Autobiographical memory reactivation in empathy

Federica Meconi*[1], Juan Linde-Domingo[2], Catarina S. Ferreira[1], Sebastian Michelmann[3], Bernhard Staresina[1,4], Ian Apperly[1], Simon Hanslmayr[1,4]

[1] School of Psychology, University of Birmingham

[2] Max Plank Institute Berlin for Human Development, Berlin

[3] Princeton Neuroscience Institute

[4] Center for Human Brain Health, University of Birmingham

*Corresponding Author: Federica Meconi, f.meconi.1@bham.ac.uk

Abstract

Empathy relies on the ability to mirror and to explicitly infer others' inner states. Studies on healthy populations show consistent evidence supporting the idea that our memories play a role in empathy when building a representation of others' inner states (Buckner & Carroll, 2007a; Spreng & Grady, 2009a; Spreng, Mar, & Kim, 2009). However, direct evidence of a reactivation of autobiographical memories (AM) when it comes to empathizing with others' inner states is yet to be shown. To address this question, we conducted two experiments where we recorded electrophysiological (Exp 1) and hemodynamic activity (Exp2) from two independent samples of participants. In Exp 1, EEG was recorded from 28 participants who performed a classic empathy task, i.e. a pain decision task in which targets for empathy were depicted in painful scenes for which participants either did or did not have an AM, followed by a task that explicitly required memory retrieval of the AM and non-AM scenes. The retrieval task acted as a 'localizer' to extract the neural fingerprints of AM and non-AM scenes, which could then be used to probe data from the empathy task. A state-of-the-art EEG pattern classifier was trained and tested across tasks and showed evidence for AM reactivation when participants were preparing their judgement in the empathy task. Participants self-reported higher empathy for people depicted in situations they had experienced themselves (for which they would have an AM) as compared to situations they had not experienced. This behavioral result was replicated in a second fMRI Experiment, where hemodynamic responses were measured from an independent sample of 28 participants. Furthermore, fMRI results showed activation in the brain networks that have been extensively shown in previous studies to underlie both AM retrieval and empathy (Amodio & Frith, 2006a; Bernhardt & Singer, 2012a; Buckner & Carroll, 2007a; Frith & Frith, 2003a; Spreng & Grady, 2009a; Spreng et al., 2009; Zaki & Ochsner, 2012a). Together, our study reports behavioral, electrophysiological and fMRI evidence that robustly supports the involvement of autobiographical memory reactivation in empathy.

Keywords: Empathy, EEG, LDA, fMRI, EEG pattern classifier, autobiographical memory

Significance statement

Human empathy is at the basis of our ability to bond with other people and interact in a meaningful way. Since we cannot fully experience what someone else is feeling, a big challenge in psychology is to understand how empathy occurs. We might use our own mind and experience as a model to represent others' inner states. To shed new light onto this debate, we used established and novel paradigms and analysis tools in the fields of social and cognitive neuroscience. In two experiments, we present compelling evidence that shows that

participants reactivate AMs when asked to empathize with another person. The current study has the potential to increase scientific knowledge to ultimately impact a wide range of professions working with people in need in the health and social sectors.

Introduction

Does our reaction to a friend who just broke her leg depend upon whether or not we have broken a leg ourselves? It is intuitively compelling that our empathy towards other people draws upon memory of our first-hand experiences. However, this intuition is challenged by evidence that patients with memory impairments seem to have preserved empathic abilities. The present studies addressed this puzzle directly by seeking direct evidence of the re-activation of autobiographical memories in the service of empathizing with others.

Empathy is a building block of human social cognition; it allows to share and to understand others' inner states and thoughts. It relies on the simulation of others' inner states (i.e. "affective" or "hot" empathy) and on the explicit reasoning about others' minds (i.e. "cognitive" or "cold" empathy); processes that are at least partly dissociable (Fan & Han, 2008; Reniers, Corcoran, Drake, Shryane, & Völlm, 2011; Sessa, Meconi, & Han, 2014; Shamay-Tsoory, Aharon-Peretz, & Perry, 2009; Zaki & Ochsner, 2012b).

Evidence from functional neuroimaging indicates that empathy and autobiographical memory draw on common brain networks. A meta-analysis (Spreng, Mar, & Kim, 2008) and other studies comparing autobiographical memory (AM) retrieval and theory of mind, i.e. "cold" empathy, within or across studies (Buckner & Carroll, 2007b; Rabin, Gilboa, Stuss, Mar, & Rosenbaum, 2009; Spreng & Grady, 2009b), showed that they share a network of fronto-temporo-parietal regions that includes precuneus, posterior cingulate cortex, retrosplenial cortex, medial temporal lobe, temporoparietal junction and medial prefrontal cortex (BA 10). There is also evidence from neuroimaging that empathy is enhanced by relevant personal experience. For example, inner simulation of others' actions develops along with infants' motor repertoire (van Elk, van Schie, Hunnius, Vesper, & Bekkering, 2008) and it is stronger for complex movements that adults practiced when compared to those they never learnt(Cross, Hamilton, & Grafton, 2006). Moreover, adults use their own experience of pain and emotion to infer others' inner states (Bluck, Baron, Ainsworth, Gesselman, & Gold, 2013; Mitchell, Banaji, & Macrae, 2005; Perry, Hendler, & Shamay-Tsoory, 2011) and to act prosocially (Bastian, Jetten, & Ferris, 2014; Gaesser & Schacter, 2014). In a functional magnetic resonance imaging (fMRI) study on a rare population of patients with congenital insensitivity to pain, Danziger and colleagues (Danziger, Faillenot, & Peyron, 2009) directly investigated whether these patients could share a pain for which they could have no autobiographical

memory. Intriguingly, patients did show similar activation to controls in critical neural regions involved in empathy for pain, suggesting that autobiographical memory doesn't play a critical role in such neural activations. However, explicit ratings of pain intensity to pictures depicting a body part in pain were significantly lower in patients than in control groups, suggesting that at the same patients' empathy for others' pain might be attenuated by their lack of past pain experiences. However, in contrast to the convergent evidence above, patients with autobiographical memory impairments, such as amnesic patients, show little evidence of impaired empathy, at least for cold empathy, which seems to be spared (Rabin, Braverman, Gilboa, Stuss, & Rosenbaum, 2012; Rosenbaum, Stuss, Levine, & Tulving, 2007) or only mildly impaired (Beadle, Tranel, Cohen, & Duff, 2013; Staniloiu, Borsutzky, Woermann, & Markowitsch, 2013). In sum, current neuroimaging research provides strong evidence for a link between autobiographical memory and empathy, but lacks the critical evidence that autobiographical memories are actually retrieved in the service of empathy. Neuropsychological evidence indicates that the best place to look for such evidence might be "hot" rather than "cold" empathy.

One example to highlight the potential interplay between memory and empathy is empathy for physical pain when the observer has themselves experienced the pain and so has a related AM. Bluck and colleagues(Bluck et al., 2013), showed that self-report of empathic abilities of participants with past experiences of pain increased after reading the story of a character who perceives chronic pain. Recent neuroimaging studies showed that both incidental (Forkmann, Wiech, Sommer, & Bingel, 2015) and voluntary (Fairhurst, Fairhurst, Berna, & Tracey, 2012) reinstatement of autobiographical pain involves partial reinstatement of activity in the brain areas that processed the perception of nociceptive stimuli. Recent advances in multivariate pattern analysis methods showed that brain activity patterns can be tracked during the encoding of new episodes and can be observed also during the retrieval of those episodes (Chen et al., 2017; Johnson, McDuff, Rugg, & Norman, 2009; Linde-Domingo, Treder, Kerrén, & Wimber, 2019; Michelmann, Bowman, & Hanslmayr, 2016; Staresina, Henson, Kriegeskorte, & Alink, 2012). Using this approach, we sought the first direct evidence for online reactivation of autobiographical memories when participants are required to empathize with others' physical pain.

In a series of two studies we aimed at investigating memory reactivation when participants were required to make an explicit judgment of their empathy awareness for others' inner states (Figure 1d). In Exp 1, EEG was recorded from 28 participants while performing two sequential tasks, i.e. starting with a pain decision task, classically used to prompt an empathic reaction (Figure 1a) and followed by a memory retrieval task, which was used to extract the neural fingerprints of AMs and non-AMs (Figure 1b). A linear discriminant analysis EEG

pattern classifier was trained during the retrieval task and tested on data obtained from the preceding empathy task to test for the online reactivation of the memories when participants were preparing their explicit judgment of empathy awareness for others' physical pain. In Exp 2, blood oxygen level dependent (BOLD) response was collected from an independent sample of 28 participants while performing the same empathy task to verify that the empathy task did indeed elicit activation in brain areas commonly associated with empathy and episodic memory. Empathy awareness judgements were collected as subjective report scores (Figure 1c).

