******

**Editorial policy and formatting checklist for Articles, Resources and Registered Reports**

|  |  |  |  |
| --- | --- | --- | --- |
| General | | | |
|  | **Yes** | **No** | **N/A** |
| Our manuscript is provided in .docx or .tex format |  |  |  |
| The manuscript file is organized in the following order:   1. Title page 2. Abstract 3. Introduction (no subheadings allowed) 4. Results (subheadings permitted) 5. Discussion (no subheadings allowed) 6. Methods (note that they come after the results and discussion! Subheadings permitted) 7. Protocol Registration (for Registered Reports only) 8. Data availability statement 9. Code availability statement (only if you have used custom code) 10. References 11. Acknowledgements 12. Author contributions 13. Competing interests 14. Figure legends (no Figures!) 15. Tables |  |  |  |
| All figures have been removed from the manuscript file and uploaded individually, one figure per file |  |  |  |
| If the manuscript has been accepted in principle, supporting information is provided in the following three categories:   1. Extended data 2. Supplementary information 3. Source data   Note: See end of this document for more details on supplementary material. |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Title page | | | |
|  | **Yes** | **No** | **N/A** |
| The title of our manuscript is no longer than 100 characters (including spaces) and does not contain any punctuation (except commas in lists)  *NOTE: The “title:subtitle” format is not permitted.* | x |  |  |
| The title does NOT make a novelty/priority claim (except in cases of genetic discovery) and does NOT claim causality when the data are correlational/cross-sectional. |  |  |  |
| The title does NOT interpret absence of evidence as evidence of absence (in the case of null results). |  |  |  |
| Author names are provided in full (with initials for middle names only) |  |  |  |
| Affiliations include department/unit, institution/organization, city, and country information |  |  |  |
| An asterisk is used to indicate the corresponding author(s) |  |  |  |
| An email address is provided for each corresponding author |  |  |  |
| All corresponding authors have an ORCID, which they have linked to their account on the *Nature Human Behaviour* manuscript tracking system  *NOTE: Please follow the steps below to link your account on our manuscript tracking system (MTS) with your ORCID. Per legal policy, this must be done in the manuscript system by the owner of the ORCID. If you don’t have an ORCID yet, you will be able to create one in minutes.*  *1. From the home page of the MTS* [*https://mts-nathumbehav.nature.com/cgi-bin/main.plex*](https://mts-nathumbehav.nature.com/cgi-bin/main.plex) *click on ‘Modify our Springer Nature account’ under ‘General tasks’.*  *2. In the ‘Personal profile’ tab, click on ‘ORCID Create/link an Open Researcher Contributor ID(ORCID)’. This will re-direct you to the ORCID website.*  *3a. If you already have an ORCID account, enter your ORCID email and password and click on ‘Authorize’ to link your ORCID with your account on the MTS.*  *3b. If you don’t yet have an ORCID, you can easily create one by providing the required information and then click on ‘Authorize’. This will link your newly created ORCID with your account on the MTS.* |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Abstract | | | |
|  | **Yes** | **No** | **N/A** |
| The abstract of our manuscript does not exceed 150 words |  |  |  |
| The abstract does not contain any references |  |  |  |
| It starts with a sentence that introduces the general topic and its significance for a broad audience. It then describes the specific question(s) our research addresses, what we did and what we found. It ends with a statement that encapsulates the significance of our findings for a broad audience. |  |  |  |
| The abstract does NOT contain any priority/novelty claims (except in the case of genetic discovery). Statements such as "This is the first demonstration..."; "We propose a novel model..." have been removed or reworded. |  |  |  |
| The abstract does NOT use causal language if the evidence is correlational/cross-sectional. |  |  |  |
| The abstract does NOT interpret absence of evidence as evidence of absence (in the case of null results). |  |  |  |
| The abstract does not contain qualitative characterizations of our work.  *NOTE:* Please avoid self-laudatory language and qualitative characterizations of your work. Your results will be more credible if readers are allowed to form their own opinion of the value of your work. |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Main Text: General | | | |
|  | **Yes** | **No** | **N/A** |
| Our manuscript does NOT contain any priority/novelty claims (except in the case of genetic discovery). |  |  |  |
| Our manuscript does NOT use causal language if the evidence is correlational/cross-sectional. |  |  |  |
| Our manuscript does NOT recycle text from our previous publications without proper acknowledgment and citation of the original work.  *NOTE:* text recycling from the authors’ own work is a form of plagiarism and must be avoided (see our policy here: <https://www.nature.com/authors/policies/plagiarism.html>).  When reusing text verbatim, you must clearly indicate this in your manuscript and identify the original source. If the portion of reused text exceeds 80 words, you must seek permission from the publisher of the original work to reproduce the text. |  |  |  |
