NEUROSCIENCE

A computational mechanism of cue-stimulus integration for pain in the brain

Jungwoo Kim[1,2,3], Suhwan Gim[1,2,3], Seng Bum Michael Yoo[1,2,3,4]*†, Choong-Wan Woo[1,2,3,5]*†

The brain integrates information from pain-predictive cues and noxious inputs to construct the pain experience. Although previous studies have identified neural encodings of individual pain components, how they are integrated remains elusive. Here, using a cue-induced pain task, we examined temporal functional magnetic resonance imaging activities within the state space, where axes represent individual voxel activities. By analyzing the features of these activities at the large-scale network level, we demonstrated that overall brain networks preserve both cue and stimulus information in their respective subspaces within the state space. However, only higher-order brain networks, including limbic and default mode networks, could reconstruct the pattern of participants' reported pain by linear summation of subspace activities, providing evidence for the integration of cue and stimulus information. These results suggest a hierarchical organization of the brain for processing pain components and elucidate the mechanism for their integration underlying our pain perception.

INTRODUCTION

The brain integrates diverse neurobiological and psychological components to construct pain perception, including the activation of nociceptors (1) and the influence of expectancies (2). Previous studies have already identified multiple components that contribute to our pain experience (3–5) and how the brain encodes and mediates those components (6, 7). For example, the dorsolateral prefrontal cortex (dlPFC) is involved in pain-related attention (8, 9), the ventromedial PFC and nucleus accumbens in cognitive appraisal (10), the somatosensory cortex and thalamus in processing noxious inputs (11, 12), and the anterior cingulate cortex (ACC) in the affective component of pain (13).

Despite the progress in the functional mapping of different pain components in the brain, there are still important questions to be answered. First, the neuro-computational mechanisms underlying the integration of multiple pain components are largely unknown. Here, we address this question by adopting the dynamical systems approach to formalize the mechanistic process of integration (14). Such an approach focuses on two key aspects of neuronal activities: temporal dependency and the covariability of neuronal populations. Since current neuronal activity influences subsequent activity (i.e., temporal dependencies), neuronal activities over time result in trajectories in a state space, where each coordinate axis represents the activity of an individual neuron or voxel. Because of recurrent connections among them, which result in covariability, these trajectories do not evolve randomly across the entire state space but are confined to specific, lower-dimensional subspaces relevant to the task at hand (15–17). The characterization of neuronal trajectories and subspaces can provide a mechanistic explanation of how the brain processes information (18–22), going beyond the description of where and what information is encoded or decoded. Second, previous studies

have suggested that information progressively transforms from sensory (i.e., unimodal) to integrative forms (i.e., transmodal) along the cortical hierarchy (23) in other domains, such as memory and narratives (24, 25). However, it remains unclear whether such a hierarchical organization also exists for pain processing. Thus, we investigate the integration of cue and stimulus information within each large-scale network in the hierarchy, including unimodal and transmodal brain regions, which represent the two extremes of the cortical hierarchy.

The central hypotheses of our study include: 1) The features of neural trajectories within each large-scale functional brain network underlie the computational process of information integration in pain, and 2) the neural evidence for this integration will be more pronounced within transmodal brain regions, such as the default mode and limbic networks, as opposed to unimodal brain regions, such as the visual and somatomotor networks. To test these, we used a cue-induced pain modulation task along with functional magnetic resonance imaging (fMRI), allowing us to examine the computational and organizational principles underlying the pain integration process at the whole brain level. To effectively integrate the dynamical systems framework into our analysis pipeline, we collected fMRI data with a high sampling rate, i.e., a short repetition time (TR = 460 ms), with simultaneous multi-slice imaging. We analyzed the data while human participants ($N = 56$) received noxious thermal stimulations following pain-predictive cues. To calculate subspaces for each cue and stimulus information, we used semisupervised linear dimensionality reduction (26, 27) within each large-scale network (28–30). Projecting the fMRI activities into cue and stimulus subspaces resulted in neuronal trajectories of experimental pain conditions (i.e., different levels of cue and stimulus) within each subspace. By using intercondition distances between these subspace activities, we examined the preservation of cue and stimulus information in their respective subspaces.

Our results show that both unimodal and transmodal brain regions preserve the cue and stimulus information for pain, while only transmodal brain networks, including the limbic and default mode networks, integrate this information. This integration is accomplished through the linear summation of the distances between neural trajectories from each subspace. These findings offer insights on the computational and organizational principles underlying the pain integration process.

[1]Center for Neuroscience Imaging Research, Institute for Basic Science, Suwon, South Korea. [2]Department of Biomedical Engineering, Sungkyunkwan University, Suwon, South Korea. [3]Department of Intelligent Precision Healthcare Convergence, Sungkyunkwan University, Suwon, South Korea. [4]Department of Neurosurgery and McNair Scholar Program, Baylor College of Medicine, Houston, TX 77030, USA. [5]Life-inspired Neural Network for Prediction and Optimization Research Group, Suwon, South Korea.
*Corresponding author. Email: sbyoo@g.skku.edu (S.B.M.Y); waniwoo@skku.edu (C.-W.W.)
†These authors contributed equally to this work.

RESULTS

Behavioral evidence for cue-stimulus integration in pain

In the experiment, we presented participants with a 2-s visual pain-predictive cue and instructed them that the presented cues correspond to pain intensity ratings from 25 participants who previously participated in the experiment. After a few seconds of delay, we administered heat stimulation to the participants' left forearm for 12.5 s. Participants then rated their pain intensity ratings on a semicircular scale ranging from 0° (no sensation) to 180° (strongest imaginable sensation) (Fig. 1A). We delivered five temperature levels that were calibrated for each participant to elicit comparable pain intensities (see Materials and Methods for the calibration process). We had three cue conditions: "low pain," "high pain," and "no cue" conditions. The low-pain cue was associated with stimulus levels 1 to 4, the high-pain cue with levels 2 to 5, and the no cue with levels 2 to 4, resulting in a total of 11 distinct trial types (Fig. 1B, top right). If cue-stimulus integration occurs, varying cue conditions should lead to different levels of pain ratings even for identical stimulus intensities (Fig. 1B, bottom right). Our behavioral data supported the cue-stimulus integration (Fig. 1C), as evidenced by the significant effects of both the cue and stimulus intensity on the pain ratings in the multilevel generalized linear model (GLM). In the model, we assumed cue and stimulus effects as fixed effects and subject-level differences as random effects. The resulting statistics are as follows: cue: $\hat{\beta} = 0.107$, SEM = 0.008, $z = 3.879$, $P = 1 \times 10^{-4}$, two-tailed, bootstrap test; stimulus: $\hat{\beta} = 0.107$, SEM = 0.005, $z = 3.631$, $P = 3 \times 10^{-4}$; interaction: $\hat{\beta} = 0.006$, SEM = 0.003, $z = 1.808$, $P = 0.071$.

Computational and organizational hypotheses about pain integration

Our first hypothesis, which proposes that the features of neural trajectories underlie the computational process of information integration in pain perception, can be considered in two steps: 1) There are distinct neural subspaces for encoding cue and stimulus intensity information, and 2) the geometric features of neural trajectories within each subspace account for information preservation and integration. These assumptions are grounded in previous literature that suggested subspace as an encoding dimension for experimental variables (31–34) and how the geometry of representations underlies cognitive processes (19, 35–37).

