

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
<input type="checkbox"/>	<input checked="" type="checkbox"/> The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement
<input type="checkbox"/>	<input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
<input type="checkbox"/>	<input checked="" type="checkbox"/> 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>
<input type="checkbox"/>	<input checked="" type="checkbox"/> A description of all covariates tested
<input type="checkbox"/>	<input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
<input type="checkbox"/>	<input checked="" type="checkbox"/> 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)
<input type="checkbox"/>	<input checked="" type="checkbox"/> 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>
<input type="checkbox"/>	<input checked="" type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
<input type="checkbox"/>	<input checked="" type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
<input type="checkbox"/>	<input checked="" type="checkbox"/> 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	<p>Custom Javascript code was used to build and display the experiment. This code and all software dependencies and versions can be found here, with dependences in package-lock.json: https://github.com/cbergey/typicality_front</p> <p>A Mongo database and custom Javascript and Python code were used to maintain the back-end of the web experiment. This code and all software dependencies and versions can be found here, with dependences in package-lock.json: https://github.com/cbergey/many_typicality_judgments</p>
Data analysis	<p>Analysis was conducted in R and Python with custom code using the following packages and versions:</p> <p>R version 4.3.1 (2023-06-16), ggpubr_0.6.0, ggh4x_0.3.0, udpipe_0.8.11, lubridate_1.9.4, forcats_1.0.0, stringr_1.5.1, dplyr_1.1.4, purrr_1.0.4, readr_2.1.5, tidyr_1.3.1, tibble_3.2.1, tidyverse_2.0.0, weights_1.0.4, Hmisc_5.2-2, broom.mixed_0.2.9.6, ggthemes_5.1.0, lmerTest_3.1-3, lme4_1.1-36, Matrix_1.6-1.1, papaja_0.1.3, tinylabls_0.2.4, xtable_1.8-4, tidyboot_0.1.1, scales_1.3.0, ggrridges_0.5.6, here_1.0.1, ggplot2_3.5.1, png_0.1-8</p> <p>Python version 3.9.1, torch 1.8.1, numpy 1.20.2, pickle5 0.0.11, pandas 1.2.4, scipy 1.6.3, transformers 4.21.2, seaborn 0.11.2, scikit-learn 0.24.2, matplotlib 3.5.0, nltk 3.6.2, openai 0.27.7, gensim 4.0.1, huggingface-hub 0.9.1, pytorch-pretrained-bert 0.6.2</p>

Analysis code is available at https://github.com/benjaminmorris/purple_carrots.

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 are available at https://github.com/benjaminmorris/purple_carrots.

All data from our rating task of adjective-noun pairs are available. Tokenwise data on usage of each adjective-noun pair in each corpus is available.

The full parent-child corpus is not publicly available due to privacy concerns about densely sampled conversational data that contains identifiable information about children. The extracted adjective-noun pairs as well as code used to extract them from the corpus are publicly available.

The adult corpus is publicly available here <https://ca.talkbank.org/access/CABNC.html>.

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

Data on the sex and gender of participants in our typicality ratings study was not collected. Given that participants are providing norms for the typicality of features of concrete things, participant-level variables are not a part of the research questions under investigation.

Data on the sex and gender of participants in the child corpus (LDP) and adult-adult corpus (CABNC) are available and reported in the citations provided. We do not report them in the manuscript as we do not investigate questions specifically related to sex and gender.

Reporting on race, ethnicity, or other socially relevant groupings

For the child corpus used, self-reported ethnicity and income data were recorded and are available in the original scientific reports on this corpus. We do not re-report them in this manuscript.

For the adult-adult corpus used, self-reported age, social class, and sex data were recorded and are available in the original reports on this corpus. We do not re-report them in this manuscript.

For the child corpus used, we do report and use children's ages. Ages were determined by report of birth dates by the children's parents as collected in the corpus.

Population characteristics

Participants in the typicality rating study were recruited on an online research platform and were required to be over 18 years of age.

Participants in the child corpus range between 14 and 58 months of age and were studied longitudinally such that, excepting attrition, each child has observations at 4 month intervals in that age range.

Recruitment

Participants in the typicality rating study were recruited on the online platform Amazon Mechanical Turk and required to be at least 18 years of age, located in the United States, and have an approval rating on the platform of at least 98%. Online crowdworkers may be more technologically informed than others in the population. We expect these typicality norms to generalize to most English-speaking contexts, though it is possible that the constraint that participants lived in the United States or factors influencing self-selection into crowdworker roles would mean that these results would generalize less well to less technologically informed populations or cultures and dialects more distinct from American and American English.

Families in the child corpus were recruited such that the families spoke only English at home, and to be representative of the racial and socioeconomic range of the Chicagoland area in the United States. This recruitment strategy means that we expect results to generalize well to American populations, particularly urban and suburban populations, though results may generalize less well to rural populations, non-English-speaking populations, and non-American populations.