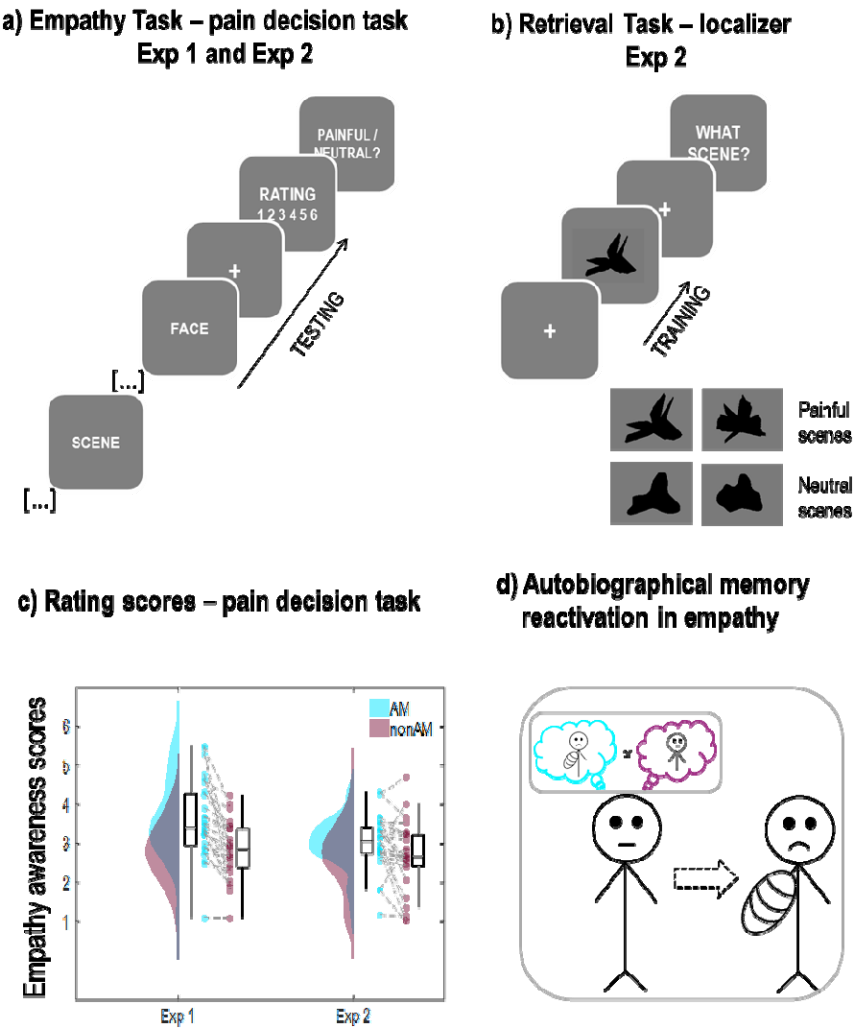

Figure 1. a) Schematic representation of the empathy task used in Exp 1 and Exp 2. Participants were required to rate how much empathy they felt for the person depicted in the preceding context. b) Schematic representation of the retrieval task used in Exp 1 that was used to train the LDA classifier. Participants first learnt to associate four abstract figures with the same sentences describing painful scenes presented during the empathy task (not shown here). In the actual task, for each trial participants were presented with one of the four figures and had to picture in their mind's eye the scene that they learnt to associate with that specific figure. c) Raincloud plots of the subjective reports of participants' empathy awareness in Exp 1 and Exp 2. d) Concept of the study; when we encounter someone who shares our same physically painful experience, memory of that experience is reactivated to empathize.

Results

Behavioural results.

Individual scores of the empathy rating revealed a main effect of the type of memory in both experiments (Exp 1: $F(1,27) = 22.319$; $p < 10^{-4}$, $\eta_p^2 = .453$; Exp 2: $F(1,27) = 6.348$; $p = .018$, $\eta_p^2 = .190$). Individuals depicted in contexts describing participants' autobiographical scenes drove enhanced explicit judgements of empathy awareness when compared to contexts describing non-autobiographical scenes. These results are shown in Figure 1c). A full description of the behavioural results is reported in the Supplementary Materials.

EEG results

LDA. We first ran a sanity check of the classifier on the retrieval task. The classifier was trained and tested with a K-fold cross-validation procedure during the presentation of the cue (Figure 1b). The square-shape of the time by time generalization matrix shown in Figure 2a showed that the task allowed the formation of stable representations associated with the figures (1 random polygon and 1 rounded shape for AMs and the same for the non-AMs) acting successfully as a localizer for the two types of memories (see the Methods section for more details). The bootstrapping analysis performed on a 0–2.5s time-window showed that the classifier could distinguish with a peak accuracy of 0.55 between AM and non-AM ($p = 0.0129$) in a sustained time-window (0 – ~2.2 secs), including a late time window that is most likely related to the representation of the memory itself rather than to any perceptual features of the stimuli. In a second step, the classifier was trained during the presentation of the cue in the retrieval task and then tested on the empathy task in a 2secs time-window starting from the onset of the face. Crucially, any consistency in the neural pattern observed across tasks would show the representation of the memories. The bootstrapping analysis revealed a significant cluster ($p = 0.0174$) in a sustained time-window (0.6–2 secs) showing evidence for the online reactivation of the memory in preparation of the empathy judgement with a peak accuracy of 0.54. The result of the classifier across tasks is shown in Figure 2b.

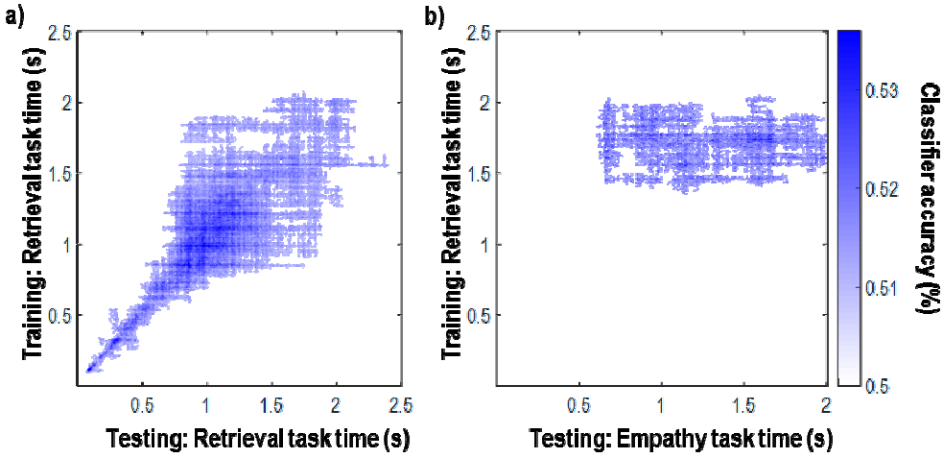

Figure 2. a) Sanity check: time by time generalization matrix showing significant classification of AM vs non-Am within the retrieval task. b) Time by time generalization matrix (i.e. training and testing at each time-point) showing significant classification of AM vs non-AM across tasks.

ERPs. Cluster analysis conducted over a 1sec time-window, from the onset of the face until the presentation of the rating, revealed one anterior and one posterior cluster of electrodes showing that ERPs significantly differ as a function of the type of memory ($p = 0.002$; anterior: min $t(27) = -2.132$; posterior: min $t(27) = 2.107$; Figure 3 depicts ERPs for AM and non-AM in the left panel and the topography of the significant clusters in right upper panel, t-values are plotted). Source analysis estimated that the neural source of this effect was the Superior Frontal Gyrus, BA 10, MNI: [-10 69 0] (Figure 3 right bottom panel).

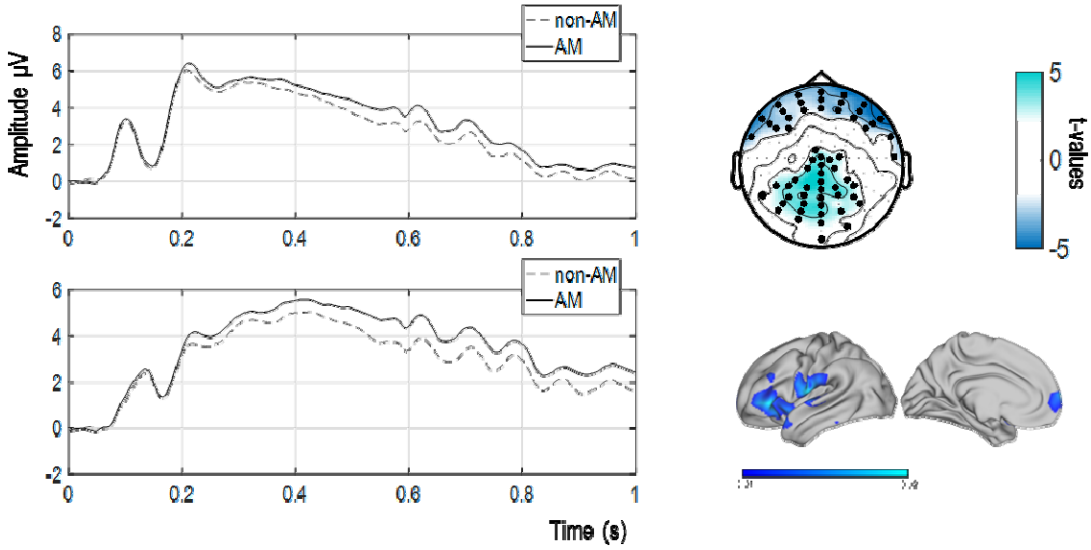

Figure 3. Left panel: ERPs time-locked to the onset of the face and reflecting AM and non-AM at the frontal and the posterior cluster. Right top panel: clusters analysis performed over all the electrodes in a 0 – 1 sec time-window. Colors code t-values. Right bottom panel: source localization of the AM vs non-AM contrast.