To test these hypotheses, we first identified distinct subspaces by applying a semisupervised linear dimensionality reduction method (26, 27) to fMRI time-series data from the heat stimulation period. This method involves factorizing the cue and stimulus information from the temporal estimates of fMRI activities by GLM (Fig. 2A) and then applying principal components analysis (PCA) on the factorized information (Fig. 2B). We conducted this analysis for each large-scale functional brain network separately, resulting in two subspaces for each network. We assumed that these subspaces encode the information for either the cue or the stimulus. Subsequently, we projected the fMRI time-series data onto these two subspaces. In each subspace, we used the inter-trajectory Euclidean distances for each time point across experimental conditions to calculate encoding performances, which is a metric indicating how well the subspace preserves (or encodes) the information over time (for details, see Materials and Methods and fig. S1). Encoding performances were calculated as R-squared from the GLM analysis
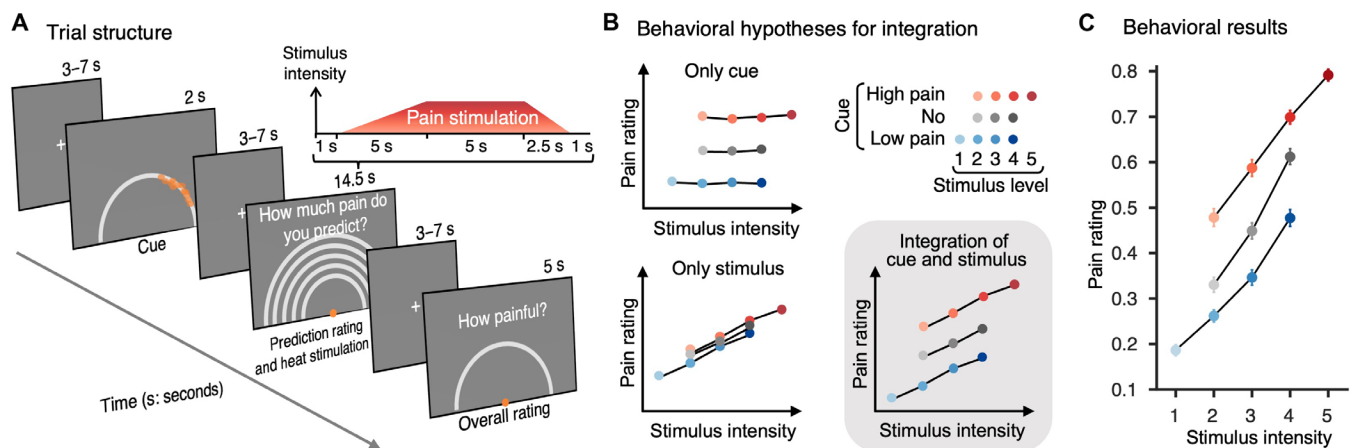
Fig. 1. Trial structure and behavioral results. (**A**) The trial structure of the cue-induced pain modulation task. The gray semicircle represents the rating scale ranging from 0° (no sensation) to 180° (strongest imaginable sensation). The distribution of orange dots during the cue period differed across the two cue conditions, for high-pain cue and for low-pain cue (see Materials and Methods for detail). Participants were informed that the dots were ratings from participants who had experienced the same intensity stimuli in other experiments. After a jittered delay, participants received heat stimulation while providing continuous ratings on their ongoing prediction of pain intensity, which were not included in the current study (for prediction rating in the figure, see Materials and Methods). After the heat stimulation, the participants rated the overall pain intensity for the trial. (**B**) Hypothetical plots for the cue and stimulus effects on pain ratings. The plot in the top left panel shows the case where only cue information is processed, whereas the plot in the bottom left panel shows when only stimulus information is processed. The plot in the bottom right panel shows the case where both cue and stimulus information are processed. Different colors indicate different cue and stimulus intensity conditions. There were five levels of stimulus intensities that were calibrated for each participant before fMRI scans (see Materials and Methods for detail). (**C**) Behavioral results. The plot shows average pain ratings across participants (N = 56) for different cue and stimulus conditions. The error bars indicate the SEM across participants. The pattern of behavioral ratings clearly shows the evidence for the integration of cue and stimulus information. All the behavioral results presented in the current study used the average of the last 1 s of the overall pain rating period.
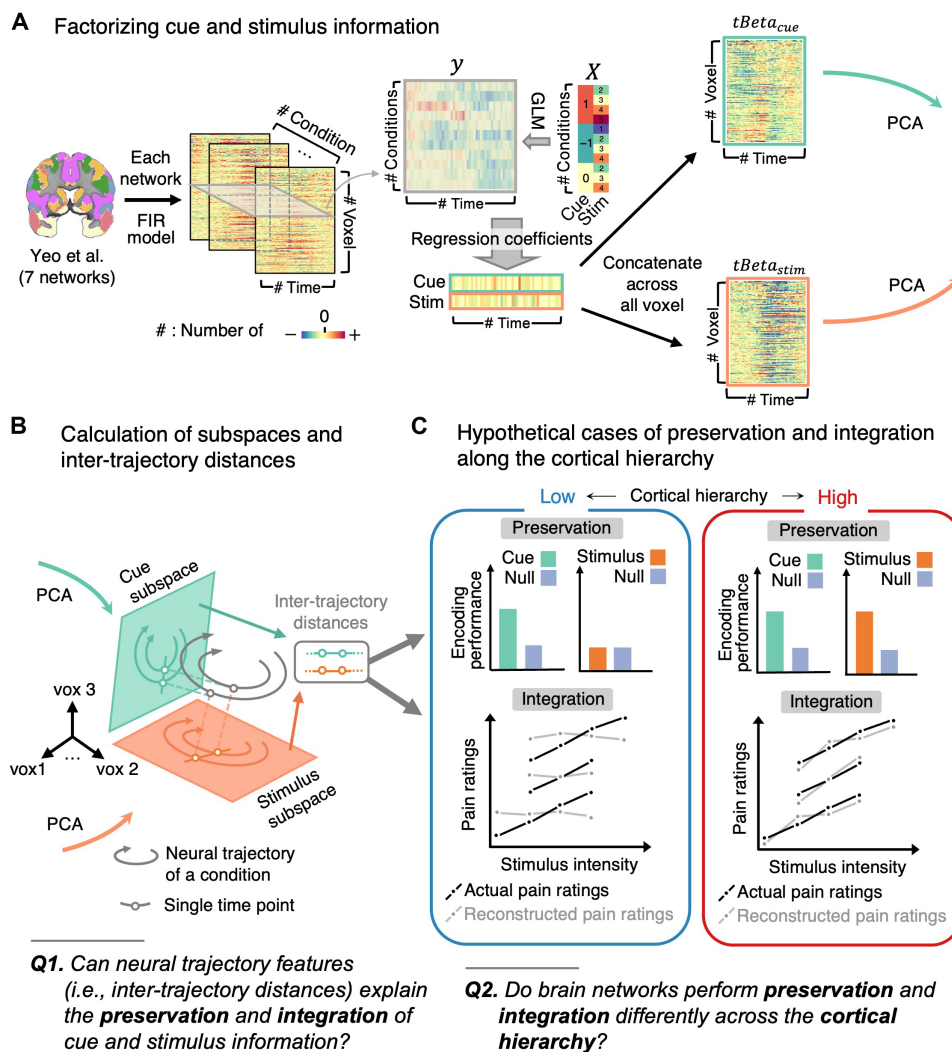
Fig. 2. Schematic of the data analysis workflow. (**A**) The schematic illustrates the process of factorizing cue and stimulus information. For each large-scale network, we estimated the temporal fMRI activities using the finite impulse response (FIR) model at the voxel level. Then, for each voxel, we performed GLM with between-subject average FIR values as the dependent variable (gray-shaded box in the figure) and cue and stimulus information as the independent variables. Specifically, we assigned values of −1, 0, and 1 to represent low cue, no cue, and high cue, respectively, and values from 1 to 5 to represent stimulus levels 1 to 5. We concatenated the resulting regression coefficients for each cue (denoted as $tBeta_{cue}$) and stimulus (denoted as $tBeta_{stim}$) across all voxels. We consider these the factorized information of the respective cue and stimulus information within each network. (**B**) Using factorized cue and stimulus information ($tBeta_{cue}$ and $tBeta_{stim}$), we performed PCA to obtain cue and stimulus subspace. We projected FIR estimates on each subspace, resulting in condition-wise temporal trajectories (colored curves in each subspace). We used inter-condition trajectory distances to calculate encoding performances for assessing information preservation (see Materials and Methods and fig. S1) and to reconstruct pain ratings for evaluating information integration (see Materials and Methods and Fig. 5A). (**C**) Hypothetical cases of networks preserving and integrating cue and stimulus information. If the encoding performance was greater than the performance with the null subspace, we consider the subspace to preserve the corresponding information. If the reconstructed ratings track the actual pain ratings, we consider the network to integrate both types of information. We hypothesized that the unimodal networks (left) would encode modality-specific information and fail to integrate the information, while the transmodal networks (right) would encode both types of information and also integrate them.

with inter-trajectory distance information as the dependent variable and either cue or stimulus information as the independent variable depending on the corresponding subspace. If the encoding performances of the subspace are significantly larger than those from the null subspace, we consider the subspace to encode the corresponding information (upper right panel of Fig. 2C). The null subspace was established by permuting conditions of independent variables during the GLM process in Fig. 2A. Encoding each piece of information in a subspace is a necessary but not sufficient condition for achieving information integration. To examine the integration of the cue and stimulus information, we summed the inter-trajectory Euclidean distances—the same metric used to calculate encoding performances—from both subspaces to reconstruct the actual pain ratings. We validated this linear operation by comparing its output (i.e., reconstructed ratings) with the actual pain ratings averaged across participants. If the reconstructed ratings align with the actual pain ratings, we considered the network to integrate both types of pain information (lower right panel of Fig. 2C).

Next, to investigate our second central hypothesis, which posits that neural evidence for integration would be more pronounced within transmodal brain regions, we calculated encoding performances and behavioral fits of the reconstructed pain ratings across large-scale functional networks. These large-scale networks were defined on the basis of previous studies (*28–30*) that segmented the cerebrum, cerebellum, and striatum into seven networks using resting-state fMRI data. We chose this network solution because it provides plausible boundaries of the cortical gradient that reflect hierarchical organization (*38*) and has demonstrated its validity for showing the brain's hierarchical properties in previous studies (*39–41*). We also confirmed the gradient organization within our dataset by assigning the first principal gradient scores (*23*) to each network using the participants' resting-state fMRI data (fig. S2). Our hypothesis predicts that unimodal networks, such as the visual and somatomotor networks, would selectively encode information related to either cue or stimulus, failing to reconstruct the actual pain ratings (left panel of Fig. 2C). In contrast, transmodal networks, such as the frontoparietal, limbic, and default mode networks, would preserve information from both cue and stimulus, but they also accurately reconstruct actual pain ratings (right panel of Fig. 2C).

Preservation of pain components in unimodal brain networks

We investigated whether the unimodal networks, including the visual and somatomotor networks, selectively preserve either the cue or the stimulus information using encoding performances within the subspaces of each unimodal network. In the visual network, encoding performances were higher for cue information compared to stimulus information ($t_{48} = 5.479$, $P = 1.553 \times 10^{-6}$, two-tailed, paired t test) (Fig. 3A). Conversely, in the somatomotor network, the results were reversed, with higher encoding performances for stimulus information than cue information (paired t test, $t_{48} = 8.014$, $P = 2.097 \times 10^{-10}$, two-tailed) (Fig. 3B).

When comparing the encoding performances with those from the null subspaces, the results showed significantly higher encoding performances in both cue and stimulus subspaces than those in the respective null subspace for all unimodal networks (see Materials and Methods for null space generation process). In the visual network (Fig. 3C), encoding performances in the cue and stimulus subspaces were significantly greater than those in the null subspaces (for cue, $t_{48} = 8.243$, $P = 9.466 \times 10^{-11}$, and for stimulus, $t_{48} = 5.818$, $P = 4.753 \times 10^{-7}$, two-tailed, paired t test). The somatomotor network (Fig. 3D) also showed significantly higher encoding performances in both cue ($t_{48} = 7.366$, $P = 2.022 \times 10^{-9}$, tailed, paired t test) and stimulus subspaces ($t_{48} = 7.451$, $P = 1.502 \times 10^{-9}$, two-tailed, paired t test) compared to the null subspaces. These results demonstrate that unimodal networks preserve both cue and stimulus information within their respective subspaces but with a certain degree of specificity—i.e., the visual network for the cue information and the somatomotor network for the stimulus information.

