Healthcare experience disrupts representational similarity between one's and others' pain in anterior insula

Corrado Corradi-Dell'Acqua[a,b,c,*], Christoph Hofstetter[c], Gil Sharvit[c,d,e], Olivier Hugli[f] and Patrik Vuilleumier[b,c,d]

[a] Theory of Pain Laboratory, Department of Psychology, Faculty of Psychology and Educational Sciences (FPSE), University of Geneva, Geneva, Switzerland.

[b] Geneva Neuroscience Center, University of Geneva, Geneva, Switzerland.

[c] Laboratory of Behavioural Neurology and Imaging of Cognition, Department of Neuroscience, University Medical Center, University of Geneva, Geneva, Switzerland.

[d] Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland.

[e] Balgrist University Hospital and University of Zurich, Zurich, Switzerland.

[f] Emergency Department, University Hospital of Lausanne (UHL), Lausanne, Switzerland

***Corresponding author.** Corrado Corradi-Dell'Acqua, University of Geneva – Campus Biotech, Ch. des Mines 9, CH-1211, Geneva, Switzerland. Tel: +41223790958. E-mail: corrado.corradi@unige.ch

Manuscript information.

Title: 13 words
Abstract: 150 words
Text Length (excluding methods): 4218 words
References: 70
Methods: 3297 words
Figures: 5 (all in color)

Abstract

Medical students and professional healthcare providers often underestimate patients' pain, an effect associated with decreased neural response of the anterior insula to pain information. However, the functional significance of these neural modulations is still debated. We recruited university medical students and emergency caregivers to test the role of healthcare experience on the behavioral/neural reactivity to other's pain, emotions, and beliefs. We confirmed that healthcare experience decreased the sensitivity to others' suffering, as measured by subjective ratings and insular response. This effect was independent from stimulus modality (pictures, texts), but specific for pain, as it did not generalize to emotions or beliefs. Critically, multivariate pattern analysis revealed that healthcare experience impacted specifically a component of the neural representation of others' pain shared with that of first-hand nociception. This suggests a reduced likelihood of appraising others' sufferance as one's own, and might offer a mechanistic explanation for pain underestimation in clinical settings.

Introduction

Unrelieved pain is a major problem worldwide, resulting in human suffering and economic costs. In medical practice, pain is difficult to quantify objectively, and it is often assessed indirectly through clinical examination and patients' self-reports. It is therefore not surprising that healthcare providers underestimate[1,2] (and undertreat[3]) patients' pain, a phenomenon that emerges during university education[4,5], and becomes more pronounced with longer experience in the field[6,7].

Neuroscience research has investigated extensively the cerebral mechanisms that underlie the appraisal of other's pain, and has begun to unveil how they may be modified by healthcare training. Imaging studies have implicated a widespread brain network, centered around the insula and cingulate cortex, in the processing and empathizing with other people's pain (as conveyed by faces, injuries, text, etc.)[8–14]. Remarkably, these activity patterns are highly similar to those measured when pain is experienced directly by oneself[15–22], and they are attenuated by those same analgesic procedures that regulate first-hand nociception, such as placebo or hypnosis[17,23,24]. These results suggest that others' pain is at least partly processed in an embodied (or empathetic) fashion, that is, by simulating its somatic and affective properties on one's own body[25,26]. Critically, medical practitioners and students exhibit decreased activity in these regions to the sight of injuries and painful expressions[4,27,28]. Indeed, continuous interaction with severe conditions and injuries might have made healthcare providers progressively desensitized towards the sight of pain[27,29], possibly due to regulatory processes protecting them from the psychological costs of repeated exposure to sufferance[30,31].

Recently, however, scholars have underscored that the neural response in the insula and cingulate cortex is not specific for pain, but responds also to a wide range of painless conditions, including various emotional events[15,16] and non-affective but intense visual/auditory stimulations[32]. This lack of specificity challengesd embodied interpretations of social cognition, as it is difficult to disentangle components of neural activity that underlie specific affective states from those coding for supra-ordinal dimensions, such as the unpleasantness, intensity, or salience of an event[16,24,33]. This problem also concerns studies investigating healthcare training, which could affect the neural response of pain-specific processes shared between oneself and others (*seeing injured patients hurts me*), or broader mechanisms that signal any emotionally or attentionally salient stimulus (*seeing injured patients captures my attention*).

Here, we combined the data from independent functional magnetic imaging (fMRI) experiments[15,34,35], in which female volunteers in their first (*Med1*: N = 15, age range 18-22 years, mean = 19.8) and fourth year of medical school (*Med4*: N = 14, age 22-29, mean = 24.14), control students from other faculties (*Controls*: N = 14, age 19-31 years, mean = 23.42) and professional nurses working in an hospital-based emergency department (*Nurses*: N = 21 age = 28-61, mean = 37.33) underwent the same set of validated experimental paradigms. These were an "Handedness" task where subjects were exposed to pictures of people's injuries, as well as painless neutral and negative control images[15,18,35]; a "Cognitive and Affective Theory of Mind" task, where participants read short narratives, and subsequently were asked to infer the protagonist's thoughts/beliefs, emotions and pain[34,36]; a "Pain Localizer", where individuals received noxious or non-noxious thermal stimulations on their own dominant hand. This allowed us to assess whether scholarly and professional healthcare experience affects the

neural responses to others' pain[4,27,28] and, most critically, whether such influence operates on state-specific representations shared with first-hand nociception[15–22] or generalizes to other emotional or attentionally salient events.

Results

Handedness Task

Participants underwent the "Empathy for Pain" task[15,18,35], in which individuals saw pictures of hands in Painful *(PF)* or Negative Painless *(PL)* situations, as well as neutral control images *(cPF, cPL)* matched with the previous two categories for visual features, but purged from any emotionally salient characteristics (see Figure 1A). Participants were required to assess whether the limb depicted was right/left, thus ensuring sustained attention to the stimuli without any explicit task demands about their aversiveness[15,18,35,37]. Following the scanning sessions, participants were also asked to rate each image across several dimensions (Arousal, Valence, Familiarity, etc.)[15,35]. Within this framework, participants' responses during these four conditions were modelled as a function of two orthogonal within-subject factors STIMULI (*Painful*, *Painless*) and EMOTIONAL AROUSAL (*Negative*, *Neutral*), and the between-subject factor GROUP (*Controls, Med1, Med4, Nurses*). Participants' AGE was always included as nuisance covariate, to account for potential confounds between biological and educational/professional seniority.

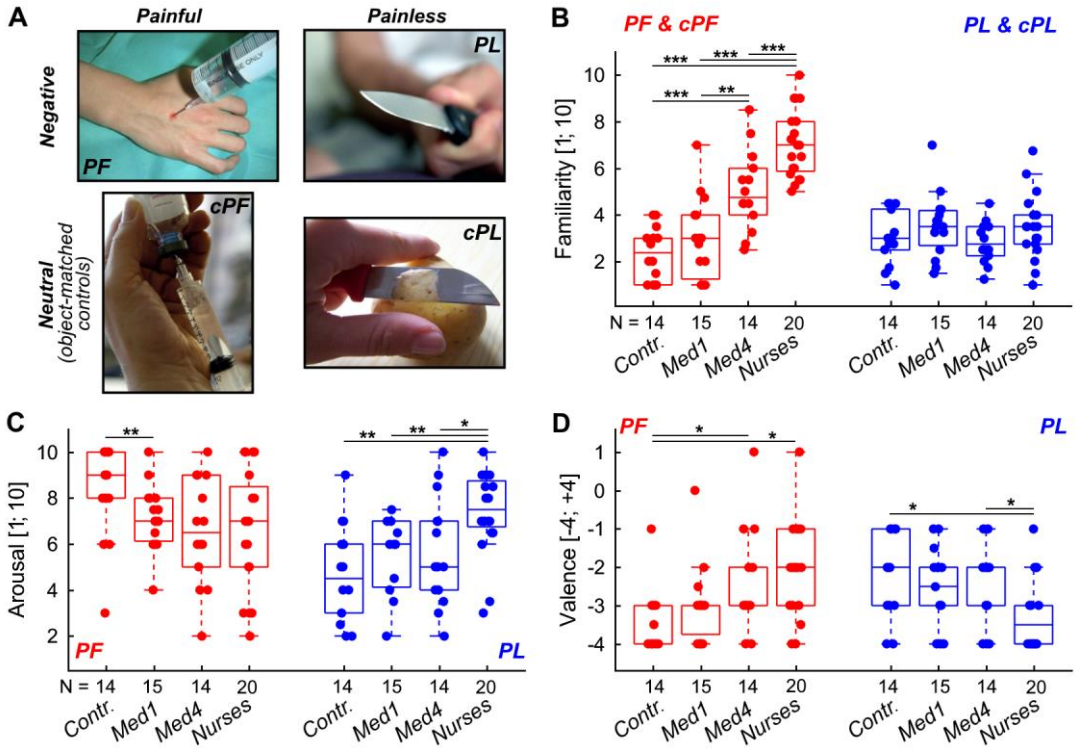

Figure 1. *Handedness task. (**A**) Example for each stimulus category. PF: Painful; PL: Negative Painless; cPF & cPL: respective neutral control conditions. (**B-D**) Boxplots displaying median post-scanning ratings of Familiarity, Arousal, and Valence for each group. Familiarity and Arousal values range from 1 (not familiar/arousing at all) to 10 (extremely familiar/arousing), whereas Valence ratings range from -4 (extremely negative) to +4 (extremely positive). For each boxplot, the horizontal line represents the median value of the distribution, the box edges refer to the inter-quartile range, and the whiskers represent the data range within 1.5 of the inter-quartile range. Individual data-points are also displayed. Red boxes and dots refer to PF/cPF stimuli, whereas blue boxes and dots refer to NPL/cNPL stimuli. Contr.: Controls; Med1 & Med4: university students enrolled at the first/fourth year of medicine; Nurses: emergency nurses. N = sample size. "***", "**", and "*" refer to significant group differences as tested through independent sample t-test at p < 0.001, p < 0.01 and p < 0.05 respectively.*

Subjective Ratings

Supplementary Tables S1-2 provide a full description of the behavioral findings. In particular the analysis of Familiarity ratings revealed a significant STIMULI*GROUP interaction ($F_{(3,55)}$ = 21.49, p < 0.001), reflecting that *PF* and *cPF* stimuli (Figure 1B, red dots) evoked progressively higher ratings from the group of *Controls* (the least familiar group) through to *Med1*, *Med4,* and

Nurses (the most familiar group). Instead, familiarity ratings for *PL* and *cPL* stimuli (blue dots) were similar and low across the four groups. Most notably, Arousal and Valence ratings were associated with a significant three-way interaction of EMOTIONAL AROUSAL*STIMULI*GROUP ($F_{(3,55)} \geq 4.40$, $p \leq 0.008$). Both the Arousal and Valence associated with *PF* stimuli (Figure 1C-D, red dots) became progressively less intense and closer to neutral values from *Controls* (the most emotionally-impacted group) through to *Med1*, *Med4,* and *Nurses* (the least emotionally-impacted group). This was not the case for *PL* stimuli (blue dots), which instead were rated as more arousing & negative across medical training and experience levels. Hence, whereas *Controls* appeared more emotionally-impacted by *PF* than *PL* images, Nurses displayed the opposite trend.

