Multiple Brain Networks Mediating Stimulus–Pain Relationships in Humans

Stephan Geuter, Elizabeth Reynolds Losin, Mathieu Roy, Lauren Atlas, Liane Schmidt, Anjali Krishnan, Leonie Koban, Tor Wager, Martin Lindquist

Multiple brain networks mediating stimulus-pain relationships in humans

Stephan Geuter[1,2,*], Elizabeth A. Reynolds Losin[3], Mathieu Roy[4], Lauren Y. Atlas[5,6], Liane Schmidt[7], Anjali Krishnan[8], Leonie Koban[2,9], Tor D. Wager[2,9,10,#], Martin A. Lindquist[1,#]

[1] Department of Biostatistics, Johns Hopkins University, USA

[2] Institute of Cognitive Science, University of Colorado Boulder, USA

[3] Department of Psychology, University of Miami, USA

[4] Department of Psychology, McGill University, Canada

[5] National Center for Complementary and Integrative Health, National Institutes of Health, USA

[6] National Center for Drug Abuse, National Institutes of Health, USA

[7] Social-and-Affective Neuroscience Team, Institute du Cerveau et de la Moelle Epinière, INSERM UMR 1127, CNRS UMR 7225, Université Pierre et Marie Curie Paris 6, France

[8] Department of Psychology, Brooklyn College of the City University of New York, USA

[9] Department of Psychology and Neuroscience, University of Colorado Boulder, USA

[10] Presidential Cluster in Neuroscience and Department of Psychological and Brain Sciences, Dartmouth College, Hanover, USA

* Corresponding author:
Stephan Geuter
Department of Biostatistics
Johns Hopkins University
615 N Wolfe Street, Baltimore, MD 21205, USA
Email: sgeuter@jhmi.edu
Phone: +1 (443) 287-8791

[#] Authors contributed equally to this work

Abstract

The brain transforms nociceptive input into a complex pain experience comprised of sensory, affective, motivational, and cognitive components. However, it is still unclear how pain arises from nociceptive input, and which brain networks coordinate to generate pain experiences. We introduce a new high-dimensional mediation analysis technique to estimate distributed, network-level patterns that formally mediate the relationship between stimulus intensity and pain. We applied the model to a large-scale analysis of functional magnetic resonance imaging data (N=284), focusing on brain mediators of the relationship between noxious stimulus intensity and trial-to-trial variation in pain reports. We identify mediators in both traditional nociceptive pathways and in prefrontal, midbrain, striatal, and default-mode regions unrelated to nociception in standard analyses. The whole-brain mediators are specific for pain vs. aversive sounds and are organized into five functional networks. Brain mediators predicted pain ratings better than previous brain measures, including the Neurologic Pain Signature (Wager et al. 2013). Our results provide a broader view of the networks underlying pain experience, as well as novel brain targets for interventions.

Introduction

The brain is central to the generation of pain; it transforms sensory input into a complex set of pain-related responses, including subjective experience, autonomic responses, avoidance behavior, and activation of linked memories and concepts. However, the boundaries of the brain systems that mediate this series of transformations have been inconsistent across studies. Traditionally, pain processing has been associated with a discrete set of brain regions targeted by spinal nociceptive afferents, including primary (S1) and secondary (S2) somatosensory, and anterior midcingulate cortices (aMCC), medial and lateral thalamus, and posterior and mid-insular cortices (Apkarian et al. 2005; Dum et al. 2009; Jensen et al. 2016). These have been referred to as the 'pain matrix', and often treated as a unitary system, though this concept has been largely abandoned as its specificity to pain has been called into question. Other studies have found that additional regions are also involved in encoding the intensity of noxious stimuli and/or correlate with pain experience under some circumstances (Bingel et al. 2002; Büchel et al. 2002; Becerra et al. 2013), making the boundaries of the 'pain matrix' elusive and context-dependent. And, though this set of regions have been grouped into subsystems (Craig et al. 2000)—for example, lateral and medial subsystems more closely related to sensory-discriminative and affective-motivational aspects of pain, respectively (Villemure et al. 2003)—empirical studies have shown that the division between sensory encoding and pain affect is not straightforward (Baliki et al. 2009; Atlas et al. 2010, 2014).

One source of complexity lies in the fact that 'pain matrix' or 'pain-processing' regions have been defined in multiple ways. Some studies identify regions based on stimulus intensity encoding, the tendency for a region to be activated by more vs. less intense noxious stimuli (Peyron et al. 2002; Wager et al. 2004). Others identify regions based on correlations with pain reports (Coghill et al. 1999; Lindquist et al. 2017). Both are relevant: A region involved in pain generation should both encode stimulus intensity and correlate with reported subjective experience, even when intensity is matched. Accordingly, one step forward lies in characterizing formal brain *mediators* of the relationship between stimulus intensity and pain report (Atlas et al. 2010, 2014). Mediation is a statistical test that links experimentally manipulated variables, brain measures, and behavioral outcome variables in a single path model. It requires both an effect of an initial (often experimentally manipulated) variable and association with an outcome (e.g., pain) controlling for stimulus intensity. Applied to pain, it can be used to identify brain regions that both encode experimental manipulations in noxious stimulus intensity and correlate with pain

experience controlling for intensity, with sufficient effect sizes in both tests to pass the more stringent test of mediation.

Previous studies have identified brain mediators of pain (Atlas et al. 2010, 2014) testing voxels one at a time. However, this kind of univariate approach can miss brain regions whose contributions to pain perception are conditional on other regions. More broadly, it is increasingly clear that much of the functional information encoded in the brain is encoded in distributed patterns across neural ensembles and systems (Pouget et al. 2000; Haxby et al. 2014), which requires brain information to be treated in a multivariate fashion (Kriegeskorte 2011; Woo et al. 2017). Mediation models with *multivariate* brain mediators are required to characterize these patterns, but have not been available until now. Here, we provide the first analysis of multivariate brain mediators of pain, using a novel statistical method called *principal directions of mediation* (PDM) (Chén et al. 2017). PDM decomposes activity across the brain into multiple networks that independently mediate stimulation effects on outcomes (Figure 1A-B). This can help identify which brain systems mediate stimulus intensity effects on pain, taking the distributed nature of information encoding in the brain into account.

Another source of complexity lies in the fact that most pain studies have been necessarily limited in sample size (typical $N < 50$). Results from small studies are increasingly recognized as variable and prone to high levels of both false positives and false negative results (Button et al. 2013). We address this issue by analyzing within-person, single-trial pain data aggregated across 8 individual studies (N=284). In addition, the test study not only included heat pain stimuli, but also two distinct types of

sounds: Physically aversive sounds (a ‘knife on plate’) and emotionally aversive sounds (gunshots, screams, etc.) This provides one of the first tests of specificity of multivariate pain-related patterns against sounds (cf. Horing et al. 2019; Liang et al. 2019), and the first test of functional specificity of brain mediators of pain.

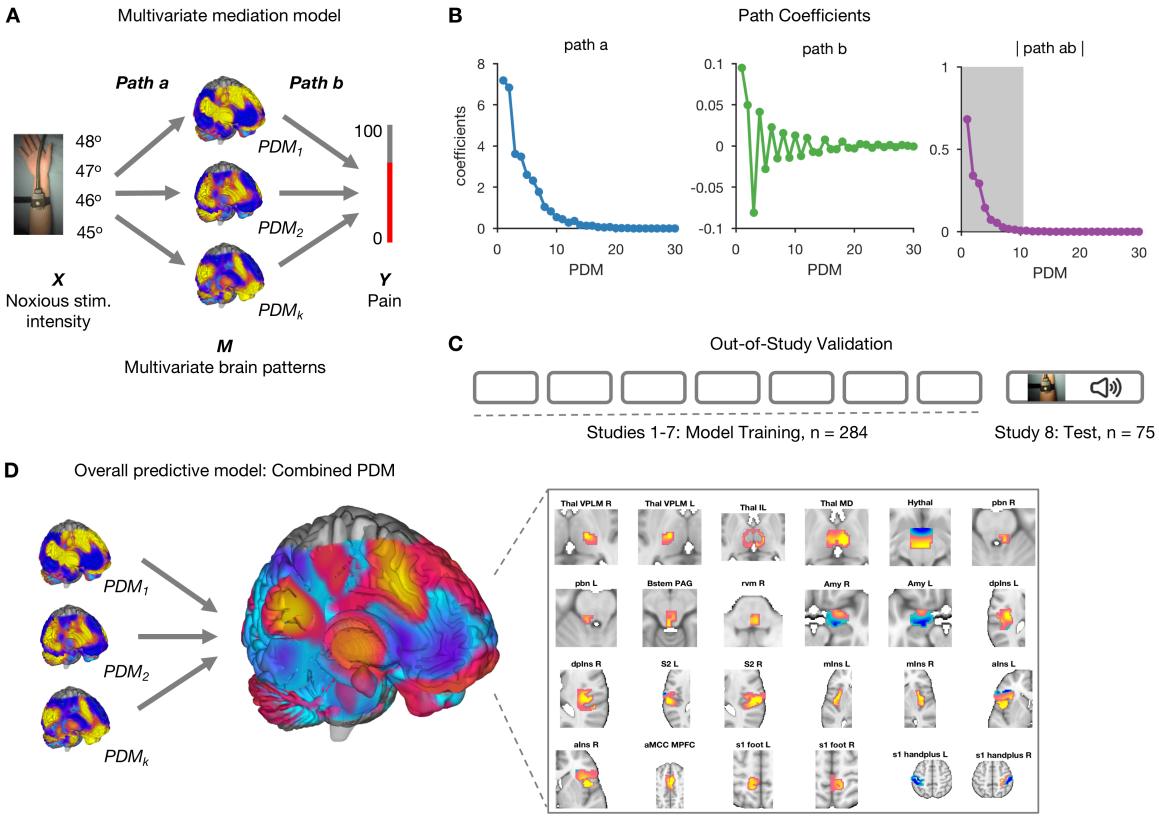

Figure 1. Mediation analysis. (A) Schematic of the mediation analysis framework. Brain activity is an intermediate, mediating variable (M) between a manipulated noxious stimulus intensity (X) and perceived pain (Y). In the high-dimensional Principal Directions of Mediation (PDM) approach, a linear combination of all brain voxels act as mediators. Multiple, orthogonal mediators can be estimated. The weight vectors w_k (or PDMs) represent the contribution of individual voxels to the k^{th} mediation pathway. Voxel weights (w_k) are fit so that the indirect, mediated effect is maximized. (B) Mediation path coefficients for all 30 PDMs are shown with signs of path a coefficients set to be positive. Path a indicates the temperature to brain (PDM) relationship, path b the PDM to pain rating relationship, and path ab the indirect, mediated effect. Positive coefficients indicate that voxels with positive weights in a given PDM are positively related with temperature and/or rating. The first 10 PDMs explain more 99% of the total indirect effect. We focus on these PDMs in the following analyses (shaded area in right panel). (C) PDMs are estimated on the training data comprising a total of 284 participants from 7 pain studies. PDM validation is done on an independent 8th data set with 75 participants. (D) Individual PDMs can be combined into a single, combined PDM (cPDM) by weighting and summing the individual, orthogonal PDMs. cPDM voxel weights are shown on the brain rendering and in key regions in the right panel.

Materials and Methods

Participants

The analysis included data from a total of 284 healthy participants from 8 independent studies, with sample sizes ranging from N = 17 to N = 75 per study. Descriptive statistics on the age, sex, and other features of the subjects in each individual study are provided in Supplementary Tables S1-S3. Further details on Studies 1-7, which were used to estimate the PDMs are provided in Lindquist et al. (Lindquist et al. 2017). Participants were recruited from New York City and Boulder/Denver Metro Areas. The institutional review board of Columbia University and the University of Colorado Boulder approved all the studies, and all participants provided written informed consent. Preliminary eligibility of participants was determined through an online questionnaire, a pain safety screening form, and a functional Magnetic Resonance Imaging (fMRI) safety screening form.

We applied several exclusion criteria for analysis purposes. Participants with psychiatric, physiological or pain disorders, neurological conditions, and MRI contraindications were excluded prior to enrollment. In addition, participants were required to have at least 30 trials with low variance inflation factors (see below), non-missing rating, and stimulation intensity data. Based on these criteria, 18 participants from Study 8 were excluded, resulting in a total of 209 participants for the primary PDM analysis and 75 participants for the validation sample.

Procedures

Overview. Participants in all studies underwent fMRI scanning while being exposed to varying levels of heat pain within-person and rating the perceived pain intensity (Supplementary Tables S1-S3). For each participant, we recorded the temperature applied, the pain rating for each trial and estimated single-trial maps of brain activity. These three variables were used in the primary mediation analysis with temperature as the initial variable, brain activity as the mediator, and pain rating as the outcome variable (Figure 1).

Using the high-dimensional mediation analysis model, we first estimated 30 whole-brain mediation patterns (PDMs). Each PDM specifies a linear combination of voxels across the brain maximizing the mediated effect from temperature to pain rating, while being orthogonal to other PDMs (Figure 1A). Each PDM (or w_p) thus represents a formal, whole brain mediator for pain. The voxel weights of each PDM inform us about the contribution of individual brain regions to the generation of a painful experience following noxious stimulation.

Furthermore, as the PDM model is linear, independent PDMs can readily be fused into a single, combined PDM that can be prospectively applied to new datasets as a predictive model. We fused the individual brain mediator maps into a combined PDM (cPDM) by estimating the weighted combination of individual PDMs that best predicted pain in the training sample (see PDM Section below and Figure 1D). Prediction performance can then be evaluated against independent test data

Thermal and aversive sound stimulation. The number of noxious stimulation trials, stimulation sites, inter-trial intervals, rating scales, and stimulus intensities and durations varied across studies, but were comparable; these variables are summarized in Tables S2 and S3. Each study also comprised a specific psychological manipulation (except Study 8), such as placebo treatment, which will be or has been reported elsewhere (Table S1).

In Studies 1-6, thermal stimulation was delivered to multiple skin sites using a TSA-II Neurosensory Analyzer (Medoc Ltd., Chapel Hill, NC) with a 16 mm Peltier thermode endplate. A PATHWAY system (Medoc Ltd., Chapel Hill, NC) was used in Studies 7 and 8. Participants rated the perceived magnitude of warmth or pain during or after each trial. Most studies applied heat to the left volar forearm. See Supplement and Lindquist et al. (2017) for details.

Study 8, which was used for validation purposes (see below), also presented aversive sounds to participants. Trials with aversive sounds were used to test the specificity of the pain PDMs. Sounds included were a physically aversive recording of nails on a chalkboard and a set of emotionally aversive sounds (attacks, screaming, and crying) from the International Affective Digital Sounds database (IADS) (Bradley and Lang 2007). Aside from these sound trials, we focus on brain mediation of pain across all trials in the present paper, irrespective of the study-specific psychological and physical manipulations that influenced pain.

fMRI data processing

Preprocessing and subject level models. We chose to retain the original preprocessing used in each published paper for two reasons: (1) to establish, and test, robustness across minor variations in processing pipelines; and (2) because study-specific analysis choices are appropriate in some cases, depending on the distribution of the data and study design. For details see Supplementary Methods and Lindquist et al. **(2017)**. Briefly, structural, T1-weighted images were co-registered to the functional mean image, then normalized to MNI space using SPM. For each study, a single trial, or single epoch model was estimated (Koyama et al. 2003;

Rissman et al. 2010; Mumford et al. 2012). The single-trial brain activation estimates served as the basis for the subsequent analysis.

