



Responsibility modulates pain-matrix activation elicited by the expressions of others in pain



Fang Cui^{a,b,c}, Abdel-Rahman Abdelgabar^a, Christian Keysers^{a,b,d}, Valeria Gazzola^{a,b,d,*}

^a Social Brain Lab, Netherlands Institute for Neuroscience, Royal Netherlands Academy for Arts and Sciences, 1105BA Amsterdam, The Netherlands

^b Department of Neuroscience, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

^c Institute of Affective and Social Neuroscience, Shenzhen University, Shenzhen, China

^d Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

ARTICLE INFO

Article history:

Received 9 December 2014

Accepted 13 March 2015

Available online 20 March 2015

Keywords:

Empathy

fMRI

Pain

Responsibility

ABSTRACT

Here we examine whether brain responses to dynamic facial expressions of pain are influenced by our responsibility for the observed pain. Participants played a flanker task with a confederate. Whenever either erred, the confederate was seen to receive a noxious shock. Using functional magnetic resonance imaging, we found that regions of the functionally localized pain-matrix of the participants (the anterior insula in particular) were activated most strongly when seeing the confederate receive a noxious shock when only the participant had erred (and hence had full responsibility). When both or only the confederate had erred (i.e. participant's shared or no responsibility), significantly weaker vicarious pain-matrix activations were measured.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Perceiving the facial expressions of pain in others has important social functions. In particular, perceiving the pain of others motivates and regulates helping behavior (Craig et al., 2001; Williams, 2002). Over the past decade, our understanding of the neural basis of this perception has been refined by a number of experiments that have exposed participants to the facial expressions of pain of others. After reviewing this evidence, we will show that a common feature of these experiments has been to show expressions of pain that were not caused by the participant him or herself. Accordingly, we will argue that an important aspect of the neural basis of pain perception has been left unexplored: how this neural activation is modulated by the degree to which the observer caused the witnessed pain.

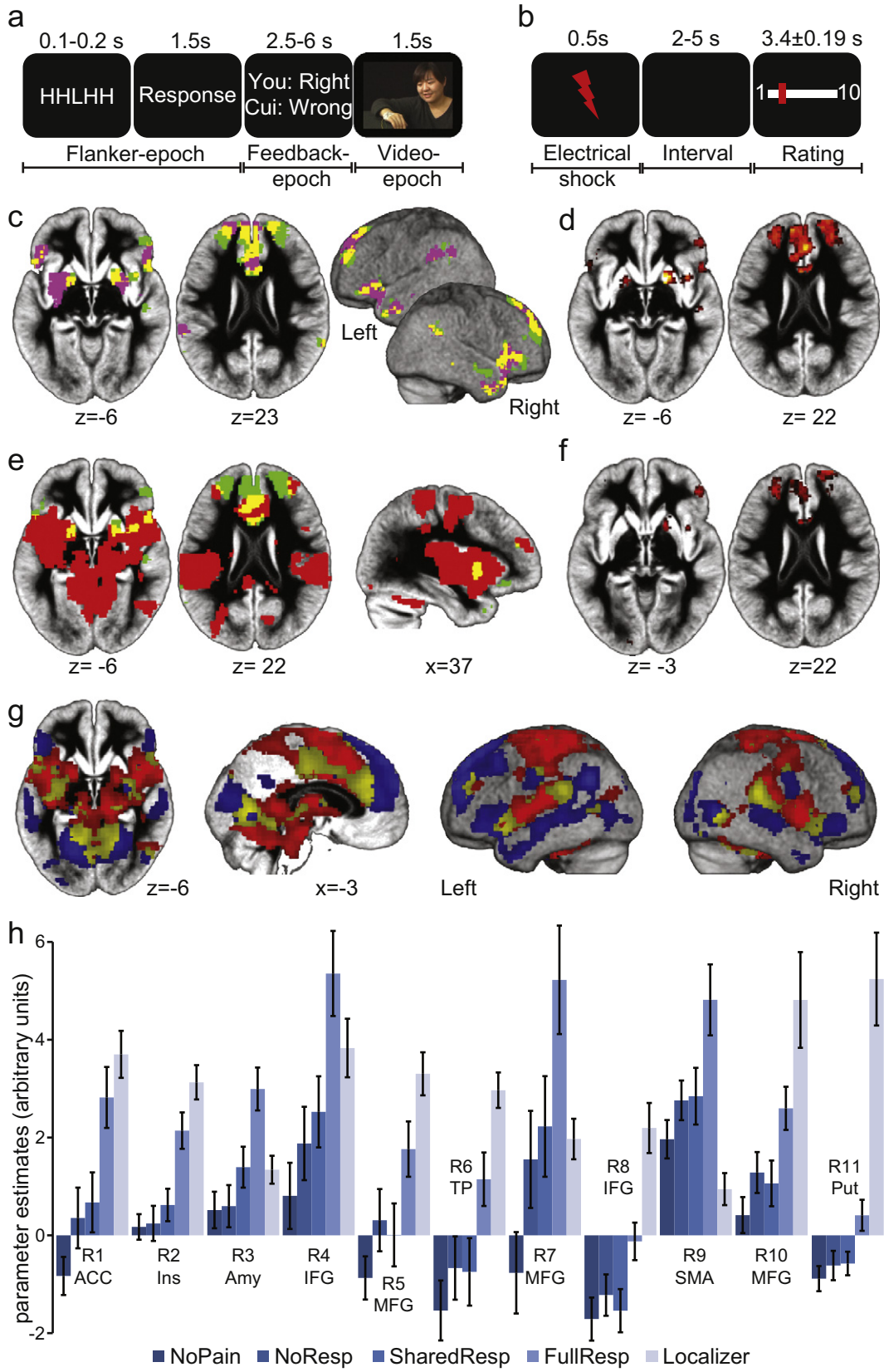
There is a long tradition of studying the neural basis of the visual processing of facial expressions in general. The observation of facial expression triggers activity in early visual cortex, in the human occipital face area (OFA) and in the middle temporal gyrus along the superior temporal sulcus (STS) (see Said et al., 2011 for a review). Additionally, facial expressions activate the frontal operculum, supplementary motor area (SMA) and somatosensory cortices that are also activated when participant produce facial expressions. The vicarious activation to viewing others' facial expressions of these sensorimotor brain regions has thus been interpreted as representing an internal simulation of the sensorimotor neural activity associated with producing the observed

