

## Multiple brain networks mediating stimulus-pain relationships in humans

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

<sup>1</sup> Department of Biostatistics, Johns Hopkins University, USA

<sup>2</sup> Institute of Cognitive Science, University of Colorado Boulder, USA

<sup>3</sup> Department of Psychology, University of Miami, USA

<sup>4</sup> Department of Psychology, McGill University, Canada

<sup>5</sup> National Center for Complementary and Integrative Health, National Institutes of Health, USA

<sup>6</sup> National Center for Drug Abuse, National Institutes of Health, USA

<sup>7</sup> 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

<sup>8</sup> Department of Psychology, Brooklyn College of the City University of New York, USA

<sup>9</sup> Department of Psychology and Neuroscience, University of Colorado Boulder, USA

\* Corresponding author:

Stephan Geuter

Department of Biostatistics

Johns Hopkins University

615 N Wolfe Street, Baltimore, MD 21205, USA

Email: [sgeuter@jhmi.edu](mailto:sgeuter@jhmi.edu)

Phone: +1 (443) 287-8791

# Authors contributed equally to this work

## 1 **Abstract**

2       The brain transforms nociceptive input into a complex pain experience comprised of  
3 sensory, affective, motivational, and cognitive components. However, it is still unclear how pain  
4 arises from nociceptive input, and which brain networks coordinate to generate pain  
5 experiences. We introduce a new high-dimensional mediation analysis technique to estimate  
6 distributed, network-level patterns mediating the relationship between stimulus intensity and  
7 pain. In a large-scale analysis of functional magnetic resonance imaging data (N=284), we  
8 identify both traditional mediators in somatosensory brain regions and additional mediators  
9 located in prefrontal, midbrain, striatal, and default-mode regions unrelated to nociception in  
10 standard analyses. The whole brain mediators are specific for pain vs. aversive sounds and are  
11 organized in five functional networks. Brain mediators explain 32% more within-subject variance  
12 of single-trial pain ratings than previous brain-based models. Our results provide a new, broader  
13 view of the networks underlying pain experience, as well as distinct targets for interventions.

## 1 Introduction

2 The brain is central to the generation of pain; it transforms sensory input from peripheral  
3 receptors into a complex set of responses, including subjective experience, autonomic and  
4 neuroendocrine responses, avoidance behavior, and new learned stimulus-outcome and action-  
5 outcome associations. Neurophysiology and neuroimaging studies have identified brain regions  
6 that are targeted by afferent nociceptive pathways (Willis and Westlund, 1997; Apkarian et al.,  
7 2005; Dum et al., 2009), which are thought to encode sensory-discriminative and affective  
8 aspects of pain experience. But pain is a complex experience that entails not only sensory and  
9 emotional aspects, but also motivational, attentional and cognitive components. A full picture of  
10 the functional brain networks supporting these components of pain experience is still lacking.  
11 Here, we address this question using a new multivariate analysis method and a large functional  
12 magnetic resonance imaging (fMRI) dataset (N=284).

13 Although the complexity of pain is widely acknowledged, the underlying brain processes  
14 are often conceptualized as a unitary system that is activated by nociceptive input. Brain regions  
15 traditionally associated with pain include primary (S1) and secondary (S2) somatosensory,  
16 anterior midcingulate cortices (amMCC), medial and lateral thalamus, and posterior and mid-  
17 insular cortices (Apkarian et al., 2005; Schweinhardt and Bushnell, 2010; Jensen et al., 2016).  
18 But regions not directly targeted by afferent pathways are also activated by acute pain stimuli  
19 (Apkarian et al., 2005; Wager et al., 2013; Jensen et al., 2016; Seminowicz and Moayedi, 2017).  
20 For example the dorsolateral prefrontal cortex (dlPFC) – a brain region involved in high-level  
21 cognitive functions – responds to painful stimulation, shows alterations in chronic pain  
22 conditions, and contributes to placebo analgesia (Krummenacher et al., 2010; Bushnell et al.,  
23 2013; Seminowicz and Moayedi, 2017; Schafer et al., 2018). Other brain regions, including the  
24 ventromedial prefrontal cortex (vmPFC) (Wager et al., 2011; Roy et al., 2012; Geuter et al., 2017b)  
25 and the nucleus accumbens (NAc) (Baliki et al., 2012; Chang et al., 2014; Lee et al., 2015; Woo  
26 et al., 2015; Ren et al., 2016), key structures for reinforcement learning, also contribute to pain  
27 modulation. However, the exact role of these regions is not clear, and they are often thought of  
28 as external modulators of activity in the core pain system (Seminowicz and Moayedi, 2017). If  
29 so, these regions may be involved in the endogenous construction and regulation of pain in the  
30 brain, but they do not mediate the effects of nociceptive input on pain, i.e., they do not link  
31 nociception with pain (Woo et al., 2017).

32 Another view, in line with the ideas originally proposed by Melzack (1999), treats these  
33 regions as part of the brain's pain system for processing cognitive-evaluative aspects of pain. If

## High-dimensional mediation analysis of pain

1 the neuronal pain system was mirroring the phenomenal complexity of the pain experience, we  
2 might expect regions processing cognitive aspects, and perhaps even other areas, e.g., those  
3 controlling attention, to be part of the broader pain system (Melzack, 1999). In this case, the  
4 dlPFC, vmPFC, NAc, and potentially parietal regions, should be true mediators of the pain  
5 response, i.e., they should be closely associated with nociceptive input and pain experience. For  
6 example, pain-evoked activity in parietal regions could link attention to the sources of pain with  
7 motor intentions (Downar et al., 2003; Oshiro et al., 2007). Furthermore, the different pain  
8 mediators should be separable into multiple different functional networks, each associated with  
9 different aspects of pain processing.

10 Here, we introduce a new multivariate mediation analytic framework that captures two  
11 important advantages in a single model. First, by analyzing spatial patterns of brain activity, our  
12 method builds on spatially distributed information across multiple spatial scales. Second, our  
13 method allows the identification of brain responses jointly linked, and interposed between,  
14 nociceptive input and pain reports. Mediation analysis (Figure 1A) has previously been applied  
15 on a voxel-by-voxel basis to investigate relationships between stimulation intensity, voxel-wise  
16 brain activation, and pain report (Figure 1B) (Atlas et al., 2014). However, as with other work on  
17 multivariate pattern classification and regression (Wager et al., 2013; Haynes, 2015), a univariate  
18 approach can miss brain regions whose contributions to pain perception are conditional on other  
19 regions. In order to capture cross-regional interactions, we use a unified high-dimensional  
20 approach that takes into account spatial co-variation of activity patterns across the brain (Figure  
21 1C,D).

22 This new approach, high-dimensional mediation analysis, identifies multiple whole brain  
23 mediators, termed *principal directions of mediation* (PDM). Each PDM represents a pattern of  
24 whole brain activity chosen because it maximizes the indirect (mediating) effect between  
25 stimulus intensity and pain report. The voxel weights of each PDM inform us about the  
26 contribution of individual brain regions to the generation of a painful experience following noxious  
27 stimulation. This approach decomposes activity across the brain into multiple networks that  
28 independently mediate stimulation effects on outcomes (i.e., pain report). Furthermore, these  
29 independent PDMs can be combined into a single, joint PDM that can be prospectively applied  
30 to new datasets as a predictive model.

31 Using data from eight different heat pain studies (N=284), we comprehensively investigate  
32 the role of brain mediators in the generation of pain experiences. Seven of the eight studies,  
33 were used as training data for the mediation analyses (N=209), and the largest individual study



## High-dimensional mediation analysis of pain

1 (N=75) was used as a test data set, using the model parameters estimated in the training data  
2 to validate model predictions in new individuals. Importantly, the test data set not only included  
3 heat pain stimuli, but also physically and emotionally aversive sounds. While brain mediators of  
4 pain should generalize to different pain data sets, they are not expected to mediate the  
5 relationship between sound stimulation levels and perceived sound intensity. This allows us to  
6 study the sensitivity and specificity of the brain mediators of pain.

7

## 8 **Results**

### 9 *Principal Directions of Mediation (PDM)*

10 Participants in all eight studies underwent fMRI scanning while being exposed to varying  
11 levels of heat pain and rating the perceived pain intensity (see Tables 1-3 for details on each  
12 study). For each participant, we recorded the temperature applied, the pain rating on a 0-100  
13 scale for each trial and estimated single-trial maps of brain activity. These three variables were  
14 used in the primary mediation analysis with temperature as the initial variable, brain activity as  
15 the mediator, and pain rating as the outcome variable (Figure 1). Using our novel high-  
16 dimensional mediation analysis model (see Chén et al., 2017 for a similar approach), we first  
17 estimated 30 whole-brain mediation patterns (PDMs). Each PDM specifies a linear combination  
18 of voxels across the brain maximizing the mediated effect from temperature to pain rating, while  
19 being orthogonal to other PDMs (Figure 1C). Each PDM (or  $w_p$ ) thus represents a whole brain  
20 mediator for pain. For each individual PDM, we obtain path coefficients for the relationship  
21 between temperature  $X$ , brain mediator  $\tilde{m}$ , and pain rating  $Y$  as in a standard mediation model.  
22 A positive path  $a$  indicates that higher temperatures lead to more activity in voxels with positive  
23 PDM weights and less activity in voxels with negative PDM weights. A positive path  $b$  indicates  
24 that voxels with positive weights contribute positively to the pain rating after controlling for  
25 temperature. This pattern of weights would be expected for regions that receive spinothalamic  
26 input, for example the posterior insula or S2 (Willis and Westlund, 1997; Dum et al., 2009), and  
27 possibly other mediating regions as well. Finally, we combine the individual brain mediator maps  
28 into a joint PDM by computing the weighted sum of the individual PDMs (Figure 1D).

29 The absolute coefficient values for the indirect  $ab$  path assess how much of the effect of  
30 the manipulated temperature on pain ratings is explained by the brain mediator, i.e., individual  
31 PDM pattern. Here, the first 10 PDMs accounted for 99.1% of the total mediation effect (Figure

## High-dimensional mediation analysis of pain

1 1E, Figure 1-supplement 1). We thus focus on the first 10 PDMs in all subsequent analyses with  
2 minimal loss of information. In order to analyze the contribution of individual brain regions to the  
3 mediation of pain, the signs of both paths  $a$  and  $b$  and the sign of the voxel weights have to be  
4 considered: Voxel weights are multiplied by the respective path coefficients to determine a  
5 region's relationship to stimulation intensity and pain rating. When considering the sign of the  
6 voxels weights, four different kinds of relationship are possible: (i) positive to temperature,  
7 positive to pain; (ii) negative to temperature, negative to pain; (iii) positive to temperature,  
8 negative to pain; and (iv) negative to temperature, positive to pain. Here, type (i) is the standard,  
9 positive mediator case and type (ii) represents a negative mediator, in which greater deactivation  
10 to the stimulus mediates increased pain (MacKinnon et al., 2000; Atlas et al., 2010). Types (iii)  
11 and (iv) are suppressor effects (MacKinnon et al., 2000), e.g., for type (iii), brain activity increases  
12 with stimulus intensity that suppress pain, and may thus be involved in stimulus-engaged  
13 regulatory processes and other negative feedback loops.