Additionally, in line with previous studies on empathy for physical pain, cluster analysis also revealed a classic ERP response associated with empathic processes (e.g.,(Sessa, Meconi, & Han, 2014), i.e. painful faces elicited more positive ERP response than neutral faces (min $t(27) = 2.164$, $p = 0.006$). Consistently, source analysis estimated that the neural source of this effect was the Inferior Frontal Gyrus, BA 9, MNI: [-62 21 30] and the Parietal Lobule, BA 7, MNI [30 -69 48].

fMRI results.

Figure 4 shows masked clusters resulting from the whole-brain analysis.

Time-locked analysis to the onset of the scene. The contrast AM>nonAM revealed a significant FWE corrected ($p <. 05$) cluster with a peak in the Precuneus, BA 7, MNI: [3 -64 38], (150 voxels), $t(27) = 5.76$ $p = 0.001$, in the Superior Parietal Lobule, BA 7, MNI: [-36 -58 59], (114 voxels), $t(27) = 4.42$, $p = 0.003$ extending to the Inferior Parietal Lobule (BA 40) and Superior Temporal Gyrus (BA 39); and a cluster with a peak in the Posterior Cingulate, BA 23, MNI: [3 -28 26], (62 voxels), $t(27) = 4.71$, $p = 0.038$. Masked clusters showing greater activation for AM as compared to non-AM are depicted in Figure 4a. The opposite contrast did not reveal any significant FWE corrected cluster.

Analysis time-locked to the onset of the face. Figure 4b shows the result of the contrast AM>nonAM. Greater activation for AM as compared to non-AM was observed in a significant FWE corrected ($p < .05$) in the Superior Frontal Gyrus, BA 10, MNI: [-18 62 23], (66 voxels), $t(27) = 5.49$ $p = 0.024$, and in a cluster in the Inferior Parietal Lobule, BA 39, MNI: [-36 61 41], (75 voxels), $t(27) = 3.67$, $p = 0.014$. This specific region of the IPL is part of the functional fractionation of the Temporoparietal Junction (TPJ) and is considered as part of the core network of the theory of mind (Schurz, Radua, Aichhorn, Richlan, & Perner, 2014). The opposite contrast revealed greater activation for non-AM than AM in a significant FWE corrected cluster in the parahippocampal gyrus (PHG), BA 36, MNI: [-18 -16 -22], (169 voxels), $t(27) = 6.22$, $p < 0.001$.

The results of the ROI analysis are reported in the Supplementary Materials

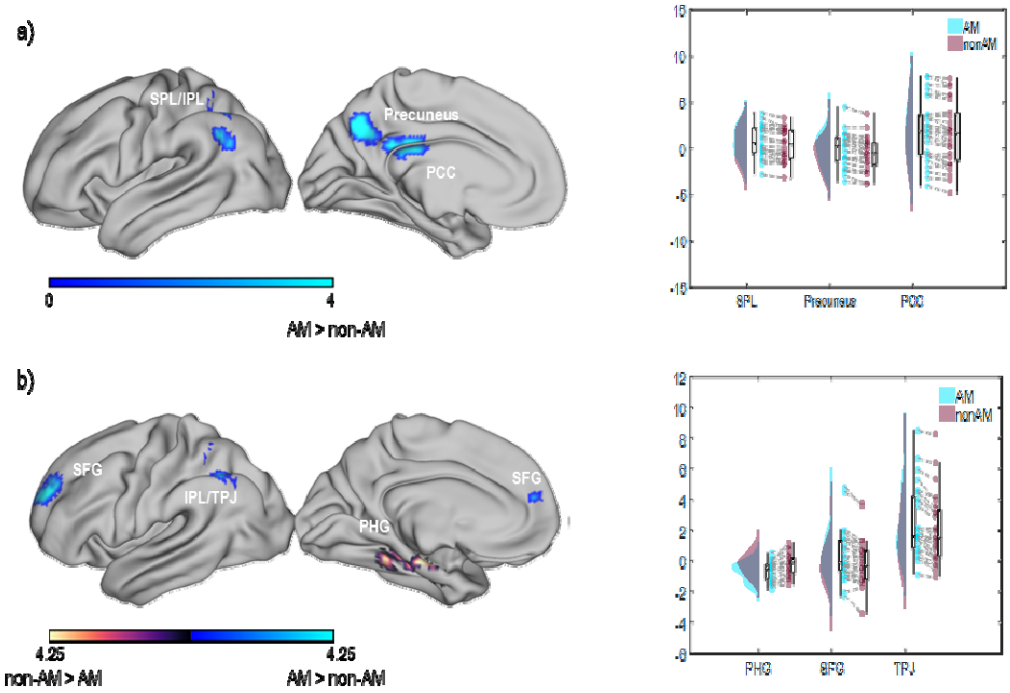

Figure 4. Whole-brain analysis results (left panel) and raincloud plots (right panel) of the activation in each condition and each cluster. A) Whole-brain analysis related to the presentation of the context. Only the contrast AM>non-AM showed significant clusters. B) Whole-brain analysis related to the presentation of the face. Figure shows significant clusters resulting from both the AM>non-AM and non-AM>AM contrasts.

Discussion

We investigated the intuitive idea that autobiographical memories for relevant pain experiences get reactivated when people empathize with others' physical pain. We conducted one EEG and one fMRI study and provided a clear-cut picture of the online reactivation of the memories when participants explicitly judge their empathy awareness and the brain areas underlying this mechanism.

In Exp 1, EEG was recorded from 28 participants while performing two sequential tasks, i.e. one pain decision task, classically used to prompt an empathic reaction and one memory retrieval task, which was used to extract the cortical EEG patterns that were the neural fingerprints of AMs and non-AMs. These data were used to probe the data from the empathy task. We applied a linear discriminant analysis EEG pattern classifier that was trained and tested across tasks. We showed direct evidence for online reactivation of the memories when participants were required to make an explicit judgement about their empathy awareness for others' physical pain. Participants self-reported higher empathy for people depicted in situations they had experienced themselves as compared to situations they had not experienced. This behavioural result was replicated in Exp 2 where we also collected BOLD responses from an independent sample of 28 participants to verify that the empathy task did

indeed elicit activation in brain areas commonly associated with empathy and episodic memory. Whole-brain analyses related to the onset of the scene showed the activation of the Precuneus (BA 7), PCC (BA 23) and left SPL (BA 7). The activation of these brain areas is consistent with previous literature showing that these brain areas underlie both the retrieval of AM and of empathic processes (Amodio & Frith, 2006b; Bernhardt & Singer, 2012b; Buckner & Carroll, 2007b; Frith & Frith, 2003b; Mazzoni et al., 2019; Spreng et al., 2008; Zaki & Ochsner, 2012b). Lastly, whole-brain analysis related to the onset of the face and contrasting BOLD response for autobiographical contexts and non-autobiographical contexts showed that the neural source of this effect was in the left SFG (BA 10), in line with the source analysis of the ERP data in Exp 1 comparing AM and non-AM in the same time-window and in a specific region of the left IPL that is part of the functional fractionation of the TPJ (BA 39), core of the theory of mind network(Schurz et al., 2014).

In summary, the present results from two independent experiments provide behavioral, electrophysiological and fMRI evidence in support of a direct engagement of autobiographical memory reactivation in empathy.