facial expressions (Bastiaansen et al., 2009; Said et al., 2011; van der Gaag et al., 2007). When witnessing facial expressions of pain, participants have been shown to additionally activate regions of the anterior insula (AI), anterior cingulate cortex (ACC) and the amygdala (Botvinick et al., 2005; Saarela et al., 2007; Simon et al., 2006). Because the AI, ACC and amygdala are part of the pain-matrix – the set of brain regions that are activated when the participants themselves are exposed to noxious stimuli on their body (Garcia-Larrea and Peyron, 2013; Melzack and Wall, 1965; Mouraux et al., 2011) – and because their level of activation during the experience of pain correlates with the unpleasantness of experienced pain (Rainville, 2002), many interpret their vicarious activation while witnessing the pain of others as the neural correlate of empathy – feeling vicariously what we see someone else to experience (Corradi-Dell'Acqua et al., 2011; Jackson et al., 2006; Koban et al., 2013; Lamm et al., 2011; Singer et al., 2004). That AI, ACC and amygdala are also vicariously activated when pain is perceived through non-facial cues (Corradi-Dell'Acqua et al., 2011; Jackson et al., 2006; Koban et al., 2013; Lamm et al., 2011; Meffert et al., 2013; Singer et al., 2004) supports the notion that these activations have less to do with the facial expressions as a motor act and more with pain as a perceived emotion. That their activation is stronger in more empathic individuals (Singer et al., 2004) and weaker in psychopaths (Meffert et al., 2013) further supports their role in empathy. Interestingly, the magnitude of vicarious activations in the AI also predicts helping behavior (Hein et al., 2010) providing evidence that vicarious activations could have behavioral significance by motivating the witness to help another. It should be noted, however, that experiencing negative emotions other than pain also activates regions such as the

* Corresponding author at: Meibergdreef 47, 1105 BA, Amsterdam, The Netherlands.
E-mail address: v.gazzola@nin.knaw.nl (V. Gazzola).

amygdala, AI and ACC. For instance the experience of disgust recruits all of these regions (Wicker et al., 2003), and so does the emotion of guilt (Jankowski and Takahashi, 2014). Accordingly, vicarious activations in

these regions cannot unambiguously be interpreted as representing vicarious pain, but could involve a mixture of emotions, such as concern or distress. Need for caution in interpreting activity in these regions as



evidence for the witness experiencing pain in the strict physical sense while witnessing the pain of others is further commanded by the fact that brain regions that track physical pain intensity most accurately during first hand pain experience (Wager et al., 2013) — although including some of the regions involved during vicarious pain (e.g. ACC) — do include numerous brain regions not typically involved during vicarious pain (e.g. mid-insula, SII, thalamus and cerebellum). Indeed, even during the experience of physical pain, not all activity in the so-called pain-matrix may be linked to a specific feeling of pain, but may rather represent a mix of processes related to salient negative events (Legrain et al., 2011).

Some factors influencing the intensity of vicarious activations have received much interest (de Vignemont and Singer, 2006): vicarious activations are stronger when attention is directed to the pain (Gu and Han, 2007), when stimuli are more realistic (Gu and Han, 2007), when the observer is socially closer to the pain-taker (Cheng et al., 2010), belongs to the same group or race (Avenanti et al., 2010; Azevedo et al., 2013; Hein et al., 2010; Xu et al., 2009) or considers the pain-taker fair (Singer et al., 2006).

Except for the work of Koban and colleagues (Koban et al., 2013) showing that the AI and dorsolateral prefrontal cortex differentiate between noxious and innocuous stimulation to another individual when the participant caused the pain by erring, but not when the other individual had caused his own pain, so far, brain activity following the perception of other people's pain has however been mostly studied in situations in which the participant witnesses pain he did not cause. Whether responsibility for the observed pain would boost the way the brain reacts to facial expressions of pain thus remains largely unknown. Based on personal experience that witnessing pain we caused is more distressing than witnessing pain we have not caused, we hypothesized that increasing levels of responsibility for observed pain should boost brain activity, particularly in regions also involved in the first-hand experience of pain (AI, ACC and amygdala). In addition, based on the diffusion of responsibility literature (Darley and Latane, 1968), we hypothesize that if the cause of the pain is shared amongst agents, vicarious activations should be reduced compared to cases in which the witness was the sole cause of the pain.

To test these hypotheses, a participant and author FC performed a difficult flanker task simultaneously, and if either or both made a mistake, FC was administered a noxious shock. In some trials the participant was thus fully responsible for causing FC's pain, in some the participant and FC shared responsibility, and in others, only FC was responsible for her own pain. We then measured, using fMRI, how brain activity in the participant's pain-matrix varied while witnessing the pain of FC, via a (supposedly live) video-feed of her facial expressions, as a function of responsibility.

Materials and methods

Participants

Thirty-three volunteers participated in this study, but three were excluded from the analyses: one because s/he felt claustrophobic and the other two because they said they didn't believe that the person in the movie received electroshocks in real time. It is known, that responses

to other people's pain can change as a function of gender (Singer et al., 2006) and ethnicity (Azevedo et al., 2013; Xu et al., 2009). To ensure that our findings are not limited to a specific gender or ethnicity, we therefore recruited our 30 final participants to be composed of a balanced number of male and female and Chinese and Caucasian (i.e. 15 Chinese, of which 7 males, and 15 Caucasians, of which 7 males). The age of our participants was $24.8 \text{ y} \pm 4.37$ (mean \pm s.d.). All participants were healthy, right-handed, had no history of neurological or psychiatric disorders, and provided a written informed consent. This study was approved by the Ethics Committee of the University of Amsterdam, the Netherlands.

General experimental setup

A confederate design was used in this study. At the beginning of the experiment, experimenter AR introduced the participant to author FC (Chinese female), who was described as another participant. The participant and FC then drew lots that were manipulated so that the participant was always assigned to the fMRI scanning. In the scanner, each trial (see Fig. 1a for a graphical illustration of each trial) started by displaying a central target letter flanked by distractors and the participant had to press the button corresponding to the central letter. This first epoch is called the Flanker-epoch. The participant was led to believe that FC would have seen, simultaneously but in another room, the same display and performed the same task. Directly after the Flanker-epoch, the participant and FC were informed about the performance of both players (Feedback-epoch). If both performed the task correctly, the participant believed FC would receive a weak, innocuous electroshock on her right hand (NoPain condition). If any erred (i.e. only the participant, both, or only FC) a stronger, noxious shock would be delivered to FC (FullResp, SharedResp or NoResp conditions, respectively, see Table 1). The participant was further led to believe that he/she would then, after a random blank interval, see in real time (through a CCTV) FC receive the electroshock, be it noxious or innocuous, depending on their joint performance (Video-epoch). A 5 to 8 s blank screen separated consecutive trials (Fig. 1a). In reality, FC was not performing the task or receiving shocks during the experiment. Instead, a computer adjusted the presentation time of the flanker task, and hence its difficulty, and simulated correct or incorrect performance of FC to ensure a minimum number of trials for each condition (Table 1, last column). During the Video-epoch pre-recorded videos of FC receiving electroshocks were shown to the participant to ensure that all participants viewed the same movies. FC only received shocks during movie recording. A total of 140 trials were presented, split in 4 runs of 35 trials each. As mentioned above, because the amount of trials for each condition depended on subject's performance in the flanker task, the number of trials differed slightly across conditions and participants and the percentage of trial presented on average for each condition is indicated in Table 1. It is important to note that a GLM comparing two conditions is valid even if the two conditions have different numbers of trials, and our main contrast (FullResp–SharedResp) includes conditions with similar numbers of trials.