Overall, these behavioral measures confirm that healthcare training leads to higher familiarity for the sight of pain (*PF*) and for neutral stimuli depicting medical devices/contexts (*cPF*; see Figure 1A). This effect was paralleled by decreased emotional susceptibility to the sight of injuries (*PF*), but a stronger sensitivity towards pictures of emotionally negative but non-painful scenarios (*PL*).

Neural responses

Supplementary Tables S3-5 provide full details about the brain regions recruited during the Handedness task. As expected, a distributed network including the middle and anterior insula (AI), the middle cingulate cortex (MCC), and the supramarginal/postcentral gyri (SMG/PCG), showed increased activity to pictures of hands in pain (contrast *PF* – *cPF*) regardless of participants' healthcare training[8–14] (Figure 2A, red blobs). Instead, arousing but painless images

(*PL – cPL*) activated only AI together with medial cortices around the supplementary motor area (SMA), dorsal to the cingulate cortex (Figure 2A, blue blobs).

We then investigated how these activations were influenced by healthcare experience. In line with behavioral results, we found that the right AI, the dorsolateral prefrontal cortex (DLPFC) and the periaqueductal grey (PAG) exhibited a three-way interaction EMOTIONAL AROUSAL*STIMULI*GROUP. Figure 2B displays the parameter estimates extracted from right AI and DLPFC, revealing that whereas this region response to painful images (*PF – cPF*, red dots) decreased linearly from *Controls* (the most sensitive group) to *Med1*, *Med4,* and *Nurses* (the least sensitive group). Importantly, such decrease was not found for the response to negative painless images (*PL – cPL*, blue dots), which, in some instances showed even an opposite trend. Overall, our data converge with previous findings that scholarly and professional healthcare experience reduces the reactivity of key brain regions implicated in the processing of others' pain, such as AI[4,27,28]. Critically, this modulation does not reflect a more global hypo-reactivity, as these same regions exhibit stable (if not enhanced) response to other categories of aversive (painless) pictures.

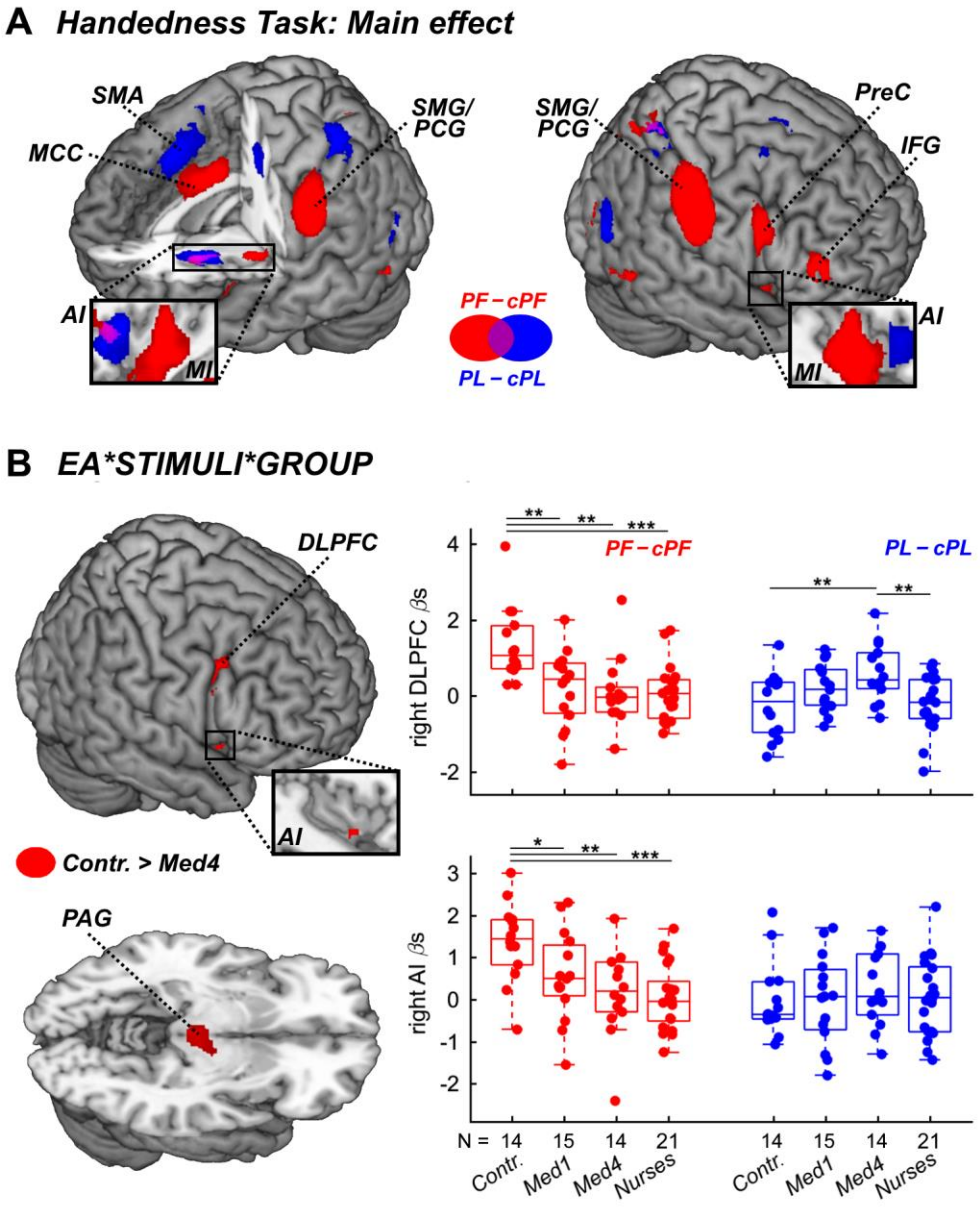

Figure 2. *Handedness Task: surface rendering of brain regions associated with (**A**) the main effect of PF – cPF (red blobs) and PL – cPL (blue blobs) across all groups, and (**B**) three-way interactions between Emotional Arousal (EA), Stimuli, and Group (where "Group" refers to the difference between Control and Med4 students). All activated regions are displayed with FWE cluster correction for multiple comparisons or small volume correction in a mask of interest (see methods). Detailed coordinates are listed in Supplementary Tables S3-5. Activity parameters extracted from the highlighted region are plotted according to groups. Individual data-points are also displayed. Red boxes/dots refer to PF – cPF activity, whereas blue boxes/dots refer to NPL – cNPL activity. SMG: Supramarginal Gyrus; PCG: Postcentral Gyrus; SMA: Supplementary Motor Area; MCC: Middle Cingulate Cortex; MI & AI: Middle & Anterior Insula; PreC: Precentral Gyrus; IFG: Inferior Frontal Gyrus; DLPFC: Dorsolateral Prefrontal Cortex; PAG: Periaqueductal Grey. Contr.: Controls; Med1 & Med4: university students enrolled at the first/fourth year of*

*medicine; Nurses: emergency nurses. N = sample size. "***", "**", and "*" refer to significant group differences (tested through independent sample t-test) at $p < 0.001$, $p < 0.01$ and $p < 0.05$ respectively.*

Vicarious Pain Signatures

To further characterize these effects of healthcare training on neural responses to observed pain, we applied a model-based approach and fed our dataset to well-established predictive neural models of vicarious pain from brain activity[21,38]. To maximize the comparability with our "Handedness" task, we considered two models derived from fMRI responses to the sight of injured limbs: *Zhou-NS*$_{2020}$[21] and *Krishnan*$_{2016}$[38] (see Methods). Both models aim at predicting the same underlying construct of pain, but do so by relying on the activity of different brain structures: whereas *Zhou-NS*$_{2020}$ relies strongly (but not exclusively) on the middle-anterior insula[21] (Figure 3A), *Krishnan*$_{2016}$ is grounded on a more widespread network involving occipital, parietal, medio-prefrontal, and subcortical structures[38] (Figure 3B). When applying these models to our data from the "Handedness" task, we found that both predicted a significant vicarious pain response to the sight of injured hands in Controls (*PF vs. cPF;* $t_{(13)} \geq 2.01$, p (1-tailed) ≤ 0.033), demonstrating a reasonable generalizability to our dataset (at least for individuals with no healthcare training).

We then assessed the degree to which the sensitivity of these models was affected by healthcare training. Supplementary Tables S6-7 describe the full results across all paradigms. In particular, the output of *Zhou-NS*$_{2020}$ was associated with a significant three-way interaction EMOTIONAL AROUSAL*STIMULI*GROUP ($F_{(3,56)} = 3.58$, $p = 0.019$), similar to the previous behavioral and neuroimaging results. Figure 3A (red boxplots) displays the output from this model, which became progressively less sensitive to injured hands (*PF – cPF*) when moving from *Controls*, to *Med1*, *Med4,* and *Nurses*. In sharp contrast, the group factor did not influence the

sensitivity of this model to the sight of negative painless images (*PL – cPL*; Figure 3A, blue boxplots). On the other hand, the *Krishnan*$_{2016}$ model revealed no significant interaction effect ($F_{(3,56)}$ = 0.27, p = 0.848), and could reliably predict responses to the sight of injured limbs in each of the four groups ($t \geq 2.21$, $p \leq 0.044$), with no significant difference among them (Figure 3B).

