nature neuroscience

ARTICLES

https://doi.org/10.1038/s41593-017-0051-7

Generalizable representations of pain, cognitive control, and negative emotion in medial frontal cortex

Philip A. Kragel[1*], Michiko Kano[2,3], Lukas Van Oudenhove[4], Huynh Giao Ly[4], Patrick Dupont[5], Amandine Rubio[6,7,8], Chantal Delon-Martin[6,7], Bruno L. Bonaz[6,7,8], Stephen B. Manuck[9], Peter J. Gianaros[9], Marta Ceko[1], Elizabeth A. Reynolds Losin[10], Choong-Wan Woo[11,12], Thomas E. Nichols[13] and Tor D. Wager[1*]

The medial frontal cortex, including anterior midcingulate cortex, has been linked to multiple psychological domains, including cognitive control, pain, and emotion. However, it is unclear whether this region encodes representations of these domains that are generalizable across studies and subdomains. Additionally, if there are generalizable representations, do they reflect a single underlying process shared across domains or multiple domain-specific processes? We decomposed multivariate patterns of functional MRI activity from 270 participants across 18 studies into study-specific, subdomain-specific, and domain-specific components and identified latent multivariate representations that generalized across subdomains but were specific to each domain. Pain representations were localized to anterior midcingulate cortex, negative emotion representations to ventromedial prefrontal cortex, and cognitive control representations to portions of the dorsal midcingulate. These findings provide evidence for medial frontal cortex representations that generalize across studies and subdomains but are specific to distinct psychological domains rather than reducible to a single underlying process.

A central aim of cognitive neuroscience is to identify how different mental processes are represented in brain activity. The medial frontal cortex (MFC), which includes multiple functionally distinct cortical areas in the superior frontal and cingulate gyri[1], is one brain region that has been linked to diverse psychological domains, i.e., sets of related psychological states with different adaptive functions[2]. Clearly, different areas within MFC encode different functions, but there is a striking convergence of overlapping functions across domains in several 'hub' areas, particularly the anterior midcingulate cortex (aMCC[3]). Research across species has linked activity in aMCC with multiple functions, including cognitive control[4,5], reward-based learning and decision making[6–9], somatic pain[10,11], and processing of emotional[12,13] and social information[14,15]. In fact, this area responds to such a variety of tasks, and so many underlying functions have been proposed to explain its responses, that it has been described as a "Rorschach test" and understanding it a "holy grail for many cognitive neuroscientists."[16]

Theories of aMCC function often explain the numerous signals in this area as components of an underlying process that operates across domains. Candidate processes have included conflict monitoring[4], adaptive control (i.e., control processes broadly engaged by negative affect and nociception[17]), cognitive effort[18], valuation of actions[19] and control[20], and detecting threats to survival[21], among others. These models have value because they offer integrative explanations for aMCC engagement across multiple domains. However, measuring brain activity across domains with functional MRI (fMRI) glosses over a potential multiplicity of different local neural circuits with distinct functions[22,23]. Electrophysiological and optogenetic studies of likely homologs of human aMCC provide evidence for distinct subpopulations of neurons with different functional properties[6,8,24]. Recent evidence suggests that multivariate patterns of fMRI activity can, in some cases, identify representations distributed across subpopulations of cells, including identifying functionally dissociable patterns within aMCC associated with different tasks[25,26].

Thus, unified accounts of aMCC function make predictions about the similarity of multivariate brain representations across domains that have not been adequately tested. If a set of domains activate representations of a single underlying process, then engaging these representations by tasks from these domain sets should produce similar patterns of brain activity in aMCC and other MFC areas. Conversely, if different domains engage an underlying pattern that is specific to each domain and not shared by other domains, this would provide evidence against a common underlying process.

Here we test these predictions using a construct-validation approach grounded in psychometric theory. We investigated three constructs that engage MFC: pain, cognitive control, and negative emotion (see Methods). We sampled human fMRI data from 18

[1]Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, CO, USA. [2]Frontier Research Institute for Interdisciplinary Sciences, Tohoku University, Sendai, Japan. [3]Department of Behavioral Medicine, Tohoku University Graduate School of Medicine, Sendai, Japan. [4]Department of Clinical and Experimental Medicine, University of Leuven, Leuven, Belgium. [5]Department of Neurosciences, University of Leuven, Leuven, Belgium. [6]Grenoble Institut des Neurosciences, GIN, Univ. Grenoble Alpes, Grenoble, France. [7]INSERM, Grenoble, France. [8]CHU Grenoble Alpes, Grenoble, France. [9]Department of Psychology, University of Pittsburgh, Pittsburgh, PA, USA. [10]Department of Psychology, University of Miami, Miami, FL, USA. [11]Center for Neuroscience Imaging Research, Institute for Basic Science, Suwon, Republic of Korea. [12]Department of Biomedical Engineering, Sungkyunkwan University, Suwon, Republic of Korea. [13]Oxford Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Population Health, University of Oxford, Oxford, UK. *e-mail: philip.kragel@colorado.edu; tor.wager@colorado.edu

studies (15 subjects per study, total $n=270$) in a balanced, hierarchical structure, with three different experimental manipulations in each domain (for example, evoked cutaneous pain, visceral nociceptive pain, and acute mechanical stimulation pain) and two independent studies for each of these experimental manipulations (i.e., subdomain; Fig. 1a). Although it is commonplace in neuroimaging studies to equate a pattern of activity from a single study with a 'representation', measurement theory and first principles dictate that representations of latent constructs must be generalizable. For instance, a representation of 'pain' must generalize across different types of painful stimuli. Our approach allowed us to develop multivariate models that localize brain representations that correspond to a single domain, rather than being driven by the particulars of a subdomain or idiosyncrasies of an individual study (Fig. 1b). In this way, these models assess the generalizability of brain representations and test the validity of the theoretical constructs of pain, cognitive control, and negative emotion.

Results

Anatomical delineation of psychological domains. Given evidence for regional specialization of cingulate function on the basis of cytoarchitecture[27], we first applied representational similarity analysis[28] within six anatomically defined cortical regions of interest: posterior midcingulate (pMCC), aMCC, perigenual anterior cingulate, subgenual anterior cingulate, ventromedial prefrontal (vmPFC), and dorsal MFC (dMFC; Fig. 2). By assessing how similar patterns of brain activity are across studies, subdomains, and domains in a single model, representational similarity analysis can provide evidence for generalizable brain representations.

This analysis revealed generalizable representations of painful stimulation in aMCC, pMCC, and dMFC that were not shared by other domains. Parameter estimates for the effect of the pain domain—across heat, mechanical, and visceral pain subdomains and controlling for study-level and subdomain-level effects—were positive within aMCC ($\hat{\beta}=0.990\pm0.266$ (s.e.m.), $z=3.72$, $P=0.0002$), pMCC ($\hat{\beta}=0.470\pm0.186$ (s.e.m.), $z=2.55$, $P=0.0107$), and dMFC ($\hat{\beta}=0.294\pm0.116$ (s.e.m.), $z=2.59$, $P=0.0097$). See Supplementary Table 1 for more information. These results indicate that patterns of pain-evoked activity in these areas are qualitatively distinct from activity patterns elicited during manipulations of cognitive control or negative emotion, independent of subdomain and study. Accordingly, in terms of aMCC activity patterns, participants in pain studies across different subdomains were more similar to each other ($r=0.1289\pm0.0039$ (s.e.m.)) than to those in studies of cognitive control ($r=0.0083\pm0.0026$ (s.e.m.); 95% confidence interval (CI) of difference = [0.1089, 0.1325]) or negative emotion ($r=0.0111\pm0.0032$ (s.e.m.), 95% CI of difference = [0.1051, 0.1297]; Fig. 2). Because these correlations are computed across subdomains, they are unlikely to be driven by similarity in any particular subdomain or study. Qualitatively similar results held for patterns of activity in pMCC and dMFC, although they were smaller in magnitude (Supplementary Tables 1 and 2). These findings are concordant with