| The main text (including results and discussion) does not contain qualitative characterizations of our work.  *NOTE:* Please avoid self-laudatory language and qualitative characterizations of your work. Your results will be more credible if readers are allowed to form their own opinion of the value of your work. |  |  |  |
| We have not reframed or changed any of our research questions/hypotheses/predictions (for confirmatory research).  *NOTE:* We invariably overrule reviewer requests to reframe or recast hypotheses or predictions (for confirmatory research). |  |  |  |
| The main text (which includes the introduction, results and discussion) does not exceed 5,000 words.  *NOTE:* *There is no limit on the Methods section, although ideally it should not exceed 3,000 words.* |  |  |  |
| References appear as superscript Arabic numerals, in order of mention. If a reference is cited more than once, the same number is used throughout the text and the reference receives a single entry in the reference list. | x |  |  |
| Our manuscript does NOT include any footnotes/endnotes. (They have been incorporated in the text or otherwise eliminated.) |  |  |  |
| Our manuscript does NOT use boldface for any reason except titles/section headings. |  |  |  |
| Our manuscript does NOT use *italics* for any other function, except for Latin terms (e.g., *a priori*) or gene symbols/functionally defined locus symbols. |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Main Text: Results & Discussion | | | |
|  | **Yes** | **No** | **N/A** |
| Frequentist inferential statistics are reported in full wherever they occur (main text, Figure captions, Tables) as follows:  statistic(degrees of freedom) = value, p = value, effect size statistic = value, % Confidence Intervals = values |  |  |  |
| P values reported are exact. We do not report ‘ns’ – we give the exact value.  *NOTE: The smallest P value that should be reported is P <0.001, except in studies of genetic associations.* |  |  |  |
| All statements or interpretations of the results are supported by appropriate, fully reported statistics. This extends to comparisons of relationships between variables – for example, interpretations pertaining to difference of differences are also supported by appropriate statistics1.  1Gelman, A. & Stern, H. [The difference between “significant” and “not significant” is not itself statistically significant](http://www.stat.columbia.edu/~gelman/research/published/signif4.pdf). Am. Stat. 60, 328–331 (2006). |  |  |  |
| We do not discuss marginally significant results as theoretically informative for the hypotheses tested. |  |  |  |
| Null effects obtained through NHST are appropriately reported and are not interpreted at all, except if appropriate statistical tests have been employed that allow for meaningful interpretation of the null results.  *NOTE: When reporting null results, you cannot say ‘there was no difference…’ or ‘there was no effect…’; instead, you must say ‘there was no statistically significant difference…’ or ‘there was no statistically significant effect…’. Unless, you have used appropriate statistics (e.g., equivalence tests, Bayes factors, etc), you cannot say anything further about null results and you must refrain from any interpretation. For example, it is inappropriate to say ‘There was no statistically significant difference between x and y (p=.342), which suggests that…’ – a null result in NHST does not suggest anything, except failure to reject the null hypothesis.* |  |  |  |
| If the main results presented in the manuscript are null, we have carried out appropriate statistical tests (e.g., equivalence tests, Bayes Factors, etc) in order to be able to interpret the null effect.  *NOTE: In order to be able to interpret a non-significant result, you must use appropriate statistics (e.g., equivalence tests, Bayes factors). Even then, there is no statistical test that can demonstrate absence of an effect and you cannot say, ‘We found no difference…’ or ‘There is no difference…’. Your reporting of the results must be in line with the statistical method you used to quantify support for the null hypothesis. For example, “We found that the data supported the null model over the alternative model with a Bayes Factor of X” (when using Bayes Factors) or “We statistically rejected effects as large or larger than those we consider meaningful” (when using equivalence tests).* |  |  |  |
| If the main results presented in the manuscript are null, we also provide statistical evidence that the study is sufficiently powered to detect the smallest theoretically or pragmatically meaningful effect. |  |  |  |
| We indicate whether the data met the assumptions of the statistical tests used, including whether normality and equal variances were formally tested. If not, we show data distribution (individual data points) and include the following statement: "Data distribution was assumed to be normal but this was not formally tested.” |  |  |  |
| We specify whether tests were one- or two-tailed, and justify the use of one-tailed tests. |  |  |  |
