SUPPLEMENTARY METHODS

Subjects and Confederates. We investigated a total of 32 healthy adult volunteers, 16 women and 16 men. One female had to be excluded from brain data analyses because of strong movement artifacts. As detailed by Supplementary Table S1, there were no significant differences in age or education between female and male participants. All subjects gave informed consent and the study was approved by the Joint Ethics Committee of the National Hospital for Neurology and Neurosurgery, London, and the Institute of Neurology, University College London. Each subject was scanned with two confederates sitting immediately beside the MRI scanner. We employed professional actors (four in total, two female and two male) to play the part of the two confederates in each experiment. This was done since the paradigm required that the same persons returned to the imaging laboratory on many occasions (to allow for full control of identity), yet on each occasion to act as if they were naïve, ordinary subjects. Moreover, they had to act in congruence to their ascribed roles. At the completion of the experiment, all subjects were contacted and fully debriefed about the aims and methods of the experiment.

Experimental Procedure. The study was divided into two main experiments and a post-scan behavioural assessment phase (for an overview, see supplementary Figure 2). The subjects were told that the two experiments were independent, one study about ‘social exchange’ and the other about ‘processing pain’. In fact, the aim of the ‘social exchange experiment’ was to induce liking and disliking of fair and unfair confederates, which subsequently served to assess modulation of empathic responses to the pain of likeable fair and dislikeable unfair players. The experiment started with one session of game play outside the scanner, followed by two sessions of the 'empathy for pain paradigm' in the scanner which was interspersed by a second session of game play.

Induction of Liking and Disliking. The game was a sequential iterated Prisoners Dilemma game, in which a first player can trust a second player by sending his/her 10 starting points (transferred to money at the end of the game) to the other player knowing that each point sent will be tripled[1]. The second player (confederate) then reciprocates by sending an amount between 0 and 10 points back, which is also tripled. Fair players reciprocate large amounts, whereas unfair players reciprocate small amounts. Two blocks were performed in each session, in which each subject played 8 and 12 individual games in the first and second session, respectively, with the two other subjects (the confederate actors) sitting in different rooms. Specifically, the response sequence of returned points for the fair player in the first run was [10 9 10 10 8 10 8 10] and in the second run [10 10 9 10 10 8 10 7 10 9 8 10] resulting in a total of 558 points. The unfair player, in contrast, reciprocated a sequence of [8 6 7 3 4 3 0 4] points in the first run and [8 6 4 0 3 5 3 6 4 0 0 6] in the second run resulting in only 240 total points returned.

Empathy for Pain Paradigm. As described previously in detail[2], we used an electrical pain stimulus applied to the dorsum of the right hand. Prior to the experiment, we determined individual current amplitudes for the high and low intensity stimulation, for the subject and both confederates, based on individual pain thresholds. Notably, there were no significant differences in stimulation intensity between male and female participants (see Supplementary Table S1).

After determining individual pain thresholds, the two confederates sat either side of the subject (positioned in the scanner) on comfortable chairs positioned at the bore of the magnet. The right hands of the subject and of the two confederates were placed on a tilted board. A mirror system allowed the subject in the scanner to see all three hands. Each trial consisted of the presentation of an anticipatory visual cue which was followed after 6 seconds by a small circle (for 2 seconds) of the same color indicating the beginning of the electrical stimulation (Figure 1c). These cues were visible to the subject and confederates by means of a large screen placed behind the subject's and confederates' hands. The cues indicated whether she/he (self), the fair player (fair), or the unfair player (unfair) would get low (no pain) or high stimulation (pain). The intensity of stimulation was indicated by the colour intensity of the arrow: light colours indicated low stimulation, and dark colours painful stimulation. All visual stimuli were of equal size and centered on a black background. A white fixation cross replaced the circle for another four seconds before the next trial started (average total duration of one trial was 12 seconds).

While the two confederates were always of the same sex, all four possible gender combinations between sex of the volunteer and sex of the confederates were used equally often throughout the study. The position (left or right) of the actor (fair or unfair) was counter-balanced across subjects.

The scanning phase consisted of two sessions of 16 minutes, interrupted by a second session of game play, for which the confederates left the scanning room to return to their game-playing rooms. This lasted approximately 12 minutes. Each empathy for pain session consisted of 10 trials of each of the six condition (pain and no-pain in the context of self, fair and unfair) and 20 null events.