Finally, as the participants from both experiments underwent a brief session in which they received painful hot or painless warm stimulations on their own hand ("Pain Localizer"; see Methods), we also assessed how well these models could predict first-hand pain experience. Our data replicate previous studies, showing that *Zhou-NS*$_{2020}$ predicted self-pain in all groups ($t \geq 2.83$, $p \leq 0.010$; Figure 3A, black/grey boxplots)[21], whereas *Krishnan*$_{2016}$ did not ($|t| \leq 1.44$, $p \geq 0.173$), suggesting that it might be more sensitive to information independent from first-hand nociception (Figure 3B)[38]. Taken together, these data highlight that information about vicarious pain is encoded in a widespread network and different neural models, relying on different brain structures for their predictions. Importantly, our findings suggest that healthcare training impacts exclusively a model that highlights activity of the insular cortex and encodes information shared with first-hand nociception.

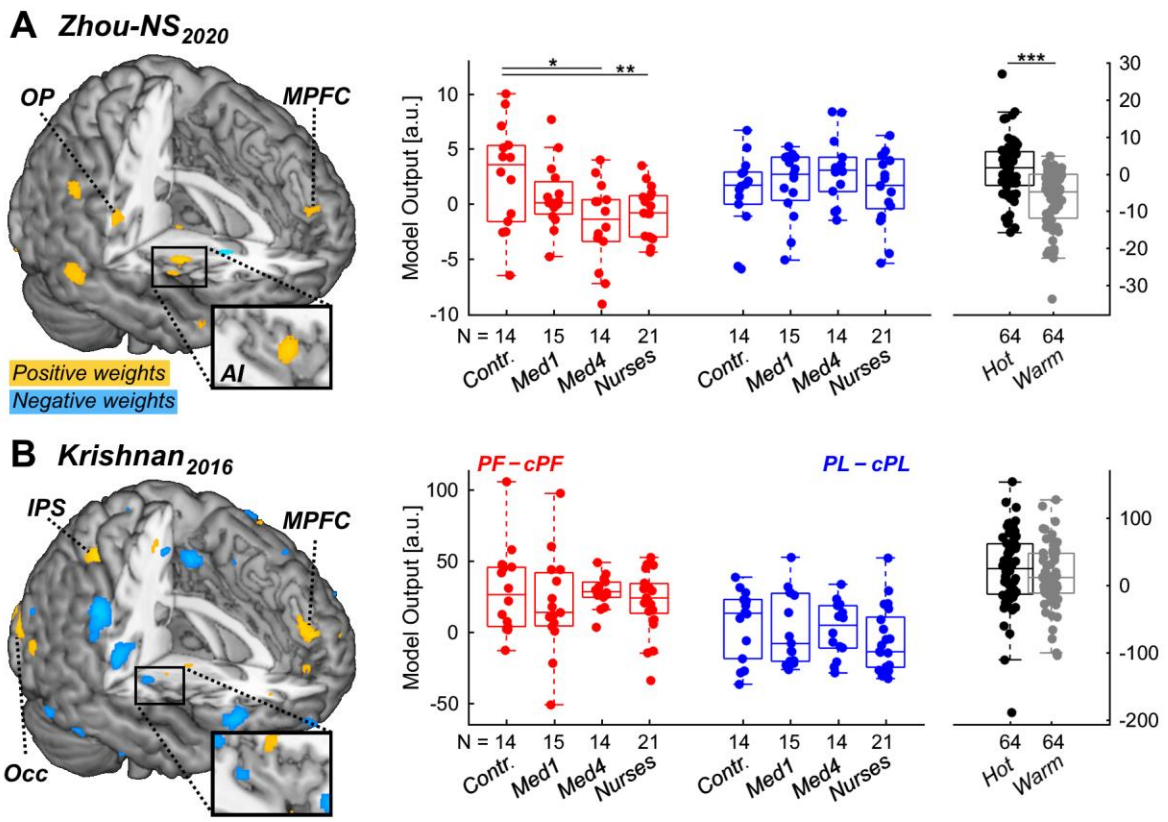

Figure 3. *Handedness Task: Vicarious Pain Signatures. Surface brain renderings displaying regions that contributed positively (yellow blobs) and negatively (cyan blobs) to the prediction of vicarious pain from the sight of injured limbs, based on the models of (**A**) Zhou-NS$_{2020}$[21] and (**B**) Krishanan$_{2016}$[38]. The contribution maps are displayed under a false discovery rate correction of q < 0.05, as provided for both models [21,38]. In each case, the estimated vicarious pain (in arbitrary units [a.u.]) from our data are also displayed. Red boxes/dots refer to the differential PF – cPF output, whereas blue boxes/dots refer to differential PL – cPL output. Black and grey boxes/dots refer to model estimates of the "Pain Localizer", in which all participants (regardless of the group/cohort) received hot (painful) or warm (painless) stimulations on their own hand. MPFC: Medial Prefrontal Cortex; AI: Anterior Insula; OP: Parietal Operculum; IPS: Intraparietal Sulcus; Occ: Occipital Cortex. N = sample size. "***", "**" and "*" refer to significant differences through t-tests at p < 0.001, p < 0.01, and p < 0.05 respectively.*

Cognitive and Affective Theory of Mind Task

The participants also underwent a previously validated "Cognitive and Affective Theory of Mind" task[34,36]. Unlike the Handedness paradigm, which measures involuntary visceral responses evoked by the sight of negative pictures, this task taps into more deliberate

inferential mechanisms about others' state in specific situations[39,40]. In particular, participants read brief story (*scenario epoch*) followed by a question probing for the protagonist's beliefs, emotions or pain (*judgment epoch*). As high-level control, there were stories with no human protagonist (photos) to account for attentional, memory and linguistic processes of no interest at play in the task.

Supplementary Tables S8-S9 provide a full description of the behavioral data in this task. These results showed no effect of healthcare experience on the explicit appraisal of beliefs, emotions, or pain as inferred from verbal scenarios. For fMRI data, we analysed brain activity during the scenarios epochs separately from that during the judgment[34,41] (see Methods). Whereas judgments require participants to focus on one specific state category, the scenarios could elicit a mix of spontaneous appraisals about different cognitive and affective states. Hence, we took advantage of normative data derived from an independent population who quantified each narrative in terms of how it triggered inferences about beliefs, emotions, or pain[34]. We applied these scores to the present study in order to identify neural structures, whose activity correlated with each mental state category[34]. Full results are described in Supplementary Tables S10-S11, and revealed that text scenarios evoking inferences about pain differentially engaged the middle cingulate cortex (MCC), supramarginal gyrus (SMG), inferior frontal gyrus (IFG), and posterior insula[34,36,42–45] (Figure 4A, red blobs). On the other hand, scenarios highlighting beliefs engaged a distinctive network implicating the bilateral temporo-parietal junction (TPJ), precuneus (PC), and dorsomedial prefrontal cortex (DMPFC, green blobs), consistent with previous meta-analyses on theory-of-mind and mentalizing[13,46–49]. Finally, scenarios referring to emotions elicited activity in the ventral medial prefrontal cortex

and middle temporal gyrus. The same networks evoked by belief and emotion scenarios were also recruited during the judgment epochs, when answering questions about beliefs/emotions as opposed to control questions about outdated pictures/photos. Instead no region showed higher activity when answering questions about pain, as opposed to the photo condition.

We then assessed the role played by healthcare training on these effects by testing for group differences in the recruitment of each of these networks. Results showed that scenarios referring to pain elicited stronger neural responses in the right AI in controls as opposed to nurses. Figure 4B displays the activity parameter estimates from this region who exhibited a progressive decrease of pain-related activity across levels of healthcare training. Interestingly, and similarly to the "Handedness" task, this decrease did not extend to scenarios referring to painless emotions which, instead, showed an opposite trend (Figure 4B, blue dots). Finally, for beliefs, we observed significant group differences during the judgment epochs (but not during scenario epochs), with cortical regions in left TPJ, left inferior temporal sulcus, right DLPFC, and supplementary motor area (SMA) showing stronger activation in young medical students (*Med1*) than other groups (Figure 4C). Hence, neural responses associated with theory-of-mind abilities did not seem to be linearly modulated by healthcare training, but rather appear to peak in the beginning of medical education and then decrease in the subsequent years. Interestingly, Figure 4C displays the parameter extracted in left TPJ from the judgment epochs, revealing that this region is more strongly recruited in *Med1*, not only in the assessment of beliefs (green dots), but also in other states like pain (red dots).

