Widespread, perception-related information in the human brain scales with levels of consciousness

Andrew D. Vigotsky [1,2,3,4]
Rami Jabakhanji [3,4]
Paulo Branco [3,4]
Gian Domenico Iannetti [5,6]
Marwan N. Baliki [4,7,8]
A. Vania Apkarian [3,4,8,9,*]

[1] Department of Biomedical Engineering, Northwestern University, Evanston, IL
[2] Department of Statistics, Northwestern University, Evanston, IL
[3] Department of Neuroscience, Northwestern University, Chicago, IL
[4] Center for Translational Pain Research, Northwestern University, Chicago, IL
[5] Division of Biosciences, University College London, London, United Kingdom
[6] Neuroscience and Behaviour Laboratory, Italian Institute of Technology, Rome, Italy
[7] Shirley Ryan AbilityLab, Chicago, IL
[8] Department of Physical Medicine & Rehabilitation, Northwestern University, Chicago, IL
[9] Department of Anesthesiology, Northwestern University, Chicago, IL

* Corresponding author:
A. Vania Apkarian
a-apkarian@northwestern.edu

Funding
This research was supported in part through the computational resources and staff contributions provided for the Quest high-performance computing facility at Northwestern University, which is jointly supported by the Office of the Provost, the Office for Research, and Northwestern University Information Technology. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1324585. GDI is supported by the Wellcome Trust and the ERC Consolidator Grant PAINSTRAT. This work is funded by the National Institute of Drug Abuse at National Institutes of Health (1P50DA044121) and the National Institute Of Neurological Disorders And Stroke of the National Institutes of Health under Award Number F31NS126012. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Acknowledgments
We would like to thank Drs. Christof Koch, Todd Parrish, and Lucas Pinto, in addition to the Apkarian Lab members, for their thoughtful comments and feedback. Our acknowledgment does not imply endorsement of our views by these colleagues, and we remain solely responsible for the views expressed herein.

ABSTRACT

How does the human brain generate coherent, subjective perceptions—transforming yellow and oblong visual sensory information into the perception of an edible banana [1]*? This is a hard problem. The standard viewpoint posits that anatomical and functional networks somehow integrate local, specialized processing across the brain to construct unique percepts. Here, we provide evidence for a novel organizational concept: we uncover task-specific information distributed across the human brain. We use functional magnetic resonance imaging (fMRI)-based decoding models to probe the presence of task-specific information. In doing so, we show that brain regions that are empirically almost entirely task-independent (considered to be "noise"—i.e.,* t-*statistics ≈ 0—in standard univariate fMRI analyses*) *can decode various tasks, even without statistical learning tools. Thus, task-specific information is present and readily accessible throughout the neocortex, challenging the sensitivity of traditional linear, univariate analytical approaches. Remarkably, even subcortical structures and the cerebellum contain task-specific signals in voxels that would univariately be considered "noise". Finally, using data from an auditory study with different levels of sedation, we show that the widespread signal in regions remote from the primary and secondary sensory cortices (auditory cortex) depends on the level of sedation, suggesting the brain's widespread, task-specific information is related to perception*[1] *rather than the encoding of the sensory stimulus. Our analysis uncovers task-specific and consciousness level-dependent information across the human brain, which we hypothesize to be the basis for coherent, subjective perceptions.*

SIGNIFICANCE

Brain imaging studies commonly rely on linear, univariate analyses to identify task-specific information. In such studies, it is assumed that such task-specific information is localized to specific "blobs." Our data challenge this notion: We show that task-specific information can be reliably uncovered across the entire human brain, even in regions that linear, univariate analyses would imply are "noise." In turn, the common, linear, univariate analysis of brain imaging data cannot detect real, widespread, task-specific information. Finally, we demonstrate that widespread, task-specific information degrades with sedation, except in the primary sensory cortex, suggesting that brain-wide information tracks perception and is a neural correlate of consciousness.

MAIN TEXT

fMRI has transformed how we study the brain, allowing the non-invasive measurement of correlates of neural activity with a spatial resolution on the scale of millimeters. This high resolution enables the comparison of blood oxygenation level-dependent (BOLD) activity within and between tasks to unravel the function of local neural circuits. Such task-based fMRI studies commonly use forward inference to identify task-related brain areas: they rely on the correlation between each voxel's activation timecourse with the task's temporal profile. Subsequently, neuroimagers contrast brain activity maps of different tasks to create a contrast map, which is then thresholded to yield localized blobs. The standard assumption of these mass-univariate analyses is that only the identified blobs reliably contain task-specific information.[2] In this report, we challenge this assumption by showing the reliable presence of task-specific information throughout

---

[1] By *perception*, we refer to the conscious experience resulting from sensory input. This experience is the product of and thus includes memory, attention, and expectations.

[2] Here, we use *information* not as an inference regarding neuronal function, but rather from a decoding perspective relating to the nature and specificity of the variance that can be extracted from our proxies of neuronal activity.

the entire neocortex—including regions identified as "noise" by univariate measures (*t*-statistics ≈ 0)—which is uncovered by integrating over large swaths of voxels. After discovering pan-neocortical information content, we probe for and uncover task-specific information in the subcortex and cerebellum. Next, we assess how different levels of sedation perturb the presence of information. We show that omnipresent information degrades with increasing levels of sedation. Rather than being task-specific, this brain-wide spread of information appears to reflect perceptual (conscious) processes and, therefore, may be involved in extracting subjective, wholistic concepts from incoming sensory inputs, as in the identification of the edible banana.

One approach to assessing the information contained in neuroimages is to decode them. Decoders transmute brain activity into a single number related to a task of interest. A suitable decoder's output will be high for the task of interest and low for the task(s) of no interest. The discriminability between the tasks of interest and no interest is indicative of the information in the data. Using six datasets with different sensory stimuli (Table S1; *N*=293 subjects) [2-7], we built simple decoding models using only the *t*-statistics from mass-univariate contrasts (Fig 1, top left; Fig S1). The first part of our study uses five of these datasets, four of which contain two stimuli and one of which includes four stimuli, totaling ten different stimulus pairs or contrasts ($4\binom{2}{2} + \binom{4}{2} = 10$). We binned the *t*-statistics by magnitude to create ten decoders for each contrast; the first decoder comprised of all voxels with the highest *t*-statistics (10th decile) and the last decoder comprised of all voxels with the lowest *t*-statistics (1st decile) (Fig 1, top right). We tested each decoder by calculating the dot product between the decoder (sign, magnitude, and location of *t*-statistics within a single decile; e.g., Fig 1, bottom right) and each brain activity map (general linear model (GLM)-derived maps of parameter estimates), yielding a weighted sum of task-related activity across all voxels comprising the decoder (Fig 1, bottom middle). We used the 0.623+ bootstrap to obtain unbiased estimates of between-subject areas under the receiver operating characteristic curve (AUC) as an indicator of discrimination performance (Fig 1, bottom left). To succinctly describe our results, we meta-analyzed the resulting AUCs and their bootstrapped variance-covariances (*Methods*).