14 PDM 1 has both positive path  $a$  and  $b$  coefficients. Brain regions with positive weights  
15 (representing positive mediators, type (i) with positive paths  $a$  and  $b$ ) are shown in warm colors  
16 in Figure 2. These include brain regions commonly associated with pain processing, such as the  
17 dorsal posterior and mid-insula, S1, S2, MCC, and the PAG (Figure 2). Significant voxels in MCC  
18 stretch into the supplementary motor area (SMA), dorsal of the cingulate sulcus. In addition, PDM  
19 1 contains negative, type (ii), mediators, including the medial prefrontal cortex (mPFC) and  
20 ipsilateral S1/M1. The negative weights indicate that these regions show less activation with  
21 increasing temperatures and less regional activation is related to higher pain ratings. Such  
22 relationships would be expected for brain regions whose function is inhibited by nociceptive  
23 input or that are deactivated with increased pain-related processing.

24 Brain regions positively mediating the relationship between temperature and pain rating  
25 (type (i)) in other PDMs are S1, M1, superior frontal gyrus (SFG), fronto-temporal operculum,  
26 temporal poles, temporal operculum, ventral insula, pons, and cerebellum (Figure 2). These  
27 positive mediators include regions, like the temporal regions, that are traditionally not considered  
28 to be pain-processing regions. Brain regions acting as negative mediators (type (ii)) in other  
29 PDMs include medial orbitofrontal cortex (mOFC), dorso-medial prefrontal cortex (dmPFC),  
30 superior parietal lobule (SPL), retrosplenial cortex (RSC), precuneus, and cuneus.

31 A more complex function is indicated by positive path  $a$  coefficients, but negative path  $b$   
32 coefficients (types (iii) and (iv), PDMs 3,5,7, and 9). Here, regions with positive voxel weights  
33 show a positive relationship with temperature, i.e., higher temperatures lead to more activity. By

## High-dimensional mediation analysis of pain

1 contrast, the negative path  $b$  indicates that these regions are negatively related to pain ratings  
2 controlling for temperature, i.e., more activity is related to lower pain ratings. Regions with such  
3 a profile fit a pain-inhibitory role as their activity increases with rising stimulus temperatures, but  
4 their increased activity mediates lower pain ratings (type (iii)). Parts of the mOFC, the cerebellum,  
5 precuneus, S1, and the left dlPFC fit this pain inhibitory profile.

6 A final set of regions shows a negative relationship with temperature (positive path  $a$ , but  
7 negative weights) and a positive relationship with pain ratings, controlling for temperature  
8 (negative voxel weights and negative path  $b$  resulting in a net positive relationship; type (iv)).  
9 Such regions show stimulus intensity-dependent deactivation, with larger de-activation  
10 mediating decreased pain, consistent with regulatory negative feedback mechanisms. Regions  
11 with this profile include parts of the mOFC, the parahippocampal gyrus, visual cortices, and the  
12 NAc. For example, NAc shows decreased activation for high temperatures, which may relate to  
13 punishment or negative reinforcement signals. At the same time, controlling for temperature,  
14 stronger NAc de-activation is related to lower pain ratings, potentially signaling reduced  
15 motivational relevance.

### 16 *Joint PDM*

17 The individual PDMs can be combined into a single, joint PDM since the individual PDMs  
18 are orthogonal to each other. Weighting each individual PDM by its indirect effect (path  $ab$ ) and  
19 summing the weighted PDMs results in a joint PDM map representing the total contribution of  
20 each voxel to the total indirect (pain mediation) effect (see Figure 1D and Methods). Significant  
21 voxel weights of the joint PDM map were determined by an additional bootstrap procedure at a  
22 false discovery rate (FDR) of  $q < 0.05$ .

23 Within the joint PDM, individually significant clusters of positive mediators included S2,  
24 MCC, SMA, PAG, insula, including anterior and dorsal-posterior parts, as well as the medial  
25 thalamus (Figure 2). Negative mediators (stimulus-induced deactivations mediating increased  
26 pain) included mPFC, SPL, S1, and M1. Many of these regions were also part of PDM 1, which  
27 accounted for the biggest proportion of the total mediated effect. However, the medial thalamus  
28 and SPL were significant in the joint PDM, but not in PDM 1.

29 While the size of the S1/M1 cluster was smaller in the joint PDM compared to PDM 1, the  
30 size of the mPFC cluster increased. Voxel weights for the mPFC and S1 were all negative in the  
31 joint PDM. The negative weights in the joint PDM indicate that the net contribution of these  
32 regions is a negative mediation of the relationship between temperature and pain, although these

## High-dimensional mediation analysis of pain

1 regions received positive weights in some of the individual PDMs. Indeed, it is possible for  
2 weights to be both positive and negative in different PDMs, because voxels may include neural  
3 ensembles participating in different distributed circuits related to either more or less pain. Thus,  
4 the individual PDMs represent a decomposition of voxels' activity into different distributed  
5 components, while the joint PDM reflects each voxel's net contribution (controlling for other  
6 voxels). Computing and analyzing the joint PDM can thus help to clarify overall relationships  
7 between regional activity and the predictor and outcome variables.

### 8 *Clustering PDMs into functional networks*

9 The PDMs provide a dimensional view of coherent distributed processes, with each PDM  
10 a distinct dimension; in addition, it is useful to cluster the regions with the highest dimensional  
11 weights, to further examine the network structure of the inter-regional relationships. To do this,  
12 we used an iterative clustering procedure to group regions based on inter-regional correlations  
13 in stimulus-evoked responses across trials without considering stimulation temperatures or pain  
14 ratings (Kober et al., 2008; Atlas et al., 2014). The cluster analysis of single-trial activity from  
15 significant voxels of all 10 PDMs revealed 26 functional regions organized into 5 different  
16 functional networks (Figure 3A,C). A functional description of these networks was determined by  
17 computing the similarity of each network with feature maps generated by the meta-analytic tools  
18 on neurosynth.org (Yarkoni et al., 2011). The top ten features for each network are shown in  
19 Table 4. Network names were chosen based on the functional associations with Neurosynth  
20 (Yarkoni et al., 2011) terms. For example, the top three feature associations for network 1 were  
21 somatosensory, motor, and stimulation. Based on these associations we labeled network 1 as  
22 'sensorimotor network'.

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

## High-dimensional mediation analysis of pain

1           We next investigated with which functional networks the individual PDMs are associated  
2 by computing pairwise Dice similarity coefficients (Figure 3B). Interestingly, the joint PDM was  
3 almost equally similar to the visual ( $D = 0.3$ ), sensorimotor ( $D = 0.27$ ), and to the executive  
4 function ( $D = 0.25$ ) networks, again stressing the diversity of brain regions contributing to pain.  
5 By contrast, PDM 1 had the greatest overall similarity with any single network, namely with the  
6 sensorimotor network ( $D = 0.7$ ). No other network was substantially associated to PDM 1 (all  
7  $D < 0.05$ ). The sensorimotor network was also associated with PDMs 2, 5, and 6. The value  
8 learning network was related to PDMs 3, 6, and 9, with the highest similarity to PDM 9 ( $D = 0.16$ ).  
9 Similarity between the default mode network and PDM 4 was highest ( $D = 0.24$ ). Parts of the  
10 DMN also overlapped with PDMs 3, 5, and 7. The executive function network was associated  
11 with PDM 2 ( $D = 0.22$ ) and PDM 3 ( $D = 0.23$ ), and, to a lesser degree, with PDMs 4 and 7. Finally,  
12 the visual network was related to PDM 3 ( $D = 0.38$ ) and to a lesser degree to PDMs 5, 7, and 9.  
13 The overall similarity pattern between functional networks and PDMs shows that in contrast to  
14 PDM1, few of the remaining joint and individual PDMs are dominated by a single network. More  
15 often PDMs were comprised of a mix of 2 or 3 networks that together act as a pain mediator,  
16 reflecting the complexity of the transformation from nociception into pain experience.

17           Projecting functional networks and regions onto the first 2 dimensions of the underlying  
18 non-metric multidimensional scaling (NMDS) space revealed that the sensorimotor network had  
19 high loadings on dimension 1, in contrast to the DMN, which had low loadings on dimension 1  
20 (Figure 3C). Dimension 1 thus spanned from the DMN to the sensorimotor network with the value  
21 learning and visual networks located between the two. Dimension 1 could thus be described  
22 approximately as an activation-deactivation gradient during pain. Interestingly, the PAG loaded  
23 relatively low on dimension 1 within the sensorimotor network and was located closest to the  
24 value learning network (Roy et al., 2014). Given the pain-modulatory role of the PAG this could  
25 indicate a flexible behavior in pain processing that differs from other regions like S1, S2, or insula,  
26 in line with previous literature (Satpute et al., 2013; Roy et al., 2014).

27           The value learning network loaded lowest on dimension 2. Default mode and sensorimotor  
28 networks scored higher than the value learning network, but still lower than the visual network.  
29 The executive function network occupied much of the top-left quadrant, with low to medium  
30 loadings on dimension 1 and medium to high loadings on dimension 2. With precuneus and LOC  
31 as the highest loading regions on dimension 2 and NAc, anterior insula, and mOFC loading  
32 lowest, this dimension can potentially be described as a gradient across processing of different  
33 time-scales, with NAc encoding transient surprises and precuneus integrating semantic

## High-dimensional mediation analysis of pain

1 information across longer time-scales (Hasson et al., 2015).

### 2 *Validation on an independent cohort*

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

9 Applying the PDMs to independent pain test data (N = 75, an independent community  
10 sample cohort of mixed races and sex), revealed significant paths  $a$  and  $b$  for all 10 PDMs and  
11 the joint PDM (Figure 4A). The indirect path was also significant for the joint PDM and all 10  
12 individual PDMs, suggesting that all 10 PDMs are reliably related to pain and generalize across  
13 cohorts. The magnitude of the indirect effects (path  $ab$ ) are monotonically decreasing for the  
14 training data (Figure 1E). On the test data, indirect path coefficients were not strictly  
15 monotonically decreasing from PDM 1 to PDM 10 (Figure 4A, Figure 4-supplement 1), indicating  
16 some variability of the PDM order across data sets, as expected. The joint PDM and the first two  
17 individual PDMs had the strongest effect in both data sets, suggesting that they capture the  
18 most important brain activity for pain across data sets. Figure 4C shows the predicted pain from  
19 the joint PDM plotted against the empirical pain ratings for pain training and test data.