PDM validation. The high-dimensional brain mediators (PDMs, see below) were estimated on the training data comprised of Studies 1-7 (Lindquist et al. 2017). Even though this data set is large (N=209) and diverse, the possibility of overfitting in the training data might reduce the generalizability of the PDMs. To test for the generalizability of the PDMs, we validated the PDMS on independent test data (Study 8, N=75). Computing the inner product of each PDM with each single-trial beta image from Study 8 resulted in 10 potential mediator variables. Each of these potential mediators was then subjected to a multi-level mediation analysis (Wager et al. 2009) with *p*-values determined by a bootstrap procedure with 5,000 iterations each. If the PDMs generalize to the new dataset, paths $\boldsymbol{a_k}$, $\boldsymbol{b_k}$, and the indirect effect $\boldsymbol{ab_k}$ should be significant for all $\boldsymbol{k = 1, \ldots, 10}$ PDMs.

We also tested whether the PDMs specifically mediate the relationship between temperature and pain intensity. To this end, we also tested the original PDMs on the aversive sound trials from Study 8. If the PDMs reflect specific patterns of brain activity involved in pain processing, they should not mediate the relationship between sound stimulation level and intensity ratings. We thus expect no significant indirect effect for the sound trials.

A further test to validate the stability of PDM estimation was conducted by switching training and test data. That is, pain PDMs were estimated on Study 8 and tested on the original training data from Studies 1-7 as described above.

Dimension reduction. The training data set consisted of a total of 13,372 single-trial beta images (i.e., activation estimate images), each consisting of 229,519 voxels, from 209 participants. To reduce the dimensionality of the data to a computationally tractable size, a generalized version of population value decomposition (PVD) (Caffo et al. 2010; Crainiceanu et al. 2011; Chén et al. 2017) was applied (using PVD.m, included in the M3 mediation toolbox available at https://github.com/canlab/MediationToolbox). This procedure is similar to singular value decomposition (SVD) but decomposes the data matrix into both participant specific and population specific components. We chose a dimensionality of $\boldsymbol{p = 30}$ based on a tradeoff between variance explained and the number of trials available for each participant. The beta images were z-scored within each participant before PVD application. The reduced data matrix used for Principal Directions of Mediation (PDM) estimation consisted of a matrix with

dimensions $\mathbf{13,372 \times 30}$.

Principal Directions of Mediation (PDM) Model

Let X_i be the temperature, Y_i the reported pain, and $M_i = (m_i^{(1)}, m_i^{(2)}, \dots m_i^{(p)})$ the brain activity over p voxels (i.e., the beta maps) measured between the application of the thermal stimuli and the pain report for observation (i.e., trial) $i = 1, \dots n$. We are interested in determining how brain activation mediates the relationship between temperature and pain report. We can estimate the parameters of this model using the following set of equations:

$$m_i^{(j)} = \alpha_{0,j} + \alpha_j X_i + \varepsilon_{ij} \qquad \text{for} \qquad j = 1, \dots p$$
$$Y_i = \beta_0 + \gamma' X_i + \beta_1 m_i^{(1)} + \beta_2 m_i^{(2)} + \cdots + \beta_p m_i^{(p)} + \eta_i \qquad (1)$$

Once the parameters have been estimated we can express the total effect γ as the sum of the direct and indirect effects as follows:

$$\gamma = \gamma' + \sum_{j=1}^{p} \alpha_j \beta_j. \qquad (2)$$

If *p* is relatively small the series of regressions described in (1) can be used to estimate the pertinent mediation effects. However, in our setting there are too many mediators to allow reasonable interpretation (unless the model coefficients are highly structured) and there are many more mediators than subjects, precluding estimation using standard procedures. To overcome these problems, we introduce a transformation of the space of mediators, determined by finding linear combinations of the original mediators that (i) are orthogonal; and (ii) are chosen to maximize the indirect effect. The first constraint allows us to fit a separate linear model for each transformed variable. The second constraint allows us to limit our analysis to only those directions that contain the most information about the indirect effect. Here, we improve and extend the approach proposed by Chén et al. (2017) by choosing a different cost function, computing a combined PDM, and analyzing an almost 10-times larger data set.

This new model, called the *principal directions of mediation* (PDM), linearly combines activity in different voxels into a smaller number of orthogonal components, with components ranked based upon the proportion of the indirect effect that each accounts for. Ideally, the components form a small number of uncorrelated mediators that represent interpretable

networks of voxels.

To illustrate, let $\widetilde{m}_i^{(k)} = \sum_{j=1}^{p} w_k^{(j)} m_i^{(j)}$ for $k = 1, \dots q$ be a set of linear transformations of the mediators with $w_k = (w_k^{(1)}, w_k^{(2)}, \dots w_k^{(p)})$. Placing these new variables into our mediation model we obtain:

$$\widetilde{m}_j^{(k)} = a_{0,k} + a_k X_j + \varepsilon_{jk} \qquad \text{for} \quad k = 1, \dots q$$
$$Y_i = b_{0,k} + c' X_i + b_k \widetilde{m}_i^{(k)} + \eta_{ik} \tag{3}$$

Now, we can decompose the total effect into direct and indirect effects as follows:

$$c = c' + \sum_{k=1}^{q} a_k b_k \tag{4}$$

The difference between this model and the standard mediation model described in (1) is that the w_k are unknown. In our approach w_1 is chosen so that it maximizes the amount of the indirect effect that is explained (i.e., $a_1 b_1$ is maximized). We refer to w_1 as the first *principal direction of mediation* (PDM). Note the first PDM corresponds to voxel-specific weights that can be mapped onto the brain, and thus provides interpretable maps of brain networks in the same manner as independent component analysis (ICA) and principal component analysis (PCA). Subsequent directions w_k , $k = 1, \dots q$, can be found that maximize the remaining indirect effect conditional on being orthogonal to previous PDMs. As the transformed mediators are ranked based upon the proportion of the indirect effect explained, one could potentially limit the number of PDMs computed to achieve dimension reduction. Hence, our approach is philosophically similar to PCA, but addresses a fundamentally different problem.

The individual, orthogonal PDMs can be fused into a combined PDM (cPDM) by computing the following weighted sum:

$$w_{\text{combined}} = \sum_{k=1}^{q} d_k \, w_k \tag{5}$$

The scalar weights d_k are estimated from the training sample using a linear model with the individual PDMs as regressors and reported pain as the response.

According to the model formulation the signs of the PDMs are not identifiable, as any change in the sign of $\widetilde{m}_i^{(k)}$ can be offset by a change in sign of both a_k and b_k. We fix the signs

of a_k to be positive for easier interpretation, i.e., positive voxel weights indicate higher brain activity for higher stimulus intensities. This is a similar constraint to the ICA approach often used in neuroimaging to detect networks. The orthogonality constraint does not reduce the total amount of variance explained by all PDMs.

The problem of finding the k^{th} PDM involves finding the vector w_k that maximizes $a_k b_k$ based on the constraint that $w_k^T w_k = 1$ and $w_k^T w_j = 0$ for all $j = 1, \ldots, k-1$. This problem can be solved using a nonlinear programming solver such as the interior-point algorithm. Inference is performed using a bootstrap procedure with 5,000 iterations, as described in Chén et al. (2017). PDM maps are thresholded at a false discovery rate (FDR) of $q < 0.05$. The cPDM map in Figure 2 displays the top 5% of voxels based on their weight parameters, yielding a more conservative display than FDR. We present results of 10 PDMs accounting for more than 99% of the total indirect effect. A software implementation is available at https://github.com/canlab/MediationToolbox (multivariateMediation.m).

In summary, we obtain scalar coefficients for paths a_k, b_k, and c'_k, as well as the indirect effect $a_k b_k$ for each PDM as in a standard, univariate mediation analysis. In addition, we obtain the voxel weight vector w_k that maximizes the indirect effect $a_k b_k$.

Cluster analysis

The voxel weight maps for the mutually independent 10 PDMs span a high-dimensional space of brain mediators of pain perception. In order to reduce the dimensionality of that space and identify brain regions with similar activation profiles, we conducted a two-stage cluster analysis. The procedure is described in detail in (Kober et al. 2008) and (Atlas et al. 2014). Briefly, for significant voxels from the 10 PDMs we extracted single-trial activity estimates, resulting in a 13,372 trials × 25,469 voxels matrix. We then used singular value decomposition (SVD) to reduce the dimensionality of the voxel space. We kept 364 components that explained 95% of the variance. Next, we clustered voxels into 250 spatial parcels using hierarchical clustering. We then computed average single-trial activity within each parcel and used non-metric multidimensional scaling (NMDS) and hierarchical clustering to further reduce the dimensionality of the data. Inspection of the Shepard plot suggested a NMDS dimensionality of 15 with stress indices below 0.05. Stress indices (S) are computed according to Shepard (1980) with

$$S = \sqrt{\frac{\sum_{h,i}(d_{hi} - \hat{d}_{hi})^2}{\sum_{h,i} d_{hi}^2}} \quad (6)$$

Here, d_{hi} is the pairwise empirical dissimilarity and $\hat{d}_{hi}$ is the distance implied by the current solution between two brain regions h and i. Hierarchical clustering was then used to cluster the 250 parcels into 33 regions that co-activate across trials. These regions were not necessarily contiguous and some spanned multiple anatomical regions, e.g., covering right mid-, and dorsal insula plus operculum. Since we used voxel-wise FDR correction on the 10 PDMs, we expect some false positive values. Accordingly, some of the functional regions were located in the cerebrospinal fluid or outside the gray matter. We thus removed 7 smaller functional clusters that were considered highly unlikely to be true gray matter region. We then averaged brain activity within the remaining 26 functional regions or nodes. NMDS was used to reduce the dimensionality again to 10 dimensions based on stress values. Applying hierarchical clustering again on the regions identified in the previous step identified large-scale functional brain networks. Permutation tests indicated that 5 networks provided the best clustering solution in terms of improvement over solutions on permuted data. Similarity of those 5 networks with the binarized PDM maps was assessed by Dice coefficients, which represents the true positive rate of the intersection between two maps.

Local pattern expression analyses

To summarize the cPDM pattern weights as a function of known regions or networks, we calculated the pattern energy within three sets of pre-defined regions. The first was a set of regions identified as contributing to nociceptive pain pathways based on prior work. We identified regions in the cortex using the atlas of Glasser et al. (2016), the thalamus using the atlas of Morel et al. (1997), and key brainstem regions based on previous papers: For parabrachial nucleus (PBN), Fairhurst et al. (2007); rostral ventral medulla (RVM), Brooks et al. (2017). For the PAG, we (T.D.W.) hand-drew a region on the 7T high-resolution group T1 of Keuken et al. (2014) and segmented out the cerebral aqueduct to exclude it. Nociceptive thalamic zones were included as defined below.

The second set of regions was based on Morel et al. (1997). We grouped the 80 or so thalamic/epithalamic regions into 17 functional zones likely to be detectable using fMRI (see Figure 3 for a complete list). Nociceptive thalamic zones included the ventral-posterior-lateral (VPL) and –medial (VPM) zone, the intralaminar group, and the mediodorsal ‘association nucleus’ (MD).

The third set of regions was a set of cortical networks defined based on resting-state connectivity, which we used to map cPDM weights onto established large-scale networks. We extracted loadings from 16 unique networks described in Schaefer et al. (2018) and manually separated them into left and right hemisphere components to examine lateralization. See Figure 4 for a complete list.

To summarize pattern weights in each local region or network, we calculated a measure of 'pattern energy', related to the absolute magnitude of predictive weights:

$$E_r = \frac{\sqrt{w^T w}}{V + 1}$$

E_r is the root-mean-square of weights in region (or network mask) *r* per cubic cm of brain tissue. *w* denotes the vector of weights for in-region voxels, and V the volume of the region in cm^3. As the variance of E_r varies inversely with region volume, the constant 1 is added to regularize the volume and thus avoid noise-driven, large-magnitude estimates for small regions. The area of the wedges in Figures 3 and 4 is proportional to E_r.

We also defined a measure of 'pattern valence' in pre-defined regions, which is the degree to which voxel weights are uniformly positive or negative. Pattern valence is defined as the cosine similarity of the pattern weights with the unit vector across in-region voxels. It is bounded at 1 and -1, where 1 indicates uniform positive weights across voxels, and -1 indicates uniform negative weights. The color of the wedges in Figures 3 and 4 are proportional to the pattern valence; red colors indicate homogenous positive weights, blue homogenous negative weights, and purple mixed, variable weights across in-region voxels.

Comparison to other multivariate models

To compare the PDM approach to other multivariate models of pain processing, we compared PDM1 and cPDM to the Neurological Pain Signature (NPS; Wager et al. 2013), to the Stimulus Intensity Independent Pain Signature 1 (SIIPS1, Woo et al. 2017), and the combination of NPS and SIIPS1. Furthermore, we compared the PDM approach against pain prediction by the Neurosynth reverse inference map for the term "pain" (https://neurosynth.org/analyses/terms/pain/). For each map, we correlated the predicted pain outcomes with the actual pain ratings on a single-trial level. In addition, we correlated the predictive brain maps with each other to evaluate similarities and differences in spatial pattern weights. Prediction outcome correlations were compared using paired t-tests of Fisher-z

transformed correlation values.

Univariate mediation analysis

In univariate mediation analyses, a mediation model is estimated separately for every brain voxel (Wager et al. 2008; Atlas et al. 2010, 2014). Univariate mediation analysis produces three sets of brain maps – one for each path – in contrast to the PDM approach, which estimates only one set of paths for each PDM map. Previous studies also used smaller sample sizes available than the present study and had thus less statistical power than the present study. We ran a univariate mediation analyses on the training data set to directly compare the univariate results to the PDM approach. Univariate multilevel mediation analysis was conducted using the Multilevel Mediation and Moderation (M3) Toolbox for Matlab (https://github.com/canlab/MediationToolbox). Voxel-wise significance was determined using a bootstrap procedure with 5,000 iterations. A false discovery rate (FDR) of $q < 0.05$ was used to control for multiple comparisons.

Results

Principal Directions of Mediation (PDM)

For each individual PDM, we estimated Path a (stimulus intensity to brain), Path b (brain to pain report), and mediation ($a * b$) effects as in a standard mediation model (Figure 1-2). A positive path a indicates that higher temperatures lead to more activity in voxels with positive PDM weights (yellow in brain figures) and less activity in voxels with negative PDM weights (blue in brain figures). A positive path b indicates that voxels with positive weights contribute positively to the pain rating after controlling for temperature. This pattern would be expected for regions that receive spinothalamic input, for example the dorsal posterior insula or S2 (Willis and Westlund 1997; Dum et al. 2009), and possibly other mediating regions as well.

The absolute coefficient values for the indirect ab path assess how much of the effect of the manipulated temperature on pain ratings is explained by the brain mediator, i.e., individual PDM pattern. Here, the first 10 PDMs accounted for 99.1% of the total mediation effect (Figure 1B, Figure S1). We thus focus on the first 10 PDMs in all subsequent analyses with minimal loss of information. In order to analyze the contribution of individual brain regions to the mediation of pain, the signs of both paths a and b and the sign of the voxel weights have to be considered: Voxel weights are multiplied by the respective path coefficients to determine a region's relationship to stimulation intensity and pain rating.