Our results are in line with previous studies suggesting that participants' past experiences interact with empathic abilities. A recent behavioural observation by Gaesser and Schacter (Gaesser & Schacter, 2014) showed that participants' explicit willingness to help increases when remembering past prosocial behaviour. Similarly, reading a story of an individual in chronic pain can increase self-reported empathic abilities of participants who have experienced a moderate amount of physical pain in their life (Bluck et al., 2013). An fMRI study showed that when participants are exposed to individuals depicted in emotional situations and are explicitly asked to reason about the individuals' inner states, the implicit recalling of participants' past experiences can occur the more similar they perceive themselves to the protagonists of the scenes (Perry et al., 2011). Our study adds to this picture the first direct evidence on the role of AM in empathy. We highlight three critical features underwriting the robustness of the study design. First, the autobiographical component of the memories used to probe empathy was unprompted in the empathy judgement task. Participants were never requested to retrieve their own memory to perform neither of the tasks. Retrieving participants' own AMs could have no impact on the rating of their empathy awareness unless participants based their judgement on their own past experience. Therefore, the output of the LDA is remarkable because participants could entirely rely on their semantic knowledge to perform both tasks, in which case the classification within and across tasks would be unsuccessful. Second, the memory retrieval task was always performed after the empathy task to avoid that participants could be primed to specifically retrieve their own memories and information about participants'

AMs and non-AMs were collected not more than three days prior to the experiment. Third, we used perceptually different stimuli to prompt empathy for specific situations and to trigger the reactivation of those episodes (i.e. sentences in the empathy task, shapes in the retrieval task). This was done to avoid any overlap in the perceptual features that cued the memories in the two tasks and ensure that the classifier could only identify the neural fingerprint of the memories reactivation per se.

The EEG pattern classifier approach has been successfully adopted to differentiate between the retrieval of perceptual and semantic information of an episodic memory (Linde-Domingo et al., 2019) and in different mechanisms of memory(Jafarpour, Horner, Fuentemilla, Penny, & Duzel, 2013). The timing of the retrieval of an autobiographical memory has been shown to occur between 400 and 600 ms even when it is only spontaneously recalled (Addante, 2015; Hebscher, Ibrahim, & Gilboa, 2019). The time by time generalization output depicted in Fig 2 is in line with this evidence. The squared shape of the output of the classifier shows that the representation is stable across time (King & Dehaene, 2014). Fig 2a shows that the representation of the memories starts between 500ms and 1sec and lasts until ~2 secs. Fig 2b shows that the representation of the memories reactivate in the time-window when empathy judgement was prepared.

Event-related potentials studies investigating empathy for physical pain have shown that an empathic reaction, reflecting the processing of a painful experience, is expressed as a positive shift of the ERP response, compared to a neutral condition with (Fan & Han, 2008; Meconi, Doro, Lomoriello, Mastrella, & Sessa, 2018; Sessa & Meconi, 2015; Sessa, Meconi, Castelli, & Dell'Acqua, 2014) or without (Sessa, Meconi, & Han, 2014; Sheng & Han, 2012) relation to explicit or implicit measures of empathy. In Exp1, we observed within a 1sec time-window a positive shift in the ERPs reflecting the processing of painful when compared to neutral faces in a cluster of centro-parietal electrodes that was estimated to be generated by the IPL and the IFG. Within the same time-window, ERPs time-locked to the onset of the faces reflecting the processing of the preceding memory showed a positive shift of the ERPs for AM as compared to non-AM. The neural source of this effect was estimated to be in the SFG. Our ERP results and the source analysis substantially replicated previous evidence(Sessa, Meconi, & Han, 2014) and are in line with our fMRI results obtained in Exp 2 and other neuroimaging studies on the neural underpinning of empathy (Amodio & Frith, 2006c; Bernhardt & Singer, 2012b; Fan, Duncan, de Greck, & Northoff, 2011; Lamm, Decety, & Singer, 2011; Shamay-Tsoory et al., 2009). Together with the behavioural evidence showing greater empathy for people depicted in AM contexts, these

results suggest that memories are processed within a similar time-window as when the empathic processes occur.

According to the multiple memory system of social cognition (MMS), prejudice and stereotyping are the result of affective and semantic associations in memory (Amodio & Ratner, 2011) resulting from autobiographical experience as well as from acquired knowledge. Racial stereotypes can directly modulate the N400 such that a larger amplitude is observed when other-race faces primed incongruent (i.e. own-race specific), when compared to congruent, personality traits of the target (Hehman, Volpert, & Simons, 2014). Newly learned associations between a human name and a non-human entity can influence empathic responses within 200 ms of exposure to the humanized non-human entity (Vaes, Meconi, Sessa, & Olechowski, 2016). Studies on cross-racial empathy for pain showed that empathic responses are more natural for own-race faces or more familiar faces when compared to other-race faces (Avenanti, Sirigu, & Aglioti, 2010; Sessa, Meconi, Castelli, et al., 2014; Xu, Zuo, Wang, & Han, 2009). These ERP studies therefore provided some parallel evidence that empathy can be affected by episodic memories that are implicitly or spontaneously accessed(Meconi, Vaes, & Sessa, 2015).

In Exp 2, participants gave higher empathy ratings for individuals depicted in autobiographical compared to non-autobiographical contexts. Drawing from the results from Exp 1, with Exp 2 we expected to find activation of the brain areas that have been found to overlap between AM retrieval and empathy (Buckner & Carroll, 2007b; Spreng & Grady, 2009b; Spreng et al., 2008) or that have been more strongly associated with empathic processes (Gaesser, 2018; Lamm et al., 2011; Zaki & Ochsner, 2012b), including the SFG (Figure 3).

Among previous evidence, a meta-analysis investigating the core network of theory of mind(Schurz et al., 2014) demonstrated that, together with the mPFC, the left TPJ is a core brain area of theory of mind, i.e. cold empathy. Damage of the left TPJ can selectively reduce theory of mind abilities but not other cognitive or executive abilities (Apperly, Samson, Chiavarino, & Humphreys, 2004; Bzdok et al., 2013; Samson, Apperly, Chiavarino, & Humphreys, 2004). Our fMRI results nicely dovetail with these expectations. Whole-brain analysis related to the onset of the faces contrasting AMs and non-AMs, revealed the activation of the left SFG and of the region of the IPL part of the functional fractionation of the left TPJ.

Studies (Buckner & Carroll, 2007b; Rabin et al., 2009; Spreng & Grady, 2009b; Spreng et al., 2008) that investigated the common neural basis of AM retrieval and theory of mind, i.e. “cold” empathy, showed that these processes share a network of fronto-temporo-parietal regions, including precuneus, posterior cingulate cortex, retrosplenial cortex, medial temporal lobe, temporoparietal junction and medial prefrontal cortex (BA

10). The parietal cortex underlying the retrieval of specific AM also involves the activation of the SPL (Addis, McIntosh, Moscovitch, Crawley, & McAndrews, 2004). Our whole-brain analysis revealed the activation of the SPL with a cluster extended to the TPJ and of the precuneus and PCC. It is important to notice that even though the parietal cortex has a substantial involvement in AM retrieval (Berryhill, Phuong, Picasso, Cabeza, & Olson, 2007), the empathy task did not explicitly require to retrieve the specific memory. This is an important difference to classic AM studies in which active and conscious search of the memory is usually requested to participants (Addis, Knapp, Roberts, & Schacter, 2012).

The brain network of conscious autobiographical memory retrieval includes also the medial temporal lobe (Addis et al., 2012, 2004; Berryhill et al., 2007; Buckner & Carroll, 2007a; Cabeza & St Jacques, 2007). In the current study, we did not observe the activation of the medial temporal lobe in the contrast AM>non-AM. It is worth underlining that participants' AMs were on average 5 years old. A recent review (Barry & Maguire, 2019) highlighted that although memories seem to become independent from hippocampal activation with remoteness, hippocampus remains involved in scene/memory reconstruction (Zeidman & Maguire, 2016) even though the original memory trace is with time stored in the neocortex. We did observe the activation of the PHG in the contrast non-AM>AM. This result was surprising and we can only speculate that the activation of the PHG in this contrast is consistent with the studies showing the involvement of the episodic simulation in performing tasks that prompt cold empathy (Gaesser, 2018; Perry et al., 2011).

The present study provides robust new evidence of re-activation of AMs in the context of empathy. However, it does not allow us to address the causal relationship between memory and empathy (i.e., whether memory directly drives empathy judgements). Addressing this would require future work that sought to disrupt memory retrieval in a time-sensitive manner. Furthermore, puzzling previous evidence showing that patients with amnesia seem to have preserved empathic abilities opens future research question on whether they could show spared ability to retrieve the information despite their inability to consciously report it.

Methods

The protocol for both experiments was approved by the University of Birmingham Research Ethics Committee (ERN-16-0101A). Written informed consent was obtained from all participants for both experiments.

Participants

Experiment 1. EEG.

Thirty-five healthy students took part to the experiment. Participants (mean age = 22, SD = 5) were recruited through the Research Participant Scheme of University of Birmingham for cash (£10/h) or course credits (1 credit/h). Four participants were left-handed, 6 were males. Seven participants were discarded from the final sample, three were not Caucasian (we showed pictures of Caucasian faces and previous studies have shown that empathic responses are subject to ethnicity bias (Avenanti et al., 2010; Sessa, Meconi, Castelli, et al., 2014; Xu et al., 2009)), two participants could not complete the task due to equipment failure and two for too low number of trials due to inaccurate responses. The final sample was composed of 28 participants (mean age = 21.96, SD = 4.82), four were males and four left-handed. All of the participants had normal or corrected-to normal vision. The eligibility criteria included native of excellent English proficiency, no history of neurological or psychiatric disorder, and having an experience of intense physical pain in their past.

