Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

Temporal dynamics of implicit moral evaluation: From empathy for pain to mentalizing processes

Kamela Cenka [a,**], Chiara Spaccasassi [a], Stella Petkovic [a,b], Rachele Pezzetta [c], Giorgio Arcara [c], Alessio Avenanti [a,d,*]

[a] Centro Studi e ricerche in Neuroscienze Cognitive, Dipartimento di Psicologia "Renzo Canestrari", Alma Mater Studiorum Università di Bologna, Cesena Campus, Cesena, Italy
[b] Sapienza, University of Rome and CLN2S@SAPIENZA, Istituto Italiano di Tecnologia, Rome, Italy
[c] IRCCS Ospedale San Camillo, Venice, Italy
[d] Centro de Investigación en Neuropsicología y Neurociencias Cognitivas, Universidad Católica del Maule, Talca, Chile

A B S T R A C T

To understand how we evaluate harm to others, it is crucial to consider the offender's intent and the victim's suffering. Previous research investigating event-related potentials (ERPs) during moral evaluation has been limited by small sample sizes and a priori selection of electrodes and time windows that may bias the results. To overcome these limitations, we examined ERPs in 66 healthy human adults using a data-driven analytic approach involving cluster-based permutation tests. Participants performed an implicit moral evaluation task requiring to observe scenarios depicting intentional harm (IHS), accidental harm (AHS), and neutral actions (NAS) while judging whether each scenario was set indoors or outdoors. Our results revealed two distinct clusters, peaking at ~170 and ~250 ms, showing differences between harm scenarios (IHS and AHS) and NAS, suggesting rapid processing of the victim's physical outcome. The difference between IHS and AHS scenarios emerged later, at ~400 ms, potentially reflecting subsequent evaluation of the agent's intentions. Source analysis showed that brain regions associated with empathy for pain were associated with the earlier peaks at ~170 and ~250 ms, while the modulation of the activity of the mentalizing network was presented at ~250 and ~400 ms. These findings advance our understanding of the neural mechanisms underlying implicit moral evaluation. Notably, they provide electrocortical new insights for models of implicit moral evaluation, suggesting an early neural response linked to empathy for pain, with subsequent integration of empathy response with mentalizing processes, followed by later cognitive evaluations, likely reflecting the assessment of the agent's moral responsibility.

1. Introduction

Mature moral reasoning is a complex process that integrates cognitive, emotional, and motivational factors (Decety and Cacioppo, 2012; Decety and Howard, 2013; Fittipaldi et al., 2023). When evaluating the morality of others' actions, particularly in situations involving harm, two key elements come into play: the intention behind the offender's action and the suffering experienced by the victim (Young et al., 2007; Cushman, 2008; Greene, 2009; Reniers et al., 2012; for a review, see Buon et al., 2016).

Central to moral evaluation is the process of mentalizing (Knobe, 2005), which involves constructing cognitive representations of others' mental states and reasoning about their thoughts, beliefs, and emotions (Frith and Frith, 2005; Mitchell, 2009; Adolphs, 2009). Neuroimaging studies have identified a network of brain regions associated with mentalizing during moral judgment tasks, including the temporo-parietal junction (TPJ) and medial prefrontal cortex (mPFC) (Young et al., 2007; Ciaramelli et al., 2012; Reniers et al., 2012; for a review, see Young and Tsoi, 2013). TPJ appears crucial for encoding and integrating others' mental states with task-relevant information (Young

and Saxe, 2009a, 2009b), while the mPFC is engaged during the perception of moral violations (Greene et al., 2001; Moll et al., 2002) and predicts the degree of condemnation toward offenders (Young and Saxe, 2009b).

Alongside mentalizing, empathy plays a role in moral evaluation, particularly when assessing harm scenarios. Empathy, defined as the ability to share and understand others' emotional experiences (de Vignemont and Singer, 2006; Avenanti et al., 2009; Zaki and Ochsner, 2012), relies on neural resonance mechanisms: when observing someone else experiencing a sensory or emotional state, the observer's brain exhibits similar patterns of activation, reflecting experience-sharing processes (Singer et al., 2004; Avenanti et al., 2005; for a review, see Zaki and Ochsner, 2012). Neural resonance encompasses experiences related to motor intentions (Tidoni et al., 2013; Avenanti et al., 2018; Patri et al., 2020), sensory experiences (Bufalari et al., 2007; Valchev et al., 2017), and emotional states (Singer et al., 2004; Azevedo et al., 2013; Paracampo et al., 2018; for reviews, see Keysers et al., 2010; Rizzolatti and Sinigaglia, 2010; Lamm and Singer, 2010). In line with that, studies have revealed that perceiving others in pain actives brain regions associated with first-hand physical pain, including the anterior insula (AI) and the anterior midcingulate cortex (aMCC) (Singer et al., 2004; Morrison et al., 2004; Botvinick et al., 2005; Jackson et al., 2005; Corradi-Dell'Acqua et al., 2011), as well as the sensorimotor cortex (Avenanti et al., 2005; Bufalari et al., 2007; Jackson et al., 2006; Benuzzi et al., 2018; for reviews, see Riečanský and Lamm, 2019; Bagnis et al., 2020; Fallon et al., 2020).

Electroencephalography (EEG) studies have provided insights into the temporal dynamics of empathic brain response. Seminal research using event-related potentials (ERPs) has investigated neural responses to stimuli depicting pain applied to different body parts (Fan and Han, 2008; Decety et al., 2010) or painful facial expressions (Sheng and Han, 2012; Sheng et al., 2016). These studies suggest an early response to painful scenes in frontal electrodes (110–190 ms), which correlates with pain ratings and likely reflects experience-sharing processes. This is followed by later modulations in central-parietal regions (>300 ms), influenced by tasks demands and cognitive evaluation, likely reflecting mentalizing processes (Fan and Han, 2008; Decety et al., 2010; for a review see Zaki and Ochsner, 2012).

These neuroscientific findings appear in line with a two-component model of moral decision-making, as previously proposed for moral dilemmas (Greene et al., 2001, 2004, 2008; Haidt, 2007). This model posits that moral judgment involves prepotent emotional responses stemming from automatic evaluation, followed by slower, controlled responses – a feature shared by general dual-systems models of decision-making (e.g., Kahneman and Frederick, 2002; Loewenstein et al., 2008). These theoretical models make a crucial prediction: the automatic evaluation of a painful outcome should occur early in time, whereas the consideration of the agent's intention should occur later.

Despite prior attempts to characterize the neural dynamics of moral cognition using EEG, to date, direct neurophysiological evidence supporting these models remains inconclusive. ERPs studies have reported a heterogenous pattern of results, with modulations observed in three distinct ERP components related to moral evaluation: N1, P2/N2, and Late Positive Potential (LPP) (Decety and Cacioppo, 2012; Yoder and Decety, 2014; Pasion et al., 2019; Fernandes et al., 2022; for a review see Wagner et al., 2017). The first component, N1, appeared to be influenced by the emotional valence of the stimulus (Olofsson et al., 2008; Yoder and Decety, 2014; Gui et al., 2016). The second component, P2, was associated with viewing moral images compared to immoral ones, and high arousal compared to low arousal pictures (Tao et al., 2022), while its negative counterpart, N2, was associated with morally good actions (Yoder and Decety, 2014). Finally, LPP appeared to be critically associated with moral reasoning (Wagner et al., 2017), as its amplitude was more pronounced for intentional harm scenarios compared to attempted harm scenarios (Gan et al., 2016).

Taken together, these ERP results fit with two-component models of moral cognition. However, the lack of consistency across the studies makes it difficult to draw clear conclusions about the neural bases of moral cognition and hampers the development of a reliable theoretical framework for understanding how the brain processes moral information. The lack of consistency across studies could be attributed to several factors, including the relatively small sample size and the heterogenous theory-driven approaches for the selection of electrodes and time windows of interest in EEG signals.

To address these limitations, in this study, we report ERPs from a substantial sample of participants (N = 66) using a data-driven approach to EEG analysis. Following the methodology of previous studies (Decety and Cacioppo, 2012; Escobar et al., 2014; Yoder and Decety, 2014; Hesse et al., 2016), using a more extensive and validated set of stimuli (Spaccasassi et al., 2023), we presented participants with three consecutive and rapid frames depicting Intentional Harm Scenarios (IHS), Accidental Harm Scenario (AHS), and Neutral Action Scenarios (NAS), which allowed us to assess experience-sharing and mentalizing processes in the brain (Baez et al., 2014; Zaki and Ochsner, 2012). Participants engaged in an implicit moral evaluation task, where they observed these scenarios while performing an unrelated attentional task. Then, at the end of the experiment, participants were required to assess the moral permissibility, malicious intentionality, and perceived physical pain associated with the actions depicted in the experimental videoclips.

We expected to find a modulation of ERPs across the earlier temporal windows identified in prior research (i.e., N1, N2/P2) when watching harmful scenarios (i.e., IHS and AHS) compared to innocuous actions (i.e., NAS), possibly reflecting fast processing of painful outcomes. In line with that, we also expected to find a slower differential response to IHS compared to AHS, reflecting later processing of the agent's intention. Finally, we utilized source analysis to estimate potential neural generators of our ERP components, aiming to provide a comprehensive understanding of the spatiotemporal dynamics of implicit moral evaluation.

2. Materials and methods

We report how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study. Materials and data for the study are available at https://osf.io/hnx9d/. No part of the study procedures or analyses was preregistered prior to the research being conducted.

2.1. Participants

66 participants (22 males; mean age ± SD: 23.75 y ± 2.14) took part in the experiment. All participants were right-handed, had normal or corrected-to-normal vision, and were naïve to the purpose of the experiment. Participants provided written informed consent. The study procedures received approval from the Bioethics committee at the University of Bologna and adhered to the ethical standards outlined in the 1964 Declaration of Helsinki (World Medical Association, 2013).

2.2. Main experiment

Participants were seated in a quiet, dimly lit room, approximately 60 cm away from a 15″ computer screen. They were instructed to watch 36 videoclips validated in a previous study (Spaccasassi et al., 2023; see its Supplementary Material for validation data). Each videoclip consisted of three frames featuring two characters in three conditions: *i*) IHS, showing an active character (the agent) intentionally harming a passive character (the victim; Intentional Harming Scenario); *ii*) AHS, where the agent unintentionally harmed the victim; Accidental Harming Scenario; *iii*) NAS, showing neutral interaction between the two characters (Neutral Action Scenario). Importantly, IHS involved the agent's

negative intent and a negative outcome for the victim, while AHS featured neutral intent but still resulted in a negative outcome for the victim. NAS had both neutral intent and outcome. The actors' faces were not visible, and the agent's action never resulted in the victim's death. Each trial began with a 3000-ms fixation cross in the center of the screen. Then, the videoclips were presented in three consecutive frames: (*i*) introducing the characters and setting (500 ms); (*ii*) revealing the agent's intention through their actions (200 ms); (*iii*) showing the outcome of the action on the victim (2000 ms) (Fig. 1). Scenarios were presented in a random order.

