p < 0.001$ (uncorrected), and survive FWE[16] or FDR[17] correction for multiple comparisons for the whole brain at the cluster level.

	SIDE	***Coordinates***			***$T_{(15)}$***	***Cluster size***
		x	***y***	***z***		
One's – Others' Painful Errors						
Anterior Insula	R	45	14	-7	5.21	58†
Putamen	R	30	5	-1	5.22	71*
Anterior Insula	L	-39	11	-4	3.91	114**
Putamen	L	-21	8	5	5.60	
Superior Frontal Gyrus	R	36	38	53	5.38	140***
Superior Frontal Gyrus	L	-21	53	26	6.60	123**
Cerebellum	R	27	-49	-28	5.24	59*
Anterior Middle Cingulate Cortex	M	6	23	23	7.49	121**
Supplementary Motor Area	M	-9	14	58	7.70	111**

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$ family-wise corrected for the whole brain; † $p < 0.05$ false-discovery-rate corrected for the whole brain

Table S4

Behavioural data from the Empathy for Pain Task. For each measure of interest, the average value associated with the four conditions is displayed together with 95% confidence intervals.

	Painful	***Control***	***Arousing***	***ArousingControl***
Resp. Times (ms)	1426 [1341, 1539]	1362 [1273, 1445]	1441 [1357, 1551]	1329 [1255, 1409]
Accuracy (%)	70.97 [64.02, 75.57]	83.97 [73.62, 89.61]	72.17 [64.81, 76.68]	81.55 [72.54, 87.06]
Arousal	6.15 [5.27, 7.04]	1.81 [1.48, 2.29]	7.01 [6.45, 7.64]	2.71 [2.24, 3.24]
Valence	-0.73 [-1.20, -0.25]	1.23 [0.94, 1.69]	-2.11 [-2.42, -1.78]	0.90 [0.65, 1.22]
Familiarity	5.29 [4.45, 6.12]	8.72 [8.27, 9.12]	2.24 [1.82, 2.75]	4.63 [4.05, 5.20]
Pain	8.92 [8.44, 9.28]	2.23 [1.58, 3.28]	3.26 [2.16, 4.69]	1.39 [1.13, 1.90]

Table S5

Neural Activations for the Empathy for Pain Task. Regions displaying differential activity for the contrast Painful – Control Images, and increased activity for Painful Images with nurses' Documentation Rate. All clusters survive correction for multiple comparisons for the whole brain at the cluster level [16], or small-volume correction for a region of interest mask (described in Table S1).

	SIDE	Coordinates			$T_{(32)}$	Cluster size
		X	Y	Z		
Painful – Control Images						
Middle/Posterior Insula	R	42	-7	-1	6.64	256***
Amygdala	R	24	-4	-16	8.39	
Middle/Posterior Insula	L	-39	-1	-7	3.89	96*
Amygdala	L	-21	-7	-13	7.15	
Inferior Frontal Gyrus	R	45	38	11	8.71	133**
Inferior Frontal Gyrus	L	-42	32	14	6.94	116**
Precentral Gyrus	R	48	8	26	6.87	160**
Precentral Gyrus	L	-45	5	23	7.38	167**
Inferior Temporal Gyrus	R	51	-55	-10	9.23	
Fusiform Gyrus	R	30	-49	-13	9.07	
Calcarine Gyrus	R	18	-94	-4	10.02	
Intraparietal Sulcus	R	27	-64	44	8.04	
Supramarginal/Postcentral Gyrus	R	63	-22	26	9.39	2983***
Inferior Temporal Gyrus	L	-45	-55	-7	7.86	
Fusiform Gyrus	L	-30	-49	-16	8.90	
Calcarine Gyrus	L	-21	-94	-5	10.24	
Periaqueductal Grey/Midbrain	M	-3	-31	-4	6.68	
Intraparietal Sulcus	L	-21	-64	44	5.24	87*
Supramarginal/Postcentral Gyrus	L	-63	-25	32	8.20	133**
Painful Images*Documentation Rate						
Postcentral Gyrus	R	60	-16	32	4.50†	18
Middle Occipital Gyrus	R	36	-73	14	4.51	89*