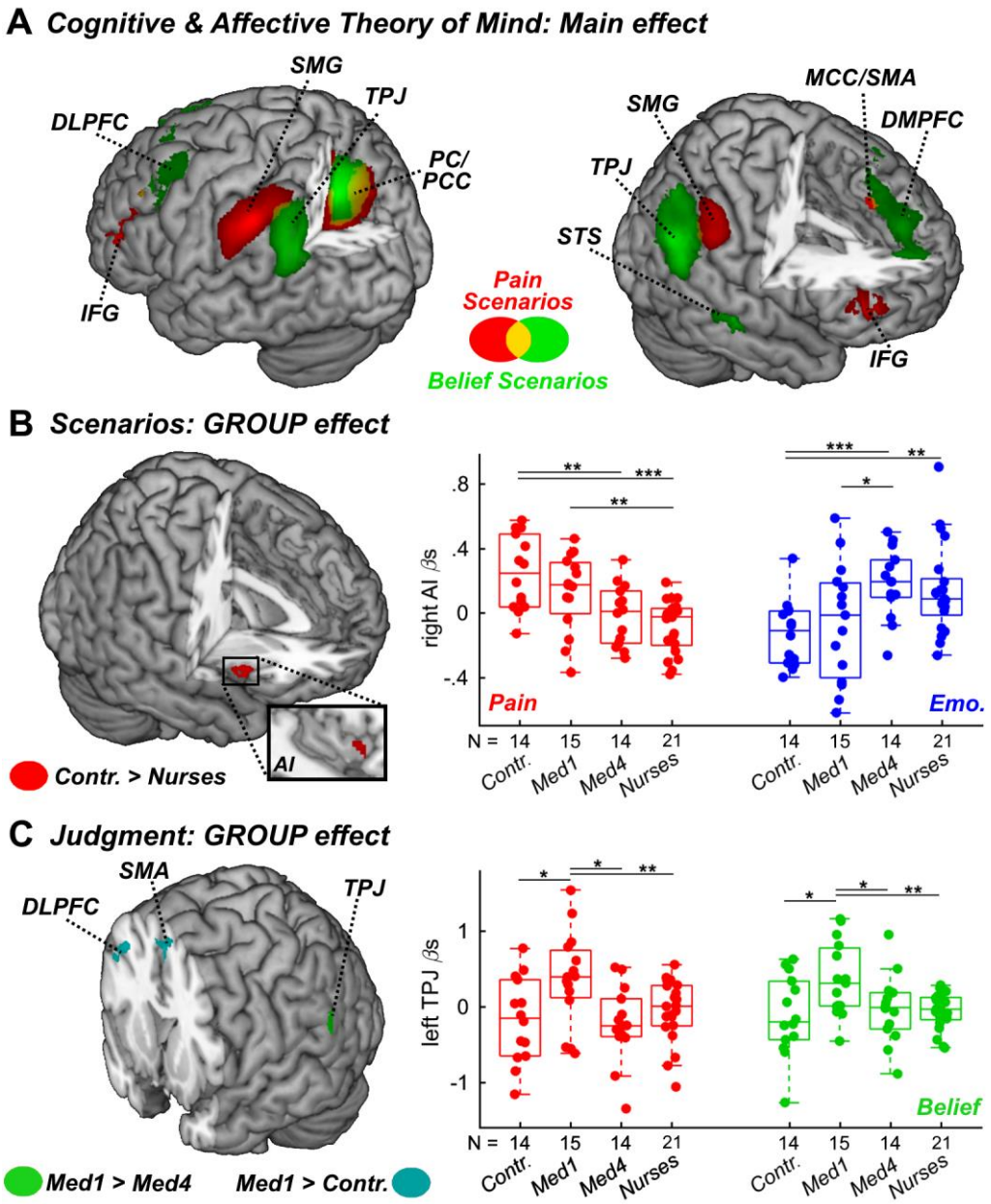

Figure 4. *Cognitive and Affective Theory of Mind Task: surface rendering of brain regions associated with (**A**) reading text-based scenarios about individual Pain (red blobs) and Beliefs (green blobs) in all groups, (**B**) group differences in Pain scenarios (red blobs refer to the contrast Control > Nurses), and (**C**) group differences in Belief judgments (green blobs refer to the contrast Med1 > Med4, whereas turquoise refer to the contrast Med1 > Control). All activated regions are displayed with FWE cluster correction for multiple comparisons or small volume correction in a mask of interest (see methods). The parameters extracted from the highlighted region are plotted across groups (with boxplots). Individual data-points are also displayed. Red boxes/lines/dots refer to pain-evoked activity, blue boxes/lines/dots refer to emotion-related activity, whereas green boxes/lines/dots refer to belief-related activity associated with scenarios and judgment epochs respectively. Detailed coordinates are listed in*

*Supplementary Tables S10-S11. DLPFC & DMPFC: Dorsolateral and Dorsomedial Prefrontal Cortex. IFG: Inferior Frontal Gyrus; SMG: Supramarginal Gyrus; TPJ: Temporo-Parietal Junction; PC: Precuenus; PCC & MCC: Posterior & Middle Cingulate Cortex; STS & ITS: Superior & Inferior Temporal Sulcus; AI: Anterior Insula; Contr.: Controls; Med1 & Med4: university students enrolled at the first/fourth year of medicine; Nurses: emergency nurses. N = sample size. "***", "**", and "*" refer to significant group differences from independent sample t-tests at $p < 0.001$, $p < 0.01$ and $p < 0.05$ respectively.*

Representational Similarity of Pain

Our participants were exposed to pain across different paradigms, some involving hot temperatures delivered to their own body ("Pain Localizer"), and others involving a representation of pain in others, either through images ("Handedness" task) or text ("Cognitive & Affective Theory of Mind" task). To seek for regions disclosing a common representation of pain across these different paradigms, we performed a multi-voxel pattern analysis of neural response allowing us to compare these conditions and test whether any representational similarity between them changed as function of healthcare experience. More specifically, we tested for regions exhibiting the highest pattern similarity whenever two painful conditions were paired together (Hot temperatures; *PF* images of wounded hands; Pain Scenarios & Judgments; Figure 5A, yellow blocks), as opposed to when they were paired with another "non-painful" event (*PL* images; Beliefs & Emotion Scenarios/Judgments; Figure 6A, blue blocks). We then estimated pain-specific information by contrasting pattern similarity in "within-pain" vs. "between-domain" pairings (Figure 5A), *via* a whole-brain searchlight analysis. This analysis highlighted a widespread network involving AI, MCC, and DLPFC in which pain-specific information was impacted by healthcare experience (see Supplementary Table S12 for full details). More specifically, comparison of *Controls* to *Med1* showed stronger within-pain

pattern similarity in AI, DLPFC and MCC (Figure 5B, orange blobs), whereas comparing *Controls* to *Nurses* indicated higher similarity in the AI and PI selectively (red blobs).

Such changes in pain-specific patterns in insula are further depicted in Figure 5C, showing activity parameters extracted from the right AI (identified in the contrast *Contr.* > *Nurses*) and summarized in a matrix form for pairwise comparisons between groups across all conditions. Detailed inspection of specific pairings underlying group differences in AI indicates that controls (unlike nurses) showed the strongest similarity between painful scenarios and both self-experienced pain (hot temperatures) and observed pain in pictures of injured hands (yellow blocks). Instead, medical students and nurses (unlike controls) showed stronger similarity of pain conditions with non-painful events from the same paradigm (e.g., *PF* & *PL* images; Pain, Beliefs & Emotion Judgments; blue blocks). Overall, these data suggest that healthcare experience modified the activity patterns evoked by others' pain in AI, making them less similar to those evoked by self-nociception, and more similar to those evoked by others' painless (mental) states.

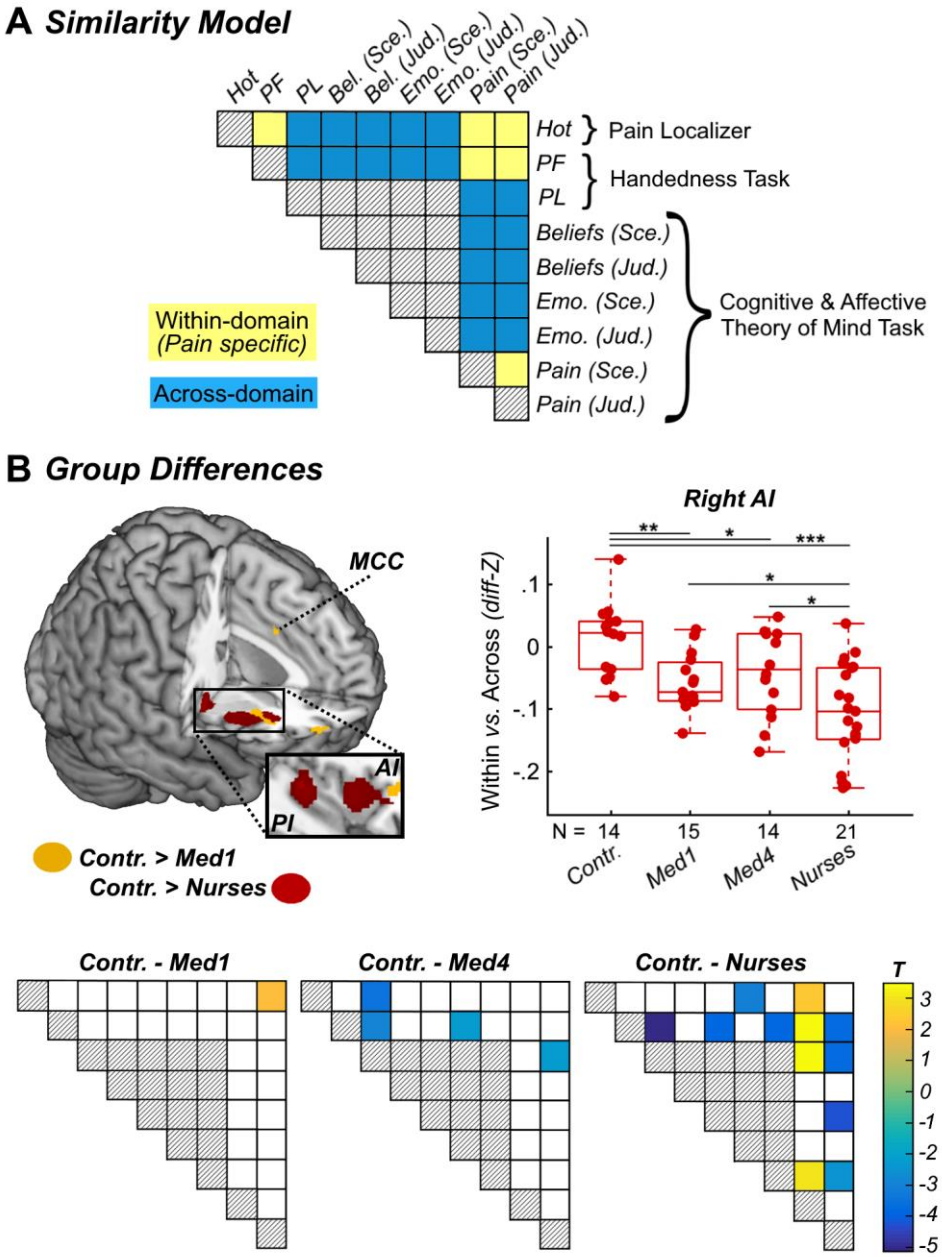

Figure 5. *Representational Similarity of Pain. (**A**) Schematic matrix of a representation model testing for pain-specific activity patterns across different tasks in our study. Within the matrix, row and column labels refer to conditions of interest: Hot temperatures from the "Pain Localizer", Painful (PF) and Negative Painless (PL) images from the "Handedness" task, and Beliefs, Emotions, and Pain Scenarios & Judgments from the "Cognitive and Affective Theory of Mind" task. Each matrix cell indicates the putative representation similarity between two conditions: yellow cells concern pairs of pain conditions (within-pain), blue cells concern pairs of a painful and painless conditions (between-domains), and black-striped cells refer to pairings of no interest. (**B**) Surface brain rendering of regions displaying significant group differences in their sensitivity to pain-specific information (within-pain vs. across domain). Detailed*

*coordinates are listed in Supplementary Table S12. Activity parameters extracted from the highlighted region are plotted across groups (with boxplots and individual data), and in matrix form. Highlighted orange/yellow matrix cells refer to pairings showing higher similarity in Controls than in healthcare providers. Turquoise/blue cells refer to pairings showing higher similarity in healthcare providers than in Controls. PI & AI: Posterior & Anterior Insula; MCC: Middle Cingulate cortex. Contr.: Controls; Med1 & Med4: university students enrolled at the first/fourth year of medicine; Nurses: emergency nurses. N = sample size. "***", "**", and "*" refer to significant group differences from independent sample t-tests at $p < 0.001$, $p < 0.01$ and $p < 0.05$ respectively.*

Discussion

We investigated the role of healthcare experience in individuals' sensitivity to others' pain, and whether this may extend to other forms of social cognition. We confirmed previous evidence that healthcare experience decreases the brain response to others' pain, including neural activity in the anterior insula (AI)[4,6,7,27,28]. We also demonstrated that such attenuation is not restricted to paradigms using pictures[4,5,27–29], but generalizes to situations where pain must be inferred from short narratives. Instead, healthcare experience did not influence in the same fashion the processing of others' painless states, such as emotions or thoughts/beliefs. Finally, multivariate pattern analyses revealed that healthcare experience affects specifically a component of the neural representation of other's pain that is common with first-hand nociceptive stimulations. Information about others' pain could still be discriminated from neural data in all participants regardless of their scholarly/professional background, when using a multivariate brain model of pain perception independent from direct somatic experience. Accordingly, representational similarity analysis revealed a reduced correlation between of others' pain-evoked patterns with self-experienced pain, but greater overlap with cognitive judgments about other painless mental states such as emotions and beliefs.