Figure 3 (E and F) shows the neural trajectories within subspaces from which the encoding performances were calculated for the visual network and somatomotor network, respectively. The actual encoding performances were calculated on the basis of the subspaces consisting of 20 dimensions, but the trajectories are displayed in
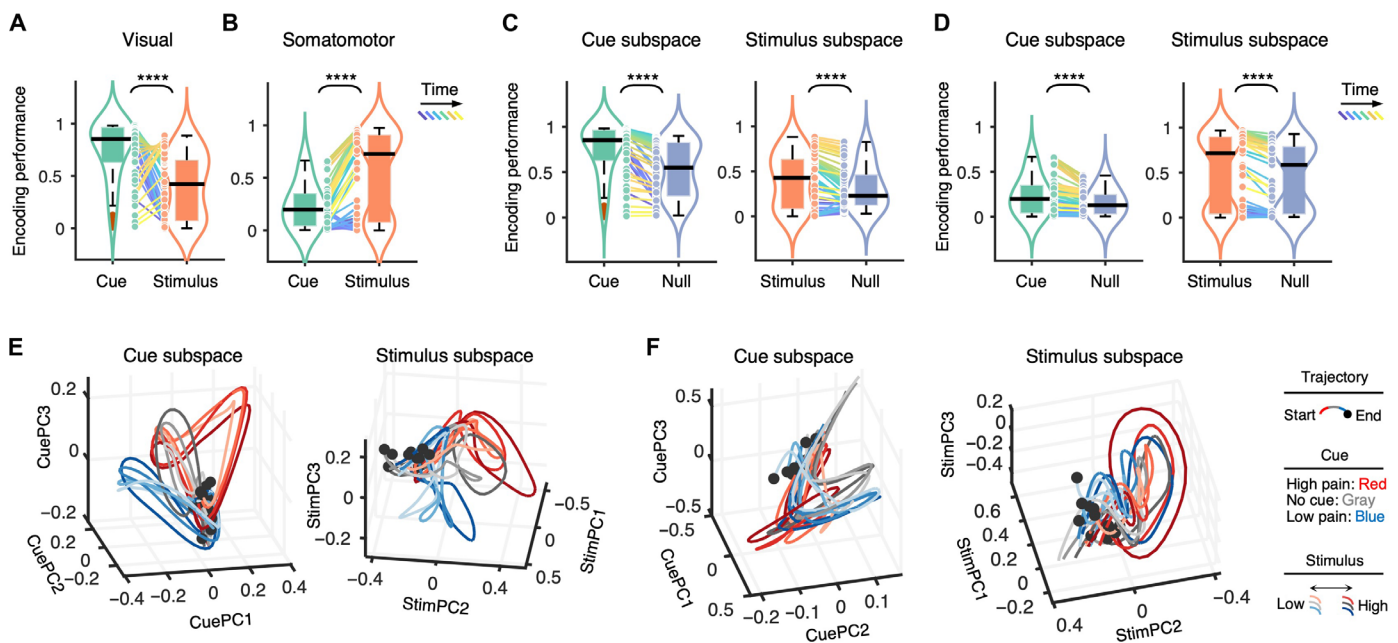
Fig. 3. Encoding performances and neural trajectories in unimodal networks. (**A** and **B**) Encoding performances of the visual (A) and somatomotor networks (B). The encoding performance was calculated for each time point, and the plots show the time information with the graded colors ranging from cool (early) to warm (late) colors. (**C** and **D**) Encoding performances based on the actual versus null subspaces for the visual network (C) and the somatomotor network (D). (**E**) Neural trajectories of the visual network. The plots show the neural trajectories of the experimental conditions within the cue (left) and the stimulus subspaces (right). The axes were the top three PCs of the temporal regression coefficients, as depicted in Fig. 2B. The trajectories were smoothed with a Gaussian kernel (SD: 3 TRs) for visualization purposes. The unsmoothed trajectories were used for the actual analysis. The trajectories' colors (cue levels) and brightness (stimulus intensity levels) indicate different experimental conditions. The trajectory's starting point aligns with the onset of pain stimulation, while its ending point corresponds to 10 s after the pain offset. A black dot marks the end of each trajectory. (**F**) Neural trajectories of the somatomotor network. The layout is the same with (E). ****$P < 0.0001$, two-tailed, paired t test.

3 dimensions for visualization. The number of dimensions was selected on the basis of the explained variance of principal components (PCs) (fig. S3), but the results did not significantly differ when tested with different numbers of dimensions (table S1).

In sum, unlike our initial hypothesis, we found that the unimodal brain networks also preserved both cue and stimulus information. However, a preference in encoding was noted: The visual network preferentially encoded cue information (Fig. 3A), whereas the somatomotor network showed a preference for stimulus information (Fig. 3B). This nuanced distinction indicates that while both types of information are preserved within these networks, each network exhibits a bias toward one type of information over the other.

## Higher effect sizes of pain component preservation in transmodal brain networks

Next, we also investigated whether the transmodal networks, including the limbic, frontoparietal, and default mode networks, preserve both cue and stimulus intensity information. First, we examined whether the networks showed different encoding performances between the cue and stimulus information. In the limbic network, there were no significant differences in encoding performances between cue and stimulus information (Fig. 4A; $t_{48} = 1.046$, $P = 0.301$, two-tailed, paired t test, Bayes factor favoring the null hypothesis = 3.85). For the frontoparietal and default mode networks, encoding performances for stimulus information were more pronounced than for cue information, similar to the results of the somatomotor network (frontoparietal: $t_{48} = 6.065$, $P = 1.992 \times 10^{-7}$, Fig. 4B; default mode: $t_{48} = 3.872$, $P = 3.260 \times 10^{-4}$, Fig. 4C).

Second, when comparing the encoding performances with those from the null subspaces, the encoding performances of both the cue and stimulus subspaces were significantly higher than those from the respective null subspaces in all transmodal networks. As shown in Fig. 4D, the limbic network showed significantly higher encoding performances for both cue ($t_{48} = 24.801$, $P = 5.089 \times 10^{-29}$) and

stimulus information ($t_{48} = 13.718$, $P = 3.148 \times 10^{-18}$). Similar results were observed in the frontoparietal network (Fig. 4E; for the cue subspace: $t_{48} = 7.322$, $P = 2.362 \times 10^{-9}$; for the stimulus subspace: $t_{48} = 8.220$, $P = 1.026 \times 10^{-10}$) and the default mode network (Fig. 4F; for the cue subspace: $t_{48} = 7.126$, $P = 4.717 \times 10^{-9}$; for the stimulus subspace: $t_{48} = 9.135$, $P = 4.495 \times 10^{-12}$; all P values were two-tailed and derived from paired t tests). These results demonstrate that the transmodal networks also preserve both cue and stimulus intensity information in their respective subspaces. Figure 4 (G to I) shows the neural trajectories within subspaces from which the encoding performances were calculated for the networks.

Further analyses revealed significant differences in effect size for encoding performances between the unimodal versus transmodal networks (fig. S4). We derived effect sizes by subtracting the null results from the actual encoding performances. Then, we calculated average effect sizes for cue and stimulus intensity across transmodal and unimodal networks and compared them between the two network types. The results show that the transmodal networks exhibited greater effect sizes for cue and stimulus intensity than the unimodal networks (for cue, $t_{48} = 9.118$, $P = 4.754 \times 10^{-12}$; for stimulus, $t_{48} = 11.566$, $P = 1.755 \times 10^{-15}$, two-tailed, paired t test). The greater effect sizes observed in the transmodal networks indicate their prominent roles in processing multiple pain-related information.

The results from the dorsal attention (dAttention) and ventral attention (vAttention) networks, which are positioned between the unimodal and transmodal networks in the cortical hierarchy (fig. S2), also indicated significant encoding of both cue (for dAttention: $t_{48} = 5.142$, $P = 4.962 \times 10^{-6}$; for vAttention: $t_{48} = 14.483$, $P = 3.830 \times 10^{-19}$; all P values are two-tailed, paired t tests) and stimulus information (for dAttention: $t_{48} = 7.977$, $P = 2.388 \times 10^{-10}$; for vAttention: $t_{48} = 10.211$, $P = 1.273 \times 10^{-13}$). The figures for encoding performances and neural trajectories of these networks are detailed in fig. S5. Note that the networks' cue and stimulus effect sizes show no significant correlation with their number of
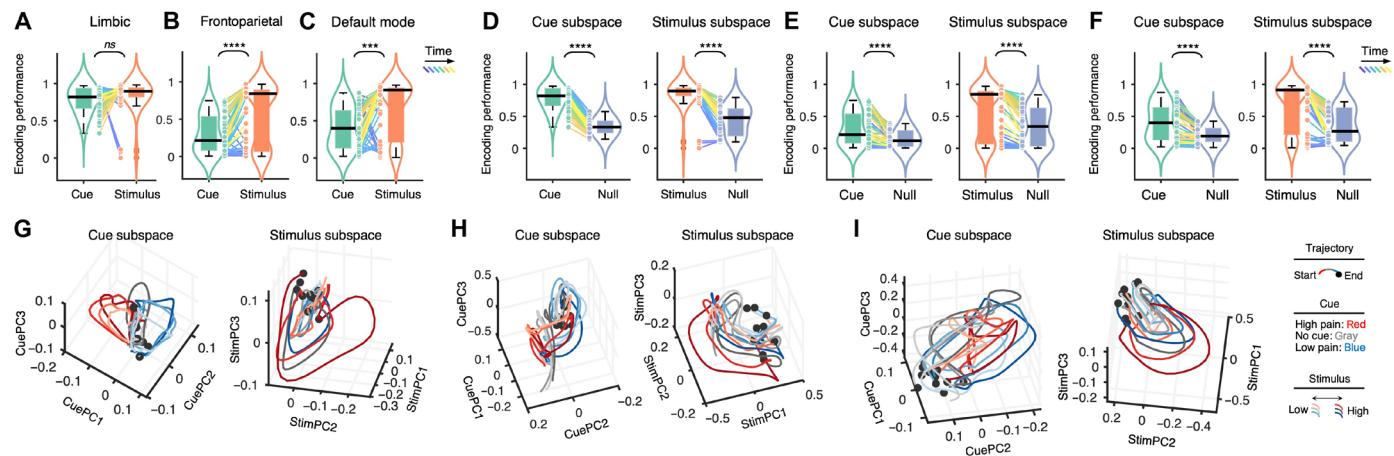
Fig. 4. Encoding performances and neural trajectories in transmodal networks. (**A** to **C**) Encoding performances of the limbic (A), frontoparietal (B), and default mode networks (C). The encoding performance was calculated for each time point, and the plots show the time information with the graded colors ranging from cool (early) to warm (late) colors. (**D** to **F**) Encoding performances based on the actual versus null subspaces for the limbic (D), frontoparietal (E), and default mode networks (F). (**G** to **I**) Neural trajectories of the limbic (G), frontoparietal (H), and default mode networks (I). The plots show the neural trajectories of the experimental conditions within the cue (left) and the stimulus subspaces (right). The axes were the top three PCs of the time-series regression coefficients. The trajectories were smoothed with a Gaussian kernel (SD: 3 TRs) for visualization purposes. The unsmoothed trajectories were used for the actual analysis. The trajectories' colors (cue levels) and brightness (stimulus intensity levels) indicate different experimental conditions. The trajectory's starting point aligns with the onset of pain stimulation, while its ending point corresponds to 10 s after the pain offset. A black dot marks the end of each trajectory. *ns*, $P > 0.05$; ***$P < 0.001$; ****$P < 0.0001$, two-tailed, paired t test.

voxels (cue effects: $r_s = -0.143$, $P = 0.783$; stimulus effects: $r_s = -0.179$, $P = 0.713$; two-tailed, Spearman's rank correlation; fig. S6).

Neural evidence for cue-stimulus integration in the limbic and default mode networks

Contrary to our initial hypothesis, which posited that the unimodal networks would selectively encode either cue or stimulus information and the transmodal networks would encode both, our results presented a different scenario. We observed that all networks across the whole brain preserved information about both cue and stimulus. Such distributed encoding of cue and stimulus led us to question whether these networks also integrate the two pieces of information they preserve. We predicted that the transmodal network would integrate the information, while the unimodal network would not. To investigate cue-stimulus integration within these networks, we reconstructed pain ratings from the neural trajectories within subspaces of each network, and calculated reconstruction fits that measure how well the reconstructed outcome aligned with the pain ratings of the participants (Fig. 5A).