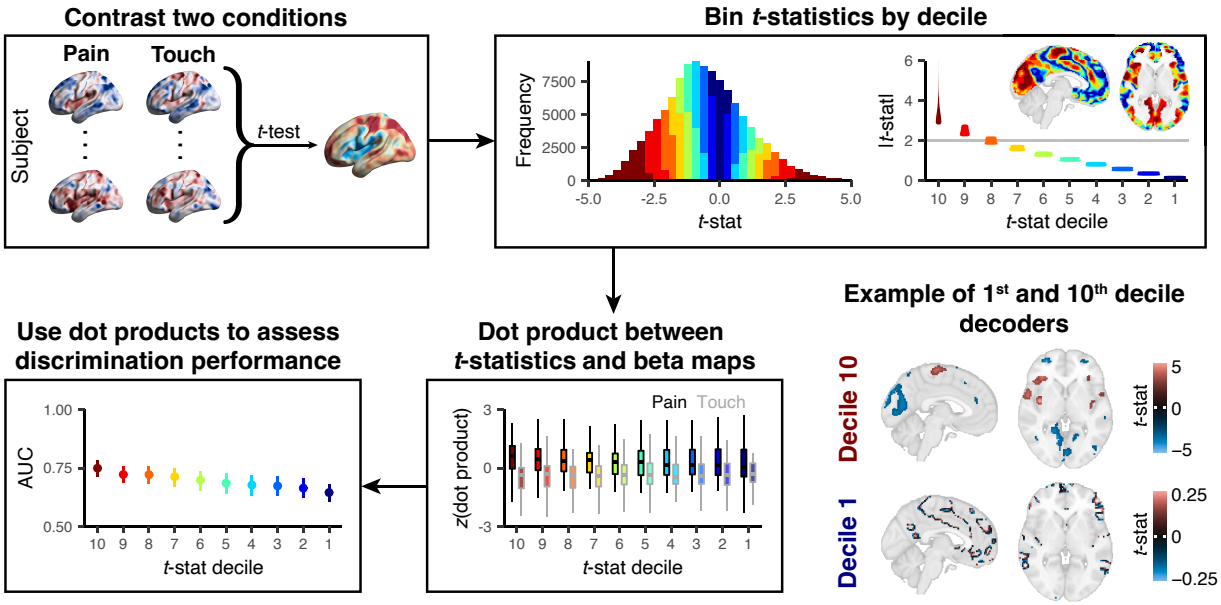

Figure 1. ***Generation of decoders to assess the presence of task-related signals.*** Voxel-wise paired *t*-tests were performed on subjects' brain activity maps using a task of interest (e.g., pain) and no interest (e.g., touch). These *t*-statistics were then binned into deciles based on their absolute magnitudes. The grey line depicts the standard minimum threshold used to dichotomize brain regions that discriminate between tasks (univariate "statistical significance" at α=0.05; uncorrected for multiple comparisons). Each decile of *t*-statistics and their locations in the brain became a decoder. On the bottom right, we show examples of these decoders for the 10th (comprised of large blobs) and 1st deciles (scattered voxels). Although the deciles were derived using the absolute value of *t*-statistics, the decoders incorporated the *t*-statistics' signs. We then calculated dot products between the decoder derived from each decile and brain activity maps. These dot products are analogous to 'linear predictors' from a regression model. Here, we *z*-scored the dot products within each decile for visualization purposes. We calculated AUCs based on these dot products, where higher dot products were assumed to correspond to the task of interest. We used the 0.632+ bootstrap to obtain unbiased AUCs when testing our decoders.

Task-specific information is widespread across the human brain

Decoding performance was consistently above chance (AUC > 0.5) for all deciles across nine out of ten contrasts. Despite univariate *t*-statistics in the lowest decile being close to zero, decoding performance was only marginally poorer in the lowest decile as compared to the highest decile (Fig 2, Table S2). Therefore, regions in neocortical grey matter commonly thought to be orthogonal to the task in univariate analyses in fact contain robust task-related information. These findings complement recent work using statistical learning to optimize voxel weights for predictive performance [8-12] and demonstrate that the presence of information is far more distributed across the brain than previously thought. Moreover, our analyses establish for the first time how accessible this information truly is: our models simply use mass-univariate *t*-statistics without any regularization or consideration of the *t*-statistics' joint distribution. Evidently, regularization and multivariable modeling are not necessary and even voxels with *t*-statistics close to zero can jointly discriminate tasks from one another. Therefore, our results indicate the presence of task-related information throughout the neocortex, which degrades slowly as a function of the univariate signal-to-noise metric (*t*-statistic).

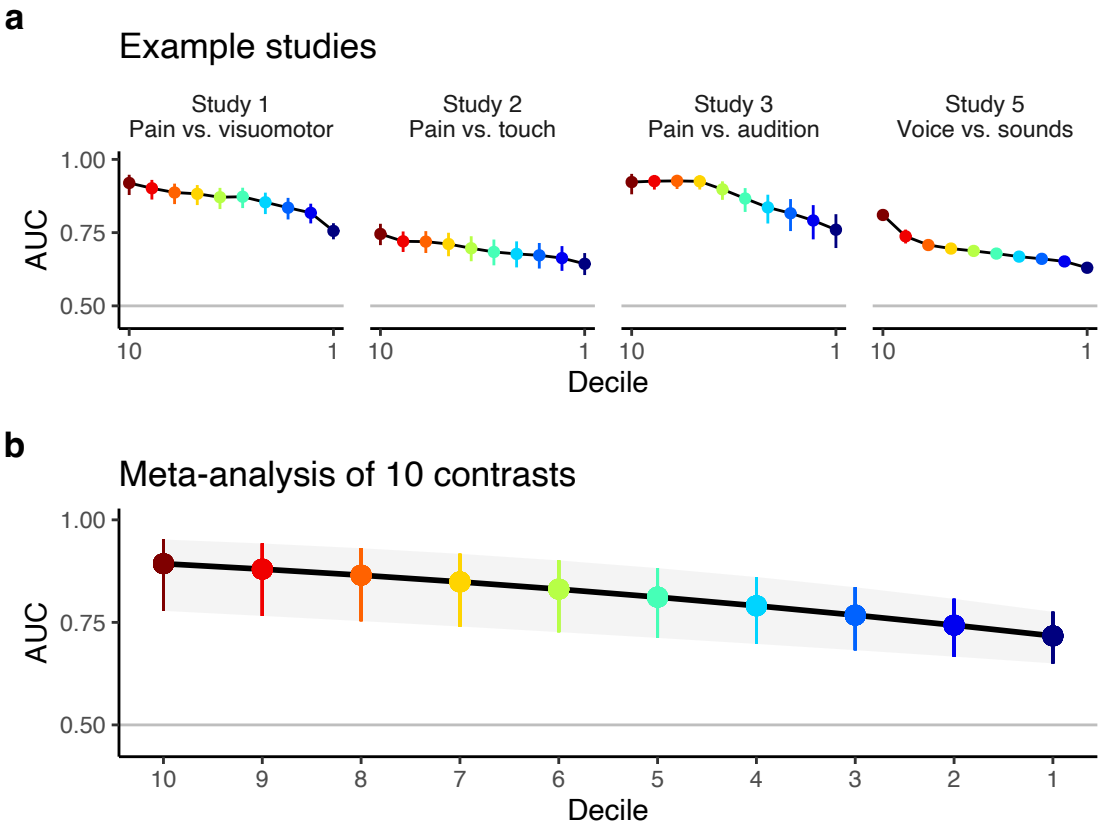

Figure 2. ***Voxels across the entire neocortex contain task-specific information.*** (**A**) Four example studies of neocortical decoding performance across all *t*-statistic deciles. Although all four studies contain task-specific information in all 10 deciles, the degree to which the tasks can be discriminated differs slightly between studies. Error bars indicate ± SE. (**B**) A mixed-effects meta-analysis across 10 contrasts reveals that all ten deciles can, on average, discriminate between tasks. The ability to discriminate between tasks implies the presence of task-specific information, meaning that even voxels with *t*-statistics close to zero (decile 1) contain marked task-specific information. Error bars indicate ± $CI_{95\%}$.

