

1 **Empathic pain evoked by sensory and emotional-communicative cues share common**
2 **and process-specific neural representations**

3

4 Feng Zhou^{a,b#}, Jialin Li^a, Weihua Zhao^a, Lei Xu^a, Xiaoxiao Zheng^a, Meina Fu^a, Shuxia Yao^a, Keith
5 M. Kendrick^a, Tor D. Wager^b, Benjamin Becker^{a#}

6

7 ^aClinical Hospital of Chengdu Brain Science Institute, MOE Key Laboratory for Neuroinformation,
8 University of Electronic Science and Technology of China, Chengdu, China

9 ^bDepartment of Psychological and Brain Sciences, Dartmouth College, New Hampshire, United

10 States of America

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25 **#Correspondence**

26 Feng Zhou (zhou.feng@live.com) and Benjamin Becker (ben_becker@gmx.de)

27 Clinical Hospital of Chengdu Brain Science Institute, MOE Key Laboratory for Neuroinformation

28 University of Electronic Science and Technology of China

29 Chengdu 611731, China

30 Tel.: +86 2861 830 811

31 **Abstract**

32 Pain empathy can be evoked by multiple cues, particularly observation of acute pain inflictions or
33 facial expressions of pain. Previous studies suggest that these cues commonly activate the insula
34 and anterior cingulate, yet vicarious pain encompass pain-specific responses as well as unspecific
35 processes (e.g., arousal) and overlapping activations are not sufficient to determine
36 process-specific shared neural representations. We employed multivariate pattern analyses to
37 fMRI data acquired during observation of noxious stimulation of body limbs (NS) and painful
38 facial expressions (FE) and found spatially and functionally similar cross-modality (NS versus
39 FE) whole-brain vicarious pain-predictive patterns. Further analyses consistently identified
40 shared neural representations in the bilateral mid-insula. The vicarious pain patterns were not
41 sensitive to respond to non-painful high-arousal negative stimuli but predicted self-experienced
42 thermal pain. Finally, a domain-general vicarious pain pattern predictive of self-experienced pain
43 but not arousal was developed. Our findings demonstrate shared pain-associated neural
44 representations of vicarious pain.

45

46 **Impact statements**

47 Machine learning analyses reveal that the observation of acute pain inflictions and facial

48 expressions of pain evoke shared pain-specific neural representations.

49

50 **Introduction**

51 Pain empathy, the capacity to resonate with, relate to, and share others' pain, is an essential part of
52 human experience. Among other functions, it motivates helping and cooperative behaviors and
53 aids in learning to avoid harmful situations . Vicarious pain can be triggered by observing or
54 imagining another individual's painful state and can be elicited by multiple types of social cues,
55 particularly the observation of an inflicted physical injury or a facial expression of pain (Decety &
56 Ickes, 2009; Jauniaux et al., 2019; Vachon-Presseau et al., 2012; Yesudas & Lee, 2015). While
57 stimuli depicting the noxious stimulation of body limbs [i.e., observation of noxious stimulation
58 (NS) induced vicarious pain (NS vicarious pain)] provides objective cues about the sensory
59 component of the observed pain, the observation of facial expressions of pain [i.e., facial
60 expressions induced vicarious pain (FE vicarious pain)] is considered more subjective and indirect
61 as the pain experience of the expresser needs to be interpreted by the observer (Hadjistavropoulos
62 et al., 2011; Vachon-Presseau et al., 2012). Functional magnetic resonance imaging (fMRI) studies
63 employing corresponding pictorial stimuli have identified distinct and common neural substrates
64 of pain empathy across vicarious pain induction procedures (Jauniaux et al., 2019). For example,
65 Vachon-Presseau et al. (2012) demonstrated that NS vicarious pain increased activity in core
66 regions of the mirror neuron system, specifically inferior frontal and posterior regions engaged in
67 coding sensory-somatic information (Rizzolatti & Craighero, 2004) while the presentation of FE
68 vicarious pain led to stronger increases in the medial prefrontal cortex and precuneus which have
69 been associated with social cognitive processes such as mentalizing and theory of mind (Amft et
70 al., 2015; Gallo et al., 2018; Mitchell, 2009; Northoff et al., 2006). Despite the different
71 psychological domains engaged in the pain empathic response induced by NS and FE both elicit

72 vicarious pain experience (Timmers et al., 2018), encompassing pain-specific processes such as
73 recognizing and understanding the painful state of the other person and affective sharing of pain
74 but also non-specific processes that are shared between pain and other non-painful experiences
75 such as arousal and negative affect (Zaki et al., 2016). In line with the shared underlying mental
76 processes previous neuroimaging meta-analyses revealed that the observation of acute pain
77 infliction and painful facial expressions commonly activate core empathy and nociceptive pain
78 regions specifically the insular and cingulate cortex (Jauniaux et al., 2019; Lamm et al., 2011;
79 Timmers et al., 2018). The overlapping activations have been suggested to reflect shared neural
80 representations of vicarious pain (Jauniaux et al., 2019; Lamm et al., 2011; Timmers et al., 2018).

81 However, overlapping functional activations within these regions do not necessarily reflect
82 shared underlying neural representations of a specific mental process (Zaki et al., 2016), given that
83 (1) due to local spatial dependencies the main focus of traditional mass-univariate fMRI analytic
84 approaches (i.e., conducting massive number of tests on brain voxels one at a time) is not on
85 single-voxel activity, but on smoothed, regional differences in brain activity across multiple tasks
86 or stimuli (Haynes, 2015), and (2) brain regions may contain multiple, distinct populations of
87 neurons and averaging across those neuron populations yields nonspecific signals (Haxby et al.,
88 2014; Zaki et al., 2016). For instance, electrophysiological and optogenetic studies have identified
89 distinct neuronal populations in the anterior cingulate and insular cortex that activate during
90 several functional domains, including pain- and empathy-related processes as well as attention,
91 salience, social observation learning and reward expectancy (Allman et al., 2011; Chen, 2018;
92 Kvitsiani et al., 2013; Sakaguchi et al., 2018; Shidara & Richmond, 2002; Shura et al., 2014; Sikes

93 & Vogt, 1992). Studies employing mass-univariate fMRI analyses suggest that both regions are
94 engaged by various experimental paradigms including not only experienced and observed pain,
95 but also reward, arousal, salience and attention (Cauda et al., 2012; Shackman et al., 2011; Uddin,
96 2015; Wager et al., 2016; Yarkoni et al., 2011). Despite the overlapping fMRI activation in
97 response to different experimental manipulations the underlying brain representations may be
98 separable (Corradi-Dell'Acqua et al., 2016; Krishnan et al., 2016; Woo et al., 2014), emphasizing
99 that more fine-grained analyses are required to determine process-specific shared or distinct neural
100 representations (Zaki et al., 2016).

101 In an effort to overcome these limitations recent studies have proposed several strategies to
102 investigate the “shared representation” question, including pharmacological (see e.g., Rütgen et
103 al., 2015) and multivariate pattern analysis (MVPA) approaches. Compared to conventional
104 analytic approaches MVPA can be effective in extracting information at much finer spatial
105 scales (e.g., below the intrinsic resolution determined by the voxel size by pooling together weak
106 feature-selective signals in each voxel; Kamitani & Tong, 2005; Woo et al., 2017) and represents
107 a more suitable approach to support or reject claims about neural mechanisms that are shared
108 between mental processes (Chikazoe et al., 2014; Peelen & Downing, 2007; Zaki et al., 2016). In
109 support of this view, using MVPA approaches researchers have demonstrated shared neural
110 representations across mental processes (including self-experienced and observed pain) in both
111 humans and animals (Carrillo et al., 2019; Corradi-Dell'Acqua et al., 2011; Corradi-Dell'Acqua
112 et al., 2016). Moreover, a growing number of recent studies have demonstrated functional
113 independence of overlapping univariate activation in these brain regions using MVPA (Krishnan

114 et al., 2016; Peelen et al., 2006; Woo et al., 2014), including separable neural representations of
115 physical and social rejection pain within the dorsal anterior cingulate cortex (Woo et al., 2014)
116 and of modality-specific aversive experience in the anterior insular cortex (Krishnan et al., 2016).

117 Nevertheless, shared multivariate patterns do not necessarily imply process-specific common
118 neural representations per se given that the shared neural representations could simply reflect
119 common demands on basal processing domains such as attention or arousal (Corradi-Dell'Acqua
120 et al., 2011; Corradi-Dell'Acqua et al., 2016; Krishnan et al., 2016). For instance,
121 Corradi-Dell'Acqua et al. (2016) found shared neural patterns between vicarious and
122 self-experienced pain in the left anterior insula and further demonstrated that the common local
123 patterns were not specific to pain-related processing, but also represented disgust and unfairness
124 suggesting modality-unspecific processing of aversive and arousing experiences.

125 This leads to the questions of (1) whether or not NS and FE induced vicarious pain share
126 pain-associated common neural representations, and further (2) whether a general (i.e., across NS
127 and FE vicarious pain modalities) neural signature of vicarious pain, which is specific to pain
128 empathic response rather than capturing unspecific processes such as negative emotional
129 experience or arousal, can be determined. More specifically, we examined the following three
130 questions in this study: (i) whether NS and FE induced vicarious pain-predictive signatures share
131 spatially (correlation and distribution) and functionally (predictions of cross-modality vicarious
132 pain versus corresponding non-painful control stimuli) similar neural representations, (ii) whether
133 a general and specific vicarious pain-predictive neural signature, which should (1) generalize
134 across different vicarious pain stimuli and (2) not be sensitive to predict unspecific negative affect

135 or arousal induced by non-painful negative stimuli, and (3) be ‘activated’ by the direct experience
136 of somatic pain as reflected by an accurate prediction of self-experienced somatic pain, can be
137 determined.

138 To this end we employed MVPA to fMRI data from an experiment during which participants
139 were presented with stimuli depicting the infliction of noxious stimulation of body limbs (NS
140 vicarious pain) or painful facial expressions (FE vicarious pain) as well as corresponding
141 non-painful control stimuli (**Figure 1A**). Given that relative to their control stimuli both sets of
142 painful stimuli were perceived as more painful in terms of recognized and shared pain as well as
143 more arousing and negative (details see **Results and Figure 1B**), we additionally asked
144 participants to undergo an emotion processing paradigm with non-painful high-arousal negative
145 stimuli and low-arousal neutral stimuli from the International Affective Picture System (IAPS)
146 (details see **Methods**) to further test the specificity of the shared neural representations with
147 respect to the vicarious experience of pain rather than emotional arousal or negative affect. To
148 determine the association of the vicarious pain signature with direct pain experience we included
149 an independent fMRI dataset that collected ratings of self-experienced pain during thermal pain
150 induction (details are provided in **Methods** and Wager et al., 2013; Woo et al., 2015).

151 Given that small sample sizes may lead to a large cross-validation error which is the
152 discrepancy between the prediction accuracy measured by cross-validation and the expected
153 accuracy on new data (Varoquaux, 2018) and fMRI-based inferences on regions that are most
154 predictive substantially benefit from larger samples (Chang et al., 2015) we included a comparably
155 large sample of n = 238 individuals (details see Methods and Li et al., 2018; Xu et al., 2020).

156

157 **Results**

158 **Evaluation of the stimuli**

159 To match the instructions between the vicarious pain and negative emotion fMRI paradigms an

160 implicit instruction was provided (attentively view the pictorial stimuli) (details see **Methods**).

161 Affective ratings of the stimuli in an independent sample confirmed that both sets of painful

162 stimuli were rated as considerably more painful compared to their respective control stimuli,

163 both in terms of recognizing and sharing pain, and additionally were rated as more arousing and

164 negative. Specifically, both categories of painful stimuli elicited a substantial level of pain

165 intensity perceived for the person displayed as well as in the observer. The NS vicarious pain

166 stimuli were rated considerably higher on both dimensions (mean \pm SE pain intensity displayed

167 = 6.73 ± 0.27 ; mean \pm SE pain intensity self-experienced = 6.14 ± 0.36) as compared to the

168 corresponding NS control stimuli (mean \pm SE pain intensity displayed = 1.37 ± 0.11 ; mean \pm SE

169 pain intensity self-experienced = 1.54 ± 0.14 ; $t_{37} = 18.11, P < 0.001$; $t_{37} = 12.71, P < 0.001$,

170 respectively). Similarly, the FE vicarious pain stimuli were also rated substantially higher on

171 both pain-related dimensions (mean \pm SE pain intensity displayed = 6.20 ± 0.25 ; mean \pm SE pain

172 intensity self-experienced = 5.05 ± 0.31) as compared to the corresponding FE control stimuli

173 (mean \pm SE pain intensity displayed = 1.78 ± 0.16 ; mean \pm SE pain intensity self-experienced =

174 2.20 ± 0.18 ; $t_{37} = 14.58, P < 0.001$; $t_{37} = 9.00, P < 0.001$, respectively) (**Figure 1B**). Moreover,

175 both categories of painful stimuli were rated as considerably more negative and induced stronger

176 arousal in the participants as compared to their respective control stimuli (NS vicarious pain

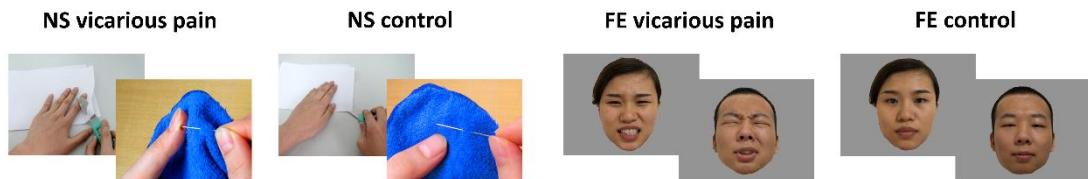
177 stimuli: mean \pm SE valence = 3.14 ± 0.22 ; mean \pm SE arousal = 5.81 ± 0.32 ; NS control stimuli:
178 mean \pm SE valence = 5.12 ± 0.15 ; mean \pm SE valence = 2.68 ± 0.25 ; $t_{37} = -7.99$, $P < 0.001$; $t_{37} =$
179 9.02 , $P < 0.001$, respectively; FE vicarious pain stimuli: mean \pm SE valence = 3.57 ± 0.21 ; mean
180 \pm SE arousal = 5.03 ± 0.29 ; FE control stimuli: mean \pm SE valence = 4.83 ± 0.12 ; mean \pm SE
181 valence = 3.34 ± 0.24 ; $t_{37} = -5.24$, $P < 0.001$; $t_{37} = 6.50$, $P < 0.001$, respectively) (**Figure 1B**).

182 Likewise, the non-painful negative IAPS pictures were rated as considerably more arousing
183 and negative as compared to the corresponding neutral stimuli. Specifically, negative stimuli
184 elicited substantial negative affect and arousal on numerical rating scales (mean \pm SE valence =
185 2.41 ± 0.16 ; mean \pm SE arousal = 6.34 ± 0.22) compared with neutral stimuli (mean \pm SE valence
186 = 5.35 ± 0.08 ; mean \pm SE arousal = 3.22 ± 0.25 ; $t_{36} = -16.09$, $P < 0.001$; $t_{36} = 12.65$, $P < 0.001$,
187 respectively).