20 In order to further corroborate the generalizability and robustness of the PDMs, we also  
21 estimated 10 PDMs on the original test data set (Study 8) and cross-validated the new PDMs on  
22 the original training data set (Studies 1-7). The results were similar to the main results presented  
23 here. Six out of ten indirect paths were significant when PDM estimation was done on the smaller  
24 sample. The indirect  $ab$  path coefficients for the first four PDMs were highest when applying the  
25 new PDMs to the original training data (Figure 4-supplement 2). Generalization thus does not  
26 depend strongly on the choice of the training data.

27 In order to test whether PDMs are mediators specifically for somatic pain, we also applied  
28 the original PDMs to other aversive stimuli that are not painful. The test data of Study 8 also  
29 included trials with physically (fingernails on chalkboard) and emotionally (screaming, crying,  
30 etc.) aversive sounds with three pre-defined intensity levels of each stimulus type. Study 8 was  
31 designed to test specificity vs. generalizability to aversive sounds and matched in duration and  
32 approximate aversiveness ratings based on pilot studies; trials were randomly intermixed with



## High-dimensional mediation analysis of pain

1 heat pain trials. Application of the original PDMs on the sound data revealed no significant  
2 indirect effects (Figure 4B, Figure 4-supplement 3) and only nine significant paths  $a$  or  $b$  in total.  
3 Thus, pain PDMs do not mediate the relationship between sound intensity and intensity ratings  
4 for either type of sound. These results indicate specificity to somatic pain vs. sound.

### 5 *Comparison to the Neurological Pain Signature (NPS)*

6 Previous studies have investigated the direct relationship between brain responses and  
7 pain reports, both using univariate (Coghill et al., 1999; Bornhövd et al., 2002; Ploner et al., 2010;  
8 Atlas et al., 2014) and multivariate approaches (Marquand et al., 2010; Brodersen et al., 2012;  
9 Schulz et al., 2012; Wager et al., 2013; Woo et al., 2017). One study trained a multivariate pattern,  
10 termed the Neurological Pain Signature (NPS), that predicts pain reports with high accuracy from  
11 brain activity that can be easily applied to new data sets (Wager et al., 2013; Krishnan et al.,  
12 2016; Geuter et al., 2017a). In contrast to the present approach, the estimation of the NPS did  
13 not account for temperature-brain relationships; its goal was to predict pain intensity without  
14 demonstrating mediation (Wager et al., 2013). We compared our mediation approach to the  
15 predictive power of the NPS by computing the variance explained in single-trial pain ratings by  
16 both models. The joint PDM explained a total of 10.5% of the single-trial rating variance in the  
17 training data (Studies 1-7) while the NPS explained a total of 4% of the rating variance within  
18 subjects (Figure 4D). The variance uniquely explained by the joint PDM was 7%, while the NPS  
19 only explained 0.5% unique rating variance and 3.5% of the rating variance was jointly explained  
20 by the joint PDM and the NPS. On the test data set (Study 8), the joint PDM explained a unique  
21 share of 6.1% of the rating variance. NPS and PDMs explained an additional 3.7% of variance  
22 together and the NPS explained additional 0.9% alone (Figure 4D). Together, this indicates that  
23 including temperature-brain relationships in the PDM approach captures additional pain variance  
24 not explained by the NPS. Here one should note that single-trial data are extremely noisy (Woo  
25 et al., 2017), but these numbers indicate high accuracy if an application can average across trials  
26 within a person as shown in previous studies results (Wager et al., 2013; Woo et al., 2017).

### 27 *Comparison to univariate mediation analysis*

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

## High-dimensional mediation analysis of pain

1 the training data set (Studies 1-7).

2 This analysis identified the MCC, cerebellum, posterior and mid-insula, S2, and S1 as brain  
3 mediators defined as the intersection of the coefficient maps for paths  $a$ ,  $b$ , and  $ab$  at FDR  $q <$   
4  $0.05$  (Figure 5). Comparing these results to the joint PDM, which estimates a joint mediating  
5 pattern across voxels, revealed both similarities and some notable differences (Figure 5). The  
6 joint PDM included additional regions not in the univariate model, including mPFC, PAG, SPL,  
7 and S1. By contrast, the univariate mediation results included a part of the cerebellum that was  
8 not included in the joint PDM. Overall, the high-dimensional approach identified more regions  
9 than the univariate approach, including regions outside the classic pain processing network like  
10 mPFC and SPL. Furthermore, the PAG, a region known to be involved in descending pain  
11 control, is part of the joint PDM, but not part of the univariate mediators. Such results are  
12 expected if some brain regions make detectable contributions only after controlling for the  
13 influences of other brain regions; this is an advantage of multivariate predictive approaches to  
14 neuroimaging analysis and multiple regression generally.

15 Computing the cosine similarities of PDMs and NPS to both univariate path  $a$  and  $b$  maps  
16 revealed an interesting pattern (Figure 6). Path  $a$  represents the relationship between  
17 temperature and brain responses, while path  $b$  represents the relationship between brain  
18 responses and rating, controlling for temperature. Projecting all maps on the space defined by  
19 temperature and pain rating related brain responses, revealed a linear ordering of components  
20 along these dimensions. The map representing the univariate mediation effect (path  $ab$ ) was  
21 most similar to path  $b$  (purple dot in Figure 6). The joint PDM (red) was most similar to the  
22 univariate path  $a$  and located in close proximity to the univariate mediation effect map with  
23 respect to the path  $a$  and  $b$  maps. However, the cosine similarity of the joint PDM with the  
24 univariate path  $ab$  map was only 0.45, indicating that the two maps reflect substantially different  
25 brain processes. Similarities between individual PDMs and the two univariate maps were ordered  
26 according to their order of estimation (and the variance explained in the training dataset), with  
27 the exception of PDM 3, which was negatively related to both maps. The NPS (black dot) was  
28 positioned between PDMs 2 and 4. However, the low overall similarity values for PDMs 3, 5-10,  
29 suggest that the space defined by univariate maps of temperature and pain rating related brain  
30 responses do not capture all components involved in pain processing. A higher dimensional  
31 representation may be more consistent with psychological theories of pain experience.

32



## 1 **Discussion**

2       Using a novel high-dimensional mediation analysis approach (*Principal Directions of*  
3 *Mediation* [PDM]), we identified brain networks that mediate the relationship between stimulus  
4 intensity and pain reports. Importantly, the PDM mediators generalized to independent pain test  
5 data but not to aversive sound data, suggesting at least some specificity for pain. A parcellation  
6 of the brain mediators into functional networks revealed distinct contributions of classic  
7 somatosensory brain regions, but also motor regions, value learning, executive control, default  
8 mode, and visual regions. This diversity of mediators shows that pain involves many brain  
9 regions in addition to somatosensory regions. The observation that the joint PDM map, which  
10 integrates the brain mediators, is equally related to the executive function, visual, and  
11 sensorimotor networks further supports the importance of non-somatosensory regions in pain  
12 processing. In this way, the diversity of brain mediators mirrors the multi-dimensional nature of  
13 pain including of sensory, affective, motivational, and cognitive components (Melzack, 1999;  
14 Turk and Melzack, 2011).

15       The new, high-dimensional mediation approach provides a more comprehensive picture of  
16 pain processing in the human brain than previous studies using univariate analyses, or studies  
17 focusing solely on the stimulation-brain or brain-outcome relationships. This is reflected by the  
18 higher share of pain rating variance explained compared to the NPS (Wager et al., 2013), by the  
19 higher sensitivity at the brain-voxel level compared to univariate mediation, and by the  
20 involvement of brain regions not observed in recent meta-analyses of pain (e.g., mPFC, PAG,  
21 and M1) (Duerden and Albanese, 2013; Jensen et al., 2016). The higher sensitivity is  
22 demonstrated by the direct comparison of univariate and multivariate maps (Figure 5), and by  
23 the PDMs not contained in the space defined by the univariate maps (Figure 6). Together, these  
24 results highlight the importance of a broad and methodologically advanced approach to studying  
25 pain and related affective processes.

26       Our results parallel observations in animals and humans that have stressed the importance  
27 of psychological and neural processes underlying motivation, learning, attention, and cognition  
28 for pain (Melzack, 1999; Atlas et al., 2014; Navratilova and Porreca, 2014; Kucyi and Davis, 2015;  
29 Wiech, 2016; Seminowicz and Moayedi, 2017). Functional and structural changes in regions  
30 strongly involved in learning, valuation, and executive functions occur during the development  
31 of chronic pain and also contribute to chronic pain (Bushnell et al., 2013; Seminowicz and  
32 Moayedi, 2017). For example, structural changes observed in the NAc, insula, dlPFC, and  
33 sensorimotor cortex distinguish healthy individuals and those suffering from chronic pain (Baliki

## High-dimensional mediation analysis of pain

1 et al., 2012; Chang et al., 2014; Schwartz et al., 2014; Seminowicz and Moayedi, 2017).  
2 Furthermore, altered communication between mPFC and NAc contributes to the development  
3 of chronic pain and regulation of acute pain (Baliki et al., 2012; Lee et al., 2015; Woo et al., 2015).  
4 Studies on large-scale functional brain connectivity have also shown that the brain  
5 switches dynamically between different states as indexed by state-dependent changes in the  
6 communication patterns within and between different brain networks (Cribben et al., 2012;  
7 Hutchison et al., 2013). These spontaneous state changes influence perception and cognition  
8 (Boly et al., 2007; Sadaghiani et al., 2015) and are known to affect the perception of noxious  
9 stimuli (Ohara et al., 2008; Ploner et al., 2010). These observations have led to the hypothesis of  
10 a 'pain connectome' in which the functional connectivity between networks determines pain  
11 experiences (Kucyi and Davis, 2015). The DMN (including mPFC, precunues, and temporal  
12 regions), the salience network (including anterior insula, PFC, and TPJ), and the anti-nociceptive  
13 network (including mPFC and PAG) have been proposed to be particularly important for pain  
14 perception (Kucyi and Davis, 2015). All of these regions were also part of the functional networks  
15 mediating pain processing in our multivariate analyses (Figure 3), supporting the notion that pain  
16 depends on activation and co-activation patterns (or functional connectivity) between all these  
17 regions and not just on the activation level in a unitary core pain system. The high-dimensional  
18 mediation approach further allows us to analyze the relationships between activity in individual  
19 brain regions with stimulus intensity and pain experience in more detail. In the following we will  
20 discuss contributions of brain regions based on their functional relationships with stimulation  
21 intensity and pain reports.

22 Activity in brain regions receiving afferent nociceptive input, including the medial thalamus,  
23 PAG, S2, insula, MCC, SMA, and ipsilateral S1 (Dum et al., 2009), increased due to increasing  
24 temperatures and higher activity was related to stronger pain, controlling for temperature. This  
25 set of commonly pain-associated regions (Apkarian et al., 2005; Bushnell et al., 2013; Duerden  
26 and Albanese, 2013; Jensen et al., 2016) was complemented by anterior temporal regions and  
27 the cerebellum, which share the same functional response profile. A positive relationship with  
28 both temperature and pain rating is in line with a traditional, feedforward encoding view of  
29 nociception (Bushnell et al., 2013; Atlas et al., 2014; Geuter et al., 2017a). Because the mediation  
30 analysis statistically controls the effects of temperature, our results show that fluctuations in  
31 regional activity also contribute to pain perception beyond the regional activity driven by direct  
32 afferent input. Activity in these positive mediator regions is thus not only determined by  
33 nociceptive input, but the processing and transformation of nociceptive input in these regions

## High-dimensional mediation analysis of pain

1 contributes to the perceived pain (Büchel et al., 2002).