Post-Scan Questionnaires. After scanning, subjects completed two questionnaires: an empathy scale[3] and a general questionnaire. The latter comprised subjective intensity ratings (on a scale ranging from 0 to 10) for low and high stimulation in first and second sessions. In addition it contained scales (a) to check whether our induction of liking and disliking was successful (see Figure 1B and Supplementary Table 1) and (b) for the assessment of the subjective desire for revenge (range from -2 to + 2).

The following three questions were designed to assess the subjective desire for revenge: (a) ‘When playing, how angry were you towards the person on your left?’; (b) ‘When the person on the left/right received a shock, how sorry did you feel?’; (c) ‘How much did the person on your left/right deserve to get a shock?’ For average gender differences and intercorrelation of these scales see Supplementary Figure 1.

Finally, we assessed individual differences in empathy using the Empathic Concern Scale of the Interpersonal Reactivity Index (IRI)[3]. Although on average men showed lower scores on the empathic concern scale (15.44 +- 0.91; mean +- SEM) than women (17.56 +- 0.89), this difference did not reach significance (two-sample two-tailed t-test: $t(30) = -1.66$, $P = 0.10$, 2-tailed; $P < 0.05$, one-tailed).

Image Acquisition and Analysis. A 1.5 Tesla Siemens Sonata MRI scanner was used to acquire gradient-echo, echoplanar T2*-weighted echo-planar images (EPI) with blood oxygenation level dependent (BOLD) contrast. We used an optimized EPI sequence in which the influence of in-plane susceptibility gradients is reduced by tilting the imaging slice by 30 deg from axial to coronal orientation. Through-plane susceptibility gradients were compensated by means of a moderate preparation gradient pulse similar to z-shimming. This method has been described in detail elsewhere[4.]

Each volume comprised 35 axial slices of 2 mm thickness with 1 mm inter-slice gap and 3 x 3 mm in-plane resolution. Volumes were acquired continuously every 3.15 s. Each session began with 6 saturation volumes (discarded from the analyses). At the end of each scanning session a T1-weighted structural image was acquired for each subject.

The images were analysed using SPM2 (Wellcome Department of Imaging Neuroscience, London, UK) using an event-related model[5]. To correct for head movements all functional volumes were realigned to the first volume[6] spatially normalized to a standard template with a resampled voxel size of 3 x 3 x 3mm, and smoothed using a Gaussian kernel with a full width at half maximum (FWHM) of 10 mm. To remove low-frequency drifts from the data, high pass temporal filtering with a cut-off of 128s was applied. After pre-processing, statistical analysis was carried out using the general linear model[7]. To create regressors of interest, each condition was modeled by convolving a delta function at each trial onset (presentation of the anticipatory cue) and at each pain onset (presentation of the circle) with a canonical hemodynamic response function (HRF). Residual effects of head motion were corrected for by including the six estimated motion parameters for each subject as regressors of no interest in the design matrix. Contrast images were then calculated by applying appropriate linear contrasts to the parameter estimates for the regressor of each event.

These contrast images were then entered into one-sample *t*-tests across the 15 females and the 16 males separately to instantiate random effects group analyses[8].

The experiment was based on a 2 x 3 x 2 factorial design with the first factor representing 'intensity of stimulation' (pain vs. no pain), the second factor being 'addressee' (self, fair and unfair) and the third factor being 'gender' (male, female subjects). Statistical parametric maps were estimated for the simple effects of pain - no pain for self, for the fair and for the unfair player, separately for women and men. To assess shared networks of pain-related activation in self, fair and unfair, we carried out a conjunction analysis, and an additional (more conservative) inclusive masking procedure[9] in which we masked the contrast pain – no pain in fair or unfair condition with the contrast pain – no pain in self. To assess specific effects due to modulation of empathic responses to pain when unfair compared to fair players were observed to receive painful stimulation, we computed simple contrasts between painful trials in fair and unfair conditions, respectively, as well as the pain x fairness interaction. Finally, we used a regression analysis to explore which brain regions showed (a) a correlation between individual differences in empathy-related activity (pain – no pain in fair) and individual empathic character traits as assessed by post-scan empathy questionnaires and (b) a correlation between pain-related activity in the unfair compared to the fair condition (i.e. pain in unfair – pain in fair) and individual tendencies to seek revenge as assessed by subjective revenge scales. To implement these analyses, the appropriate first-level contrast images were entered into a regression analysis at the second level with two regressors: one regressor modeling the mean of brain activation and the other the mean-corrected (a) empathy scores or (b) revenge scores.