When decoding tasks from neuroimages, one cannot draw inferences about the causal role of the predictors (voxels) in the decoding outcome (task) [13]. There are many reasons why information may be present in acausal structures. For example, physiological and non-physiological noise may have task specificity [14]. We attempted to rule out such confounds by decoding tasks using white matter and cerebrospinal fluid (cf. grey matter). In general, decoding performance was poor in these regions (Fig S2), suggesting these negative controls contain less signal than the neocortex. In addition, white matter and cerebrospinal fluid decoding performance covaried substantially ($r = 0.7$), but they only weakly covaried with grey matter, implying vastly different signal sources between these structures (Fig S2). Finally, head motion artifact—another potential candidate of task-related noise that contaminates the BOLD signal—only minimally correlated with decoder responses; orthogonalizing the decoders' dot products to head motion only slightly decreased discrimination (maximum $\Delta AUC < 0.05$). Thus, our observed effects are unlikely to be attributable to task-specific, non-neural confounds.

To assess the general sensitivity of the decoders, we built them using different numbers of voxels and different amounts of added noise. Decoders with fewer voxels performed poorly and

were more sensitive to added noise (Fig S3). As one might expect, the ability to successfully decode using voxels with low *t*-statistics is principally attributable to the number of included voxels. There is a continuum of explanations as to why this might be the case. On one extreme, each voxel may contain a tiny amount of task-specific information. Integrating many small bits of information allows one to accumulate enough information to discriminate between tasks. On the other extreme, since *t*-statistics are derived empirically and thus may not be stable, "signal" voxels may mix with "noise" voxels, creating instability in our deciles. In this case, by sampling more voxels, we are more likely to capture "signal" voxels in our "noise" decile(s), which would enable successful decoding. Where our findings fall along this continuum remains to be determined.

Next, we tested whether information content is specific to the neocortex. Repeating our analyses in the subcortex and cerebellum revealed that information is present throughout both regions, even where *t*-statistics are approximately zero (Fig S4, Fig S5). Cerebellar information varied more between task pairs than the neocortex (three examples shown in Fig 3A). These results complement recent work by Nakai and Nishimoto [15], who used the subcortex and cerebellum to decode 103 cognitive tasks using a within-subject approach based on more complex models which were trained using statistical learning. In contrast, we used *t*-statistics from regional activity maps to decode across- rather than within-subjects. Our meta-analysis across contrasts showed that the performance of the subcortical and cerebellar decoders was slightly inferior to the performance of the neocortex-based decoders, even after controlling for the number of voxels (Fig 3B, Fig S6). Overall, we observed that subcortical and cerebellar structures contain widespread, task-specific information, evidencing that information spread is not restricted to the neocortex but is present across the entire human brain.

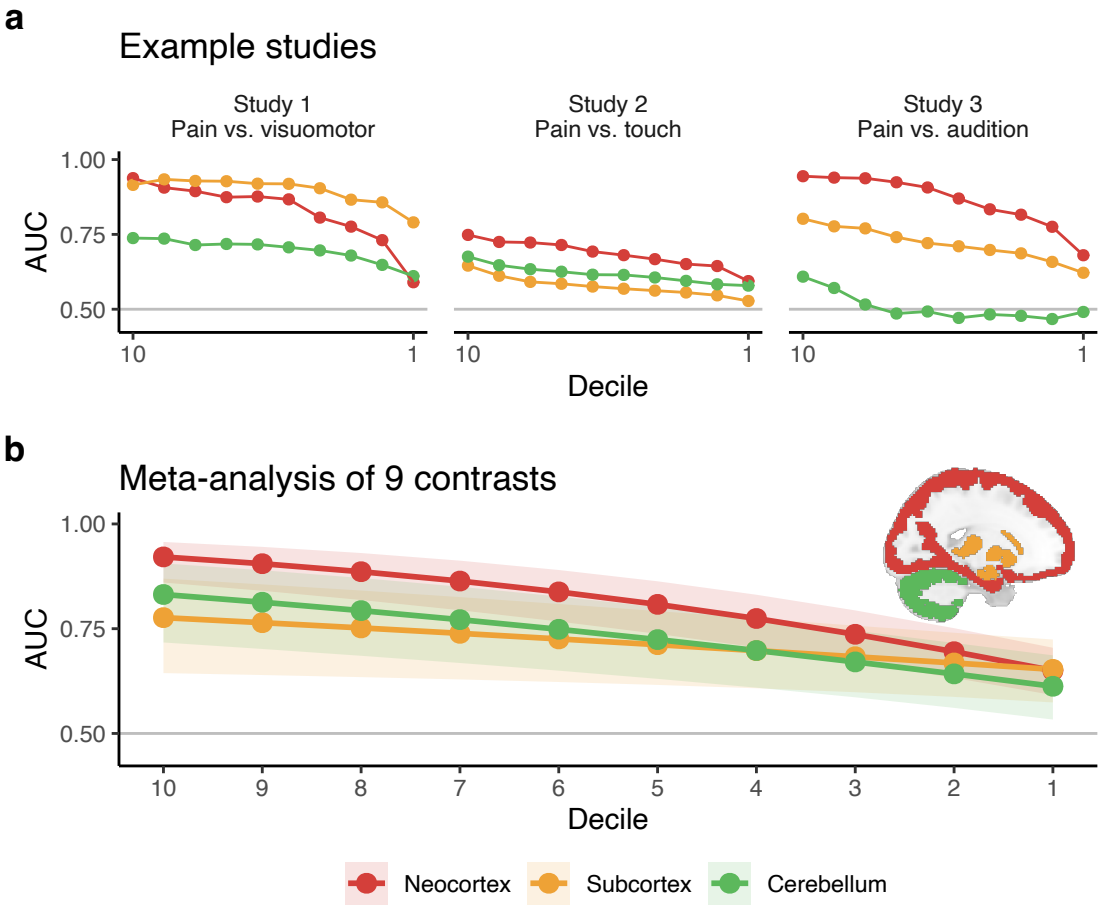

Figure 3. ***Task-relevant information is pervasively present throughout the subcortex and cerebellum but to a lesser extent than in the neocortex.*** (**A**) Three example studies demonstrate marked differences in regional task-specific information. In Study 1, the relative task-specific information in the neocortex shifts from being closer to the subcortex to the cerebellum. In Study 2, all three regions are similar, while in Study 3, there is a consistent pattern with the three areas being starkly different. (**B**) After controlling for the number of voxels in the neocortex, subcortex, and cerebellum, a meta-analysis across 9 contrasts (excluding Study 5) reveals that task-related information exists across all regions and deciles. Task-related information in the neocortex dominates for higher deciles, but this superiority vanishes in smaller deciles. Error ribbons indicate ± $CI_{95\%}$.

The notion of widespread, task-specific cortical dynamics is gaining traction across multiple fields of neuroscience. Human fMRI work demonstrates that 100 repetitions of the same task (three participants, 9–10 sessions over three months) can uncover neocortex-wide information [16]. Decoding studies, which rely on statistical learning approaches, evidence the existence of task-specific information outside of GLM areas [8-12]. Similarly, recent fMRI evidence from macaque monkeys demonstrates the presence of retinotopic tuning in cortical and subcortical regions remote from the visual cortex [17]. To complement fMRI evidence, wide-field calcium imaging and Neuropixels recordings in rodent models have been shown to capture mesoscopic neocortical and subcortical dynamics on a moment-by-moment basis, revealing brain-wide, task-specific activity across several cognitive domains [18-21]. Our results complement this prior work by demonstrating the presence of brain-wide, task-specific information in human brain fMRI and the ease with which this information can be extracted.