| If our manuscript reports Bayesian analyses, we have:   * Specified our priors and how they were selected; * Described the statistical model and techniques used in the analyses; * Summarized the posterior distribution with a measure of central tendency and a credibility interval; * Assessed the sensitivity of the analyses to different priors. |  |  |  |
| Our discussion includes a transparent discussion of limitations. |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Methods | | | |
| |  |  |  |  | | --- | --- | --- | --- | |  | No | Yes | N/A | | **Yes** | **No** | **N/A** |
| If our manuscript reports the results of research with human participants, the Methods section starts with a statement confirming that our research complies with all relevant ethical regulations; naming the board and institution that approved the study protocol; and confirming that informed consent was obtained from all human participants. Information on participant compensation is also included. |  |  |  |
| If our manuscript reports the results of research with non-human animals, the Methods section starts with a statement confirming that our research complies with all relevant ethical regulations; naming the board and institution that approved the study protocol; and confirming that the ARRIVE guidelines were used to report the research. |  |  |  |
| If our manuscript reports the results of a clinical trial, the Methods section includes the trial registration number from ClinicalTrials.gov or an equivalent agency. |  |  |  |
| If our manuscript reports the results of a Phase 2 or 3 randomized controlled trial, we are also attaching the CONSORT checklist with our submission. |  |  |  |
| The Methods include a statement indicating how the sample sizes were chosen.  *NOTE: If a power analysis was used, provide the details of this analysis. If you did not use a power analysis, the following is sufficient: "No statistical methods were used to pre-determine sample sizes but our sample sizes are [similar to/larger than] those reported in previous publications (ref x,y,z)."* |  |  |  |
| The Methods include a statement on randomization. We indicate whether the data collection was randomized or appropriately blocked, how subjects/samples were assigned to the various experimental groups and whether there was any randomization in the organization of the experimental conditions or stimulus presentations. |  |  |  |
| The Methods include a statement indicating whether blinding was used.  *NOTE: If there was no blinding, this must be clearly stated in the manuscript, as follows: "Data collection and analysis were not performed blind to the conditions of the experiments.”* |  |  |  |
| We disclose whether any subjects or data points were excluded from the analyses for any reasons and note the rationale for the exclusions. |  |  |  |
| If any of our studies was pre-registered, we provide a link to the pre-registration in the Methods section and state the date of pre-registration. We disclose all deviations from the pre-registered protocol and explain the rationale for deviation (e.g., flaw, suboptimality, or reviewer/editorial request). In cases of deviation in our analysis plan, the originally planned analyses are reported in Supplementary Information. |  |  |  |
| The sex, number and age of participants is indicated. |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Protocol registration statement (Stage 2 Registered Reports only) | | | |
|  | **Yes** | **No** | **N/A** |
| We have provided a protocol registration statement for our Stage 1 in-principle accepted protocol, under the heading ‘Protocol registration’.  *NOTE: This statement must take the following form:*  The Stage 1 protocol for this Registered Report was accepted in principle on [DATE]. The protocol, as accepted by the journal, can be found at [URL]. |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Data availability statement | | | |
|  | **Yes** | **No** | **N/A** |
| We have provided a data availability statement as a separate section, under the heading ‘Data availability’.  *NOTE: This section should inform readers about the availability of the data used to support the conclusions of your study. This information includes accession codes to public repositories (data banks for protein, DNA or RNA sequences, microarray, proteomics data etc…), references to source data published alongside the paper, unique identifiers such as URLs to data repository entries, or data set DOIs, and any other statement about data availability. We strongly encourage authors to publicly share their data unless there are ethical or legal restrictions that prevent you from sharing. For Registered Reports, public sharing of data is mandatory. For all other manuscripts, at a minimum, you should include the following statement: “The data that support the findings of this study are available from the corresponding author upon request”, mentioning any restrictions on availability. If DOIs are provided, we also strongly encourage including these in the Reference list (authors, title, publisher (repository name), identifier, year). For more guidance on how to write this section please see:* [*http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf*](http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf) |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Code availability statement | | | |