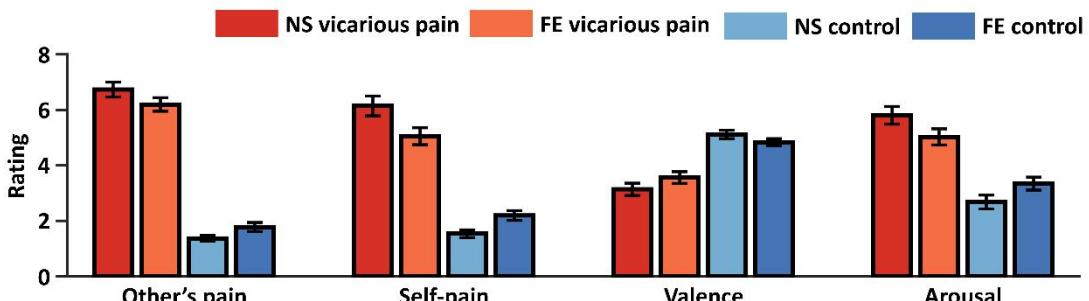
188 Post-fMRI ratings further confirmed that vicarious pain stimuli elicited higher recognizing
189 pain and arousal as compared to the control stimuli (**Supplementary File 1**). Of note, although we
190 found that the two vicarious pain evoking stimulus sets were not fully matched in terms of
191 vicarious pain intensity, arousal and valence, the differences between the stimulus sets might have
192 a small effect on our findings given that (1) this study focused on common rather than different
193 empathic pain responses elicited by the two stimulus sets and (2) both categories of vicarious pain
194 stimuli elicited substantial levels of pain empathy.

195

A Examples of vicarious pain-evoking and control images



B Behavioural ratings of the stimuli



196

197 **Figure 1.** Examples and behavioural ratings of the experimental stimuli. (A) Examples of stimuli for
198 NS and FE vicarious pain as well as corresponding non-painful control stimuli. Of note, examples of
199 the facial expressions (FE) were not included in the original stimulus set and written consent was
200 obtained from the two volunteers. (B) Behavioural ratings of the stimuli from an independent
201 sample ($n = 38$). Error bars represent standard errors of the mean. “Other’s pain” indicates “how
202 much pain do you think the person in the photo is feeling”, “self-pain” indicates “how much pain do
203 you experience when watching the picture”. All ratings were assessed by nine-point Likert scales
204 ranging from ‘1 = not painful at all or very negative or very low arousing’ to ‘9 = extremely painful
205 or very positive or very high arousing’. NS vicarious pain, observation of noxious stimulation of
206 body limbs induced vicarious pain; FE vicarious pain, observation of facial expressions of pain
207 induced vicarious pain; NS control stimuli depict body limbs in similar but innocuous situations, FE
208 control stimuli show neutral facial expressions.

209

210 **Univariate approach - shared activations for NS and FE vicarious pain**

211 To test whether NS and FE vicarious pain share similar activation patterns as determined by

212 traditional mass-univariate analyses, we performed a permutation-based correlation analysis to

213 compare the spatial similarity between the unthresholded group-level NS vicarious pain

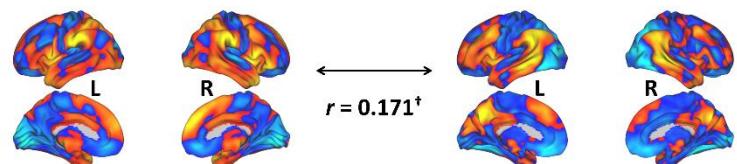
214 activation (NS vicarious pain > NS control) and the FE vicarious pain activation (FE vicarious

215 pain > FE control). We found that activation in response to NS vicarious pain was spatially

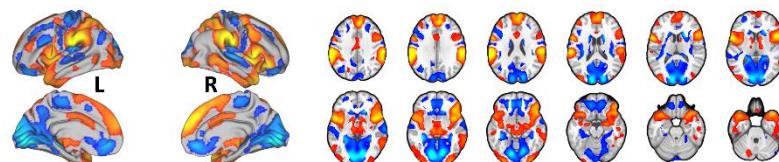
216 correlated with that to FE vicarious pain ($r = 0.171, P < 0.1$ based on permutation tests) (**Figure**

217 2A). Moreover, after multiple comparisons correction (FDR corrected, $q < 0.05$, two-tailed)
218 (**Figure 2B and 2C**), distributed regions of overlapping activation were identified, including a
219 network exhibiting increased activation during both modalities encompassing the bilateral
220 anterior and mid-insula, dorsomedial prefrontal cortex, inferior parietal lobule, middle frontal
221 gyrus and middle temporal gyrus, as well as a network of decreased activation, including the
222 rostral and ventral anterior cingulate cortices, ventromedial and orbitofrontal cortices, and
223 lingual and parahippocampal gyri (**Figure 2D**).
224

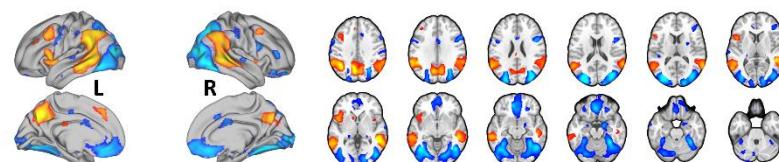
A Activation similarity between NS and FE vicarious pain



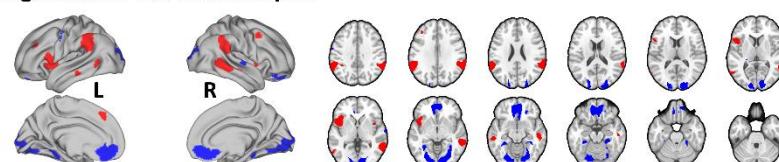
B NS vicarious pain > control (FDR $q < 0.05$)



C FE vicarious pain > control (FDR $q < 0.05$)



D Overlapping regions from the GLM analyses



225
226 **Figure 2.** Results from the conventional univariate analyses. (A) The NS vicarious pain activation
227 pattern was spatially correlated with the FE vicarious pain pattern. (B) Results from the conventional
228 univariate analysis comparing NS vicarious pain with the NS control stimuli thresholded at FDR $q <$
229 0.05 (two-tailed). (C) Results from the univariate analysis comparing FE vicarious pain with the FE
230 control stimuli thresholded at FDR $q < 0.05$ (two-tailed). (D) Overlapping activation between NE and
231 FE vicarious pain as determined by the conventional univariate approach. ${}^{\dagger}P < 0.1$. NS vicarious pain,

232 observation of noxious stimulation of body limbs induced vicarious pain; FE vicarious pain,
233 observation of facial expressions of pain induced pain; NS control stimuli depict limbs in similar
234 but innocuous situations, FE control stimuli show neutral facial expressions.

235

236 **Multivariate approach – modality general vicarious pain-predictive patterns**

237 Previous studies suggest that pain and negative emotional processes are distributed across brain

238 regions (Chang et al., 2015; Krishnan et al., 2016; Wager et al., 2013) and that compared to

239 whole-brain predictive model local regions explain considerably less variance in predicting these

240 processes (Kragel et al., 2018; Woo et al., 2017). In an initial step we therefore developed novel

241 whole-brain patterns to decode NS and FE vicarious pain separately. The NS vicarious

242 pain-predictive pattern yielded an average classification accuracy of $88 \pm 1.5\%$ standard error

243 (SE), $P < 0.001$, $d = 2.13$; d indicates effect size in terms of Cohen's d (accuracy = $96 \pm 1.2\%$

244 SE, $P < 0.001$, $d = 2.17$ based on a two-alternative forced-choice test) and the FE vicarious

245 pain-predictive pattern discriminated FE vicarious pain versus FE control with $80 \pm 1.8\%$ SE

246 accuracy, $P < 0.001$, $d = 1.64$ (accuracy = $88 \pm 2.1\%$ SE, $P < 0.001$, $d = 1.57$ based on a

247 two-alternative forced-choice test) with a 10-fold cross-validation procedure which was repeated

248 10 times, yielding 10 random partitions of the original sample.

249 Next permutation-based correlation analysis was employed to determine the similarity

250 between the whole-brain patterns of NS and FE vicarious pain which confirmed that the

251 modality-specific patterns were spatially correlated ($r = 0.170$, $P < 0.001$ based on permutation

252 tests) (**Figure 3A**). To further qualitatively determine shared but also distinct vicarious pain

253 signatures we analyzed the spatial covariation between the unthresholded weight maps for NS and

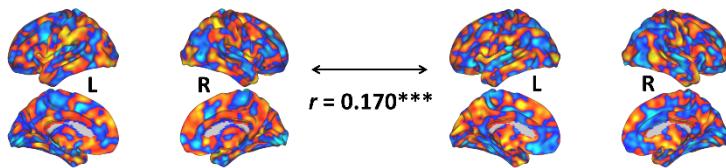
254 FE vicarious pain. To this end we plotted the joint distribution of normalized (z-transformed)

255 voxel weights of the FE vicarious pain-predictive pattern on the *x*-axis and the NS vicarious
256 pain-predictive patterns on the *y*-axis in **Figure 3B** (for similar approach see Koban et al., 2019;
257 Yu et al., 2019). Briefly, pattern weights in any given voxel are expressed as positive, negative or
258 near-zero values for each of the vicarious pain-predictive patterns, which allows to divide voxels
259 into eight equally-sized Octants depending on the relative weights in each pattern. For visual
260 presentation the Octants were color-coded with different colors indicating either voxels of shared
261 positive or shared negative weight (Octants 2 and 6, respectively), selectively positive weights for
262 NS (Octant 1) and FE (Octant 3) vicarious pain-predictive patterns, selectively negative weights
263 for either NS (Octant 5) or FE (Octant 7) vicarious pain-predictive patterns, or opposite weights in
264 the two decoders such that voxels in Octants 4 and 8 express positive and negative weights for the
265 FE vicarious pain-predictive pattern but negative and positive weights for NS vicarious
266 pain-predictive pattern, respectively. Furthermore, to provide an overall measure for voxels in
267 each Octant, we computed the sum of squared distances (SSD) from the origin, which accounts for
268 both, absolute numbers of voxels in each Octant and their (squared) distance from the origin. This
269 analysis of the spatial coactivation of NS and FE vicarious pain-predictive patterns revealed peak
270 SSDs in Octant 2 and 6 as compared to other Octants, suggesting that a considerable number of
271 voxels express positive or negative weights for both vicarious pain-predictive patterns. Overall,
272 this analysis provides further supports largely shared, but also non-shared, neural representations
273 for NS and FE vicarious pain. In support of this, between-modality classification showed that the
274 NS vicarious pain-predictive pattern could reliably discriminate FE vicarious pain versus FE
275 control with 69% accuracy ($\pm 3.0\%$ SE, $P < 0.001$, $d = 0.65$) and that the FE vicarious

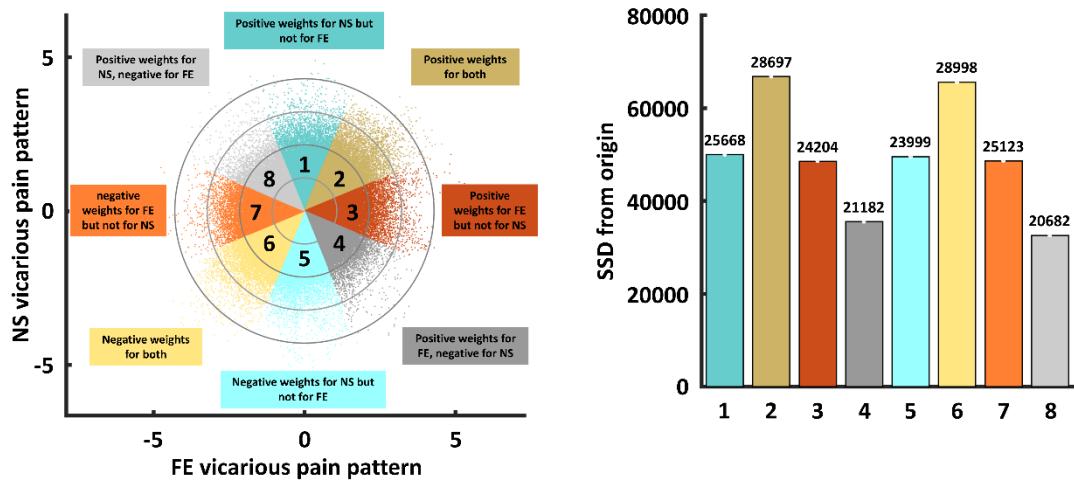
276 pain-predictive pattern could discriminate NS vicarious pain versus NS control with 78%
277 accuracy ($\pm 2.7\%$ SE, $P < 0.001$, $d = 1.00$) based on two-alternative forced-choice tests with a
278 repeated 10-fold cross-validation procedure (**Figure 3C**). Taken together, our results confirmed
279 shared neural representations between the different vicarious pain modalities at the whole-brain
280 level, yet the reduced between-modality prediction effect sizes as compared to within-modality
281 prediction effect sizes (< 50%) additionally suggest distinguishable neural representations.
282 Results remained significant after correcting for multiple comparisons using Bonferroni
283 correction.

284

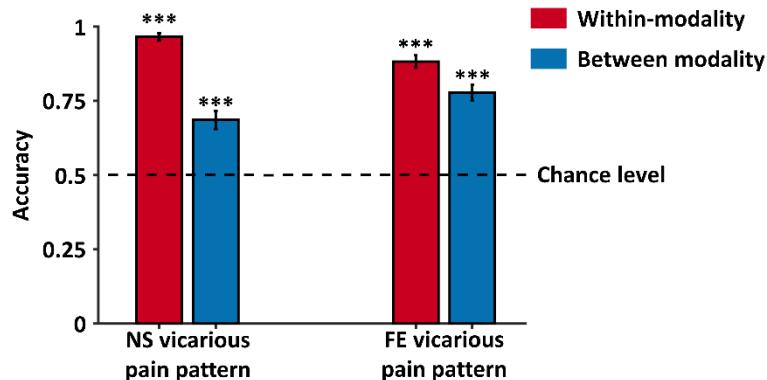
A Spatial similarity between whole-brain NS and FE vicarious pain patterns



B Joint distribution of normalized whole-brain weights of NS and FE vicarious pain patterns



C Cross-validated accuracy in two-choice classification tests (whole-brain patterns)



285

286 **Figure 3.** Results from the whole-brain multivariate pattern analyses. (A) The NS vicarious
287 pain-predictive pattern was spatially correlated with the FE vicarious pain-predictive pattern. (B)
288 Scatter plot displaying normalized voxel weights for NS (y-axis) and FE (x-axis) vicarious
289 pain-predictive patterns. Bars on the right represent the sum of squared distances from the origin (0,0)
290 for each Octant. Different colors are assigned to the eight Octants that reflect voxels of shared
291 positive or shared negative weights (Octants 2 and 6, respectively), selectively positive weights for
292 the NS (Octant 1) or for FE (Octant 3) vicarious pain patterns, selectively negative weights for the
293 NS (Octant 5) or FE (Octant 7) vicarious pain patterns, and voxels with opposite weights for the two
294 neural signatures (Octants 4 and 8). (C) Cross-validation accuracy as determined by two-alternative
295 forced-choice classification tests based on the whole-brain patterns. The results demonstrated
296 significant within- and between- modality classifications for both NS and FE vicarious
297 pain-predictive patterns. The dashed line indicates the chance level (50%), and the error bars
298 represent the standard error of the mean across subjects. *** $P < 0.001$. SSD, sum of squared distances.
299 Error bar indicates standard error. NS vicarious pain, observation of noxious stimulation of body

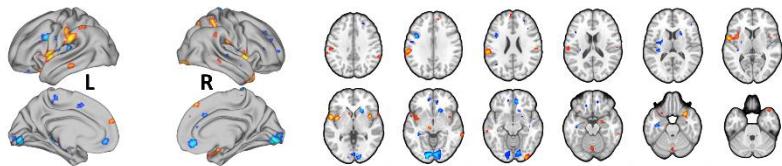
300 limbs induced vicarious pain; FE vicarious pain, observation of facial expressions of pain
301 induced pain.