As our hypothesis predicted, unimodal network regions did not successfully reconstruct pain ratings (Fig. 5, B and C). For example, ratings reconstructed using the visual network reflected only cue information (Fig. 5B), and the reconstruction fits were not significant [md (median) = 0.615, $n = 49$, $z = 1.432$, $P = 0.076$, one-tailed, Wilcoxon signed-rank test]. In addition, ratings reconstructed on

the basis of the somatomotor network largely captured stimulus information rather than cue (Fig. 5C), and their reconstruction fits were also not statistically significant (md = 0.507, $n = 49$, $z = 0.975$, $P = 0.165$).

Meanwhile, the limbic and the default mode networks successfully reconstructed the pain ratings (Fig. 5, E and F), demonstrating significant reconstruction fits (limbic: md = 0.911, $n = 49$, $z = 6.088$, $P = 5.726 \times 10^{-10}$; default mode: md = 0.791, $n = 49$, $z = 3.183$, $P = 7.285 \times 10^{-4}$; all P values were one-tailed and derived from Wilcoxon signed-rank test). In particular, the limbic network, which showed the highest encoding performance and effect size (Fig. 4D), also showed the highest reconstruction fit among all networks (Fig. 5G). Note that the reconstruction fits were also significant for the vAttention network (md = 0.846, $n = 49$, $z = 1.900$, $P = 0.029$) but not for the frontoparietal network (md = 0.424, $n = 49$, $z = 1.592$, $P = 0.056$; Fig. 5D). The significant reconstruction fits in the limbic, default mode, and vAttention networks remained consistent even when we used a different number of PCs for subspaces (fig. S7). Note that the networks' reconstruction fits show no significant correlation with their number of voxels ($r_s = -0.071$, $P = 0.906$; two-tailed, Spearman's rank correlation; fig. S6).

To assess the robustness of the results without using the spatial information of the predefined large-scale networks, we also conducted a whole-brain searchlight analysis. This involved repeating the analysis procedure of calculating the encoding performances and reconstruction fits for each spherical searchlight (radius = 5 voxels)
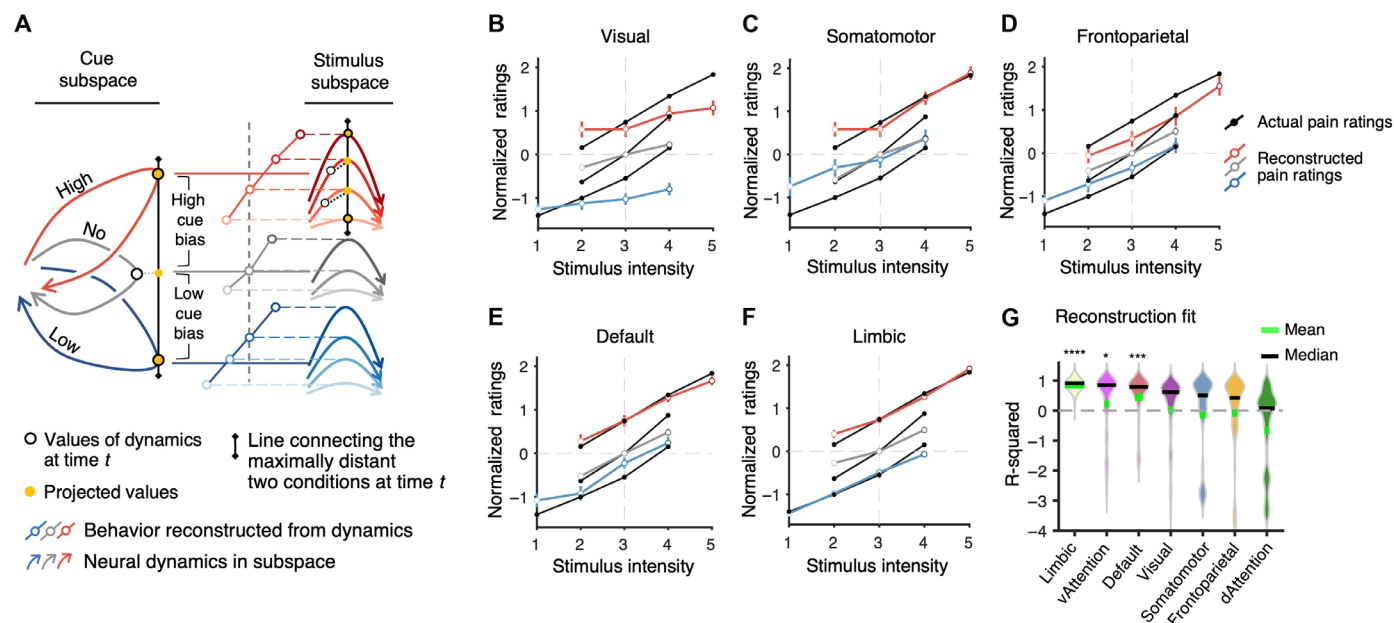
Fig. 5. Reconstructing pain ratings from neural trajectories. (**A**) The schematic illustrates how we reconstructed pain ratings from neural trajectories. For each time point in each subspace, we projected the neural trajectories onto the line that connects the maximally distant two conditions (the same procedure described in fig. S1), resulting in 11 distance values corresponding to 11 experimental conditions. In the cue subspace, we averaged the distances within each cue condition, resulting in three values. We then added the distance values from the stimulus subspace to these values in such a way that stimulus conditions corresponded to their respective cue conditions. We made the middle value (i.e., no cue and stimulus intensity level 3) as a reference by subtracting it from others and then divided the output by its SD. This procedure resulted in 11 values represented as straight colored lines of the plot, and we further compared them with the actual pain ratings. For actual pain ratings, we averaged the pain ratings of conditions across participants and set the same reference point with the reconstructed pain ratings and divided the value by its SD, which made the scale of the reconstructed and actual pain ratings comparable. (**B** to **F**) Reconstructed pain ratings (colored lines) and actual pain ratings (black lines) in the unimodal [(B) and (C)] and transmodal networks [(D) to (F)]. The error bars indicate the SEM across time. (**G**) Reconstruction fits for large-scale functional brain networks. We used the R-squared as a measure of reconstruction fit. Networks are ordered in descending order of the median. *$P < 0.05$, ***$P < 0.001$, ****$P < 0.0001$.

around each voxel across the whole brain. The results demonstrated significant encoding performances for both cue and stimulus information across nearly all brain regions (paired t test, $df = 48$, Bonferroni corrected $P < 0.05$). However, like prior results, significant reconstruction fits were primarily observed in regions within the limbic and default mode networks (fig. S8).

Considering that all the presented results included network parcellations of cerebral, cerebellar, and striatal regions, we conducted the same analysis, examining the encoding performances and reconstruction fits, only using the cerebral cortex regions (28). The analysis yielded a similar result to the whole-brain version, i.e., significant encoding of cue and stimulus information across all networks, but significant reconstruction fits only within the limbic and default mode networks (limbic: md = 0.870, $n = 49$, $z = 6.088$, $P = 5.726 \times 10^{-10}$, and default mode: md = 0.660, $n = 49$, $z = 2.626$, $P = 0.004$, one-tailed, Wilcoxon signed-rank test; fig. S9).

Overall, our findings suggest that all networks preserved information about cues and stimuli, but only the limbic and default mode networks demonstrated the integration of this information. This integration was evident by linearly combining inter-trajectory distance information from the pain component subspaces.

DISCUSSION

In this study, we used the dynamical systems framework to analyze whole-brain fMRI data, revealing differences in the preservation and integration of cue and stimulus information along the cortical hierarchy. Using semisupervised linear dimensionality reduction (26, 27), we derived two distinct subspaces, each encoding information about the pain-predictive cue and stimulus intensity. We then quantified the distances between the trajectories of different experimental conditions (i.e., inter-trajectory distances) within these subspaces and showed that all functional brain networks, ranging from unimodal to transmodal, preserved information about cues and stimuli. In addition, we reconstructed pain ratings by linearly combining inter-trajectory distances in the subspaces and found successful reconstruction in the transmodal networks, specifically the limbic and the default mode networks. This result remained robust when using different number of PCs for subspaces, conducting voxel-level searchlight analysis, and using spatial masks that included only the cerebrum. This implies the importance of the limbic and default mode networks for the cue-stimulus integration.

Although prior research has provided a broad understanding of information processing and integration along the hierarchy (39, 42, 43), they have not detailed the computational mechanisms underlying information integration within specific brain regions or networks. Conversely, studies that directly measured neuronal responses have identified behaviorally relevant neural subspaces (32, 34, 44–50) and population dynamics underlying behaviors (18, 51–54), but their scope was limited to a few brain regions due to measurement constraints. Our study, in contrast, applied the dynamical systems framework to investigate the computational mechanisms of information integration across the whole brain using fMRI.

Specifically, we focused on pain processing to understand the computational mechanism of integrating information from pain-predictive cues and noxious stimuli. However, pain inherently involves multisensory integration, where inputs from multiple sensory modalities, such as visual cues, sensory inputs, and interoceptive signals, collectively contribute to shaping the overall pain experience (55). Drawing insights from studies examining neural populations to explain general computational principles (31, 35, 51) and considering the translatability of neuronal signals into fMRI signals (56), our study proposes that linear operations on subspaces within populations of fMRI voxels could explain integrated pain behaviors. This opens an intriguing opportunity to explore whether such straightforward linear operations on subspaces could generalize to information integration in other tasks at the whole-brain level.

In terms of hierarchical differences of integration, we observed significant reconstruction fit only in the limbic and default mode networks, but not in the frontoparietal network, which is considered to be situated higher in the cortical hierarchy in resting state and other cognitive task fMRI data (23, 24, 57). The frontoparietal network, such as the dlPFC and posterior parietal cortex, is known to interact with other brain regions to modulate pain (58–62). The null results on this network may suggest that the pain modulatory process by the frontoparietal network regions could be distinct from the information integration process. Whether this different level of computation exists, and if so, the underlying mechanism should be addressed in future studies.

Furthermore, hierarchical differences of preservation and integration suggest a compromising perspective on the two prevailing hypotheses of brain functional organization—distributed processing (63–67) and functional specialization (68, 69). Our results suggest that although all brain networks preserve both types of information (i.e., cue and stimulus), the degree of information preservation varies across networks, demonstrating that there is a computational process (i.e., preservation) following the distributed processing model, but another process exists (i.e., integration) that follows the functional specialization model.

This further suggests that regions conventionally labeled as "unimodal" are not strictly unimodal, aligning with previous studies on pain (70, 71). In addition, there is a possibility that the functional hierarchical structure changes during pain processing. Many previous findings identifying the cortical hierarchy were based on resting-state fMRI data (23, 39, 72), but some studies have shown that hierarchical structure can change under task conditions, such as during movie watching (40). This suggests that the functional hierarchy may also shift during pain processing, indicating a need for a more comprehensive theoretical framework to explain the hierarchical organization of the brain.