2 By contrast, the mPFC, SPL, RSC, precuneus, and contralateral S1 and M1 were negatively  
3 related to both temperature and pain. The mPFC, RSC, and precuneus are part of the DMN,  
4 which has been associated with mind-wandering and internal thoughts (Andrews-Hanna et al.,  
5 2010; Kucyi and Davis, 2015). The negative mediating role of the DMN regions could be related  
6 to the disruption of ongoing thought processes by the painful stimulation or attentional  
7 refocusing from internal to external sensations. Similarly to the DMN response profile, activity in  
8 contralateral M1 was negatively related to stimulus intensity and pain. Motor cortex activity has  
9 been associated with painful stimuli in some neuroimaging studies (Apkarian et al., 2005;  
10 Schweinhardt and Bushnell, 2010). Along with premotor areas such as SMA, activation in M1 is  
11 sometimes interpreted in terms of motor function. However, if M1 activity would represent a  
12 motor planning response, we would expect a positive relationship with stimulus intensity and  
13 pain ratings. By contrast, the negative relationship of the contralateral M1 with pain is in line with  
14 reports of reductions in clinical pain following the inhibition of M1 by transcranial magnetic  
15 stimulation (TMS) of M1 (Passard et al., 2007; Mori et al., 2010; Moisset et al., 2016) suggesting  
16 a pain modulatory role of M1, potentially via the PAG and the ACC (Pagano et al., 2011).  
17 However, further studies are needed to test a potential causal pain inhibitory function of these  
18 negative mediator regions in the DMN and sensorimotor cortices.

19 Pain has also strong motivational implications – humans and animals avoid pain when  
20 possible because pain is usually associated with tissue damage (Navratilova and Porreca, 2014;  
21 Geuter et al., 2016). It is thus important to learn which stimuli cause pain in order to minimize  
22 future harm. However, the role of value learning regions like NAc in pain are complex and not  
23 well understood, yet (Becerra et al., 2013; Woo et al., 2015). Unraveling them could contribute a  
24 great deal to pain characterization and treatment because of its prominent role in persistent pain  
25 in animal models (Chang et al., 2014; Navratilova and Porreca, 2014; Schwartz et al., 2014; Ren  
26 et al., 2016) and humans (Baliki et al., 2010, 2012). The present study offers some constraints on  
27 interpreting NAc function in pain by demonstrating opposing relationships of NAc activity with  
28 stimulus intensity (negative) and pain (positive). Here, NAc shows stimulus intensity-dependent  
29 deactivation, with larger de-activation mediating decreased pain, consistent with regulatory  
30 negative feedback mechanisms. The NAc might exert its control in this feedback loop indirectly  
31 via its connections with the hypothalamus or mPFC as indicated by studies in humans and  
32 animals (Baliki et al., 2012; Schwartz et al., 2014; Lee et al., 2015; Woo et al., 2015). However,  
33 the exact contribution of the NAc to pain perception might rely on more complex temporal

## High-dimensional mediation analysis of pain

1 dynamics that cannot be resolved in the current data set. For example, the direction of the  
2 valence encoding at pain onset and offset is still a matter of debate (Baliki et al., 2010; Becerra  
3 et al., 2013) as is its role in aversive learning more generally (Roy et al., 2014; Matsumoto et al.,  
4 2016). Elucidating the specific contributions of the NAc in different contexts in future studies will  
5 further help our understanding of motivational and learning aspects for pain perception.

6 An advantage of the present multivariate mediation approach is that it controls for the  
7 effects of stimulation intensity when estimating the relationship between brain activity and  
8 reported outcomes. Compared to approaches that do not take into account the stimulus-brain  
9 relationship when predicting pain (Wager et al., 2013; Krishnan et al., 2016; Lindquist et al.,  
10 2017), the present mediation approach yields higher predictive accuracy. Both approaches may  
11 yield whole brain maps that can be used as predictive models of acute pain that can be applied  
12 prospectively to new data. Application of the PDMs to new datasets can be used to (i) further  
13 evaluate the sensitivity and specificity of the model, and (ii) to evaluate the effects of  
14 psychological or medical interventions on the brain processes supporting pain. Testing the  
15 PDMs on a large, independent data set showed that the PDMs generalize to other pain data, but  
16 do not generalize to aversive sounds. This speaks against the notion that the brain mediators  
17 are completely driven by stimulus independent features, such as general feelings of aversiveness  
18 or unpleasantness.

19 In summary, the new high-dimensional mediation analysis revealed a comprehensive  
20 picture of brain responses underlying the complex, multi-faceted pain experience. Several brain  
21 regions, such as the mPFC, NAc, and M1, are shown to directly and formally mediate stimulus-  
22 to-pain relationships. The functional diversity of the brain mediators observed here offers a better  
23 understanding of the brain responses underlying the complexity of the pain experience.

24

### 25 **Acknowledgements**

26 We are grateful for the funding support of the DFG (GE 2774/1-1 to S.G.) and the NIH,  
27 which supported this work under grants R01DA035484 (T.D.W.), 2R01MH076136 (T.D.W.),  
28 R01DA027794 (T.D.W.), R01 EB016061 (M.A.L.) and P41 EB015909 (M.A.L.).

29

### 30 **Author Contributions**

31 S.G., T.D.W., and M.A.L. designed the study. M.A.L. contributed unpublished analytical

## High-dimensional mediation analysis of pain

- 1 tools. S.G. conducted data analysis and drafted the manuscript. S.G., T.D.W., M.A.L., and
- 2 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
- 3 neuroimaging data and provided comments on the manuscript.

## 1 **Materials and Methods**

### 2 *Participants*

3       The analysis included a total of 284 healthy participants from 8 independent studies, with  
4 sample sizes ranging from N = 17 to N = 75 per study. Descriptive statistics on the age, sex, and  
5 other features of the subjects in each individual study are provided in Tables 1-3. Further details  
6 on Studies 1-7, which were used to estimate the PDMs are provided in Lindquist et al. (2017).  
7 Participants were recruited from New York City and Boulder/Denver Metro Areas. The  
8 institutional review board of Columbia University and the University of Colorado Boulder  
9 approved all the studies, and all participants provided written informed consent. Preliminary  
10 eligibility of participants was determined through an online questionnaire, a pain safety screening  
11 form, and a functional Magnetic Resonance Imaging (fMRI) safety screening form.

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

### 18 *Procedures*

19       In all studies, participants received a series of contact-heat stimuli and rated their  
20 experienced pain following or during each stimulus. The number of trials, stimulation sites, inter-  
21 trial intervals, rating scales, and stimulus intensities and durations varied across studies, but  
22 were comparable; these variables are summarized in Tables 2 and 3. Each study also comprised  
23 a specific psychological manipulation (except Study 8), such as placebo treatment, which will be  
24 or has been reported elsewhere (Table 1). Study 8, which was used for validation purposes (see  
25 below), also presented aversive sounds to participants. Trials with aversive sounds were used  
26 to test the specificity of the pain PDMs. Sounds included were a physically aversive recording of  
27 nails on a chalkboard and a set of emotionally aversive sounds (attacks, screaming, and crying)  
28 from the International Affective Digital Sounds database (IADS) (Bradley and Lang, 2007). Aside  
29 from these sound trials, we focus on brain mediation of pain across all trials in the present paper,  
30 irrespective of the study-specific psychological and physical manipulations that influenced pain.

## 1 *Thermal stimulation*

2 In each study, except Studies 7 and 8, thermal stimulation was delivered to multiple skin  
3 sites using a TSA-II Neurosensory Analyzer (Medoc Ltd., Chapel Hill, NC) with a 16 mm Peltier  
4 thermode endplate. A PATHWAY system (Medoc Ltd., Chapel Hill, NC) was used in Studies 7  
5 and 8. Study 7 used a circular CHEPS Peltier endplate (diameter: 32 mm) and study 8 used a 16  
6 mm ATS Peltier endplate. On every trial, after the offset of stimulation, participants rated the  
7 magnitude of the warmth or pain they had felt during the trial on a visual analog scale.  
8 Participants in Study 8 rated their pain continuously during stimulation. The maximum rating of  
9 each trial was used in the following analyses. Other thermal stimulation parameters varied across  
10 studies, with stimulation temperatures ranging from 40.8 °C to 50 °C and stimulation durations  
11 from 1.85 to 12.5 s. Most studies applied thermal stimulation to the forearm. See Table 2 for  
12 stimulation intensity levels, mean temperature for each intensity level, and details of the rating  
13 scales. See Table 3 for stimulation duration, duration of inter-stimulus interval, number and  
14 location of stimulation sites, and number of trials per subject.

## 15 *fMRI data processing*

### 16 **Preprocessing**

17 Structural T1-weighted images were co-registered to the mean functional image for each  
18 subject using the iterative mutual information-based algorithm implemented in SPM (Ashburner  
19 and Friston, 2005), and then normalized to MNI space using SPM. The version of SPM used  
20 varied across studies (Studies 1 and 6 used SPM5; while all other studies used SPM8;  
21 <http://www.fil.ion.ucl.ac.uk/spm/>). Following normalization, Studies 1 and 6 included an  
22 additional step of normalization to the group mean using a genetic algorithm-based  
23 normalization (Wager and Nichols, 2003; Atlas et al., 2010, 2014).

24 For each functional dataset, initial volumes were removed to allow for image intensity  
25 stabilization (see Lindquist et al. (2017) for details). In addition, volumes with signal values that  
26 were outliers within the time series (i.e., “spikes”) were removed. To identify outliers, both the  
27 mean and the standard deviation of intensity values across each slice were computed for each  
28 image. The Mahalanobis distances for the matrix of (concatenated) slice-wise mean and  
29 standard deviation values by functional volumes (over time) were computed, and values with a  
30 significant  $\chi^2$  value (corrected for multiple comparisons based false discovery rate) were  
31 considered outliers. In practice, less than 1% of images were deemed outliers. The output of this  
32 procedure was later included as nuisance covariates in the subject level models. Next, functional



## High-dimensional mediation analysis of pain

1 images were corrected for differences in the acquisition timing of each slice (except for multiband  
2 data with a short TR of 480 ms in Study 8) and were motion-corrected (realigned) using SPM.  
3 The functional images were warped to SPM's normative atlas (warping parameters estimated  
4 from co-registered, high-resolution structural images), interpolated to  $2 \times 2 \times 2 \text{ mm}^3$  voxels,  
5 and smoothed with an 8 mm FWHM Gaussian kernel.