To localize regions involved in the painfulness of the participant's own nociceptive experiences, after the main experiment, the participant went through another fMRI scanning session (Pain-localizer) in

Fig. 1. (a–b) Experimental task. Trial structure for (a) the responsibility task, with a screenshot taken from one of the painful videos (see also Movies 1 and 2); and (b) the Pain-localizer session. (c–g) Whole brain results. (c) Whole brain effects of responsibility on the processing of the pain of others. Purple: VideoFullResp–VideoNoResp. Green: VideoFullResp–VideoSharedResp. Yellow: overlap between the other two colors shown in the same render. (d) Axial cuts at the indicated z coordinates for the VideoFullResp–VideoSharedResp contrasts. Colors go from dark red for $t = 3.13$ to white for $t > 5$. (e) Overlap between the effects of responsibility and the pain localizer. Red(+yellow): Pain-localizer. Green(+yellow): VideoFullResp–VideoSharedResp. Yellow: overlap. (f) Interaction between responsibility and the Feedback- and Video-epochs (VideoFullResp–VideoSharedResp)–(FeedbackFullResp–FeedbackSharedResp). Thresholds and color code as for (d). (g) Effect of viewing painful facial expressions independently of responsibility and its overlap with the pain localizer. Blue(+yellow): (VideoFullResp + VideoSharedResp + VideoNoResp)–3 × VideoNoPain. Red(+yellow): Pain-localizer. Yellow: overlap. Colors go from $t = 3.13$ to $t = 8$. All the images shown from c to g were thresholded at $p_{unc} < 0.001$, $k > 10$ and survived $q_{FDR} < 0.05$. (h) ROI results. Signal extracted, for the indicated conditions against baseline, from the 11 clusters (shown in 1e and listed in Table 1) resulting from the contrast VideoFullResp–VideoSharedResp masked with the Pain-localizer.

Table 1

Experimental design and conditions. From left to right: participant's and author FC's flanker task performance; condition name from participant's responsibility point of view; type of electrical stimulation given to FC during video-recording for each condition; average number of trials for each condition included in the analysis expressed in % and in absolute number. Note that we had at least 17 repetitions of each of the four conditions in all participants.

Participant's performance	FC's performance	Condition name	Electrical stimulation	% of trials (average \pm s.e.m.)	Average amount of trials
Correct	Correct	NoPain	Innoxious	32.4 \pm 0.82	45
Incorrect	Correct	FullResp	Noxious	18.6 \pm 0.71	26
Incorrect	Incorrect	SharedResp	Noxious	17.8 \pm 0.69	25
Correct	Incorrect	NoResp	Noxious	31.2 \pm 0.71	44

which noxious and innocuous electroshocks were delivered to the participant's right hand in the scanner (see Fig. 1b and *Pain-localizer procedure* paragraph). An anatomical scan was finally acquired. After all the scanning, a debriefing was given and the participant was asked to indicate how much he/she believed that FC was actually being shocked based on their joint performance during the experiment.

Flanker-task in details

During the Flanker-epoch, one of the following five-letter-strings appeared on the screen randomly: “HHHHH”, “LLLLL”, “HHLHH” and “LLHLL”. In order to achieve a similar task difficulty across participant, the duration of each string was initially set at 150 ms, and was changed to reach a minimum of 100 ms and a maximum of 200 ms based on the participant's previous performance: if the participant gave two consecutive correct responses, the time of the next string was shortened by 10 ms; if the participant gave two consecutive incorrect responses the time was prolonged by 10 ms. The participant and author FC were instructed to simultaneously respond to the letter in the middle (H or L). To give the response the participant had to press one of two pre-assigned buttons on an MRI-compatible button-box. The fake setup for FC required her to press the “H” and “L” buttons on a keyboard. As mentioned above, in reality only the participant was running the task and he/she had 1.5 s to give a response. If no button press was recorded within this duration, the participant's performance in this trial would be considered incorrect, and indicated as such during Feedback-epoch.

Videos preparation

The videos used in this study showed author FC seated at a table. Her face and upper body were clearly visible, and two electrodes, used to deliver the electroshock, were attached and visible on the back of her right hand, which was resting in front of her on the table (see Fig. 1a for an example of video screenshot). During the video recording, we first tested FC's pain threshold (see section “Stimulation and Pain Threshold” for details). Afterwards 70 unique video-clips were recorded while FC received the noxious or innocuous electroshocks (35 videos each). Each video was cut to last 1.5 s and started with 0.3 s in which the experimenter was sitting still with a neutral face, followed by 0.5 s of electroshock to trigger FC's natural facial expression. All the settings, including the background and FC's look, were kept unchanged relative to the recording day during all the experimental days. For each of the 4 runs, at the end of each trial a movie was randomly picked (without replacement) from the 35 movies of the appropriate category (painful/painless), so that no movie would be seen twice per run.

Pain-localizer procedure

Sixteen noxious and sixteen innocuous 0.5 s electroshocks were applied, in a pseudo-randomized order (i.e. no more than two shocks of

the same intensity were delivered consecutively), on the participant's right hand using a MRI-compatible electrical stimulation system. After a random interval ranging from 2 to 5 s, the participant was asked to evaluate how painful the received electroshock was by button press. The participant was instructed to use three buttons of a MRI compatible button-box placed next to their left hand. Two buttons were used to move the slider left and right on the visual scale on the screen and the third button was for confirmation. The pain intensity scale was a 10 point scale (1: not painful at all; 10: most intense imaginable pain), with the starting point set randomly for each trial to disentangle the number of button presses from the rating (Fig. 1b). A random interval ranging from 8 to 12 s separated the response from the next stimulation.