302

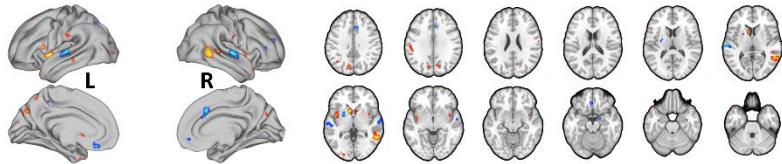
303 **Shared local representations for NS and FE vicarious pain**

304 To identify brain regions which made reliable contributions to both whole-brain NS and FE
305 vicarious pain-predictive patterns we thresholded the corresponding neural patterns at FDR $q <$
306 0.05 (two-tailed, bootstrap tests with 10,000 iterations) separately and found overlapping regions
307 in the bilateral mid-insula, left putamen and left inferior parietal lobule (**Figure 4A, 4B and 4C**),
308 emphasizing the importance of these regions for encoding both NS and FE vicarious pain.
309 Moreover, we employed a searchlight-based approach to locate regions which could predict both
310 within-modality vicarious pain (e.g., NS vicarious pain-predictive patterns to predict NS
311 vicarious pain versus NS control with a cross-validation procedure) as well as between-modality
312 vicarious pain (e.g., NS vicarious pain-predictive patterns to predict FE vicarious pain versus FE
313 control) using a cross-validation procedure. We found that a bilateral network encompassing the
314 insula, striatum as well as the ventromedial prefrontal cortex (see **Figure 4D**, $q < 0.05$, FDR
315 corrected, two-tailed) demonstrated significant within-modality cross-validation and
316 between-modality cross-prediction accuracies between NS and FE vicarious pain, implying
317 shared representation at the local pattern level. We additionally re-ran searchlight analyses with
318 two different searchlight sizes (4mm- and 10mm-radius spheres) and found that the overlapping
319 vicarious pain networks remained robust across different searchlight sizes (details see **Figure 4 –**
320 **figure supplement 1**).
321

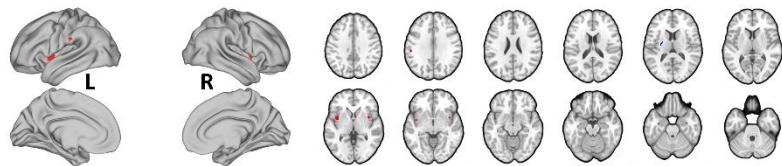
A NS vicarious pain-predictive pattern (FDR $q < 0.05$)



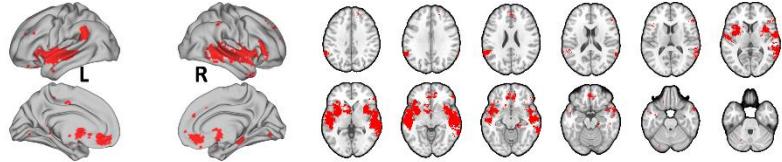
B FE vicarious pain-predictive pattern (FDR $q < 0.05$)



C Overlapping regions from whole-brain multivariate pattern analyses (FDR $q < 0.05$)



D Overlapping regions from the searchlight-based multivariate pattern analyses (FDR $q < 0.05$)



322

323

Figure 4. Brain regions that made reliable contributions to decoding vicarious pain. NS (A) and FE (B) vicarious pain-predictive patterns and (C) overlapping reliable predictive voxels (bootstrap thresholded at FDR $q < 0.05$, two-tailed). (D) Brain regions exhibiting significant within-modality cross-validation and between-modality cross-prediction accuracies between NS and FE vicarious pain (thresholded at FDR $q < 0.05$, two-tailed). NS vicarious pain, observation of noxious stimulation of body limbs induced vicarious pain; FE vicarious pain, observation of facial expressions of pain induced pain.

324

The following figure supplement is available for figure 4:

325

Figure supplement 1. Searchlight analyses with different searchlight sizes (**Figure 4-figure supplement 1**).

326

327 **Shared representations in the mid-insula**

335 Across the analyses we observed overlapping activation and shared representations in the
336 mid-insula (see also **Figure 4 – figure supplement 1** for convergent findings across searchlight
337 sizes). Accumulating evidence suggest a critical role of the mid-insula in pain-related processes,
338 including self-experienced as well as vicarious pain. In line with functional anatomical studies
339 suggesting that the mid-insula receives nociceptive information from thalamic nuclei (Craig et al.,
340 1994; Craig et al., 2000) intracerebral electrical stimulation of the mid-insula evokes pain
341 sensations (Afif et al., 2010) and previous MVPA studies demonstrated distinct neural
342 representations between pain and non-pain negative stimuli in the (right) mid-insula yet shared
343 representations across self-experienced and vicarious pain (Corradi-Dell'Acqua et al., 2011),
344 while a recent meta-analysis of conventional fMRI empathy studies reported that vicarious pain
345 uniquely activates the bilateral mid-insula and MCC as compared to empathy for non-pain
346 negative affective states (Timmers et al., 2018). Based on the specific role of the mid-insula in
347 pain-related processes we further explored whether the mid-insula shared neural representations
348 of NS and FE could be sufficient to predict vicarious pain. The mid-insula was defined based on
349 the Human Connectome Project (HCP) multi-modal parcellation atlas (Glasser et al., 2016)
350 (encompassing PoI2, FOP2, FOP3 and MI and available from the Cognitive and Affective
351 Neuroscience Laboratory Github repository at
352 https://github.com/canlab/Neuroimaging_Pattern_Masks; Figure 5 – figure supplement 1 displays
353 the mid-insula mask). We found that NS vicarious pain activation in the insula was strongly
354 positively correlated with FE vicarious pain activation ($r = 0.737, P = 0.006$ based on
355 permutation tests) and consistent with this, that the NS vicarious pain-predictive and FE

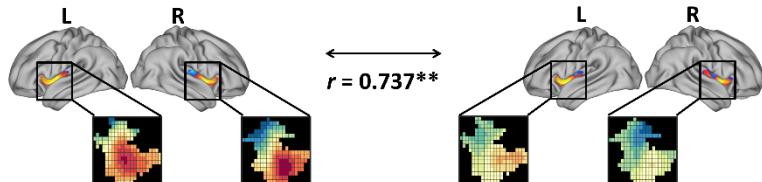
356 vicarious pain-predictive pattern weights within the mid-insula were also strongly positively
357 correlated ($r = 0.538$, $P < 0.001$ based on permutation tests) (**Figure 5A and 5B**). Moreover,
358 plotting the amount of shared positive, shared negative, and unique positive and negative voxel
359 weights (z-scored) within the mid-insula for NS and FE vicarious pain-predictive patterns
360 indicated that most voxels in the mid-insula exhibited shared positive weights (Octant 2) or
361 negative weights (Octant 6), whereas only few voxels exhibited opposite weights directions
362 (Octants 4 and 8) (**Figure 5C**). Consistent with the voxel-wise weight distribution
363 two-alternative forced-choice tests revealed that the mid-insula partial NS vicarious
364 pain-predictive pattern classified above chance for both, NS vicarious pain versus NS control
365 ($71 \pm 2.9\%$ SE, $P < 0.001$, $d = 0.72$; within-modality) and FE vicarious pain versus FE control
366 ($61 \pm 3.2\%$ SE, $P < 0.001$, $d = 0.36$; between-modality prediction) in out-of-sample participants
367 through a repeated 10-fold cross-validation procedure. In line with this, the mid-insula partial FE
368 vicarious pain-predictive pattern discriminated NS vicarious pain versus NS control with 65%
369 accuracy ($\pm 3.1\%$ SE, $P < 0.001$, $d = 0.58$; between-modality) and FE vicarious pain versus FE
370 control with 60% accuracy ($\pm 3.2\%$ SE, $P = 0.004$, $d = 0.27$; within-modality) (**Figure 5D**).
371 Together, these findings converge on common representations of vicarious pain in the
372 mid-insula across univariate and multivariate patterns for NS and FE vicarious pain. However,
373 although statistically significant, thus reflecting that the mid-insula plays important roles in
374 encoding NS and FE vicarious pain and that the neural representations of NS and FE vicarious
375 pain in this region are similar, the much lower effect sizes (as compared with the whole-brain
376 predictions) indicate that the mid-insula is not sufficient to capture vicarious pain processing alone.

377 Results remained significant after correcting for multiple comparisons using Bonferroni

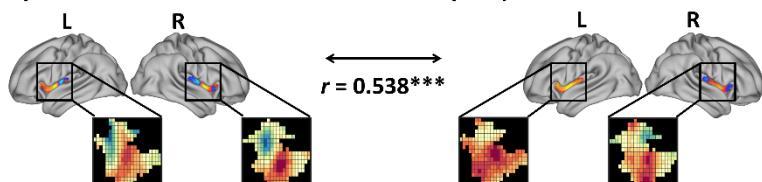
378 correction.

379

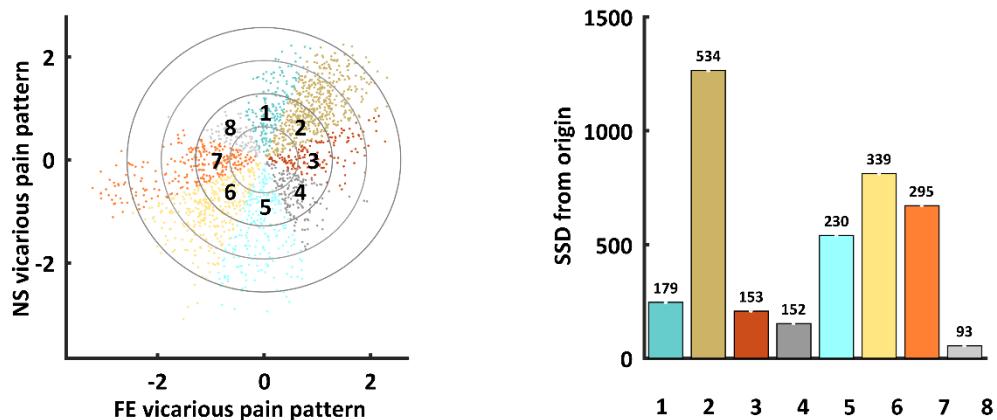
A Spatial similarity between mid-insula NS and FE vicarious pain activations



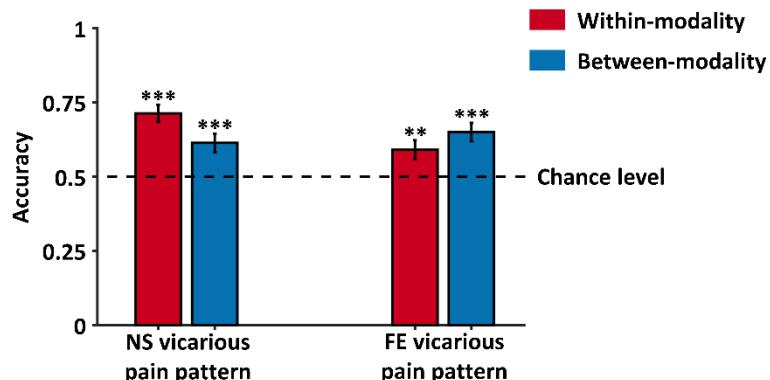
B Spatial similarity between mid-insula NS and FE vicarious pain patterns



C Joint distribution of normalized mid-insula weights of NS and FE vicarious pain patterns



D Cross-validated accuracy in two-choice classification tests (mid-insula patterns)



380

381 **Figure 5.** Results of the mid-insula focused analyses. (A) Mid-insula activation to NS vicarious pain
382 was highly similar to activation to FE vicarious pain. (B) NS vicarious pain-predictive pattern in the
383 mid-insula was spatially similar to the FE vicarious pain-predictive pattern. (C) Examining
384 voxel-level similarity in bilateral mid-insula revealed that that the majority of mid-insula voxels

385 exhibited shared positive or negative weights (Octants 2 and 6, respectively). Selective weights are
386 depicted as: selective positive weights for NS (Octant 1) and for FE (Octant 3) vicarious pain patterns,
387 selective negative weights for NS (Octant 5) and for FE (Octant 7) vicarious pain patterns. Voxels
388 with opposite weights for the two signatures are depicted in Octants 4 and 8. (D) Cross-validation
389 accuracy from the two-choice classification tests with mid-insula partial patterns. The results
390 demonstrated significant within- and between-modality classifications for both NS and FE vicarious
391 pain-predictive patterns. The dashed line indicates the chance level (50%), and error bars represent
392 standard error of the mean across subjects. ** $P < 0.01$; *** $P < 0.001$. SSD, sum of squared distances.
393 Error bar indicates standard error.

394 The following figure supplement is available for figure 5:

395 **Figure supplement 1.** The mid-insula mask used in the current study (**Figure 5-figure supplement**
396 **1).**

397

398 **Shared vicarious pain representations are not sensitive to arousal or negative affect**

399 One key question is whether the developed vicarious pain-predictive patterns are specific to the
400 vicarious sharing of pain or are rather generally sensitive to emotional arousal or negative affect.

401 To test the functional specificity whole-brain patterns were separately employed to discriminate
402 processing of high-arousal non-painful negative from low-arousal neutral stimuli from the IAPS
403 database with two-alternative forced-choice tests through a repeated 10-fold cross-validation
404 procedure. This approach revealed statistically significant yet comparably low accuracies and
405 small effect sizes (NS vicarious pain-predictive pattern: $58 \pm 3.2\%$, $P = 0.024$, $d = 0.34$; FE
406 vicarious pain-predictive pattern: $61 \pm 3.2\%$ SE, $P = 0.001$, $d = 0.42$). In contrast testing whether
407 shared representations in the mid-insula could discriminate negative versus neutral stimuli
408 revealed that neither of the insula partial patterns could classify negative stimuli above chance
409 level (NS: $56 \pm 3.2\%$ SE, $P = 0.079$, $d = 0.11$; FE: $56 \pm 3.2\%$ SE, $P = 0.111$, $d = 0.09$),
410 suggesting a pain-specific representation in this region.