When considering the signs of the voxels weights, four different kinds of relationship are possible: (*i*) positive to temperature, positive to pain; (*ii*) negative to temperature, negative to pain; (*iii*) positive to temperature, negative to pain; and (*iv*) negative to temperature, positive to pain. Here, type (*i*) is the standard, positive mediator case expected from nociceptive coding regions and type (*ii*) represents a negative mediator, in which greater deactivation to the stimulus mediates increased pain (MacKinnon et al. 2000). Types (*iii*) and (*iv*) are suppressor effects (MacKinnon et al. 2000), e.g., for type (*iii*), brain activity increases with stimulus intensity that suppress pain, and may thus be involved in stimulus-engaged regulatory processes and other negative feedback loops. Note that the values of path coefficients shown in Figure 1 depend on the scaling of the predictor (X), mediator (M), and outcome (Y). The fact that path a coefficients are an order of magnitude larger than path b is solely related to differences in scaling and does not relate to their relevance. Please note that mixing of signals from distinct neural populations within fMRI voxels is common in similar types of analysis such as Independent Component

Analysis (ICA) and manifests itself in different weight patterns across PDMs, eg. for left S1.

PDM 1 has both positive path a and b coefficients. Brain regions with positive weights (representing positive mediators, type (*i*) with positive paths a and b) are shown in warm colors in Figure 2. These include brain regions commonly associated with pain processing, such as the dorsal posterior and mid-insula, S1, S2, MCC, and the PAG (Figure 2). Significant voxels in MCC stretch into the supplementary motor area (SMA), dorsal of the cingulate sulcus. In addition, PDM 1 contains negative, type (*ii*), mediators, including the medial prefrontal cortex (mPFC) and left S1/M1. The negative weights indicate that these regions show less activation with increasing temperatures and lower regional activation is related to higher pain ratings. Such relationships are to be expected for brain regions whose function is inhibited by nociceptive input or that are deactivated with increased pain-related processing.

Brain regions positively mediating the relationship between temperature and pain rating (type (*i*)) in other PDMs are S1, M1, superior frontal gyrus (SFG), fronto-temporal operculum, temporal poles, temporal operculum, ventral insula, pons, and cerebellum (Figure 2; yellow regions in, e.g., PDMs 1, 2, 4, and 6 in particular). These positive mediators include regions, like the temporal regions, that are traditionally not considered to be pain-processing regions. Brain regions acting as negative mediators (type (*ii*)) in other PDMs include medial orbitofrontal cortex (mOFC), dorso-medial prefrontal cortex (dmPFC), superior parietal lobule (SPL), retrosplenial cortex (RSC), precuneus, and cuneus (blue regions in PDMs 1, 2, 4, and 6). Those regions belong to systems that are deactivated consequent to pain or nociception (e.g., systems mediating competing functions).

A more complex function is indicated by positive path a coefficients, but negative path b coefficients (types (*iii*) and (*iv*), PDMs 3, 5, 7, and 9). In type (*iii*), regions with positive voxel weights show a positive relationship with temperature, i.e., higher temperatures lead to more activity. However, the negative path b indicates that these regions are negatively related to pain ratings controlling for temperature, i.e., more activity is related to lower pain ratings. Regions with such a profile fit a pain-inhibitory role, as they are activated by painful stimulation but serve to dampen pain—a negative feedback loop. Parts of the mOFC/vmPFC, the cerebellum, precuneus, S1, temporal-parietal junction, and the left dlPFC fit this pain inhibitory profile (see yellow areas in PDMs 3, 5, and 7 in particular).

A final set of regions shows a negative relationship with temperature (positive path a, but negative weights; blue in PDMs 3, 5, and 7) and a positive relationship with pain ratings, controlling for temperature (negative voxel weights and negative path b resulting in a net positive

relationship; type (*iv*)). Such regions show stimulus intensity-dependent deactivation, with larger de-activation mediating decreased pain, consistent with regulatory negative feedback mechanisms. Regions with this profile include parts of the mOFC, the parahippocampal gyrus, visual cortices, and the NAc. For example, NAc shows decreased activation for high temperatures, which may relate to punishment or negative reinforcement signals. At the same time, controlling for temperature, stronger NAc de-activation is related to lower pain ratings, potentially signaling reduced motivational relevance.

In the individual PDMs, each voxel is assigned a weight value—as in Independent Components Analysis (ICA), a voxel can thus participate in multiple components or 'networks', potentially revealing multiple functional roles of a voxel. Some regions participated in multiple PDMs in this fashion—most notably, mOFC appears to play roles as a type (*ii*), (*iii*), and (*iv*) mediator in PDMs 2, 5, and 7. This may reveal a complex function of the mOFC in pain and a mixing together of signal from multiple distinguishable neural populations.

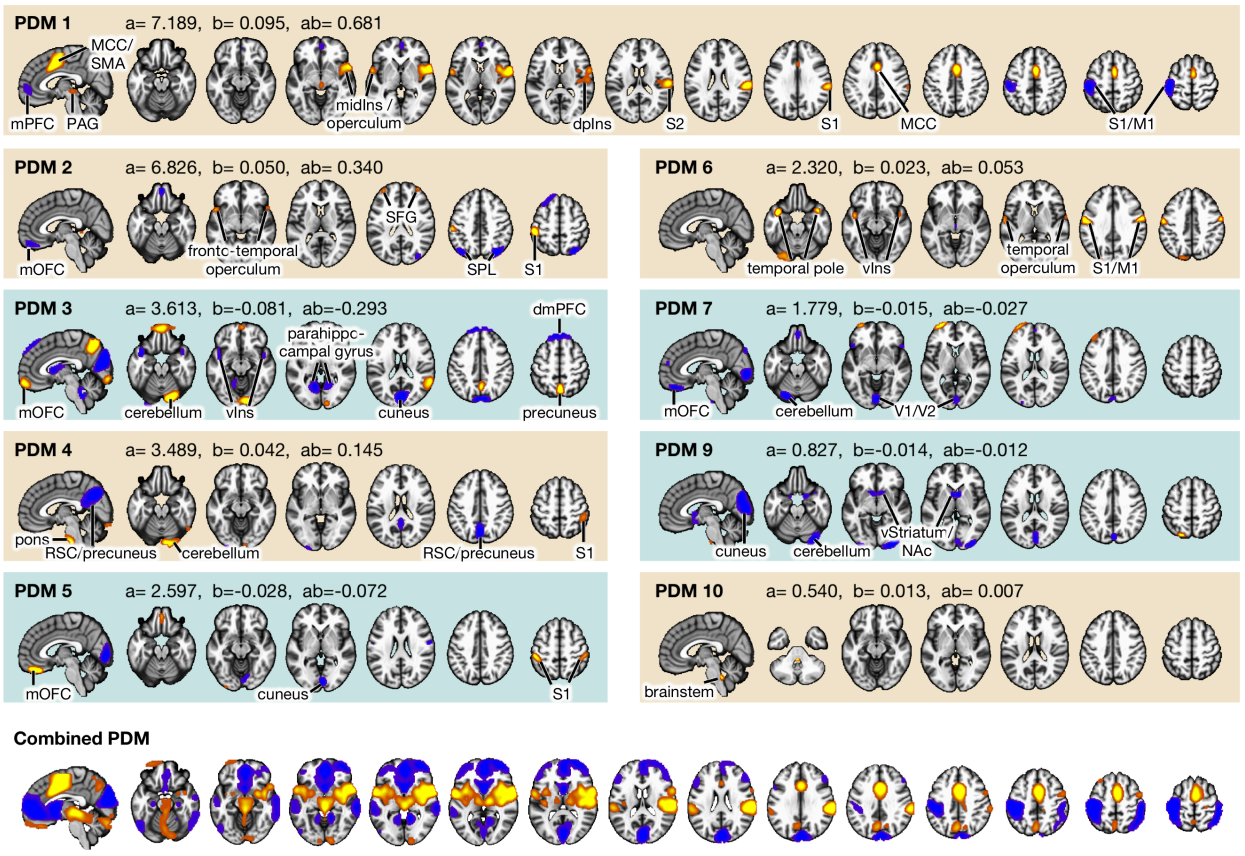

Figure 2. Principal Directions of Mediation. Voxel maps for PDMs with individually significant voxels at *FDR* $q < 0.05$. Tan backgrounds indicate PDMs with positive paths a and b. Blue backgrounds indicate PDMs with positive path a and negative path b. Brain activity increases in voxels with positive weights (warm colors) with higher temperatures. Higher brain activity in these voxels is related to higher pain ratings in PDMs with positive path b (tan panels) and negatively with negative path b (blue panels). No voxels are individually significant in PDM 8. Bottom panel shows the combined PDM, a weighted linear combination of the above 10 PDMs. The top 5% of voxels based on voxel weights are shown since almost all voxels survived the significance testing. All brain figures are displayed in neurological convention (left is left) and thresholded at FDR $q < 0.05$. MCC=midcingulate cortex, SMA=supplementary motor area, mPFC=medial prefrontal cortex, PAG=periaqueductal gray, midIns=mid-insula, dpIns=dorsal posterior insula, S2=secondary somatosensory cortex, S1=primary somatosensory cortex, M1=primary motor cortex, mOFC=medial orbitofrontal cortex, RSC=retrosplenial cortex, SFG=superior frontal gyrus, vIns=ventral insula, dmPFC=dorsomedial prefrontal cortex, V1=primary visual cortex, V2=secondary visual cortex, vStriatum=ventral striatum, NAc=nucleus accumbens, mThal=medial thalamus, aIns= anterior insula, SPL=superior parietal lobule.

Combined PDM

The individual PDMs can be fused into a single, combined PDM since the individual PDMs are orthogonal to each other. The weights are estimated from the training sample using a linear model with the individual PDMs as regressors and reported pain as the response. Summing the weighted PDMs results in a combined PDM (cPDM) map (see Figure 1D and Methods). In doing so, we lose information about multiple functional roles played by each voxel or region, but we obtain a single overall characterization of each voxel and a map that can be applied as a predictive model. Voxel weights may be both positive and negative in different PDMs, because voxels may include neural ensembles participating in different distributed circuits related to either more or less pain. Thus, the individual PDMs represent a decomposition of voxels' activity into different distributed components, while the combined PDM reflects each voxel's net contribution (controlling for other voxels). Computing and analyzing the cPDM can thus help to clarify overall relationships between regional activity and the predictor and outcome variables.

Within the cPDM, individually significant clusters of positive mediators included S2, MCC, SMA, PAG, insula (including anterior and dorsal-posterior parts), and the medial thalamus (Figure 1D, Figure 2). Negative mediators (stimulus-induced deactivations mediating increased pain) included mPFC, SPL, S1, and M1.

To further characterize the weights of the cPDM in nociception- and pain-related regions of interest, we identified 24 distinct anatomical regions that encode pain in human and animal literature, and examined the cPDM weights in each of these regions (Figure 1D, right and Figure 3). The regions were divided according to the thalamic atlas of Morel et al. (1997) and cortical atlas of Glasser et al. (2016). Positive weights were found in elements of the spinothalamic tract (bilateral VPL thalamus and dorsal posterior insula [dpINS], and also S1), spino-parabrachial tract (bilateral parabrachial nucleus and amygdala), spinoreticular tract (rostral ventral medulla [RVM] and PAG), spinohypothalamic tract (posterior hypothalamus), and spinolimbic tract (mediodorsal [MD] and intralaminar [IL] nuclei of the thalamus, and aMCC). Many of these subcortical regions have not been consistently identified in human studies, but are crucial mediators of pain in animal models. These results do not tell us that the activity in question is due to direct spinal input, but they robustly identify a set of targets in areas containing known pathways. Activity was bilateral in most regions, though the amygdala and spinothalamic regions (S1, S2, dpINS) showed right-hemisphere dominance (most studies involved left-sided stimulation).

Figure 3 shows the pattern energy (root-mean-square weights per cubic cm of tissue) for positive and negative weights in yellow and blue, respectively. It shows that weights are mixed

in the hypothalamus, RVM, amygdala, and S1, suggesting that the local pattern weights in these regions are particularly important, as they make the pattern response different from the average across the anatomic region. Figure 3B shows the pattern energy for all subdivisions of the thalamus, not only those related to nociception. The pattern energy is high (large wedge area) and weights fairly uniformly positive (red color) in thalamic nuclei known to have nociceptive inputs: VPL, VPM, IL, and MD nuclei, and the habenula (Hb). Roles for all of these in pain are well documented (Willis and Westlund 1997; Shelton et al. 2012). Pattern energy is low and with mixed signs on weights (purple color) in other sensory nuclei, including lateral/medial geniculate and pulvinar nuclei, along with other nuclei. Though a perfect match between any anatomical atlas and functional imaging data cannot be guaranteed, these findings suggest that the distribution of weights even across the relatively small volume of the thalamus is meaningful. They also confirm specific nuclei known mainly from animal and invasive human studies as pain-related in human fMRI data, and suggest new potential pain-related nuclei to be confirmed and further characterized, such as ventrolateral (VL), ventromedial (VM), and anterior medial (AM) nuclei.

Examining the cPDM weights in established resting-state cortical networks also yielded a selective profile across networks, with weights concentrated in a few networks (Figure 4). cPDM weights were high and relatively uniformly positive (red wedges in Figure 4) in ‘Somatomotor B’, ‘Ventral Attention A’, and ‘Control C’ networks. (We adopt these names are by convention only, and do not suggest that the networks’ functions map onto these labels). The first two broadly match previous analyses of nociceptive activity, though they provide additional information on mapping to sub-networks. The latter is more surprising, as it includes regions that are not typically nociceptive such as precuneus and parts of posterior cingulate. ‘Somatomotor’ and ‘Ventral Attention’ weights were left-lateralized, and ‘Somatomotor’ most strongly so. ‘Control C’ weights were right-lateralized.

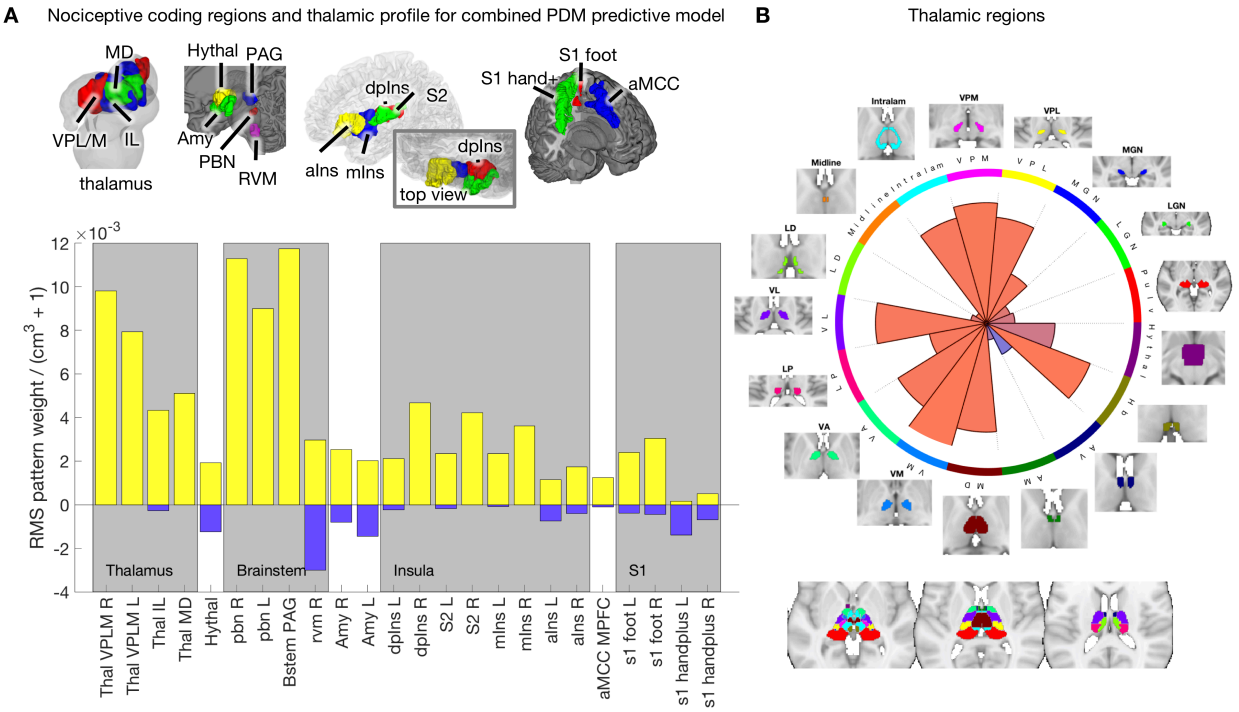

Figure 3. Combined PDM weights in nociceptive pathways. **(A)** Pattern weights in 24 cerebral targets of nociceptive pathways. The results show strong contributions from nociceptive subregions in the thalamus and brainstem, including both ventrolateral (sensory) and medial (affective) thalamic zones, and the parabrachial complex (PBN). These are critical pathways in animal models, but have been seldom identified in human pain studies. **(B)** Pattern weights across thalamic regions as defined by the Morel atlas. Predictive weights are distributed unevenly across the thalamus, and positive weights (red) are concentrated in sensory (VPM/VPL) and medial nociceptive targets (MD and intralaminar), and anterior nuclei (VA, VM, VL). They are also positive in the habenula (Hb). By contrast, they are absent in visual/auditory regions (MGN, LGN, Pulvinar). This demonstrates selectivity for expected pain-related nuclei.