EEG was continuously recorded throughout the experiment. During the third frame, in a timing varying between 200 and 1200 ms from its onset, a pulse of transcranial magnetic stimulation (TMS) was administered over a target area (right and left TPJ or right primary somatosensory cortex). Importantly, there was an approximately 10-s interpulse interval between trials to prevent cumulative changes in neural activity due to TMS delivery (Valero-Cabré et al., 2017). ERPs analyzed in the presented study focused on EEG signal uncontaminated by TMS (see below), and TMS effects on EEG signal will be the focus of separate publications. Furthermore, additional analyses were conducted to ensure that our ERP results were not influenced by any TMS-related effects (see Supplementary Material for further details).

After presenting each scenario, participants had to report where the scene took place, indoors or outdoors, by pressing two keys of a keyboard with their right hand. All participants scored with >95% accuracy ensuring they paid attention to the task. Following the response, feedback was displayed on the screen for 2000 ms, followed by a 1000 ms blank screen, marking the end of the trial. Each videoclip was repeated 8 times, resulting in 288 trials (96 per condition). Every 72 trials participants could have a break. The whole experiment lasted 50 min. The study was programmed using PsychtoolBox (Brainad &Vision, 1997) within Matlab R2019b (MathWorks Inc.) to control EEG events and trigger the TMS.

2.3. Subjective assessment

After the main experiment, participants were reintroduced to the same scenarios, presented in a random order, and asked to evaluate them using a modified version of the Empathy for Pain Task (EPT) (Baez et al., 2012, 2013, 2014; Couto et al., 2013; Spaccasassi et al., 2023). Each trial started with a white fixation cross lasting 300 ms, followed by the first frame (500 ms), the second frame (200 ms), and the third frame (1000 ms). At the end of each videoclip, participants were asked to rate, on a 9 points Likert scale, the degree to which they thought (*i*) the passive character felt physical pain as a result of the agent's action (Victim's physical Pain; 1 = not at all, 9 = completely); (*ii*) the agent's intention was malicious (Malicious Intention; 1 = not at all, 9 = completely); (*iii*) the agent's action was morally permissible (Moral Permissibility; 1 = not at all, 9 = completely). The whole assessment lasted 10 min. The experiment was programmed using PsychToolBox running in Matlab.

2.4. EEG recording

EEG signal was acquired using a BrainAmp DC EEG amplifier (BrainProducts GmBH, Germany) with 64 sintered active electrodes (ActiCHamp ActiCAP, Germany) mounted on an elastic cap according to the 10/20 coordinate system. Reference and ground electrodes corresponded to FCz and Az electrodes, respectively. The impedance level was kept below 10Ω for each electrode. The sampling rate was set at 1000 Hz.

2.5. EEG preprocessing

EEG data were analyzed using Matlab R2019b and Fieldtrip (http://www.fieldtriptoolbox.org) (Oostenveld et al., 2011), following the processing pipeline outlined by Herring et al. (2015) (http://www.fieldtriptoolbox.org/tutorial/tms-eeg/). Initially, the EEG signal was segmented into 3000-ms epochs, ranging from −2000 ms to 1000 ms
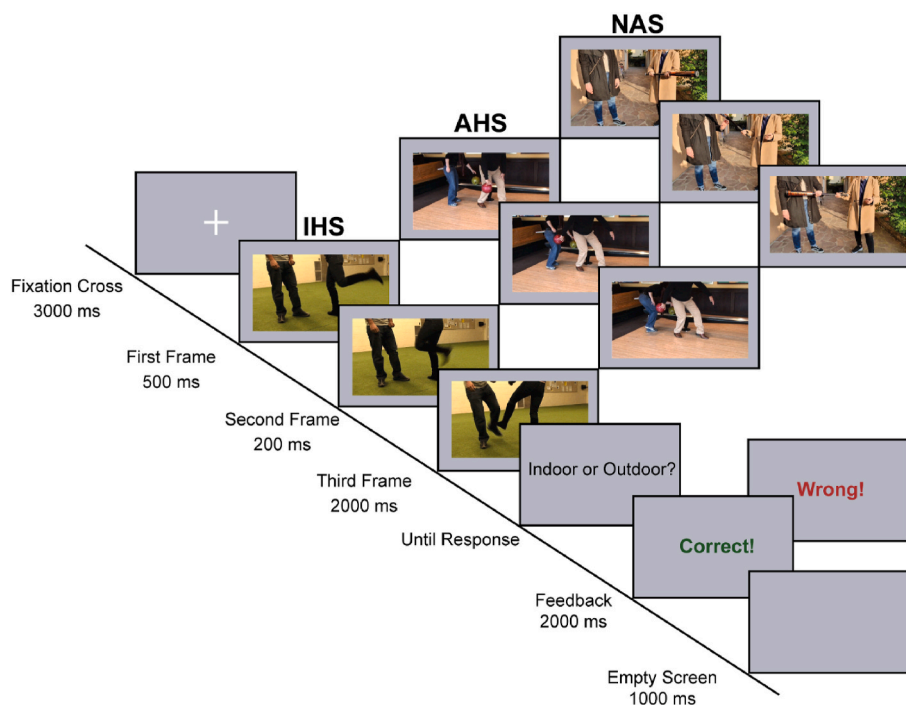

Fig. 1. Experimental procedure. Each trial began with a 3000-ms fixation cross, followed by the videoclip consisting of three frames: first frame (500 ms), second frame (200 ms) and third frame (2000 ms). Subsequently, an unrelated question appeared on the screen and remained visible until participants provided their responses. Afterward, visual feedback was displayed for 2000 ms. Finally, each trial ended with a 1000 ms blank screen. The three frames of the scenarios depict, from left to right: IHS, AHS, and NAS.

after the onset of the first frame. These epochs fully covered the presentation of the first and second frames and included only the initial 300 ms of the third frame (for further details refer to the Main Experiment section). This approach minimized the impact of TMS pulse-related EEG noise, as TMS pulse could occur randomly between 200 ms and 1200 ms after the onset of the third scenario. Approximately 8% of trials containing TMS artifacts were excluded, relying on the verification of TMS delivery timing in each trial, ensuring precise artifact removal. Additionally, 6% of trials were rejected using the criterium of 10^4 variance thresholds (for similar procedures see Zanon et al., 2018; Spaccasassi et al., 2022). Independent Component Analysis (FastICA) was employed to identify and discard bad components reflecting cranial muscle artifacts (Korhonen et al., 2011), eye-blinks, and saccades (Jung et al., 2000) through visual inspection (total rejection rate: <5%). The EEG signal was re-referenced to the channel average and downsampled to 500 Hz. A band-pass filter was applied (0.01–35 Hz). Subsequently, time-locked analysis was performed on the processed EEG signal to extract event-related potentials (ERPs). ERPs were baseline-corrected using a 300-ms absolute baseline window (from −301 to −10 ms) taken from the fixation cross. Finally, grand averages across trials and participants were computed for each experimental condition.

2.6. Subjective evaluations analysis

To analyze subjective evaluations, three repeated-measures analyses of variance (ANOVAs) were employed, with the within-subject factor Scenario with 3 levels (IHS, AHS, and NAS). Post-hoc comparisons were conducted using the Bonferroni correction, while effect sizes were estimated by computing partial-eta^2 (η_p^2). All behavioral analyses were performed with R software 3.6.2 (R Core Team, 2021) using Jamovi 2.3.12 (Şahin and Aybek, 2019).

2.7. EEG analysis

We employed within-subject cluster-based permutation tests. Specifically, we compared IHS vs. AHS, IHS vs. NAS, and AHS vs. NAS using Montecarlo computation in Fieldtrip (Maris and Oostenveld, 2007). Clusters were defined as collections of two or more neighboring electrodes where the t-statistic at a specific time point exceeded a threshold of $p < 0.001$ (two-tailed paired samples t-test). We performed cluster-based permutation tests for each comparison based on 1000 random permutations, maintaining an alpha level below 0.05. We conducted these analyses by focusing on the time window corresponding to the first frame of the video (i.e., 0–500 ms). While prior research has explored how naturalistic viewing conditions influence neural processing of dynamic stimuli (Welke and Vessel, 2022), our approach aimed to minimize temporal variability that could influence ERP responses. Although this approach may reduce the ecological validity of the results, employing static images ensured consistent temporal alignment across all responses, thereby increasing the signal-to-noise ratio.

For effect size estimation (*Cohen's d*) related to each significant cluster, we selected data from specific channels and time-points identified as significant in the permutation test.

2.8. Regression analyses

As detailed in the results section, we identified three main ERP peaks within the significant clusters: two occurring around 170 and 250 ms, which differentiated painful (AHS and IHS) from non-painful (NAS) scenes; and a third peak around 400 ms, which distinguished between IHS and AHS. Specifically, for the early indices, we calculated the mean voltage of IHS and AHS relative to NAS at 170 and 250 ms using the following formulas:

$$\Delta 170 = \frac{IHS_{170} + AHS_{170}}{2} - NAS_{170}$$

$$\Delta 250 = \frac{IHS_{250} + AHS_{250}}{2} - NAS_{250}$$

For the late index, associated with the discrimination of intentions, we computed the mean voltage at 400 ms related to the IHS vs. AHS difference as it follows:

$$\Delta 400 = IHS_{400} - AHS_{400}$$

Thus, the first two indices ($\Delta 170$ and $\Delta 250$) captured early neural responses to the perception of the victims' pain, tapping into experience-sharing processes, whereas the later index ($\Delta 400$) was associated with understanding the agents' intentions, thereby reflecting mentalizing processes.

To explore the relationships between electrophysiological and behavioral data, we conducted three regression models, using physical pain and malicious intentionality ratings as predictors for each of the three ERP indices. For each model, we computed a behavioral index corresponding to the respective ERP index. Specifically, to test whether physical pain and malicious intentionality ratings predicted ERP peaks associated with the perception of others' pain ($\Delta 170$ and $\Delta 250$), we used the difference between the mean ratings of painful scenes (IHS and AHS) and non-painful scenes (NAS) for both measures. Conversely, to test whether physical pain and malicious intentionality predicted ERP peaks associated with understanding others' intentions ($\Delta 400$) we used the difference between IHS and AHS scores for each rating as predictors.