Previous studies on the effects of cues and stimuli on pain using electroencephalogram (EEG) have shown that local brain oscillations encode stimulus intensity, whereas interregional connectivity encodes both intensity and expectancy at specific frequency levels (73, 74). Such observations indicate that different frequency bands might be involved in distinct computational processes for encoding cue and intensity information of pain. By leveraging the higher spatial resolution of fMRI than EEG, our findings propose functional subspaces within a network as a complementary-encoding mechanism for this information. Future research that combines the spatial resolution of fMRI with the temporal resolution of EEG would be valuable for further elucidating the computational mechanisms of pain.

Our study, while yielding notable insights, does have limitations that necessitate further investigation. First, the low signal-to-noise ratio inherent in fMRI data, along with the limited number of trials typically available in pain experiments, limited the application of single-trial analyses and the exploration of individual differences. Future studies should use advanced imaging techniques (e.g.,

multi-echo imaging and multimodal imaging) to improve data quality and detect reliable neural dynamics at a single trial level. With technical advancements, trial-by-trial dynamics of pain processing from sophisticated task designs (70, 75, 76) and computational models [e.g., Bayesian models (77)] will offer further insights into the computational mechanisms underlying pain perception, like some recent studies (78, 79). Second, our experiment only involved a specific pain modality (i.e., heat stimulation) with a specific type of cue (i.e., social cue based on visual input), but different pain and cue modalities, such as mechanical or cold pain, could induce distinct neural representations (80–82). Previous literature on multisensory integration has shown that a generative process model (i.e., Bayesian integration) generalizes to visual-haptic (83), visual-vestibular (84), and audio-visual information integration (85). These suggest a possibility for the generalizability of our findings while also highlighting the need for further investigation. Third, our approach used geometrical distance as a common computational unit for information integration, which required comparisons across multiple trajectories. However, we did not account for variations in dynamic features, such as velocity and curvature, across experimental conditions. Recent findings suggest that such features (e.g., curvature) are important for multiple cognitive functions, such as Bayesian priors (19) or context switching (36). In addition to these findings, our results pose intriguing questions that could enhance the understanding of not only pain processing but also general neural computation: 1) Do features of dynamics change across different subspaces? 2) How do underlying dynamical systems vary across seemingly similar neural dynamics observed in different brain regions or networks? Addressing these questions will be crucial in future research endeavors.

Neural trajectories within a network and the geometrical features derived from them may not be the only explanations for the preservation and integration of pain-related information. Although we demonstrated network-level analysis, pain processing may involve coordinated activities and interactions across multiple brain networks (11, 86). Such dynamic network interactions induce temporally evolving connectivities between different brain networks (87), which might contribute to the evolving dynamics of individual networks. Communication subspaces (88), which refer to dedicated neural subspaces that facilitate information transfer between brain regions, could provide a computational explanation. In the context of our study, we expect that incorporating additional subspaces within a network dedicated to communications would offer a comprehensive computational perspective of pain information processing. Further research is needed to understand whether such subspaces exist for pain and how these computations are implemented at the biology level, which will provide deeper insights into the neural basis of pain.

In addition, our analysis method is based on several assumptions during both the identification of the subspace (i.e., semisupervised linear dimensionality reduction) and the calculation of encoding performance. Regarding the subspace identification, we assumed that the subspaces that encode cue and stimulus are stationary over time, although they might change within a trial and/or across experimental conditions. In addition, our method relies on predefined information about cues and stimuli (i.e., predefined regressors in the GLM of Fig. 2A), which might introduce biases, particularly if actual neural representations are more complex. Regarding the calculation of encoding performance, our approach assumes a linear, quantitative relationship between conditions within each subspace (i.e., low to high cue and stimulus level 1 to 5), which may not

capture information about variables if they are nonlinearly represented. While these assumptions may introduce biases due to the simplification of neural representations, previous electrophysiology studies have used linear assumptions and dissociated functional subspaces of various task variables within populations of mixed selective neurons—neurons that respond to multiple task variables—and have demonstrated how subspaces and their dynamics can explain cognitive behaviors (19, 22, 35). Given that voxels from fMRI data also exhibit mixed selectivity, we believe that our approach could be conceptually valid for fMRI data. However, the application of such methods to fMRI data is relatively recent (89–91), and, hence, the generalizability of these methods in fMRI should be further investigated.

Overall, our study showed how and where the information about multiple pain components is processed and integrated. We achieved this by applying both the dynamical systems framework and the concept of cortical hierarchy using human fMRI data. This approach allowed us to investigate the computation of pain perception at the whole-brain level. Furthermore, our findings have the potential to offer valuable insights into the computational underpinnings of multisensory integration across a spectrum of cognitive and affective functions, extending well beyond pain.

MATERIALS AND METHODS
Participants
We used the same dataset as Gim *et al.* (92), which addressed different research questions from those of the current study. Gim *et al.* (92) provided comprehensive details of the entire experimental procedure. Here, we focused on and detailed the experimental procedures relevant to the current study. We recruited 84 healthy, right-handed individuals without neurological disorders from the Suwon area, South Korea, who consented to participate in the pain study. From this group, data from 28 participants were excluded for various reasons: thirteen because they met the exclusion criteria during the pain calibration task (as elaborated in the following section), eight withdrew during the initial pain session, one exhibited substantial discrepancies in ratings compared to the pain calibration tasks, three requested to quit during the fMRI scan, and one was excluded because of technical issues with the MRI scanner. In addition, two participants were excluded because of the absence of stimulus onsets in a trial, one for a missing field map, and two more because of technical errors in the preprocessing step. This resulted in a final sample size of $N = 56$ [age = 22.07 ± 2.48 (mean $\pm$ SD), 24 females]. Participants were recruited through university websites and flyers. All participants provided written informed consent before participating in both the first and second sessions and received a monetary reward for their participation (approximately US$10 per hour for behavioral tasks and about US$20 per hour for fMRI tasks). The institutional review board of Sungkyunkwan University approved the studies (IRB 2017-05-001).

Thermal stimulation
Thermal stimulation, with temperatures ranging from 40° to 49.2°C with a baseline of 32°C, was administered to the left forearm using a 16 mm–by–16 mm ATS thermode (Medoc, Israel). The duration of the thermal stimulation was 12.5 s, consisting of a 5-s ramp up, a 5-s plateau, and a 2.5-s ramp down (Fig. 1A). During the pain calibration task, we switched the stimulation sites among four skin sites on

the forearm for each trial. Three of the four skin sites were used in the fMRI experiment, where we switched the stimulation sites for every run. These three sites were selected on the basis of the linear regression results in the pain calibration task. We applied the highest temperature of heat stimulation before the pain calibration task (i.e., 49.2°C) and each run during the fMRI experiment (i.e., temperature corresponding to the stimulus intensity level 5) to avoid the initial habituation of the skin site (*93*).

Semicircular rating scale

To collect pain ratings, we used a semicircular rating scale. An important feature of the semicircular rating scale is that the starting point is equidistant from all possible ratings (i.e., gray outer circle in Fig. 1A), addressing an issue of intensity rating being confounded with the distance of cursor movement. At the start of the pain rating, an orange-colored dot was located at the center of the scale, and participants could move the dot using a joystick. Participants were asked to report their ratings by adjusting the angle of the dot, with 0° representing the left end of the semicircle and 180° indicating the right end. The angles indicated no sensation (0°), weak (18°), moderate (50°), strong (108°), very strong (162°), and the strongest imaginable sensation (180°). The rating trajectory was recorded by both the *x* and *y* coordinates of the dot and the angle at a sampling rate of 60 Hz. To define the anchors of the scale, we modified the generalized labeled magnitude scale (*94*) to make participants use the entire space of the scale.

Experimental procedure

The entire experiment consisted of 2-day sessions. The first session included the pain calibration task to provide similar levels of pain experience across participants (*95*, *96*). The second session included the pain task with fMRI scans. Psychtoolbox (http://psychtoolbox.org) and Matlab (MathWorks) were used to present stimulus, record ratings, and deliver heat stimulation.

Session 1: Pain calibration task

When participants arrived at the laboratory, we provided them with a detailed overview of the experimental procedure and obtained their written informed consent. Following the completion of a set of individual difference self-report questionnaires, such as demographic information, personality, and emotional states and traits, etc. (not analyzed in this study), we proceeded with the pain calibration task. This task served three purposes: 1) to familiarize participants with the heat stimulation and the experimental environment for rating, including the use of a joystick to navigate the rating scale; 2) to assess individual's subjective pain intensity to ensure similar levels of pain experience among participants; and 3) to identify and exclude individuals who exhibited either an excessively low or high sensitivity to heat stimulation. During the pain calibration task, we asked participants to report their pain intensity ratings for 12 heat stimulations.

The task began with the administration of three predetermined heat temperatures (43.4°, 45.4°, and 47.4°C) in a pseudo-random order. Pain ratings for these stimuli were used to fit a regression line predicting pain ratings based on temperatures. This linear regression line was then used to determine the subsequent stimulus temperature. For example, after the initial three stimuli and their corresponding pain ratings, a first linear regression estimation was carried out, which established three temperature levels indicative of low, medium, and high stimulus intensities, corresponding to 30%

(54°), 50% (90°), and 70% (126°) on the semicircle rating scale. These temperatures were then administered in a pseudo-random order (i.e., to avoid consecutive high stimulus intensity) across 12 trials. Linear regressions, after the first three trials, were conducted cumulatively by adding the previous trials in the regression process.

Following these trials, the final linear regression model determined five calibrated temperatures corresponding to five intensity levels for each participant (30, 40, 50, 60, and 70% of the scale, or 54°, 72°, 90°, 108°, and 126°, respectively). These temperatures were designed to evoke comparable pain intensities across participants during the fMRI experiment. We excluded participants if the estimated temperatures fell outside the predefined temperature range (40° to 49.2°C) or if the *R*-squared value of the final regression model was below 0.4. Since the stimulations were applied to multiple skin sites, we calculated *R*-squared values for each site, selecting the top three for use in the fMRI experiment. The ambient conditions (lighting and temperature) of the room where the pain calibration task took place were controlled to match those of the MRI room. To ensure consistency in the application of heat stimulations to the same skin sites in the subsequent fMRI task, photographs of each participant's left forearm were taken after the completion of the pain calibration task.

Session 2: fMRI experiment

On the second day, we conducted an fMRI experiment, which included one structural scan (T1), one resting run, two simple motor task runs, and six pain prediction task runs. Structural images were acquired while participants engaged in practice trials of the continuous pain prediction task to familiarize themselves with the task requirements. During the resting-state run, participants underwent scanning for approximately 6 min and 10 s (810 TRs), with the instruction to fixate on a cross displayed on a screen. We also administered a simple motor task, wherein participants were instructed to move the orange-colored dot to a specific location on a semicircular rating scale using an MR-compatible joystick. The data from the motor task were not used in the current study. For the pain task runs, each run comprised 18 trials, and there were 6 runs. Each trial consisted of a sequence of the following events: 1) pain-predictive cues, 2) a continuous pain prediction rating while undergoing heat stimulation, and 3) a post-heat pain intensity rating (Fig. 1A). The fMRI data during the heat stimulation period and the post-heat pain intensity rating data were used in the current study.