Electrical stimulation and pain threshold

A 100 Hz train of electrical pulses (2 ms each) was applied for 0.5 s using an MRI-compatible electrical stimulator attached on the back of the right hand on the 4th musculus interossei (stimulation area: 16 mm²) through two bipolar surface electrodes. Before the scanning we measured the pain threshold from both FC and the participant. We started from a 0.2 mA current that was then increased until maximally 6.0 mA in 0.1 mA steps (Singer et al., 2004). Participants were instructed to evaluate how painful the stimulation was on a 10-point scale (same as in Pain-localizer). We then chose the current corresponding to a rating of 7 for the painful condition and of 2 for the painless condition (Singer et al., 2004). The current selected was 0.75 \pm 0.14 mA (mean \pm s.e.m) for the painless and 2.12 \pm 0.77 mA for the painful condition. The same procedure was used only on FC during the video recording session.

Data acquisition

A Phillips Achieva 3.0 T MRI scanner was used for image acquisition. We used a T2*-weighted echo-planar sequence with 32 interleaved 3.5 mm thick axial slices and a 0.35 mm gap for functional imaging (TR = 1700 ms, TE = 27.6 ms, flip angle = 73°, FOV = 240 mm \times 240 mm, 80 \times 80 matrix of 3.5 mm isotropic voxels). At the end of the functional scanning, a T1-weighted anatomical image (1 \times 1 \times 1 mm) covering the whole brain, was acquired.

Image pre-processing

fMRI data was pre-processed using SPM8 (www.fil.ion.ucl.ac.uk). All echo planar images (EPIs) were slice-time corrected and realigned to the participant's mean EPI. T1 images were then co-registered to the mean EPI, segmented, and the gray matter was used to estimate the normalization parameters which were then applied to all EPIs. Normalized (2 \times 2 \times 2) EPIs were smoothed with an 8 mm isotropic FWHM Gaussian kernel.

General linear models

Two separate general linear models (GLM) were applied at the single subject level, one for the four runs of the responsibility task and one for the Pain-localizer. Predictors were modeled using a standard boxcar function convolved with the hemodynamic response function (HRF). For each of the four runs of the responsibility task, we included the following predictors. First, because each run started with the participant indicating their readiness with a button press, one predictor collected this initial press. The predictor was aligned with the presentation of the initial screen and lasted until the button press. Another predictor contained all the Flanker-epochs from the appearance of the string until the end of the participants' button press, independently of performance. Four separate predictors, one for each experimental condition (NoPain, FullResp, SharedResp, NoResp) were then used for the Feedback-epoch. These Feedback-epoch predictors were aligned to the

presentation of the feedback screen and lasted for a random duration (chosen by the presentation program) between 2.5 and 6 s. Four predictors finally captured the Video-epoch separately for each experimental condition. Each video-predictor lasted for 1.5 s, corresponding to the videos' actual duration. For the Pain-localizer session, we modeled one predictor, lasting 0.5 s, for all 32 electrical stimulations with a parametric modulator for the subjective rating of pain intensity. A second predictor contained the rating period, from onset of the rating screen until the end of the button presses. Six additional predictors of no interest, resulting from the realignment procedure, were entered for each of the five runs to account for translations and rotations of the head (none of the included participants had head motions parameters exceeding the acquired voxel-size). Data was then analyzed at the second level using a within-subject repeated measurement ANOVA with 8 variables, and computing pairwise comparisons between the conditions of interests using directed planned comparisons (so called *t* contrasts). The ANOVA included the parameter estimates of the four conditions during the Video-epoch (VideoNoPain, VideoFullResp, VideoSharedResp, VideoNoResp), and the four conditions during the Feedback-epoch (FeedbackNoPain, FeedbackFullResp, FeedbackSharedResp, FeedbackNoResp). Results were thresholded at $p_{unc} < 0.001$ (uncorrected) with a minimum cluster size of 10. All results presented also survived $q_{FDR} < 0.05$ (false discovery rate). We decided to control the false discovery rate at the voxel level rather than the family-wise error rate, because *fd*r is (a) a valid form of controlling the multiple comparison problem (Benjamini, 2010; Genovese et al., 2002), (b) leads to thresholds that are close to what has been found to provide optimal reproducibility (Thirion et al., 2007), and provides a better compromise between Type I and Type II error than the much more conservative family-wise error correction (Lieberman and Cunningham, 2009). In addition, we used voxel-wise *fd*r rather than topographical *fd*r (Chumbley and Friston, 2009), because we look for overlap at the voxel-level between activations during the responsibility task and a pain localizer. Although topographical *fd*r has some advantages (Chumbley and Friston, 2009), voxel-wise interpretations are problematic in the topographical approach, as the significance of a cluster does not automatically imply the significance of individual voxels within that cluster.

Results

Behavioral results

Average performance at the Flanker-task was 62.8% correct. We had at least a total of 17 repetitions of each of the four conditions in all participants.

After telling the participants the truth about the experimental design, they were asked: "Do you think the experimental setup was realistic enough to believe it (1 = strongly disagree 7 = strongly agree)?" in a feedback questionnaire. The average rating was 6.2 ± 0.7 (s.d.) and none of the 30 included participants even somewhat disagreed with the statement, demonstrating the credibility of our design. Two of the initial 33 participants had voiced doubts about the experiments before debriefing, and were excluded from the analysis. They were the only participants that selected "somewhat disagree".

Our experimental design was motivated by the assumption that participants would perceive varying degrees of responsibility based on who erred, with full-responsibility > shared > no-responsibility. However, we had not directly asked the participants if that were true. Accordingly, we tried to contact all participants again to ask them "Please rate how responsible you felt for the pain of the other in each condition, on a scale from 1 = not responsible at all to 9 = extremely responsible". Only eighteen of the participants could be contacted (the other students had changed their email address and phone number since). These 18 reported having perceived more responsibility for the pain in the FullResp (mean \pm s.e., 7.5 ± 0.06) condition, less in the SharedResp condition (3.5 ± 0.06) and close to none in the NoResp condition (1.3 ± 0.02),

with all pair-wise differences significant (*t*-test, $p < 0.001$). Given that 8 months lapsed on average between the experiment and these reports, these numbers should be interpreted with care, but they indicates that the participants' recollection of the experiment is in line with our aim to manipulate perceived responsibility and are compatible with the names given to the different conditions.