Experiment 2. fMRI.

Thirty three healthy students took part to the experiment. Participants (mean age = 25, SD = 5.9) were recruited through the Research Participants Scheme of University of Birmingham for monetary compensation (£10/h) or for course credits (1 credit/h). Participants were all right-handed, 15 were males. Five participants were discarded from the final sample, two served as pilots to adjust the timing of the paradigm and make it suitable for the fMRI environment; one participant could not complete the acquisition session in the scanner, two were discarded for excessive movements (more than 1 voxel size, 3mm). The final sample was composed of 28 participants (mean age = 24.71, SD = 5.86), eleven were males. All of the participants had normal or corrected-to normal vision.

Questionnaires

With the aim of providing a validated score of the self-perceived empathy range of our sample of participants, the Empathy Quotient (EQ (Baron-Cohen & Wheelwright, 2004) was collected at the end of the experimental session. The EQ comprises 60 items including 20 filler items. Participants respond on a 4-points scale that ranges from “strongly agree” to “strongly disagree”, the score ranges from 0 to 80 (low: 0–32, middle: 33–52, high 53–63, extremely high: 64–79, and maximum empathy: 80). The Interpersonal Reactivity Index (IRI(Davis, 1983) was also administered to collect a finer-grained picture of the empathy traits. The IRI comprises 28 items subdivided in four subscales, two underpinning the affective component of empathy (i.e., the empathic concern, EC, and the personal distress, PD) and two underpinning the cognitive component of empathy (i.e., the perspective-taking, PT, and fantasy subscales, FS). Responses are given on a 1 to 5-points scale; no ranges of normal or abnormal empathy are identified.

The 20 items Toronto Alexithymia Scale (Bagby, Parker, & Taylor, 1994) was administered to participants with the aim of showing that our participants had on average normal ability of describing their own emotions. Responses are given on a 1-5 points scale and the maximum possible score is 80. Alexithymia is shown with scores above 61. This was important as previous studies showed that alexithymia patients have also impaired others' emotions processing.

Participants from both experiments fell in the normal range (Baron-Cohen & Wheelwright, 2004) of the EQ (Exp 1: M = 51.14 SD = 9.99; Exp 2: M = 46.89 SD = 12.37), and had on average normal ability to describe their emotions as showed by the TAS score (Exp 1: M = 45.96, SD = 12.22; Exp2: M = 43.96, SD = 12.01). The IRI scores for both experiments are reported in supplementary materials (Table S1).

The autobiographical memories reported by the participants were on average 4.65 years old (SD = 5 y) for Exp 1; 4.46 years old (SD = 5.72 y) for Exp 2.

These measures were also used to explore any relation with the neural responses.

During the screening phase, the Autobiographical Memory Questionnaire (AMQ) (Rubin, Schrauf, & Greenberg, 2003) was collected for both the physically painful and the neutral memories of the participants. The AMQ is a useful tool to collect ratings about the vividness of the memory and the accuracy with which participants can recollect and relieve the event. It also requests to date the memory.

Stimuli and Procedure

Experiment 1. EEG

As soon as participants signed up for the study, they were contacted by the experimenter and screened to identify an experience of intense physical pain and an emotionally and physically neutral experience for which participants completed the AMQ(Fink et al., 1996; Maguire, Henson, Mummery, & Frith, 2001). During this phase, a physically painful and a neutral episode that didn't belong to participants' autobiographical experiences were identified.

Participants performed two tasks in the same experimental session. For all participants, the first of these tasks was a variant of the pain decision task (Meconi et al., 2018; Sessa, Meconi, & Han, 2014) and the second one was a retrieval task. As for the previous studies using this variant of the pain decision task, stimuli were sentences and faces. The sentences described autobiographical or non-autobiographical contexts of painful and neutral episodes. The sentences always followed the structure "This person got – […]" or "This person did – […]" so that all the sentences had the same syntactic complexity. The faces were a set of 16 identities, 8 males

and 8 females with a painful or a neutral facial expression. The faces were in shades of grey and they were equalized for luminance with the SHINE toolbox (Willenbockel et al., 2010).

For the retrieval task a set of 8 pairs of figures were created. 8 random polygons with equal number of black pixels were created ad hoc. The figures were then blurred with a Gaussian filter and all the pixels in shades of grey were made black. As a last step, the number of black pixels was equalized across all the figures (i.e., random polygons and the rounded shapes obtained from the polygons). The pair of figures (1 random polygon and its rounded shape) was coupled with another pair of figures; the full combinations of this coupling were 28 pairs of coupled figures. The figures were shown on a grey background (Figure 1b).

The pain decision task. For each trial participants had to read the sentence and then see the face and they were asked to rate the empathy they felt for that person depicted in that particular context on a subjective scale of 6 points. During the practice session to get familiar with the task, each participant was told the definition of empathy in the present context "Empathy is the ability to share the others' emotions in a way that if you observe somebody being sad, you feel sad as well. You don't feel sorry for that person because that would be compassion. In the same way, if you see somebody not expressing any particular emotion, you are not asked to say whether this person looks nice to you because that would be sympathy. Empathy is then the ability to mirror the others' emotions". The task was composed of 48 trials per condition that were pseudo-randomized in a way that the conditions were balanced over the full session. There were 192 trials in total subdivided in 4 blocks that took participants up to 30 minutes to complete. An illustration of the task is shown in Figure 1a.

The memory retrieval task. The memory retrieval task was composed of three phases: a) a learning phase, b) the practice session and c) the actual task. During the learning phase, participants learnt to associate one of the four sentences they read in the pain decision task with one of the four figures. The pairs were shown to the participants for the studying phase 8 times. After the studying phase, participants were tested on their memory; they could press one of four response keys on the computer keyboard. Words cueing to the four episodes were used to allow participants to answer the memory task. Each word could appear at one of four locations, each word location was randomized with a Latin square in a way that each word had the same likelihood to appear at one of the four locations. The learning phase could be ended after 8 trials, i.e. 2 repetitions per pair, when full accuracy was reported. One error within a block of 8 trials made the task start another block of trials until full accuracy was reached. Participants could start the practice session only when perfect performance was obtained in the learning phase. The practice session was performed to familiarize with the actual task and was a block of

8 trials of the exact same task. In the retrieval task, participants were only shown the figures. In each trial, one figure was shown and participants had to picture in their mind's eye the scene that was associated with that figure. They were required to press a button as soon as they could picture the scene accordingly to the vividness of the image in their mind's eye. They could press one of six response keys "s", "d", "f", "j", "k", "l" with "s" for "not vivid at all" and "l" for "very vivid". The figure was shown for 3 secs and participants had to press one of the 6 buttons within the time of the presentation of the figure. If they did not press any button, a "No Response" was recorded and the trial excluded from the analysis. The figure remained on the screen for 3 secs even if a response was collected. After the figure, participants were asked to say which scene they had pictured in their mind's eye using "d" key, if the episode was cued by the first word on the left, "f", for the second word, "j", for the third word and "k" for the last word that was displayed on the right. The cue-words could appear with equal likelihood at each of the four locations following a Latin square randomization. If they could not remember what episode was associated with that figure, they could press the space bar for "forgotten" and go to the next trial. Only correct trials were included in the analysis. There were 60 trials per condition that were pseudo-randomized to balance the distribution of all the conditions over the total of 290 trials that constituted the full session. The task, depicted in Figure 1b, right panel, was subdivided in 4 blocks that could last up to 30 minutes.

Stimuli were presented in black font on a grey background of a 17'' computer screen with a refreshing rate of 70 Hz. The tasks were programmed using Psychtoolbox.

Experiment 2. fMRI

The screening phase, the questionnaires and the procedure were the same as those used in Experiment 1 with the only exception of the adjustments in the timing of the events made to the pain decision task in order to make it suitable for the fMRI environment.

In the adjusted paradigm, each trial started with a fixation cross of a variable duration (2, 2.2, 2.4, 2.5, 2.6, 2.8, 3secs) followed by the sentence that was on screen for 3 secs. The fixation cross between the face and the sentence was jittered between 1.5 and 2.5 secs in steps of 100ms. Both the face and the following fixation cross were lengthened up to 1 sec and both the rating and the pain decision task were self-paced but if no response was given within 2 seconds the task would continue to the next trial.

Data acquisition and analysis

Experiment 1. EEG data acquisition and analysis

The EEG was recorded using a BioSemi Active-Two system from 128 Ag/AgCl active electrodes. The EEG was re-referenced offline to the average reference. Three additional external electrodes were placed below the left eye and on the lateral canthi of each eye to record vertical EOG. EEG and vertical EOG signals were digitized at a sampling rate of 1024 Hz via ActiView recording software (BioSemi, Amsterdam, the Netherlands). Individual electrodes' coordinates were logged with a Polhemus FASTRAK device (Colchester, Vermont, USA) in combination with Brainstorm implemented in MATLAB 2014b (MathWorks). For three participants the standard electrode coordinates were used due to technical problems during the experimental session.