In the pain-predictive cue phase, 25 orange-colored dots were presented on the semicircular scale. Participants were instructed that each dot represented other participants' pain ratings from previous experiments about the upcoming stimulus. These dots served as pain-predictive cues based on social information. In practice, we displayed 25 dots randomly sampled from a normal distribution for high-pain cue (mean: 138.6°; SD: 9°) and low-pain cue (mean: 39.6°; SD: 9°).

The second heat stimulation and continuous rating phase involved a rating scale with multiple semicircles and an orange dot at the center, appearing 1 s before the heat stimulation onset. Participants were asked to continuously report their ratings in response to the following question, "How much pain do you predict?" with the instruction that the angle, rather than the distance from the starting point, indicated their pain prediction. They were also instructed to report their predictions from the trial onset and continue throughout, with a movement speed limit imposed on the dot to prevent premature maxing out of the scale. The continuous rating data from this phase were not used in the current study.

In the final rating phase, participants were given 5 s to provide their overall pain rating in response to the following questions: "How painful was it?" (self-pain question) or "How painful would it be for others?" (other-pain question). Within each run, 11 trials included the self-pain question and 7 trials included the other-pain question, and they were given in a randomized order. We exclusively used ratings from trials with the self-pain questions, as they allowed us to analyze the cue and stimulus effects on overall pain ratings. A fixation cross was displayed during the jittered inter-trial and inter-stimulus intervals, which lasted between 3 and 7 s, totaling 15 s for each trial. Trial sequences were pseudo-randomized to avoid consecutive repetition of the same condition. The average ratings over the last second of the overall pain rating for self-pain questions were used for analysis.

fMRI data acquisition and preprocessing

Both functional and structural images were acquired on a Siemens 3.0 Tesla Magnetom Prisma at Sungkyunkwan University using a 64-channel head coil. High-resolution anatomical T1-weighted images were obtained with the MPRAGE protocol (TR = 2400 ms; TE = 2.34 ms; flip angle = 8°; field of view = 224 mm; voxel size = 0.7 mm). Functional images were obtained using the T2*-weighted echo planar image protocol (TR = 460 ms; TE = 27.20 ms; multi-band factor = 8; field of view = 220 mm; flip angle = 44°; the number of slices = 56; voxel size = 2.7 mm^3). For functional images, the initial set of images (18 TRs) was discarded to ensure signal stabilization.

The functional and structural data were preprocessed with Statistical Parametric Mapping (SPM12; Wellcome Trust Centre for Neuroimaging, UCL) and FMRIB Software Library (FSL 6.0.5.1). The preprocessing included realignment, distortion correction, coregistration between structural images and functional images, spatial normalization, and spatial smoothing. Functional images were realigned to the first single-band reference image, and six movement parameters (*x*, *y*, *z*, *pitch*, *roll*, and *yaw*) were estimated for each run. Susceptibility-induced distortion correction was performed using FSL TOPUP. The coregistered functional images, aligned by the structural image, were normalized to the standard MNI brain template. These normalized images were then smoothed with a 5-mm full width at half maximum Gaussian kernel. All preprocessing steps, except for distortion correction, were performed using SPM. We then removed motion-related artifacts using ICA-AROMA (*97*).

To analyze the temporal dynamics of brain activity during heat stimulation, a finite impulse response (FIR) model was used. The first-level model included condition-specific regressors, and high-pass filtering was applied with a cutoff of 0.005 Hz. Nuisance regressors included five PCs of cerebrospinal fluid signal and white matter signals and linear drift. To fully capture information for the entire trial and to accommodate the hemodynamic response delay, our model spanned from stimulus onset to 10 s after stimulus offset, encompassing 49 TRs (22.5 s) per condition.

Multilevel GLM analysis

To assess the effects of pain-predictive cues and stimulus intensity on ratings, multilevel GLM analyses were conducted using trial-level behavioral data. In our dataset, we repeatedly measured ratings for noxious heat within each participant. To account for the data dependence within each participant, we used a multilevel model, treating participants as a random effect. To determine whether our dataset was appropriate for using a multilevel GLM, we examined

the normality of the residuals of pain ratings after accounting for the effects of cues, stimulus intensities, their interactions, and participants (*98*). We performed the Shapiro-Wilk test, which failed to reject the null hypothesis of non-normality ($W = 0.999$, $P = 0.213$), supporting the normality of residual pain ratings. Furthermore, we visualized the residuals with a histogram and quantile-quantile plot (fig. S10), demonstrating the plausibility of using a multilevel GLM. In the model predicting pain ratings, the cue conditions were coded as −1 for low-pain cue, 0 for no cue, and +1 for high-pain cue. The five levels of stimulus intensity were coded with the integers 1 through 5. The model also included an interaction to examine the interaction effects of cue condition and stimulus intensity. Averaged ratings over the last 1 s of the pain rating trajectory for trials involving self-pain questions were used for the analysis. To determine statistical significance, bootstrap tests with 10,000 resamples were used.

Cue and stimulus subspaces

Using the averaged FIR time-series data across participants, we applied the GLM at each time point to factorize the information regarding cue and stimulus for each large-scale network. In the GLM, the FIR data at each time point were treated as the dependent variable, while cue and stimulus information were the independent variables. This model provided us with temporal GLM weights. In the model, high-pain cue, low-pain cue, and no cue conditions were coded as 1, −1, and 0, respectively, while stimulus intensity levels were coded as 1 through 5. We used *z*-scores of both the cue and stimulus codes as regressors and added an intercept in the GLM. The temporal GLM weights for cue and stimulus regressors, which we termed $tBeta_{cue}$ and $tBeta_{stim}$, respectively, resulted in a matrix in $\mathbb{R}^{v \times t}$, where *v* is the number of voxels in the network, and *t* is the number of TRs (= 49). Then, we applied PCA to the matrices and used spatial PCs of $tBeta_{cue}$ as a cue subspace and $tBeta_{stim}$ as a stimulus subspace. We determined the number of PCs of the subspaces with a scree plot (fig. S3). We opted to use 20 PCs, which accounted for more than 70% of the variance across all networks. However, note that varying the number of PCs used did not substantially change the results (table S1 and fig. S7).

Encoding performances in subspaces

For each large-scale network, we projected the FIR time series onto cue and stimulus subspaces, resulting in condition-level trajectories as in Figs. 3 and 4. We then calculated the encoding performance in each subspace to assess how well each subspace encoded the relevant information. Each time point of trajectories within subspaces is a matrix in $\mathbb{R}^{c \times k}$, where *c* is the number of conditions and *k* is the number of dimensions of the subspace. We selected the most distant two points at a single time point and drew a line connecting the two points. We then projected the same time point of other trajectories onto the line, which resulted in the metric we referred to as the inter-trajectory distances. We assumed that the information of experimental conditions would be linearly encoded in the subspaces, representing differences among conditions as inter-trajectory distances among conditions. Thus, projecting within-subspace trajectories onto the line connecting the most distant points in the subspaces would represent the quantitative relationship (from the smallest to the largest) of the information that the subspace encodes.

We used these inter-trajectory distance values for each time point as the dependent variable to calculate *R*-squared values and used it as the encoding performance of each subspace. The independent

variables were either *z*-scores of the cue or stimulus conditions (i.e., cues coded as −1, 0, 1, and stimulus as 1 to 5), depending on the subspace, with an intercept. Encoding performances were calculated for each subspace at every time point. To statistically test these performances, we generated null subspaces by deriving spatial PCs of null temporal GLM weights. We obtained these weights using the same temporal GLM procedure as for calculating subspaces, but with permuted condition labels as independent variables. We used the top 20 PCs of the resulting null GLM weights for each cue and stimulus subspace. Then, we projected the FIR time series onto each null cue and stimulus subspaces, calculated the encoding performances, and repeated the permutation of condition labels 5000 times. The process resulted in 5000 encoding performances for the null cue and stimulus subspace, and we averaged the values across the iterations, resulting in null encoding performance for each subspace. Null encoding performances were also calculated for each subspace at every time point. To determine the effect sizes in encoding performance, we subtracted the null encoding performances from the encoding performances at the same time point.

Reconstructing ratings from inter-trajectory distance

To examine how the neural trajectories in subspaces were related to actual pain ratings, we used the inter-trajectory distances to reconstruct the average pain ratings across participants for each condition. The calculation of inter-trajectory distances follows the same procedures as those calculated for the encoding performances. In both the cue and stimulus subspaces, the inter-trajectory distances at a single time point resulted in a vector with a size of the number of conditions. We ensured that this vector maintained a consistent quantitative relationship across time points so that the high cue value was larger than the low cue value in the cue subspace, and the stimulus level 5 value was larger than the stimulus level 1 value in the stimulus subspace. The inverse relationship of such conditions might occur from the projection process, as the direction of the line connecting the maximally distant two conditions, on which the values were projected, was solely determined by the distance information and ignorant of the condition information. This was addressed by multiplying −1 in such cases. Note that this process was unnecessary for the calculation of the encoding performance, as multiplying −1 does not change the resulting *R*-squared.

In the cue subspace, we averaged the values corresponding to the same cue, resulting in three values for high pain, low pain, and no cue conditions, respectively. Subsequently, we added the distance values from the stimulus subspace to these three values, aligning stimulus conditions with their respective cue conditions. For instance, values in the stimulus subspace corresponding to a high cue with stimulus levels 2 to 5 were added to the high cue value in the cue subspace. The resulting output reflects the distance information of both cue and stimulus conditions (Fig. 5A). Furthermore, we subtracted the value of the middle condition (i.e., the no cue condition with stimulus intensity level 3) from the output and divided it by its SD. For the actual pain ratings as well, we extracted the middle condition value and divided it by its SD to ensure a consistent scale across both the reconstructed and actual ratings. To compare them, we calculated the *R*-squared value of the reconstruction fit as follows

$$1 - \frac{\sum_c \left(y - \hat{y}\right)^2}{\sum_c \left(y - \bar{y}\right)^2}$$

In the formula, *y* indicates the across-participants averaged pain ratings, $\bar{y}$ indicates the average of the values, $\hat{y}$ indicates the reconstructed outcome, and *c* indicates the condition. Both *y* and $\hat{y}$ are in the dimension of conditions, which is 11. This metric was calculated for every time point.