The effect of responsibility on brain activation

To identify the areas that respond to the display of FC's pain more in the condition in which the participant is the main responsible for FC's getting the electroshock compared to the case in which FC is causing the electroshock herself, we used the contrast VideoFullResp–VideoNoResp. This revealed stronger activations in conditions in which the witnessed pain was entirely due to the participant's mistake than those that were entirely due to author FC's mistake. Regions showing this difference ($p_{unc} < 0.001$, $k > 10$, $T > 3.13$; also $q_{FDR} < 0.05$) included the middle temporal gyrus around the superior temporal sulcus, the inferior frontal gyrus, ACC, AI, amygdala, striatum and right superior frontal gyrus (see Inline Supplementary Table S1; and Fig. 1c, purple).

Inline Supplementary Table S1 can be found online at <http://dx.doi.org/10.1016/j.neuroimage.2015.03.034>.

The reverse contrast revealed no activations (at $q_{FDR} < 0.05$). Note that the contrasts (VideoFullResp–VideoNoPain)–(VideoNoResp–VideoNoPain) that would isolate the part of empathy triggered by the sight of pain are mathematically identical to the contrast we report above (VideoFullResp–VideoNoResp) because VideoNoPain is canceled out. A similar logic applies to the following contrasts. To explore whether sharing the responsibility for FC getting the electroshock would suffice to reduce the response of the participant while witnessing FC's display of pain, we contrasted VideoFullResp–VideoSharedResp. This revealed a similar circuit (See Inline Supplementary Table S2; Fig. 1d, and green in Fig. 1c and e) that overlapped (yellow in Fig. 1c) with VideoFullResp–VideoNoResp, and included again the ACC, AI, amygdala, striatum, higher level visual areas of the temporal lobe and more cognitive regions including temporal pole and the superior frontal gyrus ($p_{unc} < 0.001$, $k > 10$, $T > 3.42$; also $q_{FDR} < 0.05$). Again, the reverse contrast revealed no activations (at $q_{FDR} < 0.05$). Because the contrast VideoFullResp–VideoSharedResp only includes trials in which the participant received the same negative feedback about his/her performance differences in activation cannot be related to the participant's self-error-monitoring, which is known to activate regions similar to those of pain experience (Carter et al., 1998; Taylor et al., 2007; Ullsperger and von Cramon, 2003). Finally, to explore if there was a further decrease in the response to seeing the painful videos if the participant had no rather than shared responsibility, we computed the VideoSharedResp–VideoNoResp contrast, but this revealed no significant differences at $q_{FDR} < 0.05$, nor did the reverse contrast, in line with the similarity between the contrasts of these respective conditions and VideoFullResp. Reducing the threshold to $p_{unc} < 0.005$ ($k > 10$) revealed bilateral amygdala and hippocampus, left prefrontal, right inferior temporal gyrus and left middle temporal gyrus.

Inline Supplementary Table S2 can be found online at <http://dx.doi.org/10.1016/j.neuroimage.2015.03.034>.

Consistent with the literature (Garcia-Larrea and Peyron, 2013; Melzack and Wall, 1965; Mouraux et al., 2011), our Pain-localizer revealed areas associated with the "pain-matrix", including the bilateral cingulate cortex, bilateral insula, sensorimotor strip (Brodmann Area, BA, 2, 3b, 4a), striatum, premotor cortex (BA6, and inferior frontal gyrus), inferior parietal cortex, and cerebellum (all $p_{unc} < 0.001$, $k > 10$, $T > 3.42$; also $q_{FDR} < 0.05$, See Inline Supplementary Table S3; and red in Figs. 1e and g).

Inline Supplementary Table S3 can be found online at <http://dx.doi.org/10.1016/j.neuroimage.2015.03.034>.

Because an overlap between self- and other-emotions is considered a defining feature of the neural proxy of empathy in the literature

Table 2

Activation table of the overlap between VideoFullResp–VideoSharedResp and the Pain-localizer. Both responsibility effect and Pain-localizer were individually thresholded at $p_{unc} < 0.001$, $k > 10$, and survived $q_{FDR} < 0.05$. The T values shown in this table are from VideoFullResp–VideoSharedResp.

#	Cluster size	MIN coordinates (mm) X, Y, Z			T-values	Hem	Anatomical description
1	383	4	36	24	4.73	R	Anterior cingulate cortex
		−8	28	20	4.5	L	Anterior cingulate cortex
2	367	18	8	−6	5.04	R	Putamen
		38	8	−2	4.5	R	Insula lobe
		30	16	−12	3.99	R	Insula lobe
		28	10	−14	3.98	R	Olfactory cortex
		20	−4	−14	3.73	R	Hippocampus
3	142	−24	−4	−14	4.37	L	Amygdala
		−18	6	−8	3.74	L	Putamen
4	76	54	20	−4	3.89	R	Inferior frontal gyrus (p. Orbitalis)
		58	20	2	3.61	R	Inferior frontal gyrus (p. Triangularis)
5	25	−30	46	20	3.74	L	Middle frontal gyrus
6	20	−40	2	−20	3.83	L	Temporal pole
7	16	30	52	22	3.53	R	Middle frontal gyrus
8	16	−50	20	−6	3.41	L	Inferior frontal gyrus (p. Orbitalis)
9	12	14	24	58	3.46	R	SMA
10	12	32	42	22	3.47	R	Middle frontal gyrus
11	10	−24	−4	12	4.09	L	Putamen

(Gazzola and Keysers, 2009; Keysers et al., 2004; Singer et al., 2004; Wicker et al., 2003), to explore which of the regions with BOLD signals modulated by responsibility during the Video-epoch might be interpreted as a proxy of empathy, we inclusively masked the activated areas resulting from the contrast VideoFullResp–VideoSharedResp, with those from the Pain-localizer. We found a set of regions that overlapped with pain experience (yellow in Fig. 1e), and a set that did not. The former (Table 2; Fig. 1e, yellow) includes the ACC, right AI, bilateral putamen, left amygdala, right and left inferior frontal gyrus and SMA. These regions are supposed to be of particular relevance when it comes to empathy (Fan et al., 2011; Lamm et al., 2011; Singer et al., 2004). A number of clusters however clearly fell outside of the pain localizer, including high-level visual regions of the temporal lobe around the STS and the dorsolateral prefrontal cortex (See Inline Supplementary Table S4; and Fig. 1e, green).

Inline Supplementary S4 can be found online at <http://dx.doi.org/10.1016/j.neuroimage.2015.03.034>.