In our analysis, we focused on a priori regions of interest (ROIs) that were shown by previous studies[10] to be critical for processing of noxious stimuli (SI, SII, ACC, insula cortex, thalamus, brainstem; we do not report cerebellar activations because the chosen field of view did not cover the cerebellum in all subjects). Based on previous findings[11-13] we expected two regions to be of particular importance for pain-related empathy: fronto-insular cortex (FI) and ACC. The term FI was originally created by von Economo to delineate the region at the transition between orbito-frontal and insular cortex (see also the recent work by John Allman and co-workers [14]). Here we use the term FI in a topographical sense, referring to anterior insular cortex extending into orbital and frontal gyrus.

In the analysis focusing on activity specific to perceiving pain in unfair as compared to fair players we extended our regions of interest to include areas known to be involved in reward processing including ventral striatum (nucleus accumbens) and the orbito-frontal cortex. We report results at $P < 0.005$ uncorrected for multiple comparisons in the a-priori regions of interest. In the tables, we indexed results significant at $P < 0.05$ corrected for multiple comparisons across the whole brain (false discovery rate (FDR)[15] corrected) by an asterisk (*). For the conjunction analysis we adopted a significance level of $P < 0.001$ for regions of interest, and also used an asterisk (*) in the tables when significant at $P < 0.05$, whole brain FDR-corrected. Correlations between questionnaire measures and brain activity in regions of interest are reported at $P < 0.05$.

References

1. Berg, J., Dickhaut, J. & McCabe, K. Trust, reciprocity and social history. *Games Econ. Behav.* **10**, 122-142 (1995).

2. Singer, T. *et al.* Empathy for pain involves the affective but not sensory components of pain. *Science* **303**, 1157-1162 (2004).

3. Davis, M. A. A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selectived Documents in Psychology* **10**, 85 (1980).

4. Deichmann, R., Gottfried, J. A., Hutton, C. & Turner, R. Optimized EPI for fMRI studies of the orbitofrontal cortex. *NeuroImage* **19**, 430-441 (2003)

5. Josephs, O., Turner, R. & Friston, K. Event-related fMRI. *Hum. Brain Mapp.* **5**, 243-248 (1997).

6. Friston, K. J. *et al.* Spatial Registration and Normalization of Images. *Human Brain Mapping* **2**, 165-189 (1995).

7. Friston, K. J. *et al.* Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping* **2**, 189-210 (1995).

8. Penny, W., & Holmes, A.P. Random-Effects Analysis. In: Human Brain Function., 843-850. Elsevier, San Diego (2004).

9. Nichols T, Brett M, Andersson J, Wager T, Poline JB. Valid conjunction inference with the minimum statistic. *Neuroimage* **25**, 653-60 (2005).

10. Peyron, R., Laurent, B. & Garcia-Larrea, L. Functional imaging of brain responses to pain. A review and meta-analysis. *Neurophysiol. Clin.* **30**, 263-288 (2000).

11. Singer, T. *et al.* Empathy for pain involves the affective but not sensory components of pain. *Science* **303**, 1157-1162 (2004).

12. Jackson, P. L., Meltzoff, A. N. & Decety, J. How do we perceive the pain of others: A window into the neural processes involved in empathy. *NeuroImage* **3**, 771-779 (2005).

13. Morrison, I., Lloyd, D., di Pellegrino, G. & Roberts, N. Vicarious responses to pain in anterior cingulate cortex: is empathy a multisensory issue? *Cogn Affect. Behav. Neurosci.* **4**, 270-278 (2004).

14. Allman, J. M., Watson, K. K., Tetreault, N. A. & Hakeem, A. Y. Intuition and Autism: Role of Von Economo Neurons. *Trends Cogn Sci.* (2005).

15. Genovese CR, Lazar NA, & Nichols T. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* **15**, 870-878 (2002).