Previous studies have reported how AI and middle cingulate cortex (MCC) respond to others' suffering[8–14], as conveyed through facial expressions[4,20,21,23,24], photos/videos of injuries[15,17,18,21,27,28,35,38,50], symbolic cues[16,19,22,51], and short narratives[34,36,42–45]. Yet the functional interpretation of these neural responses remains debated. One proposal is that a large part of this activity might encode supraordinal features, such as unpleasantness, arousal, or salience[15,16,52,53], rather than pain-specific signals. Another view, based on studies carefully controlling for these supraordinal confounders, is that both self and vicarious pain are represented in the insular cortex in a pain-selective manner, perhaps parallel to other supraordinal signals[16,24,32,33,54,55]. Our results therefore converge with, but also extend, these findings by showing that healthcare experience changes AI responses to others' suffering in a pain-specific fashion. Although AI also responded to painless emotional pictures and narratives, this sensitivity was impacted by healthcare experience in different ways, sometimes exhibiting an increased (rather than decreased) response in senior students/nurses (see also emotional ratings).

Our results therefore appear to agree with the idea that healthcare experience modifies neuronal populations sensitive specifically to pain. Alternatively, it is theoretically possible that healthcare experience affects neuronal populations in AI that code for broad properties like unpleasantness or salience. In this case, however, the influence of healthcare experience would still be dependent of the nature of the state observed in others (pain vs. other affective states), and cannot be explained by a general hyporeactivity or inhibition. For instance, it is conceivable that healthcare experience could modulate the "access" to these neurons, by regulating the likelihood for a specific type of stimuli to engage this region. Indeed, seminal models suggest

that AI is part of the so-called "salience network", responding to events sufficiently relevant to capture attention and reset action goals[52,53,56]. Our data are also compatible with this interpretation, as professional exposure to injuries, cuts, or burns would make these categories less novel or unusual, and therefore less likely to trigger neurons responding to salience, at the advantage of other events outside this domain of expertise.

In our study, we also performed comprehensive multivariate pattern analyses to examine the similarity between the neural response to one's and others' pain. By employing previously established whole-brain signatures of vicarious pain, we found that healthcare experience impacted exclusively a predictive model (*Zhou-NS*$_{2020}$) based on the neural similarity with self-pain and strongly grounded on AI[21]. Another model (*Krishnan*$_{2016}$), derived from activity in networks independent from first-hand nociception[38], was not affected by healthcare experience, such that it could correctly predict responses to pain pictures in all groups to the same degree. These results were complemented by a representational similarity analysis, which revealed that healthcare experience disrupts the neural overlap between pain signals in AI and MCC across our different paradigms (hot temperatures, pain pictures, and verbal scenarios). Taken together, these results further support the role of an insular-cingulate network in the representation of others' pain through the recruitment of neuronal populations that are also involved in processing pain in one's own body[15–22]. Critically, individuals with longer medical training do not appear to process others' pain through this pathway, but rely on information encoded in other brain networks.

Representational similarity between one's and others' pain in regions such AI has been often interpreted as the neural substrate for the ability to embody others' suffering, and

possibly to empathize with the observed condition[25,26]. Within this framework, it has been argued that healthcare experience might induce a greater regulation of empathetic responses, allowing repetitive interactions with patients without contagion from their sufferance[30,31]. However, we advise caution in interpreting the modulations observed here as reflective of a change in empathy. Empathy is a multidimensional process and neuroimaging results from paradigms like ours do not systematically correlate with scores from validated questionnaires (see[12] for meta-analytic evidence). Furthermore, in contrast with the consistent neuroimaging results[4,5,27–29], behavioral research investigating how healthcare experience impacts on trait empathy has led to mixed results, with some studies suggesting that medical students and healthcare providers have lower empathy[57–60], and others reporting no change[61] or even an increase[60,62,63]. This could reflect heterogeneity in definitions and measures of empathy, which vary extensively across studies. For these reasons, we favor a more parsimonious explanation, suggesting that healthcare experience impacts negatively the embodied processes used to appraise others' pain, whereby observed somatic/emotional experiences are simulated on oneself. However, this does not imply a broader reduction in sensitivity to others' affect, in accordance with the multidimensional psychological construct of empathy captured by questionnaires and other measures.

Remarkably, we found that brain activity evoked by processing and judging beliefs was also influenced by healthcare experience. In particular, unlike AI for pain, the left TPJ (Figure 4C, green dots) showed increased activity during judgments about other's beliefs, which predominated in early medical students relative to both controls and emergency nurses. TPJ is a key component of the core theory-of-mind network, repeatedly implicated in tasks involving

the assessment of others' thoughts or beliefs[13,46–49]. Unfortunately, this neural modulation was not paralleled by documented behavioral changes, precluding any definite conclusion about whether young medical students might differ from other groups in processing the point of view of others. Interestingly, however, the same TPJ area showed enhanced activity in young medical students also for assessment of pain states (Figure 4C, red dots), possibly hinting to a more general role of the neural processes underlying theory-of-mind also when confronted with others' pain. This would be consistent with engagement of more cognitive and deliberative components of empathy and perspective taking, replacing more automatic activation of embodied and salient signals in AI. Future research will need to investigate more thoroughly the role of healthcare experience in theory-of-mind abilities.

There are few important limitations in our study that need to be underlined. First, we investigated the role of healthcare experience by aggregating data from different populations with different levels of experience. As such, our study shares the weaknesses of cross-sectional investigations[64], as the role of healthcare experience was not tested longitudinally in the same population. Second, the overall sample was sizeable (N = 64) but it was split into smaller groups (N = from 14 to 21). Hence, although our effects appear reliable across various measures, we cannot rule out that their size might be inflated by the low power. Third, as data from cohort 2 (nurses) were obtained in a larger project involving a wide range of tasks[35], each paradigm was administered in a shorter version relative to cohort 1 (students). To account for this, our analyses included only a portion of data from cohort 1, to obtain an equal amount of trials/blocks per condition. Finally, we rigorously controlled for age in all analyses, to insure that differences between cohorts were not confounded by biological seniority. Still, some

differences could not be fully equated, such as the properties of thermal stimulations, the duration of each scanning session, or the functional MRI sequence parameters. However, we believe that the impact of these confounds was negligible. Indeed, the data from nurses was generally very similar to those from senior medical students, and both groups differed greatly from controls. Hence, it is the healthcare experience, and not the protocol, that most strongly drove our effects.

Notwithstanding its limitations, our study extends previous investigations about the role of healthcare experience in social cognition in several ways. First, it shows how experience in healthcare influences negatively neural reactivity in AI to others' pain, both from visual and text information. Second, it shows how this effect is specific for pain, but does not generalize to other forms of social cognition, such as painless affect or theory-of-mind abilities. Third, it shows how the neural signature of vicarious pain modified by healthcare experience impacts a representation that is shared with first-hand nociception. Fourth, it demonstrates that, in contrast, information about vicarious pain encoded in other brain structures is unaffected by healthcare experience, such that it can be proficiently used by predictive multivariate models to detect the sight of injuries. Overall, we conclude that extensive professional healthcare experience may result in lower likelihood of processing others' sufferance as one's own, accompanied with lower neural reactivity of pain-responsive areas such as the insula, but possibly associated with more cognitive appraisals of others' mental states. These results may contribute to better understand how pain is evaluated and often underestimated in real-life clinical settings[6,7,65,66].

Methods

Participants

Our study involved 64 participants (age range: 18-61 years, mean = 27.30), obtained by combining two independent cohorts. Cohort 1 included 43 female university students from the University of Geneva. Part of these participants were recruited in previous studies[15,34], although they were tested as a unique group of students, without taking into account the faculty in which they were enrolled. Here, we included only those individuals who fell into the following three groups: students enrolled in the first year of medical school (*Med1*: N = 15, age 18-22 years, mean = 19.8); students enrolled in the fourth year of medical school (*Med4*: N = 14, age 22-29 years, mean = 24.14); and students attending other university faculties or high schools except medicine, infirmary, dentistry, or kinesi-physiotherapy (*Controls*: N = 14, age 19-31 years, mean = 23.42).

Cohort 2 included 21 female *Nurses* from the Emergency Department of the University Hospital of Lausanne (age range = 28-61, mean = 37.33; post-graduate experience = 3-33 years, mean = 11.03). These were part of a larger sample used in a previous study[35], from which we included only females to allow better compatibility with cohort 1. None of our participants had any history of neurological or psychiatric illness. Written informed consent was obtained from all subjects. The studies were approved by the ethical committee of the University Hospital of Geneva (cohort 1) and by the Ethical Cantonal Commission of Canton Vaud (cohort 2), and were conducted according to the declaration of Helsinki.