Supplementary Materials

This PDF file includes:
Fig. S1 to S10
Table S1
References

REFERENCES AND NOTES

1. A. I. Basbaum, D. M. Bautista, G. Scherrer, D. Julius, Cellular and molecular mechanisms of pain. *Cell* **139**, 267–284 (2009).
2. L. Y. Atlas, T. D. Wager, How expectations shape pain. *Neurosci. Lett.* **520**, 140–148 (2012).
3. H. L. Fields, How expectations influence pain. *Pain* **159**, S3–S10 (2018).
4. B. Seymour, F. Mancini, Hierarchical models of pain: Inference, information-seeking, and adaptive control. *Neuroimage* **222**, 117212 (2020).
5. K. Wiech, Deconstructing the sensation of pain: The influence of cognitive processes on pain perception. *Science* **354**, 584–587 (2016).
6. N. Fallon, C. Roberts, A. Stancak, Shared and distinct functional networks for empathy and pain processing: A systematic review and meta-analysis of fMRI studies. *Soc. Cogn. Affect. Neurosci.* **15**, 709–723 (2020).
7. V. A. Apkarian, J. A. Hashmi, M. N. Baliki, Pain and the brain: Specificity and plasticity of the brain in clinical chronic pain. *Pain* **152**, S49–S64 (2011).
8. A. Watson, W. El-Deredy, G. D. Iannetti, D. Lloyd, I. Tracey, B. A. Vogt, V. Nadeau, A. K. P. Jones, Placebo conditioning and placebo analgesia modulate a common brain network during pain anticipation and perception. *Pain* **145**, 24–30 (2009).
9. K. Wiech, M. Ploner, I. Tracey, Neurocognitive aspects of pain perception. *Trends Cogn. Sci.* **12**, 306–313 (2008).
10. C.-W. Woo, M. Roy, J. T. Buhle, T. D. Wager, Distinct brain systems mediate the effects of nociceptive input and self-regulation on pain. *PLOS Biol.* **13**, e1002036 (2015).
11. R. C. Coghill, C. N. Sang, J. M. Maisog, M. J. Iadarola, Pain intensity processing within the human brain: A bilateral, distributed mechanism. *J. Neurophysiol.* **82**, 1934–1943 (1999).
12. C. J. Vierck, B. L. Whitsel, O. V. Favorov, A. W. Brown, M. Tommerdahl, Role of primary somatosensory cortex in the coding of pain. *Pain* **154**, 334–344 (2013).
13. P. Rainville, G. H. Duncan, D. D. Price, B. Carrier, M. C. Bushnell, Pain affect encoded in human anterior cingulate but not somatosensory cortex. *Science* **277**, 968–971 (1997).
14. S. Vyas, M. D. Golub, D. Sussillo, K. V. Shenoy, Computation through neural population dynamics. *Annu. Rev. Neurosci.* **43**, 249–275 (2020).
15. C. J. MacDowell, T. J. Buschman, Low-dimensional spatiotemporal dynamics underlie cortex-wide neural activity. *Curr. Biol.* **30**, 2665–2680.e8 (2020).
16. S. Saxena, J. P. Cunningham, Towards the neural population doctrine. *Curr. Opin. Neurobiol.* **55**, 103–111 (2019).
17. R. B. Ebitz, B. Y. Hayden, The population doctrine in cognitive neuroscience. *Neuron* **109**, 3055–3068 (2021).
18. A. Nair, T. Karigo, B. Yang, S. Ganguli, M. J. Schnitzer, S. W. Linderman, D. J. Anderson, A. Kennedy, An approximate line attractor in the hypothalamus encodes an aggressive state. *Cell* **186**, 178–193.e15 (2023).
19. H. Sohn, D. Narain, N. Meirhaeghe, M. Jazayeri, Bayesian computation through cortical latent dynamics. *Neuron* **103**, 934–947.e5 (2019).
20. C. Langdon, M. Genkin, T. A. Engel, A unifying perspective on neural manifolds and circuits for cognition. *Nat. Rev. Neurosci.* **24**, 363–377 (2023).
21. L. M. Boyle, L. Posani, S. Irfan, S. A. Siegelbaum, S. Fusi, Tuned geometries of hippocampal representations meet the computational demands of social memory. *Neuron* **112**, 1358–1371.e9 (2024).
22. Y. Xie, P. Hu, J. Li, J. Chen, W. Song, X.-J. Wang, T. Yang, S. Dehaene, S. Tang, B. Min, L. Wang, Geometry of sequence working memory in macaque prefrontal cortex. *Science* **375**, 632–639 (2022).
23. D. S. Margulies, S. S. Ghosh, A. Goulas, M. Falkiewicz, J. M. Huntenburg, G. Langs, G. Bezgin, S. B. Eickhoff, F. X. Castellanos, M. Petrides, E. Jefferies, J. Smallwood, Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12574–12579 (2016).
24. U. Hasson, J. Chen, C. J. Honey, Hierarchical process memory: Memory as an integral component of information processing. *Trends Cogn. Sci.* **19**, 304–313 (2015).
25. J. M. Huntenburg, P.-L. Bazin, D. S. Margulies, Large-scale gradients in human cortical organization. *Trends Cogn. Sci.* **22**, 21–31 (2018).

26. V. Mante, D. Sussillo, K. V. Shenoy, W. T. Newsome, Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78–84 (2013).

27. M. C. Aoi, V. Mante, J. W. Pillow, Prefrontal cortex exhibits multidimensional dynamic encoding during decision-making. *Nat. Neurosci.* **23**, 1410–1420 (2020).

28. B. T. T. Yeo, F. M. Krienen, J. Sepulcre, M. R. Sabuncu, D. Lashkari, M. Hollinshead, J. L. Roffman, J. W. Smoller, L. Zöllei, J. R. Polimeni, B. Fischl, H. Liu, R. L. Buckner, The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).

29. E. Y. Choi, B. T. T. Yeo, R. L. Buckner, The organization of the human striatum estimated by intrinsic functional connectivity. *J. Neurophysiol.* **108**, 2242–2263 (2012).

30. R. L. Buckner, F. M. Krienen, A. Castellanos, J. C. Diaz, B. T. T. Yeo, The organization of the human cerebellum estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 2322–2345 (2011).

31. M. T. Kaufman, M. M. Churchland, S. I. Ryu, K. V. Shenoy, Cortical activity in the null space: Permitting preparation without movement. *Nat. Neurosci.* **17**, 440–448 (2014).

32. S. B. M. Yoo, B. Y. Hayden, The transition from evaluation to selection involves neural subspace reorganization in core reward regions. *Neuron* **105**, 712–724.e4 (2020).

33. I. Voitov, T. D. Mrsic-Flogel, Cortical feedback loops bind distributed representations of working memory. *Nature* **608**, 381–389 (2022).

34. C. Tang, R. Herikstad, A. Parthasarathy, C. Libedinsky, S.-C. Yen, Minimally dependent activity subspaces for working memory and motor preparation in the lateral prefrontal cortex. *eLife* **9**, 58154 (2020).

35. Z. Fu, D. Beam, J. M. Chung, C. M. Reed, A. N. Mamelak, R. Adolphs, U. Rutishauser, The geometry of domain-general performance monitoring in the human medial frontal cortex. *Science* **376**, eabm9922 (2022).

36. G. Okazawa, C. E. Hatch, A. Mancoo, C. K. Machens, R. Kiani, Representational geometry of perceptual decisions in the monkey parietal cortex. *Cell* **184**, 3748–3761.e18 (2021).

37. S. Bernardi, M. K. Benna, M. Rigotti, J. Munuera, S. Fusi, C. D. Salzman, The geometry of abstraction in the hippocampus and prefrontal cortex. *Cell* **183**, 954–967.e21 (2020).

38. V. J. Sydnor, B. Larsen, D. S. Bassett, A. Alexander-Bloch, D. A. Fair, C. Liston, A. P. Mackey, M. P. Milham, A. Pines, D. R. Roalf, J. Seidlitz, T. Xu, A. Raznahan, T. D. Satterthwaite, Neurodevelopment of the association cortices: Patterns, mechanisms, and implications for psychopathology. *Neuron* **109**, 2820–2846 (2021).

39. R. V. Raut, A. Z. Snyder, M. E. Raichle, Hierarchical dynamics as a macroscopic organizing principle of the human brain. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 20890–20897 (2020).

40. A. Samara, J. Eilbott, D. S. Margulies, T. Xu, T. Vanderwal, Cortical gradients during naturalistic processing are hierarchical and modality-specific. *Neuroimage* **271**, 120023 (2023).

41. Z. Huang, G. A. Mashour, A. G. Hudetz, Functional geometry of the cortex encodes dimensions of consciousness. *Nat. Commun.* **14**, 72 (2023).

42. T. Ito, J. D. Murray, Multitask representations in the human cortex transform along a sensory-to-motor hierarchy. *Nat. Neurosci.* **26**, 306–315 (2023).

43. R. V. Raut, A. Z. Snyder, A. Mitra, D. Yellin, N. Fujii, R. Malach, M. E. Raichle, Global waves synchronize the brain's functional systems with fluctuating arousal. *Sci. Adv.* **7**, eabf2709 (2021).

44. G. F. Elsayed, A. H. Lara, M. T. Kaufman, M. M. Churchland, J. P. Cunningham, Reorganization between preparatory and movement population responses in motor cortex. *Nat. Commun.* **7**, 13239 (2016).

45. M. D. Golub, P. T. Sadtler, E. R. Oby, K. M. Quick, S. I. Ryu, E. C. Tyler-Kabara, A. P. Batista, S. M. Chase, B. M. Yu, Learning by neural reassociation. *Nat. Neurosci.* **21**, 607–616 (2018).

46. J. Weber, G. Iwama, A.-K. Solbakk, A. O. Blenkmann, P. G. Larsson, J. Ivanovic, R. T. Knight, T. Endestad, R. Helfrich, Subspace partitioning in the human prefrontal cortex resolves cognitive interference. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2220523120 (2023).

47. A. Libby, T. J. Buschman, Rotational dynamics reduce interference between sensory and memory representations. *Nat. Neurosci.* **24**, 715–726 (2021).

48. A. C. Snyder, B. M. Yu, M. A. Smith, A stable population code for attention in prefrontal cortex leads a dynamic attention code in visual cortex. *J. Neurosci.* **41**, 9163–9176 (2021).

49. D. L. Kimmel, G. F. Elsayed, J. P. Cunningham, W. T. Newsome, Value and choice as separable and stable representations in orbitofrontal cortex. *Nat. Commun.* **11**, 3466 (2020).

50. J. A. Gallego, M. G. Perich, S. N. Naufel, C. Ethier, S. A. Solla, L. E. Miller, Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. *Nat. Commun.* **9**, 4233 (2018).

51. M. M. Churchland, J. P. Cunningham, M. T. Kaufman, J. D. Foster, P. Nuyujukian, S. I. Ryu, K. V. Shenoy, K. V. Shenoy, Neural population dynamics during reaching. *Nature* **487**, 51–56 (2012).