The FullResp and SharedResp trials not only differ in the degree of responsibility perceived by the participant, but also based on the performance of the confederate FC, which is correct in FullResp but incorrectly in SharedResp. Because it has been shown that the errors and successes of others can vicariously activate regions encoding errors and successes in the self (Heldmann et al., 2008; Mathalon et al., 2003; Monfardini et al., 2013; Shane et al., 2008), the greater activity in those areas resulting from the VideoFullResp–VideoSharedResp contrast could reflect a spill-over from the vicarious processing of the success of another triggered in the Feedback-epoch. Because in our design the Feedback-epoch informing the participants about their errors was separate in time from the Video-epoch triggering the processing of the facial expressions, we would expect the activations associated with error monitoring for the contrast FullResp–SharedResp to be greater during the Feedback-epoch than the Video-epoch. If VideoFullResp–VideoSharedResp however reflects a modulation of facial expression processing by responsibility, we would expect the difference to be larger during the Video-epoch when facial expressions are shown than during the Feedback-epoch when errors are revealed. An interaction analysis (VideoFullResp–VideoSharedResp) > (FeedbackFullResp–FeedbackSharedResp) confirms that the AI and ACC showed a larger

modulation during the Video-epoch compared to the Feedback-epoch (Fig. 1f; and Inline Supplementary Table S5), and that this overlaps with the Pain-localizer. The inverse contrasts did not show any significant activations (at $q_{FDR} < 0.05$). This result suggests a spillover from a vicarious error processing to be insufficient to explain the difference during the Video-epoch. Note that this interaction analysis will pitch brain activity time-locked to the presentation of the feedback screen against that time-locked to the presentation of the video. In theory, participants could have processes errors at other points in time as well. Such ‘free floating’ error related activity, will either go into error (and thus play against the interaction analysis) or into the response to both the feedback and video stimuli (as it is not time locked to either), and hence also not show up in the interaction (but as a main-effect). Calculating the contrast FeedbackFullResp–FeedbackSharedResp, within the whole brain or within the Pain-localizer failed to yield any significant activation (at $q_{FDR} < 0.05$), reinforcing that a spill-over from the Feedback phase is unlikely.

Inline Supplementary Table S5 can be found online at <http://dx.doi.org/10.1016/j.neuroimage.2015.03.034>.

The majority of empathy for pain studies so far contrasted stimuli depicting another person in pain against stimuli representing the same person going through a non-painful experience – independently of responsibility. To explore whether we can reproduce the existing results with our data-set, we performed a contrast between all videos illustrating pain and those illustrating no pain (i.e. VideoFullResp + VideoSharedResp + VideoNoResp–3 × VideoNoPain; blue + yellow in Fig. 1g) and overlapped (yellow and Table S6) this contrast with the Pain-localizer (red + yellow). Results evidenced a network of brain regions (yellow in Fig. 1g, and Inline Supplementary Table S6) recruited in both the observation of the displays of pain of others and the subjective experience of pain (independently of the responsibility the observer has for causing the other person’s pain), consistent with that found in the literature, including the AI, ACC and amygdala (Lamm et al., 2011). We finally extracted the mean signal time course in clusters common to VideoFullResp–VideoSharedResp and the Pain-localizer (see Table 2). To explore how the responsibility-dependent brain activation reported above develops in time, compares to baseline and to the level of activation in first person pain experience, we ran two additional analyses. For the first analysis, we extracted, for each of the 11 clusters listed in Table 2 (see also Fig. 1e), the mean signal time course from the normalized functional images of each participant. Using Marsbar (<http://marsbar.sourceforge.net/>, Brett et al., 2002), we then calculated parameter estimates for all four video conditions (NoPain, NoResp, SharedResp and FullResp) against baseline. Fig. 1h shows these parameter estimates averaged across participants and illustrates that the activation was always numerically lowest in the condition in which no pain was witnessed, and clearly above baseline for the VideoFullResp condition in most of the ROIs, except in ROI8 located in the temporal pole. We additionally calculated parameter estimates for the pain localizer, and plotted the sum of the parameter common to all shocks and the parametric modulator, to represent the activation level to the noxious shocks. This revealed that in most of the ROIs, except ROI8 and 11, the first-person experience of a noxious shock triggered activity that is roughly commensurable to witnessing another experience a shock under full responsibility (Fig. 1h). We did not perform a statistical analysis of these parameter estimates because the clusters were selected based on responses during the Video epoch and the Pain-Localizer, biasing statistical comparisons. For the second analysis, we extracted the time course (peri-stimulus time histogram, PSTH; rfxplot toolbox for SPM, <http://rfxplot.sourceforge.net/index.html>, Glascher, 2009) for each video condition and the Localizer (for high and low intensity stimuli separately), to explore whether the BOLD signal in response to painful facial expressions has a time course similar to that during the first-person experience of pain (see Inline Supplementary Fig. S1). The analysis illustrates the similarity of time courses of the video and localizer responses.

Inline Supplementary Table S6 and Fig. S1 can be found online at <http://dx.doi.org/10.1016/j.neuroimage.2015.03.034>.

Discussion

In contrast to most studies investigating neural responses to the pain of others, in which the participant witnesses a pain he did not cause (Lamm et al., 2011), we investigate how the brain reacts to facial expressions of the pain of others as a function of the participant's responsibility. We find the brain response to witnessing facial expressions of pain to be increased when the observer had full responsibility for the pain inflicted to the other. Shared or no responsibility lead to relatively reduced activations. The brain regions modulated by responsibility include the AI, ACC, the putamen and amygdala. Part of the BOLD activity of these regions additionally correlated with stimulus painfulness ratings during the participant's first-hand experience of pain (pain localizer). Our results confirm the existing literature on empathy, by showing that the AI and ACC, the two most consistent regions across studies of empathy for pain (Lamm et al., 2011) are indeed activated both while experiencing pain and witnessing the pain of others (VideoFullResp + VideoSharedResp + VideoNoResp – 3 × VideoNoPain). However, our findings extend the literature by showing how activations to the pain of others are modulated by the observer's responsibility.

That having no responsibility for the pain reduces vicarious activations dovetails with a recent study (Koban et al., 2013) showing that the AI and dorsolateral prefrontal cortex differentiate between noxious and innocuous stimulation to another individual when the participant caused the pain by erring, but not when the other individual had caused his own pain. Because in Koban et al.'s study a task was performed by the witness in the full- but not in the no-responsibility condition, it was not possible to determine an effect of sharing responsibility. In addition, the modulation ascribed to agency, could have been due to differences in attention because participants could disengage attention during the no-responsibility trials in which they were not required to perform a task. By having our participant always do the task together with the pain-taker, and including a pain localizer, we were able to extend their results and (a) demonstrate a strong effect of sharing responsibility and (b) make the attention explanation unlikely. In addition, by showing that some voxels modulated by responsibility (AI, ACC, putamen and amygdala) fall within our participant's pain localizer while others (STS/MTG and dorsolateral prefrontal cortex) do not, we help decompose the neural effects of responsibility, in some that are more likely to relate to negative affective reactions that are common to pain observation and experience (AI, ACC, putamen and amygdala), and some that are more likely to reflect other, less affective processes (STS/MTG, dIPFC).

