

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	The current study reanalyzed a publicly available dataset (Hippocampus; http://aka.ms/hippocampus) and did not involve new data collection. The Hippocampus dataset was originally collected online using HTML and JavaScript packages, including Bootstrap v4.0.0, d3 v5.16.0, and jQuery v3.2.1.
Data analysis	Data analysis was performed using publicly available Python and R packages. For text similarity analysis, we used a pretrained model (all-mpnet-base-v2) implemented in Sentence Transformers (v2.2.0). For sentiment analysis, we used the VADER sentiment analysis tool implemented in the NLTK library (v3.8.1). For mixed-effects logistic regression, we used lme4 (v1.1-35.5) in R. We used the psych package (v2.5.6) to compute Williams tests, cocor (v1.1.4) to obtain confidence intervals and Cohen's <i>q</i> , and irr (v0.84.1) to estimate Fleiss' <i>κ</i> in R. For other statistical tests, we used Pingouin (v0.5.1), Scipy (v1.12.0), and Scikit-learn (v1.2.2) in Python.

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](#)

The original Hippocampus dataset is publicly available on the Microsoft website (<https://www.microsoft.com/en-us/download/details.aspx?id=105291>). A preprocessed version of the subset used in the current study is publicly available on the Open Science Framework (<https://osf.io/a65bg>).

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	Among the 210 participants included in our study, 118 self-identified as female, 90 as male, 1 as nonbinary, and 1 did not report their gender in the demographic survey. Because we did not have specific hypotheses regarding sex or gender, these variables were not included in our analyses.
Reporting on race, ethnicity, or other socially relevant groupings	The current study did not use any socially constructed or socially relevant categorization variables.
Population characteristics	All participants were based in the United States and could read and write English. The final sample consisted of 210 participants, with the following age distribution: 14 aged 18–24, 42 aged 25–29, 49 aged 30–34, 38 aged 35–39, 15 aged 40–44, 19 aged 45–49, 17 aged 50–54, and 16 aged 55 or older.
Recruitment	Crowdsourced participants were recruited via Amazon Mechanical Turk. Individuals employed by, or enrolled as students at, the University of Washington or Microsoft were not eligible to participate. No additional details were provided in the Hippocampus dataset or in the original publication describing it (Sap et al., 2020, Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics).
Ethics oversight	The Hippocampus dataset was generated in accordance with protocols approved by the Institutional Review Board at Microsoft Research.

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](https://www.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	We conducted quantitative analyses on an existing dataset (Hippocampus; Sap et al., 2020) consisting of written autobiographical memory narratives and associated Likert-scale ratings.
Research sample	The Hippocampus dataset involved Amazon Mechanical Turk workers based in the United States. According to the original publication describing the dataset (Sap et al., 2020), the workers were 47% male, 52% female (<1% non-binary, <1% other), with an average age of 36 years (SD = 10 years), and were predominantly White (73%, with 10% Black, 6% Hispanic, and 5% Asian).
Sampling strategy	Participants for the Hippocampus dataset were randomly sampled. The original publication (Sap et al., 2020) did not describe methods for predetermining the sample size. In the current study, we selected a subset of participants who met our pre-established criteria (see Data exclusions below for details), resulting in 210 participants included in our analyses.
Data collection	The Hippocampus dataset was collected online via Amazon Mechanical Turk. Although the dataset and the original publication (Sap et al., 2020) did not provide specific details, in online studies participants typically use their own electronic devices to access and complete the task at a location of their choosing, without the researcher present, and other individuals may or may not be present during participation.

Timing	The exact start and stop dates of data collection for the Hippocampus dataset were not reported in either the dataset or the original publication describing it (Sap et al., 2020).
Data exclusions	We selected 210 participants from the 1,319 individuals who completed both the first and second recall sessions in the original Hippocampus dataset, based on the following pre-established criteria: (1) Participants were required to report no distraction (i.e., a distraction rating of 1) during both recall sessions. (2) Only participants who described the same specific event occurring within a single day across both recalls were included. Participants who recalled different events across sessions or whose narratives were disorganized and lacked a clear episode were excluded. Selection was performed manually by one of the authors (C.K.), who reviewed the content of both recalls. (3) Participants were required to recall an event that occurred within three months (90 days) prior to the first recall session. This criterion was intended to focus on relatively recent memories, as older, more stabilized memories typically allow less room for transformation between recall sessions.
Non-participation	Because we analyzed a subset of participants from an existing dataset, response rate is not applicable to the current study. The response and drop-out rates for the full Hippocampus dataset were not reported in the original publication describing the dataset (Sap et al., 2020).
Randomization	The selected 210 participants were not allocated into experimental groups in the current study.

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

Plants

Seed stocks	Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.
Novel plant genotypes	Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.
Authentication	Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.