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$ family-wise corrected for the whole brain; † $p < 0.05$ family-wise corrected for small volume

Table S6

Behavioural data from the Balloon Analog Risk Task. In keeping with previous studies, we measured the average number of inflations in each trial (excluding trials associated with negative outcome), the response times associated with each choice, and the overall money gained during the experimental session. Furthermore, we also considered subjects' median anxiety rating collected along the whole experimental session. For each measure of interest, the average value is displayed together with 95% confidence intervals.

# inflations	***Response Times (ms)***	***Money gained (CHF)***	***Anxiety [0-100]***
6.03 [5.79, 6.26]	584 [536, 643]	50.08 [45.23, 55.02]	31.25 [25.33, 37.42]

Table S7

Neural Activations for the Balloon Analog Risk Task. Regions displaying differential activity for the contrast monetary loss – win, and increased activity for Monetary Loss with nurses' CI Rate.

	SIDE	*Coordinates*			$T_{(31)}$	*Cluster size*
		x	*Y*	*z*		
Monetary Loss – Win						
Anterior Insula	R	39	17	-4	9.79	
Ventral insula	R	39	-1	-16	10.97	603***
Posterior Insula	R	39	-10	-7	6.36	
Anterior Insula	L	-36	20	-7	11.74	
Ventral Insula	L	-27	8	-22	9.54	493***
Posterior Insula	L	-39	-4	-13	5.74	
Temporo-Parietal Junction	R	60	-40	20	4.65	68*
Middle Occipital Gyrus	L	-30	-88	14	4.93	168***
Inferior Occipital Gyrus	L	-45	-73	2	5.51	
Precentral/Postcentral Gyrus	R	39	-16	50	3.97	
Supplementary Motor Area	M	9	8	62	8.19	1025***
Middle Cingulate Cortex	M	-6	17	35	7.78	
Calcarine Cortex	R	18	-64	14	5.06	
Lingual Gyrus	R	27	-61	-7	8.74	
Middle Temporal Gyrus	R	46	-64	-4	5.09	1802***
Calcarine Cortex	L	-15	-70	11	6.10	
Fusiform Gyrus	L	-27	-67	-7	7.67	
Thalamus	M	9	-28	-7	8.08	
Monetary Loss*CI Rate						
Ant. Insula/Inf. Frontal Gyrus	R	24	32	-10	4.96	170***
Mid. Insula-Opercular Junction	R	21	-13	14	5.06	69*
Putamen	R	30	-4	11	4.98	
Ant. Insular-Opercular Junction	L	-27	29	20	4.90	156**
Mid. Insula-Opercular Junction	L	-36	5	17	5.87	128**
Hippocampus	R	30	-28	-7	4.55	104**
Thalamus	R	21	-34	-1	4.21	
Angular Gyrus	R	39	-52	20	4.45	63*
Middle Occipital Gyrus	R	45	-76	23	4.95	120**
Middle Occipital Gyrus	L	-33	-64	17	5.02	88*
Cerebellum	R	12	-37	-34	5.49	759***
Cerebellum	L	-42	-58	-37	4.91	

Middle Cingulate Cortex	M	-6	-4	29	5.33	261***

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$ family-wise corrected for the whole brain.

Table S8

Behavioural data from the SHAME: online measures. For each measure of interest, the average value associated with the four conditions is displayed together with 95% confidence intervals.