### 6 **Single trial analysis (Except Study 3 and Study 6)**

7 For each study, a single trial, or “single-epoch”, design and analysis approach was used  
8 to model the data. Quantification of single-trial response magnitudes was done by constructing  
9 a GLM design matrix with separate regressors for each trial (Rissman et al., 2010; Mumford et  
10 al., 2012). First, boxcar regressors, convolved with the canonical hemodynamic response  
11 function (HRF), were constructed to model cue and rating periods in each study. Regressors for  
12 each trial, as well as several types of nuisance covariates were also included. Because each trial  
13 consisted of relatively few volumes, trial estimates could be strongly affected by acquisition  
14 artifacts that occur during that trial (e.g. sudden motion, scanner pulse artifacts, etc.). Therefore,  
15 trial-by-trial variance inflation factors (VIFs; a measure of design-induced uncertainty due, in this  
16 case, to collinearity with nuisance regressors) were calculated, and any trials with VIFs exceeding  
17 2.5 were excluded from the analyses (VIF threshold for Study 8 was 3.5 as in the primary  
18 publication). For Study 1, global outliers (trials that exceeded three standard deviations (SDs)  
19 above the mean) were also excluded, and a principal component based denoising step was  
20 employed during preprocessing to minimize artifacts. This generated single trial estimates that  
21 reflect the amplitude of the fitted HRF on each trial and refer to the magnitude pain-period activity  
22 for each trial in each voxel.

### 23 **Single trial analysis (Only Study 3 and Study 6)**

24 For Studies 3 and 6, single trial analyses were based on fitting a set of three basis functions,  
25 rather than the standard canonical HRF used in the other studies. This flexible strategy allowed  
26 the shape of the modeled hemodynamic response function (HRF) to vary across trials and voxels.  
27 This procedure differed from that used in other studies because it maintains consistency with  
28 the procedures used in the original publications. For both Study 3 and Study 6, the pain period  
29 basis set consisted of three curves shifted in time and was customized for thermal pain  
30 responses based on previous studies (Lindquist et al., 2009; Atlas et al., 2010). To estimate cue-  
31 evoked responses for Study 6, the pain anticipation period was modeled using a boxcar epoch  
32 convolved with a canonical HRF. This epoch was truncated at 8 s to ensure that fitted



## High-dimensional mediation analysis of pain

1 anticipatory responses were not affected by noxious stimulus-evoked activity. As in the other  
2 studies, nuisance covariates were included and trials with VIFs larger than 2.5 were excluded. In  
3 Study 6 trials that were global outliers (those that exceeded 3 SDs above the mean) were also  
4 excluded. The fitted basis functions from the flexible single trial approach were used to  
5 reconstruct the HRF and compute the area under the curve (AUC) for each trial and in each voxel.  
6 These trial-by-trial AUC values were used as estimates of trial-level pain-period activity.

### 7 **Data sets and PDM validation**

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

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

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

### 26 **Dimension reduction**

27 The training data set consisted of a total of 13,372 single-trial beta images, each consisting  
28 of 229,519 voxels, from 209 participants. To reduce the dimensionality of the data to a  
29 computationally tractable size, a generalized version of population value decomposition (PVD)  
30 (Caffo et al., 2010; Crainiceanu et al., 2011; Chén et al., 2017) was applied (using PVD.m as part  
31 of the M3 mediation toolbox available at <https://github.com/canlab/MediationToolbox>). This  
32 procedure is similar to singular value decomposition (SVD) but decomposes the data matrix into

## High-dimensional mediation analysis of pain

1 both participant specific and population specific components. We chose a dimensionality of  $p =$   
2 30 based on a tradeoff between variance explained and the number of trials available for each  
3 participant. The beta images were z-scored within each participant before PVD application. The  
4 reduced data matrix used for Principal Directions of Mediation (PDM) estimation consisted of a  
5 matrix with dimensions  $13,372 \times 30$ .

### 6 *Principal Directions of Mediation (PDM)*

7 Let  $X_i$  be the temperature,  $Y_i$  the reported pain, and  $\mathbf{M}_i = (m_i^{(1)}, m_i^{(2)}, \dots, m_i^{(p)})$  the brain  
8 activity over  $p$  voxels (i.e., the beta maps) measured between the application of the thermal  
9 stimuli and the pain report for observation (i.e., trial)  $i = 1, \dots, n$ . We are interested in determining  
10 how brain activation mediates the relationship between temperature and pain report, which is  
11 illustrated using the three-variable path model shown in Figure 1. We can estimate the  
12 parameters of this model using the following set of equations:

13

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

16

17 Once the parameters have been estimated we can express the total effect  $\gamma$  as the sum of the  
18 direct and indirect effects as follows:

19

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

21

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

## High-dimensional mediation analysis of pain

1 computing the joint PDM, and analyzing an almost 10-times larger data set.

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

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

$$\begin{aligned} 10 \quad \tilde{m}_j^{(k)} &= a_{0,k} + a_k X_j + \varepsilon_{jk} && \text{for } k = 1, \dots, q \\ 11 \quad Y_i &= b_{0,k} + c' X_i + b_k \tilde{m}_i^{(k)} + \eta_{ik} && (3) \end{aligned}$$

12

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

14

$$15 \quad c = c' + \sum_{k=1}^q a_k b_k \quad (4)$$

16

17 The difference between this model and the standard mediation model described in (1) is  
18 that the  $\mathbf{w}_k$  are unknown. In our approach  $\mathbf{w}_1$  is chosen so that it maximizes the amount of the  
19 indirect effect that is explained (i.e.,  $a_1 b_1$  is maximized). We refer to  $\mathbf{w}_1$  as the first *principal*  
20 *direction of mediation* (PDM). Note the first PDM corresponds to voxel-specific weights that can  
21 be mapped onto the brain, and thus provides interpretable maps of brain networks in the same  
22 manner as independent component analysis (ICA) and principal component analysis (PCA).  
23 Subsequent directions  $\mathbf{w}_k$ ,  $k = 1, \dots, q$ , can be found that maximize the remaining indirect effect  
24 conditional on being orthogonal to previous PDMs. As the transformed mediators are ranked  
25 based upon the proportion of the indirect effect explained, one could potentially limit the number  
26 of PDMs computed to achieve dimension reduction. Hence, our approach is philosophically  
27 similar to PCA, but addresses a fundamentally different problem.

28 The individual, orthogonal PDMs can be combined into a joint PDM by computing the  
29 following weighted sum:

30

$$31 \quad w_{joint} = \sum_{k=1}^q a_k b_k w_k \quad (5)$$

32

## High-dimensional mediation analysis of pain

1 According to the model formulation the signs of the PDMs are not identifiable, as any  
2 change in the sign of  $\tilde{m}_i^{(k)}$  can be offset by a change in sign of both  $a_k$  and  $b_k$ . We fix the signs  
3 of  $a_k$  to be positive for easier interpretation, i.e., positive voxel weights indicate higher brain  
4 activity for higher stimulus intensities. This is a similar constraint to the ICA approach often used  
5 in neuroimaging to detect networks. Note this does not impact the joint PDM as the sign of  
6  $a_k b_k$  is unchanged if both  $a_k$  and  $b_k$  change signs.

7 The problem of finding the  $k^{th}$  PDM involves finding the vector  $\mathbf{w}_k$  that maximizes  $a_k b_k$   
8 based on the constraint that  $\mathbf{w}_k^T \mathbf{w}_k = 1$  and  $\mathbf{w}_k^T \mathbf{w}_j = 0$  for all  $j = 1, \dots, k - 1$ . This problem can be  
9 solved using a nonlinear programming solver such as the interior-point algorithm. Inference is  
10 performed using a bootstrap procedure with 5,000 iterations, as described in Chén et al. (2017).  
11 We also test individual voxel weights for the joint PDM for significance using the bootstrap  
12 procedure above. All PDM maps are thresholded at a false discovery rate (FDR) of  $q < 0.05$ . We  
13 present results of 10 PDMs accounting for more than 99% of the total indirect effect (Figure 2).  
14 The PDM implementation is available at <https://github.com/canlab/MediationToolbox>  
15 (multivariateMediation.m).

16 In summary, we obtain scalar coefficients for paths  $a_k$ ,  $b_k$ , and  $c'_k$ , as well as the indirect  
17 effect  $ab_k$  for each PDM as in a standard, univariate mediation analysis. In addition, we obtain  
18 the voxel weight vector  $\mathbf{w}_k$  that maximizes the indirect effect  $ab_k$ .

### 19 *Cluster analysis*

20 The voxel weight maps for the mutually independent 10 PDMs span a high-dimensional  
21 space of brain mediators of pain perception. In order to reduce the dimensionality of that space  
22 and identify brain regions with similar activation profiles, we conducted a two-stage cluster  
23 analysis. The procedure is described in detail in Kober et al. (2008) and Atlas et al. (Atlas et al.,  
24 2014). Briefly, for significant voxels from the 10 PDMs we extracted single-trial activity estimates,  
25 resulting in a  $13,372 \text{ trials} \times 25,469 \text{ voxels}$  matrix. We then used singular value decomposition  
26 (SVD) to reduce the dimensionality of the voxel space. We kept 364 components that explained  
27 95% of the variance. Next, we clustered voxels into 250 parcels using hierarchical clustering.  
28 We then computed average single-trial activity within each parcel and used non-metric  
29 multidimensional scaling (NMDS) and hierarchical clustering to further reduce the dimensionality  
30 of the data. Inspection of the Shepard plot suggested a NMDS dimensionality of 15 with stress  
31 indices below 0.05. Stress indices ( $S$ ) are computed according to Shepard (1980) with

32

## High-dimensional mediation analysis of pain

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

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. 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. The position of the 5 networks and their constituent brain regions were projected on the first 2 dimensions of the NMDS space to visualize relationships and functional connectivity. 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.

### *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.

## 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, 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.
- Boly M, Balteau E, Schnakers C, Degueldre C, Moonen G, Luxen A, Phillips C, Peigneux P, Maquet P, Laureys S (2007) Baseline brain activity fluctuations predict somatosensory perception in humans. *PNAS* 104:12187–12192.
- 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.
- 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.