Understanding the nature of this brain-wide information is more challenging than identifying its existence. Recent work in mice demonstrates widespread cortical dynamics to be necessary for behavior—preventing local clusters of activation impairs performance, suggesting a functional rather than epiphenomenal role [19]. If activation across the entire brain is necessary for task performance, it is more likely that the information we detected across brain regions is complementary than redundant. In other words, different brain regions capture distinct properties of the task. However, we remain agnostic as to the role of this widespread information in conscious perception instead of simply being a task correlate. To address this, we will now link these findings to consciousness.

Widespread, task-specific information scales with consciousness

Neurophysiological theories of consciousness rely on brain-wide information sharing, posited to be necessary but not sufficient for consciousness [22]. Conceivably, the association between information sharing and consciousness [23] suggests that task-specific brain-wide information should attenuate with increasing levels of sedation. If information is not readily shared across the brain, it cannot be omnipresent. But how do states of consciousness interact with task-specific, brain-wide information content? To assess this, we analyzed a dataset in which individuals listened to an auditory stimulus (five-minute audio from a movie) while under different levels of sedation [6,7]. Since there was no task vector associated with the auditory stimulus, we averaged participants' auditory cortex time courses to serve as the task vector and used a separate resting-state scan as a negative control.

Consistent with our analyses above, information was omnipresent across the neocortex when participants were awake. However, this information degraded with increasing levels of sedation and was partially restored while recovering from sedation (Fig 4A and B). We performed a region-of-interest (ROI)-based analysis to complement the decile analysis. In the awake state, different regions exhibited distinct abilities to discriminate the task from resting-state, with the auditory cortex exhibiting the greatest discrimination. Moreover, the auditory cortex's task-specific signal was invariant to sedation level, but task-specific information degraded with deeper levels of sedation across all other ROIs (posterior, anterior, visual, and motor cortices) (Fig 4C). These results imply that brain-wide, task-specific information content is related to the perception rather than the encoding of the sensory stimulus.

The brain's modularity and interconnected functional and structural networks must underly our observed brain-wide distribution of information [24-27]. In particular, long-range connections and small-world networks to which they give rise provide a mechanism for efficient information sharing. The properties of these networks are thus likely to be critical for how information is communicated and captured across the brain. We elucidated the relevance of functional connectivity to our findings by examining intersubject functional connectivity (ISFC) using the auditory cortex as a seed. We quantified the temporal relationship between a participant's auditory cortex and all other participants', say, posterior cortex. ISFC decreased with more sedation, much like decoding performance (Fig 4D). However, within-subject functional connectivity remained largely unperturbed (Fig S7). Thus, ISFC seems to track sedation-dependent perceptual states.

Our results complement previous work in which transcranial magnetic stimulation (TMS) was used to induce electrical potentials that differentially propagated over the neocortex as a function of consciousness state [23]. Here, we leveraged passive sensory stimuli, but our findings are consistent: loss of consciousness degrades brain-wide information content via a downregulation in corticocortical information sharing. Therefore, consciousness seems a necessary condition for the

presence of widespread task-related cortical information. Practically, our simple decoding approach may be sufficient to identify neural correlates of consciousness using natural sensory stimuli without artificially stimulating the brain (cf. TMS [23]).

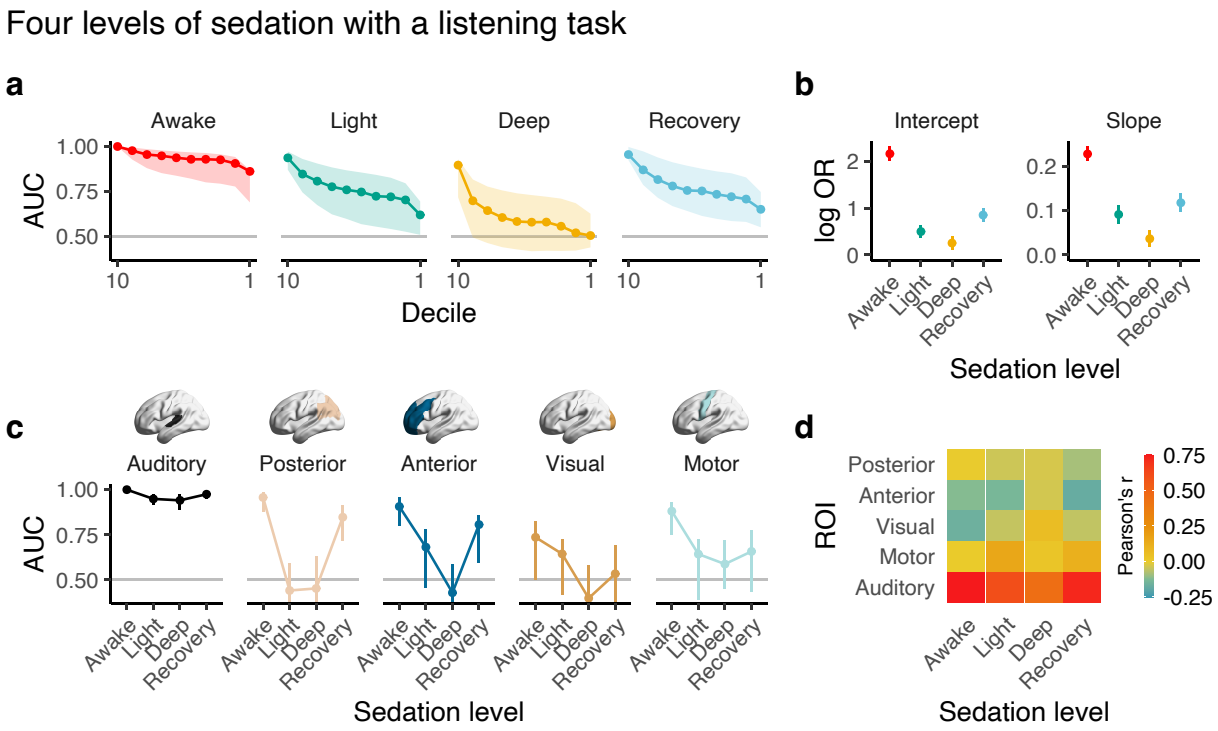

Figure 4: ***Sedation affects the neocortical distribution of task-relevant information in a region-specific manner.*** (**A**) Widespread cortical task-relevant information decreases with deeper levels of sedation, as indicated by the decrease in AUCs from awake to light to deep, along with the increase in AUC from deep to recovery. (**B**) Intercepts (defined by decile = 1) and slopes of the curves in (**A**) reveal stark decoding differences between levels of anesthesia. Awake's first decile (intercept) has the strongest performance of the different levels of sedation. Its slope (on the logit scale; OR = odds ratio) is also greatest since it is more difficult to improve performance as AUC → 1. (**C**) The primary and secondary sensory cortices (auditory cortex) retain information across sedation levels. In contrast, other cortical regions' information content drops off with increasing levels of sedation. (**D**) Task-specific intersubject functional connectivity decreases with deeper sedation. All error bars indicate ± $CI_{95\%}$.