	Difficulty (dots diff)	***Accuracy (%)***	***Resp. Times (ms)***	***Pain Threshold (deg)***
Subject in the scanner (self)	4.75 [4.57, 4.88]	52.01 [47.28, 54.27]	928 [848, 1001]	---
Subject outside the scanner (other)	4.84 [4.61, 4.97]	51.85 [50.11, 55.17]	953 [877, 1024]	48.31 [47.78, 48.75]

Table S9

Behavioural data from the SHAME: post-scan rating measures. For each measure of interest, the average value associated with the four kinds of errors is displayed together with 95% confidence intervals. “Self” refers to the case in which the participant in the MRI scanner made a mistake, whereas “Other” refers to the case in which the participants in the MRI scanner observed the confederate making a mistake. “Pain” and “NoPain” refer to errors with painful and painless outcomes respectively.

	Pain	***Fear***	***Shame***	***Guilt***	***Sadness***	***Anger***
Self-Pain	2.61 [2.07, 3.35]	1.93 [1.50, 2.50]	2.81 [2.29, 3.30]	3.77 [3.13, 4.23]	2.42 [1.93, 2.94]	3.26 [2.68, 3.80]
Other-Pain	2.61 [2.03, 3.39]	2.45 [1.91, 3.03]	1.48 [1.22, 1.97]	1.42 [1.16, 1.85]	2.29 [1.76, 2.88]	1.74 [1.35, 2.29]
Self-NoPain	1.64 [1.28, 2.07]	1.45[1.16, 1.94]	2.52 [2.03, 3.03]	2.84 [2.28, 3.37]	1.74 [1.37, 2.23]	2.58 [2.06, 3.10]
Other-NoPain	1.45 [1.16, 1.85]	1.64 [1.30, 2.17]	1.42 [1.16, 1.91]	1.42 [1.14, 1.94]	1.84 [1.45, 2.34]	1.35 [1.07, 1.85]

Table S10

Neural Activations for the SHAME. Regions displaying differential activity for the contrast One's – Other's Painful Errors, and increased activity for One's Painful Errors with nurses' CI Rate. All clusters are displayed with a height threshold corresponding to p < 0.001 (uncorrected), and survive correction for multiple comparisons for the whole brain, or small-volume correction for a region of interest mask (described in Table S3).

	SIDE	***Coordinates***			***$T_{(30)}$***	***Cluster size***
		x	***y***	***z***		
One's – Others' Painful Errors						
Anterior Middle Cingulate Cortex	M	-3	26	29	4.35†	25
One's Painful Errors*CI Rate						
Middle Frontal Gyrus	L	-33	8	53	6.79	104*
Anterior Middle Cingulate Cortex	M	3	44	14	5.94	131**

$^{**}p < 0.01$ $^{*}p < 0.05$ family-wise corrected for the whole brain; † $p < 0.05$ family-wise corrected for small volume

Table S11

Multivariate modelling of behavioural data. For each task, we modelled brain activity with multivariate regression based on LASSO and Random Forest (RF) approaches. The analysis of each task is described in terms of number of features fed to the multivariate regression, as well as measures of model's proficiency (Mean Square Error [*MSE*]) at predicting each of the three indexes from the delegated analgesia protocol. Significant predictions are highlighted in bold, and significance cut-offs (5th percentile of a permutation-based *MSE* distribution) are displayed in squared brackets. Full details in Supplementary Methods.

Task	*# Features*	*Algorithm*	*Docum Rate*	*CI Rate*	*Treatment App.*	
From the neuroimaging session						
Pain Images	6	*LASSO*	2.48 [2.23]	1.79 [1.55]	1.73 [1.49]	$\cdot 10^{-3}$
		RF	2.28 [2.13]	1.92 [1.44]	**1.29*** [1.37]	
Monetary Loss	4	*LASSO*	3.44 [3.10]	1.81 [1.71]	2.12 [1.87]	
		RF	3.91 [2.91]	1.93 [1.60]	2.24 [1.79]	
One's Painful Errors	9	*LASSO*	3.23 [3.14]	1.93 [1.68]	2.18 [1.98]	
		RF	**2.86*** [2.97]	2.10 [1.61]	2.30 [1.82]	

* error lower than chance at $p < 0.05$