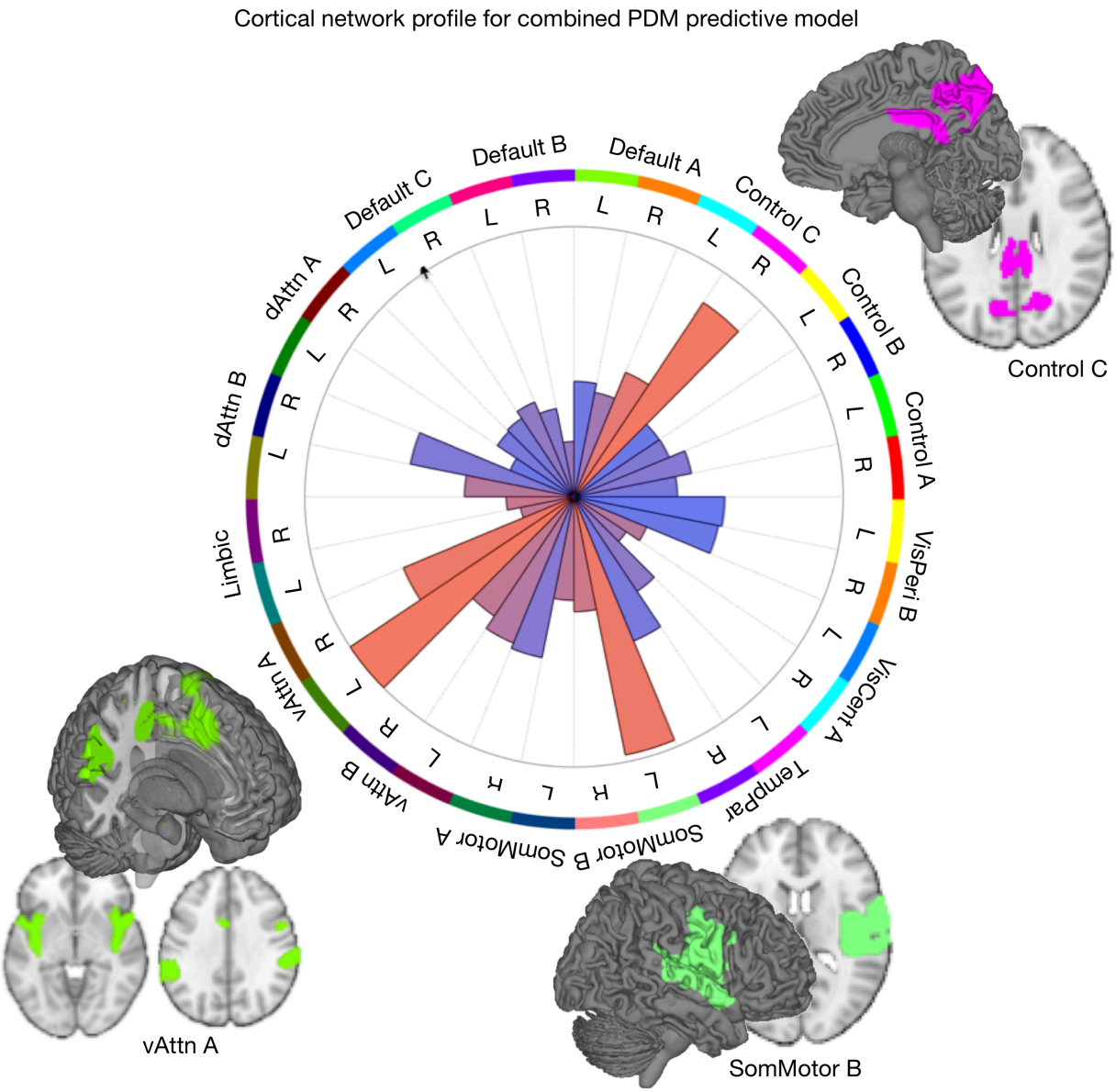

Figure 4. Cortical network profile for the cPDM. Pattern energy in cortical resting state networks are distributed unevenly with strong, positive weights (red wedges) present in Somatomotor B, Ventral attention A, and Control C networks. The first two match broadly on known nociceptive processing areas, while parts of Control C (e.g., precuneus) are less known for pain processing.

Clustering PDMs into functional networks

While the previous analyses examined the relationship of cPDM weights with independently defined anatomical regions and functional brain networks, in the next step we analyzed the spatial clustering arising from the individual PDMs themselves. The PDMs provide a dimensional view of coherent distributed processes, with each PDM a distinct dimension; clustering the PDMs can reveal the network structure of the inter-regional relationships. To do this, we used an iterative clustering procedure to group regions based on inter-regional correlations in stimulus-evoked responses across trials without considering stimulation temperatures or pain ratings (Kober et al. 2008; Atlas et al. 2014). The cluster analysis of single-trial activity from significant voxels of all 10 PDMs revealed 26 functional regions organized into 5 different functional networks (Figure 5A,B). Though somewhat limited, for exploratory purposes a functional description of these networks was determined by computing the similarity of each network with feature maps generated by the meta-analytic tools on neurosynth.org (Yarkoni et al. 2011). The top ten features for each network are shown in Table 1. Network names were chosen based on the functional associations with neurosynth.org terms. For example, the top three feature associations for network 1 were somatosensory, motor, and stimulation. Based on these associations we labeled network 1 as 'sensorimotor network'.

Network 1 ('sensorimotor') included somatosensory regions like dpIns, mid-insula, S2, S1, but also the PAG, MCC, SMA, M1, and cerebellum. The second network ('value learning') included the NAc, ventral anterior insula, frontal operculum, and temporal poles. Network 3 consisted of regions that are part of the default mode network (DMN), including mPFC, mOFC, and retrosplenial cortex. The fourth network ('executive function') included precuneus, inferior parietal lobule (IPL), superior parietal lobule (SPL), dorsal lateral occipital cortex (dLOC), temporal-parietal junction (TPJ), superior frontal gyrus (SFG), and dlPFC. Finally, network 5 ('visual') included mostly occipital, visual areas and parts of the parahippocampal gyrus. The variety of functions ascribed to the five networks mediating pain indicate that pain processing involves multiple, distinct brain networks in addition to somatosensory systems.

We next investigated with which functional networks the individual PDMs are associated by computing pairwise Dice similarity coefficients across voxels, estimating the spatial similarity of the PDMs and network maps (Figure 5B). By contrast, PDM 1 (type *(i)*/(*ii*) mediators) had the greatest overall similarity with any single network, namely with the sensorimotor network ($D = 0.7$). No other network was substantially associated to PDM 1 (all $D < 0.05$). PDMs 2, 5, and 6 were also spatial similarity to the sensorimotor network (PDMs 2 and 6 are type (*i*)/(*ii*) mediators).

The value learning network was related to PDMs 3, 6, and 9, with the highest similarity to PDM 9 ($D = 0.16$)—thus, mainly components with type (*iii*)/(*iv*) mediation (PDM 3 and 9). Similarity between the default mode network and PDM 4 was highest ($D = 0.24$, mainly type (*ii*) mediators). Parts of the DMN also overlapped with PDMs 3, 5, and 7. The executive function network was associated with PDM 2 ($D = 0.22$) and PDM 3 ($D = 0.23$), and, to a lesser degree, with PDMs 4 and 7. Finally, the visual network was related to PDM 3 ($D = 0.38$) and to a lesser degree to PDMs 5, 7, and 9. The overall similarity pattern between functional networks and PDMs shows that in contrast to PDM1, few of the remaining PDMs are dominated by a single network. More often PDMs were comprised of a mix of 2 or 3 networks that together act as a pain mediator, reflecting the complexity of the transformation from nociception into pain experience.

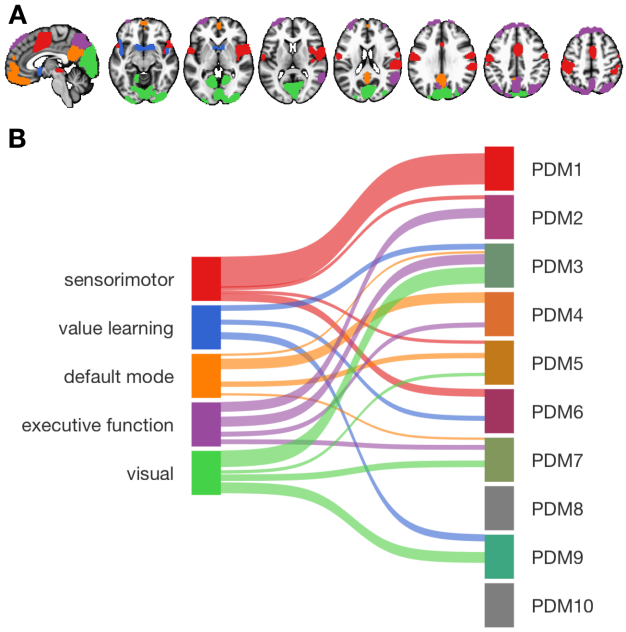

Figure 5. Functional networks mediating pain processing. (**A**) Five functional networks based on the clustering of brain activity in significant voxels from the PDM analysis. Labels for colors are shown in B. (**B**) Associations between functional networks and PDMs. Ribbon width represents Dice-coefficient similarity between networks and PDMs.

Table 1. *Neurosynth.org network associations*

Sensorimotor		**Value-learning**		**Default mode**		**Executive function**		**Visual**	
r	*features*	*r*	*features*	*r*	*features*	*r*	*features*	*r*	*features*
0.369	somatosensory	0.313	reward	0.207	self-referential	0.139	mental	0.223	visual
0.304	motor	0.255	money	0.202	person	0.124	intention	0.152	eye
0.301	stimulation	0.252	anticipation	0.201	self	0.117	stories	0.141	eyes
0.272	sensorimotor	0.252	rewards	0.197	default	0.115	attention	0.137	color
0.266	muscle	0.251	incentive	0.176	autobiographical	0.115	visuospatial	0.126	shape
0.257	sensory	0.240	monetary	0.157	resting state	0.114	story	0.108	shapes
0.256	pain	0.236	outcome	0.149	social	0.108	reasoning	0.105	spatial
0.245	movements	0.196	outcomes	0.149	mentalizing	0.107	default	0.102	development
0.245	production	0.185	dopamine	0.148	personal	0.106	calculation	0.097	distractor
0.240	painful	0.179	reinforcement	0.135	thought	0.106	retrieval	0.097	target

Note: Top ten features from neurosynth.org showing the highest Pearson's correlation (*r*) with each network.

Validation on an independent cohort

Although we estimated PDMs on a large and diverse data set, there is a risk that the PDMs may over-fit noise inherent in the training data, potentially preventing generalization to other data sets. We thus applied the PDMs to an independent test data set, without re-estimating any model parameters. The resulting vectors of potential mediators ($\widetilde{m}_i^{(k)}$) were then entered into standard multi-level mediation models. If the PDMs generalize to the new data, the indirect ab effects should be significant on the test data.

Applying the PDMs to independent pain test data (N = 75, an independent community sample cohort of mixed races and sex), revealed significant paths a and b for all 10 PDMs and the cPDM (Figure 6A). The indirect path ab was also significant for the cPDM and all individual PDMs, suggesting that all PDMs are reliably related to pain and generalize across cohorts. The magnitude of the indirect effects (path ab) are monotonically decreasing for the training data (Figure 1B). On the test data, indirect path coefficients were not strictly monotonically decreasing from PDM 1 to PDM 10 (Figure 6A, Figure S2), indicating some variability of the PDM order across data sets, as expected. The cPDM and the first two individual PDMs had the strongest effect in both data sets, suggesting that they capture the most important brain activity for pain across data sets. Figure 6C shows the predicted pain from the cPDM plotted against the empirical pain ratings for pain training and test data.

To further corroborate the generalizability and robustness of the PDMs, we also estimated 10 PDMs on the original test data set (Study 8) and cross-validated the new PDMs on the original training data set (Studies 1-7). The results were similar to the main results presented here. Six out of ten indirect paths were significant when PDM estimation was done on the smaller sample. The indirect ab path coefficients for the first four PDMs were highest when applying the new PDMs to the original training data (Figure S3). Generalization thus does not depend strongly on the choice of the training data.

In order to test whether PDMs are mediators specifically for somatic pain, we also applied the original PDMs to other, non-painful aversive stimuli in Study 8—physically (knife on plate) and emotionally (screaming, crying, etc.) aversive sounds with three pre-defined intensity levels of each stimulus type. Study 8 was designed to test specificity vs. generalizability to aversive sounds and matched in duration and approximate aversiveness ratings based on pilot studies; trials were randomly intermixed with heat pain trials. Application of the original PDMs on the sound data revealed no significant indirect effects (Figure 6B, Figure S4) and only nine significant

paths a or b in total. Thus, pain-derived PDMs do not mediate the relationship between sound intensity and intensity ratings for ether type of sound. However, they are not perfectly selective as the expression of some PDMs correlates positively with ratings of sound unpleasantness (Path b). In summary, these results nevertheless indicate some degree of specificity to somatic pain vs. sound.

Comparison to other multivariate models

Previous studies have investigated the direct relationship between brain responses and pain reports, both using univariate (Coghill et al. 1999; Bornhövd et al. 2002; Ploner et al. 2010) and multivariate approaches (Marquand et al. 2010; Brodersen et al. 2012; Wager et al. 2013; Geuter et al. 2014; Woo et al. 2017). One study trained a multivariate pattern, termed the Neurological Pain Signature (NPS), which predicts pain reports from brain activity that can be easily applied to new data sets (Wager et al. 2013). In contrast to the present approach, the estimation of the NPS did not account for temperature-brain relationships; its goal was rather to predict pain intensity without demonstrating mediation. In addition, we compared it to the Stimulus Intensity Independent Pain Signature (SIIPS1), which was trained to predict pain ratings after removing linear effects of stimulus intensity on brain activity and ratings (Woo et al. 2017). Additionally, the combination of NPS and SIIPS1 as well as the Neurosynth reverse inference map for the term "pain" were compared.