EEG—analysis. EEG data was analyzed with MATLAB (©Mathworks, Munich, Germany) using the open-source FieldTrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011) (http://fieldtrip.fcdonders.nl/) and in-house Matlab routines.

Univariate ERP analysis

Pain Decision task. EEG data were first segmented into epochs of 4 s, starting 1 sec before the onset of the face. The epoched data was visually inspected to discard large artifacts from further analysis. Further preprocessing steps included Independent Component Analysis for ocular artifacts correction and re-referencing to average reference. After removing trials which were contaminated by eye and muscle artefacts, an average of 45 trials (range: 34-48) remained for autobiographical memory and 45 trials (range: 37-48) for non-autobiographical memory condition.

Cluster-based permutation tests were performed on the event-related potentials time-locked to the onset of the face to account for significant differences between painful and neutral facial expressions in order to replicate previous findings and show an ERP empathic response to faces. Additionally, preliminary analysis was carried out to test for any involvement of memory in the empathy task. To this end, cluster-based permutation tests were performed on the ERPs time-locked to the onset of the face regardless of the facial expression contrasting autobiographical vs. non-autobiographical memory.

Retrieval task. EEG data were first segmented into epochs of 6 s, starting 2 seconds before the onset of the cue. The epoched data was visually inspected to discard large artifacts from further analysis. Further preprocessing steps included Independent Component Analysis for ocular artifacts correction and re-referencing to average reference. After removing trials which were contaminated by eye and muscle artefacts, an average of 51 trials (range: 42-60) remained for autobiographical memory and 50 trials (range: 38-58) for non-autobiographical memory condition.

Linear Discriminant Analysis EEG pattern classifier.

Linear Discriminant Analysis (LDA) is a multivariate pattern analysis method that finds a decision boundary that allows to distinguishing the pattern of brain activity associated with one category of stimuli from the pattern of brain activity that is associated with another category of stimuli. This is based on specified features of the EEG signal. It can then estimate with certain accuracy whether the pattern of brain activity in data that was not used to find the decision boundary, is more similar to one or the other category of stimuli.

In order to reduce unwanted noise and computational time, the signal was filtered between 0.1 and 40 Hz and down sampled to 128Hz before classification with a baseline correction window of 500ms before the onset of the stimuli.

The LDA was trained and tested on the EEG raw patterns (i.e. amplitude of the signal on each of the 128 electrodes), for each participant and at each time point and regularized with shrinkage (Blankertz, Lemm, Treder, Haufe, & Müller, 2011).

The classifier was trained on the signal acquired while participants were performing the retrieval task in the time-window including the presentation of the cue to detect systematic differences between the EEG patterns reflecting the representation of AM and non-AM scenes. It was then tested on the signal independently acquired while participants were performing the preceding pain decision task in the time-window from the onset of the face until the rating was presented. The aim of the LDA was to test for the online reactivation of the memory in preparation of the explicit judgement of participants' empathy awareness.

Before training and testing the LDA in two different datasets, we trained and tested the classifier on the retrieval task during the presentation of the cue to show that the task was successful to act as a localizer of the representation of the AM and non-AM scenes. A K-fold cross-validation procedure with 5 repetitions was used to train and test the classifier. The output of this analysis is the accuracy with which the classifier could distinguish between the two memory scenes for each time-point over all trials and electrodes. Therefore, the LDA reduces the data into a single decoding time course per dataset. To make sure that the output was not biased by the signal to noise ratio due to the different amount of trials, we equalized the number of trials for AM and non-AM before train the classifier.

Source Analysis

For source reconstruction, a time-domain adaptive spatial filtering linear constrained minimum variance (LCMV) beamformer (van Veen, van Drongelen, Yuchtman, & Suzuki, 1997), as implemented in fieldtrip was applied. As the source model a standardized boundary element model was used, which was derived from an averaged T1-weighted MRI dataset (MNI, www.mni.mcgill.ca). That was used in combination with individual

electrode positions logged into a Polhemus system. Source analysis was carried out for the time-domain ERP components that revealed significant results on the scalp level.

Statistical analysis

Average reaction time to the painful/neutral facial expression task was calculated for each participant. Mean proportions of accurate responses given within +/-2.5SD from the average reaction time of each participant and mean proportions of the empathy rating scores were computed for each condition and inserted in two repeated measures ANOVAs with a 2 (Emotional memory: Painful vs. Neutral) x 2 (Memory: Autobiographical vs. non-Autobiographical) x 2 (Facial expression: Painful vs. Neutral) as within-subject factors. Bonferroni corrected paired-sample t-tests were conducted when appropriate to explore significant interactive effects.

For the classifier analysis, an empirical null distribution was created with a combined permutation and bootstrapping approach (Stelzer, Chen, & Turner, 2013) that tested whether the maximum cluster of accuracy values above the chance level was statistically significant. We used the LDA in 100 matrices with pseudo-randomly shuffled labels independently for each participant and created a null distribution of accuracy values that we contrasted with the LDA outputs obtained with the real data. This was done by sampling with replacement 100000 times from the real and random data of each subject and computing a group average. This procedure resulted in an empirical chance distribution, which allowed us to investigate whether the results from the real-labels classification had a low probability of being obtained due to chance ($p < 0.05$) (i.e., exceeding the 95^{th} percentile).

Experiment 2. fMRI Acquisition and analysis

Data acquisition was performed with 3T Philips Medical Systems Achiva MRI scanner of the Birmingham University Imaging Centre using a 32-channel head coil. Functional T2-weighted images were acquired with isotropic voxels of 3mm, repetition time [TR] = 1750 msec, echo time [TE] = 30 msec, field of view [FOV] = 240x240x123 mm, and flip angle = 78°. Each volume comprised 33 sequentially ascending axial slices with an interslice gap of 0.75mm). Each participant underwent four blocks of scan series, one full block comprised 410 volumes. A high-resolution T1-weighted anatomical scan was acquired with an MPRAGE sequence (TR = 7.4 msec, TE = 3.5 msec, isotropic voxel size of 1mm, FOV = 256x256x176, flip angle = 7°) after the first two functional scanning blocks. The MR scanner was allowed to reach a steady state by discarding the first three volumes in each of the four scan series block.

fMRI Analyses

The analyses were performed using the SPM12 toolbox (University College London, London, UK; http://www.fil.ion.ucl.ac.uk/spm/). For each scanning block, a motion realignment of each slice to the first slice was carried out before time realignment (slices corrected to the middle one). Data was then linearly detrended, using a Linear Model of Global Signal algorithm (Macey, Macey, Kumar, & Harper, 2004) to remove any minimal fluctuation due to the physical setting. Functional images served as reference for the co-registration with the anatomical image. The data was further normalized to an MNI template, and finally, images were spatially smoothed with an 8-mm FWHM Gaussian kernel.

Two separate univariate analyses were carried out for two different time-windows, one analysis was time-locked to the onset of the face, and the other was time-locked to the onset of the context. This was only done to parallel Exp 1 and not to test for any functional dissociation between the two time-windows. In both cases statistical parametric maps were created for each participant's block of trials.

AM and non-AM conditions were directly contrasted in paired-sample t-tests on a group-level analysis.

A Region of interest (ROI) for the hippocampus was manually drawn for each participant for a direct investigation of the activation of this brain area in both autobiographical and non-autobiographical memory conditions against baseline. Baseline was manually computed as the mean activity of the whole trial.

Time-lock to the onset of the context. Regressors were defined for autobiographical and non-autobiographical memories time-locked to the onset of the contexts regardless of the emotional content of the scene described by the sentence. Additional regressors of no interest were again included in the design matrix to explain variance in the data not due to the experimental manipulation under investigation (first and third fixation crosses, presentation of the face, onset of the rating and of the pain decision question, the button presses) and the 6 motion parameters obtained during the realignment phase of the preprocessing. Sixty statistical parametric maps were created (4 blocks × 15 regressors) for each participant.

Time-lock to the onset of the face. Regressors were defined for autobiographical and non-autobiographical memories time-locked to the onset of the faces regardless of their emotional expression. Additional regressors of no interest were included in the design matrix to explain variance in the data not due to the experimental manipulation under investigation (first and second fixation crosses, presentation of the sentence, the onset of the pain decision question, the 2 button presses) and the 6 motion parameters. Fifty-four statistical parametric maps were created (4 blocks × 14 regressors) for each participant.

For both time-windows, a within-subject analysis was carried out on the data set of each participant to obtain the mean statistical parametric map for each experimental condition. Finally, a group-level paired-sample t-test

contrasting AM and non-AM was performed. A cluster-wise analysis was performed with uncorrected $p = .001$ and then Family Wise Error correction was applied for multiple comparison (cluster p threshold = .05). Peak voxel MNI are reported in brackets.