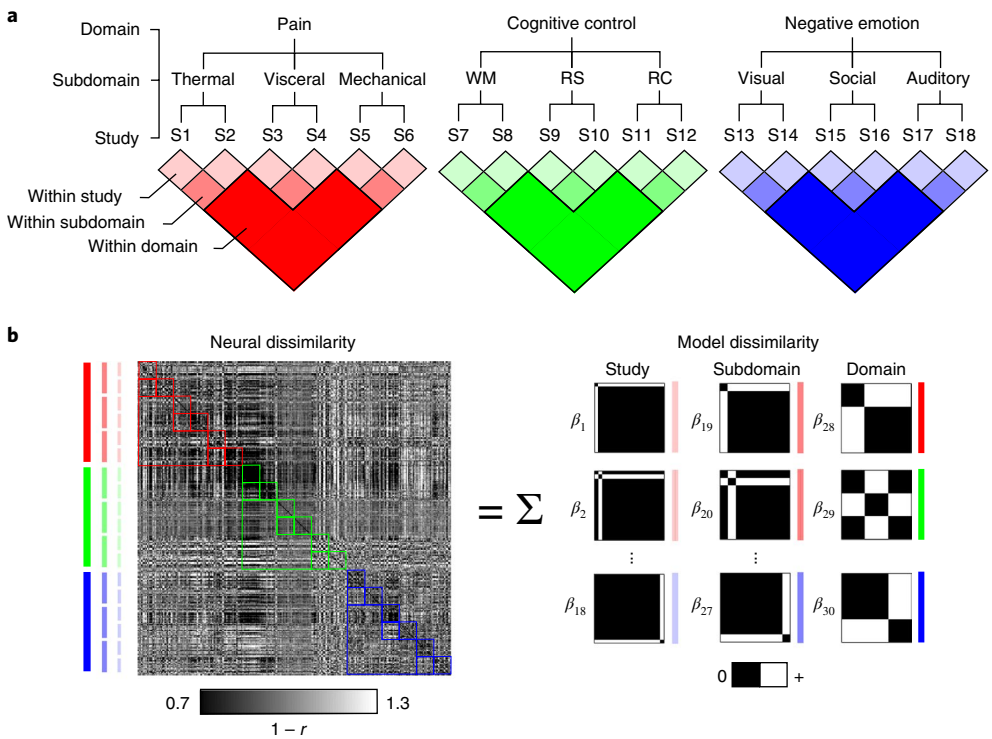

Fig. 1 | Study selection and multivariate modeling. a, Hierarchical structure of studies and tasks. Dendrograms convey theoretical groupings of fMRI activity at levels of study (level 1: studies S1–S18), subdomain (level 2: thermal, visceral, mechanical, working memory (WM), response selection (RS), response conflict (RC), visual, social, and auditory), and domain (level 3: pain, cognitive control, and negative emotion). Colored regions illustrate model-based partitioning of neural similarity into components that generalize across subjects (unique to a study, top 18 squares), studies (unique to a subdomain, middle nine squares), and subdomains (unique to a domain, bottom three regions). **b**, Decomposing multivariate pattern similarity into study-, subdomain-, and domain-specific components. The matrix in the left panel shows the dissimilarity of fMRI patterns across all subjects ($n=270$) in the entire medial frontal cortex. Each row represents one individual participant, and each element the dissimilarity (1–Pearson's correlation coefficient) in brain activity patterns for two individuals. Colored bars to the left indicate corresponding levels in the functional hierarchy. The right panel shows how the observed neural dissimilarity across pairs of images from the 18 studies is modeled as a weighted summation of theoretical dissimilarity matrices constructed according to study (18 parameters), subdomain (9 parameters), and domain (3 parameters) membership, in addition to a constant term (not shown).

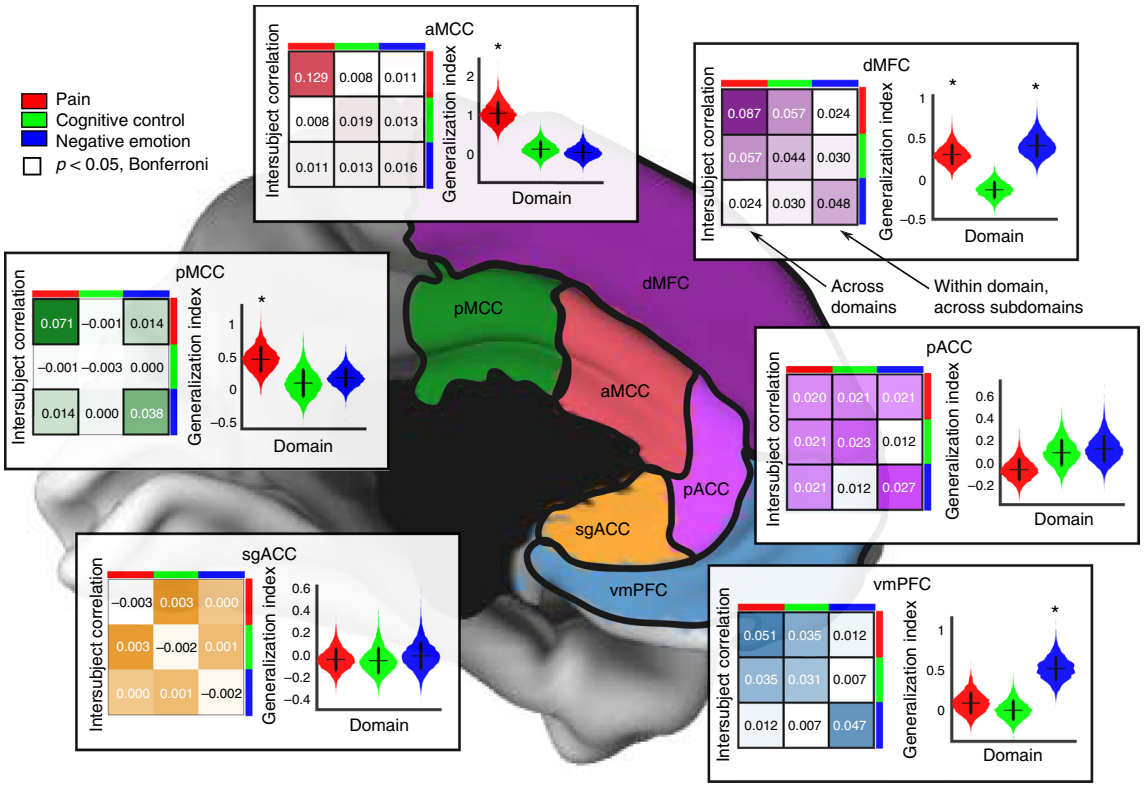

Fig. 2 | Regional assessment of generalizable representations in MFC. Brain rendering depicts the anatomical parcellation of MFC using the four-region model of cingulate cortex[27] comprising aMCC, perigenual anterior cingulate cortex (pACC), subgenual anterior cingulate cortex (sgACC), and pMCC, in addition to vmPFC and dMFC. Surrounding panels depict the average intersubject Pearson correlations (36,315 pairwise correlations computed from $n=270$ participants) both within domains but from different subdomains ($n=2{,}700$ pairwise correlations) and between different domains ($n=8{,}100$ pairwise correlations, left), in addition to bootstrap distributions of the generalization index computed from the full sample ($b=5{,}000$ bootstrap samples, right). Elements depicted inside the violin plots indicate the bootstrap standard error. Correlations with significantly positive values (bootstrap test, Bonferroni corrected $P<0.05$) have solid black borders. The generalization index quantifies the extent to which brain activity within a region is similar within a domain (across different subdomains) but is different across domains. Asterisks (*) indicate FDR-corrected $q<0.05$.

theoretical models that implicate the aMCC in pain[29,30] and with studies identifying nociceptive circuits in dorsal anterior cingulate cortex[31]. Observations of aMCC activity during noxious stimulation have often been attributed to more general mechanisms, such as directing attention, response selection, or responding to salient events. However, we identified representations of evoked pain distinct from those related to cognitive and emotional domains, which are also attention-demanding, salient, and involve motor preparation, ruling out such general explanations as the primary drivers of aMCC responses during painful stimulation.

The regional analysis also revealed generalizable representations of negative emotion—across social emotion, emotional pictures, and emotional sounds— in vmPFC ($\widehat{\beta}=0.514\pm0.140$ (s.e.m.), $z=3.65$, $P=0.0003$) and dMFC ($\widehat{\beta}=0.404\pm0.133$ (s.e.m.), $z=3.03$, $P=0.0024$; Supplementary Table 2). Within vmPFC, patterns of activation from different subdomains of negative emotion were more similar to each other ($r=0.0474\pm0.0027$ (s.e.m.)) than they were to evoked pain ($r=0.0117\pm0.0023$ (s.e.m.), 95% CI of difference = [0.0270, 0.0440]) or cognitive control ($r=0.0072\pm0.0020$ (s.e.m.), 95% CI of difference = [0.0316, 0.0484]) studies (Supplementary Table 4). These observations agree with those of recent neuroimaging studies identifying representations of cross-modal subjective value[32] and perceived emotion[33] in vmPFC. By revealing representations of negative emotion that generalize across stimulus modality and social contexts, these results further substantiate the notion that vmPFC integrates emotional value across diverse stimuli[34,35]. Further, these data suggest that, although painful and unpleasant emotional events can engage a shared negative affective component, vmPFC representations evoked by these two types of stimuli are qualitatively distinct. Recent meta-analytic work has suggested that this difference may be related to the generation of affective meaning[36], in which information about environmental cues, memories of past events, and evaluations of potential outcomes are combined into an integrated representation of an organism's well-being in the current environment. This integrative processing would stand in contrast to affective representations that are not conceptually driven, such as pain. We note that these data do not directly assess the generalizability of vmPFC representations to positive emotion or to internally generated states elicited through memory retrieval, as we focused on inductions using negative stimuli.