## High-dimensional mediation analysis of pain

- 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.
- 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* Available at: <https://academic.oup.com/biostatistics/article/doi/10.1093/biostatistics/kxx027/3868977/High-dimensional-multivariate-mediation-with> [Accessed August 8, 2017].
- 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.
- 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.
- Cribben I, Haraldsdottir R, Atlas LY, Wager TD, Lindquist MA (2012) Dynamic connectivity regression: Determining state-related changes in brain connectivity. *NeuroImage* 61:907–920.
- Downar J, Mikulis DJ, Davis KD (2003) Neural correlates of the prolonged salience of painful stimulation. *NeuroImage* 20:1540–1551.
- 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.
- Geuter S, Boll S, Eippert F, Büchel C (2017a) Functional dissociation of stimulus intensity encoding and predictive coding of pain in the insula. *eLife* 6:e24770.
- Geuter S, Cunningham JT, Wager TD (2016) Disentangling opposing effects of motivational states on pain perception: *PAIN Reports* 1:e574.
- Geuter S, Koban L, Wager TD (2017b) The Cognitive Neuroscience of Placebo Effects: Concepts, Predictions, and Physiology. *Annu Rev Neurosci* 40:167–188.
- Hasson U, Chen J, Honey CJ (2015) Hierarchical process memory: memory as an integral component of information processing. *Trends in Cognitive Sciences* 19:304–313.
- Haynes J-D (2015) A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives. *Neuron* 87:257–270.
- Hutchison RM, Womelsdorf T, Allen EA, Bandettini PA, Calhoun VD, Corbetta M, Della Penna S,



## High-dimensional mediation analysis of pain

- Duyn JH, Glover GH, Gonzalez-Castillo J, Handwerker DA, Keilholz S, Kiviniemi V, Leopold DA, de Pasquale F, Sporns O, Walter M, Chang C (2013) Dynamic functional connectivity: Promise, issues, and interpretations. *NeuroImage* 80:360–378.
- 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.
- 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.
- Krishnan A, Woo C-W, Chang LJ, Ruzic L, Gu X, López-Solà M, Jackson PL, Pujol J, Fan J, Wager TD (2016) Somatic and vicarious pain are represented by dissociable multivariate brain patterns. *eLife* 5:e15166.
- Krummenacher P, Candia V, Folkers G, Schedlowski M, Schönbacher G (2010) Prefrontal cortex modulates placebo analgesia. *Pain* 148:368–374.
- Kucyi A, Davis KD (2015) The dynamic pain connectome. *Trends in Neurosciences* 38:86–95.
- 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.
- 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* 145, Part B:274–287.
- Lindquist MA, Meng Loh J, Atlas LY, Wager TD (2009) Modeling the hemodynamic response function in fMRI: Efficiency, bias and mis-modeling. *NeuroImage* 45:S187–S198.
- 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.
- Melzack R (1999) From the gate to the neuromatrix: *Pain* 82:S121–S126.
- Moisset X, de Andrade D c., Bouhassira D (2016) From pulses to pain relief: an update on the mechanisms of rTMS-induced analgesic effects. *Eur J Pain* 20:689–700.
- Mori F, Codecà C, Kusayanagi H, Monteleone F, Buttari F, Fiore S, Bernardi G, Koch G, Centonze D (2010) Effects of Anodal Transcranial Direct Current Stimulation on Chronic



## High-dimensional mediation analysis of pain

- Neuropathic Pain in Patients With Multiple Sclerosis. *The Journal of Pain* 11:436–442.
- 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.
- Navratilova E, Porreca F (2014) Reward and motivation in pain and pain relief. *Nature Neuroscience* 17:1304–1312.
- Ohara S, Crone NE, Weiss N, Kim JH, Lenz FA (2008) Analysis of synchrony demonstrates that the presence of “pain networks” prior to a noxious stimulus can enable the perception of pain in response to that stimulus. *Exp Brain Res* 185:353–358.
- Oshiro Y, Quevedo AS, McHaffie JG, Kraft RA, Coghill RC (2007) Brain Mechanisms Supporting Spatial Discrimination of Pain. *J Neurosci* 27:3388–3394.
- Pagano RL, Assis DV, Clara JA, Alves AS, Dale CS, Teixeira MJ, Fonoff ET, Britto LR (2011) Transdural motor cortex stimulation reverses neuropathic pain in rats: A profile of neuronal activation. *European Journal of Pain* 15:268.e1-268.e14.
- Passard A, Attal N, Benadhira R, Brasseur L, Saba G, Sichere P, Perrot S, Januel D, Bouhassira D (2007) Effects of unilateral repetitive transcranial magnetic stimulation of the motor cortex on chronic widespread pain in fibromyalgia. *Brain* 130:2661–2670.
- 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.
- 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.
- 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.
- Roy M, Shohamy D, Wager TD (2012) Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends in Cognitive Sciences* 16:147–156.
- Sadaghiani S, Poline J-B, Kleinschmidt A, D’Esposito M (2015) Ongoing dynamics in large-scale functional connectivity predict perception. *PNAS* 112:8463–8468.
- Satpute AB, Wager TD, Cohen-Adad J, Bianciardi M, Choi J-K, Buhle JT, Wald LL, Barrett LF (2013) Identification of discrete functional subregions of the human periaqueductal gray. *PNAS* 110:17101–17106.
- Schafer SM, Geuter S, Wager TD (2018) Mechanisms of placebo analgesia: A dual-process model informed by insights from cross-species comparisons. *Progress in Neurobiology* 160:101–122.

## High-dimensional mediation analysis of pain

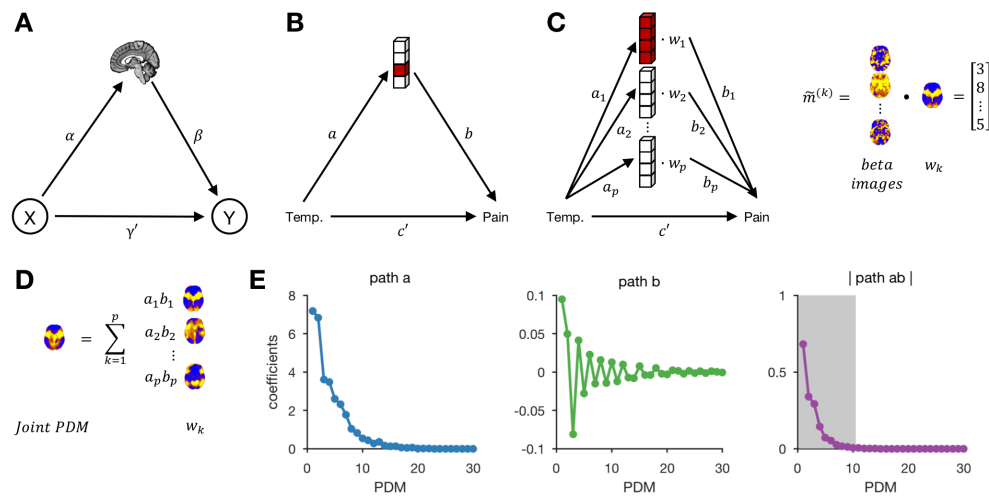
- Schulz E, Zherdin A, Tiemann L, Plant C, Ploner M (2012) Decoding an Individual's Sensitivity to Pain from the Multivariate Analysis of EEG Data. *Cereb Cortex* 22:1118–1123.
- Schwartz N, Temkin P, Jurado S, Lim BK, Heifets BD, Polepalli JS, Malenka RC (2014) Decreased motivation during chronic pain requires long-term depression in the nucleus accumbens. *Science* 345:535–542.
- Schweinhart P, Bushnell MC (2010) Pain imaging in health and disease — how far have we come? *J Clin Invest* 120:3788–3797.
- Seminowicz DA, Moayedi M (2017) The Dorsolateral Prefrontal Cortex in Acute and Chronic Pain. *The Journal of Pain* 18:1027–1035.
- Shepard RN (1980) Multidimensional Scaling, Tree-Fitting, and Clustering. *Science* 210:390–398.
- Turk DC, Melzack R (2011) The Measurement of Pain and the Assessment of People Experiencing Pain. In: *Handbook of Pain Assessment*, 3rd ed. (Turk DC, Melzack R, eds), pp 3–16. New York: Guilford Press.
- Wager TD, Atlas LY, Leotti LA, Rilling JK (2011) Predicting Individual Differences in Placebo Analgesia: Contributions of Brain Activity during Anticipation and Pain Experience. *The Journal of Neuroscience* 31:439–452.
- 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, Nichols TE (2003) Optimization of experimental design in fMRI: a general framework using a genetic algorithm. *NeuroImage* 18:293–309.
- 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 subregions of the medial prefrontal cortex and heart-rate reactivity. *NeuroImage* 47:821–835.
- Wiech K (2016) Deconstructing the sensation of pain: The influence of cognitive processes on pain perception. *Science* 354:584–587.
- 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

## High-dimensional mediation analysis of pain

synthesis of human functional neuroimaging data. Nat Meth 8:665–670.

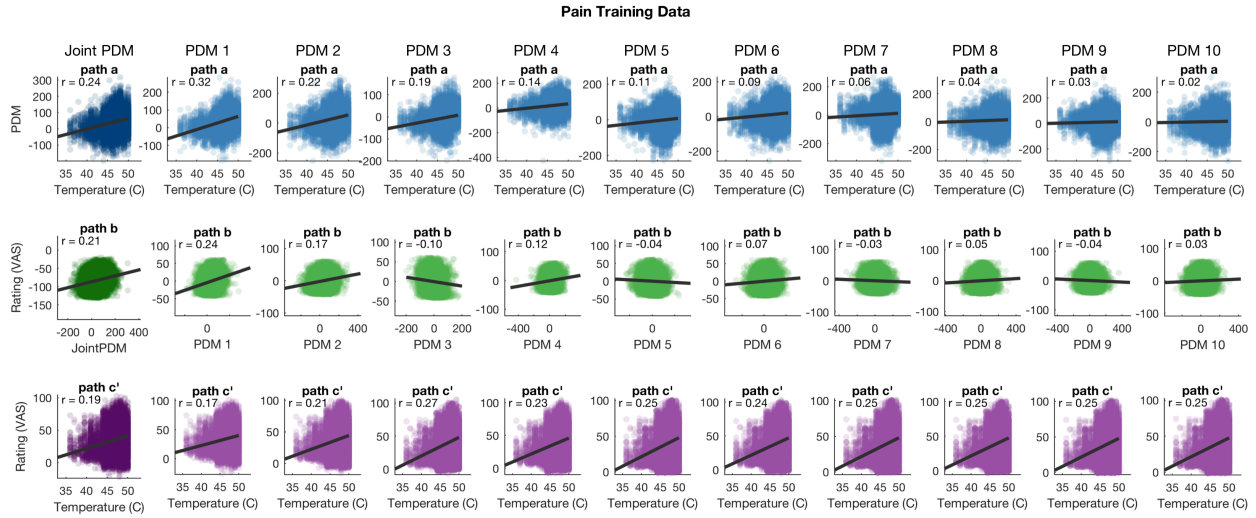
## High-dimensional mediation analysis of pain