To examine relationships among the PDM models and other established models, we compared prediction-outcome correlations (see Fig. 6D) Correlations were calculated across individual differences in response values for each model. The first PDM performed best (mean r=0.53), followed by the cPDM (mean r=0.50). The NPS, the combination of NPS and SIIPS1, and the Neurosynth reverse inference map performed roughly equivalently (mean r=0.26, 0.28, and 0.25, respectively). Finally, SIIPS1 performed the worst (mean r=0.13). Prediction performance between PDM1 and cPDM did not differ significantly, while both PDM-based models significantly outperformed the remaining models (q_{FDR}<0.05, all p<0.2e^{-11}, all $t_{(74)}$>8.4).

The cPDM was highly correlated with PDM1 (r=0.95), but only moderately correlated with NPS (r=0.64) and NeuroSynth (r=0.66). The NPS, by contrast, was more strongly correlated with Neurosynth (r=0.82) than the PDM models. The SIIPS1 pattern was distinct, and essentially uncorrelated with either the PDM models, NPS, or NeuroSynth (r<=0.15). This is expected, as the SIIPS was designed to be independent of stimulus intensity.

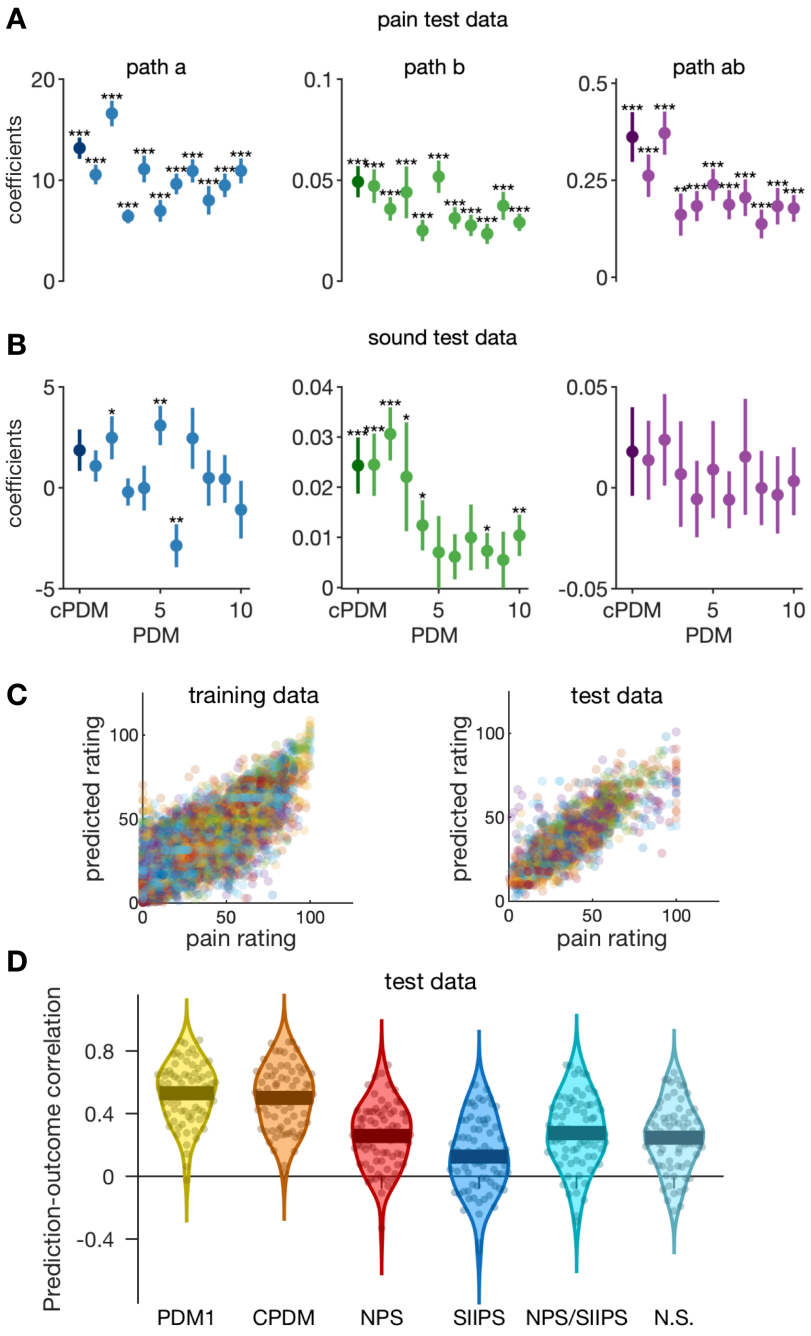

Figure 6. Validation on independent data (N=75). (**A**) The cPDM (dark circle) and all 10 individual PDMs (lighter circles) are significant mediators for independent pain test data. (**B**) PDMs show specificity with respect to aversive sounds because no indirect effect is significant here. (**C**) Scatter plots of pain predicted from the cPDM against empirical pain ratings for training (left) and test (right) pain data. Individual trials from all subjects are shown. Colors indicate different subjects. (**D**) The prediction-outcome correlations between reported pain and pattern responses for the first PDM (PDM1), the cPDM, the Neurological Pain Signature (NPS), the Stimulus Intensity Independent Pain Signature 1 (SIIPS1), the combination of NPS and SIIPS1, and the Neurosynth reverse inference map for the term “pain”.

Comparison to univariate mediation analysis

In contrast to the present multivariate PDM approach, mass-univariate mediation analyses of fMRI data estimate independent mediation models for each voxel (Wager et al. 2008; Atlas et al. 2014). The intersection of voxels with significant paths a, b, and ab is then interpreted as a set of mediating brain regions. In order to compare the novel high-dimensional PDM approach to the univariate mediation analysis, we first computed a mass-univariate mediation analysis on the training data set (Studies 1-7).

This analysis identified the MCC, cerebellum, posterior and mid-insula, S2, and S1 as brain mediators defined as the intersection of the coefficient maps for paths a, b, and ab at FDR $q < 0.05$ (Figure 7). Comparing these results to the cPDM revealed both similarities and some notable differences (Figure 2,7). Both maps include somatosensory regions in aMCC, insula, and S2, as well as cerebellum. Additional regions with positive weights in the cPDM included thalamus, PAG, and other midbrain regions like the PBN not included in the univariate model. Furthermore, negative contributions in SPL and S1 were not identified in the univariate model. Such results are expected if some brain regions make detectable contributions only after controlling for the influences of other brain regions; this is an advantage of multivariate predictive approaches to neuroimaging analysis.

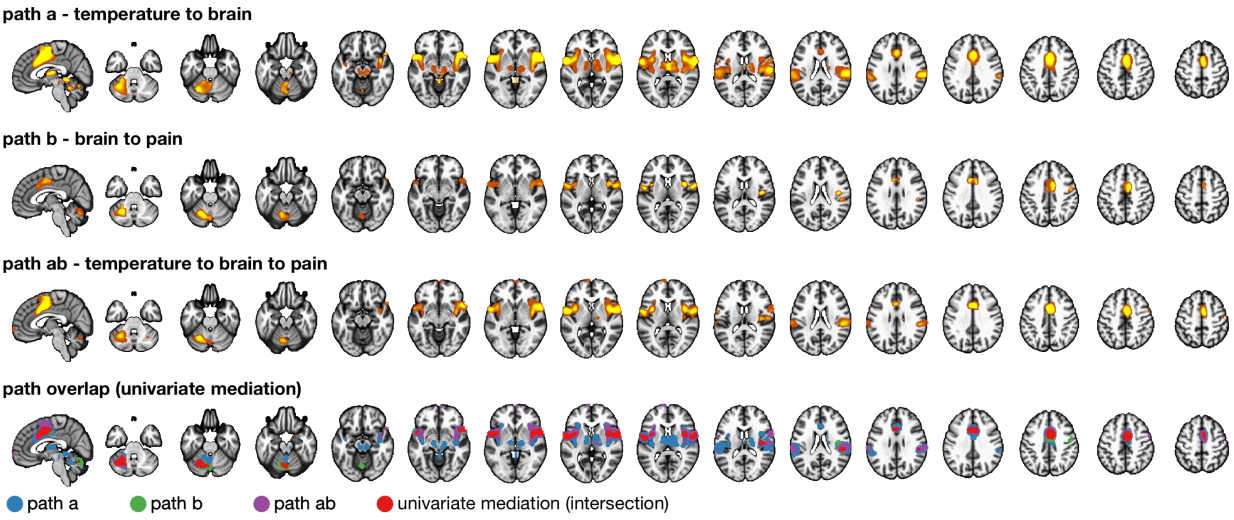

Figure 7. Comparison to univariate mediation analysis. Top three panels show individually significant voxels for paths a (blue), b (green), and ab (purple) from a univariate mediation analysis at FDR $q < 0.05$. Panel 4 shows voxels mediating the relationship between temperature and pain, i.e., the overlap between the three paths (red).

Discussion

In brief, our analyses identified brain mediators of pain that extend substantially beyond the boundaries of traditional 'pain matrix' regions, including prefrontal and midbrain regions previously thought to play an 'extra-nociceptive' or modulatory role (Kucyi et al. 2012; Geuter et al. 2013; Seminowicz and Moayedi 2017). Integrated activity across these pathways, as reflected in the Combined PDM, predict human pain intensity more accurately than previous pain predictive models.

Brain regions primarily associated with motivational, learning, or executive functions have been considered to modulate activity in a pain processing brain system. The involvement of those brain regions in pain processing has been shown in a multitude of studies (eg. Bushnell et al. 2013; Seminowicz and Moayedi 2017). Here, we show that many of these regions, including NAc, dlPFC, mPFC, and mOFC, are formal mediators of the stimulus-pain relationship. Their activity is directly related to stimulus intensity and at the same time to pain when controlling for stimulus intensity. The broad range of psychological functions associated with the regions serving as formal pain mediators is in line with the idea that brain regions related to motivational, learning, and executive functions are much more directly involved in the generation of pain. Their mediating role suggests that they are not necessarily external modulators to a distinct pain system but that primarily non-pain processing brain regions play more a direct role in generating the pain experience, further blurring the boundaries of a so-called 'pain-matrix'.

One of the most prominent functions of pain is its motivational drive since it is associated with tissue damage (Navratilova and Porreca 2014; Geuter et al. 2016). Learning about painful stimuli is thus important to learn to minimize future harm. Pain stimulation relates to activity in the NAc, a brain region associated with motivational learning (Becerra et al. 2013; Woo et al. 2015). In line with the NAc's role in pain in humans (Baliki et al. 2010, 2012) and animal models (Chang et al. 2014; Navratilova and Porreca 2014; Schwartz et al. 2014; Ren et al. 2016), NAc also acts as a formal mediator between nociceptive stimuli and pain. Furthermore, we show that NAc function for pain is based on opposing relationships of NAc activity with stimulus intensity (negative) and pain (positive) – NAc shows stimulus intensity-dependent deactivation, with larger de-activation mediating decreased pain, consistent with regulatory negative feedback mechanisms. The NAc might exert its control in this feedback loop indirectly via its connections with the hypothalamus or mPFC as indicated by studies in humans and animals (Baliki et al. 2012; Schwartz et al. 2014; Lee et al. 2015; Woo et al. 2015). However, the exact contribution of

the NAc to pain perception might rely on more complex temporal dynamics that cannot be resolved in the current data set and are still a matter of debate (Baliki et al. 2010; Becerra et al. 2013) as is its role in aversive learning more generally (Roy et al. 2014; Matsumoto et al. 2016).

Notably, another novel feature of the present cPDM map is that it contains positive weights in the bilateral PBN and specific parts of the amygdala, as well as RVM and PAG. Though definitive localization to these nuclei is difficult with any human method, activation is consistent with their locations in atlases and previous studies, and identifying them robustly in humans could provide an important step forward in the ability to study both bottom-up and top-down effects on crucial nociceptive and pain-modulatory pathways.

The PAG and RVM receive nociceptive afferents and form a major descending bulbospinal that which controls the balance of descending pain-inhibitory and facilitatory projections to the spinal dorsal horn (Fields 2004; De Felice et al. 2011; Wager and Atlas 2015; Geuter, Koban, et al. 2017). Imaging studies have identified PAG and RVM activation during both evoked pain and pain-modulatory conditions like placebo analgesia (Tracey et al. 2002; Eippert, Bingel, et al. 2009; Tinnermann et al. 2017). But these regions have not, to our knowledge, been identified as mediators of human stimulus-pain relationships.

The PBN has rarely been reported in neuroimaging studies, but is an emerging target of great importance in representing danger signals related to pain and other bodily input. The PBN is a major center for pain and other forms of interoception and chemical sensation, including taste, itch, dyspnea, and vagally mediated immune surveillance and sickness behavior (Goehler et al. 2000; Kelley et al. 2003). A major pathway composed of CGRP neurons projects from the PBN to the central nucleus of the amygdala. This pathway is activated in response to danger signals across multiple sensory modalities, including visceral and cutaneous pain (chemical, mechanical, thermal and electrical) and itch, across ascending trigeminal, spinal and vagal sensory pathways(Han et al. 2015; Campos et al. 2018). It is crucial for representing danger signals that produce avoidance behavior, including joint control of learned pain avoidance and food intake, which is strongly inhibited by its activation (Han et al. 2015; Sato et al. 2015; Campos et al. 2018). PBN also plays a role in pain regulation by sending non-CGRP projections to the rostral ventral medulla (RVM) (Roeder et al. 2016). PBN stimulation in humans can reduce chronic pain (Katayama et al. 1985). FMRI activation consistent with human PBN has been found to respond to noxious stimulation, correlate with human parasympathetic activity (Napadow et al. 2008), and respond to vagal stimulation (Frangos et al. 2015) and acupuncture (Napadow et al. 2009). Our results suggest it may be possible to measure activity in human PBN-central

amygdala and other rubrospinal pathways.

Among regions commonly associated with pain in neuroimaging studies, including the medial thalamus, PAG, S2, insula, MCC, SMA, and S1 (Dum et al. 2009), activity increased due to increasing temperatures and higher activity was related to stronger pain, controlling for temperature. This set of pain associated regions (Apkarian et al. 2005; Bushnell et al. 2013; Duerden and Albanese 2013; Jensen et al. 2016) was complemented by anterior temporal regions and the cerebellum, which share the same functional response profile. A positive relationship with both temperature and pain rating is in line with a traditional, feed-forward encoding view of nociception (Bushnell et al. 2013; Atlas et al. 2014; cf. Geuter, Boll, et al. 2017).

By contrast, the mPFC, SPL, RSC, precuneus, and parts of S1 and M1 were negatively related to both temperature and pain. The mPFC, RSC, and precuneus are part of the DMN, which has been associated with mind-wandering and internal thoughts (Andrews-Hanna et al. 2010; Kucyi and Davis 2015). The negative mediating role of the DMN regions could be related to the disruption of ongoing thought processes by the painful stimulation or attentional refocusing from internal to external sensations. The observation that some regions, like S1, participate in both, positive and negative, mediation relationships, may indicating the mixing of signals from distinct neuronal populations within single fMRI voxels. Such mixing of activity patterns is also observed across components from ICA.

In addition, there are multiple sources of endogenous variation in drivers of pain beyond stimulus intensity, including attention, arousal, and endogenous variation in ascending spinal afferents due to processing within the spinal cord (Eippert, Finsterbusch, et al. 2009; Geuter and Büchel 2013; Tinnermann et al. 2017). Because each study included a different psychological manipulation to modulate pain, our approach will only identify brain mediators common to all studies. Differences across studies thus increase the robustness of the identified brain mediators, but also reduce the amount of variance that will be explained in each single study. Additional factors introducing variance across studies include context-effects (Leknes et al. 2013), temporal effects within and across trials (Jepma et al. 2014), and pre-stimulus fluctuations in brain activity (Ploner et al. 2010). Variation in pain related to variation in afferent input, e.g., endogenous trial-to-trial variation in spinal cord processing, will be captured in path b in the mediation model. Other endogenous sources of variation unrelated to afferent input may be captured in the direct effect (path c') and will not be reflected in a brain mediation model that seeks to connect noxious stimulus intensity with pain. Similarly, potential non-linear relationships between stimulus intensity and pain or differences across participants might not be adequately

represented by linear brain mediators. However, using linear models instead of non-linear models offers better interpretability.