411 In addition, using the emotional processing data, we developed a negative
412 emotion-predictive pattern which could accurately classify non-painful negative vs. neutral

413 stimuli (accuracy = $86 \pm 1.6\%$ SE, $p < 0.001$, $d = 2.07$ using repeated 10-fold cross-validation
414 procedures). The negative emotion-predictive pattern could significant discriminate NS
415 vicarious pain versus its control (cross-validated accuracy = $70 \pm 3.0\%$ SE, $P < 0.001$, $d = 0.88$)
416 and FE vicarious pain versus its control (cross-validated accuracy = $61 \pm 3.2\%$ SE, $P < 0.001$, d
417 = 0.28). Of note, accuracy and effect size are lower as compared to FE vicarious pain pattern's
418 prediction of NS vicarious pain (cross-validated accuracy = $78 \pm 2.7\%$ SE, $P < 0.001$, $d = 1.00$)
419 and vice versa (cross-validated accuracy = $69 \pm 3.0\%$ SE, $P < 0.001$, $d = 0.65$) and the
420 mid-insula negative-predictive pattern did not predict vicarious pain (accuracies = $40 \pm 3.2\%$ SE,
421 $48 \pm 3.2\%$ SE for NS and FE vicarious pain, respectively). Moreover, in contrast to the
422 pain-predictive patterns (see below for details) neither the whole-brain or the mid-insula
423 negative-predictive pattern could predict thermal pain intensity (whole-brain, $r_{196} = 0.101$, $P =$
424 0.157 ; mid-insula, $r_{196} = -0.319$), which additionally emphasizes the functional specificity of the
425 pain-predictive pattern in the domain of pain-related processing. Together these findings suggest
426 that negative emotional processing might share some neural representations with vicarious pain,
427 but that the whole-brain and mid-insula vicarious pain representations are more specific to the
428 pain-related information. Results remained stable after correcting for multiple comparisons using
429 Bonferroni correction.

430

431 **A vicarious pain-predictive pattern that predicts both NS and FE vicarious pain**

432 Given that the NS and FE vicarious pain-predictive patterns shared similar whole-brain as well
433 as local neural representations, we developed a general vicarious pain pattern which yielded a

434 classification accuracy of $82 \pm 1.2\%$ SE, $P < 0.001$, $d = 1.77$ (accuracy = $91 \pm 1.3\%$ SE, $P <$
435 0.001 , $d = 1.74$ based on a two-alternative forced-choice test) in discriminating vicarious pain
436 versus non-painful control. More specifically, the pattern could accurately predict both NS
437 vicarious pain from the NS control ($95 \pm 1.4\%$ accuracy, $P < 0.001$, $d = 2.10$) and FE vicarious
438 pain from the FE control ($87 \pm 2.1\%$ accuracy, $P < 0.001$, $d = 1.45$), but performed considerably
439 worse classifying non-painful negative versus neutral stimuli ($59 \pm 2.1\%$ accuracy, $P = 0.01$, $d =$
440 0.30), in forced-choice classifications. In line with the spatially overlapping modality-specific
441 vicarious pain patterns the general vicarious pain pattern was highly similar with both, the NS
442 vicarious pain pattern ($r = 0.587$, permuted $P < 0.001$ based on permutation tests) and FE
443 vicarious pain pattern ($r = 0.702$, $P < 0.001$ based on permutation tests). To functionally
444 characterize the general vicarious pain-predictive pattern the Neurosynth decoder function was
445 used to assess its similarity to the reverse inference meta-analysis maps generated for the entire
446 set of terms included in the Neurosynth dataset. The most relevant features were “painful” and
447 “pain” [for the top 50 terms (excluding anatomical terms) ranked by the correlation strengths
448 between the vicarious pain pattern and the meta-analytic maps (see word cloud, size of the font
449 scaled by correlation strength, **Figure 6A**). After thresholding and correction for multiple
450 comparisons (bootstrapping 10,000 samples, FDR $q < 0.05$, two-tailed), the general vicarious
451 pain-predictive pattern revealed a distributed network engaged in vicarious pain processing
452 encompassing the bilateral mid-insula, inferior parietal lobule and ventromedial prefrontal cortex
453 (**Figure 6B**), further emphasizing the importance of these regions for encoding vicarious pain.

454 All conclusions remained the same after controlling for multiple comparisons using the

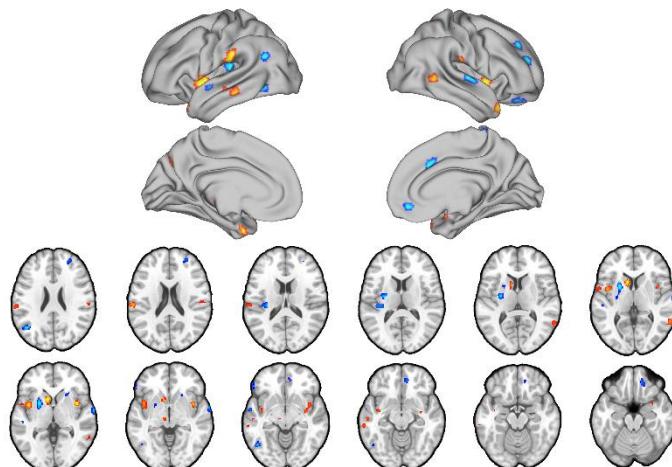
455 Bonferroni correction.

456

A Meta-analytic decoding of the general vicarious pain-predictive pattern



B The general vicarious pain-predictive pattern (FDR $q < 0.05$)



457

458 **Figure 6.** A general vicarious pain-predictive pattern which predicts both observation of noxious
459 stimulation of body limbs and facial expressions of pain induced vicarious pain. (A) Word cloud
460 showing the top 50 relevant terms (excluding anatomical terms) for the meta-analytic decoding of the
461 general vicarious pain-predictive pattern. The size of the font was scaled by correlation strength. (B)
462 When thresholded at FDR $q < 0.05$, two-tailed (bootstrapped 10,000 samples) the general vicarious
463 pain-predictive pattern revealed a distributed network of vicarious pain empathy representation
464 including bilateral mid-insula and ventromedial prefrontal cortex.

465

466 **Association of the vicarious pain-predictive pattern with self-experienced somatic pain**

467 To test the associations between the vicarious pain representation with directly experienced pain
468 we applied the whole-brain general vicarious pain-predictive pattern to self-experienced thermal
469 pain data using dot-product of vectorized activation maps with the pattern classifier weights. We
470 found that the general vicarious pain-predictive pattern expressions were highly correlated with
471 both overall objective temperature levels ($r_{196} = 0.538, P < 0.001$) and subjective pain ratings
472 ($r_{196} = 0.507, P < 0.001$). Moreover, the general pain-predictive pattern discriminated high
473 thermal pain versus low thermal pain with a 94% accuracy ($\pm 4.2\% \text{ SE}, P < 0.001, d = 2.00$),
474 high thermal pain versus medium thermal pain with a 91% accuracy ($\pm 5.0\% \text{ SE}, P < 0.001, d =$
475 1.56) and medium thermal pain versus low thermal pain with an 82% accuracy ($\pm 6.7\% \text{ SE}, P =$
476 0.001, $d = 1.20$) using two-alternative forced-choice tests (**Figure 7A**).

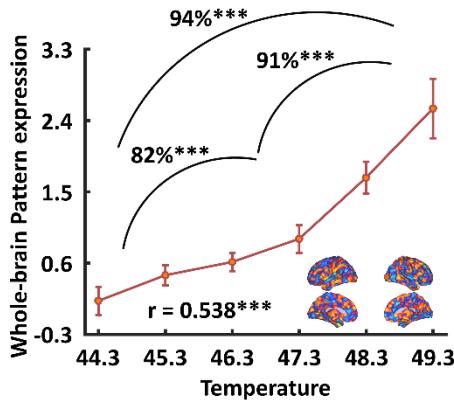
477 When prediction focused on the mid-insula the general vicarious pain-predictive local
478 pattern could discriminate high thermal pain versus low thermal pain (accuracy = $88 \pm 5.7\% \text{ SE}$,
479 $P < 0.001, d = 1.56$), high thermal pain versus medium thermal pain (accuracy = $88 \pm 5.7\% \text{ SE}$,
480 $P < 0.001, d = 1.23$) and medium thermal pain versus low thermal pain (accuracy = $82 \pm 6.7\%$
481 $\text{SE}, P < 0.001, d = 1.49$) above chance levels (**Figure 7B**). In addition, the mid-insula partial
482 pattern expressions (i.e., focusing on the mid-insula pattern) were highly correlated with
483 temperature levels ($r_{196} = 0.454, P < 0.001$) as well as individual pain ratings ($r_{196} = 0.440, P <$
484 0.001). Together with the predictions using NS and FE vicarious pain-predictive patterns
485 separately (**Figure 7 – figure supplement 1**), our results demonstrate that the vicarious pain
486 patterns respond to self-experienced somatic pain, confirming that the vicarious pain patterns

487 reflect pain-associated information. All findings remained significant after correcting multiple

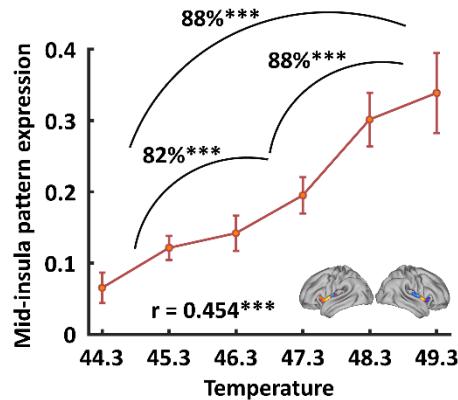
488 comparisons via Bonferroni correction.

489

A The general vicarious pain whole-brain pattern
expressions of thermal pain



B The general vicarious pain mid-insula
pattern expressions of thermal pain



490

491 **Figure 7.** Generalizability of the general (across NS and FE) vicarious pain-predictive pattern. Both
492 whole-brain (A) and mid-insula (B) vicarious pain-predictive patterns could accurately predict the
493 severity and classify the levels of self-experienced pain in an independent dataset. *** $P < 0.001$. Error
494 bar indicates standard error.

495 The following figure supplements are available for figure 7:

496 **Figure supplement 1.** Generalizability of the NS and FE vicarious pain-predictive patterns (**Figure**
497 **7-figure supplement 1**).

498 **Figure supplement 2.** Effects of sample size on prediction accuracy of thermal pain (**Figure**
499 **7-figure supplement 2**).

500

501 Discussion

502 Several studies have explored the neural underpinnings of vicarious pain in humans and

503 suggested overlapping univariate fMRI activity in the anterior cingulate and insular cortex

504 across different vicarious pain induction procedures (for meta-analyses see e.g., Jauniaux et al.,

505 2019; Timmers et al., 2018). However, the conventional univariate approach lacks anatomical

506 and functional specificity to test the question of whether vicarious pain across different

507 modalities shares common and process-specific neural representations (Iannetti et al., 2013;
508 Krishnan et al., 2016; Woo et al., 2014; Zaki et al., 2016). Here we employed a fine-grained
509 MVPA approach which is sensitive and specific to particular types of mental processes including
510 pain (Kragel et al., 2018; Peelen et al., 2006; Wager et al., 2013; Woo et al., 2017) to explore (1)
511 whether shared neural representations of vicarious pain can be determined across different
512 induction procedures (FE, NS) and (2) whether the shared neural representation is sensitive to
513 pain-unspecific components of the vicarious pain response (arousal, negative affect) and related
514 to the experience of somatic pain. We demonstrated that shared multivariate patterns encoding
515 NS and FE vicarious pain can be determined at the whole-brain level and that across different
516 analytic approaches the mid-insular cortex was consistently engaged across induction procedures.
517 Furthermore, we demonstrated that these patterns were not sensitive to respond to the processing
518 of non-painful high-arousal negative stimuli in the same sample, together with the findings
519 showing that NS vicarious pain predicted FE vicarious pain (and vice versa) more accurately as
520 compared with the predictions using a negative emotion decoder, suggesting that the common
521 vicarious pain representations do not simply reflect shared unspecific processes of negative
522 affect or arousal. Moreover, the shared vicarious pain representations predicted self-experienced
523 thermal pain in an independent sample, suggesting an association between the neural expression
524 and processes directly related to the experience of pain. Together these results provide evidence
525 for a generalized neural representation of vicarious pain, particularly in the mid-insula, and
526 demonstrated that the shared neural signature may specifically capture pain-associated aspects of
527 the vicarious pain response rather unspecific processes such as aversive experience or arousal.

528 The idea that vicarious pain across different induction procedures shares common neural
529 representations has been supported by meta-analyses covering previous fMRI pain empathy
530 studies that demonstrated overlapping activation in the insular and cingulate cortex (Jauniaux et
531 al., 2019; Timmers et al., 2018). In line with these meta-analytic findings we found that these
532 regions were consistently engaged during both NS and FE vicarious pain. However, the insular
533 and anterior cingulate cortex are involved in a wide range of mental processes including
534 representation of interoceptive and affective states as well as salience detection (Craig & Craig,
535 2009; Critchley et al., 2004; Timmers et al., 2018; Uddin, 2015), suggesting that the overlapping
536 activity might be due to common underlying mental processes such as detecting and orienting
537 attention towards salient stimuli or unspecific emotional arousal (Corradi-Dell'Acqua et al., 2011;
538 Corradi-Dell'Acqua et al., 2016; Valentini & Koch, 2012).

539 To systematically test whether vicarious sharing of pain elicited by different social cues
540 shares common neural representations we developed and compared multivariate patterns that
541 predicted NS and FE vicarious pain evoking stimuli respectively. While mass-univariate analysis
542 results reflect the presence of intermingled neuronal populations related to stimulus-specific
543 representations, MVPA investigates whether idiosyncratic spatial variations in the fMRI signal
544 are shared or dissociated across different conditions and thus might be more suitable to
545 determine process-specific representations in meso-scale neural circuits (Kamitani & Tong, 2005;
546 Kriegeskorte et al., 2006; Peelen & Downing, 2007). Moreover, previous studies suggest that
547 whole-brain predictive models could better capture emotional processes compared to regional
548 approaches, such as decoding of a single brain region or searchlight-based methods (Kragel et al.,

549 2018; Woo et al., 2017). To this end, we first identified whole-brain fMRI patterns that
550 accurately predicted NS and FE vicarious pain, respectively. We found that the NS and FE
551 vicarious pain-predictive patterns were spatially correlated and both could classify within- and
552 between-modality painful versus non-painful stimuli at the whole-brain level, suggesting that NS
553 and FE vicarious pain share distributed processing across multiple systems and component
554 processes. In line with previous studies demonstrating that while NS vicarious pain provides
555 objective cues about the sensory component of the observed pain the FE vicarious lacks such
556 information and is more subjective and indirect as the pain experience of the expresser need to
557 be interpreted by the observer (Hadjistavropoulos et al., 2011; Vachon-Presseau et al., 2012), the
558 decreased accuracies and effect sizes in the cross-modality predictions additionally suggest
559 partly distinguishable neural representations of NS and FE vicarious pain possibly reflecting the
560 engagement of different component processes.