2.9. Source estimation

To estimate ERP activity at the source level we used the Brainstorm toolbox (http://neuroimage.usc.edu/brainstorm; Tadel et al., 2011). For the forward solution modeling, we used the OpenMEEG BEM algorithm (Gramfort et al., 2010) with 15002 vertices and the ICBM 152 as a template anatomy. The noise covariance matrix was calculated using the baseline epochs (from −301 to −10 ms). Inverse solution estimation was conducted using Minimum Norm Imaging (MNI) normalization (Baillet et al., 2001) with sLORETA (Standardized Low Resolution Brain Electromagnetic Tomography; Pascual-Marqui, Michel and Lehmann, 1994), and was restricted to the cortical surface using a constrained dipole direction. Data were rectified and spatially smoothed with a Gaussian kernel with a Full-Width Half Maximum (FWHM) of 3 mm. To identify the source of the effects observed at the sensor level, we conducted paired-sample t-tests on selected regions of interest (ROIs) and whole-brain levels at three specific time points: 170 ms, 250 ms, and 400 ms following stimulus onset, corresponding to the three main ERP peaks (see Results). The first two timepoints reflect the discrimination between painful (IHS and AHS) and non-painful (NAS) scenarios, while the later timepoint reflects the discrimination between IHS and AHS. To account for multiple comparisons, we applied the False Discovery Rate (FDR) correction.

We selected 10 different ROIs subdivided into two sets: *i*) the bilateral aMCC, insula and supramarginal gyrus (SMG), as these brain regions are mostly associated with self-pain and pain empathy, following Fallon's meta-analysis (Fallon et al., 2020); *ii*) the bilateral TPJ and vmPFC, as these areas are associated with mentalizing and, consequently, with the processing of moral cognition, as reported by Bzdok's meta-analysis (Bzdok et al., 2012). To determine our ROIs, we identified coordinates reported in previous meta-analyses and selected brain regions based on the use of the Brainnetome Atlas (Fan et al., 2016). For ROI analysis, we conducted a comparison based on the maximum values among all vertices contained within the specific ROI (Arutiunian et al., 2022). Due to potential ambiguities in interpreting signs in source estimation, significance was interpreted based on temporal and spatial effects, without considering the direction (Stropahl et al., 2018). Consequently, all statistical values will be reported in absolute terms.

3. Results

3.1. Cluster-based Permutation Tests on ERPs

We identified six significant clusters when comparing conditions. In each comparison, we detected two clusters with opposite polarity, one with positive voltages located posteriorly and another with negative voltages located anteriorly.

IHS vs. NAS: we observed a left posterior cluster in the 121–419 ms time window ($p < 0.001$; *Cohen's d = −0.68*; Fig. 2a), and a mostly left central cluster emerging later in the 241–500 ms time window ($p < 0.001$; *Cohen's d = 0.54*; Fig. 2b), both showing larger voltages for NAS compared to IHS (mean voltage ± S.D. in the posterior cluster: IHS: 4.6

± 2.2 μV; NAS: 5.1 ± 2.3; central cluster: IHS: −3.5 ± 1.7 μV; NAS: −3.9 ± 1.7 μV).

AHS vs. NAS: we observed a posterior cluster in the 169–500 ms time window ($p < 0.001$; *Cohen's d = −0.94*; Fig. 2c), and a central cluster in the 159–500 ms time window ($p < 0.001$; *Cohen's d = 0.99*; Fig. 2d), both showing larger voltages for NAS than for AHS (posterior cluster: AHS: 5.5 ± 2.4 μV; NAS: 6.3 ± 2.5 μV; central cluster: AHS: −2.8 ± 1.3 μV; NAS: −3.3 ± 1.3 μV).

IHS vs. AHS: we observed a mostly left posterior cluster ($p < 0.001$; *Cohen's d = 0.66*), and a right central cluster ($p < 0.001$; *Cohen's d = −0.50*), both starting at 333 ms, continuing throughout the time window of interest (333–500 ms) and showing greater voltages for IHS than for AHS (posterior cluster: IHS: 7.0 ± 3.1 μV; AHS: 6.39 ± 3.1 μV; Fig. 2e;
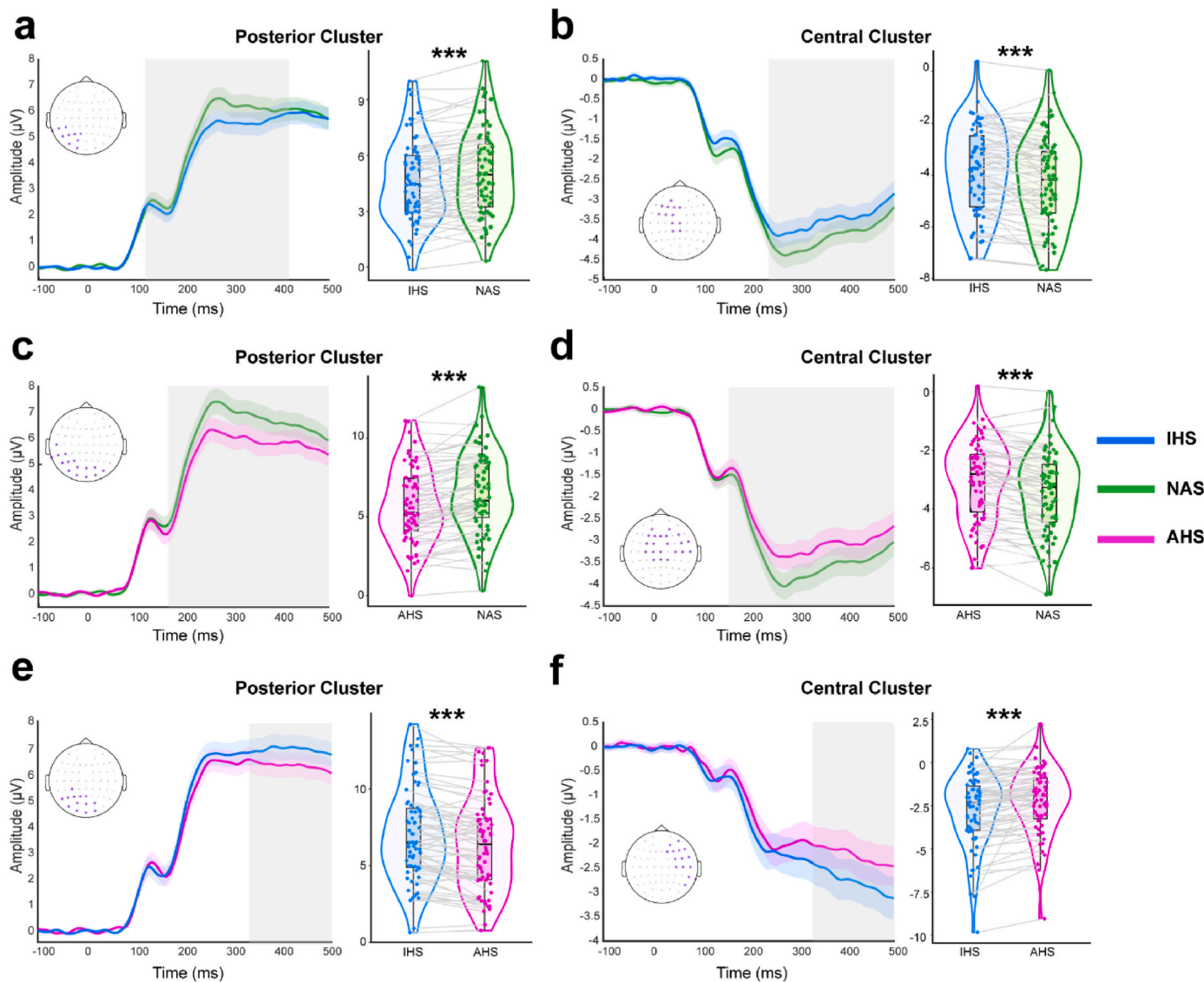

Fig. 2. Significant ERPs emerging from cluster-based permutation tests. The IHS vs NAS comparisons showed two significant clusters **(a)** one over left posterior electrodes (TP9, CP5, P3, P7, O1, TP7, P1, P5, PO7, PO3) in the 121–419 ms time-window and **(b)** the other over left central electrodes (Fz, F3, FC5, FC1, CP1, Cz, AF3, F1, F5, FC3, C1, CPz) in the 241–500 ms time-window. In both clusters, IHS ERPs were smaller than NAS ERPs. The AHS vs NAS comparisons showed two significant clusters **(c)** one over posterior electrodes (T7, TP9, CP5, P3, P7, O1, Oz, O2, TP7, P5, PO7, PO3, POz, PO4, PO8) in the 169–500 ms time window and **(d)** the other over central electrodes (Fz, F3, FC1, C3, CP1, CP2, Cz, C4, T8, FC6, FC2, AF3, F1, FC3, C1, CPz, C6, FT8, F6, F2, AF4). In both clusters, AHS ERPs were smaller than NAS ERPs. The IHS vs AHS comparisons showed two significant clusters in the 333–500 ms time-window, **(e)** over posterior electrodes (Pz, P3, P7, O1, Oz, O2, CP3, P1, P5, PO7, PO3, POz, PO4) and (f) over right central electrodes (P8, AF4, AF8, F6, FT8, FC4, C6, TP8, F8, F4, FC6, T8, C4, TP10). In both clusters, IHS ERPs were larger than AHS ERPs. In all panels, shaded bands indicate standard errors around the mean and the grey areas highlight the time windows in which the analysis showed significant differences. Significant sensors are highlighted in the scalp maps, and violin plots and boxplots of ERP amplitudes are shown, with light grey lines connecting subject observations in the two conditions. IHS is represented in blue color, AHS in magenta, and NAS in green. The significance level is shown by asterisks (***$p < 0.001$).

central cluster: IHS: -2.8 ± 2.1 μV; AHS: -2.2 ± 2.0 μV; Fig. 2f).

3.2. Source analysis

Source analysis was conducted on the three main ERP peaks identified within the six clusters. The ROI analysis conducted on the first peak (i.e., at 170 ms, corresponding to the initial EEG activity distinguishing between painful vs. non-painful scenarios; Fig. 2a–c, d) revealed modulation in two empathy-related ROIs, specifically the left SMG ($t = 3.5$) and the right insula ($t = 2.7$), where the mean activity during IHS and AHS differed from NAS (all $p < 0.05$, *FDR corrected*; Fig. 3a). No other significant modulation was observed across empathy or mentalizing ROIs. Whole brain analysis conducted on the same peak did not yield significant results (Fig. 4a).