Handedness Task

Full details about the paradigm are described in previous studies[15,18,35]. Participants saw a randomized sequence of colour pictures (768x768 pixels, corresponding to 14.25°x14.25° degrees of visual angle) of human hands, organized in four categories. The "Painful" category (*PF*) was composed of 60 (cohort 1) or 30 (cohort 2) images depicting hands in painful situations, inferable by either the presence of wounds/burns on the skin and/or an external object (scalpel, syringe, etc.) acting on the skin surface. The negative "Painless" category (*PL*) was composed of 30 pictures of hands in emotionally aversive, but painless situations (hands holding knifes/guns, hands with handcuffs, etc.). For both *PF* and *NPL* stimuli, we also selected neutral control stimuli (*cPF* and *cPL*) that were matched with the previous two categories for hand laterality (right/left) orientation (angular distance from the viewer's own hand position at rest), and for visual features (presence of objects, human bodies, etc.), but purged from any emotionally-salient (painful, arousing) features. This yielded a 2 x 2 design with STIMULI (*Painful*, *Painless*) and EMOTIONAL AROUSAL (*Negative*, *Neutral*) as factors. All images were equated in luminance.

For each experimental trial, one of the 180 (Experiment 1) or 120 (Experiment 2) hand stimuli was presented for 2500 ms, followed by an inter-trial interval that ranged from 2500 to 4100 ms (mean and median = 3300 ms) with incremental steps of 320 ms. Participants were asked to perform a handedness task: if the stimulus depicted a right hand they had to press a key corresponding to right hands, whereas if the stimulus was a left hand they had to press another key corresponding to left hands. This task is known to be accomplished by mentally imagining to move one's own hand until it is aligned with the viewed hand[67], hence favouring

an embodied perspective, but it did not make any explicit demand to process the painful or emotional cues in pictures. Participants were instructed to respond as fast as possible and to ignore other image features (e.g., blades, wounds) that were irrelevant to the task. The 4 experimental conditions of this task were presented in a randomized order together with 30 null-events, in which an empty screen replaced the stimuli. All trials were presented in a unique scanning session which lasted about 21 (cohort 1) or 15 (cohort 2) minutes.

Cognitive and Affective Theory of Mind Task

Full details about the paradigm are described in previous studies[34,36]. For this task, we used 36 short French-written narratives describing a person engaged in various situations, followed by questions probing for the reader's awareness of the protagonist's beliefs, emotions, or pain in this situation. As high-level control condition, we included 12 additional stories with no human protagonist but describing physical entities with changing properties on visual maps or photographs (photos). The full list of narratives is provided in Corradi-Dell'Acqua et al.[34].

The task was organized in experimental sessions containing a random sequence of 24 narratives (6 per conditions) of about 10 minutes. In cohort 1, participants underwent two sessions, covering the full set of 48 scenarios, whereas in cohort 2, participants underwent only one session. Within each session, the scenarios were presented for 12 seconds, followed by a judgment epoch of 5 seconds. During the judgment phase, a question was presented together with two possible answers, each located on a different side of the screen. Participants made responses by pressing one of two possible keys, corresponding to the side of the answer they believed to be correct. The position of the correct response on the screen was counterbalanced across narratives. Judgments were followed by an inter-trial interval of 10 seconds.

Importantly, in the "judgment" stage, participants were asked to evaluate only one dimension (beliefs, emotions, pain), but it is likely that during the reading stage, the "scenarios" could elicit spontaneous considerations about multiple dimensions at the time (e.g., a story about someone's sadness often describes also what this person might be thinking). The likelihood of scenarios to elicit inferences about each dimension was obtained from validation data of our story database, as described in Corradi-Dell'Acqua et al.[34].

Pain Localizer

In this task, participants received either noxious or non-noxious thermal stimulations to their hand palm, delivered by using a computer controlled thermal stimulator with an MRI compatible 25x50 mm fluid cooled Peltier probe (MSA Thermotest, SOMEDIC Sales AB, Sweden). The non-noxious temperature was fixed to 36°C (cohort 1) or 38°C (cohort 2). The noxious temperature varied on a participant-by-participant basis and ranged between 41-52°C (cohort 1: average = 46.42°C, *Hot – Warm* difference = 10.89°C; cohort 2: 49.25°C, *Hot – Warm* = 11.25°C). This temperature was selected, through a dedicated thresholding procedure as described in previous studies[15,35]. In cohort 1, thermal events were delivered in two separate runs, involving the stimulation of the right or left palm respectively. In each run, we delivered ten trials, five with noxious temperature and five with a non-noxious temperature. Each trial was organized into two consecutive thermal shifts, each lasting 9 seconds (3 seconds of temperature increase, 3 seconds of plateau, and 3 seconds of temperature decrease), followed by an inter-trial interval of 18 seconds. In cohort 2, thermal stimuli were delivered in a single run, with stimulation of the dominant palm. In this run, we delivered twelve trials, six with noxious temperature and six with non-noxious temperature. Trials were organized in a unique

thermal shift lasting approximately 9 seconds (3 second of temperature increase, 2 seconds of plateau, and 3 seconds of decrease), followed by an inter-trial interval ranging between 9-14.5 sec (average 11.75). In both experiments, a visual cue (identical for noxious and non-noxious shifts) informed participants of each of these shifts.

Post-Scanning rating session

After scanning, participants were asked to rate each of the stimuli from the "Handedness" task in terms of familiarity ("*how much is the content described in this picture familiar to you?*"), emotional intensity ("*how intense is the emotion triggered by this image?*"), emotional valence (*"does this image elicit positive or negative emotions?"*), and pain ("*how intense is the pain felt by the hand depicted on this image?*"). The ratings were divided in four blocks, one for each question, during which all stimuli were rated on a Likert scale ranging from 1 to 10 (with the exception of valence for which a Likert scale from -4 to +4 was used). To avoid habituation biases due to the presentation of the same stimuli four times, the order of the blocks and the order of the stimuli within each block were randomized across participants. See previous reports for more details about the rating session[15]. Post-scanning data from one subject of cohort 2 is missing.

Scanning Procedure

Participants lay supine with their head fixated by firm foam pads. Stimuli were presented using E-Prime 2.0 (Psychology Software Tools, Inc.) on a LCD projector (CP-SX1350, Hitachi, Japan) outside the scanner bore, subtending about 14.25° (vertical) x 19° degrees of visual angle. Participants saw the monitor through a mirror mounted on the MR headcoil. Key-presses were

recorded on an MRI-compatible bimanual response button box (HH-2×4-C, Current Designs Inc., USA). During the "Pain Localizer", the button box was replaced by the thermode Peltier probe attached on participant's palm. The order of each task was counterbalanced across participants. In cohort 2, these were intermingled with other paradigms which are described elsewhere[35].

Data Processing

For the purpose of the present study, we combined the data from the two cohorts together by assessing effect of healthcare training in terms of four-level GROUP (*Controls*, *Med1*, *Med4*, *Nurses*). Critically, as cohort 2 underwent overall shorted version of the paradigms (see above) we considered only a sub-selection of the data from cohort 1 in order to match the two experiments in terms of power. For the "Handedness" task, we included only those 120 trials (out of 180) which were used in both experiments. For the "Pain Localizer", we included only the functional run in which thermal events were delivered on the right/dominant hand. Finally, for the "Cognitive and Affective Theory of Mind" task, we used data only of the first run in chronological order. For the analysis of behavioural responses, this meant discarding trials/sessions in excess prior to the analysis. For the analysis of brain activity, this meant creating *ad hoc* models where the trials in excess were modelled separately as conditions of no interest (see below).

Behavioural Data

Data from the "Handedness" task were analysed through a factorial schema, where the four conditions of interest (*PF*, *PL*, *cPF*, *cPL*) were organized into two orthogonal within-subject factors of STIMULI (*Painful, Painless*) and EMOTIONAL AROUSAL (*Negative, Neutral*). For each subject and condition, we calculated the average accuracy, median Response Times of correct

responses, and median post-scanning ratings, and then fed them to a Repeated Measures ANOVA with these two within-subject factors. For the "Cognitive and Affective Theory of Mind" task, we modelled the average accuracy and median Response Times of correct responses with an ANOVA with STORY CATEGORY (*Beliefs*, *Emotion*, *Pain*, *Photos*) as unique within-subject factor. In all these models, we also tested the effect of healthcare training by modelling a GROUP factor with four levels (*Controls, Med1, Med4, Nurses*). AGE was always included as nuisance convariate to account for potential confounds between biological and healthcare seniority (participants with more experience are usually older).

Imaging processing

Data Acquisition. A Siemens Trio 3-T whole-body scanner was used to acquire both T1-weighted anatomical images and gradient-echo planar T2*-weighted MRI images with blood oxygenation level dependent (BOLD) contrast. For cohort 1, the scanning sequence was a trajectory-based reconstruction sequence with a repetition time (TR) of 2100 msec, an echo time (TE) of 30 msec, a flip angle of 90 degrees, in-plane resolution 64x64 voxels (voxel size 3 x 3 mm), 32 slices, a slice thickness of 3 mm, with no gap between slices. For cohort 2, we used a multiplex sequence[68], with TR = 650 ms, TE = 30 ms, flip angle = 50°, 36 interleaved slices, 64 x 64 in-slice resolution, 3 x 3 x 3 mm voxel size, and 3.9 mm slice spacing. The multiband accelerator factor was 4, and parallel acquisition techniques (PAT) was not used.

Preprocessing. Statistical analysis was performed using the SPM12 software (http://www.fil.ion.ucl.ac.uk/spm/). For each subject, all functional images were realigned to correct for head movement. For cohort 2, where a field map image was available, we accounted for geometric distortions related to magnetic field inhomogeneity through an unwrapping procedure. The resulting images were then slice-time corrected, to account for

temporal delays within the acquisition of a whole brain volume, and normalized to a template based on 152 brains from the Montreal Neurological Institute (MNI) with a voxel-size resolution of 2 x 2 x 2 mm. The deformation field for the normalization procedure was estimated on the average functional image, which minimizes the effect of geometric distortions of the magnetic field even in absence of prior distortion correction[69]. Finally, the normalized images were smoothed by convolution with an 8 mm full-width at half-maximum Gaussian kernel.

First-Level analysis. Preprocessed images from each task were analysed using the General Linear Model (GLM) framework implemented in SPM, as in previous studies using the same paradigms[15,18,34,35]. For the "Handedness Task", trial time onsets from each of the four conditions of interest were modelled with a delta function. For cohort 1, we modelled 60 trials in excess (30 *PF*, 30 *cPF*; not implemented in cohort 2) as additional conditions of interest. For each condition we also included an additional vector in which individual Response Times were modelled parametrically[15,18,35]. For the "Pain Localizer", each thermal stimulation was modelled based on the time during which temperature was at plateau.