561 In the context of previous studies suggesting that pain empathy deficits are mediated by
562 regional-specific brain lesions and functional dysregulations (Leigh et al., 2013; Shamay-Tsoory
563 et al., 2009; Xu et al., 2019) the question for the contribution of specific brain regions arises.
564 Thresholding the vicarious pain patterns (at FDR $q < 0.05$, two-tailed) allowed us to identify
565 voxels that reliably contributed to the respective decoders and revealed that specifically the
566 bilateral mid-insula provided important features to predict both NS and FE vicarious pain.
567 Moreover, the mid-insula partial vicarious pain patterns were highly spatially correlated and
568 both could significantly predict within- and between-modality vicarious pain-related experience.
569 Consistent with this, searchlight-based classification analyses also demonstrated that mid-insula

570 local patterns produced significant within- and between-modality predictions of vicarious pain.
571 Our results are in line with a previous meta-analysis showing that the mid-insula responds
572 specifically to empathy for pain across different task paradigms compared to empathy for
573 non-pain negative affective states (Timmers et al., 2018), which together with the present
574 findings suggests that the mid-insula represents a core neural substrate for vicarious pain.

575 Although multivariate predictive models can capture information at much finer spatial
576 scales and consequently anatomical specificity (Kamitani & Tong, 2005; Woo et al., 2017) the
577 question of the specific mental processes captured by our vicarious pain predictive patterns
578 remains unclear. Pain empathy is a multi-component process that includes the vicarious sharing
579 of pain but may also evoke emotional arousal and negative affect and these unspecific processes
580 can be captured by the decoders. To determine the functional specificity of the neural
581 representations we applied the vicarious pain-predictive patterns to data from an emotion
582 processing paradigm acquired in the same sample as well as to data from a thermal pain
583 induction experiment in an independent sample and found that (1) the vicarious pain patterns
584 performed only modest for discriminating high-arousal (non-painful) negative stimuli from
585 low-arousal neutral stimuli, and (2) the whole-brain and mid-insula patterns predicted levels of
586 self-experienced thermal pain with high accuracies. Finally, we developed a general vicarious
587 pain-predictive pattern across NS and FE vicarious pain induction procedures and demonstrated
588 that it accurately predicted both NS and FE vicarious pain (accuracies > 87%) as well as thermal
589 pain intensities (accuracies > 82%), yet classified non-painful negative versus neutral stimuli
590 with comparably low accuracy (59%). In line with the prediction results, meta-analytic decoding

591 analysis revealed that this general vicarious pain pattern was highly correlated with the domains
592 of “painful” and “pain”, but not with “arousal”, “valence” or “negative” (not shown in the top
593 100 relevant terms). Together these findings suggest a shared neural representation of vicarious
594 pain and a high-specificity of the whole-brain and specifically the mid-insula patterns for the
595 vicarious experience of pain. A previous study developed a vicarious pain signature (VPS) that
596 was sensitive and specific to NS vicarious pain, but not sensitive to the intensity of
597 self-experienced somatic pain (Krishnan et al., 2016). Examining similarities with our general
598 vicarious pain-predictive pattern revealed only modest spatial correlations between the two
599 patterns ($r = 0.04$). The different instructions employed in the experiments might have
600 contributed to the low overlap such that participants in the previous study were required to
601 explicitly rate their emotional response to the stimuli whereas we decided for an implicit
602 processing (passive viewing) paradigm across the vicarious pain and negative emotional
603 processing paradigm and additionally to control for cognitive processes which can modulate
604 empathic reactivity and painful experience as well as the specific neural networks engaged
605 (Jauniaux et al., 2019; Urien & Wang, 2019). Moreover, we found that the present pattern could
606 successfully predict pain experience during thermal heat stimulation while the VPS was not
607 sensitive to self-experienced pain. The observed differences might be explained in terms of (1)
608 the considerably larger sample size included in the present study and prediction accuracy (as
609 reflected by prediction-outcome correlation) of self-experienced pain experience increased as a
610 function of sample size used to develop the NS vicarious pain decoder (see additional analysis
611 presented in **Figure 7 – figure supplement 2**), and (2) differences between paradigms and

612 instructions such that, for example, a recent meta-analysis of empathy for pain studies showed
613 that the mid-cingulate gyrus was more activated by explicit cognitive/evaluative paradigms
614 while the right inferior frontal gyrus and anterior insula were more activated by implicit
615 perceptual/affective paradigms (Timmers et al., 2018).

616 Our results highlighted the mid-insula as a key region sharing similar neural representations
617 across NS and FE vicarious pain suggesting that it may contribute to the core vicarious pain
618 experience that characterizes pain empathy. Consistent with whole-brain results, the shared
619 information in the mid-insula was specific to vicarious pain rather than negative affect or arousal.

620 Previous non-human primate and human studies indicate that the posterior and mid-insula
621 receive nociceptive information from thalamic nuclei (Craig et al., 1994; Craig et al., 2000)
622 which are in turn conveyed to the anterior insula for progressive integration with higher level
623 affective and interoceptive experience (Corradi-Dell'Acqua et al., 2011; Corradi-Dell'Acqua et
624 al., 2016; Singer et al., 2009). Although overarching models of the neural basis and neuroimaging
625 meta-analysis (Jauniaux et al., 2019; Timmers et al., 2018) emphasize the role of the anterior
626 insula in pain empathy processing, accumulating evidence from studies examining shared and
627 process-specific representations of vicarious pain suggest a specific role of the mid-insula in
628 vicarious pain (Corradi-Dell'Acqua et al., 2011; Krishnan et al., 2016), whereas the (left) anterior
629 insula also responded to negative stimuli in general (Corradi-Dell'Acqua et al., 2011) and across
630 modalities (Corradi-Dell'Acqua et al., 2016). Importantly, the peak anterior insula coordinates
631 identified in these previous studies did not overlap with our mid-insula mask or the mid-insula
632 region that exhibited reliable predictive features in both NS and FE vicarious pain whole-brain

633 patterns determined in the present study, suggesting a more specific role of the mid-insula in
634 pain-related components of the vicarious pain response (see also recent meta-analysis by Timmers
635 et al., 2018 demonstrating a specific role of the mid-insula in pain empathy). In support of our
636 findings a previous study employed a similar whole-brain MVPA approach to predict NS
637 vicarious pain induced by a evaluative paradigm also identified bilateral mid-insula as reliable (q
638 < 0.05 , FDR corrected) predictive regions (Krishnan et al., 2016), further conforming the
639 reliable contribution of this region in encoding vicarious pain. Studies examining the functional
640 and anatomical organization of the insular cortex with intracerebral electrical stimulation have
641 demonstrated that painful sensations can be elicited by stimulation of the middle but not the
642 anterior insula (Affif et al., 2010). Together with the functional relevance of the mid-insula to
643 predict objective and subjective pain experience in an independent sample and the contribution of
644 this region to nociception as well as vicarious pain (Botvinick et al., 2005; Krishnan et al., 2016;
645 Lamm et al., 2011; Timmers et al., 2018; Wager et al., 2013) our findings suggest that the shared
646 representations in the mid-insula across vicarious pain induction procedures may specifically code
647 the automatic pain sharing which resonates with embodies conceptualizations of vicarious pain
648 (see e.g., Corradi-Dell'Acqua et al., 2011 for a convergent interpretation). However, consistent
649 with previous evidence that (NS) vicarious pain representation is distributed across brain regions
650 and single local regions exhibit considerably lower effect sizes compared to whole-brain
651 predictive models (Krishnan et al., 2016) we found that the prediction effect sizes for the
652 mid-insula were smaller than those observed in our whole brain analyses. These findings suggest

653 that despite the key role of the mid-insula in vicarious pain experience this region is not
654 sufficient to fully capture this process.

655 Consistent with previous studies suggesting that the anterior cingulate cortex represents a
656 core brain region for emotional empathy in general and pain empathy in particular (Fan et al.,
657 2011; Jauniaux et al., 2019; Timmers et al., 2018), we found overlapping deactivations in the
658 rostral and ventral anterior cingulate cortex in mass-univariate analyses and shared patterns in
659 the dorsal, rostral and ventral anterior cingulate cortex in searchlight-based prediction analyses
660 between NS and FE vicarious pain. However, no overlapping reliable predictive voxels for
661 whole-brain NS and FE pain-predictive patterns were found in cingulate regions suggesting a
662 differential involvement of this region during FE and NS vicarious pain induction procedures.

663 From a methodological perspective these results may reflect that the whole-brain predictive
664 model could provide a more specific neural description of a behavior or mental process (Kragel
665 et al., 2018; Woo et al., 2017). In line with our findings previous studies also show that
666 significant activation and searchlight-based prediction in local regions do not necessarily imply
667 reliable predictive features in whole-brain predictive models (Krishnan et al., 2016; Woo et al.,
668 2014). From a brain systems perspective these findings may indicate that the anterior cingulate
669 cortex is not specifically involved in vicarious pain elicited across induction procedures.

670 Although the anterior cingulate cortex has been reliably identified in meta-analytic studies
671 covering brain activation patterns during (pain) empathy induction procedures (see e.g.,
672 Jauniaux et al., 2019; Timmers et al., 2018) it has also been associated with a number of basal

673 processes, including arousal and salience, and activation in this region may reflect rather
674 unspecific neural responses.

675 The present study has limitations that should be addressed in future studies. Compared to
676 the homogeneous stimuli within the conditions of the vicarious pain and the self-experienced
677 pain paradigm the stimuli displaying emotional evocative scenes from the IAPS database may
678 have led to a higher inter-trial variance in the negative processing experiment. Although the
679 inter-stimulus variance should not systematically differ between the experimental conditions
680 employed to develop the corresponding decoder we cannot fully exclude that this may have
681 partly contributed to the low accuracies of the emotional processing decoder with respect to
682 predicting self-experienced pain ratings. Moreover, the current study employed a passive
683 observation paradigm and a recent meta-analysis revealed that vicarious pain induced by
684 cognitive/evaluative and affective/perceptual paradigms elicited activations in overlapping yet
685 also different brain regions (Timmers et al., 2018). Whether the present conclusions could
686 generalize to more “active” engagements in empathy (e.g., explicitly asking subjects to imagine
687 that the injury occurring in the picture displayed was happening to them) remains to be
688 determined.

689 In conclusion, by applying a novel whole-brain as well as local-region based MVPA
690 approaches in a large sample of healthy adults, our results provide the first neuroimaging
691 evidence that NS and FE vicarious pain shares common neural representations, especially in the
692 mid-insula which may specifically encode the vicarious sharing of pain that specifically
693 characterizes pain empathy. Moreover, we also provide a general vicarious pain-predictive

694 pattern (across NS and FE vicarious pain stimuli), which may be employed in future studies to
695 facilitate inferences about pain empathy across modalities as well as self-experienced pain. Our
696 study offers a new approach to better understand pain empathy by exploring common neural
697 representations and linking these shared representations to felt pain.

698

699 **Methods**

700 **Key Resources Table**

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
software, algorithm	Matlab R2015b	MathWorks	RRID: SCR 001622	
software, algorithm	SPM12	Wellcome Trust Centre for Neuroimaging	RRID:SCR_007037	
software, algorithm	CANLab Core Tools	CANlab	https://github.com/canlab	
Other	Thermal pain data	Wager et al. (2013)	https://ndownloader.figshare.com/files/12708989	
Other	Vicarious pain signatures	This paper	https://neurovault.org/collections/6332/	Deposited multivariate patterns
Other	Data and codes	This paper	https://figshare.com/articles/Vicarious_pain_dataset/11994498	Deposited fMRI data and scripts for figures

701

702 **Participants**

703 N = 252 healthy young participants were enrolled in the current study and underwent a previously
704 validated NS and FE vicarious pain empathy fMRI paradigm. The fMRI data on the basic group
705 activation maps for NS and FE vicarious pain contrasts were previously published in a study
706 examining dimensional associations with trait autism and alexithymia (Li et al., 2018) and a study
707 investigating network-level communication during pain empathic processing using an exploratory
708 inter-subject phase synchronization approach (Xu et al., 2020). Of note, the aim, methodological
709 approach and hypotheses of the current study were independent from these previous publications;
710 here we focus on identifying an fMRI multivariate pattern for NS and FE vicarious pain separately
711 and assessing their relationship. To further examine the specificity of the determined pain patterns
712 from general negative emotion processing the data from an emotion processing paradigm from the
713 same subjects was additionally used. Due to technical issues during data acquisition (incomplete
714 data, n = 6), left-handedness (n = 4) or excessive head motion (> 3 mm translation or 3° rotation; n
715 = 4) data from 14 participants were excluded leading to a sample of n = 238 participants (118
716 females; mean ± SD age = 21.58 ± 2.32 years) for the pain empathy analyses; data from 15
717 participants (incomplete data n = 8; left-handedness, n = 4; excessive head motion, n = 3) was
718 excluded from the emotion processing paradigm analyses (n = 237; 120 females; mean ± SD age =
719 21.55 ± 2.30 years). Participants provided written consent, the study was approved by the ethics
720 committee at the University of Electronic Science and Technology and was in accordance the
721 Declaration of Helsinki. Consent authorization for publication has been obtained from individuals
722 in Figure 1.
723

724 **Experimental stimuli**

725 The main aim of the present study was to determine (1) shared neural representations of pain
726 empathy and (2) to further differentiate the specificity of the neural representation of shared
727 vicarious pain from unspecific arousal and negative processing. For aim (1) we employed two
728 different sets of validated pain empathy experimental stimuli displaying noxious stimulation of
729 body limbs (NS vicarious pain) and facial expressions of pain (FE vicarious pain) as well as
730 respective non-painful control stimuli (see **Figure 1A** for examples). The NS vicarious stimuli
731 displayed a person's hand or foot in painful or non-painful everyday situations from the first-person
732 perspective (e.g., the painful stimulus displays cutting a hand with a knife whereas the matched
733 non-painful control stimulus shows cutting vegetables with a knife; for an evaluation of the stimuli
734 see also Meng et al., 2012). The FE vicarious stimuli incorporated painful and neutral facial
735 expressions from 16 Chinese actors (8 males; for an evaluation of the stimuli see also Sheng &
736 Han, 2012). To further validate the stimulus properties we recruited an independent sample of 40
737 subjects (two of them were excluded due to incomplete data; 17 females; Mean \pm SD age = 20.45 \pm
738 1.43 years) to rate the intensity of pain the depicted person is experiencing, the intensity of
739 (vicarious) pain they experience while seeing the picture, valence and arousal for each stimulus on
740 nine-point Likert scales (1 = "not painful at all", "very negative" or "lowly arousing", 9 =
741 "extremely painful", "very positive" or "highly arousing"). In line with previous studies
742 employing these stimulus sets (Meng et al., 2012; Sheng & Han, 2012) ratings in the present
743 sample confirmed that both sets of painful stimuli were rated as considerably more painful in
744 terms of the perceived level of pain the person in the picture is experiencing as well as level of

745 vicarious pain experience in the observer (all P s < 0.001). As expected, both sets of painful stimuli
746 were also rated as more negative and stronger arousing than the control stimuli (all P s < 0.001)
747 (details see **Figure 1B** and **Results**). To determine whether the shared higher arousal and negative
748 affect of both painful stimuli relative to their control stimuli may have contributed to the identified
749 shared neural expression (aim 2) we additionally employed a stimulus set with non-painful
750 high-arousal negative pictures and low-arousal neutral control stimuli. All stimuli were from the
751 International Affective Picture System (IAPS) database. We recruited another independent sample
752 of 37 subjects (16 females; Mean \pm SD age = 23.60 ± 2.86 years) to rate the valence and arousal for
753 each stimulus with nine-point Likert scales (1 = “very negative” or “lowly arousing”, 9 = “very
754 positive” or “highly arousing”). Given that the IAPS stimuli we selected were non-painful we did
755 not ask subject to rate pain intensity. Negative stimuli elicited substantial negative affect and arousal
756 on numerical rating scales as compared with neutral stimuli (details see **Results**).
757

758 **Presentation of the stimuli**

759 The pain empathy paradigm employed a blocked design incorporating condition-specific blocks
760 presenting the validated visual stimuli displaying painful everyday scenes (NS vicarious pain) and
761 painful facial expressions (FE vicarious pain) as well as modality-specific control stimuli
762 displaying non-painful scenes (NS control) or neutral facial expressions (FE control). A total of 16
763 blocks (4 blocks per condition) were presented in a pseudo-randomized order and interspersed by a
764 jittered red fixation cross (8, 10, or 12s). Each block (16s) incorporated four condition-specific
765 stimuli (each presented for 3s) separated by a white fixation cross (1s). An implicit processing

766 paradigm (passive viewing) was employed. To this end participants were instructed to attentively
767 watch the presented stimuli.

