nature portfolio

Corresponding author(s): Greta Tuckute (gretatu@mit.edu), Evelina Fedorenko (evelina9@mit.edu)

Last updated by author(s): 2023/10/22

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 (*n*) 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 *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ 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. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ 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 *d*, Pearson's *r*), 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	For fMRI data collection, the materials and scripts to localize the language and multiple demand system are available from the Fedorenko Lab website (https://evlab.mit.edu/funcloc). The software used to collect fMRI data was MATLAB 2014b.
Data analysis	fMRI preprocessing was performed using SPM12 (release 7487) and custom MATLAB scripts, some of which build on the CONN toolbox (CONN EvLab module, release 19b). Event-related fMRI first-level modeling was performed using GLMsingle (Prince et al., 2022) in a Python environment (version 3.8.12). The code is available at the following repository: https://github.com/gretatuckute/GLMsingle Blocked fMRI first-level modeling was performed using SPM12 (release 7487) and custom MATLAB scripts, some of which build on the CONN toolbox (CONN EvLab module, release 19b). For further data analysis, code was written in Python (version 3.8.11), making heavy use of the numpy (Harris et al., 2020; version 1.21.2), scipy (Virtanen et al., 2020; version 1.7.3), scikit-learn (Pedregosa et al., 2011; version 0.24.2), pandas (McKinney et al., 2010; version 1.4.2) and transformers (Wolf et al., 2020; version 4.11.3) libraries. For statistical analyses, linear mixed effects models were implemented using the lmer function from the lme4 package (Bates et al., 2015; version 1.1-31) in R (version 4.2.2). Analysis code is available at the following repository: https://github.com/gretatuckute/drive_suppress_brains

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

Data can be downloaded via the following repository: https://github.com/gretatuckute/drive_suppress_brains

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	Information about participants' sex was collected and reported. Gender information was not collected.
Reporting on race, ethnicity, or other socially relevant groupings	Information about race, ethnicity, or other socially relevant groupings is not reported.
Population characteristics	A total of 14 neurotypical adults (9 female), aged 21 to 31 (mean 25.3; std 3). 12 participants (~86%) were right-handed, as determined by the Edinburgh handedness inventory and self-report 2 (~14%) were left-handed. All participants were native speakers of English.
Recruitment	Participants were recruited from MIT and the surrounding Cambridge/Boston, MA community.
Ethics oversight	The protocol for these studies was approved by MIT's Committee on the Use of Humans as Experimental Subjects (COUHES). All participants gave written informed consent in accordance with the requirements of this protocol.

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

Behavioural & social sciences study design

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

Study description	The study contains fMRI neuroimaging data and behavioral data. Data were quantitative.
Research sample	Participants were from MIT and the surrounding Cambridge/Boston, MA community. Participants were native English speakers.
Sampling strategy	The sample size was based on those used for previous fMRI semantic decoding experiments (Huth et al., 2016; Pereira et al., 2018).
Data collection	Neuroimaging data were collected at the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at MIT. Behavioral data were collected using the Prolific and Amazon Mechanical Turk crowd-sourcing platforms.
Timing	Neuroimaging data were collected between October 2021 and December 2022. Behavioral data were collected between February 2023 and April 2023.
Data exclusions	For neuroimaging data, no participants were excluded from the study based on data quality considerations. For behavioral data, participants were excluded based on the following pre-defined criteria: 1. Native speaker status: Participants were excluded based on their native speaker status self report as well as the Prolific/mTurk language and location filters. 2. Sentence completions: Participants were excluded if their sentence completions were ungrammatical, contained spelling errors (that were not obvious typos) or if the completions were deeply nonsensical. 3. Average response time: Participants were excluded if the average response time per question was less than 3 seconds (i.e., for the survey that contained two questions, the threshold was 6 seconds). 4. Lack of variance in ratings: Participants were excluded if they only used a total of 2 unique rating values (out of 7) for all items in the survey. In addition, for the 2-question "form and meaning" survey, participants were excluded if they always provided the same rating for two questions across all items.