Finally, for the "Cognitive and Affective Theory of Mind Task" we ran a model similar to that of Corradi-Dell'Acqua et al.[34], using a boxcar function for the 12 s long blocks during which a scenario was presented, separately from the subsequent 5 s long blocks during which the judgment took place. We modelled four "judgment" vectors, one for each of the four kinds of stories (*Beliefs*, *Emotion*, *Pain*, *Photos*), whereas we modelled only one "scenario" vector, in which all stories were treated as a unique condition. The latter were complemented by four further parametric regressors, one referring to the number of characters in the scenarios, and the other three describing the likelihood to elicit mental attributions beliefs, emotions, and

pain, respectively, based on previous validation of the database[34]. The scenarios of the photo stories, which had no human protagonists and therefore were not qualified by these dimensions [34], were associated with an artificial value of 0. To avoid biases related to the order of parametric predictors, and to ensure that each effect was uniquely interpretable, the story epochs were modelled after removing the serial orthogonalization option from SPM settings. Text-length was also controlled for the "judgment" blocks by including a parametric regressor which, however, made no distinction between the four kinds of stories.

For all tasks, we accounted for putative habituation effects of neural responses in each condition by using the time-modulation option implemented in SPM, which creates a regressor in which the trial order is modulated parametrically. Furthermore, each regressor was convolved with a canonical hemodynamic response function and associated with its first order temporal derivative. To account for movement-related variance, we included six differential movement parameters as covariates of no interest. Low-frequency signal drifts were filtered using a cut-off period of 128 sec. In cohort 1, serial correlation in the neural signal were accounted for through first-order autoregressive model AR(1). In cohort 2, we used instead an exponential covariance structure, as implemented in the 'FAST' option of SPM12 for rapid sequences. Global scaling was applied, with each fMRI value rescaled to a percentage value of the average whole-brain signal for that scan.

Second-level Analyses. Functional contrasts, comparing differential parameter estimate images associated in one experimental condition *vs.* the other, were then fed in a second level factorial design, where a given effect (described corresponding first-level contrast image) is tested across the four groups. In modelling the variance components, we allowed grouping

factor to have unequal variance across their levels, while subjects were assumed to have equal variance. Within these designs, activated voxels were considered as significant only if they passed a threshold corresponding to $p < 0.05$ FWE corrected at the cluster level [70], with an underlying height threshold corresponding to $p < 0.001$ (uncorrected).

Volume of interest analysis. In addition to the whole-brain analysis, we also constrained our hypothesis by focusing on voxels from brain regions of theoretical interest. For contrasts probing individual ability at assessing others' pain, we considered the recent meta-analysis of Kogler et al.[11], specifically the map describing the conjoint activities between self- and others-pain (Figure 15 in Kogler et al.[11]), available at the following repository https://anima.fz-juelich.de/studies/Kogler_Empathy_2020. For contrasts probing theory-of-mind we considered the meta-analysis of Mar[46], specifically, the map representing convergent activity in story-based theory-of-mind tasks, provided in the Supplementary Materials of the published study[46]. From each of these meta-analytical maps, we created a binary mask in which the outlined voxels were labelled as "1", and the remaining parts of the brain were labelled as "0". Regions within these masks were considered significant if exceeding FWE small volume correction for multiple comparisons.

Vicarious pain signatures. We also submitted data from the "Handedness" task to two multivariate models predictive of vicarious pain from brain activity. To maximize the comparability with our experiment, we choose models derived from datasets in which individuals were shown pictures of limb injuries. The first was developed by Zhou and colleagues[21] (hereafter $Zhou\text{-}NS_{2020}$), and the second by Krishnan and colleagues[38] (hereafter $Krishnan_{2016}$). Both models are freely available at the following repository

https://github.com/canlab/Neuroimaging_Pattern_Masks/tree/master/Multivariate_signature_patterns/, comprising whole brain maps (files: *"NS_vicarious_pain_pattern_unthresholded.nii"* and *"bmrk4_VPS_unthresholded.nii"*, respectively) where values at each voxel describe the relative linear contribution to the prediction of pain observation. Our brain-based estimation of vicarious pain was obtained as the dot-product of the brain activity from our studies with that of the weight maps, as implemented in the code provided together with the models. The resulting vicarious pain estimates associated with each subject/condition were then fed to the same ANOVA scheme used for behavioural measures. To our knowledge, in the current literature, there is no neural model of vicarious pain response based on verbal material, as such no model-based approach was applied to the "Cognitive and Affective Theory of Mind" task

Representational Similarity of Pain. We complemented the above analysis by running a correlation-based Representation Similarity Analysis to identify the presence of any common representation of pain in oneself (from the "Pain Localizer") and pain in others as perceived through pictures ("Handedness") and text ("Cognitive and Affective Theory of Mind"). For this purpose, we run first-level GLMs that were identical to those of the standard univariate analysis, except that they were modelled on preprocessed images without normalization and smoothing[15,18,34]. Following previous studies, we performed a searchlight approach that does not rely on a priori assumptions about informative brain regions, but searches for predictive information throughout the whole brain[15,16,18,34]. For each coordinate in the native brain image, a spherical volume-of-interest was defined around it (5 voxels diameter, 81 voxels total). Then, for each individual subject, we extracted the parameter estimates associated with 9 conditions of interest within this sphere (*Hot* temperatures, *PF* & *PL* images, *Belief*, *Emotion* and *Pain*

scenarios & judgments). Thus, each of the 9 conditions was associated with a unique multivoxel pattern of *β*s in the volume of-interest. These patterns were then correlated one with another, thus resulting in a symmetrical 9 x 9 correlation matrix. The correlation coefficients *r* in this matrix were Fisher transformed $z = 0.5 * \log_e[(1 + r)/(1 - r)]$[15,34] and then assigned to the center voxel of the sphere.

For the purpose of the present study, we considered the average of the *z*-transformed correlation coefficients obtained when pairing together different pain conditions (*Hot* temperatures, *PF* pictures, *Pain* scenarios & judgments). Such "within-pain" similarity was compared to the average correlation obtained when pairing one pain condition with a painless events form the same paradigms. This resulted in different *z*-map for each individual, which was then normalized to the MNI template to the MNI single-subject template and smoothed using an 8 mm FWHM Gaussian kernel. The resulting maps were then fed to the same flexible factorial routine used for the standard univariate analysis.

References

1. Ruben, M. A., van Osch, M. & Blanch-Hartigan, D. Healthcare providers' accuracy in assessing patients' pain: A systematic review. *Patient Education and Counseling* **98**, 1197–1206 (2015).
2. Ruben, M. A., Blanch-Hartigan, D. & Shipherd, J. C. To Know Another's Pain: A Meta-analysis of Caregivers' and Healthcare Providers' Pain Assessment Accuracy. *Annals of Behavioral Medicine* **52**, 662–685 (2018).
3. Rupp, T. & Delaney, K. A. Inadequate analgesia in emergency medicine. *Ann Emerg Med* **43**, 494–503 (2004).
4. Dirupo, G., Totaro, S., Richard, J. & Corradi-Dell'Acqua, C. Medical education and distrust modulate the response of insular-cingulate network and ventral striatum in pain diagnosis. *eLife* **10**, e63272 (2021).
5. Xie, J., Yang, H., Xia, X. & Yu, S. The Influence of Medical Professional Knowledge on Empathy for Pain: Evidence From fNIRS. *Front. Psychol.* **9**, (2018).
6. Davoudi, N., Afsharzadeh, P., Mohammadalizadeh, S. & Haghdoost, A. A. A comparison of patients' and nurses' assessments of pain intensity in patients with coronary artery disease. *International Journal of Nursing Practice* **14**, 347–356 (2008).
7. Choinière, M., Melzack, R., Girard, N., Rondeau, J. & Paquin, M.-J. Comparisons between patients' and nurses' assessment of pain and medication efficacy in severe burn injuries. *Pain* **40**, 143–152 (1990).
8. Ding, R. *et al.* Domain-general and domain-preferential neural correlates underlying empathy towards physical pain, emotional situation and emotional faces: An ALE meta-analysis. *Neuropsychologia* 107286 (2019) doi:10.1016/j.neuropsychologia.2019.107286.
9. Fan, Y., Duncan, N. W., de Greck, M. & Northoff, G. Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. *Neuroscience & Biobehavioral Reviews* **35**, 903–911 (2011).
10. Jauniaux, J., Khatibi, A., Rainville, P. & Jackson, P. L. A meta-analysis of neuroimaging studies on pain empathy: investigating the role of visual information and observers' perspective. *Soc Cogn Affect Neurosci* **14**, 789–813 (2019).
11. Kogler, L., Müller, V. I., Werminghausen, E., Eickhoff, S. B. & Derntl, B. Do I feel or do I know? Neuroimaging meta-analyses on the multiple facets of empathy. *Cortex* (2020) doi:10.1016/j.cortex.2020.04.031.
12. Lamm, C., Decety, J. & Singer, T. Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage* **54**, 2492–2502 (2011).

13. Schurz, M. *et al.* Toward a hierarchical model of social cognition: A neuroimaging meta-analysis and integrative review of empathy and theory of mind. *Psychol Bull* (2020) doi:10.1037/bul0000303.

14. Timmers, I. *et al.* Is Empathy for Pain Unique in Its Neural Correlates? A Meta-Analysis of Neuroimaging Studies of Empathy. *Front. Behav. Neurosci.* **12**, (2018).

15. Corradi-Dell'Acqua, C., Hofstetter, C. & Vuilleumier, P. Felt and Seen Pain Evoke the Same Local Patterns of Cortical Activity in Insular and Cingulate Cortex. *J. Neurosci.* **31**, 17996–18006 (2011).

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

17. Braboszcz, C., Brandao-Farinelli, E. & Vuilleumier, P. Hypnotic analgesia reduces brain responses to pain seen in others. *Scientific Reports* **7**, 9778 (2017).

18. Qiao-Tasserit, E., Corradi-Dell'Acqua, C. & Vuilleumier, P. The good, the bad, and the suffering. Transient emotional episodes modulate the neural circuits of pain and empathy. *Neuropsychologia* **116**, 99–116 (2018).

19. O'Connell, K. *et al.* Increased similarity of neural responses to experienced and empathic distress in costly altruism. *Sci Rep* **9**, 10774 (2019).