|  | **Yes** | **No** | **N/A** |
| Right after the Data availability statement and before the References, we have provided a code availability statement as a separate section, under the heading ‘Code Availability’.  *NOTE: You only need to include a separate “Code availability” statement on the availability of your analysis routines/code if you developed any custom code. We strongly encourage you to provide this custom code as a separate SI file titled “Supplementary Software”, and include a call-out to it in your Methods section. If possible, please also provide a link (e.g. GitHub) to a live version of your code so that readers can track its updates. At a minimum, you should include the following statement: “Custom code that supports the findings of this study is available from the corresponding author upon request.”* |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| References | | | |
|  | **Yes** | **No** | **N/A** |
| Only one publication is given for each number. |  |  |  |
| Only articles that have been published or accepted by a named publication or recognized preprint server are in the numbered list.  *NOTE: Published conference abstracts, numbered patents and research datasets that have been assigned a digital object identifier may be included in the reference list.* |  |  |  |
| Unpublished manuscripts, submitted manuscripts, and manuscripts under review/revision are not included in the numbered list.  *NOTE: If a manuscript is under consideration, not yet submitted or not yet accepted, it should be mentioned in the main text only in parentheses, as follows: (Up to five author names, et al., unpublished manuscript). No major claims can rely solely on unpublished manuscripts or data.* |  |  |  |
| References are provided in the following format:  Journal article:  Rosenzweig, C. et al. Attributing physical and biological impacts to anthropogenic climate change. Nature ****453,**** 353–357 (2008).  Book:  Jones, R. A. L. Soft Machines: Materials and Life (Oxford Univ. Press, 2004).  Research dataset:  Hao, Z., AghaKouchak, A., Nakhjiri, N. & Farahmand, A. Global Integrated Drought Monitoring and Prediction System (GIDMaPS) data sets. figshare <http://dx.doi.org/10.6084/m9.figshare.853801> (2014).  Preprint:  VanderWeele, T. J., Mathur, M. B. & Chen, Y. Outcome-wide longitudinal designs for causal inference: a new template for empirical studies. Preprint at arXiv <http://arxiv.org/abs/1810.10164> (2019). |  |  |  |
| All references are complete.  Please ensure that ref 33 is complete | x |  |  |
| (For TeX/LaTeX) The references are included within the manuscript itself.  *NOTE: If you wish to use BibTeX, please copy the reference list from the .bbl file, paste it into the main manuscript .tex file, and delete the associated \bibliography and \bibliographystyle commands. Before submission, please ensure that the complete .tex file compiles successfully on your own system with no errors or warnings.* |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Acknowledgements | | | |
|  | **Yes** | **No** | **N/A** |
| We have indicated all sources of funding that supported our work.  *NOTE: If no specific funding supported the work, include the following statement: “The authors received no specific funding for this work.”* |  |  |  |
| We have indicated what role the funder(s) had in the conceptualization, design, data collection, analysis, decision to publish, or preparation of the manuscript.  *NOTE: If any of this information could be perceived as a competing interest, ensure that it is also included in your competing interests statement. If the funder(s) had no role, please include the following statement: “The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.”* | x |  |  |
| The Acknowledgements do not include thanks to anonymous referees or editors, or effusive comments. |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Author contributions | | | |
|  | **Yes** | **No** | **N/A** |
| We have included an ‘Author contributions’ statement of individual author contributions to the paper (such as experimental work, project planning, data analysis, etc.)  *NOTE: The statement should be short, and refer to authors by their initials. For details please see the Authorship section of our joint Editorial policies at* [*http://www.nature.com/authors/editorial\_policies/authorship.html*](http://www.nature.com/authors/editorial_policies/authorship.html) |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Competing interests | | | |
|  | **Yes** | **No** | **N/A** |
| We have declared all financial or non-financial competing interests.  *NOTE: We ask authors to declare both financial and non-financial competing interests. See* [*https://www.nature.com/authors/policies/competing.html*](https://www.nature.com/authors/policies/competing.html) *for more details. If you have no financial or non-financial competing interests, please use the following standard statement: “The authors declare no competing interests.”* |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Figures & Tables | | | |
|  | **Yes** | **No** | **N/A** |
| We followed the guidelines detailed in the following document to prepare our figures: <https://www.nature.com/documents/NRJs-guide-to-preparing-final-artwork.pdf> |  |  |  |