The ROI analysis conducted on the second peak (i.e., at 250, aligning EEG activity discriminating between painful vs. non-painful scenarios, Fig. 2a, b, c, d) showed modulation in several empathy ROIs, including the left aMCC ($t = 2.5$), the left ($t = 6.4$) and right ($t = 4.8$) SMG, and the right AI ($t = 2.3$), where the mean activity during IHS and AHS differed from NAS (all $p < 0.05$; *FDR corrected*). We also found modulation in two mentalizing ROIs, namely the left ($t = 8.1$) and right ($t = 4.9$) TPJ (all $p < 0.05$; *FDR corrected*) (Fig. 3b). The whole-brain analysis also revealed an extended bilateral network involved in higher-order visual processing of social cues, encompassing the lateral occipitotemporal cortex and the inferotemporal cortex, and the concurrent activation of the left inferior parietal lobule overlapping with the ventral attentional network, as well as activation in the sensorimotor cortices; on the medial surface, we observed bilateral activation of the posterior cingulate and medial parietal and temporal cortex, overlapping with the DMN; additionally, we observed an activation of the left aMCC and bilateral insula extending into the frontal and parietal opercula, involving sensorimotor and affective regions classically associated with first-person and

vicarious experience of pain, with some regions implicated in salience processing (all $p < 0.05$; *FDR corrected*; Fig. 3b).

The ROI analysis conducted on the third peak (i.e., at 400 ms, corresponding to the onset of EEG activity distinguishing between the IHS and AHS; Fig. 2e and f) did not show any differences across empathy ROIs, while results on the moral ROIs, showed differences in the left TPJ ($t = 3.4$, $p < 0.05$, *FDR corrected*) (Fig. 3c). Whole-brain analysis on the third timepoint showed an activation of the posterior cingulate and retrosplenial cortex, the precuneus and medial parietal regions overlapping with the DMN (Fig. 4c).

3.3. Subjective evaluations

Physical Pain. We found a significant main effect of Scenario ($F_{2,130} = 1398$; $p < 0.001$; $\eta_p^2 = 0.96$; Fig. 5a), showing that pain ratings for AHS (M. $\pm$ S.D. = 6.24 ± 1.00) and IHS (6.11 ± 1.01) were comparable ($p = 0.25$) and larger than ratings for NAS (1.03 ± 0.12; all $p < 0.001$).
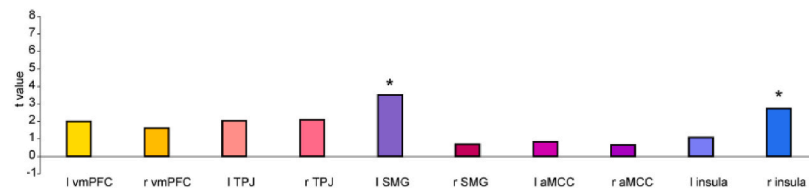
Moral Permissibility. We found a significant main effect of Scenario ($F_{2,130} = 692$; $p < 0.001$; $\eta_p^2 = 0.91$; Fig. 5b), with the three scenarios significantly differing from each other (all $p < 0.001$); the lowest permissibility ratings were observed for IHS (2.36 ± 0.95), followed by AHS (4.56 ± 1.57), and then NAS (8.58 ± 0.50).

Malicious Intentionality. We found a significant main effect of Scenario ($F_{2,130} = 971$; $p < 0.001$; $\eta_p^2 = 0.94$; Fig. 5c), indicating that the three scenarios significantly differed from each other (all $p < 0.001$); the highest malicious intentionality ratings were attributed to IHS (7.51 ± 0.88), followed by AHS (3.01 ± 0.94), and then NAS (1.31 ± 1.03).
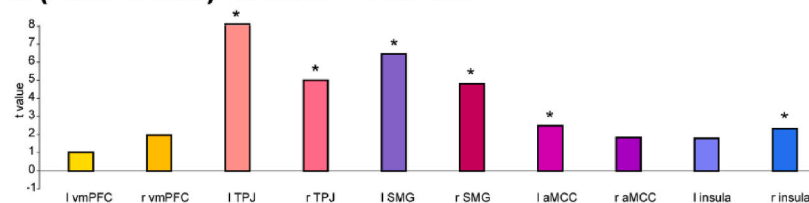
3.4. Relation between electrophysiological results and subjective ratings

We identified three main ERP peaks across the significant clusters. Two occurred at ~170 and ~250 ms and distinguished painful (IHS and
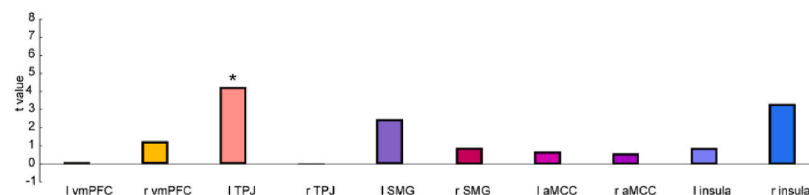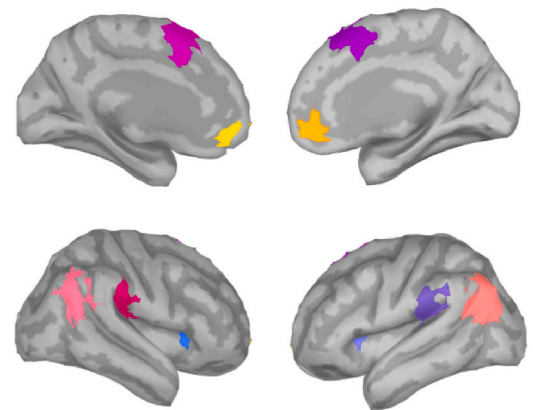

Fig. 3. Source analysis on ROIs. (a) Histograms of the significant activation (absolute t values) resulting from the paired sample *t*-test on ROIs comparing painful (AHS & IHS) and non-painful (NAS) interactions at ~170 ms. (b) Histograms of the significant activation (absolute t values) resulting from the paired sample *t*-test on ROIs comparing painful (AHS & IHS) and non-painful (NAS) interactions at ~250 ms. (c) Histograms of the significant activation (absolute t values) resulting from the paired sample *t*-test on ROIs, comparing IHS and AHS at ~400 ms. Asterisks indicate statistical significance ($p < 0.05$). (d) ROIs selected from the Brainnetome atlas.
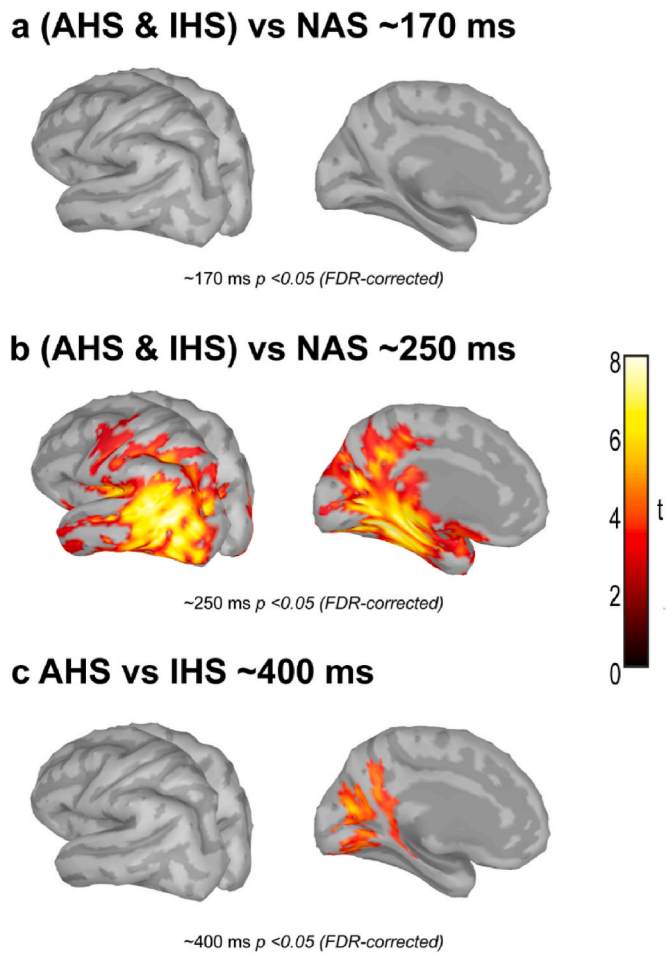
a (AHS & IHS) vs NAS ~170 ms

~170 ms p <0.05 (FDR-corrected)

b (AHS & IHS) vs NAS ~250 ms

~250 ms p <0.05 (FDR-corrected)

c AHS vs IHS ~400 ms

~400 ms p <0.05 (FDR-corrected)

Fig. 4. Whole-brain source analysis. (a) Results of paired sample *t*-test comparing painful (AHS & IHS) and non-painful (NAS) interactions at ~170 ms. **(b)**. Results of paired sample *t*-test comparing painful (AHS & IHS) and non-painful (NAS) interactions at ~250 ms. **(c)** Results of paired sample *t*-test comparing IHS and AHS at ~400 ms. The absolute values of the significant activations (*t* values, $p < 0.05$, *FDR corrected*) are depicted on a template cortex, revealing significant broad cortical activation at 250 ms in the comparison between painful and non-painful interactions and a subsequent smaller activation at 400 ms when comparing IHS and AHS.

AHS) from non-painful (NAS) scenes, thus reflecting early neural responses to the perception of the victims' pain (Δ170, Δ250). A third, later peak at ~400 ms distinguished between IHS and AHS, thus reflecting understanding of agents' intentions (Δ400).

To explore the relationship between our ERP data and behavioral results, each of the three indices was entered into a regression model with corresponding indices of physical pain and malicious intentionality as predictors.

The initial regression model for the earliest peak (Δ170) did not reach statistical significance ($R^2 = 0.06$, $F_{2,63} = 1.98$, $p = 0.146$). However, after excluding four statistical outliers with standard residuals greater than 2 SD, the regression model became significant ($R^2 = 0.12$, $F_{2,59} = 3.92$, $p = 0.025$). Specifically, the pain rating index positively predicted the modulation of Δ170 ($\beta = 0.33$, $p = 0.010$, Fig. 6a), while the malicious intentionality index did not ($\beta = 0.06$, $p = 0.61$, Fig. 6b). Neither the regression model for Δ250 ($R^2 = 0.06$, $F_{2,63} = 2.32$, p = 0.11) nor the model for the late peak index (Δ400) reached statistical significance ($R^2 = 0.02$, $F_{2,63} = 0.92$, $p = 0.404$).