With regard to those visual activations falling within the pain-localizer, the dominant interpretation in the literature has been to interpret them as evidence for empathy (Lamm et al., 2011). The rationale is that these voxels have BOLD activations predicting pain unpleasantness during shock experience, and they are activated while witnessing the pain of another individual, therefore their activation in the latter case causes the negative affective response to pain to be felt on behalf of the person that is seen to suffer. If interpreted in this way, what our experiment suggests is that the observers felt more vicarious pain when they witnessed a pain they caused. However, it is well known, that other negative affective responses also recruit similar voxels in the brain. In particular the experience and observation of disgust (Wicker et al., 2003) and the experience of guilt (Jankowski and Takahashi, 2014) also activates the AI, ACC and amygdala, and so do other salient stimuli (Legrain et al., 2011). Accordingly, some argue that activations in the 'pain-matrix' should not so much be viewed as unequivocal evidence for pain, but rather as evidence that an event triggered a complex reaction involving saliency detection, deployment of attention and prioritization of protective actions that also occurs following a noxious

stimulus (Legrain et al., 2011). This complex reaction is of course closely related to what it feels like to be in pain, but is not specifically linked to noxious stimuli. Hence, the boosts in activity associated with full responsibility, that we observe in voxels of the observer's brain involved in the first and third person pain perception, may reflect a mix of attentional and motor readiness states that are common to different negative emotions, including vicarious pain and guilt. This increase in activity may therefore be conservatively interpreted as reflecting a boost in the attentional and affective reaction to the pain of others, rather than as evidence for a specific boost in vicarious pain. Further studies will be needed to explore if the exact function to be associated with the vicarious activation of these voxels can be further decomposed, for instance, by mapping the activation pattern of these voxels with a larger spectrum of emotions (pain, guilt, salience etc.) to then explore using multi-voxel pattern classification, which vicarious patterns can be specifically associated with one negative form of affect, and which reflect more generic aspects (Wagner et al., 1998).

A further limitation to be kept in mind is that in our design the responsibility of the witness is inversely proportional to that of the victim: cases of full-responsibility for the observer were cases of full innocence of the victim. It might thus be that stronger brain activity in the Full-Responsibility situation was due to the observer's perceived responsibility, or to the fact that the observer perceived that an innocent victim was experiencing pain. We cannot dissociate these alternatives. Experiments with two confederates (a passive victim and a second player) would help overcome this limitation.

What functional implications might such a responsibility modulation of pain-matrix activity during pain observation have? Neural activity in the pain-matrix following noxious stimuli on our own body arguably serves to motivate us to stay out of harm's way in the future. By extension, activation in this matrix while witnessing the pain of others could motivate us to keep others out of harm's way. This vicarious activation, as a learning signal, should then be maximal if the observer caused the pain, because this is the condition with maximal causal relationship to the observer's own behavior. That pain-matrix activation reflects responsibility could thus serve as a mechanism that optimizes learning in a social context. Merely sharing responsibility was sufficient to decrease pain-matrix activity, and being the only responsible triggered the largest brain activity. This indirectly also shed light on the bystander effect (Darley and Latane, 1968), and advocates that if we want people to process the pain of others, it might be essential to maximize brain activations by actively emphasizing each potential viewer's personal responsibility.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.neuroimage.2015.03.034>.

Acknowledgments

VG was supported by a VENI grant 451-09-006 from the Netherlands Organization for Scientific Research (N.W.O.). CK and the research was supported by a VIDI grant of NWO 452-04-305 and a Marie Curie Excellence Grant MEXT-CT-2005-023253 and European Research Council grant 312511 of the European Commission. FC was supported by the China Scholarship Council for State Scholarship Fund.

References

- Avenanti, A., Sirigu, A., Aglioti, S.M., 2010. Racial bias reduces empathic sensorimotor resonance with other-race pain. *Curr. Biol.* 20, 1018–1022.
- Azevedo, R.T., Macaluso, E., Avenanti, A., Santangelo, V., Cazzato, V., Aglioti, S.M., 2013. Their pain is not our pain: Brain and autonomic correlates of empathic resonance with the pain of same and different race individuals. *Hum. Brain Mapp.* 34, 3168–3181.
- Bastiaansen, J.A., Thioux, M., Keysers, C., 2009. Evidence for mirror systems in emotions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 2391–2404.
- Benjamini, Y., 2010. Discovering the false discovery rate. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* 72, 405–416.