| Our manuscript does not contain more than 8 display items (i.e., combination of Figures and Tables). |  |  |  |
| All figures have been removed from the main manuscript file and uploaded as individual files, one figure per file.  *NOTE: Figure captions should remain in the main manuscript file.* |  |  |  |
| Data are presented in a format that shows data distribution (dot-plots or box-and-whisker plots), with all box-plot elements (e.g. center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers) defined. If bar graphs are used, the corresponding dot plots are overlaid. |  |  |  |
| Graph axes start at zero and have not be altered in scale to exaggerate effects.  *NOTE: A ‘discontinuity’ can be used for the y axis if absolutely necessary due to sizing constraints, but the break must be visually evident and should not impinge on any data points.* |  |  |  |
| All relevant figures have scale bars (rather than numerical descriptions of magnification). |  |  |  |
| All relevant figures have defined error bars. |  |  |  |
| All images comply with Nature Research image integrity policy (<https://www.nature.com/authors/policies/image.html>)  *NOTE: Unprocessed data must be provided upon request. Please double-check figure assembly to ensure that all panels are accurate (e.g. all labels are correct, no inadvertent duplications have occurred during preparation, etc.).* |  |  |  |
| For publication of identifiable images of research participants, we confirm that consent to publish was obtained and is noted in the Methods. |  |  |  |
| All text figures and tables, including supplementary figures and tables, are cited in the text in numerical order. |  |  |  |
| Figure legends are concise and no longer than 350 words. They begin with a brief title and then describe what is presented in the figure and detail all relevant statistical information (redundantly with the main text). The number of participants/observations in each figure (panel) is specified. |  |  |  |
| For Tables that report the results of statistical models, full statistical details are reported (including exact p values, effect sizes, and confidence intervals).  *NOTE: Do not use asterisks to denote different levels of statistical significance.* |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| If your manuscript has been accepted in principle,  you must also provide the following items | | | |
|  | **Yes** | **No** | **N/A** |
| Provide an updated version of the Reporting Summary with your final files. This reporting summary will be published along with your paper.  *NOTE: The Reporting Summary can be found here:* [*https://www.nature.com/documents/nr-reporting-summary.pdf*](https://www.nature.com/documents/nr-reporting-summary.pdf)*. Due to its advanced features, the file cannot be viewed in a browser – please download, save, and view in Adobe Reader or other suitable software.* | x |  |  |
| Please sign and upload with your final submission the [Licence to Publish form](http://www.nature.com/documents/snl-ltp.docx)  Or, if the corresponding author is either a Crown government employee (including Great Britain and Northern Ireland, Canada and Australia), or a US Government employee, please sign and return the [Licence to Publish form for Crown government employees](http://www.nature.com/documents/snl-ltp-crown.docx), or a [Licence to Publish form for US government employees](http://www.nature.com/documents/snl-ltp-govus.docx).  For more information on our licence policy, please consult <http://npg.nature.com/authors>.  *NOTE: The licence to publish form must be hand-signed.* |  |  |  |
| Should your manuscript contain any items (figures, tables, images, videos or text boxes) that are the same as (or are adaptations of) items that have previously been published elsewhere and/or are owned by a third party, please note that it is your responsibility to obtain the right to use such items and to give proper attribution to the copyright holder. This includes pictures taken by professional photographers and images downloaded from the internet. If you do not hold the copyright for any such item (in whole or part) that is included in your paper, please complete and return this [Third Party Rights Table](http://www.nature.com/documents/thirdpartyrights-origres.doc), and attach any grant of rights that you have collected. |  |  | x |
| A brief (maximum 250 characters, including spaces) summary of the main findings of the paper to be used on our website and in our E-Alerts. The summary should be written in the third person in language suitable for a broad audience. The summary may be edited by the editors prior to publication. Please provide this summary in your cover letter. | x |  |  |
| Twitter handles for authors (if desired).  *NOTE: In an effort to disseminate your work to the broadest possible audience, we suggest that you provide your twitter handle and the twitter handles of any co-authors in your cover letter, so that we can be sure to tag you when we tweet your paper from the journal’s account.* | x |  |  |
| Candidates for the cover image (300 dpi resolution at 8.5 by 11 inches), if desired. |  |  | x |
| A completed version of this checklist (including the Supporting Information inventory below). |  |  |  |