768 In line with the pain empathy paradigm the emotion processing paradigm employed a block
769 design incorporating three experimental conditions (positive, negative and neutral pictures). A
770 total of 19 blocks (neutral, 7 blocks; negative, 6 blocks; positive, 6 blocks) were presented in a
771 pseudo-randomized order and interspersed by a jittered red fixation cross (8, 10, or 12s). In each
772 block (16s) four condition-specific stimuli (3s) were presented and separated by a white fixation
773 cross (1s). An implicit processing (passive viewing) paradigm was employed and participants
774 were asked to attentively watch the stimuli. To ensure attentive processing, participants were
775 required to press a button when a stimulus with a white frame (one in each block) was presented.

776

777 **Thermal pain paradigm**

778 Thirty-three healthy (22 females; mean \pm SD age = 27.9 ± 9.0 years), right-handed subjects
779 participated in the thermal pain study (details see Wager et al., 2013; Woo et al., 2015). Six levels
780 of temperature (ranging from 44.3 to 49.3°C in increments of 1°C) were delivered to the volar
781 surface of the left inner forearm using a TSA-II Neurosensory Analyzer (Medoc Ltd.) with a
782 16-mm Peltier thermode during fMRI acquisition. The fMRI task included 7 passive
783 experience runs and 2 regulation runs where subjects were asked to cognitively “increase”
784 (regulate-up) or “decrease” (regulate-down) pain intensity with each run encompassing 11 trials.
785 Each trial consisted of a 12.5s stimulus (3s ramp-up and 2s ramp-down periods and 7.5s at the
786 target temperature), a jittered 4.5-8.5 sec delay, a 4 sec painful/non-painful decision period, a 7s

787 continuous warmth or pain rating period (on a visual analogue scale) and 23-27s rest. For the
788 current study we incorporated the data from the passive experience runs.

789

790 **MRI data acquisition and preprocessing**

791 MRI data were collected on a 3.0-T GE Discovery MR750 system (General Electric Medical
792 System, Milwaukee, WI, USA). Functional MRI data was acquired using a T2*-weighted
793 echo-planar imaging (EPI) pulse sequence (repetition time = 2s, echo time = 30ms, 39 slices, slice
794 thickness = 3.4mm, gap = 0.6mm, field of view = 240 × 240mm, resolution = 64 × 64, flip angle =
795 90°, voxel size = 3.75 × 3.75 × 4mm). To improve spatial normalization and exclude participants
796 with apparent brain pathologies a high-resolution T1-weighted image was acquired using a 3D
797 spoiled gradient recalled (SPGR) sequence (repetition time = 6ms, echo time = minimum, 156 slices,
798 slice thickness = 1mm, no gap, field of view = 256 × 256mm, acquisition matrix = 256 × 256,
799 flip angle = 9°, voxel size = 1 × 1 × 1mm). OptoActive MRI headphones
800 (<http://www.optoacoustics.com/>) were used to reduce acoustic noise exposure for the participants
801 during MRI data acquisition.

802 Functional MRI data was preprocessed using Statistical Parametric Mapping (SPM12;
803 RRID:SCR_007037; <https://www.fil.ion.ucl.ac.uk/spm/software/spm12/>). The first 10 volumes of
804 each run were discarded to allow MRI T1 equilibration and active noise cancelling by the
805 headphones. The remaining volumes were spatially realigned to the first volume and unwarped to
806 correct for nonlinear distortions related to head motion or magnetic field inhomogeneity. The
807 anatomical image was segmented into grey matter, white matter, cerebrospinal fluid, bone, fat and

808 air by registering tissue types to tissue probability maps. Next, the skull-stripped and bias
809 corrected structural image was generated and the functional images were co-registered to this
810 image. The functional images were subsequently normalized the Montreal Neurological Institute
811 (MNI) space (interpolated to $2 \times 2 \times 2$ mm voxel size) by applying the forward deformation
812 parameters that were obtained from the segmentation procedure, and spatially smoothed using an
813 8-mm full-width at half maximum (FWHM) Gaussian kernel.

814

815 **Pain empathy - univariate general linear model (GLM) analyses**

816 A two-level random effects GLM analysis was conducted on the fMRI signal to determine shared
817 modality-specific activation patterns using a mass-univariate GLM approach. The first-level
818 model included four condition-specific (NS vicarious pain, NS control, FE vicarious pain, and
819 FE control) box-car regressors logged to the first stimulus presentation per block that were
820 convolved with SPM12's canonical hemodynamic response function (HRF). The fixation cross
821 epoch during the inter-block interval served as implicit baseline, and a high-pass filter of 128
822 seconds was applied to remove low frequency drifts. Regressors of non-interest (nuisance
823 variables) included (1) six head movement parameters and their squares, their derivatives and
824 squared derivatives (leading to 24 motion-related nuisance regressors in total); and (2) motion
825 and signal-intensity outliers (based on Nipype's rapidart function). Single-subject voxel-wise
826 statistical parametric maps for the empathy modality-specific contrasts (NS vicarious pain > NS
827 control and FE vicarious pain > FE control) were obtained and subjected to group-level
828 one-sample t-tests. The corresponding analyses were thresholded and corrected for multiple

829 comparisons within a grey matter mask based on false discovery rate (FDR $q < 0.05$, two-tailed)
830 with a minimum extent of 100mm³. The resulting thresholded activation maps were next used to
831 identify common regions of activation across the modalities (NS and FE vicarious pain; i.e.,
832 masking the overlapping significant voxels).

833 To determine the activation similarity of NS and FE vicarious pain a permutation-based
834 correlation analysis was employed (Hong et al., 2019). Specifically, we (1) calculated Pearson's
835 correlation (r) between the modality-specific unthresholded statistical maps (NS vicarious pain >
836 NS control versus FE vicarious pain > FE control), (2) shuffled the condition labels for the NS
837 stimuli, obtained a new group-level statistical map for "NS vicarious pain > NS control" and
838 calculated the activation similarity of FE and the "modelled" NS vicarious pain, (3) repeated
839 step (2) 10,000 times, (4) repeated steps (2-3) with shuffled labels for FE instead of NS stimuli,
840 and finally (5) calculate the probability of observing the activation similarity between the true
841 NS and FE pain given the null distribution of permuted activation similarity. A P value < 0.05
842 was being considered statistically significant and between 0.05 - 0.1 was being considered as
843 marginal significant.

844

845 **Pain empathy - multivariate pattern analyses**

846 For the multivariate pattern analyses, nuisance regression (24 head motion parameters, motion
847 and signal-intensity outliers, and linear trend) and high-pass filtering (cut off at 128s) were
848 initially simultaneous conducted on the preprocessed fMRI data. Next, the fMRI signal was
849 averaged within the 4 condition-specific blocks (shifted by 3 TRs to account for the delay of the

850 HRF). In line with previous studies (e.g., Krishnan et al., 2016; Wager et al., 2013; Woo et al.,
851 2014) we used normalized and smoothed (8-mm FWHM Gaussian kernel) data to develop the
852 population-level vicarious pain-predictive patterns as previous studies suggested that this
853 smoothing level could improve inter-subject functional alignment while retaining sensitivity to
854 mesoscopic activity patterns that are consistent across subjects (de Beeck, 2010; Shmuel et al.,
855 2010). Linear support vector machines (SVMs, $C = 1$) were then employed to the whole-brain
856 maps (restrict to a grey matter mask) to train multivariate pattern classifiers on the cleaned
857 averaged fMRI signal to discriminate NS vicarious pain versus NS control and FE vicarious pain
858 versus FE control separately. The classification performance was evaluated by a 10-fold
859 cross-validation procedure during which all participants were randomly assigned to 10
860 subsamples of 23 or 24 participants using MATLAB's cvpartition function. The optimal
861 hyperplane was computed based on the multivariate pattern of 214 or 215 participants (training
862 set) and evaluated by the excluded 24 or 23 participants (test set). The training set was linearly
863 scaled to $[-1, 1]$, and the test set was next scaled using the same scaling parameters before
864 applying SVM (Hsu et al., 2003). This procedure was repeated 10 times with each subsample
865 being the testing set once. To avoid a potential bias of training-test splits, the cross-validation
866 procedures throughout the study were repeated 10 times by producing different splits in each
867 repetition and the resultant accuracy and P values were averaged to produce a convergent
868 estimation (Zhou et al., 2018). In line with the mass-univariate analyses and to identify which
869 brain regions made reliable contributions to the decoders (Wager et al., 2013; Zhou et al., 2019),
870 the pattern maps were thresholded at FDR $q < 0.05$ (two-tailed) with a minimum extent of

871 100mm³ using bootstrap procedures with 10,000 samples. Next the thresholded maps were
872 subjected to a conjunction analysis to identify regions that robustly contributed to both NS and
873 FE vicarious pain classifiers by masking overlapping significant voxels. Statistical maps were
874 visualized using the Connectome Workbench provided by the Human Connectome Project
875 (<https://www.humanconnectome.org/software/connectome-workbench>).

876 Similarity patterns between the modality-specific neural patterns were determined
877 employing (1) Pearson's correlation between the whole-brain unthresholded classifier weights
878 using a permutation test (similar to the activation similarity analysis), and (2) "between -
879 modality classification" tests encompassing the following two steps: (a) pattern classifiers were
880 trained separately for NS vicarious pain versus NS control and FE vicarious pain versus FE
881 control with a 10-fold cross-validation procedure (repeated 10 times), and next (b) applying the
882 identified patterns of NS and FE vicarious pain to out-of-sample participants for the FE
883 vicarious pain versus FE control and NS vicarious pain versus NS control respectively using a
884 two-alternative forced choice test, where pattern expression values were compared for two
885 conditions with the image exhibiting the higher expression being determined as pain.

886

887 **Pain empathy – within- and between- modality classification analyses employing local**
888 **classifiers**

889 To further identify regions with shared neural expressions across NS and FE vicarious pain, a
890 local pattern-based classification approach with three-voxel radius spherical searchlights around
891 center voxels was employed (Corradi-Dell'Acqua et al., 2011; Kriegeskorte et al., 2006; Woo et

892 al., 2014). Specifically, (1) multivariate pattern classifiers using a defined local region were
893 trained to discriminate vicarious pain versus control within each modality (i.e., NS and FE
894 stimuli) separately, (2) the patterns obtained were next applied to out-of-sample participants for
895 within-modality cross-validation and between-modality cross-prediction. Steps (1) and (2) were
896 repeated for each local region across the whole-brain. It was hypothesized that shared neural
897 representations for NS and FE pain within a local region would be reflected by significant
898 cross-validation and cross-prediction accuracies for each classifier. Given that the specific
899 results of searchlight based-approaches strongly depend on the searchlight size if information is
900 not present and detected equally at all spatial frequencies (Etzel et al., 2013) we repeated our
901 analyses with two additional searchlight sizes (4mm- and 10mm-radius spheres).

902

903 **Specificity of the NS and FE vicarious pain-predictive patterns**

904 To test whether the observed NS and FE vicarious pain-predictive patterns were specific to pain
905 processing or rather reflect general aspects of negative emotional processing the two pain
906 predictive patterns were applied to the data from the emotional task paradigm. The first-level
907 model for the emotion processing data included the four experimental conditions (positive,
908 negative, neutral and white framed stimuli) and high-pass filter and nuisance regressors were
909 identical to the pain empathy GLM analysis. The two pain-predictive patterns were next applied
910 to negative and neutral contrasts (via dot-products) using a repeated 10-fold cross-validation
911 procedure separately, and subsequently two-alternative forced choice tests were employed to
912 discriminate negative versus neutral stimuli.

913

914 **Generalized vicarious pain-predictive pattern**

915 Given that we found shared neural representations between NS and FE vicarious pain (see
916 Results for details), a general vicarious pain pattern was developed by classifying vicarious pain
917 (NS and FE) versus control stimuli and further evaluated by predicting NS vicarious pain versus
918 NS control and FE vicarious pain versus FE control separately through 10-fold cross validation
919 procedures. We next constructed 10,000 bootstrap sample sets to visualize the voxels that made
920 the most reliable contribution to the classification and to decode the cognitive relevance of the
921 classifier with the resultant Z map using the Neurosynth (Yarkoni et al., 2011). Moreover, to
922 compare the general vicarious pain pattern with the NS and FE vicarious patterns, we examined
923 the similarities between this general vicarious pain pattern and the NS and FE vicarious pain
924 patterns, respectively.

925

926 **Generalizability of the vicarious pain pattern**

927 To test the functional relevance and generalizability of the empathic-induced neural pain pattern
928 the unthresholded whole-brain pattern of the general across NS and FE vicarious pain was
929 applied to determine the behavioral and neural responses during actual pain induction. To this
930 end data from a previous study employing different levels of thermal pain induction during
931 fMRI scanning (MRI data acquisition and preprocessing details see Wager et al., 2013; Woo et
932 al., 2015). First-level GLM analysis included regressors for stimulation periods for each of the 6
933 levels and the 11-sec rating periods as well as nuisance regressors including intercept for each

934 run, linear drift across time within each run, indicator vectors for outliers and head movement.

935 The general vicarious pain pattern from the current study was used to estimate the pattern

936 expressions of each participant in each condition (stimulation period) and next the neural pattern

937 expressions of the 6 pain levels were (1) correlated with the temperature levels (1-6) as well as

938 the subjective pain ratings separately and, (2) employed to discriminate high thermal pain

939 stimulation (average of 48.3 and 49.3°C) versus low stimulation (average of 44.3 and 45.3°C),

940 high stimulation versus medium stimulation (average of 46.3 and 47.3°C), as well as medium

941 stimulation versus low stimulation. Moreover, we conducted the same analyses with NS and FE

942 vicarious pain patterns to determine the robustness of the prediction.