20. Wagner, I. C., Rütgen, M. & Lamm, C. Pattern similarity and connectivity of hippocampal-neocortical regions support empathy for pain. *Soc Cogn Affect Neurosci* **15**, 273–284 (2020).

21. Zhou, F. *et al.* Empathic pain evoked by sensory and emotional-communicative cues share common and process-specific neural representations. *eLife* **9**, e56929 (2020).

22. Berluti, K. *et al.* Reduced Multivoxel Pattern Similarity of Vicarious Neural Pain Responses in Psychopathy. *Journal of Personality Disorders* **34**, 628–649 (2020).

23. Rütgen, M. *et al.* Placebo analgesia and its opioidergic regulation suggest that empathy for pain is grounded in self pain. *PNAS* **112**, E5638–E5646 (2015).

24. Rütgen, M. *et al.* Beyond Sharing Unpleasant Affect—Evidence for Pain-Specific Opioidergic Modulation of Empathy for Pain. *Cerebral Cortex* (2021) doi:10.1093/cercor/bhaa385.

25. Bastiaansen, J. A. C. J., Thioux, M., & Christian Keysers. Evidence for mirror systems in emotions. *Philos. Trans. R. Soc. Lond., B, Biol. Sci* **364**, 2391–2404 (2009).

26. Bernhardt, B. C. & Singer, T. The neural basis of empathy. *Annu. Rev. Neurosci.* **35**, 1–23 (2012).

27. Cheng, Y. *et al.* Expertise modulates the perception of pain in others. *Curr. Biol* **17**, 1708–1713 (2007).

28. Cheng, Y., Chen, C. & Decety, J. How Situational Context Impacts Empathic Responses and Brain Activation Patterns. *Front. Behav. Neurosci.* **11**, (2017).

29. Decety, J., Yang, C.-Y. & Cheng, Y. Physicians down-regulate their pain empathy response: An event-related brain potential study. *Neuroimage* (2010) doi:10.1016/j.neuroimage.2010.01.025.

30. Gleichgerrcht, E. & Decety, J. The relationship between different facets of empathy, pain perception and compassion fatigue among physicians. *Front. Behav. Neurosci.* **8**, (2014).

31. Vaes, J. & Muratore, M. Defensive dehumanization in the medical practice: a cross-sectional study from a health care worker's perspective. *Br J Soc Psychol* **52**, 180–190 (2013).

32. Liang, M., Su, Q., Mouraux, A. & Iannetti, G. D. Spatial Patterns of Brain Activity Preferentially Reflecting Transient Pain and Stimulus Intensity. *Cerebral Cortex* **29**, 2211–2227 (2019).

33. Sharvit, G., Lin, E., Vuilleumier, P. & Corradi-Dell'Acqua, C. Does inappropriate behavior hurt or stink? The interplay between neural representations of somatic experiences and moral decisions. *Science Advances* **6**, (2020).

34. Corradi-Dell'Acqua, C., Hofstetter, C. & Vuilleumier, P. Cognitive and affective theory of mind share the same local patterns of activity in posterior temporal but not medial prefrontal cortex. *Soc Cogn Affect Neurosci* **9**, 1175–1184 (2014).

35. Corradi-Dell'Acqua, C. *et al.* Pain management decisions in emergency hospitals are predicted by brain activity during empathy and error monitoring. *Br J Anaesth* **123**, e284–e292 (2019).

36. Corradi-Dell'Acqua, C. *et al.* Deficits in cognitive and affective theory of mind relate to dissociated lesion patterns in prefrontal and insular cortex. *Cortex* **128**, 218–233 (2020).

37. Gu, X. *et al.* Functional dissociation of the frontoinsular and anterior cingulate cortices in empathy for pain. *J. Neurosci* **30**, 3739–3744 (2010).

38. Krishnan, A. *et al.* Somatic and vicarious pain are represented by dissociable multivariate brain patterns. *eLife* **5**, e15166 (2016).

39. Shamay-Tsoory, S. G. The Neural Bases for Empathy. *Neuroscientist* **17**, 18–24 (2011).

40. Stietz, J., Jauk, E., Krach, S. & Kanske, P. Dissociating Empathy From Perspective-Taking: Evidence From Intra- and Inter-Individual Differences Research. *Front. Psychiatry* **10**, (2019).

41. Aichhorn, M. *et al.* Temporo-parietal Junction Activity in Theory-of-Mind Tasks: Falseness, Beliefs, or Attention. *Journal of Cognitive Neuroscience* **21**, 1179–1192 (2008).

42. Bruneau, E., Pluta, A. & Saxe, R. Distinct roles of the 'Shared Pain' and 'Theory of Mind' networks in processing others' emotional suffering. *Neuropsychologia* **50**, 219–231 (2012).

43. Jacoby, N., Bruneau, E., Koster-Hale, J. & Saxe, R. Localizing Pain Matrix and Theory of Mind networks with both verbal and non-verbal stimuli. *NeuroImage* **126**, 39–48 (2016).

44. Bruneau, E. G., Dufour, N. & Saxe, R. Social cognition in members of conflict groups: behavioural and neural responses in Arabs, Israelis and South Americans to each other's misfortunes. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**, 717–730 (2012).

45. Bruneau, E. G., Jacoby, N. & Saxe, R. Empathic control through coordinated interaction of amygdala, theory of mind and extended pain matrix brain regions. *NeuroImage* **114**, 105–119 (2015).

46. Mar, R. A. The neural bases of social cognition and story comprehension. *Annu Rev Psychol* **62**, 103–134 (2011).

47. Molenberghs, P., Johnson, H., Henry, J. D. & Mattingley, J. B. Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews* **65**, 276–291 (2016).

48. Schurz, M., Radua, J., Aichhorn, M., Richlan, F. & Perner, J. Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews* **42**, 9–34 (2014).

49. Van Overwalle, F. Social cognition and the brain: A meta-analysis. *Human Brain Mapping* **30**, 829–858 (2009).

50. Karjalainen, T. *et al.* Dissociable Roles of Cerebral μ-Opioid and Type 2 Dopamine Receptors in Vicarious Pain: A Combined PET–fMRI Study. *Cereb Cortex* **27**, 4257–4266 (2017).

51. López-Solà, M., Koban, L., Krishnan, A. & Wager, T. D. When pain really matters: A vicarious-pain brain marker tracks empathy for pain in the romantic partner. *Neuropsychologia* **145**, 106427 (2020).

52. Iannetti, G. D. & Mouraux, A. From the neuromatrix to the pain matrix (and back). *Exp Brain Res* **205**, 1–12 (2010).

53. Mouraux, A., Diukova, A., Lee, M. C., Wise, R. G. & Iannetti, G. D. A multisensory investigation of the functional significance of the "pain matrix". *NeuroImage* **54**, 2237–2249 (2011).

54. Horing, B., Sprenger, C. & Büchel, C. The parietal operculum preferentially encodes heat pain and not salience. *PLOS Biology* **17**, e3000205 (2019).

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

56. Uddin, L. Q. Salience processing and insular cortical function and dysfunction. *Nature Reviews Neuroscience* **16**, 55–61 (2015).

57. Hojat, M. *et al.* The Devil is in the Third Year: A Longitudinal Study of Erosion of Empathy in Medical School. *Academic Medicine* **84**, 1182–1191 (2009).

58. Neumann, M. *et al.* Empathy decline and its reasons: a systematic review of studies with medical students and residents. *Acad Med* **86**, 996–1009 (2011).

59. Bellini, L. M. & Shea, J. A. Mood Change and Empathy Decline Persist during Three Years of Internal Medicine Training. *Academic Medicine* **80**, 164–167 (2005).

60. Smith, K. E., Norman, G. J. & Decety, J. The complexity of empathy during medical school training: evidence for positive changes. *Medical Education* **51**, 1146–1159 (2017).

61. Cameron, C. D. & Inzlicht, M. Empathy choice in physicians and non-physicians. *British Journal of Social Psychology* **59**, 715–732 (2020).

62. Handford, C., Lemon, J., Grimm, M. C. & Vollmer-Conna, U. Empathy as a Function of Clinical Exposure - Reading Emotion in the Eyes. *PLOS ONE* **8**, e65159 (2013).

63. Kataoka, H. U., Koide, N., Ochi, K., Hojat, M. & Gonnella, J. S. Measurement of Empathy Among Japanese Medical Students: Psychometrics and Score Differences by Gender and Level of Medical Education. *Academic Medicine* **84**, 1192–1197 (2009).

64. Wang, X. & Cheng, Z. Cross-Sectional Studies: Strengths, Weaknesses, and Recommendations. *Chest* **158**, S65–S71 (2020).

65. Kappesser, J., Williams, A. C. de C. & Prkachin, K. M. Testing two accounts of pain underestimation. *Pain* **124**, 109–116 (2006).

66. Teske, K., Daut, R. L. & Cleeland, C. S. Relationships between nurses' observations and patients' self-reports of pain. *Pain* **16**, 289–296 (1983).

67. Corradi-Dell'Acqua, C., Tomasino, B. & Fink, G. R. What is the position of an arm relative to the body? Neural correlates of body schema and body structural description. *J. Neurosci* **29**, 4162–4171 (2009).

68. Feinberg, D. A. *et al.* Multiplexed Echo Planar Imaging for Sub-Second Whole Brain FMRI and Fast Diffusion Imaging. *PLOS ONE* **5**, e15710 (2010).

69. Calhoun, V. D. *et al.* The impact of T1 versus EPI spatial normalization templates for fMRI data analyses. *Human Brain Mapping* **38**, 5331–5342 (2017).

70. Friston, K. J., Worsley, K. J., Frackowiak, R. S. J., Mazziotta, J. C. & Evans, A. C. Assessing the significance of focal activations using their spatial extent. *Hum. Brain Mapp.* **1**, 210–220 (1993).

Acknowledgments

This study was conducted at the Brain and Behaviour Laboratory (BBL) at the University of Geneva and benefited from the support of the BBL technical staff.

Authors Contribution

CCD, OH, and PV conceived the design. CCD & CH acquired the data from cohort 1. CCD, GS & EF acquired the data from cohort 2. CCD analysed the data. OH, PV contributed to the interpretation of the results. CCD drafted the manuscript. All authors revised critically the manuscript.

Funding

CCD, is supported by the Swiss National Science Foundation (SNSF) grant nos. PP00O1_157424 and PP00P1_183715. PV is supported by the SNSF grant no 32003B_138413.

Competing Interests

The authors declare they have no competing interests.