4. Discussion

In the present study, we used ERPs to investigate the brain dynamics associated with implicit moral evaluation of naturalistic scenarios depicting harm to a victim. Our findings revealed consistent electrocortical differences across scenarios, manifesting in two clusters with opposite polarities: one distributed over posterior electrodes, displaying a positive voltage, and the other over central electrodes, exhibiting a negative voltage. Notably, we observed an early ERP difference between scenarios involving pain (IHS and AHS, which share a common painful outcome) and the scenarios not involving pain (NAS), peaking at ~170 and ~250 ms over posterior and central clusters. In contrast, the difference between the IHS and AHS (which distinguish from each other in terms of intentionality) emerged later, with clear differences at a latency of ~400 ms evident across both clusters. These electrophysiological results provide support to two-component models of moral decision-making (Greene et al., 2001, 2004, 2008; Haidt, 2007), suggesting early automatic evaluation and emotional response triggered by the observation of painful outcomes, leading to a clear distinction between the painful and non-painful scenes around the ~170 and ~250 ms peaks, and, subsequently, a later cognitive evaluation of the agents' mental states (i.e., intentions), resulting in a pronounced difference around 400 ms from stimulus onset.
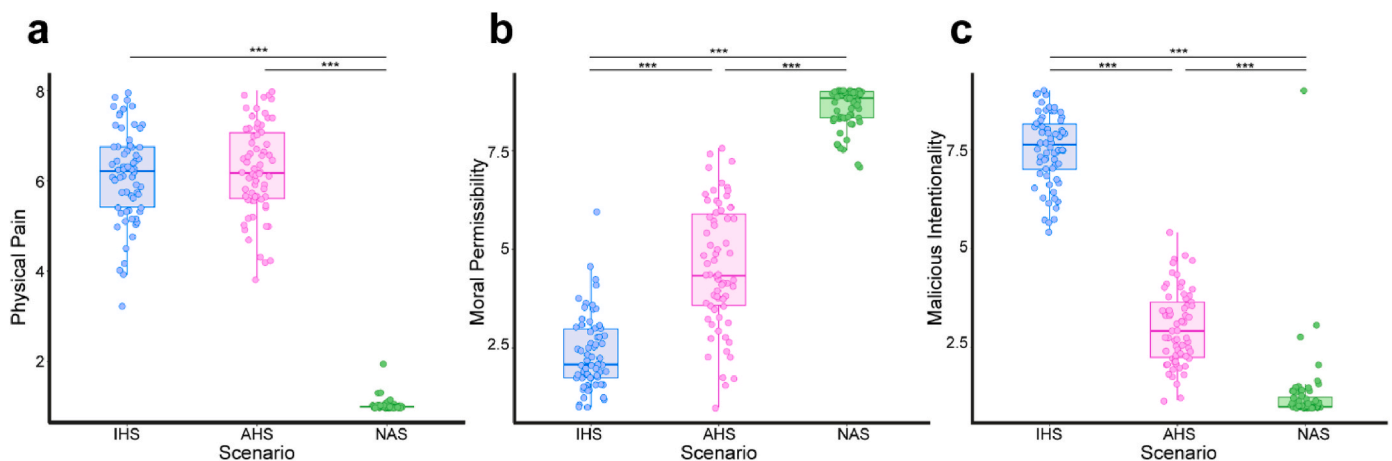

Fig. 5. Box plots depicting subjective evaluations (a.) Physical Pain ratings. **(b.)** Moral Permissibility ratings. **(c.)** Malicious Intentionality ratings. For each graph, the x-axis represents the three different scenarios (IHS, AHS, NAS) while the y-axis depicts the specific rating score. Significance levels are shown by asterisks (***$p < 0.001$).
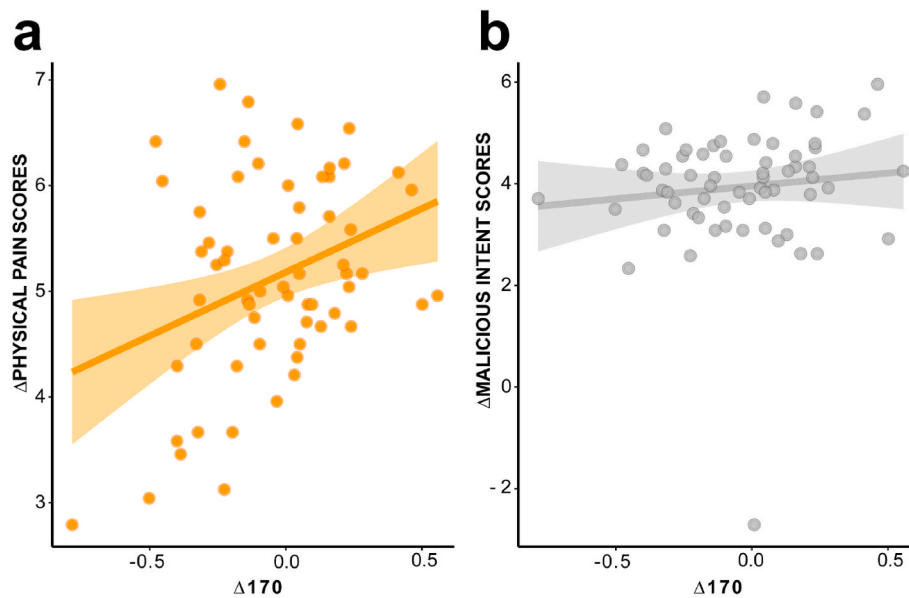
Fig. 6. Regression analysis between neurophysiological and behavioral data. The graph illustrates the relationship between early neural responses to the victims' pain (Δ170) and subjective evaluations of physical pain (a) and malicious intentionality (b). All indices were calculated as the differences between the mean responses to painful scenes (IHS and AHS) and non-painful scenes (NAS).

4.1. From experience sharing to mentalizing processes

Correlation analyses showed that the early ERP modulation at ~170 ms positively predicted physical pain ratings, consistent with prior work on pain empathy (e.g., Fan and Han, 2008). Moreover, source reconstruction of this early pain-related response highlighted the involvement of the right insula and left SMG, suggesting an activation of affective and sensorimotor components of experience-sharing during the vicarious perception of others' pain (Avenanti et al., 2006; Costantini et al., 2008; Benuzzi et al., 2018; Gallo et al., 2018; Zhou and Han, 2021; Vitale et al., 2023; for reviews and meta-analyses, see Lamm et al., 2011; Jauniaux et al., 2019; Fallon et al., 2020). These findings are consistent with studies showing that the insula and sensorimotor regions predict both the intensity of one's own pain (Kong et al., 2006; Moulton et al., 2012) and the pain attributed to others (Bufalari et al., 2007; Avenanti et al., 2009; Gu et al., 2012).

Interestingly, our source reconstruction also revealed a nuanced pattern of dynamic transition between experience-sharing and mentalizing processes. The second peak (~250 ms) showed a broader recruitment of affective and sensorimotor empathy networks, including modulation of the bilateral SMG, left aMCC, and the right insula (Singer et al., 2004; Jackson et al., 2006; Ionta et al., 2020; for a meta-analysis see Lamm et al., 2011; for a review see Fallon et al., 2020). Additionally, we observed bilateral recruitment of the TPJ and midline regions, suggesting that mentalizing networks are engaged in parallel with vicarious pain processing (Akitsuki and Decety, 2009; Krishnan et al., 2016), at a stage in which outcomes and intentions are likely integrated (Young et al., 2007; Young and Saxe, 2009a, 2009b). The concurrent activation of SMG and TPJ is particularly significant, as recent discussions propose that the SMG, a region implicated in vicarious pain (e.g., Ionta et al., 2020; Fallon et al., 2020) and anatomically proximate to the TPJ, may play a key role in integrating and balancing self and other perspectives within experience-sharing mechanisms (for a review see Gamble et al., 2024). The broad engagement of distinct brain networks at ~250 ms is consistent with research on cross-network interactions in naturalistic social cognition tasks, which require the simultaneous processing of multiple information streams (for a meta-analysis see Maliske et al., 2023).

During the third peak (~400 ms), which was associated with differences between IHS and AHS, we observed selectively involvement of

the left TPJ, a brain region critically implicated in the understanding of social intentions (Samson et al., 2004; Ciaramidaro et al., 2007), along with other midline regions involved in mental state attribution. This pattern possibly reflects the discrimination between malicious and innocuous intentions (Young et al., 2007; Young and Saxe, 2009b). These results are consistent with the notion that later engagement of the TPJ might be related to internal updates based on the contextual sensory information (Geng and Vossel, 2013), thereby aiding the integration of contextual understanding required for comprehending another person's perspective (Carter et al., 2013; Doricchi et al., 2022; Masina et al., 2022).

Taken together, these findings expand our understanding of implicit moral cognition by illustrating a complex interplay between neural mechanisms responsible for experience-sharing and mentalizing. The brain's response is initially dominated by affective and sensorimotor resonance with the victim's pain through experience-sharing mechanisms, gradually integrating mentalizing aspects, and moving towards a comprehensive evaluation that encompasses both an emotional alignment with the victim and a cognitive appraisal of the agent's intention. Importantly, our results, although seemingly consistent with the two-component model of moral evaluation (Greene et al., 2001, 2004, 2008; Haidt, 2007), provide significant new insights into a partial temporal overlap between these processes. Notably, our findings suggest that mentalizing processes, while strongly engaged during discrimination between IHS and AHS occurring later in time (~400 ms), begin earlier (~250 ms) and seem to contribute, alongside with the experience-sharing processes, in distinguishing painful and non-painful interactions (Krishnan et al., 2016). The prompt recruitment of brain regions related to the processing of others' intentions, with the TPJ potentially serving as the core area, aligns with a model of a feedback loop between experience-sharing and mentalizing systems (Gamble et al., 2024) and could be explained by the primeval necessity to predict agents' beliefs by analyzing those social cues depicted in the scenarios that can reveal the underlying others' mental states (e.g., negative state in the two painful scenarios).

The extensive simultaneous modulation of brain networks critical for mentalizing and experience-sharing likely reflects the complexity of social scenarios involving two interacting agents (Akitsuki and Decety, 2009; for a review, see Gamble et al., 2024). A growing body of research indicates that processing rich social scenarios engages multiple

cognitive and affective processes simultaneously (for a meta-analysis, see Maliske et al., 2023). Studies using naturalistic emotional scenarios have similarly underscored the involvement of experience-sharing and mentalizing networks (Zaki et al., 2007, 2009; Decety et al., 2008; Cheetham et al., 2009), suggesting these neural systems co-activate and reciprocally influence each other during social perception (Paracampo et al., 2017, 2018).

4.2. Comparison with prior ERP research

Prior research has used EEG to investigate the temporal brain dynamics during the observation of painful and non-painful stimuli, suggesting that experience-sharing processes are associated with early ERP modulations over fronto-central electrodes (N1, P2/N2), which are further influenced by the emotional valence and arousal conveyed by the stimuli (Fan and Han, 2008; Olofsson et al., 2008; Yoder and Decety, 2014; Gui et al., 2016; Tao et al., 2022), while later modulations (P3/LPP) are associated with executive control processes that regulate the affective experiences (Fan and Han, 2008; Han et al., 2008; Decety et al., 2010; Meng et al., 2013; Fan et al., 2014; Galang et al., 2020; Chen et al., 2023; for a meta-analysis see Coll et al., 2018). Similar components have also been associated with moral evaluation, though findings remain heterogeneous across studies (Yoder and Decety, 2014; Gan et al., 2016; Cui et al., 2016; Tao et al., 2022; Jiang et al., 2022). While these studies reported distinct ERP components, either positive or negative, over specific electrode clusters and time windows, our study revealed a pattern of both positive deflections over the posterior sensors and negative deflections over the central sensors, persisting throughout the entire investigation window.