943

944 **Data availability**

945 Statistical and pattern weight images are available on Neurovault

946 (<https://neurovault.org/collections/6332/>). Vicarious pain dataset as well as numerical data and

947 Matlab scripts that were used to generate the figures are available on figshare

948 (https://figshare.com/articles/Vicarious_pain_dataset/11994498). Other data can be obtained

949 from the corresponding authors upon reasonable request.

950

951 **Code availability**

952 Code is available at <https://github.com/canlab> and from the corresponding authors upon

953 reasonable request.

954

955 **Acknowledgements**

956 This work was supported by the National Key Research and Development Program of China (Grant
957 No. 2018YFA0701400), National Natural Science Foundation of China (NSFC, No 91632117,
958 31700998, 31530032); Fundamental Research Funds for Central Universities (ZYGX2015Z002),
959 Science, Innovation and Technology Department of the Sichuan Province (2018JY0001); National
960 Institute of Mental Health (R01 MH116026) and National Institute of Biomedical Imaging and
961 Bioengineering (R01EB026549).

962

963 **Competing interests**

964 The authors declare that they have no conflict of interest.

965

966 **References**

- 967 Afif, A., Minotti, L., Kahane, P., & Hoffmann, D. (2010). Anatomofunctional
968 organization of the insular cortex: a study using intracerebral electrical
969 stimulation in epileptic patients. *Epilepsia*, 51(11), 2305-2315.
- 970
- 971 Allman, J. M., Tetreault, N. A., Hakeem, A. Y., Manaye, K. F., Semendeferi, K.,
972 Erwin, J. M., Park, S., Goubert, V., & Hof, P. R. (2011). The von Economo
973 neurons in fronto-insular and anterior cingulate cortex. *Ann N Y Acad Sci*,
974 1225, 59.
- 975
- 976 Amft, M., Bzdok, D., Laird, A. R., Fox, P. T., Schilbach, L., & Eickhoff, S. B. (2015).
977 Definition and characterization of an extended social-affective default
978 network. *Brain Structure and Function*, 220(2), 1031-1049.
- 979
- 980 Botvinick, M., Jha, A. P., Bylsma, L. M., Fabian, S. A., Solomon, P. E., & Prkachin,
981 K. M. (2005). Viewing facial expressions of pain engages cortical areas
982 involved in the direct experience of pain. *Neuroimage*, 25(1), 312-319.
- 983
- 984 Carrillo, M., Han, Y., Migliorati, F., Liu, M., Gazzola, V., & Keysers, C. (2019).
985 Emotional mirror neurons in the rat's anterior cingulate cortex. *Current
986 Biology*, 29(8), 1301-1312. e1306.
- 987
- 988 Cauda, F., Costa, T., Torta, D. M., Sacco, K., D'Agata, F., Duca, S., Geminiani, G.,
989 Fox, P. T., & Vercelli, A. (2012). Meta-analytic clustering of the insular
990 cortex: characterizing the meta-analytic connectivity of the insula when
991 involved in active tasks. *Neuroimage*, 62(1), 343-355.
- 992
- 993 Chang, L. J., Gianaros, P. J., Manuck, S. B., Krishnan, A., & Wager, T. D. (2015). A
994 sensitive and specific neural signature for picture-induced negative affect.
995 *PLoS biology*, 13(6), e1002180.
- 996
- 997 Chen, L. M. (2018). Cortical representation of pain and touch: evidence from
998 combined functional neuroimaging and electrophysiology in non-human
999 primates. *Neuroscience bulletin*, 34(1), 165-177.
- 1000
- 1001 Chikazoe, J., Lee, D. H., Kriegeskorte, N., & Anderson, A. K. (2014). Population
1002 coding of affect across stimuli, modalities and individuals. *Nat Neurosci*,
1003 17(8), 1114.
- 1004

- 1005 Corradi-Dell'Acqua, C., Hofstetter, C., & Vuilleumier, P. (2011). Felt and seen pain
1006 evoke the same local patterns of cortical activity in insular and cingulate
1007 cortex. *Journal of Neuroscience*, 31(49), 17996-18006.
- 1008
- 1009 Corradi-Dell'Acqua, C., Tusche, A., Vuilleumier, P., & Singer, T. (2016).
1010 Cross-modal representations of first-hand and vicarious pain, disgust and
1011 fairness in insular and cingulate cortex. *Nature communications*, 7, 10904.
- 1012
- 1013 Craig, A., Bushnell, M., Zhang, E.-T., & Blomqvist, A. (1994). A thalamic nucleus
1014 specific for pain and temperature sensation. *Nature*, 372(6508), 770.
- 1015
- 1016 Craig, A. D., Chen, K., Bandy, D., & Reiman, E. M. (2000). Thermosensory
1017 activation of insular cortex. *Nat Neurosci*, 3(2), 184.
- 1018
- 1019 Craig, A. D., & Craig, A. (2009). How do you feel--now? The anterior insula and
1020 human awareness. *Nature reviews neuroscience*, 10(1).
- 1021
- 1022 Critchley, H. D., Wiens, S., Rotshtein, P., Öhman, A., & Dolan, R. J. (2004). Neural
1023 systems supporting interoceptive awareness. *Nat Neurosci*, 7(2), 189.
- 1024
- 1025 de Beeck, H. P. O. (2010). Against hyperacuity in brain reading: spatial smoothing
1026 does not hurt multivariate fMRI analyses? *Neuroimage*, 49(3), 1943-1948.
- 1027
- 1028 Decety, J. E., & Ickes, W. E. (2009). *The social neuroscience of empathy*. MIT Press.
- 1029
- 1030 Etzel, J. A., Zacks, J. M., & Braver, T. S. (2013). Searchlight analysis: promise,
1031 pitfalls, and potential. *Neuroimage*, 78, 261-269.
- 1032
- 1033 Fan, Y., Duncan, N. W., de Grecq, M., & Northoff, G. (2011). Is there a core neural
1034 network in empathy? An fMRI based quantitative meta-analysis.
1035 *Neuroscience & Biobehavioral Reviews*, 35(3), 903-911.
- 1036
- 1037 Gallo, S., Paracampo, R., Müller-Pinzler, L., Severo, M. C., Blömer, L.,
1038 Fernandes-Henriques, C., Henschel, A., Lammes, B. K., Maskaljunas, T., &
1039 Suttrup, J. (2018). The causal role of the somatosensory cortex in prosocial
1040 behaviour. *Elife*, 7, e32740.
- 1041
- 1042 Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E.,
1043 Ugurbil, K., Andersson, J., Beckmann, C. F., & Jenkinson, M. (2016). A
1044 multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615), 171.
- 1045

- 1046 Hadjistavropoulos, T., Craig, K. D., Duck, S., Cano, A., Goubert, L., Jackson, P. L.,
1047 Mogil, J. S., Rainville, P., Sullivan, M. J., & Williams, A. C. d. C. (2011). A
1048 biopsychosocial formulation of pain communication. *Psychological bulletin*,
1049 137(6), 910.
- 1050
- 1051 Haxby, J. V., Connolly, A. C., & Guntupalli, J. S. (2014). Decoding neural
1052 representational spaces using multivariate pattern analysis. *Annual review of*
1053 *neuroscience*, 37, 435-456.
- 1054
- 1055 Haynes, J.-D. (2015). A primer on pattern-based approaches to fMRI: principles,
1056 pitfalls, and perspectives. *Neuron*, 87(2), 257-270.
- 1057
- 1058 Hong, Y.-W., Yoo, Y., Han, J., Wager, T. D., & Woo, C.-W. (2019). False-positive
1059 neuroimaging: Undisclosed flexibility in testing spatial hypotheses allows
1060 presenting anything as a replicated finding. *Neuroimage*, 195, 384-395.
- 1061
- 1062 Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2003). A practical guide to support vector
1063 classification.
- 1064
- 1065 Iannetti, G. D., Salomons, T. V., Moayedi, M., Mouraux, A., & Davis, K. D. (2013).
1066 Beyond metaphor: contrasting mechanisms of social and physical pain.
1067 *Trends Cogn Sci*, 17(8), 371-378.
- 1068
- 1069 Jauniaux, J., Khatibi, A., Rainville, P., & Jackson, P. L. (2019). A meta-analysis of
1070 neuroimaging studies on pain empathy: investigating the role of visual
1071 information and observers' perspective. *Social cognitive and affective*
1072 *neuroscience*, 14(8), 789-813.
- 1073
- 1074 Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the
1075 human brain. *Nat Neurosci*, 8(5), 679.
- 1076
- 1077 Koban, L., Jepma, M., López-Solà, M., & Wager, T. D. (2019). Different brain
1078 networks mediate the effects of social and conditioned expectations on pain.
1079 *Nature communications*, 10(1), 1-13.
- 1080
- 1081 Kragel, P. A., Koban, L., Barrett, L. F., & Wager, T. D. (2018). Representation,
1082 pattern information, and brain signatures: from neurons to neuroimaging.
1083 *Neuron*, 99(2), 257-273.
- 1084
- 1085 Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional
1086 brain mapping. *Proceedings of the National Academy of Sciences*, 103(10),
1087 3863-3868.

- 1088
1089 Krishnan, A., Woo, C.-W., Chang, L. J., Ruzic, L., Gu, X., Lopez-Sola, M., Jackson,
1090 P. L., Pujol, J., Fan, J., & Wager, T. D. (2016). Somatic and vicarious pain are
1091 represented by dissociable multivariate brain patterns. *Elife*, 5, e15166.
1092
1093 Kvitsiani, D., Ranade, S., Hangya, B., Taniguchi, H., Huang, J., & Kepecs, A. (2013).
1094 Distinct behavioural and network correlates of two interneuron types in
1095 prefrontal cortex. *Nature*, 498(7454), 363.
1096
1097 Lamm, C., Decety, J., & Singer, T. (2011). Meta-analytic evidence for common and
1098 distinct neural networks associated with directly experienced pain and
1099 empathy for pain. *Neuroimage*, 54(3), 2492-2502.
1100
1101 Leigh, R., Oishi, K., Hsu, J., Lindquist, M., Gottesman, R. F., Jarso, S., Crainiceanu,
1102 C., Mori, S., & Hillis, A. E. (2013). Acute lesions that impair affective
1103 empathy. *Brain*, 136(8), 2539-2549.
1104
1105 Li, J., Xu, L., Zheng, X., Fu, M., Zhou, F., Xu, X., Ma, X., Li, K., Keith, K., & Becker,
1106 B. (2018). Common and dissociable contributions of alexithymia and autism
1107 to domain-specific interoceptive dysregulations-a dimensional neuroimaging
1108 approach. *bioRxiv*, 432971.
1109
1110 Meng, J., Hu, L., Shen, L., Yang, Z., Chen, H., Huang, X., & Jackson, T. (2012).
1111 Emotional primes modulate the responses to others' pain: an ERP study.
1112 *Experimental brain research*, 220(3-4), 277-286.
1113
1114 Mitchell, J. P. (2009). Inferences about mental states. *Philosophical Transactions of
the Royal Society B: Biological Sciences*, 364(1521), 1309-1316.
1115
1116 Northoff, G., Heinzel, A., De Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp,
1117 J. (2006). Self-referential processing in our brain—a meta-analysis of imaging
1118 studies on the self. *Neuroimage*, 31(1), 440-457.
1119
1120 Peelen, M. V., & Downing, P. E. (2007). Using multi-voxel pattern analysis of fMRI
1121 data to interpret overlapping functional activations. *Trends Cogn Sci*, 11(1),
1122 4-4.
1123
1124 Peelen, M. V., Wiggett, A. J., & Downing, P. E. (2006). Patterns of fMRI activity
1125 dissociate overlapping functional brain areas that respond to biological motion.
1126 *Neuron*, 49(6), 815-822.
1127
1128

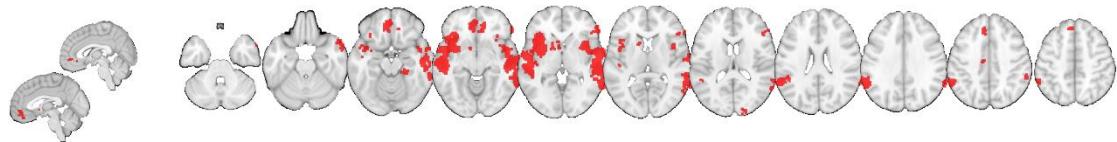
- 1129 Rütgen, M., Seidel, E.-M., Silani, G., Riečanský, I., Hummer, A., Windischberger, C.,
1130 Petrovic, P., & Lamm, C. (2015). Placebo analgesia and its opioidergic
1131 regulation suggest that empathy for pain is grounded in self pain. *Proceedings*
1132 *of the National Academy of Sciences*, 112(41), E5638-E5646.
- 1133
- 1134 Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annu. Rev.*
1135 *Neurosci.*, 27, 169-192.
- 1136
- 1137 Sakaguchi, T., Iwasaki, S., Okada, M., Okamoto, K., & Ikegaya, Y. (2018). Ethanol
1138 facilitates socially evoked memory recall in mice by recruiting pain-sensitive
1139 anterior cingulate cortical neurons. *Nature communications*, 9(1), 3526.
- 1140
- 1141 Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J., &
1142 Davidson, R. J. (2011). The integration of negative affect, pain and cognitive
1143 control in the cingulate cortex. *Nature reviews neuroscience*, 12(3), 154.
- 1144
- 1145 Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for
1146 empathy: a double dissociation between emotional and cognitive empathy in
1147 inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain*, 132(3),
1148 617-627.
- 1149
- 1150 Sheng, F., & Han, S. (2012). Manipulations of cognitive strategies and intergroup
1151 relationships reduce the racial bias in empathic neural responses. *Neuroimage*,
1152 61(4), 786-797.
- 1153
- 1154 Shidara, M., & Richmond, B. J. (2002). Anterior cingulate: single neuronal signals
1155 related to degree of reward expectancy. *Science*, 296(5573), 1709-1711.
- 1156
- 1157 Shmuel, A., Chaimow, D., Raddatz, G., Ugurbil, K., & Yacoub, E. (2010).
1158 Mechanisms underlying decoding at 7 T: ocular dominance columns, broad
1159 structures, and macroscopic blood vessels in V1 convey information on the
1160 stimulated eye. *Neuroimage*, 49(3), 1957-1964.
- 1161
- 1162 Shura, R. D., Hurley, R. A., & Taber, K. H. (2014). Insular cortex: structural and
1163 functional neuroanatomy. *The Journal of neuropsychiatry and clinical*
1164 *neurosciences*, 26(4), iv-282.
- 1165
- 1166 Sikes, R. W., & Vogt, B. A. (1992). Nociceptive neurons in area 24 of rabbit cingulate
1167 cortex. *Journal of neurophysiology*, 68(5), 1720-1732.
- 1168
- 1169 Singer, T., Critchley, H. D., & Preuschoff, K. (2009). A common role of insula in
1170 feelings, empathy and uncertainty. *Trends Cogn Sci*, 13(8), 334-340.