- Botvinick, M., Jha, A.P., Bylsma, L.M., Fabian, S.A., Solomon, P.E., Prkachin, K.M., 2005. Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain. *NeuroImage* 25, 312–319.
- Brett, M., Anton, J., Valabregue, R., Poline, J., 2002. Region of interest analysis using an SPM toolbox. The 8th International Conference on Functional Mapping of the Human Brain, Sendai, Japan.
- Carter, C.S., Braver, T.S., Barch, D.M., Botvinick, M.M., Noll, D., Cohen, J.D., 1998. Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science* 280, 747–749.
- Cheng, Y., Chen, C., Lin, C.P., Chou, K.H., Decety, J., 2010. Love hurts: an fMRI study. *NeuroImage* 51, 923–929.
- Chumbley, J.R., Friston, K.J., 2009. False discovery rate revisited: FDR and topological inference using Gaussian random fields. *NeuroImage* 44, 62–70.
- Corradi-Dell'Acqua, C., Hofstetter, C., Vuilleumier, P., 2011. Felt and seen pain evoke the same local patterns of cortical activity in insular and cingulate cortex. *J. Neurosci.* 31, 17996–18006.
- Craig, K., Prkachin, K., Grunau, R., 2001. The facial expression of pain. In: Turnk, D., Melzack, R. (Eds.), *Handbook of Pain Assessment*, 2nd ed. Guilford, New York.
- Darley, J.M., Latane, B., 1968. Bystander intervention in emergencies: diffusion of responsibility. *J. Pers. Soc. Psychol.* 8, 377–383.
- de Vignemont, F., Singer, T., 2006. The empathic brain: how, when and why? *Trends Cogn. Sci.* 10, 435–441.
- Fan, Y., Duncan, N.W., de, G.M., Northoff, G., 2011. Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. *Neurosci. Biobehav. Rev.* 35, 903–911.
- García-Larrea, L., Peyron, R., 2013. Pain matrices and neuropathic pain matrices: a review. *Pain* 154, S29–S43.
- Gazzola, V., Keysers, C., 2009. The observation and execution of actions share motor and somatosensory voxels in all tested subjects: single-subject analyses of unsmoothed fMRI data. *Cereb. Cortex* 19, 1239–1255.
- Genovese, C.R., Lazar, N.A., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage* 15, 870–878.
- Glascher, J., 2009. Visualization of group inference data in functional neuroimaging. *Neuroinformatics* 7, 73–82.
- Gu, X., Han, S., 2007. Attention and reality constraints on the neural processes of empathy for pain. *NeuroImage* 36, 256–267.
- Hein, G., Silani, G., Preuschoff, K., Batson, C.D., Singer, T., 2010. Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. *Neuron* 68, 149–160.
- Heldmann, M., Russeler, J., Munte, T.F., 2008. Internal and external information in error processing. *BMC Neurosci.* 9, 33.
- Jackson, P.L., Rainville, P., Decety, J., 2006. To what extent do we share the pain of others? Insight from the neural bases of pain empathy. *Pain* 125, 5–9.
- Jankowski, K.F., Takahashi, H., 2014. Cognitive neuroscience of social emotions and implications for psychopathology: Examining embarrassment, guilt, envy, and schadenfreude. *Psychiatry Clin. Neurosci.* 68, 319–336.
- Keysers, C., Wicker, B., Gazzola, V., Anton, J.L., Fogassi, L., Gallese, V., 2004. A touching sight: SII/PV activation during the observation and experience of touch. *Neuron* 42, 335–346.
- Koban, L., Corradi-Dell'Acqua, C., Vuilleumier, P., 2013. Integration of error agency and representation of others' pain in the anterior insula. *J. Cogn. Neurosci.* 25, 258–272.
- Lamm, C., Decety, J., Singer, T., 2011. Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *NeuroImage* 54, 2492–2502.
- Legrain, V., Iannetti, G.D., Plaghki, L., Mouraux, A., 2011. The pain matrix reloaded: a salience detection system for the body. *Prog. Neurobiol.* 93, 111–124.
- Lieberman, M.D., Cunningham, W.A., 2009. Type I and Type II error concerns in fMRI research: re-balancing the scale. *Soc. Cogn. Affect. Neurosci.* 4, 423–428.
- Mathalon, D.H., Whitfield, S.L., Ford, J.M., 2003. Anatomy of an error: ERP and fMRI. *Biol. Psychol.* 64, 119–141.
- Meffert, H., Gazzola, V., den Boer, J.A., Bartels, A.A., Keysers, C., 2013. Reduced spontaneous but relatively normal deliberate vicarious representations in psychopathy. *Brain* 136, 2550–2562.
- Melzack, R., Wall, P.D., 1965. Pain mechanisms: a new theory. *Science* 150, 971–979.
- Monfardini, E., Gazzola, V., Boussaoud, D., Brovelli, A., Keysers, C., Wicker, B., 2013. Vicarious neural processing of outcomes during observational learning. *Plos One* 8, e73879.
- Mouraux, A., Diukova, A., Lee, M.C., Wise, R.G., Iannetti, G.D., 2011. A multisensory investigation of the functional significance of the "pain matrix". *NeuroImage* 54, 2237–2249.
- Rainville, P., 2002. Brain mechanisms of pain affect and pain modulation. *Curr. Opin. Neurobiol.* 12, 195–204.
- Saarela, M.V., Hlushchuk, Y., Williams, A.C., Schurmann, M., Kalso, E., Hari, R., 2007. The compassionate brain: humans detect intensity of pain from another's face. *Cereb. Cortex* 17, 230–237.
- Said, C.P., Haxby, J.V., Todorov, A., 2011. Brain systems for assessing the affective value of faces. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 366, 1660–1670.
- Shane, M.S., Stevens, M., Harenski, C.L., Kiehl, K.A., 2008. Neural correlates of the processing of another's mistakes: a possible underpinning for social and observational learning. *NeuroImage* 42, 450–459.
- Simon, D., Craig, K.D., Miltner, W.H., Rainville, P., 2006. Brain responses to dynamic facial expressions of pain. *Pain* 126, 309–318.
- Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R.J., Frith, C.D., 2004. Empathy for pain involves the affective but not sensory components of pain. *Science* 303, 1157–1162.
- Singer, T., Seymour, B., O'Doherty, J.P., Stephan, K.E., Dolan, R.J., Frith, C.D., 2006. Empathic neural responses are modulated by the perceived fairness of others. *Nature* 439, 466–469.
- Taylor, S.F., Stern, E.R., Gehring, W.J., 2007. Neural systems for error monitoring: recent findings and theoretical perspectives. *Neuroscientist* 13, 160–172.
- Thirion, B., Pinel, P., Meriaux, S., Roche, A., Dehaene, S., Poline, J.B., 2007. Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. *NeuroImage* 35, 105–120.
- Ullsperger, M., von Cramon, D.Y., 2003. Error monitoring using external feedback: specific roles of the habenular complex, the reward system, and the cingulate motor area revealed by functional magnetic resonance imaging. *J. Neurosci.* 23, 4308–4314.
- van der Gaag, C., Minder, R.B., Keysers, C., 2007. Facial expressions: what the mirror neuron system can and cannot tell us. *Soc. Neurosci.* 2, 179–222.
- Wager, T.D., Atlas, L.Y., Lindquist, M.A., Roy, M., Woo, C.W., Kross, E., 2013. An fMRI-based neurologic signature of physical pain. *N. Engl. J. Med.* 368, 1388–1397.
- Wagner, A.D., Schacter, D.L., Rotte, M., Koutstaal, W., Maril, A., Dale, A.M., Rosen, B.R., Buckner, R.L., 1998. Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. *Science* 281, 1188–1191.
- Wicker, B., Keysers, C., Plailly, J., Royet, J.P., Gallese, V., Rizzolatti, G., 2003. Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. *Neuron* 40, 655–664.
- Williams, A.C., 2002. Facial expression of pain: an evolutionary account. *Behav. Brain Sci.* 25, 439–455.
- Xu, X., Zuo, X., Wang, X., Han, S., 2009. Do you feel my pain? Racial group membership modulates empathic neural responses. *J. Neurosci.* 29, 8525–8529.