5. Correlation with other participants: Participants were excluded if the average Pearson correlation with the ratings of remaining participants fell below 2 standard deviations below the mean inter-participant correlation. The inter-participant correlated was computed by correlating a vector of responses for a given participant with the vector of responses for each of the remaining participants and taking the average of these pairwise correlation values.

Non-participation

No participants dropped out.

Randomization

Participants were not allocated to experimental groups. There was no randomization procedure for participant selection or enrollment.

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

Magnetic resonance imaging

Experimental design

Design type

Localizer experiments were blocked.
The critical experiment consisted of both an event-related design and a blocked design. n=10 participants took part in the event-related experiment (5 participants completed two sessions each, and 5 participants completed 3 sessions each). n=4 participants took part in the blocked experiment (one session each).

Design specifications

For the language localizer task: Blocked design with a counterbalanced condition order across runs. Each stimulus consisted of 12 words/nonwords. Stimuli were presented in the center of the screen, one word/nonword at a time, at the rate of 450ms per word/nonword. Each stimulus was preceded by a 100ms blank screen and followed by a 400ms screen showing a picture of a finger pressing a button, and a blank screen for another 100ms, for a total trial duration of 6s. Experimental blocks lasted 18s (with 3 trials per block), and fixation blocks lasted 14s. Each run (consisting of 5 fixation blocks and 16 experimental blocks) lasted 358s. Participants completed 2 runs.
For the multiple demand localizer task: MD localizer task: Blocked design with a counterbalanced condition order across runs. The runs consisted of easy and hard arithmetic conditions. The arithmetic task (numbers) were presented in the center of the screen for 1,450ms, followed by the response choices presented for 1,450ms and an inter-stimulus interval of 100ms. Experimental blocks lasted 15 s (with 5 trials per block), and fixation blocks lasted 15s. Each run consisted of 16 experimental blocks—8 blocks per condition—and 5 fixation blocks; a fixation block appeared at the beginning of the run and after each set of four experimental blocks, and lasted 315s. Participants completed 2 runs.
For the critical event-related experiment: Participants passively read each sentence once, in a condition-rich, event-related fMRI design (each sentence is effectively a condition). Sentences were presented (in black font) on a light grey background one at a time for 2s with a 4s inter-stimulus interval (ISI) consisting of a fixation cross. Each run contained 50 unique sentence trials and three 12s fixation blocks (in the beginning, middle (i.e., after 25 sentences) and end of each run). Each run lasted 336s (5:36 minutes). Participants completed either a total of 20 runs (n=5 participants across two sessions) or 30 runs (n=5 participants across three sessions).
For the critical blocked experiment: Sentences were grouped into blocks of 5 sentences from the same condition (baseline, drive, suppress) and were presented on the screen (in black font on a light grey background) one at a time for 2s with a 400ms ISI consisting of a fixation cross (for a total block duration of 12s). Each run consisted of 24 blocks with 8 blocks (40 sentences) per condition. There were five 12s fixation blocks: in the beginning and end of each run, as well as after 6, 12, and 18 blocks. Each run lasted 348s (5:48 minutes). Participants (n=4) completed a total of six runs.

Behavioral performance measures

Button press in the language localizer task. Forced-choice in the multiple demand localizer task. Behavioral performance measures were not analyzed for the localizers.
For the critical experiment, no task was performed during data acquisition. However, participants completed a memory task at the end of each scanning session (outside of the scanner) to incentivize attention throughout the session. Participants were informed ahead of time that they would be asked to perform a memory task after the scanning session. Behavioral responses were analyzed: The average accuracy (sum of correct responses divided by total number of responses was 70.4% (SD across sessions: 11.4%) for the event-related participants (n=24 sessions – responses for

one session were not saved due to an error in the script), and 61.7% (SD across sessions: 10%) for the blocked participants (n=4 sessions).

