

Reporting Summary

Nature Research 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 Research policies, see [Authors & Referees](#) 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

The main data set consists of the 2016-12-01 Wikipedia database dump publicly provided by Wikimedia (<https://dumps.wikimedia.org/backup-index.html>. Accessed 2017-12-15). We used a publicly available database of word embeddings called fastText (<https://fasttext.cc/docs/en/pretrained-vectors.html>. Accessed 2017-07-01). Survey data was collected using Google Forms (Survey 1) and Qualtrics (Survey 2).

Data analysis

Analysis was conducted using Python 3.6 and R3.5. Natural language processing was done primarily using the Python package "NLTK" Version 3.4. Structural equation modeling was performed using R package "lavaan" 0.6-3. Article quality was assessed with Python package wikiclass.

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [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 list of figures that have associated raw data
- A description of any restrictions on data availability

Data is publicly available at Wikipedia's database dumps (<https://dumps.wikimedia.org/>).

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	The study consists primarily of a quantitative analysis of observational, publicly available data from Wikipedia. A small survey was conducted with Wikipedia users to validate quantitative measures.
Research sample	The editing activity on Wikipedia is recorded using users' usernames or, for unregistered users, their IP addresses.
Sampling strategy	The bulk of the analysis used observational data. For the survey of editors, we created random samples of n=500 from the set of editors for whom we had calculated ideological alignment scores (those editing at least one page in our Politics corpus). For the first survey, a survey link was posted to the public user page for each of the 500 individuals; for the second survey, only n=327 such links were posted, due to the labor-intensity of the posting procedure.
Data collection	<i>Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.</i>
Timing	The observational data came from the 2016-12-01 Wikipedia database dump. The surveys were conducted in November 2017 and April 2018.
Data exclusions	Unregistered users (whose usernames are IP addresses) were excluded from the observational dataset, because IP addresses are not reliable as identity. Survey responses from respondents who reported living outside of the USA were dropped, as the political culture we focus on is USA-specific.
Non-participation	n/a
Randomization	n/a

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
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

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

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Demographic characteristics of Wikipedia users are not available in the observational data. Regarding the surveys, we did not collect demographic characteristics beyond location (see above)
Recruitment	Observational data included the entirety of Wikipedia editors. Regarding surveys, see above.
Ethics oversight	University of Chicago's Institutional Review Board (IRB17-0679)

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