Several factors may have contributed to these effects, which were not readily apparent with the heterogeneous hypothesis-driven approaches typically used in prior ERP studies on moral cognition. These factors include a larger sample of participants, a data-driven approach, and the use of different scenarios. Regarding this last aspect, previous work on empathy for pain often investigated brain responses to decontextualized scenarios depicting isolated body parts (e.g., hands) in painful and non-painful conditions (Fan and Han, 2008; Han et al., 2008; Meng et al., 2013 Fan et al., 2014; Galang et al., 2020; Chen et al., 2023) without considering the influence of the agent's intention in moral evaluation. Importantly, our study builds on and extends these findings by using stimuli depicting painful and non-painful social interactions within different contexts (Decety and Cacioppo, 2012; Baez et al., 2012, 2013, 2014; Spaccasassi et al., 2023), allowing us to investigate brain responses related to observing pain and inferring the agent's responsibility and intentions. Our findings reveal an early recruitment of experience-sharing mechanisms that distinguish painful (IHS and AHS) from non-painful scenarios (NAS), followed by a later engagement of mentalizing processes that discriminate between harmful actions performed with (IHS) or without (AHS) a malicious intention. These results appear consistent with the notion that the perception of other's pain modulates both early ERP components reflecting automatic emotional responses, and later components associated with cognitive evaluation (e. g., Fan and Han, 2008; Decety et al., 2010). We show that this progression is bridged by a transitional phase, where emotional resonance and cognitive evaluation are integrated during perception of harmful scenarios.

Our subjective evaluation results using the modified EPT task (Decety et al., 2011; Baez et al., 2012, 2013, 2014; Couto et al., 2013) are in line with a previous study (Spaccasassi et al., 2023). Both IHS and AHS were perceived to evoke similar levels of pain intensity, unlike NAS, which elicited minimal perceived pain. This suggests that any difference observed between IHS and AHS in the present study should not be attributed to the pain attributed to the victims. Furthermore, consistent with prior research, our behavioral findings suggest that participants integrated both intention and outcome information when explicitly evaluating moral scenarios (Young et al., 2007; Young and Saxe, 2009b;

Gan et al., 2016; Spaccasassi et al., 2023). Indeed, although the intentions were neutral in both AHS and NAS, the former was perceived to have a greater level of malicious intent compared to the latter, possibly suggesting that the negative outcomes retrospectively influenced the perception of the agents' intentions. In line with this, participants consistently rated IHS as less morally permissible than AHS in moral permissibility ratings, despite the presence of a negative outcome in both scenarios.

4.3. Limitations and future directions

A key limitation of this study is the lower spatial resolution of EEG source estimation compared to other neuroimaging techniques. While EEG captures temporal dynamics, its ability to localize brain activity is limited. Thus, source reconstructions should be interpreted cautiously, and future studies using higher-resolution techniques are recommended.

Our study did not show involvement of the prefrontal cortex, a region crucial for moral reasoning (for a review, see Forbes and Grafman, 2010) and mentalizing (Cristiano et al., 2023; for a review, see Raichle, 2015). Neither ROI nor whole-brain analyses showed activation in the vmPFC or dorsal sectors of the prefrontal cortex, which are also implicated in moral cognition and functionally linked to TPJ (Young et al., 2007; Krueger and Hoffman, 2016). This null finding may be due to methodological factors, as our source estimation focused on three relatively early time points (~170 ms, ~250 ms, and ~400 ms), potentially preceding full prefrontal cortex involvement. Moreover, we employed an implicit moral judgment task, while significant prefrontal cortex involvement is often linked to explicit moral judgment (Moll et al., 2005). Given the limitations of EEG source estimation and the challenges in interpreting null findings, we refrain from drawing strong conclusions. Future research could further directly address this issue by using an explicit moral judgment task with EEG, as partially different neural mechanisms are likely involved in implicit vs explicit tasks (Tao et al., 2022; Borgomaneri et al., 2023, 2024).

5. Conclusion

The current study aimed to explore the temporal dynamics of moral cognition by recording EEG signal while participants viewed moral scenarios. Our findings revealed that painful scenarios, where an agent causes physical suffering to a victim either intentionally (IHS) or accidently (AHS), result in early ERP modulations. These early modulations are linked to the prompt engagement of sensorimotor and affective experience-sharing mechanisms, which are then integrated with mentalizing processes. In contrast, the difference between IHS and AHS, distinguished solely by the agent's intention, emerges as a later modulation in ERP responses. This later modulation reflects the activation of the mentalizing network, highlighting the role of intention in moral evaluation. Overall, our findings support a dual model of moral cognition, which posits that both intention and outcome processing shape moral judgments. Our study enriches this model by providing insights into the temporal dynamics and spatial localization of these processes, suggesting an early integration of experience-sharing and mentalizing mechanisms. This blending of neural processes underscores the complexity of moral cognition and highlights the nuanced interplay between different cognitive components in moral judgment.

CRediT authorship contribution statement

Kamela Cenka: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation. **Chiara Spaccasassi:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation. **Stella Petkovic:** Writing – review & editing, Investigation. **Rachele Pezzetta:** Writing – review & editing, Supervision, Funding acquisition. **Giorgio Arcara:** Writing – review & editing,

Supervision, Resources, Funding acquisition. **Alessio Avenanti:** Writing – review & editing, Writing – original draft, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Open practices section

The study in this article earned Open Data and Open Material badges for transparent practices. The data and material that support the findings of this study are openly available at: https://osf.io/hnx9d/. No analysis code was used.

Declaration of competing interest

The authors have nothing to declare.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuropsychologia.2024.109033.

Data availability

We have shared data in a public repository, this is reported in the manuscript

References

Adolphs, R., 2009. The social brain: neural basis of social knowledge. Annu. Rev. Psychol. 60, 693–716.

Akitsuki, Y., Decety, J., 2009. Social context and perceived agency affects empathy for pain: an event-related fMRI investigation. Neuroimage 47 (2), 722–734.

Arutiunian, V., Arcara, G., Buyanova, I., Gomozova, M., Dragoy, O., 2022. The age-related changes in 40 Hz Auditory Steady-State Response and sustained Event-Related Fields to the same amplitude-modulated tones in typically developing children: a magnetoencephalography study. Hum. Brain Mapp. 43 (17), 5370–5383.

Avenanti, A., Bueti, D., Galati, G., Aglioti, S.M., 2005. Transcranial magnetic stimulation highlights the sensorimotor side of empathy for pain. Nat. Neurosci. 8 (7), 955–960.

Avenanti, A., Minio-Paluello, I., Bufalari, I., Aglioti, S.M., 2006. Stimulus-driven modulation of motor-evoked potentials during observation of others' pain. Neuroimage 32 (1), 316–324.

Avenanti, A., Minio-Paluello, I., Bufalari, I., Aglioti, S.M., 2009. The pain of a model in the personality of an onlooker: influence of state-reactivity and personality traits on embodied empathy for pain. Neuroimage 44 (1), 275–283.

Avenanti, A., Paracampo, R., Annella, L., Tidoni, E., Aglioti, S.M., 2018. Boosting and decreasing action prediction abilities through excitatory and inhibitory tdcs of inferior frontal cortex. Cerebr. Cortex 28 (4), 1282–1296.

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 (12), 3168–3181.

Baez, S., Rattazzi, A., Gonzalez-Gadea, M.L., Torralva, T., Vigliecca, N.S., Decety, J., Manes, F., Ibanez, A., 2012. Integrating intention and context: assessing social cognition in adults with Asperger syndrome. Front. Hum. Neurosci. 6, 302.

Baez, S., Herrera, E., Villarin, L., Theil, D., Gonzalez-Gadea, M.L., Gomez, P., Mosquera, M., Huepe, D., Strejilevich, S., Vigliecca, N.S., Matthäus, F., Decety, J., Manes, F., Ibañez, A.M., 2013. Contextual social cognition impairments in schizophrenia and bipolar disorder. PLoS One 8 (3), e57664.

Baez, S., Manes, F., Huepe, D., Torralva, T., Fiorentino, N., Richter, F., Huepe-Artigas, D., Ferrari, J., Montañes, P., Reyes, P., Matallana, D., Vigliecca, N.S., Decety, J., Ibanez, A., 2014. Primary empathy deficits in frontotemporal dementia. Front. Aging Neurosci. 6, 262.

Bagnis, A., Celeghin, A., Diano, M., Mendez, C.A., Spadaro, G., Mosso, C.O., Avenanti, A., Tamietto, M., 2020. Functional neuroanatomy of racial categorization from visual perception: a meta-analytic study. Neuroimage 217, 116939.

Baillet, S., Mosher, J.C., Leahy, R.M., 2001. Electromagnetic brain mapping. IEEE Signal Proc. Mag. 18 (6), 14–30.

Benuzzi, F., Lui, F., Ardizzi, M., Ambrosecchia, M., Ballotta, D., Righi, S., Pagnoni, G., Gallese, V., Porro, C.A., 2018. Pain mirrors: neural correlates of observing self or others' facial expressions of pain. Front. Psychol. 9, 1825.

Borgomaneri, S., Zanon, M., Di Luzio, P., Cataneo, A., Arcara, G., Romei, V., Tamietto, M., Avenanti, A., 2023. Increasing associative plasticity in temporo-occipital back-projections improves visual perception of emotions. Nat. Commun. 14 (1), 5720.

Borgomaneri, S., Vitale, F., Battaglia, S., de Vega, M., Avenanti, A., 2024. Task-related modulation of motor response to emotional bodies: a TMS motor-evoked potential study. Cortex 171, 235–246.

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 (1), 312–319.

Brainard, D.H., Vision, S., 1997. The psychophysics toolbox. Spatial Vis. 10, 433–436.

Bufalari, I., Aprile, T., Avenanti, A., Di Russo, F., Aglioti, S.M., 2007. Empathy for pain and touch in the human somatosensory cortex. Cerebr. Cortex 17 (11), 2553–2561.

Buon, M., Seara-Cardoso, A., Viding, E., 2016. Why (and how) should we study the interplay between emotional arousal, Theory of Mind, and inhibitory control to understand moral cognition? Psychon. Bull. Rev. 23 (6), 1660–1680.

Bzdok, D., Schilbach, L., Vogeley, K., Schneider, K., Laird, A.R., Langner, R., Eickhoff, S. B., 2012. Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. Brain Struct. Funct. 217 (4), 783–796.