- 1171
1172 Timmers, I., Park, A. L., Fischer, M. D., Kronman, C. A., Heathcote, L. C.,
1173 Hernandez, J. M., & Simons, L. E. (2018). Is empathy for pain unique in its
1174 neural correlates? A meta-analysis of neuroimaging studies of empathy.
1175 *Frontiers in Behavioral Neuroscience*, 12, 289.
- 1176
1177 Uddin, L. Q. (2015). Salience processing and insular cortical function and
1178 dysfunction. *Nature reviews neuroscience*, 16(1), 55.
- 1179
1180 Urien, L., & Wang, J. (2019). Top-Down Cortical Control of Acute and Chronic Pain.
1181 *Psychosomatic medicine*, 81(9), 851-858.
- 1182
1183 Vachon-Presseau, E., Roy, M., Martel, M.-O., Albouy, G., Chen, J., Budell, L.,
1184 Sullivan, M. J., Jackson, P. L., & Rainville, P. (2012). Neural processing of
1185 sensory and emotional-communicative information associated with the
1186 perception of vicarious pain. *Neuroimage*, 63(1), 54-62.
- 1187
1188 Valentini, E., & Koch, K. (2012). Fine-grained analysis of shared neural circuits
1189 between perceived and observed pain: implications for the study of empathy
1190 for pain. *Journal of neurophysiology*, 108(7), 1805-1807.
- 1191
1192 Varoquaux, G. (2018). Cross-validation failure: small sample sizes lead to large error
1193 bars. *Neuroimage*, 180, 68-77.
- 1194
1195 Wager, T. D., Atlas, L. Y., Botvinick, M. M., Chang, L. J., Coghill, R. C., Davis, K.
1196 D., Iannetti, G. D., Poldrack, R. A., Shackman, A. J., & Yarkoni, T. (2016).
1197 Pain in the ACC? *Proceedings of the National Academy of Sciences*, 113(18),
1198 E2474-E2475.
- 1199
1200 Wager, T. D., Atlas, L. Y., Lindquist, M. A., Roy, M., Woo, C.-W., & Kross, E.
1201 (2013). An fMRI-based neurologic signature of physical pain. *New England
1202 Journal of Medicine*, 368(15), 1388-1397.
- 1203
1204 Woo, C.-W., Chang, L. J., Lindquist, M. A., & Wager, T. D. (2017). Building better
1205 biomarkers: brain models in translational neuroimaging. *Nat Neurosci*, 20(3),
1206 365.
- 1207
1208 Woo, C.-W., Koban, L., Kross, E., Lindquist, M. A., Banich, M. T., Ruzic, L.,
1209 Andrews-Hanna, J. R., & Wager, T. D. (2014). Separate neural
1210 representations for physical pain and social rejection. *Nature communications*,
1211 5, 5380.
- 1212

- 1213 Woo, C.-W., Roy, M., Buhle, J. T., & Wager, T. D. (2015). Distinct brain systems
1214 mediate the effects of nociceptive input and self-regulation on pain. *PLoS*
1215 *biology*, 13(1), e1002036.
- 1216
- 1217 Xu, L., Bolt, T., Nomi, J. S., Li, J., Zheng, X., Fu, M., Kendrick, K. M., Becker, B., &
1218 Uddin, L. Q. (2020). Inter-subject phase synchronization differentiates neural
1219 networks underlying physical pain empathy. *Social cognitive and affective*
1220 *neuroscience*. <https://doi.org/10.1093/scan/nsaa025>
- 1221
- 1222 Xu, X., Dai, J., Liu, C., Chen, Y., Xin, F., Zhou, F., Zhou, X., Huang, Y., Wang, J., &
1223 Zou, Z. (2019). Common and Disorder-Specific Neurofunctional Markers of
1224 Dysregulated Empathic Reactivity in Major Depression and Generalized
1225 Anxiety Disorder. *Psychotherapy and psychosomatics*, 1.
- 1226
- 1227 Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011).
1228 Large-scale automated synthesis of human functional neuroimaging data.
1229 *Nature methods*, 8(8), 665.
- 1230
- 1231 Yesudas, E. H., & Lee, T. (2015). The role of cingulate cortex in vicarious pain.
1232 *BioMed Research International*, 2015.
- 1233
- 1234 Yu, H., Koban, L., Chang, L. J., Wagner, U., Krishnan, A., Vuilleumier, P., Zhou, X.,
1235 & Wager, T. D. (2019). A generalizable multivariate brain pattern for
1236 interpersonal guilt. *bioRxiv*, 835520.
- 1237
- 1238 Zaki, J., Wager, T. D., Singer, T., Keysers, C., & Gazzola, V. (2016). The anatomy of
1239 suffering: understanding the relationship between nociceptive and empathic
1240 pain. *Trends Cogn Sci*, 20(4), 249-259.
- 1241
- 1242 Zhou, F., Geng, Y., Xin, F., Li, J., Feng, P., Liu, C., Zhao, W., Feng, T., Guastella, A.
1243 J., Ebstein, R. P., Kendrick, K. M., & Becker, B. (2019). Human Extinction
1244 Learning Is Accelerated by an Angiotensin Antagonist via Ventromedial
1245 Prefrontal Cortex and Its Connections With Basolateral Amygdala. *Biol*
1246 *Psychiatry*, 86(12), 910-920. <https://doi.org/10.1016/j.biopsych.2019.07.007>
- 1247
- 1248 Zhou, F., Zimmermann, K., Xin, F., Scheele, D., Dau, W., Banger, M., Weber, B.,
1249 Hurlemann, R., Kendrick, K. M., & Becker, B. (2018). Shifted balance of
1250 dorsal versus ventral striatal communication with frontal reward and
1251 regulatory regions in cannabis-dependent males. *Human Brain Mapping*,
1252 39(12), 5062-5073.
- 1253
- 1254

1255

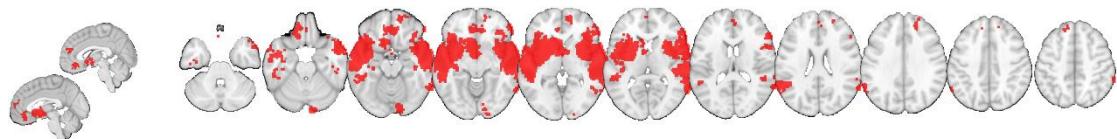
A Overlapping regions from the searchlight-based prediction analyses (4mm-radius sphere)



B Overlapping regions from the searchlight-based prediction analyses (6mm-radius sphere)

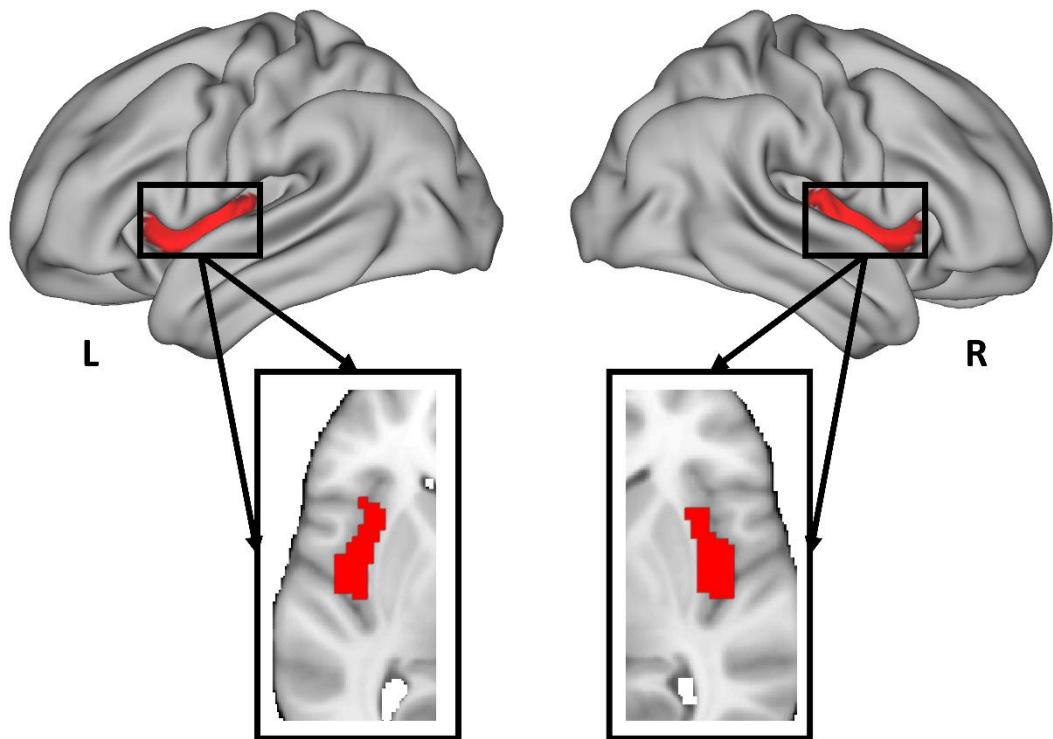


C Overlapping regions from the searchlight-based prediction analyses (10mm-radius sphere)



1256

1257 **Figure 4 – figure supplement 1.** Searchlight analyses with different searchlight sizes. This figure
1258 shows the results for the analyses in which we ran searchlight analyses with (A) 4mm-, (B) 6mm-,
1259 and (C) 10mm-radius spheres.
1260

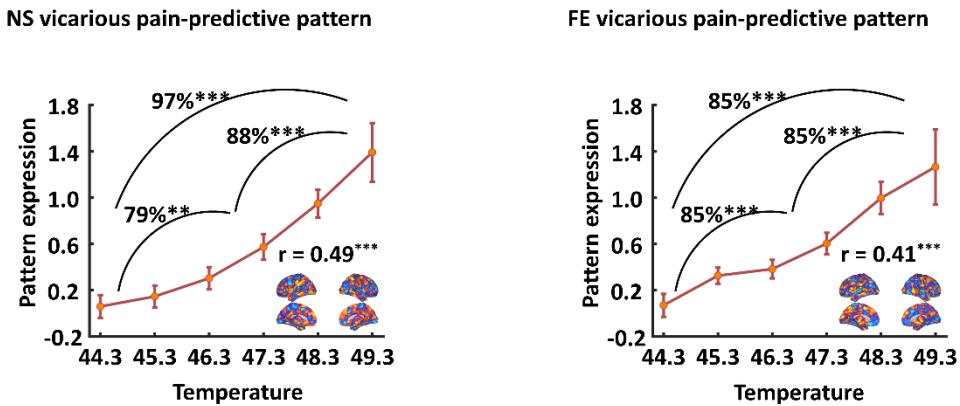


1261

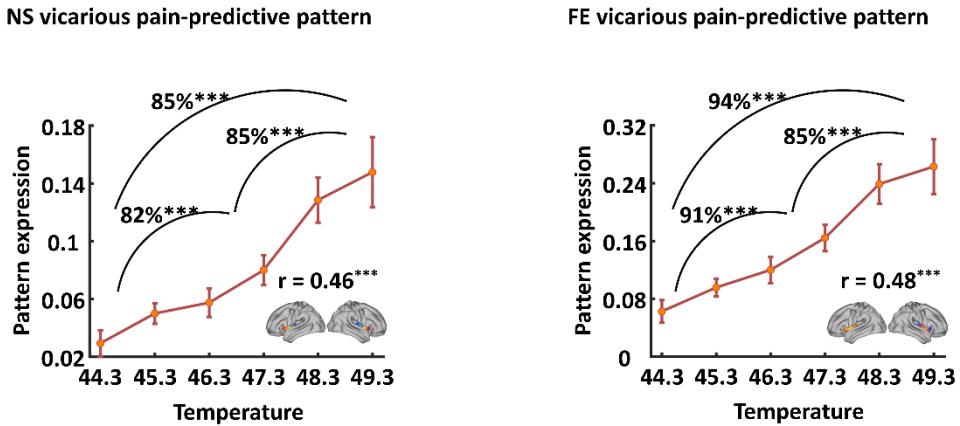
1262 **Figure 5 – figure supplement 1.** The mid-insula mask used in the current study.

1263

A Whole-brain vicarious pain-predictive pattern expressions of thermal pain



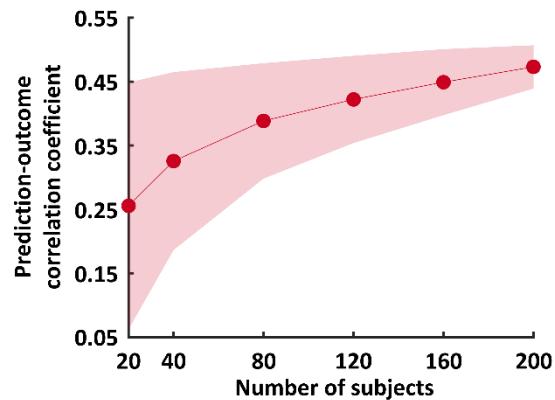
B Mid-insula vicarious pain-predictive pattern expressions of thermal pain



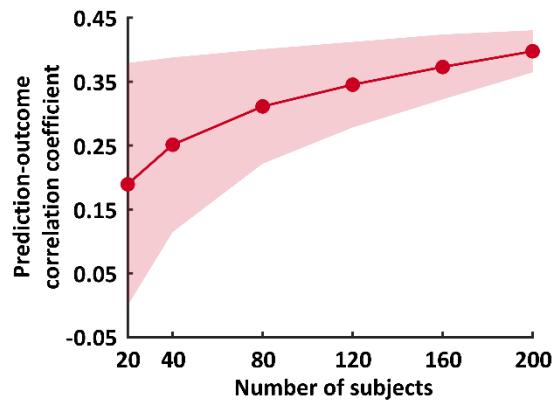
1264

1265 **Figure 7 – figure supplement 1.** Generalizability of the NS and FE vicarious pain-predictive
1266 patterns. Both whole-brain (A) and mid-insula (B) vicarious pain-predictive patterns could accurately
1267 predict the severity and classify the levels of self-experienced pain in an independent dataset. NS
1268 vicarious pain, observation of noxious stimulation of body limbs induced vicarious pain; FE vicarious
1269 pain, observation of facial expressions of pain induced vicarious pain. r indicates correlation
1270 coefficient between pattern expression and temperature levels. ** $P < 0.001$; *** $P < 0.001$.
1271

A NS vicarious pain-predictive pattern predicts thermal pain levels



B NS vicarious pain-predictive pattern predicts thermal pain ratings



1272 **Figure 7 – figure supplement 2.** Varying sample size prediction. This figure depicts the results for
1273 the analysis in which we predicted thermal pain levels (A) and ratings (B) using randomly selected n
1274 = 20, 40, 80, 120, 160 and 200 subjects' NS data. Shaded area indicates standard deviation. NS
1275 vicarious pain, observation of noxious stimulation of body limbs induced vicarious pain.
1276

1277 **Supplementary file 1.** Table shows post-fMRI subjective ratings for vicarious pain evoking
1278 stimuli (Mean \pm SD). NS vicarious pain, observation of noxious stimulation of body limbs
1279 induced vicarious pain; FE vicarious pain, observation of facial expressions of pain induced
1280 vicarious pain; NS control stimuli depict body limbs in similar but innocuous situations, FE
1281 control stimuli show neutral facial expressions.