No cingulate or other areas within MFC exhibited a generalizable representation specific to cognitive control across working memory (*N*-back tasks), response selection (stop-signal tasks), or response conflict (Flanker and Simon tasks) subdomains (see "Study and contrast selection" in Methods for citations to included studies). However, we did identify a generalizable representation of response selection, a subdomain of cognitive control particularly involved in motor inhibition, in vmPFC (Supplementary Table 1). As we observed deactivation in this area during task performance (Fig. 3; like others[37]), this representation may reflect a pattern of deactivation not shared by other domains or other cognitive control subdomains. Patterns of vmPFC activation from different response selection studies were more similar to one another ($r=0.0828\pm0.0033$ (s.e.m.)) than to those during manipulations

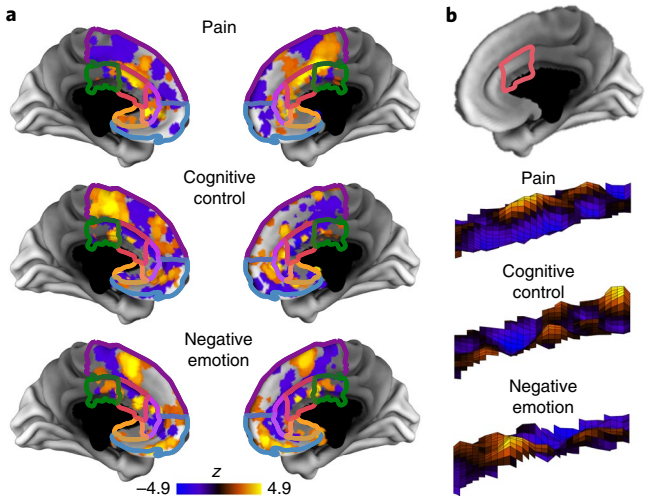

Fig. 3 | Identifying latent brain representations that predict the occurrence of distinct functional domains in each region of interest. **a**, Latent patterns of activity that generalize across studies and subdomains but are specific for the domains of pain, cognitive control, and negative emotion, extracted using partial-least-squares separately for each region and thresholded at $P<0.05$, uncorrected, for display ($n=270$ participants). **b**, Expanded view of latent patterns in aMCC in the left hemisphere. Images are displayed using radiological convention.

of negative emotion ($r=0.0052\pm0.0023$ (s.e.m.), 95% CI of difference = [0.0674, 0.0876]), evoked pain ($r=0.0286\pm0.0024$ (s.e.m.), 95% CI of difference = [0.0443, 0.0647]), working memory ($r=0.0514\pm0.0045$ (s.e.m.), 95% CI of difference = [0.0185, 0.0442]), or response conflict ($r=0.0254\pm0.0043$ (s.e.m.), 95% CI of difference = [0.447, 0.0701]; Supplementary Table 4). Thus, though generalizable representations of both negative emotion and response selection were observed within vmPFC, these representations appear to be distinct. It is also possible that control-related representations are highly dependent on individual study parameters, as we found strong study-specific effects in multiple regions, including aMCC (Supplementary Table 2).

Analysis of activation spanning the full extent of MFC (combining the six regions of interest) produced similar results, with effects of painful stimulation and negative emotion that generalize within but not across domains (Supplementary Fig. 1 and Supplementary Table 2). Confirmatory analyses that directly contrasted the spatial similarity of brain activity within domains against spatial correlations across domains further supported these results (Supplementary Fig. 2 and Supplementary Tables 3–5). Additional confirmatory analyses using different model parameterizations produced qualitatively similar results (Supplementary Fig. 3).

To quantify the weight of evidence favoring generalizable representations specific to each of the three domains, we additionally conducted model comparisons using the Bayesian information criterion in each region of interest (see Methods for details). Results of this analysis corroborate inferences drawn on individual parameter estimates (Table 1). aMCC representations were best explained by a model including the domain of pain (in addition to terms for study and subdomain), but not cognitive control or negative emotion. vmPFC representations, on the other hand, were best explained by modeling the domain of negative emotion but not pain or cognitive control. The best fitting models of dMFC and full MFC representations included all three domains, indicative of diverse coding in these regions. Additional model comparisons using the Brainnetome atlas, a parcellation based on functional and anatomical connectivity[38], provide evidence for generalizable representations in other brain regions outside the MFC as well (Supplementary Table 6 and Supplementary Fig. 4).

Searchlight mapping of psychological domains. As there is well-established variability in the anatomy of the cingulate sulcus[39], we additionally conducted searchlight mapping[40] to localize domain-specific representations without strongly relying on the boundaries between regions and to lessen the impact of anatomical variability. In this approach, we modeled the similarity structure of spherical volumes (radius = 8 mm) centered at each voxel in MFC, identifying areas wherein local patterns of brain activity contain generalized representations of pain, cognitive control, and negative emotion. By examining patterns of activation in small spherical volumes, these searchlights provide a smooth estimate of pattern information[40] that is not constrained by fixed boundaries that may not match the anatomy of every subject (of importance here, as ~40% of the population has a paracingulate gryus[39], which extends the spatial extent of MCC).

The results of the searchlight analysis were largely concordant with those based on anatomical parcellation (Fig. 4a,d). Generalizable representations of painful stimulation were found in aMCC within the cingulate sulcus ($z_{peak}=4.88$, Montreal Neurological Institute coordinates (MNI_{xyz}) = [2, 14, 24], $P=1.06\times10^{-6}$, $q<0.05$ false discovery rate (FDR)-corrected) and extending into dMFC ($z_{peak}=5.58$, $MNI_{xyz}=[2, 8, 46]$, $P=2.41\times10^{-8}$, $q<0.05$ FDR-corrected). Also consistent with the regional analysis, representations of negative emotion were found in dMFC, above the dorsal bank of the cingulate sulcus in pre-supplementary motor area (SMA; $z_{peak}=3.00$, $MNI_{xyz}=[-10, 10, 52]$, $P=0.0027$) and vmPFC ($z_{peak}=3.78$, $MNI_{xyz}=[0, 48, -10]$, $P=1.57\times10^{-4}$), albeit at lower (uncorrected) thresholds.

Unlike the regional analyses, the searchlight analysis revealed domain-specific representations of cognitive control along the

Table 1 | Bayesian information criterion (BIC) weights and adjusted R^2 for selected models

Region	Study and subdomain (28)	Pain (29)	Cognitive control (29)	Negative emotion (29)	Full model (31)	Adjusted R^2 (optimal model)
pMCC	< 0.0001	0.0673	< 0.0001	< 0.0001	**0.9327**	0.0220
aMCC	< 0.0001	**0.9901**	< 0.0001	< 0.0001	0.0099	0.0342
pACC	**0.4686**	0.0342	0.1411	0.3484	0.0077	0.0134
sgACC	**0.7986**	0.0704	0.0770	0.0536	0.0004	0.0006
vmPFC	< 0.0001	< 0.0001	< 0.0001	**0.9669**	0.0331	0.0567
dMFC	< 0.0001	< 0.0001	< 0.0001	< 0.0001	**1.0000**	0.0831
MFC	< 0.0001	< 0.0001	< 0.0001	< 0.0001	**1.0000**	0.0934

Bold font indicates models with highest BIC weights, and adjusted R^2 values for these optimal models in each region are listed based on the total variation in the data. BIC weights sum to 1 for each region. The number of free parameters in each model is listed in parentheses.

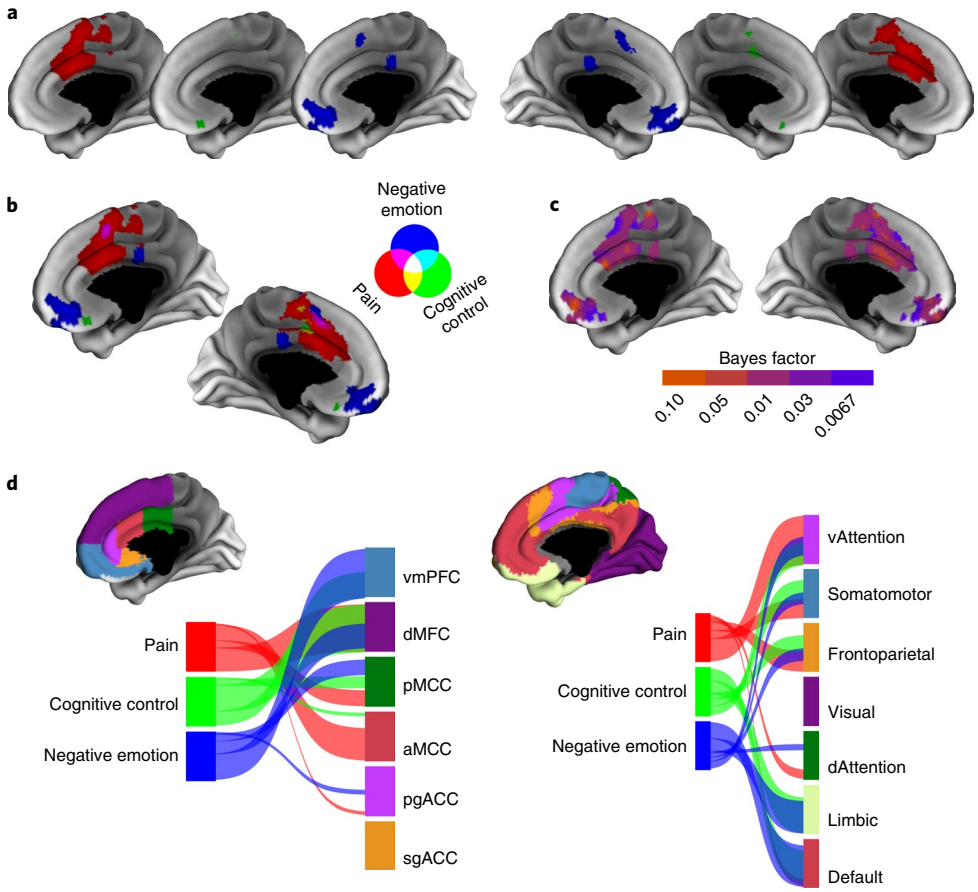

Fig. 4 | Representational mapping of pain, cognitive control, and negative emotion in MFC. a, Searchlight maps display where local patterns of brain activity are consistent with domain-specific representation of pain (red), cognitive control (green), and negative emotion (blue; $n = 270$ participants). **b**, Additive conjunction of searchlight maps, with each domain mapped onto orthogonal dimensions in the red–green–blue (RGB) color space. Overlap between pain and cognitive control is depicted in yellow; overlap between pain and negative emotion is colored magenta. Maps are thresholded at $P < 0.05$, two-tailed, uncorrected cutoff to highlight any possible overlap ($n = 270$ participants). **c**, Brain maps of Bayes factors indicating relative evidence against overlap among the three domains at each voxel. Smaller values indicate evidence against overlap; values less than 0.1 are considered strong evidence ($n = 270$ participants). **d**, River plots depict the similarity between searchlight maps and anatomical parcellation of MFC (left) and functional parcellation of cortical regions from resting-state data[48] (right). Line thickness indicates the degree of correspondence between sets. vAttention, ventral attention; dAttention, dorsal attention. Images are displayed using radiological convention.