Carter, R.M., Huettel, S.A., 2013. A nexus model of the temporal-parietal junction. Trends Cognit. Sci. 17 (7), 328–336.

Cheetham, M., Pedroni, A.F., Antley, A., Slater, M., Jäncke, L., 2009. Virtual milgram: empathic concern or personal distress? Evidence from functional MRI and dispositional measures. Front. Hum. Neurosci. 3, 29.

Chen, T., Li, Q., Peng, M., Li, X., 2023. Moral transgression modulates empathy for pain: evidence from ERP and EEG data. Biol. Psychol. 176, 108467.

Ciaramelli, E., Braghittoni, D., di Pellegrino, G., 2012. It is the outcome that counts! Damage to the ventromedial prefrontal cortex disrupts the integration of outcome and belief information for moral judgment. J. Int. Neuropsychol. Soc. 18 (6), 962–971.

Ciaramidaro, A., Adenzato, M., Enrici, I., Erk, S., Pia, L., Bara, B.G., Walter, H., 2007. The intentional network: how the brain reads varieties of intentions. Neuropsychologia 45 (13), 3105–3113.

Coll, M.P., 2018. Meta-analysis of ERP investigations of pain empathy underlines methodological issues in ERP research. Soc. Cognit. Affect. Neurosci. 13 (10), 1003–1017.

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 (49), 17996–18006.

Costantini, M., Galati, G., Romani, G.L., Aglioti, S.M., 2008. Empathic neural reactivity to noxious stimuli delivered to body parts and non-corporeal objects. Eur. J. Neurosci. 28 (6), 1222–1230.

Couto, B., Sedeño, L., Sposato, L.A., Sigman, M., Riccio, P.M., Salles, A., Lopez, V., Schroeder, J., Manes, F., Ibanez, A., 2013. Insular networks for emotional processing and social cognition: comparison of two case reports with either cortical or subcortical involvement. Cortex 49 (5), 1420–1434.

Cristiano, A., Finisguerra, A., Urgesi, C., Avenanti, A., Tidoni, E., 2023. Functional role of the theory of mind network in integrating mentalistic prior information with action kinematics during action observation. Cortex 166, 107–120.

Cui, F., Ma, N., Luo, Y.J., 2016. Moral judgment modulates neural responses to the perception of other's pain: an ERP study. Sci. Rep. 6, 20851.

Cushman, F., 2008. Crime and punishment: distinguishing the roles of causal and intentional analyses in moral judgment. Cognition 108 (2), 353–380.

de Vignemont, F., Singer, T., 2006. The empathic brain: how, when and why? Trends Cogn. Sci. 10 (10), 435–441.

Decety, J., Cacioppo, S., 2012. The speed of morality: a high-density electrical neuroimaging study. J. Neurophysiol. 108 (11), 3068–3072.

Decety, J., Howard, L.H., 2013. The role of affect in the neurodevelopment of morality. Child Dev. Perspect. 7 (1), 49–54.

Decety, J., Michalska, K.J., Akitsuki, Y., 2008. Who caused the pain? An fMRI investigation of empathy and intentionality in children. Neuropsychologia 46 (11), 2607–2614.

Decety, J., Yang, C.Y., Cheng, Y., 2010. Physicians down-regulate their pain empathy response: an event-related brain potential study. Neuroimage 50 (4), 1676–1682.

Doricchi, F., Lasaponara, S., Pazzaglia, M., Silvetti, M., 2022. Left and right temporal-parietal junctions (TPJs) as "match/mismatch" hedonic machines: a unifying account of TPJ function. Phys. Life Rev. 42, 56–92.

Escobar, M.J., Huepe, D., Decety, J., Sedeño, L., Messow, M.K., Baez, S., Rivera-Rei, Á., Canales-Johnson, A., Morales, J.P., Gómez, D.M., Schröeder, J., Manes, F., López, V., Ibáñez, A., 2014. Brain signatures of moral sensitivity in adolescents with early social deprivation. Sci. Rep. 4, 5354.

Fallon, N., Roberts, C., Stancak, A., 2020. Shared and distinct functional networks for empathy and pain processing: a systematic review and meta-analysis of fMRI studies. Soc. Cognit. Affect. Neurosci. 15 (7), 709–723.

Fan, Y., Han, S., 2008. Temporal dynamic of neural mechanisms involved in empathy for pain: an event-related brain potential study. Neuropsychologia 46 (1), 160–173.

Fan, Y.T., Chen, C., Chen, S.C., Decety, J., Cheng, Y., 2014. Empathic arousal and social understanding in individuals with autism: evidence from fMRI and ERP measurements. Soc. Cognit. Affect. Neurosci. 9 (8), 1203–1213.

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A. R., Fox, P.T., Eickhoff, S.B., Yu, C., Jiang, T., 2016. The human Brainnetome Atlas: a

new brain atlas based on connectional architecture. Cerebr. Cortex 26 (8), 3508–3526.

Fernandes, C., Pasion, R., Gonçalves, A.R., Almeida, R., Garcez, H., Ferreira-Santos, F., Barbosa, F., Marques-Teixeira, J., 2022. Awareness to utilitarian responses in later life: an ERP study with moral dilemmas. Neurosci. Lett. 787, 136824.

Fittipaldi, S., Armony, J.L., García, A.M., Migeot, J., Cadaveira, M., Ibáñez, A., Baez, S., 2023. Emotional descriptions increase accidental harm punishment and its cortico-limbic signatures during moral judgment in autism. Sci. Rep. 13 (1), 1745.

Forbes, C.E., Grafman, J., 2010. The role of the human prefrontal cortex in social cognition and moral judgment. Annu. Rev. Neurosci. 33, 299–324.

Frith, C., Frith, U., 2005. Theory of mind. Curr. Biol. 15 (17), R644–R646.

Galang, C.M., Jenkins, M., Obhi, S.S., 2020. Exploring the effects of visual perspective on the ERP components of empathy for pain. Soc. Neurosci. 15 (2), 186–198.

Gallo, S., Paracampo, R., Müller-Pinzler, L., Severo, M.C., Blömer, L., Fernandes-Henriques, C., Henschel, A., Lammes, B.K., Maskaljunas, T., Suttrup, J., Avenanti, A., Keysers, C., Gazzola, V., 2018. The causal role of the somatosensory cortex in prosocial behaviour. Elife 7, e32740.

Gamble, R.S., Henry, J.D., Decety, J., Vanman, E.J., 2024. The role of external factors in affect-sharing and their neural bases. Neurosci. Biobehav. Rev. 157, 105540.

Gan, T., Lu, X., Li, W., Gui, D., Tang, H., Mai, X., et al., 2016. Temporal dynamics of the integration of intention and outcome in harmful and helpful moral judgment. Front. Psychol. 6, 2022.

Geng, J.J., Vossel, S., 2013. Re-evaluating the role of TPJ in attentional control: contextual updating? Neurosci. Biobehav. Rev. 37 (10 Pt 2), 2608–2620.

Gramfort, A., Papadopoulo, T., Olivi, E., Clerc, M., 2010. OpenMEEG: opensource software for quasistatic bioelectromagnetics. Biomed. Eng. Online 9, 45.

Greene, J.D., 2009. Dual-process morality and the personal/impersonal distinction: a reply to McGuire, Langdon, Coltheart, and Mackenzie. J. Exp. Soc. Psychol. 45 (3), 581–584.

Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M., Cohen, J.D., 2001. An fMRI investigation of emotional engagement in moral judgment. Science 293 (5537), 2105–2108.

Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M., Cohen, J.D., 2004. The neural bases of cognitive conflict and control in moral judgment. Neuron 44 (2), 389–400.

Greene, J.D., Morelli, S.A., Lowenberg, K., Nystrom, L.E., Cohen, J.D., 2008. Cognitive load selectively interferes with utilitarian moral judgment. Cognition 107 (3), 1144–1154.

Gu, X., Gao, Z., Wang, X., Liu, X., Knight, R.T., Hof, P.R., Fan, J., 2012. Anterior insular cortex is necessary for empathetic pain perception. Brain 135 (9), 2726–2735.

Gui, D.Y., Gan, T., Liu, C., 2016. Neural evidence for moral intuition and the temporal dynamics of interactions between emotional processes and moral cognition. Soc. Neurosci. 11 (4), 380–394.

Haidt, J., 2007. The new synthesis in moral psychology. Science 316 (5827), 998–1002.

Han, S., Fan, Y., Mao, L., 2008. Gender difference in empathy for pain: an electrophysiological investigation. Brain Res. 1196, 85–93.

Herring, J.D., Thut, G., Jensen, O., Bergmann, T.O., 2015. Attention modulates TMS-locked alpha oscillations in the visual cortex. J. Neurosci. 35 (43), 14435–14447.

Hesse, E., Mikulan, E., Decety, J., Sigman, M., Garcia, M.delC., Silva, W., Ciraolo, C., Vaucheret, E., Baglivo, F., Huepe, D., Lopez, V., Manes, F., Bekinschtein, T.A., Ibanez, A., 2016. Early detection of intentional harm in the human amygdala. Brain 139 (Pt 1), 54–61.

Ionta, S., Costantini, M., Ferretti, A., Galati, G., Romani, G.L., Aglioti, S.M., 2020. Visual similarity and psychological closeness are neurally dissociable in the brain response to vicarious pain. Cortex 133, 295–308.

Jackson, P.L., Meltzoff, A.N., Decety, J., 2005. How do we perceive the pain of others? A window into the neural processes involved in empathy. Neuroimage 24 (3), 771–779.

Jackson, P.L., Brunet, E., Meltzoff, A.N., Decety, J., 2006. Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain. Neuropsychologia 44, 752–761.

Jauniaux, J., Khatibi, A., Rainville, P., Jackson, P.L., 2019. A meta-analysis of neuroimaging studies on pain empathy: investigating the role of visual information and observers' perspective. Soc. Cognit. Affect. Neurosci. 14 (8), 789–813.

Jiang, Q., Zhuo, L., Wang, Q., Lin, W., 2022. The neural basis of moral judgement for self and for others: Evidence from event-related potentials. Front. Hum. Neurosci. 16, 919499.

Jung, T.P., Makeig, S., Humphries, C., Lee, T.W., Mckeown, M.J., Iragui, V., Sejnowski, T. J., 2000. Removing electroencephalographic artifacts by blind source separation. Psychophysiol. 37 (2), 163–178.

Kahneman, D., Frederick, S., 2002. Representativeness revisited: attribute substitution in intuitive judgment. In: Gilovich, T., Griffin, D., Kahneman, D. (Eds.), Heuristics and Biases: the Psychology of Intuitive Judgment. Cambridge University Press, pp. 49–81.