Task-related confounds, such as head motion, are likely greater when individuals are awake. However, our sedation level-dependent findings were unrelated to head motion (Fig S8). Moreover, we observed similar results whether we used auditory cortex activity from the awake or deep anesthesia conditions (Fig S9). This latter point is remarkable: auditory cortex activity with deep sedation is arguably "purer" than that in the awake condition, as higher-level processes and feedback loops will not modulate it, and similarly, head motion should be negligible. The generalizability of our findings across participants and levels of sedation reinforces that our findings represent consciousness-related neural information rather than task-related artifacts.

The neocortex is not the only neural structure involved in consciousness. Much has been discussed and debated regarding the role of the thalamus, other subcortical structures, and the cerebellum [22]. Like the neocortex, the subcortex's task-specific information content demonstrates a dose-dependence on the level of sedation. Similarly, so does the cerebellum, although its dependence on sedation level is more complex (Fig S10). We should emphasize that the presence

of brain-wide information reflects the state of consciousness rather than demonstrating its mechanism(s).

Concluding Remarks

Our results render the brain-wide presence of information indisputable and elucidate the ease with which this information can be decoded, even in brain regions that mass-univariate analyses suggest are approximately orthogonal to the tasks being studied. We also show that the ubiquity of this information is not without bounds—consciousness is itself a necessary condition for the brain-wide spread of task-related information. Many of the tasks that we compared are trivially simple, only involving passive sensory stimuli; yet, related information is spread across the entire brain. Neurocognitively, our results imply that perceptual states engage the entire brain. We speculate that the details of the distribution of information may define the nuanced properties of perception; for example, the edibility of the oblong, yellow object. Finally, these results strongly challenge the notion of localization of information in the brain without precluding regional specialization of function. For example, although language-specific information can be uncovered across the entire neocortex [28], the functional role of Broca's area is incontrovertible [29]. Unraveling the unique contribution of diverse brain regions to perceptual states requires future investigation, including the necessity of some and the sufficiency of others, including methods beyond traditional linear, univariate analyses of brain activity.

MATERIALS AND METHODS

Datasets

Six datasets were used in this paper; all are part of published studies and were either provided by their authors (Datasets 1–4) or downloaded from public repositories (Datasets 5 & 6). Datasets 1–4 consist of voxel-wise, whole brain, task-dependent general linear model (GLM) analysis activation maps (ftp://openpain.org/LimitsDecoding/Data/Beta_maps/). Datasets 5 & 6 consist of BOLD timeseries which were processed using standard fMRI pre- and post-processing methods described below.

Dataset 1

Fifteen (15), right-handed, adult subjects (mean age: 35 ± 11 years, 7 females). Subjects had no history of pain, psychiatric, or neurological disorders. fMRI data were collected while subjects received thermal stimuli across 3 temperatures: 47, 49, and 51 degrees Celsius. Subjects continuously rated, using a finger span device [30,31], their pain from 0 (not painful) to 100 (worst imaginable pain) ("pain rating" task). A control scan was performed while subjects used the finger span device to track a moving bar projected on the screen ("visual rating" task; moving bar replicated for each subject the specific pain rating task temporal pattern). The dataset includes one GLM beta map per subject per stimulus type. The dataset was previously described in [32].

Dataset 2

Fifty-one (51) healthy, right-handed, adult subjects (age = 24 ± 2 years, 34 females). Subjects had no history of brain injuries, pain disorders, or psychiatric or neurological diseases. fMRI data was collected while subjects received painful heat stimuli on the right foot dorsum using a CO_2 laser, as well as tactile stimuli to the same area using electrical stimulation. Stimuli were not delivered at the same time. Perceived intensities were recorded for every stimulus and only the stimuli with matched perceived intensity for painful heat and touch were selected for GLM analysis. The dataset includes one activation map per subject per stimulus modality – painful heat and touch. The dataset was previously described in [4,33].

Dataset 3

Fourteen (14) healthy, right-handed, adult subjects (age = 20–36 years old, 6 females). fMRI data were collected while subjects received painful heat stimuli on the right foot dorsum using a CO_2 laser, tactile stimuli to the same area using electrical stimulation, visual stimuli using a white disk presented above the right foot, and auditory stimuli delivered via pneumatic earphones. Stimuli were not delivered at the same time. Perceived intensities were recorded for every stimulus and only the stimuli with matched perceived intensity across the four modalities were selected for GLM analysis. The dataset includes one activation map per subject per stimulus modality – painful heat, tactile, auditory, and visual. The dataset was previously described and published in [4].

Dataset 4

Thirty-three (33) healthy, right-handed, adult subjects (age = 28 ± 9 years, 22 females). Subjects had no history of pain, psychiatric, or neurological disorders. fMRI data was collected while subjects received thermal stimuli that varied in one-degree Celsius increments across six temperatures from 44.3 up to 49.3°C. Subjects then evaluated each stimulus as warm, and scored it from 0, not perceived up to 99, about to become painful, or as painfully hot, and scored it from 100 (no pain) to 200 (worst imaginable pain). The dataset includes an average GLM activation map per subject per stimulus temperature, as well as the corresponding average stimulus ratings. When this dataset was applied dichotomously (pain vs. no pain), we averaged the brain activity maps from the painful and nonpainful conditions; we omitted subjects who had fewer than two brain activity maps for each condition, resulting in 29 subjects for dichotomous ratings. The dataset was previously described in [3,34].

Dataset 5

Two-hundred thirteen (213) healthy, adult subjects (age = 24 ± 7 years, 101 females). Subjects had no history of physical or mental health conditions. fMRI data were collected while subjects performed a voice localizer task. Forty blocks of vocal sounds (20) and non-vocal sounds (20) interspersed with periods of silence were presented while the subjects lay silent and passively listened with their eyes closed in the scanner. This dataset was previously described in [35]. Raw fMRI data were downloaded from OpenNeuro (ds000158). We used minimal pre-processing for this study which was performed using the FMRIB 5.0.8 software library (FSL) [36], MATLAB2018a and in-house scripts. The following steps were performed: motion correction, intensity normalization, nuisance regression of 6 motion vectors, signal-averaged overall voxels of the eroded white matter and ventricle region, and global signal of the whole brain, and band-pass filtering (0.008–0.1 Hz) by applying a 4th-order Butterworth filter. All pre-processed fMRI data were registered to the 2×2×2 mm MNI152 template using a two-step procedure: the mean of preprocessed fMRI data was registered with a 7-degrees-of-freedom affine transformation (x, y, z, α, β, γ, and scale factor k) to its corresponding T_1 brain (FLIRT); next, transformation parameters were computed by nonlinearly registering individual T_1 brains to the MNI152 template (FNIRT). Combining the two transformations yielded a mapping from the preprocessed fMRI data to standard space. Task-related activation maps (vocal vs. silence, and non-vocal vs. silence) were derived from a whole-brain GLM regression analysis using the FMRIB Software Library (FSL) [36-38].