Acquisition

Imaging type(s): Structural and functional.

Field strength: 3T.

Sequence & imaging parameters: T1-weighted, Magnetization Prepared RApid Gradient Echo (MP-RAGE) structural images were collected in 176 sagittal slices with 1 mm isotropic voxels (TR = 2,530 ms, TE = 3.48 ms, TI = 1100 ms, flip = 8 degrees).
Functional, blood oxygenation level dependent (BOLD), data were acquired using an SMS EPI sequence (with a 90 degree flip angle and using a slice acceleration factor of 2), with the following acquisition parameters: fifty-two 2 mm thick near-axial slices acquired in the interleaved order (with 10% distance factor) 2 mm × 2 mm in-plane resolution, FoV in the phase encoding (A ≫ P) direction 208 mm and matrix size 104 × 104, TR = 2,000 ms and TE = 30 ms, and partial Fourier of 7/8. The first 10 s of each run were excluded to allow for steady state magnetization.

Area of acquisition: Whole brain.

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software: SPM12 and custom MATLAB scripts.

Normalization: All functional scans were coregistered and resampled using B-spline interpolation to the first scan of the first session.

Normalization template: SPM12 default Montreal Neurological Institute (MNI) template.

Noise and artifact removal: For the blocked experiments, data were high-pass filtered at 128s.
For the event-related experiments, the data-driven analysis method GLMdenoise and the statistical technique of ridge regression were used. These methods can account for a variety of sources of noise (e.g., physiological, motion, scanner artifacts, effects of collinearity).

Volume censoring: For the blocked experiments, time points classified as outliers based on the motion data each had a regressor included in the GLM but were not removed (outliers were identified from the resulting subject-motion estimates as well as from BOLD signal indicators using default thresholds in CONN preprocessing pipeline: 5 standard deviations above the mean in global BOLD signal change, or framewise displacement values above 0.9 mm; Nieto-Castanon, 2020).
For the event-related experiment, no volume censoring was performed.

Statistical modeling & inference

Model type and settings: Single-trial BOLD response amplitudes were estimated for individual voxels in individual participants. Voxels were aggregated by averaging across voxels in pre-defined regions of interest (see "Anatomical location(s)" below). Univariate and predictive (encoding model) analyses were performed on these ROI-level responses.

Effect(s) tested: Main effects were estimated using LME models with the following formulae:
BOLD response ~ variable_of_interest + (1 | sentence) + run_within_session + trial_within_run
Where "variable_of_interest" is either the condition (drive, suppress, baseline) or the behavioral sentence property of interest.

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

Anatomical location(s): Functionally-located language regions were the primary brain ROIs. These were defined as follows: Each individual map for the sentences > nonwords contrast from the language localizer was intersected with a set of 10 binary parcels (both hemispheres) derived from a probabilistic activation overlap map for the same contrast in a large set of participants (n=220) using watershed parcellation (an approach developed in Fedorenko et al., 2010).
For supplementary analyses, we further defined ROIs based on the multiple demand localizer, also intersected with a set of binary parcels derived using the approach described in Fedorenko et al., 2010.
Finally, in an attempt to cover a large part of the cortex, we obtained ROIs from the Glasser parcellation (Glasser et al., 2016).

Statistic type for inference (See Eklund et al. 2016): The Eklund paper concerns traditional group-level random effect analyses, in the current paper all the analyses are performed within individuals and then the extracted responses are analyzed with linear mixed effects models and correlation measures.

Correction: Not applicable, as voxel-wise statistical significance inferences are not included in this paper.

Models & analysis

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

Multivariate modeling and predictive analysis

For the encoding model analyses, we mapped from language model representations to brain responses using a ridge regression mapping model (L2 regularized). The regularization parameter was estimated using leave-one-out cross-validation implemented using the scikit-learn Python library function RidgeCV (Pedregosa et al., 2011; version 0.24.2). No dimensionality reduction was performed.