References

Addante, R. J. (2015). A critical role of the human hippocampus in an electrophysiological measure of implicit memory. *NeuroImage*, *109*, 515. https://doi.org/10.1016/j.neuroimage.2014.12.069

Addis, D. R., Knapp, K., Roberts, R. P., & Schacter, D. L. (2012). Routes to the past: Neural substrates of direct and generative autobiographical memory retrieval. *Neuroimage*, *59*(3), 2908–2922. https://doi.org/10.1016/j.neuroimage.2011.09.066

Addis, D. R., McIntosh, A. R., Moscovitch, M., Crawley, A. P., & McAndrews, M. P. (2004). Characterizing spatial and temporal features of autobiographical memory retrieval networks: A partial least squares approach. *NeuroImage*, *23*(4), 1460–1471. https://doi.org/10.1016/j.neuroimage.2004.08.007

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*(4), 268. https://doi.org/10.1038/nrn1884

Amodio, D. M., & Ratner, K. G. (2011). A Memory Systems Model of Implicit Social Cognition. *Current Directions in Psychological Science*, *20*(3), 143–148. https://doi.org/10.1177/0963721411408562

Apperly, I. A., Samson, D., Chiavarino, C., & Humphreys, G. W. (2004). Frontal and Temporo-Parietal Lobe Contributions to Theory of Mind: Neuropsychological Evidence from a False-Belief Task with Reduced Language and Executive Demands. *Journal of Cognitive Neuroscience*, *16*(10), 1773–1784. https://doi.org/10.1162/0898929042947928

Avenanti, A., Sirigu, A., & Aglioti, S. M. (2010). Racial Bias Reduces Empathic Sensorimotor Resonance with Other-Race Pain. *Current Biology*, *20*(11), 1018–1022. https://doi.org/10.1016/j.cub.2010.03.071

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia scale—I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research*, *38*(1), 23–32. https://doi.org/10.1016/0022-3999(94)90005-1

Baron-Cohen, S., & Wheelwright, S. (2004). The Empathy Quotient: An Investigation of Adults with Asperger Syndrome or High Functioning Autism, and Normal Sex Differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163–175. https://doi.org/10.1023/B:JADD.0000022607.19833.00

Barry, D. N., & Maguire, E. A. (2019). Remote Memory and the Hippocampus: A Constructive Critique. *Trends in Cognitive Sciences*, *23*(2), 128–142. https://doi.org/10.1016/j.tics.2018.11.005

Bastian, B., Jetten, J., & Ferris, L. J. (2014). Pain as Social Glue: Shared Pain Increases Cooperation. *Psychological Science*, *25*(11), 2079–2085. https://doi.org/10.1177/0956797614545886

Beadle, J. N., Tranel, D., Cohen, N. J., & Duff, M. (2013). Empathy in Hippocampal Amnesia. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00069

Bernhardt, B. C., & Singer, T. (2012). The Neural Basis of Empathy. *Annual Review of Neuroscience*, *35*(1), 1–23. https://doi.org/10.1146/annurev-neuro-062111-150536

Berryhill, M. E., Phuong, L., Picasso, L., Cabeza, R., & Olson, I. R. (2007). Parietal Lobe and Episodic Memory: Bilateral Damage Causes Impaired Free Recall of Autobiographical Memory. *Journal of Neuroscience*, *27*(52), 14415–14423. https://doi.org/10.1523/JNEUROSCI.4163-07.2007

Blankertz, B., Lemm, S., Treder, M., Haufe, S., & Müller, K.-R. (2011). Single-trial analysis and classification of ERP components — A tutorial. *NeuroImage*, *56*(2), 814–825. https://doi.org/10.1016/j.neuroimage.2010.06.048

Bluck, S., Baron, J. M., Ainsworth, S. A., Gesselman, A. N., & Gold, K. L. (2013). Eliciting Empathy for Adults in Chronic Pain through Autobiographical Memory Sharing. *Applied Cognitive Psychology*, *27*(1), 81–90. https://doi.org/10.1002/acp.2875

Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences*, *11*(2), 49–57. https://doi.org/10.1016/j.tics.2006.11.004

Bzdok, D., Langner, R., Schilbach, L., Jakobs, O., Roski, C., Caspers, S., … Eickhoff, S. B. (2013). Characterization of the temporo-parietal junction by combining data-driven parcellation, complementary connectivity analyses, and functional decoding. *NeuroImage*, *81*, 381–392. https://doi.org/10.1016/j.neuroimage.2013.05.046

Cabeza, R., & St Jacques, P. (2007). Functional neuroimaging of autobiographical memory. *Trends in Cognitive Sciences*, *11*(5), 219–227. https://doi.org/10.1016/j.tics.2007.02.005

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, *20*(1), 115–125. https://doi.org/10.1038/nn.4450

Cross, E. S., Hamilton, A. F. de C., & Grafton, S. T. (2006). Building a motor simulation de novo: Observation of dance by dancers. *NeuroImage*, *31*(3), 1257–1267. https://doi.org/10.1016/j.neuroimage.2006.01.033

Danziger, N., Faillenot, I., & Peyron, R. (2009). Can We Share a Pain We Never Felt? Neural Correlates of Empathy in Patients with Congenital Insensitivity to Pain. *Neuron*, *61*(2), 203–212. https://doi.org/10.1016/j.neuron.2008.11.023

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126.

Fairhurst, M., Fairhurst, K., Berna, C., & Tracey, I. (2012). An fMRI Study Exploring the Overlap and Differences between Neural Representations of Physical and Recalled Pain. *PLOS ONE*, *7*(10), e48711. https://doi.org/10.1371/journal.pone.0048711

Fan, Y., Duncan, N. W., de Greck, M., & Northoff, G. (2011). Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. *Neuroscience & Biobehavioral Reviews*, *35*(3), 903–911. https://doi.org/10.1016/j.neubiorev.2010.10.009

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. https://doi.org/10.1016/j.neuropsychologia.2007.07.023

Fink, G. R., Markowitsch, H. J., Reinkemeier, M., Bruckbauer, T., Kessler, J., & Heiss, W.-D. (1996). Cerebral Representation of One's Own Past: Neural Networks Involved in Autobiographical Memory. *Journal of Neuroscience*, *16*(13), 4275–4282. https://doi.org/10.1523/JNEUROSCI.16-13-04275.1996

Forkmann, K., Wiech, K., Sommer, T., & Bingel, U. (2015). Reinstatement of pain-related brain activation during the recognition of neutral images previously paired with nociceptive stimuli: *PAIN*, *156*(8), 1501–1510. https://doi.org/10.1097/j.pain.0000000000000194

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1431), 459–473. https://doi.org/10.1098/rstb.2002.1218

Gaesser, B. (2018). *Episodic Mind Reading: Mentalizing Guided by Imagined and Remembered Scenes*. https://doi.org/10.31234/osf.io/acv9r

Gaesser, B., & Schacter, D. L. (2014). Episodic simulation and episodic memory can increase intentions to help others. *Proceedings of the National Academy of Sciences*, *111*(12), 4415–4420. https://doi.org/10.1073/pnas.1402461111

Hebscher, M., Ibrahim, C., & Gilboa, A. (2019). Precuneus stimulation alters the neural dynamics of autobiographical memory retrieval. *BioRxiv*, 592436. https://doi.org/10.1101/592436

Hehman, E., Volpert, H. I., & Simons, R. F. (2014). The N400 as an index of racial stereotype accessibility. *Social Cognitive and Affective Neuroscience*, *9*(4), 544–552. https://doi.org/10.1093/scan/nst018

Jafarpour, A., Horner, A. J., Fuentemilla, L., Penny, W. D., & Duzel, E. (2013). Decoding oscillatory representations and mechanisms in memory. *Neuropsychologia*, *51*(4), 772–780. https://doi.org/10.1016/j.neuropsychologia.2012.04.002

Johnson, J. D., McDuff, S. G. R., Rugg, M. D., & Norman, K. A. (2009). Recollection, Familiarity, and Cortical Reinstatement: A Multivoxel Pattern Analysis. *Neuron*, *63*(5), 697–708. https://doi.org/10.1016/j.neuron.2009.08.011

King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210. https://doi.org/10.1016/j.tics.2014.01.002

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. https://doi.org/10.1016/j.neuroimage.2010.10.014

Linde-Domingo, J., Treder, M. S., Kerrén, C., & Wimber, M. (2019). Evidence that neural information flow is reversed between object perception and object reconstruction from memory. *Nature Communications*, *10*(1), 179. https://doi.org/10.1038/s41467-018-08080-2

Macey, P. M., Macey, K. E., Kumar, R., & Harper, R. M. (2004). A method for removal of global effects from fMRI time series. *NeuroImage*, *22*(1), 360–366. https://doi.org/10.1016/j.neuroimage.2003.12.042

Maguire, E. A., Henson, R. N., Mummery, C. J., & Frith, C. D. (2001). Activity in prefrontal cortex, not hippocampus, varies parametrically with the increasing remoteness of memories. *Neuroreport*, *12*(3), 441–444.