Dataset 6

Seventeen (17) healthy, adult subjects (4 women; age = 24 ± 5 years) partook in this study, which involved listening to a natural stimulus (5 min plot-driven audio story) and resting-state (first 5 min of 8 min scan) while under different levels of anesthesia [6,7]. Sedation levels under propofol were determined by the Ramsey scale (awake, no propofol; light sedation, Ramsey = 3; deep sedation, Ramsey = 5; recover, Ramsey = 2, approximately 11 minutes after cessation of propofol) [6]. We used minimal pre-processing for this study which was performed using the FMRIB 5.0.8 software library (FSL) [36], MATLAB2018a and in-house scripts. The following steps were performed: motion correction, intensity normalization, nuisance regression of 6 motion vectors, signal-averaged overall voxels of the eroded white matter and ventricle region, and global signal of the whole brain, and band-pass filtering (0.008–0.1 Hz) by applying a 4th-order Butterworth filter. All pre-processed fMRI data were registered to the 2×2×2 mm MNI152 template using a two-step procedure: the mean of preprocessed fMRI data was registered with a 7-degrees-of-freedom affine transformation (x, y, z, α, β, γ, and scale factor k) to its corresponding T_1 brain (FLIRT); next, transformation parameters were computed by nonlinearly registering individual T_1 brains to the MNI152 template (FNIRT). Combining the two transformations yielded a mapping from the preprocessed fMRI data to standard space. Task-related activation maps (vocal vs. silence, and non-vocal vs. silence) were derived from a whole-brain GLM regression analysis using R.

Decoder Construction and Evaluation

Brain activity maps were masked to include only neocortical grey matter voxels using the Harvard-Oxford neocortical mask thresholded at 0.5. For each contrast, we performed a voxel-wise paired t-test using two brain activity maps from each subject, resulting in a t-statistic for each voxel in the grey matter. The t-statistic map was then binned into deciles by $|t|$—decile 10 contained the highest absolute value t-statistics (the "most significant") and decile 1 contained the lowest absolute value t-statistics (the "least significant"). These deciled t-statistic maps served as our decoders.

We evaluated the decoders ($\mathbf{D} \in \mathbb{R}^{p\times 10}$) by multiplying them with the brain activity maps of interest ($\mathbf{B}_{\boldsymbol{I}} \in \mathbb{R}^{n\times p}$) and no interest ($\mathbf{B}_{\boldsymbol{NI}} \in \mathbb{R}^{n\times p}$), for p voxels and n subjects. This resulted in two matrices of dot products between the decoders and brain activity maps: one matrix of dot products from the activity maps of interest ($\mathbf{R}_{\boldsymbol{I}} = \mathbf{B}_{\boldsymbol{I}}\mathbf{D}$) and one matrix of dot products from the activity maps of no interest ($\mathbf{R}_{\boldsymbol{NI}} = \mathbf{B}_{\boldsymbol{NI}}\mathbf{D}$). The columns of $\mathbf{R}_{\boldsymbol{I}}$ and $\mathbf{R}_{\boldsymbol{NI}}$ were then compared to calculate an AUC via the Mann-Whitney U-statistic ($AUC = U_1/n^2$). That is, column 1 in $\mathbf{R}_{\boldsymbol{I}}$ was compared with column 1 in $\mathbf{R}_{\boldsymbol{NI}}$, column 2 in $\mathbf{R}_{\boldsymbol{I}}$ was

compared with column 2 in $\mathbf{R}_{NI}$, and so on for all 10 columns, producing 10 AUCs—one for each decile. In doing so, we treated the subjects as dependent for decoder training (paired *t*-test) but independent for testing.

We constructed and tested all decoders using the 0.632+ bootstrap method with 100 replicates, which provides unbiased estimates of out-of-sample performance [39]. Briefly, the 0.632+ bootstrap was performed as follows:

1. Train and test a model using the original sample. Let the resulting AUC be called the "apparent" AUC, $\hat{\theta}^{app}$.
2. Generate b bootstrap samples by resampling the original sample with replacement. Note, each bootstrap sample contains approximately $1 - {}^{1}\!/_{e} = 63.2\%$ of the original sample. For each of the b bootstrap samples, we train the model on the bootstrap sample and test the model on the ~36.8% of individuals not part of the bootstrap sample. Let this AUC estimate be the "leave-one-out" (out-of-sample) bootstrap AUC, $\hat{\theta}_i^{boot}$.
3. Average the resulting out-of-sample bootstraps, $\hat{\theta}_{\bullet}^{boot} = \frac{1}{b}\sum_{i=1}^{b} \hat{\theta}_i^{boot}$.
4. Obtain the 0.632+ estimate.
    a. Calculate the relative overfitting rate,

$$\hat{R} = \begin{cases} 1, & \hat{\theta}_{\bullet}^{boot} \leq 0.5 \\ \left(\hat{\theta}^{app} - \hat{\theta}_{\bullet}^{boot}\right) / \left(\hat{\theta}^{app} - 0.5\right), & \hat{\theta}^{app} > \hat{\theta}_{\bullet}^{boot}, \\ 0, & \text{otherwise} \end{cases}$$

where 0.5 indicates no information in the decoder.

    b. Calculate the weight for adjusting the 0.632 estimate,

$$\hat{w} = \frac{0.632}{1 - 0.368\hat{R}}.$$

    c. Calculate the 0.632+ estimate,

$$\hat{\theta}^{0.632+} = \left(1 - \hat{w}\right) \cdot \hat{\theta}^{app} + \hat{w} \cdot \hat{\theta}_{\bullet}^{boot}.$$

Note, we did not adjust $\hat{\theta}_{\bullet}^{boot}$ with $\max\left(\hat{\theta}_{\bullet}^{boot}, 0.5\right)$ in 4c as is common [39,40], since this would create a floor effect such that $\hat{\theta}^{0.632+} \geq 0.5$, which would downwardly bias our variance estimates in the next step. This results in $\hat{\theta}^{0.632+}$ estimates that are identical to estimates with the adjustment when $\hat{\theta}^{0.632+} > 0.5$, but allows our estimates to dip below chance (AUC = 0.5) since it removes the floor effect.

Variances and covariances of the AUCs were estimated using a nested bootstrap with 500 replicates in the outer loop and 100 replicates in the inner loop [40], totaling 500×100 = 50,000 replicates of each study. All inner and outer bootstraps were performed on the subject level. This sampling was carried out on Northwestern University's High Performance Computing clusters (Quest), which took ~12 hours to complete using 50 cores.

Meta-analysis

We performed a single-paper meta-analysis to consolidate our results [41]. First, all AUCs were "squeezed" or shrunken toward 0.5 to avoid boundary effects [42],

$$\hat{\theta}^{*} = \frac{(n-1) \cdot \hat{\theta}^{0.632+} + 0.5}{n},$$

where n was the total number of brains used (i.e., twice the number of participants). Next, the 0.632+ bootstrap-estimated AUCs and their bootstrapped replicates were logit-transformed, and the logit-transformed bootstrap replicates were used to generate a 100×100 variance-covariance matrix of sampling errors. The logit-transformed AUCs were used as the response variable in a multivariate, multilevel linear meta-regression [43]. This allowed for within-study dependence to be properly accounted for, including the dependence between deciles in a single contrast (e.g., decile 1 and decile 2 in Study 1) and the dependence between contrasts in Study 3 (e.g., decile 1 in touch vs. pain and decile 2 in visuomotor vs. pain). We were principally interested in the effect of decile on discrimination performance; we treated decile continuously and used it as a linear moderator (fixed-effect). Similarly, decile was treated continuously in the random-effect specification, wherein contrasts were nested within studies.

Perturbations

Noise

Since voxels with low signal-to-noise ratios (i.e., low t-statistics) were capable of decoding, we aimed to evaluate this finding's boundary conditions. Each brain activity map contains a correlation coefficient r_i for each voxel i, along with a t-statistic t_i. We started with a brain of t-statistics, to which we added Gaussian noise ($\mathcal{N}(0, c|t_i|)$, where $c \in \{0,1,2,3,4,5\}$). This procedure ensured that the noise added to each voxel was proportional to its signal-to-noise ratio to avoid biasing the regions with high signal. The t-statistics with added noise were then converted to Pearson's r, on which we performed the decoding. Noise was added within each 0.632+ bootstrap replicate such that the resulting AUCs were averaged over 500 iterations (as opposed to 100 for other analyses) of added noise.