In summary, the new high-dimensional mediation analysis revealed a comprehensive picture of brain responses underlying the complex, multi-faceted pain experience. Several brain regions, such as the mPFC, thalamus, NAc, and PBN are shown to directly and formally mediate stimulus-to-pain relationships. The functional diversity of the brain mediators observed here offers a better understanding of the brain responses underlying the complexity of the pain experience.

Acknowledgements

This work was supported by the German Research Foundation (DFG) ("GE 2774/1-1" to S.G.) and the National Institutes of Health, which supported this work under grants "R01DA035484" (T.D.W.), "2R01MH076136" (T.D.W.), "R01DA027794" (T.D.W.), "R01 EB016061" (M.A.L.), "R01 EB026549" (M.A.L.), and "P41 EB015909" (M.A.L.).

Author Contributions

S.G., T.D.W, and M.A.L. designed the study. M.A.L. contributed unpublished analytical tools. S.G. conducted data analysis and drafted the manuscript. S.G., T.D.W., M.A.L., and E.A.R.L. edited and revised the manuscript. E.A.R.L., M.R., L.Y.A., L.S., A.K., and L.K. curated neuroimaging data and provided comments on the manuscript.

References

Andrews-Hanna JR, Reidler JS, Huang C, Buckner RL. 2010. Evidence for the Default Network's Role in Spontaneous Cognition. Journal of Neurophysiology. 104:322–335.

Apkarian AV, Bushnell MC, Treede R-D, Zubieta J-K. 2005. Human brain mechanisms of pain perception and regulation in health and disease. European Journal of Pain. 9:463–484.

Atlas LY, Bolger N, Lindquist MA, Wager TD. 2010. Brain Mediators of Predictive Cue Effects on Perceived Pain. The Journal of Neuroscience. 30:12964–12977.

Atlas LY, Lindquist MA, Bolger N, Wager TD. 2014. Brain mediators of the effects of noxious heat on pain. PAIN. 155:1632–1648.

Baliki MN, Geha PY, Apkarian AV. 2009. Parsing pain perception between nociceptive representation and magnitude estimation. J Neurophysiol. 101:875–887.

Baliki MN, Geha PY, Fields HL, Apkarian AV. 2010. Predicting Value of Pain and Analgesia: Nucleus Accumbens Response to Noxious Stimuli Changes in the Presence of Chronic Pain. Neuron. 66:149–160.

Baliki MN, Petre B, Torbey S, Herrmann KM, Huang L, Schnitzer TJ, Fields HL, Apkarian AV. 2012. Corticostriatal functional connectivity predicts transition to chronic back pain. Nature Neuroscience. 15:1117–1119.

Becerra L, Navratilova E, Porreca F, Borsook D. 2013. Analogous responses in the nucleus accumbens and cingulate cortex to pain onset (aversion) and offset (relief) in rats and humans. J Neurophysiol. 110:1221–1226.

Bingel U, Quante M, Knab R, Bromm B, Weiller C, Büchel C. 2002. Subcortical structures involved in pain processing: evidence from single-trial fMRI. Pain. 99:313–321.

Bornhövd K, Quante M, Glauche V, Bromm B, Weiller C, Büchel C. 2002. Painful stimuli evoke different stimulus–response functions in the amygdala, prefrontal, insula and somatosensory cortex: a single-trial fMRI study. Brain. 125:1326–1336.

Bradley MM, Lang PJ. 2007. The International Affective Digitized Sounds (; IADS-2): Affective ratings of sounds and instruction manual. University of Florida, Gainesville, FL, Tech Rep B-3.

Brodersen KH, Wiech K, Lomakina EI, Lin C, Buhmann JM, Bingel U, Ploner M, Stephan KE, Tracey I. 2012. Decoding the perception of pain from fMRI using multivariate pattern analysis. NeuroImage. 63:1162–1170.

Brooks JCW, Davies W-E, Pickering AE. 2017. Resolving the Brainstem Contributions to Attentional Analgesia. J Neurosci. 37:2279–2291.

Büchel C, Bornhövd K, Quante M, Glauche V, Bromm B, Weiller C. 2002. Dissociable Neural Responses Related to Pain Intensity, Stimulus Intensity, and Stimulus Awareness within the Anterior Cingulate Cortex: A Parametric Single-Trial Laser Functional Magnetic Resonance Imaging Study. The Journal of Neuroscience. 22:970–976.

Bushnell MC, Čeko M, Low LA. 2013. Cognitive and emotional control of pain and its disruption in chronic pain. Nat Rev Neurosci. 14:502–511.

Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafò MR. 2013. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci. 14:365–376.

Caffo BS, Crainiceanu CM, Verduzco G, Joel S, Mostofsky SH, Bassett SS, Pekar JJ. 2010. Two-stage decompositions for the analysis of functional connectivity for fMRI with application to Alzheimer's disease risk. NeuroImage. 51:1140–1149.

Campos CA, Bowen AJ, Roman CW, Palmiter RD. 2018. Encoding of danger by parabrachial CGRP neurons. Nature. 555:617–622.

Chang P-C, Pollema-Mays SL, Centeno MV, Procissi D, Contini M, Baria AT, Martina M, Apkarian AV. 2014. Role of nucleus accumbens in neuropathic pain: Linked multi-scale evidence in the rat transitioning to neuropathic pain. PAIN. 155:1128–1139.

Chén OY, Crainiceanu C, Ogburn EL, Caffo BS, Wager TD, Lindquist MA. 2017. High-dimensional multivariate mediation with application to neuroimaging data. Biostatistics.

Coghill RC, Sang CN, Maisog JM, Iadarola MJ. 1999. Pain Intensity Processing Within the Human Brain: A Bilateral, Distributed Mechanism. Journal of Neurophysiology. 82:1934–1943.

Craig AD, Chen K, Bandy D, Reiman EM. 2000. Thermosensory activation of insular cortex. Nat Neurosci. 3:184–190.

Crainiceanu CM, Caffo BS, Luo S, Zipunnikov VM, Punjabi NM. 2011. Population Value Decomposition, a Framework for the Analysis of Image Populations. Journal of the American Statistical Association. 106:775–790.

De Felice M, Sanoja R, Wang R, Vera-Portocarrero L, Oyarzo J, King T, Ossipov MH, Vanderah TW, Lai J, Dussor GO, Fields HL, Price TJ, Porreca F. 2011. Engagement of descending inhibition from the rostral ventromedial medulla protects against chronic neuropathic pain. PAIN. 152:2701–2709.

Duerden EG, Albanese M-C. 2013. Localization of pain-related brain activation: A meta-analysis of neuroimaging data. Hum Brain Mapp. 34:109–149.

Dum RP, Levinthal DJ, Strick PL. 2009. The Spinothalamic System Targets Motor and Sensory Areas in the Cerebral Cortex of Monkeys. J Neurosci. 29:14223–14235.

Eippert F, Bingel U, Schoell ED, Yacubian J, Klinger R, Lorenz J, Büchel C. 2009. Activation of the opioidergic descending pain control system underlies placebo analgesia. Neuron. 63:533–543.

Eippert F, Finsterbusch J, Bingel U, Büchel C. 2009. Direct evidence for spinal cord involvement in placebo analgesia. Science. 326:404.

Fairhurst M, Wiech K, Dunckley P, Tracey I. 2007. Anticipatory brainstem activity predicts neural processing of pain in humans. Pain. 128:101–110.

Fields HL. 2004. State-dependent opioid control of pain. Nature Reviews Neuroscience. 5:565–575.

Frangos E, Ellrich J, Komisaruk BR. 2015. Non-invasive Access to the Vagus Nerve Central Projections via Electrical Stimulation of the External Ear: fMRI Evidence in Humans. Brain Stimul. 8:624–636.

Geuter S, Boll S, Eippert F, Büchel C. 2017. Functional dissociation of stimulus intensity encoding and predictive coding of pain in the insula. eLife. 6:e24770.

Geuter S, Büchel C. 2013. Facilitation of Pain in the Human Spinal Cord by Nocebo Treatment. The Journal of Neuroscience. 33:13784–13790.

Geuter S, Cunningham JT, Wager TD. 2016. Disentangling opposing effects of motivational states on pain perception. PAIN Reports. 1:e574.

Geuter S, Eippert F, Hindi Attar C, Büchel C. 2013. Cortical and subcortical responses to high and low effective placebo treatments. NeuroImage. 67:227–236.

Geuter S, Gamer M, Onat S, Büchel C. 2014. Parametric trial-by-trial prediction of pain by easily available physiological measures. PAIN. 155:994–1001.

Geuter S, Koban L, Wager TD. 2017. The Cognitive Neuroscience of Placebo Effects: Concepts, Predictions, and Physiology. Annu Rev Neurosci. 40:167–188.

Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, Ugurbil K, Andersson J, Beckmann CF, Jenkinson M, Smith SM, Van Essen DC. 2016. A multi-modal parcellation of human cerebral cortex. Nature. 536:171–178.

Goehler LE, Gaykema RP, Hansen MK, Anderson K, Maier SF, Watkins LR. 2000. Vagal immune-to-brain communication: a visceral chemosensory pathway. Auton Neurosci. 85:49–59.

Han S, Soleiman MT, Soden ME, Zweifel LS, Palmiter RD. 2015. Elucidating an Affective Pain Circuit that Creates a Threat Memory. Cell. 162:363–374.

Haxby JV, Connolly AC, Guntupalli JS. 2014. Decoding Neural Representational Spaces Using Multivariate Pattern Analysis. Annual Review of Neuroscience. 37:435–456.

Horing B, Sprenger C, Büchel C. 2019. The parietal operculum preferentially encodes heat pain

and not salience. PLOS Biology. 17:e3000205.

Jensen KB, Regenbogen C, Ohse MC, Frasnelli J, Freiherr J, Lundström JN. 2016. Brain activations during pain: a neuroimaging meta-analysis of patients with pain and healthy controls. PAIN. 157:1279–1286.

Jepma M, Jones M, Wager TD. 2014. The Dynamics of Pain: Evidence for Simultaneous Site-Specific Habituation and Site-Nonspecific Sensitization in Thermal Pain. The Journal of Pain. 15:734–746.

Katayama Y, Tsubokawa T, Hirayama T, Yamamoto T. 1985. Pain relief following stimulation of the pontomesencephalic parabrachial region in humans: brain sites for nonopiate-mediated pain control. Appl Neurophysiol. 48:195–200.

Kelley KW, Bluthé R-M, Dantzer R, Zhou J-H, Shen W-H, Johnson RW, Broussard SR. 2003. Cytokine-induced sickness behavior. Brain, Behavior, and Immunity, Biological Mechanisms of Psychosocial Effects on Disease: Implications for Cancer Control. 17:112–118.

Keuken MC, Bazin P-L, Crown L, Hootsmans J, Laufer A, Müller-Axt C, Sier R, van der Putten EJ, Schäfer A, Turner R, Forstmann BU. 2014. Quantifying inter-individual anatomical variability in the subcortex using 7 T structural MRI. NeuroImage. 94:40–46.

Koban L, Jepma M, López-Solà M, Wager TD. 2019. Different brain networks mediate the effects of social and conditioned expectations on pain. Nat Commun. 10:1–13.

Kober H, Barrett LF, Joseph J, Bliss-Moreau E, Lindquist K, Wager TD. 2008. Functional grouping and cortical–subcortical interactions in emotion: A meta-analysis of neuroimaging studies. NeuroImage. 42:998–1031.

Koyama T, McHaffie JG, Laurienti PJ, Coghill RC. 2003. The single-epoch fMRI design: validation of a simplified paradigm for the collection of subjective ratings. Neuroimage. 19:976–987.

Kriegeskorte N. 2011. Pattern-information analysis: From stimulus decoding to computational-model testing. NeuroImage. in press.

Kucyi A, Davis KD. 2015. The dynamic pain connectome. Trends in Neurosciences. 38:86–95.

Kucyi A, Hodaie M, Davis KD. 2012. Lateralization in intrinsic functional connectivity of the temporoparietal junction with salience- and attention-related brain networks. Journal of Neurophysiology. 108:3382–3392.

Lee M, Manders TR, Eberle SE, Su C, D'amour J, Yang R, Lin HY, Deisseroth K, Froemke RC, Wang J. 2015. Activation of Corticostriatal Circuitry Relieves Chronic Neuropathic Pain. J Neurosci. 35:5247–5259.

Leknes S, Berna C, Lee MC, Snyder GD, Biele G, Tracey I. 2013. The importance of context:

When relative relief renders pain pleasant. Pain. 154:402–410.

Liang M, Su Q, Mouraux A, Iannetti GD. 2019. Spatial Patterns of Brain Activity Preferentially Reflecting Transient Pain and Stimulus Intensity. Cereb Cortex. 29:2211–2227.

Lindquist MA, Krishnan A, López-Solà M, Jepma M, Woo C-W, Koban L, Roy M, Atlas LY, Schmidt L, Chang LJ, Reynolds Losin EA, Eisenbarth H, Ashar YK, Delk E, Wager TD. 2017. Group-regularized individual prediction: theory and application to pain. NeuroImage, Individual Subject Prediction. 145, Part B:274–287.

MacKinnon DP, Krull JL, Lockwood CM. 2000. Equivalence of the Mediation, Confounding and Suppression Effect. Prev Sci. 1:173–181.

Marquand A, Howard M, Brammer M, Chu C, Coen S, Mourão-Miranda J. 2010. Quantitative prediction of subjective pain intensity from whole-brain fMRI data using Gaussian processes. NeuroImage. 49:2178–2189.

Matsumoto H, Tian J, Uchida N, Watabe-Uchida M. 2016. Midbrain dopamine neurons signal aversion in a reward-context-dependent manner. eLife. 5:e17328.

Morel A, Magnin M, Jeanmonod D. 1997. Multiarchitectonic and stereotactic atlas of the human thalamus. Journal of Comparative Neurology. 387:588–630.

Mumford JA, Turner BO, Ashby FG, Poldrack RA. 2012. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. NeuroImage. 59:2636–2643.

Napadow V, Dhond R, Conti G, Makris N, Brown EN, Barbieri R. 2008. Brain correlates of autonomic modulation: combining heart rate variability with fMRI. Neuroimage. 42:169–177.

Napadow V, Dhond R, Park K, Kim J, Makris N, Kwong KK, Harris RE, Purdon PL, Kettner N, Hui KKS. 2009. Time-variant fMRI activity in the brainstem and higher structures in response to acupuncture. Neuroimage. 47:289–301.

Navratilova E, Porreca F. 2014. Reward and motivation in pain and pain relief. Nature Neuroscience. 17:1304–1312.

Peyron R, Frot M, Schneider F, Garcia-Larrea L, Mertens P, Barral FG, Sindou M, Laurent B, Mauguière F. 2002. Role of Operculoinsular Cortices in Human Pain Processing: Converging Evidence from PET, fMRI, Dipole Modeling, and Intracerebral Recordings of Evoked Potentials. NeuroImage. 17:1336–1346.