Mazzoni, G., Clark, A., De Bartolo, A., Guerrini, C., Nahouli, Z., Duzzi, D., … Venneri, A. (2019). Brain activation in highly superior autobiographical memory: The role of the praecuneus in the autobiographical memory retrieval network. *Cortex*. https://doi.org/10.1016/j.cortex.2019.02.020

Meconi, F., Doro, M., Lomoriello, A. S., Mastrella, G., & Sessa, P. (2018). Neural measures of the role of affective prosody in empathy for pain. *Scientific Reports*, *8*(1), 291. https://doi.org/10.1038/s41598-017-18552-y

Meconi, F., Vaes, J., & Sessa, P. (2015). On the neglected role of stereotypes in empathy toward other-race pain. *Social Neuroscience*, *10*(1), 1–6. https://doi.org/10.1080/17470919.2014.954731

Michelmann, S., Bowman, H., & Hanslmayr, S. (2016). The Temporal Signature of Memories: Identification of a General Mechanism for Dynamic Memory Replay in Humans. *PLOS Biology*, *14*(8), e1002528. https://doi.org/10.1371/journal.pbio.1002528

Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005). The Link between Social Cognition and Self-referential Thought in the Medial Prefrontal Cortex. *Journal of Cognitive Neuroscience*, *17*(8), 1306–1315. https://doi.org/10.1162/0898929055002418

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Intell. Neuroscience*, *2011*, 1:1–1:9. https://doi.org/10.1155/2011/156869

Perry, D., Hendler, T., & Shamay-Tsoory, S. G. (2011). Projecting memories: The role of the hippocampus in emotional mentalizing. *NeuroImage*, *54*(2), 1669–1676. https://doi.org/10.1016/j.neuroimage.2010.08.057

Rabin, J. S., Braverman, A., Gilboa, A., Stuss, D. T., & Rosenbaum, R. S. (2012). Theory of mind development can withstand compromised episodic memory development. *Neuropsychologia*, *50*(14), 3781–3785. https://doi.org/10.1016/j.neuropsychologia.2012.10.016

Rabin, J. S., Gilboa, A., Stuss, D. T., Mar, R. A., & Rosenbaum, R. S. (2009). Common and Unique Neural Correlates of Autobiographical Memory and Theory of Mind. *Journal of Cognitive Neuroscience*, *22*(6), 1095–1111. https://doi.org/10.1162/jocn.2009.21344

Reniers, R. L. E. P., Corcoran, R., Drake, R., Shryane, N. M., & Völlm, B. A. (2011). The QCAE: A Questionnaire of Cognitive and Affective Empathy. *Journal of Personality Assessment*, *93*(1), 84–95. https://doi.org/10.1080/00223891.2010.528484

Rosenbaum, R. S., Stuss, D. T., Levine, B., & Tulving, E. (2007). Theory of Mind Is Independent of Episodic Memory. *Science*, *318*(5854), 1257–1257. https://doi.org/10.1126/science.1148763

Rubin, D. C., Schrauf, R. W., & Greenberg, D. L. (2003). Belief and recollection of autobiographical memories. *Memory & Cognition*, *31*(6), 887–901. https://doi.org/10.3758/BF03196443

Samson, D., Apperly, I. A., Chiavarino, C., & Humphreys, G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nature Neuroscience*, *7*(5), 499. https://doi.org/10.1038/nn1223

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*, *42*, 9–34. https://doi.org/10.1016/j.neubiorev.2014.01.009

Sessa, P., & Meconi, F. (2015). Perceived trustworthiness shapes neural empathic responses toward others' pain. *Neuropsychologia*, *79*, 97–105. https://doi.org/10.1016/j.neuropsychologia.2015.10.028

Sessa, P., Meconi, F., Castelli, L., & Dell'Acqua, R. (2014). Taking one's time in feeling other-race pain: An event-related potential investigation on the time-course of cross-racial empathy. *Social Cognitive and Affective Neuroscience*, *9*(4), 454–463. https://doi.org/10.1093/scan/nst003

Sessa, P., Meconi, F., & Han, S. (2014). Double dissociation of neural responses supporting perceptual and cognitive components of social cognition: Evidence from processing of others' pain. *Scientific Reports*, *4*, 7424. https://doi.org/10.1038/srep07424

Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain*, *132*(3), 617–627. https://doi.org/10.1093/brain/awn279

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. https://doi.org/10.1016/j.neuroimage.2012.04.028

Spreng, R. N., & Grady, C. L. (2009). Patterns of Brain Activity Supporting Autobiographical Memory, Prospection, and Theory of Mind, and Their Relationship to the Default Mode Network. *Journal of Cognitive Neuroscience*, *22*(6), 1112–1123. https://doi.org/10.1162/jocn.2009.21282

Spreng, R. N., Mar, A. R., & Kim, A. S. N. (2009). The Common Neural Basis of Autobiographical Memory, Prospection, Navigation, Theory of Mind, and the Default Mode: A Quantitative Meta-analysis. *Journal of Cognitive Neuroscience*, *21*(3), 489–510. https://doi.org/10.1162/jocn.2008.21029

Spreng, R. N., Mar, R. A., & Kim, A. S. N. (2008). The Common Neural Basis of Autobiographical Memory, Prospection, Navigation, Theory of Mind, and the Default Mode: A Quantitative Meta-analysis. *Journal of Cognitive Neuroscience*, *21*(3), 489–510. https://doi.org/10.1162/jocn.2008.21029

Staniloiu, A., Borsutzky, S., Woermann, F., & Markowitsch, H. J. (2013). Social cognition in a case of amnesia with neurodevelopmental mechanisms. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00342

Staresina, B. P., Henson, R. N. A., Kriegeskorte, N., & Alink, A. (2012). Episodic Reinstatement in the Medial Temporal Lobe. *Journal of Neuroscience*, *32*(50), 18150–18156. https://doi.org/10.1523/JNEUROSCI.4156-12.2012

Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): Random permutations and cluster size control. *NeuroImage*, *65*, 69–82. https://doi.org/10.1016/j.neuroimage.2012.09.063

Vaes, J., Meconi, F., Sessa, P., & Olechowski, M. (2016). Minimal humanity cues induce neural empathic reactions towards non-human entities. *Neuropsychologia*, *89*, 132–140. https://doi.org/10.1016/j.neuropsychologia.2016.06.004

van Elk, M., van Schie, H. T., Hunnius, S., Vesper, C., & Bekkering, H. (2008). You'll never crawl alone: Neurophysiological evidence for experience-dependent motor resonance in infancy. *NeuroImage*, *43*(4), 808–814. https://doi.org/10.1016/j.neuroimage.2008.07.057

van Veen, B. D., van Drongelen, W., Yuchtman, M., & Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering*, *44*(9), 867–880. https://doi.org/10.1109/10.623056

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684. https://doi.org/10.3758/BRM.42.3.671

Xu, X., Zuo, X., Wang, X., & Han, S. (2009). Do You Feel My Pain? Racial Group Membership Modulates Empathic Neural Responses. *Journal of Neuroscience*, *29*(26), 8525–8529. https://doi.org/10.1523/JNEUROSCI.2418-09.2009

Zaki, J., & Ochsner, K. N. (2012). The neuroscience of empathy: Progress, pitfalls and promise. *Nature Neuroscience*, *15*(5), 675. https://doi.org/10.1038/nn.3085

Zeidman, P., & Maguire, E. A. (2016). Anterior hippocampus: The anatomy of perception, imagination and episodic memory. *Nature Reviews Neuroscience*, *17*(3), 173–182. https://doi.org/10.1038/nrn.2015.24

Authors contribution.

F.M. formulated the research question, collected and analysed all the data, manually drew the ROIs around the hippocampus and wrote the manuscript. F.M., S.H. and I.A.A. designed the studies. S.H. and I.A.A. supervised the analysis and substantially contributed to the writing of the manuscript. C.F. helped with the analysis of the fMRI data. B.S. supervised the fMRI analysis and the drawing of the ROIs. J.L.D. and S.M. helped with the classifier and bootstrapping analysis. All the authors gave important feedback and comments to the manuscript.

Acknowledgements.

This project has received funding from the European Union's Horizon 2020 research and innovation programme, MSCA-IF-2015, under grant agreement Nº702530, awarded to F.M. F.M. is also supported by grant from the Economic and Social Research Council (ESRC grant agreement NºES/S001964/1). S.H. is supported by grants from the European Research Council (grant agreement Nº647954), the Economic and Social Research Council (ESRC grant agreement NºES/R010072/1), and the Wolfson Society and Royal Society.

The Authors thank all the participants who took part to the studies and to Ross Wilson and Nina Salman who helped with the data collection of the fMRI data.

Conflict of interest.

None to declare.