52. D. Thura, J. F. Cabana, A. Feghaly, P. Cisek, Integrated neural dynamics of sensorimotor decisions and actions. *PLOS Biol.* **20**, e3001861 (2022).

53. E. L. Sylwestrak, Y. J. Jo, S. Vesuna, X. Wang, B. Holcomb, R. H. Tien, D. K. Kim, L. Fenno, C. Ramakrishnan, W. E. Allen, R. Chen, K. V. Shenoy, D. Sussillo, K. Deisseroth, Cell-type-specific population dynamics of diverse reward computations. *Cell* **185**, 3568–3587.e27 (2022).

54. M. Khona, I. R. Fiete, Attractor and integrator networks in the brain. *Nat. Rev. Neurosci.* **23**, 744–766 (2022).

55. D. Senkowski, M. Höfle, A. K. Engel, Crossmodal shaping of pain: A multisensory approach to nociception. *Trends Cogn. Sci.* **18**, 319–327 (2014).

56. N. Kriegeskorte, R. Cusack, P. Bandettini, How does an fMRI voxel sample the neuronal activity pattern: Compact-kernel or complex spatiotemporal filter? *Neuroimage* **49**, 1965–1976 (2010).

57. L. Cocchi, A. Zalesky, A. Fornito, J. B. Mattingley, Dynamic cooperation and competition between brain systems during cognitive control. *Trends Cogn. Sci.* **17**, 493–501 (2013).

58. J. Kong, K. Jensen, R. Loiotile, A. Cheetham, H.-Y. Wey, Y. Tan, B. Rosen, J. W. Smoller, T. J. Kaptchuk, R. L. Gollub, Functional connectivity of the frontoparietal network predicts cognitive modulation of pain. *Pain* **154**, 459–467 (2013).

59. J. Lorenz, S. Minoshima, K. L. Casey, Keeping pain out of mind: The role of the dorsolateral prefrontal cortex in pain modulation. *Brain* **126**, 1079–1091 (2003).

60. D. A. Seminowicz, M. Moayedi, The dorsolateral prefrontal cortex in acute and chronic pain. *J. Pain* **18**, 1027–1035 (2017).

61. V. Legrain, S. Van Damme, C. Eccleston, K. D. Davis, D. A. Seminowicz, G. Crombez, A neurocognitive model of attention to pain: Behavioral and neuroimaging evidence. *Pain* **144**, 230–232 (2009).

62. D. A. Seminowicz, K. D. Davis, Interactions of pain intensity and cognitive load: The brain stays on task. *Cereb. Cortex* **17**, 1412–1422 (2007).

63. M. T. Wallace, R. Ramachandran, B. E. Stein, A revised view of sensory cortical parcellation. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2167–2172 (2004).

64. S. A. Koay, A. S. Charles, S. Y. Thiberge, C. D. Brody, D. W. Tank, Sequential and efficient neural-population coding of complex task information. *Neuron* **110**, 328–349.e11 (2022).

65. L. Pinto, K. Rajan, B. DePasquale, S. Y. Thiberge, D. W. Tank, C. D. Brody, Task-dependent changes in the large-scale dynamics and necessity of cortical regions. *Neuron* **104**, 810–824.e9 (2019).

66. T. T. Rogers, J. L. McClelland, Parallel distributed processing at 25: Further explorations in the microstructure of cognition. *Cognit. Sci.* **38**, 1024–1077 (2014).

67. S. B. M. Yoo, B. Y. Hayden, Economic choice as an untangling of options into actions. *Neuron* **99**, 434–447 (2018).

68. J. A. Fodor, *The Modularity of Mind* (MIT Press, 1983).

69. B. T. T. Yeo, F. M. Krienen, S. B. Eickhoff, S. N. Yaakub, P. T. Fox, R. L. Buckner, C. L. Asplund, M. W. L. Chee, Functional specialization and flexibility in human association cortex. *Cereb. Cortex* **25**, 3654–3672 (2015).

70. F. Mancini, S. Zhang, B. Seymour, Computational and neural mechanisms of statistical pain learning. *Nat. Commun.* **13**, 6613 (2022).

71. M. Liang, A. Mouraux, L. Hu, G. D. Iannetti, Primary sensory cortices contain distinguishable spatial patterns of activity for each sense. *Nat. Commun.* **4**, 1979 (2013).

72. Y. Gu, L. E. Sainburg, S. Kuang, F. Han, J. W. Williams, Y. Liu, N. Zhang, X. Zhang, D. A. Leopold, X. Liu, Brain activity fluctuations propagate as waves traversing the cortical hierarchy. *Cereb. Cortex* **31**, 3986–4005 (2021).

73. F. S. Bott, M. M. Nickel, V. D. Hohn, E. S. May, C. Gil Ávila, L. Tiemann, J. Gross, M. Ploner, Local brain oscillations and interregional connectivity differentially serve sensory and expectation effects on pain. *Sci. Adv.* **9**, eadd7572 (2023).

74. M. M. Nickel, L. Tiemann, V. D. Hohn, E. S. May, C. Gil Ávila, F. Eippert, M. Ploner, Temporal-spectral signaling of sensory information and expectations in the cerebral processing of pain. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2116616119 (2022).

75. J. Onysk, N. Gregory, M. Whitefield, M. Jain, G. Turner, B. Seymour, F. Mancini, Statistical learning shapes pain perception and prediction independently of external cues. *elife* **12**, –RP90634 (2024).

76. M. Roy, D. Shohamy, N. Daw, M. Jepma, G. E. Wimmer, T. D. Wager, Representation of aversive prediction errors in the human periaqueductal gray. *Nat. Neurosci.* **17**, 1607–1612 (2014).

77. A. Strube, B. Horing, M. Rose, C. Büchel, Agency affects pain inference through prior shift as opposed to likelihood precision modulation in a Bayesian pain model. *Neuron* **111**, 1136–1151.e7 (2023).

78. J. Kim, A. Joshi, L. Frank, K. Ganguly, Cortical-hippocampal coupling during manifold exploration in motor cortex. *Nature* **613**, 103–110 (2023).

79. S. K. Wandelt, D. A. Bjånes, K. Pejsa, B. Lee, C. Liu, R. A. Andersen, Representation of internal speech by single neurons in human supramarginal gyrus. *Nat. Hum. Behav.* **8**, 1136–1149 (2024).

80. E. G. Duerden, M.-C. Albanese, Localization of pain-related brain activation: A meta-analysis of neuroimaging data. *Hum. Brain Mapp.* **34**, 109–149 (2013).

81. M. Čeko, P. A. Kragel, C.-W. Woo, M. López-Solà, T. D. Wager, Common and stimulus-type-specific brain representations of negative affect. *Nat. Neurosci.* **25**, 760–770 (2022).

82. L. Koban, M. Jepma, M. López-Solà, T. D. Wager, Different brain networks mediate the effects of social and conditioned expectations on pain. *Nat. Commun.* **10**, 4096 (2019).

83. M. O. Ernst, Learning to integrate arbitrary signals from vision and touch. *J. Vis.* **7**, 7.1-14 (2007).

84. C. R. Fetsch, A. H. Turner, G. C. DeAngelis, D. E. Angelaki, Dynamic reweighting of visual and vestibular cues during self-motion perception. *J. Neurosci.* **29**, 15601–15612 (2009).

85. N. W. Roach, J. Heron, P. V. McGraw, Resolving multisensory conflict: A strategy for balancing the costs and benefits of audio-visual integration. *Proc. Biol. Sci.* **273**, 2159–2168 (2006).

86. R. C. Coghill, *The Distributed Nociceptive System: A Framework for Understanding Pain*. Elsevier Ltd. [Preprint] (2020). (https://doi.org/10.1016/j.tins.2020.07.004.

87. J.-J. Lee, S. Lee, D. H. Lee, C.-W. Woo, Functional brain reconfiguration during sustained pain. *eLife* **11**, e74463 (2022).

88. J. D. Semedo, A. Zandvakili, C. K. Machens, B. M. Yu, A. Kohn, Cortical areas interact through a communication subspace. *Neuron* **102**, 249–259.e4 (2019).

89. T. Zhang, J. S. Gao, T. Çukur, J. L. Gallant, Voxel-based state space modeling recovers task-related cognitive states in naturalistic fMRI experiments. *Front. Neurosci.* **14**, 565976 (2020).

90. Y. J. John, K. S. Sawyer, K. Srinivasan, E. J. Müller, B. R. Munn, J. M. Shine, It's about time: Linking dynamical systems with human neuroimaging to understand the brain. *Netw. Neurosci.* **6**, 960–979 (2022).

91. M. Venkatesh, J. Jaja, L. Pessoa, Brain dynamics and temporal trajectories during task and naturalistic processing. *Neuroimage* **186**, 410–423 (2019).

92. S. Gim, S.-J. Hong, E. A. Reynolds Losin, C.-W. Woo, Temporal dynamics of brain mediation in predictive cue-induced pain modulation. bioRxiv 576786 [Preprint] (2024). (https://doi.org/10.1101/2024.01.23.576786.

93. M. Jepma, M. Jones, T. D. Wager, The dynamics of pain: Evidence for simultaneous site-specific habituation and site-nonspecific sensitization in thermal pain. *J. Pain* **15**, 734–746 (2014).

94. L. M. Bartoshuk, V. B. Duffy, B. G. Green, H. J. Hoffman, C.-W. Ko, L. A. Lucchina, L. E. Marks, D. J. Snyder, J. M. Weiffenbach, Valid across-group comparisons with labeled scales: The gLMS versus magnitude matching. *Physiol. Behav.* **82**, 109–114 (2004).

95. W. Yoshida, B. Seymour, M. Koltzenburg, R. J. Dolan, Uncertainty increases pain: Evidence for a novel mechanism of pain modulation involving the periaqueductal gray. *J. Neurosci.* **33**, 5638–5646 (2013).

96. L. Y. Atlas, N. Bolger, M. A. Lindquist, T. D. Wager, Brain mediators of predictive cue effects on perceived pain. *J. Neurosci.* **30**, 12964–12977 (2010).

97. R. H. R. Pruim, M. Mennes, D. van Rooij, A. Llera, J. K. Buitelaar, C. F. Beckmann, ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *Neuroimage* **112**, 267–277 (2015).

98. H. Singmann, D. Kellen, "An introduction to mixed models for experimental psychology" in *New Methods in Cognitive Psychology* (Routledge, 2019), pp. 4–31.

99. R. Vos de Wael, O. Benkarim, C. Paquola, S. Lariviere, J. Royer, S. Tavakol, T. Xu, S.-J. Hong, G. Langs, S. Valk, B. Misic, M. Milham, D. Margulies, J. Smallwood, B. C. Bernhardt, BrainSpace: A toolbox for the analysis of macroscale gradients in neuroimaging and connectomics datasets. *Commun. Biol.* **3**, 103 (2020).