cingulate sulcus, extending into SMA and motor cortex ($z_{peak} = 2.81$, $MNI_{xyz} = [-2, 2, 46]$, $P = 0.005$). This localization, which falls along the boundary between aMCC and dMFC, agrees with meta-analyses showing the epicenter of control-related activity in this area across working memory, inhibition, and attention-shifting tasks[41], and with the posterior rostral cingulate zone[42], a region classically thought to be involved in response selection. Due to its proximity to and connectivity with functionally related brain regions[17], this area is a prime candidate for integrating different types of control signals from multiple sources, such as the expected value of control[20] and value-guided behavioral adaptations[19].

Integrative views of cingulate function are in part supported by observations that overlapping activation is observed in aMCC and adjacent MFC during manipulations of pain, cognitive control, and social and evaluative processing[17]. To assess whether the domain-specific representations identified in the present study similarly overlap within the broader territory of the MFC, we performed a conjunction analysis of the searchlight maps (Fig. 4b). Results revealed that these representations were predominately dissociable; the three domains did not commonly overlap in any voxel. Minimal overlap was found in dMFC, with small clusters of activity coding for both pain and cognitive control (60 voxels, 0.57%), and for pain and negative emotion (93 voxels, 0.88%). A small degree of overlap was also observed for pain and negative emotion in vmPFC (17 voxels, 0.53%). The only overlapping effects in cingulate cortex were for pain and cognitive control, spanning the border between pMCC (6 voxels, 1.16%) and aMCC (2 voxels, 1.29%).

To evaluate evidence against overlap, we computed Bayes factors using the minimum z-score from the three-domain conjunction analysis[43,44]. In this analysis, if the minimum statistic from all three domains is less than or near zero, then there is little support for overlap. Conversely, if the minimum statistic is large and positive, it is more likely that there is overlap across the domains. Values < 1 reflect evidence in favor of the null hypothesis of no representation in all three domains, and values > 1 reflect evidence in favor of overlap. A Bayes factor < 0.1 is generally considered strong evidence against overlap, and a factor > 10 is generally considered strong evidence for overlap. This analysis revealed substantial evidence against overlap in the MFC (Fig. 4c), with a maximum Bayes factor $= 0.0898$.

Discussion

Our results reveal generalizable representations of pain, cognitive control, and negative emotion in separable patterns of MFC

activity. The limited overlap of generalizable representations across domains contrasts with conventional, univariate assessments of cingulate function, which show substantial overlap across domains[17,45,46]. Thus, our findings here highlight functional diversity within MFC, and they suggest that domain-specific representations exist in most parts of MFC, including aMCC and adjacent regions. Though domain-specific representations are limited to one domain (pain, negative emotion, and cognitive control), our design allows us to infer that they do generalize across multiple subdomains (for example, somatic thermal, somatic mechanical, and visceral pain). The generalizable representations we identify provide empirical constraints on what integrative theories of aMCC function must explain. In most of the MFC, it may not be necessary to explain pain- or affect-related and cognitive error–related signals with a single mechanism. Along these lines, it has recently been suggested that the aMCC functions to monitor for conflicts related to ever-present, survival-relevant goals[30]. Consistent with this proposal, we identified generalizable representations of pain in aMCC. However, these representations were qualitatively distinct from those evoked by negative emotional stimuli, including social rejection, that were not generalizable. This distinction makes it unlikely that the aMCC is engaged in the same way to achieve different survival-relevant goals. It is also consistent with other recent work identifying dissociations in MFC activity across tasks taken from different domains[47].

On the other hand, our findings leave room for integrative theories that explain computational mechanisms in terms of a convergence of different neural populations that interact to achieve computational goals. That is, it is possible that the same computational function may be implemented in diverse neural circuits, depending on their inputs and outputs. Our data therefore do not argue against unified computational accounts of aMCC, but rather against a unitary neural implementation of those computations.

The proximity of pain, negative emotion, and cognitive representations we identified provides a neural substrate for their comparison and integration[17]. It is possible that integration across domains could be identified in carefully controlled studies implementing within-subject designs across domains and subdomains. Further, although we included many studies and subdomains, our sampling was far from exhaustive, and testing specificity is an open-ended process. Future work combining a more diverse set of subdomains (for example, incorporating studies of chronic pain, positive and negative reward prediction error, and behavioral withdrawal) with model-based approaches will help further test the claims of integrative theories of control.

In conclusion, we identified domain-specific representations for pain and negative emotion in the aMCC and vmPFC. These representations generalized across participants and diverse subdomains (three per domain). These representations could only be identified by extracting generalizable brain patterns across studies and subdomains. Currently, conventional research investigating population-level neural representation has been conducted at the level of individual studies. Even when very large, studies that sample limited numbers of tasks (i.e., a single exemplar task for a psychological domain) are not capable of identifying generalizable brain representations in this way. Although cross-validation procedures, reliability analysis, and independent replication samples are becoming increasingly common to ensure the generalizability of results, it is evident that findings are often idiosyncratic to a particular study or experimental context, rather than truly reflecting the mental operations under study. Our results demonstrate how modeling the similarity structure of fMRI data drawn from multiple studies and subdomains can overcome this challenge and disambiguate latent brain representations of theoretical constructs.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41593-017-0051-7.

Received: 21 August 2017; Accepted: 29 November 2017;
Published online: 01 January 2018

References

1. Amodio, D. M. & Frith, C. D. Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* **7**, 268–277 (2006).
2. Cosmides, L. & Tooby, J. Origins of domain specificity: the evolution of functional organization. in *Mapping the Mind: Domain Specificity in Cognition and Culture* (eds. Hirschfeld, L.A. & Gelman, S.A.) 85–116 (1994).
3. Vogt, B. A. Midcingulate cortex: structure, connections, homologies, functions and diseases. *J. Chem. Neuroanat.* **74**, 28–46 (2016).
4. Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S. & Cohen, J. D. Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature* **402**, 179–181 (1999).
5. Ridderinkhof, K. R., Ullsperger, M., Crone, E. A. & Nieuwenhuis, S. The role of the medial frontal cortex in cognitive control. *Science* **306**, 443–447 (2004).
6. Ito, S., Stuphorn, V., Brown, J. W. & Schall, J. D. Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science* **302**, 120–122 (2003).
7. Dosenbach, N. U. et al. A core system for the implementation of task sets. *Neuron* **50**, 799–812 (2006).
8. Procyk, E., Tanaka, Y. L. & Joseph, J. P. Anterior cingulate activity during routine and non-routine sequential behaviors in macaques. *Nat. Neurosci.* **3**, 502–508 (2000).
9. Kolling, N., Behrens, T. E., Mars, R. B. & Rushworth, M. F. Neural mechanisms of foraging. *Science* **336**, 95–98 (2012).
10. Büchel, C. et al. 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. *J. Neurosci.* **22**, 970–976 (2002).
11. Rainville, P., Duncan, G. H., Price, D. D., Carrier, B. & Bushnell, M. C. Pain affect encoded in human anterior cingulate but not somatosensory cortex. *Science* **277**, 968–971 (1997).
12. Etkin, A., Egner, T., Peraza, D. M., Kandel, E. R. & Hirsch, J. Resolving emotional conflict: a role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron* **51**, 871–882 (2006).
13. Bishop, S., Duncan, J., Brett, M. & Lawrence, A. D. Prefrontal cortical function and anxiety: controlling attention to threat-related stimuli. *Nat. Neurosci.* **7**, 184–188 (2004).
14. Tomlin, D. et al. Agent-specific responses in the cingulate cortex during economic exchanges. *Science* **312**, 1047–1050 (2006).
15. Rudebeck, P. H., Buckley, M. J., Walton, M. E. & Rushworth, M. F. S. A role for the macaque anterior cingulate gyrus in social valuation. *Science* **313**, 1310–1312 (2006).
16. Ebitz, R. B. & Hayden, B. Y. Dorsal anterior cingulate: a Rorschach test for cognitive neuroscience. *Nat. Neurosci.* **19**, 1278–1279 (2016).
17. Shackman, A. J. et al. The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nat. Rev. Neurosci.* **12**, 154–167 (2011).
18. Critchley, H. D. et al. Human cingulate cortex and autonomic control: converging neuroimaging and clinical evidence. *Brain* **126**, 2139–2152 (2003).
19. Behrens, T. E., Woolrich, M. W., Walton, M. E. & Rushworth, M. F. Learning the value of information in an uncertain world. *Nat. Neurosci.* **10**, 1214–1221 (2007).
20. Shenhav, A., Botvinick, M. M. & Cohen, J. D. The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron* **79**, 217–240 (2013).
21. Eisenberger, N. I. & Lieberman, M. D. Why rejection hurts: a common neural alarm system for physical and social pain. *Trends Cogn. Sci.* **8**, 294–300 (2004).
22. Haynes, J. D. A primer on pattern-based approaches to fMRI: principles, pitfalls, and perspectives. *Neuron* **87**, 257–270 (2015).
23. Logothetis, N. K. What we can do and what we cannot do with fMRI. *Nature* **453**, 869–878 (2008).
24. Kvitsiani, D. et al. Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. *Nature* **498**, 363–366 (2013).
25. Krishnan, A. et al. Somatic and vicarious pain are represented by dissociable multivariate brain patterns. *eLife* **5**, e15166 (2016).
26. Woo, C. W. et al. Separate neural representations for physical pain and social rejection. *Nat. Commun.* **5**, 5380 (2014).
27. Vogt, B. A., Berger, G. R. & Derbyshire, S. W. Structural and functional dichotomy of human midcingulate cortex. *Eur. J. Neurosci.* **18**, 3134–3144 (2003).
28. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 4 (2008).