Keysers, C., Kaas, J.H., Gazzola, V., 2010. Somatosensation in social perception. Nat. Rev. Neurosci. 11 (6), 417–428.

Knobe, J., 2005. Theory of mind and moral cognition: exploring the connections. Trends Cogn. Sci. 9 (8), 357–359.

Kong, J., White, N.S., Kwong, K.K., Vangel, M.G., Rosman, I.S., Gracely, R.H., Gollub, R. L., 2006. Using fMRI to dissociate sensory encoding from cognitive evaluation of heat pain intensity. Hum. Brain Mapp. 27 (9), 715–721.

Korhonen, R.J., Hernandez-Pavon, J.C., Metsomaa, J., Mäki, H., Ilmoniemi, R.J., Sarvas, J., 2011. Removal of large muscle artifacts from transcranial magnetic stimulation-evoked EEG by independent component analysis. Med. Biol. Eng. Comput. 49 (4), 397–407.

Krishnan, A., Woo, C.W., Chang, L.J., Ruzic, L., Gu, X., López-Solà, M., Jackson, P.L., Pujol, J., Fan, J., Wager, T.D., 2016. Somatic and vicarious pain are represented by dissociable multivariate brain patterns. Elife 5, e15166.

Krueger, F., Hoffman, M., 2016. The emerging neuroscience of third-party punishment. Trends Neurosci. 39 (8), 499–501.

Lamm, C., Singer, T., 2010. The role of anterior insular cortex in social emotions. Brain Struct. Funct. 214 (5–6), 579–591.

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

Loewenstein, G., Rick, S., Cohen, J.D., 2008. Neuroeconomics. Annu. Rev. Psychol. 59, 647–672.

Maliske, L.Z., Schurz, M., Kanske, P., 2023. Interactions within the social brain: Co-activation and connectivity among networks enabling empathy and Theory of Mind. Neurosci. Biobehav. Rev. 147, 105080.

Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164 (1), 177–190.

Masina, F., Pezzetta, R., Lago, S., Mantini, D., Scarpazza, C., Arcara, G., 2022. Disconnection from prediction: a systematic review on the role of right temporoparietal junction in aberrant predictive processing. Neurosci. Biobehav. Rev. 138, 104713.

Meng, J., Jackson, T., Chen, H., Hu, L., Yang, Z., Su, Y., Huang, X., 2013. Pain perception in the self and observation of others: an ERP investigation. Neuroimage 72, 164–173.

Mitchell, J.P., 2009. Inferences about mental states. Philos. Trans. Royal Soc. Lond. B Biol. Sci. 364 (1521), 1309–1316.

Moll, J., de Oliveira-Souza, R., Bramati, I.E., Grafman, J., 2002. Functional networks in emotional moral and nonmoral social judgments. Neuroimage 16 (3 Pt 1), 696–703.

Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., Grafman, J., 2005. The neural basis of human moral cognition. Nat. Rev. Neurosci. 6 (10), 799–809.

Morrison, I., Lloyd, D., di Pellegrino, G., Roberts, N., 2004. Vicarious responses to pain in anterior cingulate cortex: is empathy a multisensory issue? Cognit. Affect. Behav. Neurosci. 4 (2), 270–278.

Moulton, E.A., Pendse, G., Becerra, L.R., Borsook, D., 2012. BOLD responses in somatosensory cortices better reflect heat sensation than pain. J. Neurosci. 32 (17), 6024–6031.

Olofsson, J.K., Nordin, S., Sequeira, H., Polich, J., 2008. Affective picture processing: an integrative review of ERP findings. Biol. Psychol. 77 (3), 247–265.

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intel. Neurosci. 2011, 156869.

Paracampo, R., Tidoni, E., Borgomaneri, S., di Pellegrino, G., Avenanti, A., 2017. Sensorimotor network crucial for inferring amusement from smiles. Cerebr. Cortex 27 (11), 5116–5129.

Paracampo, R., Pirruccio, M., Costa, M., Borgomaneri, S., Avenanti, A., 2018. Visual, sensorimotor and cognitive routes to understanding others' enjoyment: an individual differences rTMS approach to empathic accuracy. Neuropsychologia 116 (Pt A), 86–98.

Pascual-Marqui, R.D., Michel, C.M., Lehmann, D., 1994. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. Int. J. Psychophysiol. 18 (1), 49–65.

Pasion, R., Fernandes, C., Gonçalves, A.R., Ferreira-Santos, F., Páscoa, R., Barbosa, F., Marques-Teixeira, J., 2019. The effect of aging on the (mis) perception of intentionality-an ERP study. Soc. Neurosci. 14 (2), 149–161.

Patri, J.F., Cavallo, A., Pullar, K., Soriano, M., Valente, M., Koul, A., Avenanti, A., Panzeri, S., Becchio, C., 2020. Transient disruption of the inferior parietal lobule impairs the ability to attribute intention to action. Curr. Biol. 30 (23), 4594–4605.

Raichle, M.E., 2015. The brain's default mode network. Annu. Rev. Neurosci. 38, 433–447.

Reniers, R.L.E.P., Corcoran, R., Völlm, B.A., Mashru, A., Howard, R., Liddle, P.F., 2012. Moral decision-making, ToM, empathy and the default mode network. Biol. Psychol. 90 (3), 202–210.

Riečanský, I., Lamm, C., 2019. The role of sensorimotor processes in pain empathy. Brain Topogr. 32 (6), 965–976.

Rizzolatti, G., Sinigaglia, C., 2010. The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. Nat. Rev. Neurosci. 11 (4), 264–274.

Şahin, M.D., Aybek, E.C., 2019. Jamovi: an easy to use statistical software for the social scientists. Int. J. Asses. Tools E. 6, 670–692.

Samson, D., Apperly, I.A., Chiavarino, C., Humphreys, G.W., 2004. Left temporoparietal junction is necessary for representing someone else's belief. Nat. Neurosci. 7 (5), 499–500.

Sheng, F., Han, S., 2012. Manipulations of cognitive strategies and intergroup relationships reduce the racial bias in empathic neural responses. Neuroimage 61 (4), 786–797.

Sheng, F., Han, X., Han, S., 2016. Dissociated neural representations of pain expressions of different races. Cerebr. Cortex 26 (3), 1221–1233.

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 (5661), 1157–1162.

Spaccasassi, C., Zanon, M., Borgomaneri, S., Avenanti, A., 2022. Mu rhythm and corticospinal excitability capture two different frames of motor resonance: a TMS/ EEG co-registration study. Cortex 154, 197–211.

Spaccasassi, C., Cenka, K., Petkovic, S., Avenanti, A., 2023. Sense of agency predicts severity of moral judgments. Front. Psychol. 13, 1070742.

Stropahl, M., Bauer, A.R., Debener, S., Bleichner, M.G., 2018. Source-modeling auditory processes of EEG data using EEGLAB and Brainstorm. Front. Neurosci. 12, 309.

Tadel, F., Baillet, S., Mosher, J.C., Pantazis, D., Leahy, R.M., 2011. Brainstorm: a user-friendly application for MEG/EEG analysis. Comput. Intel. Neurosci. 2011, 879716.

Tao, D., Leng, Y., Peng, S., Xu, J., Ge, S., Deng, H., 2022. Temporal dynamics of explicit and implicit moral evaluations. Int. J. Psychophysiol. 172, 1–9.

Team, R.C., 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2012.

Tidoni, E., Borgomaneri, S., di Pellegrino, G., Avenanti, A., 2013. Action simulation plays a critical role in deceptive action recognition. J. Neurosci. 33 (2), 611–623.

Valchev, N., Tidoni, E., Hamilton, A.F.C., Gazzola, V., Avenanti, A., 2017. Primary somatosensory cortex necessary for the perception of weight from other people's action: a continuous theta-burst TMS experiment. Neuroimage 152, 195–206.

Valero-Cabré, A., Amengual, J.L., Stengel, C., Pascual-Leone, A., Coubard, O.A., 2017. Transcranial magnetic stimulation in basic and clinical neuroscience: a comprehensive review of fundamental principles and novel insights. Neurosci. Biobehav. Rev. 83, 381–404.

Vitale, F., Urrutia, M., Avenanti, A., de Vega, M., 2023. You are fired! Exclusion words induce corticospinal modulations associated with vicarious pain. Soc. Cognit. Affect. Neurosci. 18 (1), nsad033.

Wagner, N.F., Chaves, P., Wolff, A., 2017. Discovering the neural nature of moral cognition? Empirical, theoretical, and practical challenges in bioethical research with electroencephalography (EEG). J. Bioeth. Inq. 14, 299–313.

Welke, D., Vessel, E.A., 2022. Naturalistic viewing conditions can increase task engagement and aesthetic preference but have only minimal impact on EEG quality. Neuroimage 256, 119218.

World Medical Association, 2013. World Medical Association declaration of Helsinki: Ethical principles for medical research involving human subjects. JAMA 310, 2191–2194.

Yoder, K.J., Decety, J., 2014. Spatiotemporal neural dynamics of moral judgment: a high-density ERP study. Neuropsychologia 60, 39–45.

Young, L., Saxe, R., 2009a. An FMRI investigation of spontaneous mental state inference for moral judgment. J. Cogn. Neurosci. 21 (7), 1396–1405.

Young, L., Saxe, R., 2009b. Innocent intentions: a correlation between forgiveness for accidental harm and neural activity. Neuropsychologia 47, 2065–2072.

Young, L., Tsoi, L., 2013. When mental states matter, when they don't, and what that means for morality. Soc. Personal. Psychol. Compass 7 (8), 585–604.

Young, L., Cushman, F., Hauser, M., Saxe, R., 2007. The neural basis of the interaction between theory of mind and moral judgment. Proc. Natl. Acad. Sci. U.S.A. 104 (20), 8235–8240.

Zaki, J., Ochsner, K.N., 2012. The neuroscience of empathy: progress, pitfalls and promise. Nat. Neurosci. 15 (5), 675–680.

Zaki, J., Ochsner, K.N., Hanelin, J., Wager, T.D., Mackey, S.C., 2007. Different circuits for different pain: patterns of functional connectivity reveal distinct networks for processing pain in self and others. Soc. Neurosci. 2 (3–4), 276–291.

Zaki, J., Weber, J., Bolger, N., Ochsner, K., 2009. The neural bases of empathic accuracy. Proc. Natl. Acad. Sci. U.S.A. 106 (27), 11382–11387.

Zanon, M., Borgomaneri, S., Avenanti, A., 2018. Action-related dynamic changes in inferior frontal cortex effective connectivity: a TMS/EEG coregistration study. Cortex 108, 193–209.

Zhou, Y., Han, S., 2021. Neural dynamics of pain expression processing: alpha-band synchronization to same-race pain but desynchronization to other-race pain. Neuroimage 224, 117400, 2021.