The adult corpus used, the spoken portion of the British National Corpus, was recruited from across the UK and designed to be representative of the UK population across gender, age, and social class. This strategy means we expect results to generalize well to conversational UK English and similar dialects.

Ethics oversight

The University of Chicago IRB approved and oversaw this study.

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

In this study, we used corpus analysis to extract adjective-noun pairs from parent-child and adult-adult conversational corpora. These extracted pairs are used in the study described above.
We then sought to characterize these usage data by collecting quantitative data on people's beliefs about the typicality of features for these adjective-noun pairs. For example, people would report how typical it is for a cow to be a brown cow on a scale of 1 (never) to 7 (always).
We also evaluate language models' representations of these adjective-noun pairs.

Research sample

For the typicality judgment study, participants were recruited on Amazon Mechanical Turk and required to be 18 years or older and living in the United States. Sex, ethnicity, and other demographic data were not collected. According to Amazon's reporting, women are slightly over-represented and African Americans are slightly under-represented on the Mechanical Turk platform, and people with higher incomes (\$100,000+) are under-represented. Selection into online crowdwork may also select for people who are relatively technologically informed.

Families in the child corpus used were recruited such that the families spoke only English at home, and to be representative of the racial and socioeconomic range of the Chicagoland area in the United States. This recruitment strategy means that we expect results to generalize well to American populations, particularly urban and suburban populations, though results may generalize less well to rural populations, non-English-speaking populations, and non-American populations.

The adult corpus used, the spoken portion of the British National Corpus, was recruited from across the UK and designed to be representative of the UK population across gender, age, and social class. This strategy means we expect results to generalize well to conversational UK English and similar dialects.

Sampling strategy

Participants on Amazon Mechanical Turk are roughly randomly sampled from those within the platform that matched our recruitment requirements, with the constraint that crowdworkers have first-come first-served priority to tasks on the platform. We did not sample crowdworkers with any stratification or constraints other than requiring them to be 18 years or older, living in the United States, and to have a 98% or better prior approval rating on tasks. We aimed to collect five ratings for each adjective-noun pair, and thus recruited 910 participants based on the number of pairs to rate.

Families in the original child corpus study (LDP) were recruited by a mailing to approximately 5000 families in the Chicagoland area and an advertisement in a parenting magazine. Interested families were given a demographic questionnaire by phone and the final 64 families were chosen to be as closely representative of the 2000 Chicago census as possible with respect to income, education, and ethnic identity. Children also had to be reported to have no known physical or cognitive disabilities.

Individuals in the original adult corpus study (BNC) were recruited from across the UK and designed to be representative of the UK population across gender, age, and social class. Researchers sampled across a variety of locations in the UK and used initial interviews to determine demographic characteristics and select participants for demographic balance.

Data collection

Participants in the typicality rating study we collected completed the study online. They used mouse clicks to indicate their responses to typicality judgment questions. No researcher was present.

Families in the original child corpus study were recorded using video cameras with one or more experimenters present. These experimenters were not aware of our later hypotheses and analyses.

Individuals in the original adult corpus study were recorded using tape recorders they carried. No experimenters were present. The researchers involved in this corpus study were not aware of our later hypotheses and analyses.

Timing

The typicality rating task was run between May 1 and May 8 of 2023.

The child corpus study (LDP) was collected between 2000 and 2006.

The adult corpus study (BNC) was collected between 1980 and 1993.

Data exclusions

In the typicality rating study, participants were excluded if they incorrectly answered a simple attention check question that asked them to select a particular number on the rating scale. 16 participants were excluded.

Families in the original child corpus study were not excluded in our study. Note that there exists a separate part of this corpus that includes children with traumatic brain injury; we did not include this part of the corpus.

We excluded any conversations including child participants (828 transcripts) from our sample of the BNC. 140 of the original BNC recordings were missing from the final data and we did not include these in our study. This resulted in a final sample of 1052 transcripts.

Non-participation

In the typicality rating study, 39 participants abandoned the study partway through. We included any typicality judgments that were recorded from their tasks.

Families in the original child corpus study all participated in at least 4 timepoints of recording sessions, out of a possible 12. The mean number of timepoints was 11.3.

The original adult-adult corpus study does not report dropout rates of participants.

Randomization

In the typicality rating study, sets of stimuli were randomly grouped into lists of 35 adjective-noun pairs. Participants were randomly assigned a list of 35 ratings to do in each task, with the constraint that lists were preferentially assigned if they were undersampled (less than 5 ratings) as the experiment progressed. A MongoDB database was used to track which lists had been rated. No experimental manipulations were assigned beyond this balanced, randomized distribution of stimuli.

Participants in the original corpus studies were not assigned to any experimental conditions.

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
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> 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.