

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	We used the public datasets, and no software was used on our side to collect data.
Data analysis	python 3.10 pandas 1.5.3 scipy 1.10.0 pytorch 1.13.1 dalle-pytorch 1.6.4 scikit-learn 1.6.1 numpy 1.24.2 nilearn 0.10.0 nibabel 5.3.2 matplotlib 3.6.3 seaborn 0.13.2 clip code: https://github.com/openai/CLIP

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

For mouse single neuron analysis, we used the Neuropixels Visual Coding Neuropixels and two-photon calcium imaging datasets from the Allen Institute, accessible at <https://portal.brain-map.org/overview>. For resting-state human fMRI analysis, we used HCP-7T dataset from <https://www.humanconnectome.org>. We shared our EEG-fMRI dataset at <https://openneuro.org/datasets/ds003768>. For task human fMRI analysis, we used NSD dataset available at <https://naturalscenesdataset.org>. The Python code that produced the major results of this paper is available at <https://zenodo.org/records/14230834>. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Simultaneous EEG-fMRI Resting-state Dataset: Resting-state fMRI data were collected simultaneously with electroencephalography (EEG) for 27 subjects age: 22.1 ± 3.1 years; 14 females and 13 males.
Natural Scenes Dataset (NSD): Eight participants (two males and six females; age range, 19–32 years).
Human Connectome Project (HCP): We utilized the WU-UMinn HCP 7T dataset, a subset of the HCP S1200 release (28), comprising 7T fMRI data from 184 subjects within the age range of 22 to 35, including 112 females and 72 males.
This study is focused on a general phenomenon that is conserved across mice and human, and the sex/gender isn't a factor that we are trying to address here.

Reporting on race, ethnicity, or other socially relevant groupings

No applicable. This study reports a general phenomenon with focusing on race, ethnicity, and other socially relevant groups.

Population characteristics

We used the following three human datasets:
Simultaneous EEG-fMRI Resting-state Dataset: Resting-state fMRI data were collected simultaneously with electroencephalography (EEG) for 27 subjects age: 22.1 ± 3.1 years; 14 females and 13 males.
Natural Scenes Dataset (NSD): Eight participants (two males and six females; age range, 19–32 years).
Human Connectome Project (HCP): We utilized the WU-UMinn HCP 7T dataset, a subset of the HCP S1200 release (28), comprising 7T fMRI data from 184 subjects within the age range of 22 to 35, including 112 females and 72 males.

Recruitment

We used the public datasets. The detailed information regarding each one, including the recruitment, are available in their websites/papers, which we have provided in our paper. No any potential self-selection bias or other biases that we're aware of could potential impact the reported results

Ethics oversight

All the public datasets we used have obtained ethical approvals from their corresponding institutional review board (IRB). The use of these data in this study has been reviewed and approved by the Pennsylvania State University IRB and also followed the data usage agreements with data providers.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For statistical analysis, the sample size is defined by the actual number of subjects, animals, or sessions included in the analysis.
Data exclusions	No data were intentionally excluded from our current study.
Replication	We compared the results from different perspective across different datasets. The results were reproducible across datasets and analyses were cross-validated.
Randomization	This was an observational study without any intervention, and thus no randomization was performed.
Blinding	This was a study that retrospectively analyzed available observational neuroimaging and clinical data without any intervention. No blinding

was performed.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	Allen Neuropixel dataset: three transgenic lines (n = 8 Pvalb-IRES-Cre x Ai32, n = 12 Sst-IRES-Cre x Ai32, and n = 8 Vip-IRES-Cre x Ai32) Allen Two-photon Calcium Imaging dataset: 13 transgenic lines (n=14 Vip-IRES-Cre, n=25 Slc17a7-IRES2-Cre, n=9 Ntsr1-Cre_GN220, n=25 Sst-IRES-Cre, n=32 Cux2-CreERT2, n=19 Rorb-IRES2-Cre, n=20 Nr5a1-Cre, n=17 Emx1-IRES-Cre, n=21 Rbp4-Cre_KL100, n=8 Fezf2-CreER, n=6 Scnn1a-Tg3-Cre, n=7 Pvalb-IRES-Cre, n=3 Tlx3-Cre_PL56) UCL Mice Neuropixel Recording Dataset: Multiple genotypes were used, including n=2 tetO-G6s x CaMK-tTA, n=2 Vglut1-IRES2-Cre-D, n=3 RCL-GCaMP6f x VGlut-Cre, n=1 Snap25-GCaMP6s
Wild animals	Allen Neuropixel dataset: 30 C57BL/6J wild-type mice Allen Two-photon Calcium Imaging dataset: N/A UCL Mice Neuropixel Recording Dataset: 2 C57BL/6J wild-type mice
Reporting on sex	Allen Neuropixel dataset: 13 female and 45 male Allen Two-photon Calcium Imaging dataset: N/A UCL Mice Neuropixel Recording Dataset: 6 female and 4 male
Field-collected samples	The present study utilized publicly available data and did not involve data collection. Detailed settings for each dataset can be found in the respective papers.
Ethics oversight	Experiments for Allen Neuropixel dataset and Two-photon Calcium Imaging dataset were conducted in accordance with protocols approved by the Allen Institute's Institutional Animal Care and Use Committee. Experiments for UCL Mice Neuropixel Recording Dataset were conducted according to the UK Animals Scientific Procedures Act (1986) and under personal and project licenses released by the Home Office following appropriate ethics review.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A

Magnetic resonance imaging

Experimental design

Design type	Simultaneous EEG-fMRI Resting-state Dataset: resting-state Natural Scenes Dataset (NSD): task, event-related design Human Connectome Project (HCP): resting-state
Design specifications	Simultaneous EEG-fMRI Resting-state Dataset: 10 minutes resting-state session Natural Scenes Dataset (NSD): images were presented using a rapid event-related design consisting of 4-s trials Human Connectome Project (HCP): 15 minutes resting-state session
Behavioral performance measures	Natural Scenes Dataset (NSD): behavioral responses were recorded using a button box to indicate whether an image had been seen before.