Ploner M, Lee MC, Wiech K, Bingel U, Tracey I. 2010. Prestimulus functional connectivity determines pain perception in humans. Proceedings of the National Academy of Sciences. 107:355–360.

Pouget A, Dayan P, Zemel R. 2000. Information processing with population codes. Nat Rev

Neurosci. 1:125–132.

Ren W, Centeno MV, Berger S, Wu Y, Na X, Liu X, Kondapalli J, Apkarian AV, Martina M, Surmeier DJ. 2016. The indirect pathway of the nucleus accumbens shell amplifies neuropathic pain. Nat Neurosci. 19:220–222.

Rissman J, Greely HT, Wagner AD. 2010. Detecting individual memories through the neural decoding of memory states and past experience. PNAS. 107:9849–9854.

Roeder Z, Chen Q, Davis S, Carlson JD, Tupone D, Heinricher MM. 2016. Parabrachial complex links pain transmission to descending pain modulation: PAIN. 157:2697–2708.

Roy M, Shohamy D, Daw N, Jepma M, Wimmer GE, Wager TD. 2014. Representation of aversive prediction errors in the human periaqueductal gray. Nat Neurosci. 17:1607–1612.

Sato M, Ito M, Nagase M, Sugimura YK, Takahashi Y, Watabe AM, Kato F. 2015. The lateral parabrachial nucleus is actively involved in the acquisition of fear memory in mice. Mol Brain. 8:22.

Schaefer A, Kong R, Gordon EM, Laumann TO, Zuo X-N, Holmes AJ, Eickhoff SB, Yeo BTT. 2018. Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. Cereb Cortex. 28:3095–3114.

Schwartz N, Temkin P, Jurado S, Lim BK, Heifets BD, Polepalli JS, Malenka RC. 2014. Decreased motivation during chronic pain requires long-term depression in the nucleus accumbens. Science. 345:535–542.

Seminowicz DA, Moayedi M. 2017. The Dorsolateral Prefrontal Cortex in Acute and Chronic Pain. The Journal of Pain. 18:1027–1035.

Shelton L, Becerra L, Borsook D. 2012. Unmasking the mysteries of the habenula in pain and analgesia. Progress in Neurobiology. 96:208–219.

Shepard RN. 1980. Multidimensional Scaling, Tree-Fitting, and Clustering. Science. 210:390–398.

Tinnermann A, Geuter S, Sprenger C, Finsterbusch J, Büchel C. 2017. Interactions between brain and spinal cord mediate value effects in nocebo hyperalgesia. Science. 358:105–108.

Tracey I, Ploghaus A, Gati JS, Clare S, Smith S, Menon RS, Matthews PM. 2002. Imaging Attentional Modulation of Pain in the Periaqueductal Gray in Humans. J Neurosci. 22:2748–2752.

Villemure C, Slotnick BM, Bushnell MC. 2003. Effects of odors on pain perception: deciphering the roles of emotion and attention. Pain. 106:101–108.

Wager TD, Atlas LY. 2015. The neuroscience of placebo effects: connecting context, learning and health. Nat Rev Neurosci. 16:403–418.

Wager TD, Atlas LY, Lindquist MA, Roy M, Woo C-W, Kross E. 2013. An fMRI-Based Neurologic Signature of Physical Pain. New England Journal of Medicine. 368:1388–1397.

Wager TD, Davidson ML, Hughes BL, Lindquist MA, Ochsner KN. 2008. Prefrontal-Subcortical Pathways Mediating Successful Emotion Regulation. Neuron. 59:1037–1050.

Wager TD, Rilling JK, Smith EE, Sokolik A, Casey KL, Davidson RJ, Kosslyn SM, Rose RM, Cohen JD. 2004. Placebo-induced changes in FMRI in the anticipation and experience of pain. Science. 303:1162–1167.

Wager TD, Waugh CE, Lindquist M, Noll DC, Fredrickson BL, Taylor SF. 2009. Brain mediators of cardiovascular responses to social threat: Part I: Reciprocal dorsal and ventral sub-regions of the medial prefrontal cortex and heart-rate reactivity. NeuroImage, Brain Body Medicine. 47:821–835.

Willis WD, Westlund KN. 1997. Neuroanatomy of the Pain System and of the Pathways That Modulate Pain. Journal of Clinical Neurophysiology Neurophysiology of Pain. 14:2–31.

Woo C-W, Roy M, Buhle JT, Wager TD. 2015. Distinct Brain Systems Mediate the Effects of Nociceptive Input and Self-Regulation on Pain. PLoS Biol. 13:e1002036.

Woo C-W, Schmidt L, Krishnan A, Jepma M, Roy M, Lindquist MA, Atlas LY, Wager TD. 2017. Quantifying cerebral contributions to pain beyond nociception. Nature Communications. 8:14211.

Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD. 2011. Large-scale automated synthesis of human functional neuroimaging data. Nat Meth. 8:665–670.

Multiple brain networks mediating stimulus-pain relationships in humans

— Supplementary Materials

Stephan Geuter[1,2,*], Elizabeth A. Reynolds Losin[3], Mathieu Roy[4], Lauren Y. Atlas[5,6], Liane Schmidt[7], Anjali Krishnan[8], Leonie Koban[2,9], Tor D. Wager[2,9,10,#], Martin A. Lindquist[1,#]

[1] Department of Biostatistics, Johns Hopkins University, USA

[2] Institute of Cognitive Science, University of Colorado Boulder, USA

[3] Department of Psychology, University of Miami, USA

[4] Department of Psychology, McGill University, Canada

[5] National Center for Complementary and Integrative Health, National Institutes of Health, USA

[6] National Center for Drug Abuse, National Institutes of Health, USA

[7] Social-and-Affective Neuroscience Team, Institute du Cerveau et de la Moelle Epinière, INSERM UMR 1127, CNRS UMR 7225, Université Pierre et Marie Curie Paris 6, France

[8] Department of Psychology, Brooklyn College of the City University of New York, USA

[9] Department of Psychology and Neuroscience, University of Colorado Boulder, USA

[10] Presidential Cluster in Neuroscience and Department of Psychological and Brain Sciences, Dartmouth College, Hanover, USA

* Corresponding author:
Stephan Geuter
Department of Biostatistics
Johns Hopkins University
615 N Wolfe Street, Baltimore, MD 21205, USA
Email: sgeuter@jhmi.edu
Phone: +1 (443) 287-8791

[#] Authors contributed equally to this work

Supplementary Materials and Methods

Procedures

Thermal and aversive sound stimulation. The number of noxious stimulation trials, stimulation sites, inter-trial intervals, rating scales, and stimulus intensities and durations varied across studies, but were comparable; these variables are summarized in Tables S2 and S3. Each study also comprised a specific psychological manipulation (except Study 8), such as placebo treatment, which will be or has been reported elsewhere (Table S1).

In each study, except Studies 7 and 8, thermal stimulation was delivered to multiple skin sites using a TSA-II Neurosensory Analyzer (Medoc Ltd., Chapel Hill, NC) with a 16 mm Peltier thermode endplate. A PATHWAY system (Medoc Ltd., Chapel Hill, NC) was used in Studies 7 and 8. Study 7 used a circular CHEPS Peltier endplate (diameter: 32 mm) and study 8 used a 16 mm ATS Peltier endplate. On every trial, after the offset of stimulation, participants rated the magnitude of the warmth or pain they had felt during the trial on a visual analog scale. Participants in Study 8 rated their pain continuously during stimulation. The maximum rating of each trial was used in the following analyses. Other thermal stimulation parameters varied across studies, with stimulation temperatures ranging from 40.8 °C to 50 °C and stimulation durations from 1.85 to 12.5 s. Most studies applied thermal stimulation to the forearm. See Table S2 for stimulation intensity levels, mean temperature for each intensity level, and details of the rating scales. See Table S3 for stimulation duration, duration of inter-stimulus interval, number and location of stimulation sites, and number of trials per subject.

fMRI data processing

Preprocessing. Structural T1-weighted images were co-registered to the mean functional image for each subject using the iterative mutual information-based algorithm implemented in SPM **(Ashburner and Friston 2005)**, and then normalized to MNI space using SPM. The version of SPM used varied across studies (Studies 1 and 6 used SPM5; while all other studies used SPM8; http://www.fil.ion.ucl.ac.uk/spm/). Following normalization, Studies 1 and 6 included an additional step of normalization to the group mean using a genetic algorithm-based normalization **(Wager and Nichols 2003)**. We chose to retain the original preprocessing used in each published paper for two reasons: (1) to establish, and test, robustness across minor variations in processing pipelines; and (2) because study-specific analysis choices are

appropriate in some cases, depending on the distribution of the data and study design.

For each functional dataset, initial volumes were removed to allow for image intensity stabilization (see Lindquist et al. (2017) for details). In addition, volumes with signal values that were outliers within the time series (i.e., "spikes") were removed. To identify outliers, both the mean and the standard deviation of intensity values across each slice were computed for each image. The Mahalanobis distances for the matrix of (concatenated) slice-wise mean and standard deviation values by functional volumes (over time) were computed, and values with a significant $\chi 2$ value (corrected for multiple comparisons based false discovery rate) were considered outliers. In practice, less than 1% of images were deemed outliers. The output of this procedure was later included as nuisance covariates in the subject level models. Next, functional images were corrected for differences in the acquisition timing of each slice (except for multiband data with a short TR of 480 ms in Study 8) and were motion-corrected (realigned) using SPM. The functional images were warped to SPM's normative atlas (warping parameters estimated from co-registered, high-resolution structural images), interpolated to $2 \times 2 \times 2\, mm^3$ voxels, and smoothed with an 8 mm FWHM Gaussian kernel.

Single trial analysis (Except Studies 3 and 6). For each study, a single trial, or "single-epoch", design and analysis approach was used to model the data. Quantification of single-trial response magnitudes was done by constructing a GLM design matrix with separate regressors for each trial **(Koyama et al. 2003; Rissman et al. 2010; Mumford et al. 2012)**. First, boxcar regressors, convolved with the canonical hemodynamic response function (HRF), were constructed to model cue and rating periods in each study. Regressors for each trial, as well as several types of nuisance covariates were also included. Because each trial consisted of relatively few volumes, trial estimates could be strongly affected by acquisition artifacts that occur during that trial (e.g. sudden motion, scanner pulse artifacts, etc.). Therefore, trial-by-trial variance inflation factors (VIFs; a measure of design-induced uncertainty due, in this case, to collinearity with nuisance regressors) were calculated, and any trials with VIFs exceeding 2.5 were excluded from the analyses (VIF threshold for Study 8 was 3.5 as in the primary publication). For Study 1, global outliers (trials that exceeded three standard deviations (SDs) above the mean) were also excluded, and a principal component based denoising step was employed during preprocessing to minimize artifacts. This generated single trial estimates that reflect the amplitude of the fitted HRF on each trial and refer to the magnitude pain-period activity for each trial in each voxel.

Single trial analysis (Studies 3 and 6). For Studies 3 and 6, single trial analyses were based

on fitting a set of three basis functions, rather than the standard canonical HRF used in the other studies. This flexible strategy allowed the shape of the modeled hemodynamic response function (HRF) to vary across trials and voxels. This procedure differed from that used in other studies because it maintains consistency with the procedures used in the original publications. For both Study 3 and Study 6, the pain period basis set consisted of three curves shifted in time and was customized for thermal pain responses based on previous studies **(Lindquist et al. 2009; Atlas et al. 2010)**. To estimate cue-evoked responses for Study 6, the pain anticipation period was modeled using a boxcar epoch convolved with a canonical HRF. This epoch was truncated at 8 s to ensure that fitted anticipatory responses were not affected by noxious stimulus-evoked activity. As in the other studies, nuisance covariates were included and trials with VIFs larger than 2.5 were excluded. In Study 6 trials that were global outliers (those that exceeded 3 SDs above the mean) were also excluded. The fitted basis functions from the flexible single trial approach were used to reconstruct the HRF and compute the area under the curve (AUC) for each trial and in each voxel. These trial-by-trial AUC values were used as estimates of trial-level pain-period activity.

Supplementary Figures

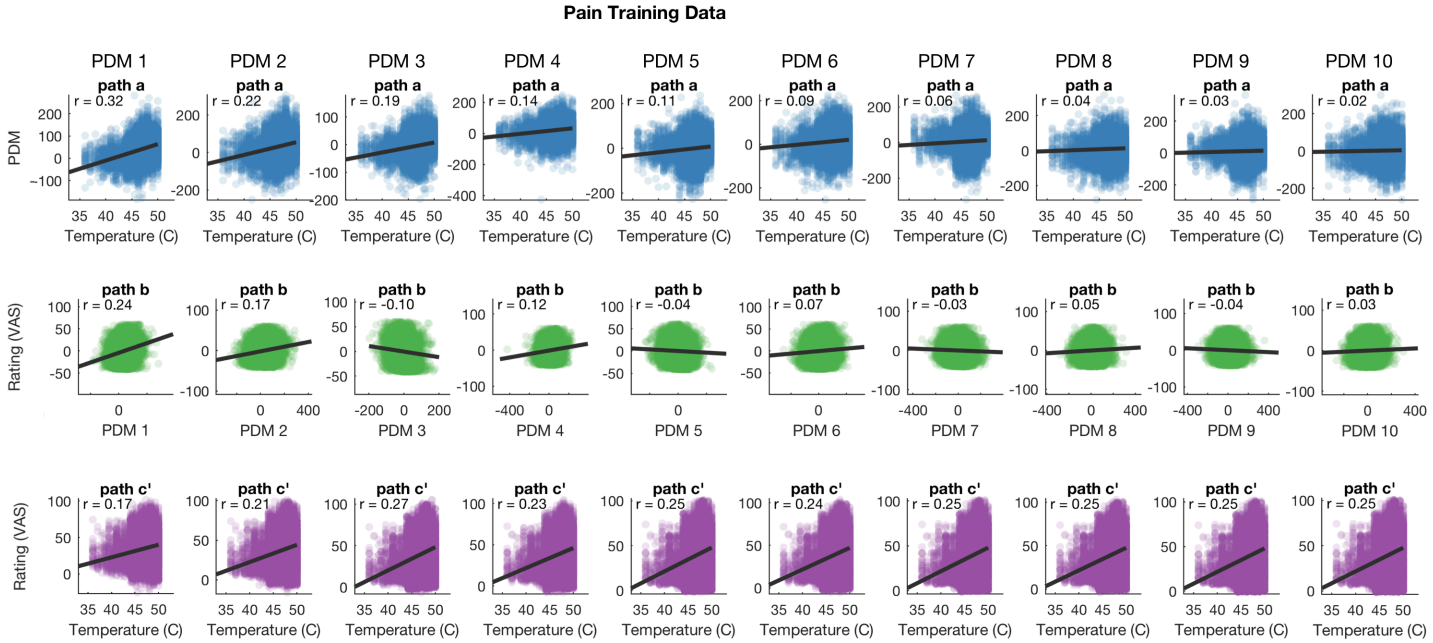

Figure S1. Bivariate relationships between temperatures, mediators (PDM expression), and pain ratings for the training data (studies 1-7). Data are adjusted according to the mediation equations, i.e., ratings in path b plots are adjusted for temperatures and PDMs, ratings in path c' plots are adjusted for PDMs, and PDMs in path b plots are adjusted for temperatures. PDMs are estimated on the training data.