29. Peyron, R., Laurent, B. & García-Larrea, L. Functional imaging of brain responses to pain. A review and meta-analysis (2000). *Neurophysiol. Clin.* **30**, 263–288 (2000).
30. Lieberman, M. D. & Eisenberger, N. I. The dorsal anterior cingulate cortex is selective for pain: results from large-scale reverse inference. *Proc. Natl. Acad. Sci. USA* **112**, 15250–15255 (2015).
31. Hutchison, W. D., Davis, K. D., Lozano, A. M., Tasker, R. R. & Dostrovsky, J. O. Pain-related neurons in the human cingulate cortex. *Nat. Neurosci.* **2**, 403–405 (1999).
32. McNamee, D., Rangel, A. & O'Doherty, J. P. Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. *Nat. Neurosci.* **16**, 479–485 (2013).
33. Peelen, M. V., Atkinson, A. P. & Vuilleumier, P. Supramodal representations of perceived emotions in the human brain. *J. Neurosci.* **30**, 10127–10134 (2010).
34. Levy, D. J. & Glimcher, P. W. The root of all value: a neural common currency for choice. *Curr. Opin. Neurobiol.* **22**, 1027–1038 (2012).
35. Montague, P. R. & Berns, G. S. Neural economics and the biological substrates of valuation. *Neuron* **36**, 265–284 (2002).
36. Roy, M., Shohamy, D. & Wager, T. D. Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends Cogn. Sci.* **16**, 147–156 (2012).
37. Laird, A. R. et al. Investigating the functional heterogeneity of the default mode network using coordinate-based meta-analytic modeling. *J. Neurosci.* **29**, 14496–14505 (2009).
38. Fan, L. et al. The Human Brainnetome Atlas: a new brain atlas based on connectional architecture. *Cereb. Cortex* **26**, 3508–3526 (2016).
39. Paus, T. et al. Human cingulate and paracingulate sulci: pattern, variability, asymmetry, and probabilistic map. *Cereb. Cortex* **6**, 207–214 (1996).
40. Kriegeskorte, N., Goebel, R. & Bandettini, P. Information-based functional brain mapping. *Proc. Natl Acad. Sci. USA* **103**, 3863–3868 (2006).
41. Van Snellenberg, J. X. & Wager, T. D. Cognitive and motivational functions of the human prefrontal cortex. in *Luria's Legacy in the 21st Century* (Christiansen, A.-L., Goldberg, E. & Bougakov, D. eds.)30–61 (2009).
42. Amiezz, C. & Petrides, M. Neuroimaging evidence of the anatomo-functional organization of the human cingulate motor areas. *Cereb. Cortex* **24**, 563–578 (2014).
43. Gallistel, C. R. The importance of proving the null. *Psychol. Rev.* **116**, 439–453 (2009).
44. Dienes, Z. Using Bayes to get the most out of non-significant results. *Front. Psychol.* **5**, 781 (2014).
45. de la Vega, A., Chang, L. J., Banich, M. T., Wager, T. D. & Yarkoni, T. Large-scale meta-analysis of human medial frontal cortex reveals tripartite functional organization. *J. Neurosci.* **36**, 6553–6562 (2016).
46. Torta, D. M. & Cauda, F. Different functions in the cingulate cortex, a meta-analytic connectivity modeling study. *Neuroimage* **56**, 2157–2172 (2011).
47. Jahn, A., Nee, D. E., Alexander, W. H. & Brown, J. W. Distinct regions within medial prefrontal cortex process pain and cognition. *J. Neurosci.* **36**, 12385–12392 (2016).
48. Yeo, B. T. et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).

Acknowledgements

We thank S. Fukudo, T. Muratsubaki and J. Morishita for assistance with data collection; K. Ochsner for sharing data from studies of negative emotion; T. Braver and J. Gray for sharing working memory data; and R. Poldrack for sharing response selection data (available at https://openfmri.org/). This research was supported by grants R01 HL089850 to P.J.G.; P01 HL040962 to S.B.M.; grants OCI-1131801, R01 DA035484, and R01 MH076136 to T.D.W.; JSPS-FWO grant VS.014.13 N to L.V.O. and S. Fukudo; JSPS-KAKENHI grant 26460898 to M.K.; R01 MH076137 and R01 AG043463 to K.O.; by the Direction de la Recherche Clinique of the University Hospital of Grenoble Alpes; and by the pharmaceutical labs Ferring and Cephalon. L.V.O. is funded by the KU Leuven Special Research Fund. T.E.N. is supported by the Wellcome Trust.

Author contributions

P.A.K. and T.D.W. designed the experiment and drafted the manuscript. P.A.K. conducted data analysis. P.A.K., T.E.N., and T.D.W. developed simulated experiments for evaluating statistical procedures. A.R., B.L.B., M.C., C.D.-M., H.G.L., E.A.R.L., L.V.O., M.K., P.D., P.J.G., S.B.M., T.D.W., and C.-W.W. contributed neuroimaging data. All authors provided feedback and revised the manuscript.

Competing interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41593-017-0051-7.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to P.A.K. or T.D.W.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations

Methods

Experimental design. We adopted a construct-validation approach to examine generalizable representations of three constructs that engage the MFC: pain, cognitive control, and negative emotion. By (i) identifying latent multivariate representations that are indicative of an underlying psychological domain rather than idiosyncrasies of one particular study or subdomain (i.e., task) and (ii) examining the similarity of those latent representations across multiple psychological domains, this framework provides a more definitive test of shared representation in the aMCC and areas within the MFC compared to conventional single-study or single-method investigations.

This approach has been taken for decades in psychometric research to assess construct validity[49]. The idea is to define a latent construct—intelligence and anxiety are classic examples—and measure it with multiple distinct indicators. Using multiple indicators permits the extraction of common factors that underlie the construct. Intercorrelations among indicators of the same construct provide evidence for convergent validity, suggesting that the indicators measure the same construct. If different indicators uniquely load on different constructs, they provide evidence for discriminant validity. Together, establishing convergence and discrimination provides strong evidence for construct validity[50].

Applied to fMRI, indicators are patterns of brain activity evoked by different tasks, and the constructs are the functional psychological domains (for example, pain or working memory) that the tasks putatively measure. The vast majority of studies, even large-scale studies like the Human Connectome Project and UK Biobank[51,52], use only one specific task variant as a single indicator for a domain. This poses a problem in inferring the similarity of psychological domains from the similarity in patterns of brain activity. For example, if a pain task differs from a negative emotion task in aMCC[25,26], is it because pain is represented differently from negative emotion, or because the particular variant of pain studied differs from the particular variant of negative emotion? A study with multiple varieties of pain and multiple varieties of negative emotion could address this question, if it showed that a brain representation common to multiple varieties of pain is distinct from a representation common to multiple varieties of negative emotion. This is the method we used in the present study, as detailed below.

Study and contrast selection. fMRI data were sampled from studies of acute thermal somatic stimulation[53,54], acute visceral stimulation[55,56], acute mechanical somatic stimulation, working memory[57,58], response selection[59,60], response conflict[61], induction of negative emotion using images of visual scenes[62,63], social rejection[64], the perception of others in pain[25], and emotionally aversive vignettes from the International Affective Digital Sounds system[65]. Together these data formed a balanced hierarchical sample, with six pain studies (two thermal, two visceral, and two mechanical), six cognitive control studies (two working memory, two response selection, and two response conflict), and six negative emotion studies (two visual, two social, and two auditory). Although negative emotion can be evoked through diverse methods, including the recollection of emotional events, the brain activity we analyze here is focused on exteroceptive processing, which has reliably been linked to overlapping MFC activity[17].