### Figures



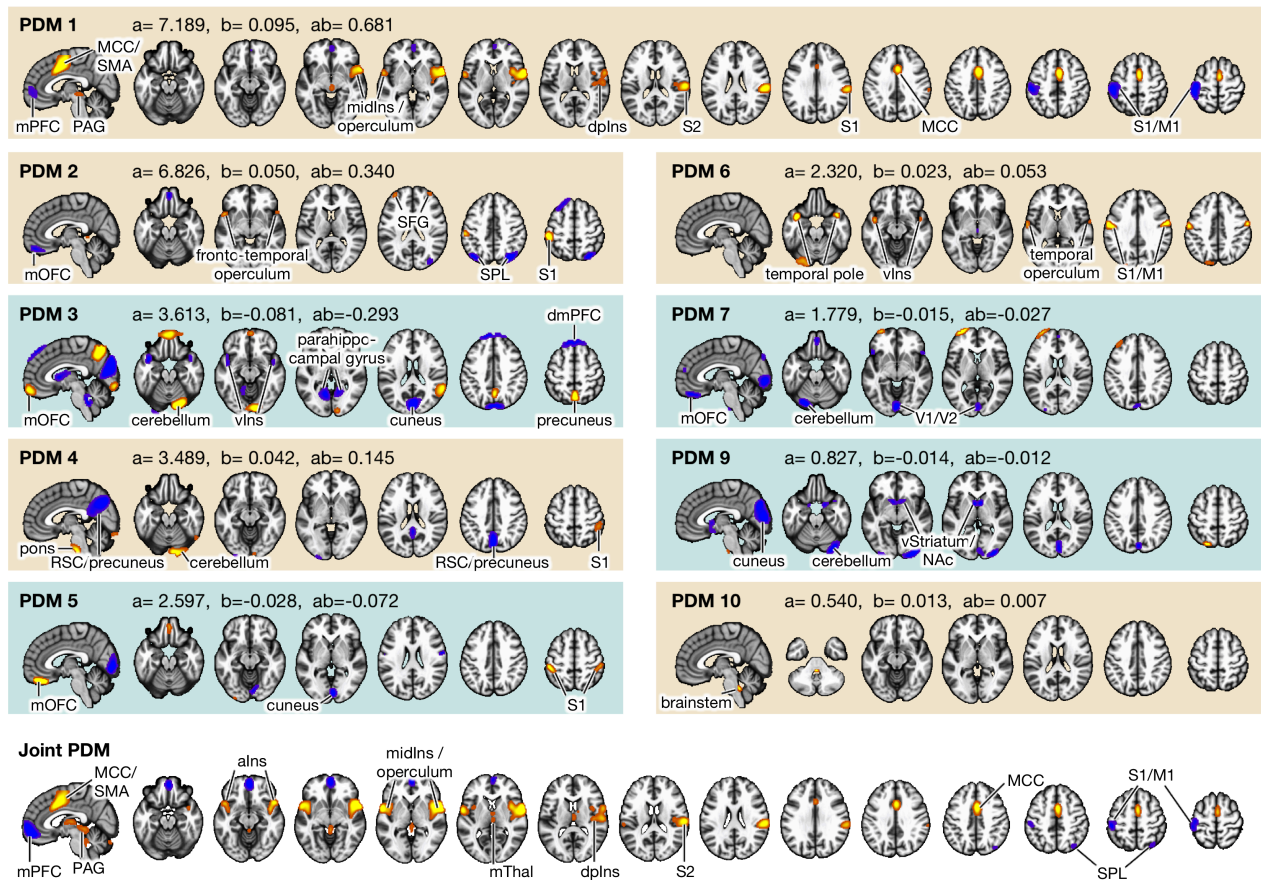
**Figure 1** Mediation analysis. (A) Schematic of the mediation analysis framework. Brain activity is an intermediate variable between a manipulated variable  $X$  and outcome  $Y$ . (B) In the univariate case, a separate mediation analysis is computed for every brain voxel to determine mediators between stimulation temperature and pain report. (C) In the high-dimensional *Principal Directions of Mediation* (PDM) approach, a linear combination of all brain voxels is used as a mediator. 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. Taking the dot product of the PDM ( $w_k$ ) and single-trial brain activation maps (beta images) results in a vector representing a potential mediator. Voxel weights ( $w_k$ ) are fit so that the indirect, mediated effect is maximal. (D) Individual PDMs can be combined into a single joint PDM by summing the individual PDMs weighted by their path coefficients because individual PDMs are orthogonal to each other. (E) 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).

## High-dimensional mediation analysis of pain



**Figure 1-supplement 1** 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.

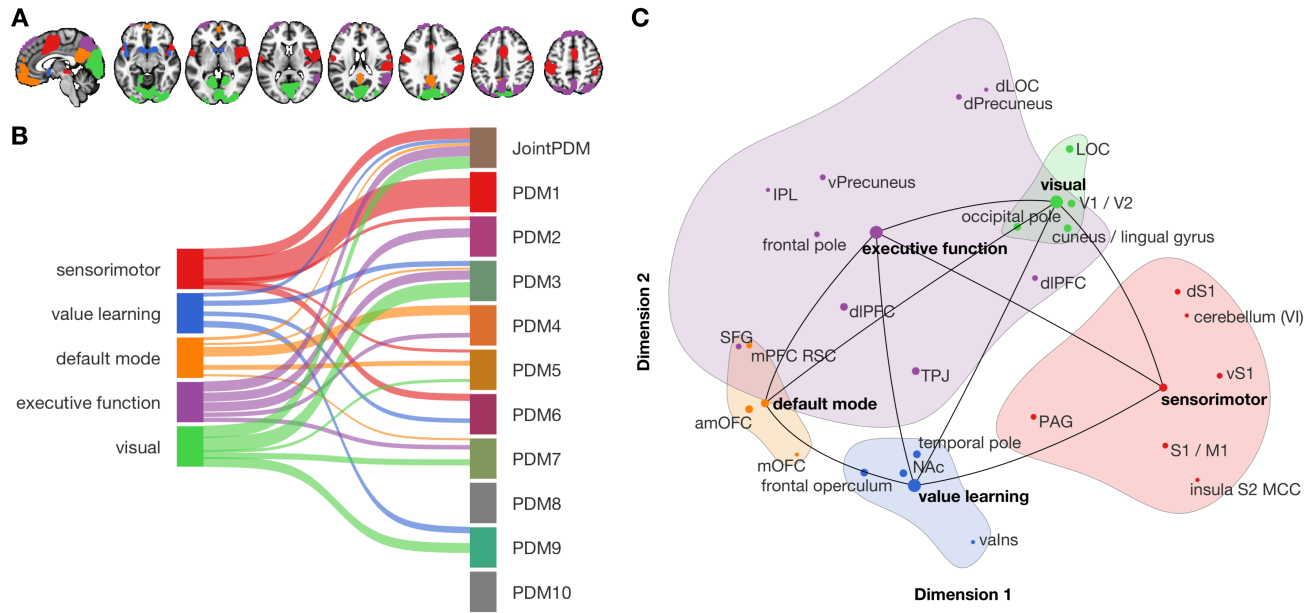
## High-dimensional mediation analysis of pain



**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 indicated 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 joint PDM, a combination of the above 10 PDMs. Regions with individually significant voxels in the joint PDM include somatosensory regions, such as S1, S2, insula, MCC, SMA, PAG, and thalamus, but also mPFC M1, and SPL. 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, dplns=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, vlns=ventral insula, dmPFC=dorsomedial prefrontal cortex, V1=primary visual cortex, V2=secondary visual cortex, vStriatum=ventral striatum, NAC=nucleus accumbens, mThal=medial thalamus, alns= anterior insula, SPL=superior parietal lobule.

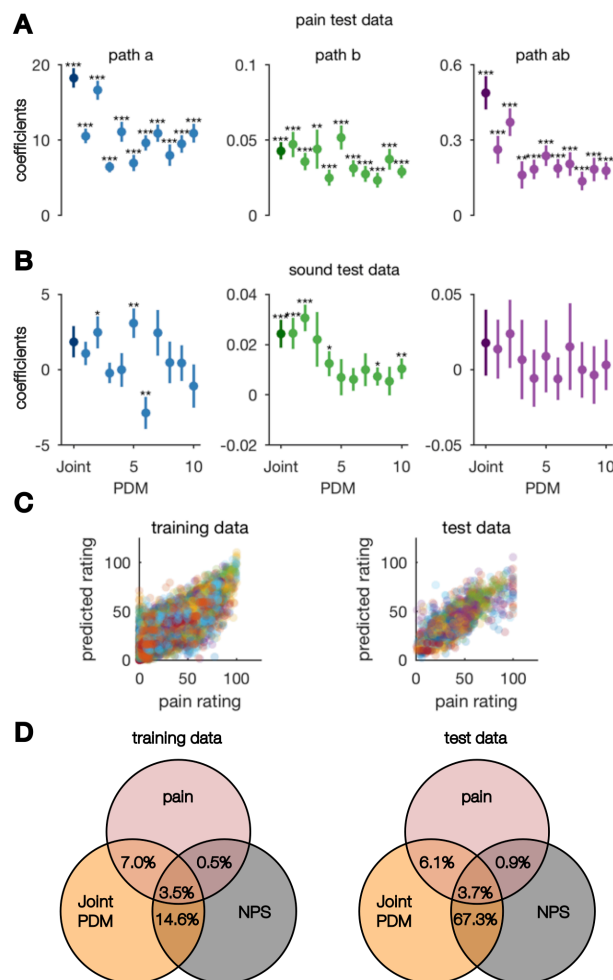


## High-dimensional mediation analysis of pain



**Figure 3** 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 the joint and individual PDMs. Ribbon width represents Dice-coefficient similarity between networks and PDMs. (C) Projection of the five functional networks and individual regions onto the first two dimensions spanned by the NMDS solution. Circle size indicates the number of significant connections for each region or network.

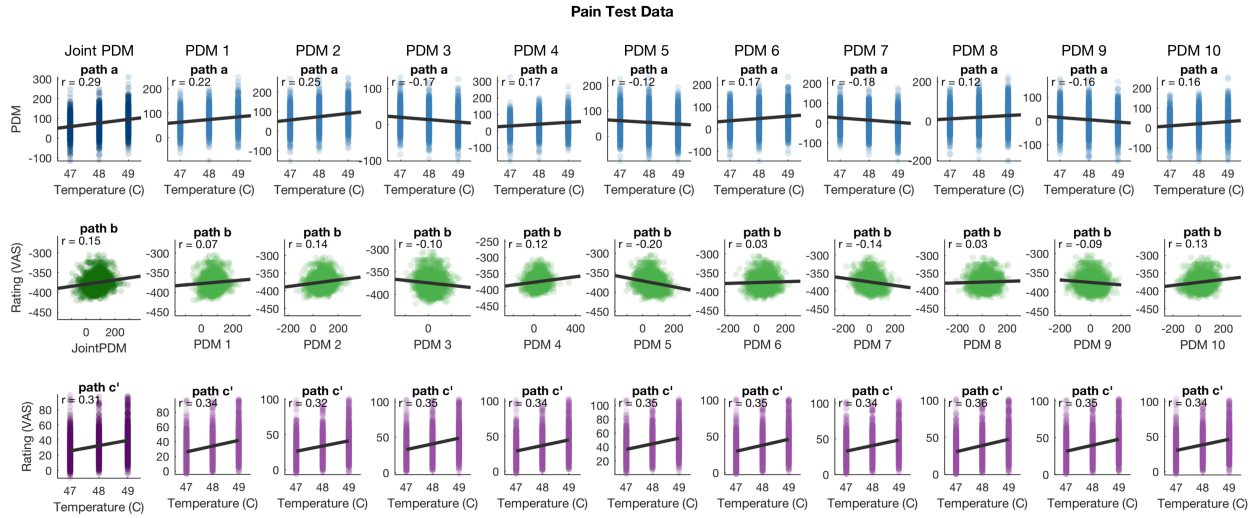
## High-dimensional mediation analysis of pain



**Figure 4** Validation on independent data (N=75). (A) The joint (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 joint PDM against empirical pain ratings for training (left) and test (right) pain data. Individual trials from all subjects are shown. Colors indicate different subjects. (D) Variance explained in single-trial pain ratings of the training and test data sets for the joint PDM and the NPS, which was only trained on pain ratings without temperature information. The joint PDM accounts for 7% and 6.1%, respectively, pain rating variance not accounted for by the NPS. Error bars indicate SEM. \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ .

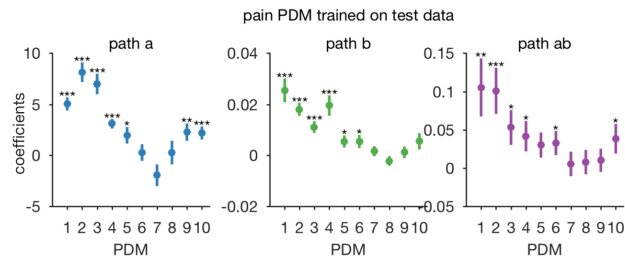


## High-dimensional mediation analysis of pain



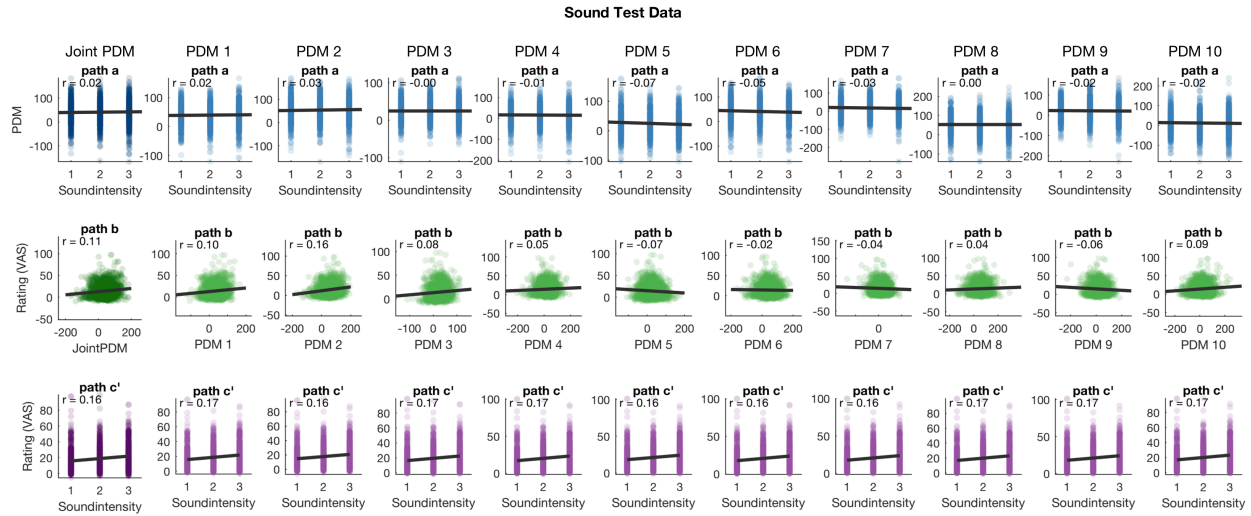
**Figure 4-supplement 1** Bivariate relationships between temperatures, mediators (PDM expression), and pain ratings for the pain test data (Study 8, N = 75). Data are adjusted according to the mediation equations, i.e. ratings in path b plots are adjusted for temperatures and PDMs, ratings in path c' 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).

## High-dimensional mediation analysis of pain



**Figure 4-supplement 2** Generalization of pain PDMs from small sample to large sample. Here, test and training data sets were switched. 10 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.

## High-dimensional mediation analysis of pain



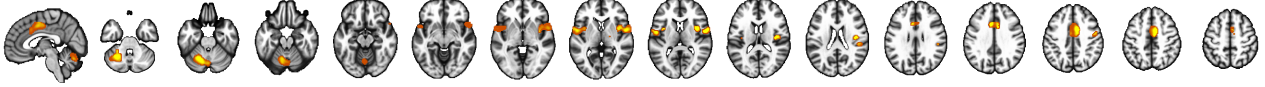
**Figure 4-supplement 3** Bivariate relationships between sound intensity levels, mediators (PDM expression), and intensity ratings for the training data (studies 1-7). Data are adjusted according to the mediation equations, i.e. ratings in path b plots are adjusted for stimulus levels and PDM, ratings 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).

## High-dimensional mediation analysis of pain

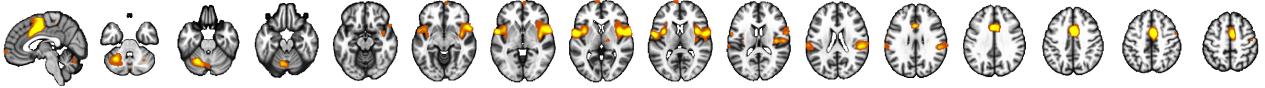
### path a - temperature to brain



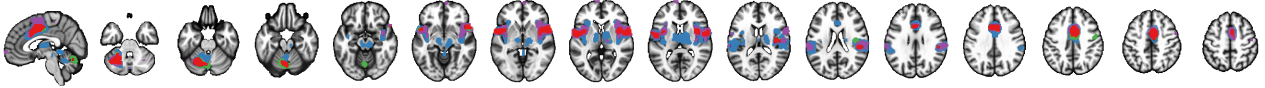
### path b - brain to pain



### path ab - temperature to brain to pain

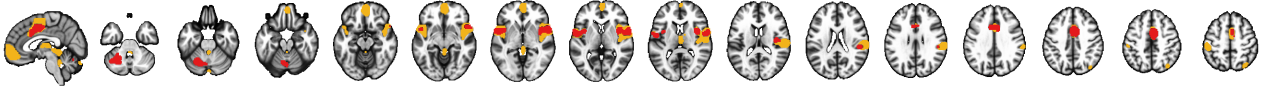


### path overlap (univariate mediation)



● path a    ● path b    ● path ab    ● univariate mediation (intersection)

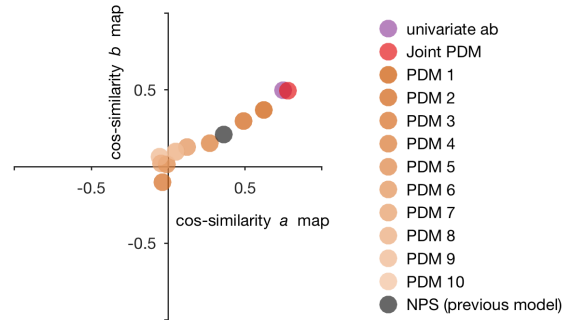
### comparison to joint PDM



● Joint PDM    ● univariate

**Figure 5** 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). The bottom panel compares the univariate mediation map (red) and the joint PDM (yellow).

## High-dimensional mediation analysis of pain



**Figure 6** Similarity of mediation maps to univariate stimulus intensity and pain rating maps. Similarity between the PDMs, univariate mediation (path  $ab$ ), and a previous pain predictive map (NPS) to the univariate maps for path  $a$  and  $b$ , respectively, measured by the cosine similarity between pairs of maps. The joint PDM and the univariate  $ab$  map most similar to the  $a$  and  $b$  maps representing stimulus intensity and pain intensity, respectively. The similarity between the joint PDM and the path  $ab$  map is only 0.45. Individual PDMs and the NPS are less and less similar to the univariate effect maps, indicating that the univariate maps do not capture all the information of the multivariate mediation maps.

Table 1. *Demographics*

Study <sup>♦</sup>	Sample Size	Sex	Mean age in Years (Std. Deviation)	Prior publications
<b>PDM Training Data</b>				
<b>Study 1 (NSF)</b>	26	9 F / 17 M	27.8	Atlas et al. (2014), <i>PAIN</i> ; Wager et al. (2013) <i>NEJM</i>
<b>Study 2 (BMRK3)</b>	33	22 F / 11 M	27.9 (9.0)	Woo et al. (2015), <i>PLoS Biology</i> Wager et al. (2013) <i>NEJM</i>
<b>Study 3 (BMRK4)</b>	28	10 F / 18 M	25.2 (7.4)	Krishnan et al. (2016) <i>eLife</i>
<b>Study 4 (IE)</b>	50	27 F / 23 M	25.1 (6.9)	Roy et al. (2014), <i>Nature Neuroscience</i>
<b>Study 5 (ILCP)</b>	29	16 F* / 12 M	20.4 (3.3)**	Schmidt et al. ( <i>in prep.</i> )
<b>Study 6 (EXP)</b>	17	9 F / 8 M	25.5	Atlas et al. (2010), <i>Journal of Neuroscience</i>
<b>Study 7 (SCEBL)</b>	26	11 F / 15 M	28 (9.3)	Koban et al. ( <i>in prep.</i> )
<b>PDM Test Data</b>				
<b>Study 8 (BMRK5)</b>	75	39 F / 36 M	28.2 (5.6)	Losin et al. ( <i>under review</i> )

*Note.* <sup>♦</sup>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.

Table 2. *Stimulation Parameters*

Study	Intensities	Mean Temperature by Intensity Level (Within Subject SE)	Rating scale	Mean Ratings by Intensity Level (Within Subject SEM)
<b>PDM Training Data</b>				
<b>Study 1 (NSF)</b>	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)
<b>Study 2 (BMRK3)</b>	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)
<b>Study 3 (BMRK4)</b>	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)
<b>Study 4 (IE)</b>	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)
<b>Study 5 (ILCP)</b>	L, H (Calibrated)	44.7, 46.7 (0)	0-8 VAS (no pain to worst pain imaginable)	24.3, 46.7 (1.14)
<b>Study 6 (EXP)</b>	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)
<b>Study 7 (SCEBL)</b>	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)
<b>PDM Test Data</b>				
<b>Study 8 (BMRK5)</b>	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.



Table 3. *Task Characteristics*

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

Table 4. *Neurosynth.org* network associations

<b>Sensorimotor</b>		<b>Value-learning</b>		<b>Default mode</b>		<b>Executive function</b>		<b>Visual</b>	
<i>r</i>	<i>features</i>	<i>r</i>	<i>features</i>	<i>r</i>	<i>features</i>	<i>r</i>	<i>features</i>	<i>r</i>	<i>features</i>
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.