Acquisition

Imaging type(s)	functional
Field strength	Simultaneous EEG-fMRI Resting-state Dataset: 3T Natural Scenes Dataset (NSD): 7T Human Connectome Project (HCP): 7T
Sequence & imaging parameters	Simultaneous EEG-fMRI Resting-state Dataset: BOLD fMRI data were acquired using an echo planar imaging (EPI) sequence with TR 2100 milliseconds, TE 25 milliseconds, flip angle 90°, slice thickness 4 mm, slices 35, FOV 240 mm, and an in-plane resolution of 3 mm. Natural Scenes Dataset (NSD): Functional data were collected using gradient-echo EPI at 1.8-mm isotropic resolution with whole-brain (including cerebellum) coverage (84 axial slices, slice thickness 1.8mm, slice gap 0mm, field-of-view 216mm (FE) × 216mm (PE), phase encode direction anterior-to-posterior, matrix size 120 × 120, TR = 1,600ms, TE = 22.0 ms, flip angle 62°, echo spacing 0.66ms, bandwidth 1,736 Hz per pixel, partial Fourier 7/8, iPAT 2, multi-band slice acceleration factor 3). Human Connectome Project (HCP): Functional data were collected using gradient-echo EPI at 1.6-mm isotropic resolution (slice thickness 1.6mm, field-of-view 208mm (RO) × 208mm (PE), TR = 1,000ms, TE = 22.2 ms, flip angle 45deg, echo spacing 0.64ms)
Area of acquisition	Whole brain
Diffusion MRI	☐ Used ☒ Not used

Preprocessing

Preprocessing software	Simultaneous EEG-fMRI Resting-state Dataset: preprocessing script from the 1000 Functional Connectomes Project Natural Scenes Dataset (NSD): preprocessing script from NSD Human Connectome Project (HCP): preprocessing script from HCP
Normalization	All individual images were spatially normalized to the Montreal Neurological Institute (MNI) space.
Normalization template	Simultaneous EEG-fMRI Resting-state Dataset: MNI 152 Natural Scenes Dataset (NSD): Native Human Connectome Project (HCP): MNI 152
Noise and artifact removal	Simultaneous EEG-fMRI Resting-state Dataset: Nuisance parameters, including linear and quadratic trends, motion parameters, white matter, and CSF signals, were regressed from the fMRI data Natural Scenes Dataset (NSD): Data underwent initial preprocessing to correct for head motion, EPI distortion, gradient nonlinearities, and alignment across scan sessions. Human Connectome Project (HCP): The resting-state HCP fMRI volumetric data were preprocessed using the minimal preprocessing pipeline (92) and artifacts were further removed with ICA-FIX denoising (93).
Volume censoring	No volume censoring was performed.

Statistical modeling & inference

Model type and settings	decoding model based on deep learning representation similarity analysis
Effect(s) tested	t-test permutation test ANOVA

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☒ Both

Anatomical location(s)

For every fMRI dataset analyzed, we utilized the Schaefer 400 Parcellations (Yeo-17 Network version)(38) to obtain cortical signals from the volumetric fMRI data. This was accomplished by averaging the standardized voxel signals located within each parcel. Similarly, signals from thalamic nuclei/regions were extracted utilizing the Morel Atlas (96), and signals from the brainstem nuclei, part of the ascending arousal network, using the Harvard AAN atlas (97).

Statistic type for inference

(See [Eklund et al. 2016](#))

Unlike conventional fMRI studies that aim to identify significantly activated brain regions, this study used fMRI data to either identify propagating waves based on relative timing of local peaks or to predict semantic information of visual stimulation. Voxel-based or cluster-based spatial inferences discussed in Eklund et. al. 2016 were not used and are also irrelevant in this context. The statistical maps were used mainly to show spatial variation in wave dynamics or decoding accuracy, rather than identify and claim the significance of certain regions

Correction

No analyses involved multiple comparisons.

Models & analysis

n/a	Involvement in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☐	☒ Multivariate modeling or predictive analysis

Multivariate modeling and predictive analysis

To quantify the level of sensory stimulus encoding, we developed a deep learning model to decode semantic information of each image stimulus based on fMRI responses it evoked. The model comprised an fMRI encoder, which extracted latent representation from the fMRI responses, i.e., the fMRI embeddings, and a caption decoder, which translated the fMRI embeddings into descriptive text captions. The fMRI encoder was trained to align the fMRI embeddings with the contrastive language-image pre-training (CLIP) embedding space through contrastive learning.

To assess the sensory information encoded by neuronal populations, we analyzed spiking data obtained during natural scene image stimulation sessions. We defined the neural code as a population response to each displayed image and employed support vector machines (SVMs) based on neural code to decode the actual images was viewed by the mouse. The sensory efficacy is quantified by the decoding accuracy of the SVMs.