Due to variability in sample size across studies (range = 15–183), data were randomly subsampled by selecting 15 participants from each study (total $n = 270$). Although no statistical methods were used to predetermine this sample size, it is similar to those reported in previous publications[7,9,11] (see also Supplementary Table 7 for maximal univariate effects in the present sample). Because our focus was to generalize across studies, no attempts to replicate individual experiments were made. No participants were excluded from the analysis. A subset of these data was previously used to validate the use of automated meta-analysis to decode cognitive states of pain, emotion, and working memory[63].

This was not a randomized study; it was a meta-analysis of multiple studies. Participants were recruited independently for each of the 18 studies being analyzed. Data collection was conducted blind to the goals of the present study. A posteriori group assignment was based on the goals of each study and the experimental manipulation being used (for example, studies involving thermal stimulation of the forearm were considered members of the pain domain). Data analysis was not performed blind to the conditions of the experiments.

Informed consent was provided by all subjects in accordance with local ethics and institutional review boards. Participants sampled from studies 5 ($n = 15$, 4 female; $M_{age} = 26.9$), 6 ($n = 15$, 8 female; $M_{age} = 24.2$), 17 ($n = 15$, 7 female; $M_{age} = 31.1$), and 18 ($n = 15$, 9 female; $M_{age} = 24.4$) provided informed consent as approved by the University of Colorado Boulder institutional review board. Participants in study 11 ($n = 21$, 9 female; $M_{age} = 30.5$) provided informed consent in accordance with the New York University institutional review board. Descriptions of ethics approvals, image acquisition and analysis, and demographics are described briefly in Supplementary Table 7 and in full detail in the corresponding references (see also the Life Sciences Reporting Summary).

Contrasts from thermal painful stimulation were between high and low levels of pain[54] or between high levels of painful stimulation and baseline[53]. Contrasts for visceral stimulation studies were between rectal distension trials and baseline. For mechanical stimulation studies, contrasts were made between pressure application to the thumb and baseline. Contrasts for both working memory studies were between *N*-back blocks and a fixation baseline. For response-selection studies, contrasts were between trials in a go/no-go task engaging response selection (as defined in the Cognitive Atlas[66]) against baseline. Response-conflict contrasts were made between congruent and incongruent trials in studies using the Eriksen Flanker and Simon tasks. Studies of visual negative emotion compared negative to neutral IAPS pictures[63] or negative pictures against baseline[62]. Social negative emotion studies compared viewing pictures of ex-partners versus friends[64] and viewing images of others in pain[25] versus baseline. Auditory negative emotion studies compared listening to unpleasant affective sounds to baseline.

fMRI analysis. We employed three converging methods to isolate generalizable brain representations: (i) we modeled how dissimilar patterns of brain activity were from another, called representational similarity analysis (RSA[28,67]); (ii) we directly compared the similarity of brain activity in studies coming from the same domain (but different studies and subdomains) to the similarity of brain activity in studies from different domains; and (iii) we used partial-least-squares regression[68] to characterize the spatial profile of generalizable brain representations in MFC.

Feature selection. Input data were defined a priori as voxels located within the anterior midline, defined as voxels within cingulate cortex and superior frontal gyrus (in the LONI Probabilistic Brain Atlas[69]), anterior to the plane $y = -22$ mm in MNI space. The primary analyses were conducted within four cingulate subregions[27], as well as ventral and dorsal aspects of medial frontal cortex. These anatomically defined regions of interest include posterior midcingulate cortex (pMCC; $y = -22$ mm to $y = 4.5$ mm, $z > 5$ mm, 822 voxels), anterior midcingulate cortex (aMCC; $y = 4.5$ mm to $y = 30$ mm, $z > 5$ mm, 815 voxels), perigenual anterior cingulate cortex (pACC; $y > 30$ mm, $z < 5$ mm, 794 voxels), and subgenual anterior cingulate cortex (sgACC; $y = 4.5$ mm to $y = 30$ mm, $z < 5$ mm, 302 voxels), as well as within superior frontal gyrus and ventromedial prefrontal cortex (split by the plane $z = 5$ mm, dMFC and vmPFC, 4,311 and 10,619 voxels respectively). As these divisions were originally defined in Talairach space, they were converted to MNI-152 space using the Lancaster transform[70]. A secondary analysis was performed using searchlight mapping[40], in which multiple analyses were conducted, each using patterns of fMRI activation within spherical regions (radius = 8 mm) centered at every voxel in the MFC as input. An additional exploratory analysis using RSA-based model comparison, described in full detail below, was conducted using a whole-brain parcellation based on structural and functional connectivity[38].

RSA: model specification. We estimated representational dissimilarity matrices (RDMs) by computing the correlation distance (1–Pearson's *r*, excluding on-diagonal elements, which have a dissimilarity value of 0) of multivoxel patterns of brain activity. Each pattern was acquired from one of 270 subjects drawn from the full sample of 18 studies ($n = 15$ per study). Next, we constructed model-based RDMs to characterize different components of a psychological hierarchy (Fig. 1). At the lowest level, the 18 studies were individually modeled to account for study-specific idiosyncrasies. Next, the nine subdomains (visceral stimulation, thermal stimulation, mechanical stimulation, response conflict, stop/go response selection, working memory, visual negative emotion, social negative emotion, and auditory negative emotion) were modeled to account for response patterns that generalize across studies. Finally, the three psychological domains (pain, cognitive control, and negative emotion) were modeled as independent predictors to account for response patterns that generalize not only across studies, but across subdomains as well, hence being generalizable within but not across the three domains. Individual RDMs were computed from binary vectors indicating membership based on study (18 RDMs), subdomain (9 RDMs), or psychological domain (3 RDMs). The unique off-diagonal elements (intersubject dissimilarities) of these 30 RDMs, in addition to a constant RDM, were vectorized to form regressors in a model. Linear regression was used with this model to fit the observed intersubject brain dissimilarity matrix. On-diagonal elements were excluded for all similarity-based analyses, as they have perfect correlations and zero dissimilarity. The general linear model assumes independence while dissimilarity matrices exhibit complex dependence; as a result we use bootstrap inference to obtain *P* values (see "Inferences on RSA model parameters" section below).

RSA: model properties and diagnostics. To assess the suitability of our models, we conducted simulated experiments using resting-state fMRI (rsfMRI) data ($n = 270$) from the 1,000 Functional Connectomes project[71]. This allowed us to test the RSA models' false positive rates, as well as the biases and variances of parameter estimates, using data with no true effects but real fMRI noise. This is the approach recently taken by Eklund et al.[72] to assess false positive rates in standard GLM analyses.

We used a Monte Carlo procedure to conduct 1,000 RSA-based analyses of rsfMRI data, each with randomly generated event-related fMRI models for the 270 participants, allowing us to estimate the distribution of RSA parameter estimates under the null hypothesis. To mirror the dependence structure (including study/site effects) in our present sample, we selected rsfMRI participants from 18 different sites, sampling 15 participants from each site (total $n = 270$). All data were subjected to standard preprocessing, including realignment to correct for motion, nonlinear warping to standard MNI space, spatial smoothing (4-mm FWHM), and high-pass filtering (128-s cutoff). Then, for each Monte Carlo iteration, for

each participant, we generated a series of 10 events with random onsets, modeled them with Dirac delta (impulse response) functions and convolved them with SPM′s standard hemodynamic response function to generate a single regressor of interest for each subject (a constant term was also included). Then, we estimated these models, generating a series of 270 event-related activation maps. These were subjected to RSA-based modeling, as described in the model specification section above. To estimate the false positive rates for RSA model parameter estimates in bootstrap-based inference, we conducted bootstrap resampling ($b=200$ bootstrap samples) and obtained P values for each RSA model parameter. We repeated this entire procedure 1,000 times, with randomly specified models on every iteration, i.e., estimating a total of 270,000 unique fMRI activation maps and 1,000 RSA model fits and bootstrap tests.

Under the assumption that the rsfMRI data contain no consistent relationship with the randomly specified models, an unbiased model should have parameter estimates centered on zero across the 1,000 Monte Carlo iterations, indicating no systematic bias. Thus, to estimate bias, we calculated the mean deviation of each parameter estimate from zero. We also estimated the variance of parameter estimates across the 1,000 iterations; lower variance indicates greater precision and power. To estimate variance, we calculated the s.d. in each RSA parameter estimate across the 1,000 iterations. The proportion of false positives was assessed using a $P<0.05$ cutoff (two-tailed). False positives were defined as parameter estimates below the 2.5th or above the 97.5th percentile on the bootstrap distribution, and the false positive rate for each RSA model parameter estimate was defined as the proportion of the 1,000 RSA models for which that parameter estimate was significant.

The results of these analyses indicated that the modeling procedure is unbiased, i.e., average null-hypothesis values were nearly exactly zero. Of 180 parameters evaluated (6 ROIs × 30 parameters) the largest effect was not significantly different from zero ($z=0.904$, $P=0.366$). In addition, false-positive rates for all RSA model parameters were at or below the nominal value of 0.05 (Supplementary Fig. 5).

We also repeated the entire RSA model simulation (500 iterations) using synthetic null-hypothesis data generated from a Wishart distribution (Supplementary Fig. 6) and with a homogeneous set of task data (using 180 subjects from study 13) and found qualitatively identical results (not shown). This simulated a case in which there are no study-level effects. Overall, the RSA model procedures provide unbiased estimates under the null hypothesis, and false positive rates are appropriately controlled.