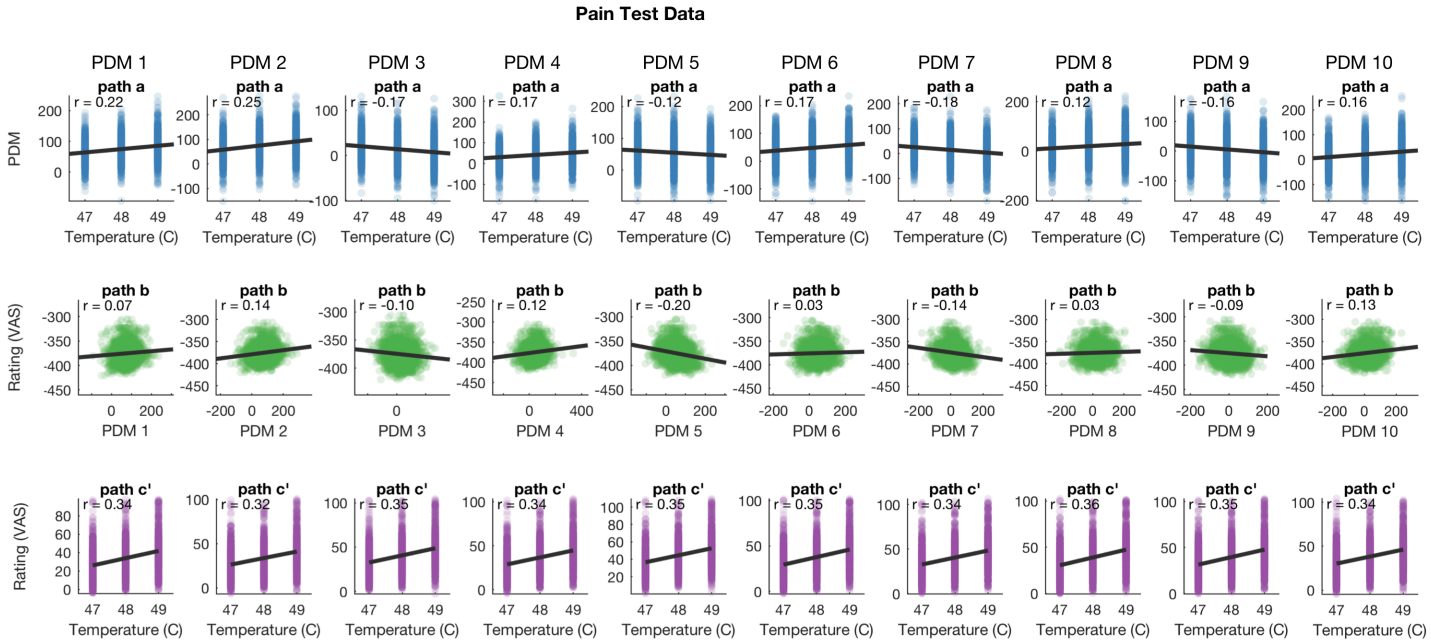

Figure S2. Bivariate relationships between temperatures, mediators (PDM expression), and pain ratings for the pain test data (Study 8, N = 75). Data are adjusted according to the mediation equations, i.e. ratings in path b plots are adjusted for temperatures and PDMs, ratings in path c' plots are adjusted for PDMs, and PDMs in path b plots are adjusted for temperatures. PDMs are estimated on the pain training data (studies 1-7).

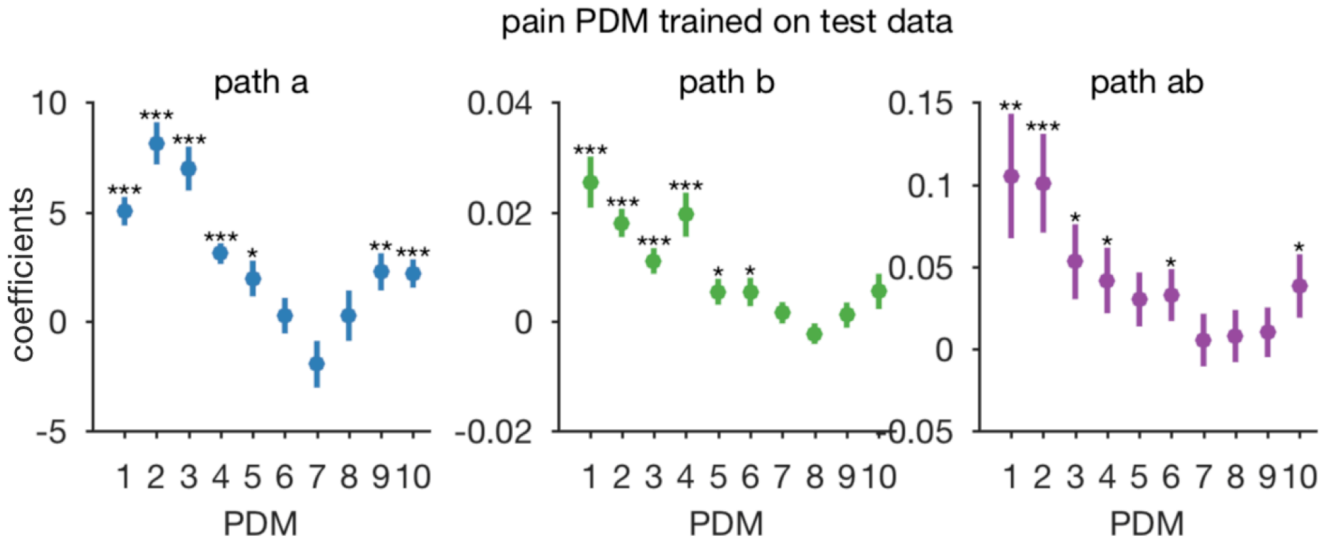

Figure S3. Generalization of pain PDMs from small sample to large sample. Here, test and training data sets were switched. Ten pain PDMs were estimated on the original test data set (study 8, N=75) and used as mediators in the original training data (studies 1-7, N=209). PDM 1-4, 6, and 10 are significant mediators for the larger set when trained on the smaller set.

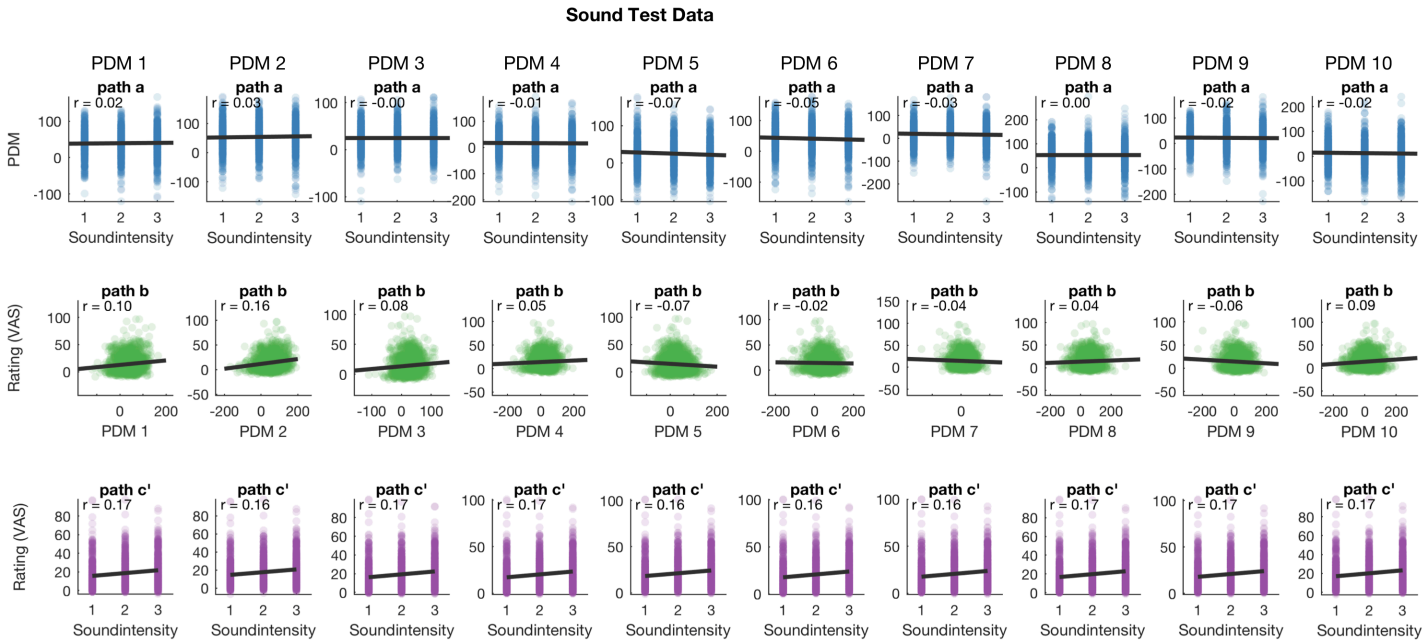

Figure S4. Bivariate relationships between sound intensity levels, mediators (PDM expression), and intensity ratings for the training data (studies 1-7). Data are adjusted according to the mediation equations, i.e. ratings in path b plots are adjusted for stimulus levels and PDM, ratings in path c' plots are adjusted for PDMs, and PDMs in path b plots are adjusted for stimulus levels. PDMs are estimated on the pain training data (studies 1-7).

Supplementary References

Ashburner J, Friston KJ. 2005. Unified segmentation. NeuroImage. 26:839–851.

Atlas LY, Bolger N, Lindquist MA, Wager TD. 2010. Brain Mediators of Predictive Cue Effects on Perceived Pain. The Journal of Neuroscience. 30:12964–12977.

Koyama T, McHaffie JG, Laurienti PJ, Coghill RC. 2003. The single-epoch fMRI design: validation of a simplified paradigm for the collection of subjective ratings. Neuroimage. 19:976–987.

Lindquist MA, Meng Loh J, Atlas LY, Wager TD. 2009. Modeling the hemodynamic response function in fMRI: Efficiency, bias and mis-modeling. NeuroImage, Mathematics in Brain Imaging. 45:S187–S198.

Mumford JA, Turner BO, Ashby FG, Poldrack RA. 2012. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. NeuroImage. 59:2636–2643.

Rissman J, Greely HT, Wagner AD. 2010. Detecting individual memories through the neural decoding of memory states and past experience. PNAS. 107:9849–9854.

Wager TD, Nichols TE. 2003. Optimization of experimental design in fMRI: a general framework using a genetic algorithm. NeuroImage. 18:293–309.

Supplementary Tables

Supplementary Table S1. *Demographics*

Study[§]	Sample Size	Sex	Mean age in Years (Std. Deviation)	Prior publications
PDM Training Data				
Study 1 (NSF)	26	9 F / 17 M	27.8	Atlas et al. (2014), *Pain* Wager et al. (2013) *NEJM*
Study 2 (BMRK3)	33	22 F / 11 M	27.9 (9.0)	Woo et al. (2015), *PLoS Biology* Wager et al. (2013) *NEJM*
Study 3 (BMRK4)	28	10 F / 18 M	25.2 (7.4)	Krishnan et al. (2016) *eLife*
Study 4 (IE)	50	27 F / 23 M	25.1 (6.9)	Roy et al. (2014), *Nat. Neurosci.*
Study 5 (ILCP)	29	16 F* / 12 M	20.4 (3.3)**	Woo et al. (*2017*) *Nat. Comms.*
Study 6 (EXP)	17	9 F / 8 M	25.5	Atlas et al. (2010), *J. Neurosci.*
Study 7 (SCEBL)	26	11 F / 15 M	28 (9.3)	Koban et al. (*2019), Nat. Comms.*
PDM Test Data				
Study 8 (BMRK5)	75	39 F / 36 M	28.2 (5.6)	Losin et al. (*in press, Nat. Hum. Beh.*)

Note. [§]Internal study codes to facilitate tracking of datasets; *Gender of one participant is unknown; **Age of one participant is unknown. Studies 1-7 have been reported on in Lindquist et al., 2017.

Supplementary Table S2. *Stimulation Parameters*

Study	Intensities	Mean Temperature by Intensity Level (Within Subject S.E.)	Rating scale	Mean Ratings by Intensity Level (Within Subject S.E.M.)
PDM Training Data				
Study 1 (NSF)	N, L, M, H (Calibrated)	40.8, 43.1, 45.1, 47.0 (0.16)	0-8 VAS (0, no sensation; 1, non-painful warmth; 2, low pain; 5, moderate pain; 8, maximum tolerable pain)	2.0, 2.8, 4.2, 6.6 (0.14)
Study 2 (BMRK3)	6 levels (Fixed)	44.3, 45.3, 46.3, 47.3, 48.3, 49.3	0-100 VAS	49.1, 56.6, 74.3, 99.4, 133.0, 159.3 (3.12)
Study 3 (BMRK4)	L, M, H (Fixed)	46.0, 47.0, 48.0	0-100 VAS (0, no sensation; 1.4, barely detectable; 6.1, weak; 17.2, moderate; 35.4, strong; 53.3, very strong; 100, strongest imaginable sensation)	UL: 31.7, 40.5, 53.6 (0.9787) LL: 31.5, 40.2, 53.3 (0.96)
Study 4 (IE)	L, M, H (Fixed)	46.0, 47.0, 48.0	0-100 VAS (0, no pain; 100, worst imaginable pain)	29.4, 38.9, 51.9 (0.64)
Study 5 (ILCP)	L, H (Calibrated)	44.7, 46.7 (0)	0-8 VAS (no pain to worst pain imaginable)	24.3, 46.7 (1.14)
Study 6 (EXP)	L, M, H (Calibrated)	41.2, 44.4, 47.2 (0.21)	0-8 VAS (0, no sensation; 1, non-painful warmth; 2, low pain; 5, moderate pain; 8, maximum tolerable pain)	2.5, 4.3, 7.4 (0.13)
Study 7 (SCEBL)	L, M, H (Fixed)	48, 49, 50	0-100 VAS (0, no pain; 100, worst imaginable pain)	26.0, 33.3, 40.4 (1.12)
PDM Test Data				
Study 8 (BMRK5)	L, M, H (Fixed)	47, 48, 49	0-100 gVAS (0, no experience; 100, strongest imaginable experience)	30.6, 39.9, 48.2 (1.64)

Note: Heat /pain levels: N = Nonpainful, L = Low, M = Medium, H = High. VAS = visual analogue scale. gVAS = generalized visual analogue scale.

Supplementary Table S3. *Task Characteristics*

Study	Duration (seconds)	Inter-heat interval (seconds)	Locations (number of sites)	Range of Number of Trials Per Subject	Mean proportion of trials excluded (Std. Deviation)	Other experimental manipulations
PDM Training Data						
Study 1 (NSF)	10	38	Left arm (3)	35-48	0.08 (0.07)	Masked emotional faces evenly crossed with temperature
Study 2 (BMRK3)	12.5	20.5-28.5	Left arm (2)	97	0.1 (0.04)	Cognitive self-regulation up and down
Study 3 (BMRK4)	11	25-27	Left arm (4), left foot (4)	81	0.08 (0.06)	Heat-predictive visual cues (low, medium, or high)
Study 4 (IE)	11	36-38	Left arm (6)	48	N/A	Heat-predictive visual cues; placebo manipulation
Study 5 (ILCP)	10	17-25	Left arm (2)	64	0.05 (0.03)	Agency (make choice, observe choice), Certainty (80% low pain, 50% low pain)
Study 6 (EXP)	10	38	Left arm (4)	61-64	0.03 (0.04)	Heat-predictive auditory cues
Study 7 (SCEBL)	1.85	26-37	Right leg (6)	96	0.04 (0.03)	Heat-predictive visual cues (low or high) and unreinforced social information
PDM Test Data						
Study 8 (BMRK5)	8, 11	11.5-32.75	Left arm (4)	30-36	0.04 (0.04)	Aversive sounds, modality-predictive cues (sound vs. heat)