Voxel Sampling

Since the ability to decode with low signal-to-noise voxels likely arises from integrating over so many small bits of information, we randomly subsampled voxels. The full decoders contained approximately 10,735 voxels per decile, which could come from anywhere within our grey matter mask. We built and assessed decoders by subsampling the brain activity maps, such that the resulting decile-based decoders contained 100, 250, 500, 1000, 2500, 5000, 7500, and 10735 voxels each. Voxels were sampled within each 0.632+ bootstrap replicate such that the resulting AUCs were averaged over 500 iterations (as opposed to 100 for other analyses) of sampled voxels.

Anatomical Specificity

Neocortex, Subcortex, and Cerebellum

Neocortical, subcortical (thalamus, striatum, hippocampus, and amygdala), and cerebellar grey matter voxels were extracted from each brain activity map. The neocortical grey matter mask contained 112,651 voxels; the subcortical mask contained 6,882 voxels; and the cerebellar cortex mask contained 17,142 voxels. Since decoding power is sensitive to the number of voxels, we randomly subsampled 6,882 voxels (or fewer for studies that were further masked) from each mask to control for voxel number. This subsampling was completed within each 0.632+ inner bootstrap replicate.

Neocortical Grey Matter, White Matter, and Cerebral Spinal Fluid

Neocortical grey matter (GM), white matter (WM), and cerebral spinal fluid (CSF) voxels were masked using the Harvard-Oxford atlas with conservative thresholds: 112,651 for GM; 61,324 for WM; and 1,926 for CSF. Within each study, we controlled for the number of voxels by resampling 1,925 voxels (since 1,926 < 61,324 < 112,651) from GM and WM within each bootstrap run.

Anesthesia Decoders

The anesthesia dataset employed a naturalistic audio stimulus and thus does not have a task vector associated with it. Moreover, this was the only task performs. As such, we compared each anesthesia level's task (naturalistic listening) to resting state. To facilitate this, we used the average auditory cortex activity

from the training sample as the task vector. Of course, this analysis is circular within the training sample, but because the training sample's brain activity was used as the vector in the testing sample and decoding was assessed based on the resulting brain activity maps, the testing is not circular. To extract the auditory cortex vector, we defined a region of interest (ROI) based on the Neurosynth association map for "auditory", which was thresholded using a *z*-score of 12.

Decile Decoders

Decile-based decoders for the anesthesia dataset were created in a similar manner to the other datasets. To summarize the performance within each decile, we fit a single generalized least squares model on the logit-transformed AUCs from all anesthesia states, $\mathbf{y}$. To do so, all AUCs were "squeezed" towards 0.5 like they were for the meta-analysis. Our weight matrix, $\mathbf{W}$, was defined as the inverse of the variance-covariance matrix of the logit-transformed bootstrapped AUCs, $\hat{\mathbf{\Sigma}}$. The parameter estimates, $\hat{\boldsymbol{\beta}}$, and their standard errors were calculated as

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{y}$$

$$\mathrm{SE}\left[\hat{\boldsymbol{\beta}}\right] = \sqrt{\mathrm{diag}\left((\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}\hat{\mathbf{\Sigma}}\mathbf{W}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1}\right)},$$

where $\mathbf{X}$ is the design matrix,

$$\mathbf{X} = [1 \quad x^{\mathrm{decile}}] \otimes [x^{\mathrm{awake}} \quad x^{\mathrm{light}} \quad x^{\mathrm{deep}} \quad x^{\mathrm{recover}}].$$

This differs from standard weighted least squares (a diagonal weight matrix) but has more favorable properties since it accounts for covariation.

Region of Interest Decoders

We created posterior cortex, anterior cortex, visual cortex, and motor cortex ROIs using the Harvard-Oxford neocortical atlas, thresholded at 25%. The posterior cortex was specified as areas 20–22; the anterior cortex, areas 1, 4, and 5; the visual cortex, areas 36, 40, and 48; and the motor cortex, area 7. In addition, we used the same auditory cortex ROI as described above. Again, the averaged auditory ROI time course from the training sample was used as the task vector. A decoder was created using the *t*-statistics from each ROI (without deciles), which was fit and tested using the same approach as the decile decoders.

Functional Connectivity

Pearson correlation coefficients were used to calculate within- and inter-subject (ISFC) functional connectivity between the auditory cortex and the other ROIs, during both the auditory task and rest within each level of anesthesia. Within-subject functional connectivity was calculated by averaging the time course between all voxels within each ROI, calculating the correlation between the auditory ROI and all other ROIs, converting from Pearson's *r* to Fisher's *z*, averaging across participants, and then converting back to Pearson's *r*. ISFC was calculated similarly to previous work [44]. When calculating subject 1's ISFC, we correlated subject 1's auditory cortex time course with the average time course of, for example, posterior cortex from subjects 2–*n*. This was repeated for all subjects and the resulting auditory cortex-posterior cortex ISFCs were averaged using Fisher's *z* to obtain the final estimate of the auditory cortex-posterior cortex ISFC. Again, this was repeated for the anterior, visual, and motor cortex; ISFC was also measured between auditory cortices across all subjects.

References

1 Edelman, G. M. & Tononi, G. *A universe of consciousness : how matter becomes imagination*. 1st edn, (Basic Books, 2000).

2 Pernet, C. R. *et al.* The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. *Neuroimage* **119**, 164-174, doi:10.1016/j.neuroimage.2015.06.050 (2015).

3 Wager, T. D. *et al.* An fMRI-based neurologic signature of physical pain. *N Engl J Med* **368**, 1388-1397, doi:10.1056/NEJMoa1204471 (2013).

4 Liang, M., Su, Q., Mouraux, A. & Iannetti, G. D. Spatial Patterns of Brain Activity Preferentially Reflecting Transient Pain and Stimulus Intensity. *Cereb Cortex* **29**, 2211-2227, doi:10.1093/cercor/bhz026 (2019).

5 Baliki, M. N., Geha, P. Y. & Apkarian, A. V. Parsing pain perception between nociceptive representation and magnitude estimation. *J Neurophysiol* **101**, 875-887, doi:10.1152/jn.91100.2008 (2009).

6 Kandeepan, S. *et al.* Modeling an auditory stimulated brain under altered states of consciousness using the generalized Ising model. *Neuroimage* **223**, 117367, doi:10.1016/j.neuroimage.2020.117367 (2020).

7 Naci, L. *et al.* Functional diversity of brain networks supports consciousness and verbal intelligence. *Sci Rep* **8**, 13259, doi:10.1038/s41598-018-31525-z (2018).

8 Kumar, S., Ellis, C. T., O'Connell, T. P., Chun, M. M. & Turk-Browne, N. B. Searching through functional space reveals distributed visual, auditory, and semantic coding in the human brain. *PLoS Comput Biol* **16**, e1008457, doi:10.1371/journal.pcbi.1008457 (2020).

9 Cox, C. R. & Rogers, T. T. Finding Distributed Needles in Neural Haystacks. *J Neurosci* **41**, 1019-1032, doi:10.1523/JNEUROSCI.0904-20.2020 (2021).