RSA: model identifiability. Here we model intersubject RDMs, so the data to be modeled comprise an $n\times n$ dissimilarity matrix, where n is the number of participants; this matrix has rank $\min(n, r)$ where r is the number of voxels used to compute the correlation. The lower triangle of the dissimilarity matrix is vectorized and fit with a linear model, with design matrices based on equivalently vectorized dissimilarity (as illustrated in Fig. 1). Thus, the outcome data is a vector of length $u=n\times(n-1)/2$. The model dimension depends on the exact parameterization used, but a saturated model for the effect of study would have dimension $k+k\times(k-1)/2$, where k is the number of studies; this is k parameters for the average intrastudy relationship (for pairs of subjects) within each study, and $k\times(k-1)/2$ parameters for each possible relationship between each pair of studies. In practice, we use a much simpler model, but confirm identifiability by checking the rank of the design matrix and variance inflation factors.

In the present work, we included $k=18$ studies with $n=270$ participants divided equally among them (15 per study). We used the constant term in the model to characterize the average similarity of data within studies, leaving a subspace of interstudy relationships spanned by $k+k\times(k-1)/2-1=170$ dimensions. In principle, the upper bound on the number of regressors for interstudy differences in an identifiable model is 170. However, we were primarily interested in specific interstudy relationships, particularly those common to subdomains (9 parameters for pairs of studies with the same subdomain) and those common to domains (three parameters for three sets of 9 studies that load on the same domain construct across three subtypes). Our full model thus contained 31 regressors: 18 for specific studies, 9 for subdomains, three for domains, and one intercept term. As the number of voxels per ROI was at least $302>31$, there was no risk of degenerate (zero-residual) models.

RSA: model rank and variance inflation factors. To confirm the identifiability and efficiency of our models, we computed the rank of our design matrix and variance inflation factors (VIFs) for each regressor. The VIFs show the degree to which variance in each parameter estimate is increased due to partial colinearity with linear combinations of other regressors. The rank of our design matrix is 31, indicating that the model is full-rank, identifiable, and not overparameterized. VIFs were finite for all regressors, consistent with the fact that model was identifiable (Supplementary Fig. 2). VIFs varied across regressors based on the partially shared variance across study, domain, and subdomain, which was unavoidable because part of the covariance common to participants in each domain is shared with the subdomains and studies that fall within it. However, the VIFs were clearly in a range that indicated reasonable ability to make inferences on the unique variance explained by each parameter, given the sample size. The regressors that were of primary interest (i.e., the three domain level terms) had VIFs = 1.66.

Statistics: inferences on RSA model parameters. Parameter estimates ($\widehat{\beta}$) from the model provide estimates of generalizability, with the interpretation of a significant (nonzero) parameter estimate depending on the nature of the regressor. Study-specific regressors test generalizability across individual participants within a study, in the sense that they capture intersubject correlations in the spatial patterns of activity. Positive values indicate similar patterns across participants for the study modeled. Subdomain-level regressors test generalizability across two studies of the same subdomain, controlling for other model parameters. Positive values indicate shared spatial patterns across studies of the modeled subdomain and, thus, evidence for a coherent subdomain. Domain-level regressors, which were of primary interest here, test generalizability across three distinct subdomains (six studies), controlling for shared patterns specific to the study and subdomain. Positive values indicate shared spatial patterns across subdomains and, thus, evidence for a coherent domain-related pattern. Thus, we refer to these parameter estimates as generalization indices (for example, Fig. 2), as they reflect the extent to which patterns of brain activation generalize across subdomains, studies, or subjects. Because the regressors were tested jointly in multiple regression modeling, significant domain-level parameter estimates imply that the pattern of brain activity shared across the domain was not reducible to a subdomain-specific or study-specific pattern.

Inference on parameter estimates ($\widehat{\beta}$) was made using bootstrap resampling of subjects. This procedure involved repeatedly resampling subjects, with replacement, for each study over 5,000 iterations for the regional analysis and over 1,000 iterations for the searchlight analysis (see below). In each resampling, a new RDM was constructed using fMRI activation for the resampled subjects and GLMs were estimated. Because samples from the same subject can be drawn multiple times in this approach, pairwise dissimilarity values from the same subject (with a dissimilarity value of 0) were excluded from the analysis. Bootstrap distributions for individual model parameters were compared against 0 using normal approximation for inference. These distributions were visually inspected and assumed to be normal, although this was not statistically tested. Unless noted otherwise, the main results reported in the manuscript (both regional and searchlight analyses) were thresholded after correcting for multiple comparisons based on the false discovery rate (FDR corrected, $q<0.05$). All tests are two-tailed unless otherwise specified.

Statistics: RSA model comparison. To formally compare the amount of evidence for domain-generalizable representations, model comparisons were made using the Bayesian information criterion (BIC). The reference model included terms for each study (18 parameters) and each subdomain (9 parameters), as well as a constant (28 parameters in total). Next, we fit three models that each included a single additional term for one of the psychological domains (29 parameters in total). Finally, a more fully specified model that contained all three psychological domains (31 parameters in total) was fit. These five models were fit for each region of interest and the full extent of MFC. BIC values were computed using the log-likelihood of fitted models and penalizing based on the number of free parameters[73]. The number of samples was set the number of participants included in the analyses (270) as opposed to the number of unique elements in the dissimilarity matrix, because of dependence between elements of the dissimilarity matrix. BIC values were converted to weights using the formulation in Wagenmakers and Farrell[74]. These weights characterize which model is most likely to have produced the observed similarity structure in each region, given the a priori set of models. Finally, the adjusted R^2 was computed for the model favored by the BIC analysis in each region of interest.

This analysis was additionally conducted for regions spanning the whole brain (as delineated in the Brainnetome Atlas[38]). Because this parcellation contains regions with fewer voxels than participants included in the analysis, the number of samples was set to the minimum of 270 and the number of voxels in each parcel.

Statistics: model-free analysis comparing spatial correlations within and between domains. To provide evidence for generalizable brain representations without using an explicit model of expected similarity relations, we tested for nonzero correlations within domains and between different subdomains, within domains and between studies, and within domains generally. Tests of nonzero correlations were performed by constructing confidence intervals ($\alpha=0.05$) using bootstrap analyses with the bias corrected and accelerated percentile method. We also conducted a series of hierarchical tests comparing: (i) intersubject correlations from the same domain but different subdomains vs. correlations from different domains; (ii) correlations from the same domain but different studies vs. correlations from the same domain but different subdomains; and (iii) correlations from the same domain vs. correlations from the same domain versus different studies. These tests were constructed to sequentially identify the average effect of domain, subdomain, and study for each of the three domains. Inferences were drawn by calculating bootstrap confidence intervals based on the mean difference in correlation coefficients (for a graphical depiction, see Supplementary Fig. 7).

Statistics: PLS estimation of latent patterns for each domain. To estimate patterns of activity associated with each domain (Fig. 2), partial-least-squares regressions[70] were run separately in each region of interest, with contrasts from

all 270 subjects forming the data matrix and dummy coded variables forming the output matrix (270 subjects by 30 parameters: 18 studies, 9 subdomains, and 3 domains, with values of +1/−1 based on inclusion/exclusion for each term). Parameter estimates were bootstrapped over 5,000 iterations, and z-scores were estimated based on the mean and standard error of the bootstrap distributions. We note that this approach is not designed to ensure that representations are uniquely specific to each domain; if some studies or subdomains have especially high covariance with a domain, they could have a large influence on domain-level patterns. For this reason, PLS-based estimation of patterns is used in conjunction with RSA and direct comparisons of intersubject correlations.

Statistics: searchlight thresholds and assessment of overlap. For the conjunction analysis and visualization of searchlight maps, an uncorrected threshold ($P<0.05$) was used to ensure that an overly conservative threshold did not obscure overlapping regions. To estimate the relative evidence for and against overlap of the domains, Bayes factors were computed[44] using the minimum statistic compared to the conjunction null[75] and a uniform distribution ranging from 0 to 10 as a prior distribution (theoretically plausible values of parameters estimates). This test evaluates whether there is more evidence in favor of overlap (i.e., that the minimum statistic of the three maps is substantially greater than zero) or against overlap (the minimum statistic of the three maps is relatively close to or less than zero). Bayes factors >3 provide evidence of overlapping representations, whereas values less than 0.33 provide evidence against overlap.

Correspondence of searchlight maps with existing parcellations. River plots were created to depict the correspondence (quantified as the cosine similarity) between the searchlight maps and existing anatomical[27] and functional[48] parcellations.

Life Sciences Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Data availability. The fMRI data for studies 9–12 are available from OpenfMRI: https://openfmri.org/dataset/ds000008/, https://openfmri.org/dataset/ds000007/, https://openfmri.org/dataset/ds000101/, and https://openfmri.org/dataset/ds000102/. fMRI data for study 13 is available at NeuroVault, http://neurovault.org/collections/503. The fMRI data that support the findings of this study are available for download at https://canlabweb.colorado.edu/files/MFC_Generalizability.tar.gz.

Code availability. Matlab code for implementing all analyses is available at https://github.com/canlab/ and in the Supplementary Software.