10 Rish, I. & Cecchi, G. A. Holographic brain: Distributed versus local activation patterns in fMRI. *IBM Journal of Research and Development* **61**, 3:1-3:9, doi:10.1147/jrd.2017.2648699 (2017).

11 Rish, I., Cecchi, G. A., Heuton, K., Baliki, M. N. & Apkarian, A. V. in *Medical Imaging 2012: Image Processing* (eds David R. Haynor & Sébastien Ourselin) (2012).

12 Mohr, H., Wolfensteller, U., Frimmel, S. & Ruge, H. Sparse regularization techniques provide novel insights into outcome integration processes. *Neuroimage* **104**, 163-176, doi:10.1016/j.neuroimage.2014.10.025 (2015).

13 Jabakhanji, R. *et al.* Limits of Decoding Mental States with fMRI. *Cortex*, doi:10.1016/j.cortex.2021.12.015 (2022).

14 Liu, T. T. Noise contributions to the fMRI signal: An overview. *Neuroimage* **143**, 141-151, doi:10.1016/j.neuroimage.2016.09.008 (2016).

15 Nakai, T. & Nishimoto, S. Preserved representations and decodability of diverse cognitive functions across the cortex, cerebellum, and subcortex. *bioRxiv*, 2021.2012.2009.471939, doi:10.1101/2021.12.09.471939 (2021).

16 Gonzalez-Castillo, J. *et al.* Whole-brain, time-locked activation with simple tasks revealed using massive averaging and model-free analysis. *Proc Natl Acad Sci U S A* **109**, 5487-5492, doi:10.1073/pnas.1121049109 (2012).

17 Klink, P. C., Chen, X., Vanduffel, W. & Roelfsema, P. R. Population receptive fields in nonhuman primates from whole-brain fMRI and large-scale neurophysiology in visual cortex. *Elife* **10**, doi:10.7554/eLife.67304 (2021).

18 Ren, C. & Komiyama, T. Characterizing Cortex-Wide Dynamics with Wide-Field Calcium Imaging. *J Neurosci* **41**, 4160-4168, doi:10.1523/JNEUROSCI.3003-20.2021 (2021).

19 Pinto, L. *et al.* Task-Dependent Changes in the Large-Scale Dynamics and Necessity of Cortical Regions. *Neuron* **104**, 810-824 e819, doi:10.1016/j.neuron.2019.08.025 (2019).

20 Stringer, C. *et al.* Spontaneous behaviors drive multidimensional, brainwide activity. *Science* **364**, 255, doi:10.1126/science.aav7893 (2019).

21 Gilad, A. & Helmchen, F. Spatiotemporal refinement of signal flow through association cortex during learning. *Nat Commun* **11**, 1744, doi:10.1038/s41467-020-15534-z (2020).

22 Tononi, G. & Koch, C. Consciousness: here, there and everywhere? *Philos Trans R Soc Lond B Biol Sci* **370**, doi:10.1098/rstb.2014.0167 (2015).

23 Casali, A. G. *et al.* A theoretically based index of consciousness independent of sensory processing and behavior. *Sci Transl Med* **5**, 198ra105, doi:10.1126/scitranslmed.3006294 (2013).

24 Eguiluz, V. M., Chialvo, D. R., Cecchi, G. A., Baliki, M. & Apkarian, A. V. Scale-free brain functional networks. *Phys Rev Lett* **94**, 018102, doi:10.1103/PhysRevLett.94.018102 (2005).

25 Knosche, T. R. & Tittgemeyer, M. The role of long-range connectivity for the characterization of the functional-anatomical organization of the cortex. *Front Syst Neurosci* **5**, 58, doi:10.3389/fnsys.2011.00058 (2011).

26 Sporns, O., Tononi, G. & Edelman, G. M. Connectivity and complexity: the relationship between neuroanatomy and brain dynamics. *Neural Netw* **13**, 909-922, doi:10.1016/s0893-6080(00)00053-8 (2000).

27 Sporns, O., Tononi, G. & Edelman, G. M. Theoretical neuroanatomy: relating anatomical and functional connectivity in graphs and cortical connection matrices. *Cereb Cortex* **10**, 127-141, doi:10.1093/cercor/10.2.127 (2000).

28 Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E. & Gallant, J. L. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* **532**, 453-458, doi:10.1038/nature17637 (2016).

29 Broca, P. Sur le siège de la faculté du langage articulé. *Bulletins de la Société d'anthropologie de Paris* **6**, 377-393, doi:10.3406/bmsap.1865.9495 (1865).

30 Apkarian, A. V., Krauss, B. R., Fredrickson, B. E. & Szeverenyi, N. M. Imaging the pain of low back pain: functional magnetic resonance imaging in combination with monitoring subjective pain perception allows the study of clinical pain states. *Neuroscience Letters* **299**, 57-60 (2001).

31 Baliki, M. N. *et al.* Chronic pain and the emotional brain: Specific brain activity associated with spontaneous fluctuations of intensity of chronic back pain. *J Neurosci* **26**, 12165-12173, doi:10.1523/Jneurosci.3576-06.2006 (2006).

32 Baliki, M. N., Geha, P. Y. & Apkarian, A. V. Parsing Pain Perception Between Nociceptive Representation and Magnitude Estimation. *J Neurophysiol* **101**, 875-887, doi:10.1152/jn.91100.2008 (2009).

33 Su, Q. *et al.* Brain regions preferentially responding to transient and iso-intense painful or tactile stimuli. *Neuroimage* **192**, 52-65, doi:10.1016/j.neuroimage.2019.01.039 (2019).

34 Woo, C. W., Roy, M., Buhle, J. T. & Wager, T. D. Distinct brain systems mediate the effects of nociceptive input and self-regulation on pain. *PLoS Biol* **13**, e1002036, doi:10.1371/journal.pbio.1002036 (2015).

35 Pernet, C. R. *et al.* The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. *Neuroimage* **119**, 164-174 (2015).

36 Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W. & Smith, S. M. Fsl. *Neuroimage* **62**, 782-790, doi:10.1016/j.neuroimage.2011.09.015 (2012).

37 Smith, S. M. *et al.* Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* **23 Suppl 1**, S208-219, doi:10.1016/j.neuroimage.2004.07.051 (2004).

38 Woolrich, M. W. *et al.* Bayesian analysis of neuroimaging data in FSL. *Neuroimage* **45**, S173-186, doi:10.1016/j.neuroimage.2008.10.055 (2009).

39 Efron, B. & Tibshirani, R. Improvements on Cross-Validation: The 632+ Bootstrap Method. *Journal of the American Statistical Association* **92**, 548-560, doi:10.1080/01621459.1997.10474007 (1997).

40 Noma, H., Shinozaki, T., Iba, K., Teramukai, S. & Furukawa, T. A. Confidence intervals of prediction accuracy measures for multivariable prediction models based on the bootstrap-based optimism correction methods. *Stat Med*, doi:10.1002/sim.9148 (2021).

41 McShane, B. B. & Böckenholt, U. Single Paper Meta-analysis: Benefits for Study Summary, Theory-testing, and Replicability. *Journal of Consumer Research*, doi:10.1093/jcr/ucw085 (2017).

42 Smithson, M. & Verkuilen, J. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods* **11**, 54-71, doi:10.1037/1082-989x.11.1.54 (2006).

43 Viechtbauer, W. Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software* **36**, doi:10.18637/jss.v036.i03 (2010).

44 Simony, E. *et al.* Dynamic reconfiguration of the default mode network during narrative comprehension. *Nat Commun* **7**, 12141, doi:10.1038/ncomms12141 (2016).