References

49. Cronbach, L. J. & Meehl, P. E. Construct validity in psychological tests. *Psychol. Bull.* **52**, 281–302 (1955).
50. Campbell, D. T. & Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.* **56**, 81–105 (1959).
51. Barch, D. M. et al. Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage* **80**, 169–189 (2013).
52. Miller, K. L. et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* **19**, 1523–1536 (2016).
53. Wager, T. D. et al. An fMRI-based neurologic signature of physical pain. *N. Engl. J. Med.* **368**, 1388–1397 (2013).
54. Atlas, L. Y., Bolger, N., Lindquist, M. A. & Wager, T. D. Brain mediators of predictive cue effects on perceived pain. *J. Neurosci.* **30**, 12964–12977 (2010).
55. Rubio, A. et al. Uncertainty in anticipation of uncomfortable rectal distension is modulated by the autonomic nervous system-a fMRI study in healthy volunteers. *Neuroimage* **107**, 10–22 (2015).
56. Kano, M. et al. Influence of uncertain anticipation on brain responses to aversive rectal distension in patients with irritable bowel syndrome. *Psychosom. Med*. https://doi.org/10.1097/PSY.0000000000000484 (2017).
57. DeYoung, C. G., Shamosh, N. A., Green, A. E., Braver, T. S. & Gray, J. R. Intellect as distinct from Openness: differences revealed by fMRI of working memory. *J. Pers. Soc. Psychol.* **97**, 883–892 (2009).
58. van Ast, V. A. et al. Brain mechanisms of social threat effects on working memory. *Cereb. Cortex* **26**, 544–556 (2016).
59. Xue, G., Aron, A. R. & Poldrack, R. A. Common neural substrates for inhibition of spoken and manual responses. *Cereb. Cortex* **18**, 1923–1932 (2008).
60. Aron, A. R., Behrens, T. E., Smith, S., Frank, M. J. & Poldrack, R. A. Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (MRI) and functional MRI. *J. Neurosci.* **27**, 3743–3752 (2007).
61. Kelly, A. M. C., Uddin, L. Q., Biswal, B. B., Castellanos, F. X. & Milham, M. P. Competition between functional brain networks mediates behavioral variability. *Neuroimage* **39**, 527–537 (2008).
62. Gianaros, P. J. et al. An inflammatory pathway links atherosclerotic cardiovascular disease risk to neural activity evoked by the cognitive regulation of emotion. *Biol. Psychiatry* **75**, 738–745 (2014).
63. Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C. & Wager, T. D. Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* **8**, 665–670 (2011).
64. Kross, E., Berman, M. G., Mischel, W., Smith, E. E. & Wager, T. D. Social rejection shares somatosensory representations with physical pain. *Proc. Natl Acad. Sci. USA* **108**, 6270–6275 (2011).
65. Bradley, M. M. & Lang, P. J. *The International Affective Digitized Sounds (IADS-2): Affective Ratings of Sounds and Instruction Manual*. (University of Florida, Gainesville, FL, 2007). Tech. Rep. B-3.
66. Poldrack, R. A. et al. The cognitive atlas: toward a knowledge foundation for cognitive neuroscience. *Front. Neuroinform.* **5**, 17 (2011).
67. Kriegeskorte, N. & Kievit, R. A. Representational geometry: integrating cognition, computation, and the brain. *Trends Cogn. Sci.* **17**, 401–412 (2013).
68. Wold, S., Sjostrom, M. & Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometr. Intell. Lab. Syst.* **58**, 109–130 (2001).
69. Shattuck, D. W. et al. Construction of a 3D probabilistic atlas of human cortical structures. *Neuroimage* **39**, 1064–1080 (2008).
70. Lancaster, J. L. et al. Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Hum. Brain Mapp.* **28**, 1194–1205 (2007).
71. Biswal, B. B. et al. Toward discovery science of human brain function. *Proc. Natl Acad. Sci. USA* **107**, 4734–4739 (2010).
72. Eklund, A., Nichols, T. E. & Knutsson, H. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proc. Natl Acad. Sci. USA* **113**, 7900–7905 (2016).
73. Schwarz, G. Estimating dimension of a model. *Ann. Stat.* **6**, 461–464 (1978).
74. Wagenmakers, E. J. & Farrell, S. AIC model selection using Akaike weights. *Psychon. Bull. Rev.* **11**, 192–196 (2004).
75. Nichols, T., Brett, M., Andersson, J., Wager, T. & Poline, J. B. Valid conjunction inference with the minimum statistic. *Neuroimage* **25**, 653–660 (2005).

nature research

Corresponding author(s): Wager, Kragel

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.	Because data from multiple datasets were combined in a multi-study framework, balanced subsampling was performed (n = 15 per study) to help equate statistical power for comparisons between studies.

2. Data exclusions

Describe any data exclusions.	No data were excluded from the analysis.

3. Replication

Describe whether the experimental findings were reliably reproduced.	Although replication was not attempted, the main findings reflect effects that are reliable across 6 different studies.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.	This was not a randomized study, it was a mega-analysis of multiple studies. Contrasts were constructed within subjects and the goal was to compare effects across studies. Participants were recruited independently for each of the 18 studies being analyzed. A posteriori group assignment was based on the goals of each study and experimental manipulation being used (e.g., studies involving thermal stimulation of the forearm were considered members of the 'pain' domain).

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.	Investigators were not aware of group comparisons during data collection. No blinding was performed for data analysis.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	Analysis was conducted using CANLab and SPM software implemented in MATLAB. Code for implementing all analyses is available at https://github.com/canlab/ and www.fil.ion.ucl.ac.uk/spm/software/.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	No unique materials were used.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	No eukaryotic cell lines were used.
b. Describe the method of cell line authentication used.	No eukaryotic cell lines were used.
c. Report whether the cell lines were tested for mycoplasma contamination.	No eukaryotic cell lines were used.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No commonly misidentified cell lines were used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	Non-human animals were not used.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	270 healthy human participants were involved in the research. The age and gender of participants in each of the 18 studies are detailed in Supplementary Table 7.

nature research

Corresponding author(s): Wager, Kragel

☐ Initial submission ☐ Revised version ☒ Final submission

MRI Studies Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Experimental design

1. Describe the experimental design.	Multiple methods are used, depending on the study. Our approach aims to generalize across these methodological factors.
2. Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.	The number of trials per subject ranged from 8 to 768. Stimulus durations ranged from 2s to 18s. Scan length ranged from 5 min to 13.5 min. Details for each study are listed in Supplementary Table 7 and the original publications for each study.
3. Describe how behavioral performance was measured.	Behavioral data are not presented.

Acquisition

4. Imaging	
a. Specify the type(s) of imaging.	BOLD fMRI
b. Specify the field strength (in Tesla).	1.5 and 3 Tesla
c. Provide the essential sequence imaging parameters.	EPI (standard and multiband) and spiral in-out sequences were used for data acquisition. Readers are referred to the original publications for more details.
d. For diffusion MRI, provide full details of imaging parameters.	Diffusion MRI was not collected.
5. State area of acquisition.	Whole brain scans were used.

Preprocessing

6. Describe the software used for preprocessing.	SPM and custom code were used for preprocessing, the particular version depending on the study. Readers are referred to the original publications for more details.
7. Normalization	
a. If data were normalized/standardized, describe the approach(es).	Non-linear normalization to MNI space was performed.
b. Describe the template used for normalization/transformation.	The templates used depend on the study, but all are in ICBM152 space.
8. Describe your procedure for artifact and structured noise removal.	Regression of motion parameters was performed in all studies (either 6 parameters based on translation and rotation or 24 with the inclusion of their derivatives, successive differences, and squared successive differences).
9. Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.	Outlier timepoints were excluded in some studies. These were identified based on Mahalanobis distance using a chi-square test.

Statistical modeling & inference

10. Define your model type and settings.	Multivariate RSA models were specified and fitted using least squares regression. Subject was treated as a random effect.
11. Specify the precise effect tested.	Dissimilarity between brain activity (1 - Pearson's correlation coefficient) was modeled as a function of study, psychological subdomain, and domain.
12. Analysis	
a. Specify whether analysis is whole brain or ROI-based.	ROI-based and whole brain searchlight
b. If ROI-based, describe how anatomical locations were determined.	Existing parcellations were used for ROI-based analyses.
13. State the statistic type for inference. (See Eklund et al. 2016.)	Voxel-wise statistics were used.
14. Describe the type of correction and how it is obtained for multiple comparisons.	FDR correction was used.
15. Connectivity	
a. For functional and/or effective connectivity, report the measures of dependence used and the model details.	Connectivity analysis was not performed.
b. For graph analysis, report the dependent variable and functional connectivity measure.	Graph analysis was not performed.
16. For multivariate modeling and predictive analysis, specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.	For multivariate models, the independent variables comprised dissimilarity matrices (1 - Pearson's correlation coefficient) estimated on data from all subjects (n = 270). Feature extraction was done in three ways: using local searchlights in the medial frontal cortex (MFC), using a priori ROIs in the MFC (Vogt parcellation), and using a set of ROIs that spans the whole brain (Brainnetome atlas). No dimension reduction was performed. RSA-based models were estimated using least squares regression. Partial Least Squares regression was also used to predict psychological domains from patterns of brain activity. Inferences were made on parameter estimates using bootstrap and Monte Carlo methods.