**Inventory of Supporting Information**

Our manuscripts include three types of supporting information:

1. Extended Data
2. Supplementary Information
3. Source Data

EXTENDED DATA: There is a limit of 8 display items in the main text (combination of Figures and Tables). However, we also allow up to 10 Extended Data figures. Extended Data are an integral part of the paper and only data that directly contribute to the main message should be presented. These figures will be integrated into the full-text HTML version of your paper and will be appended to the online PDF (for a published example, please see <https://www.nature.com/articles/s41591-019-0508-1>). There is a limit of 10 Extended Data figures, and each must be referred to in the main text. Each Extended Data figure should be of the same quality as the main figures, and should be supplied at a size that will allow both the figure and legend to be presented on a single A4 page. Each figure should be submitted as an individual .jpg, .tif or .eps file with a maximum size of 10 MB each, using the ‘Figure File’ article type in our Manuscript Tracking system. All Extended Data figure legends must be provided in the Inventory of Supporting Information provided below, not in the figure files themselves.

SUPPLEMENTARY INFORMATION: Supplementary Information is material that is essential background to the study but which it is not practical to include in the main text or extended data (for example, video files, large data sets and calculations). Each item must be referred to in the main manuscript and detailed in the attached Inventory of Supplementary Information. Tables containing large data sets should be in Excel format, with the table number and title included within the body of the table. All textual information and any additional Supplementary Figures (which should be presented with the legends directly below each figure) should be provided as a single, combined PDF. Please note that Supplementary Information is not copyedited and we cannot replace any Supplementary Information after the paper has been formally accepted unless there has been a critical scientific error.

All Extended Data Figures must be referred to in your manuscript and cited as *Extended Data Figure 1, Extended Data Figure 2*, etc. Additional Supplementary Figures and other items are not required to be called out in your manuscript text, but should be numerically numbered, starting at one, as Supplementary Figure 1, etc., not SI1.

SOURCE DATA: We encourage you to provide Source Data for your Figures (including Extended Data Figures) whenever possible. Statistical Source Data for any relevant Figures should be provided as individual Excel files (one for each figure) with the linked figure noted within the Excel file. For imaging Source Data, we encourage deposition in a relevant repository.

|  |
| --- |
| **Instructions:**  Please complete each of the Inventory Tables below to outline your Extended Data and Supplementary Information items.  There are four sections; *1. Extended Data, 2A. Supplementary Information: Flat Files, 2B. Supplementary Information: Additional Files,* and *3. Source Data.* Each section includes specific instructions. Please complete these tables as fully as possible.  Please note that titles and descriptive captions will only be lightly edited, so please ensure that you are satisfied with these prior to submission.  If you have any questions about any of the information contained in this inventory, please contact the Editorial Assistant: humanbehaviour@nature.com |

1. **Extended Data**

**Complete the Inventory below for all Extended Data figures.**

* Keep Figure Titles to one sentence only
* File names should include the Figure Number. i.e.: *Smith\_ED Fig1.jpg*
* Please be sure to include the file extension in the Filename. Note that Extended Data files must be submitted as .jpg, .tif or .eps files *only*
* All Extended Data figure legends must be provided in the Inventory below and should not exceed 300 words each *(if possible)*
* Please include Extended Data *ONLY* in this table

|  |  |  |  |
| --- | --- | --- | --- |
| **Figure #** | **Figure title**  One sentence only | **Filename**  This should be the name the file is saved as when it is uploaded to our system. Please include the file extension. i.e.: *Smith\_ED Fig1.jpg* | **Figure Legend**  If you are citing a reference for the first time in these legends, please include all new references in the Online Methods References section, and carry on the numbering from the main References section of the paper. |
| **Extended Data Fig. 1** | Sample size and demographic characteristics of Project Implicit data | ED1.jpg | a, Number of participants by country after exclusions (note that US participants are excluded from the visualization because of the large sample size; N = 545,673). Our final sample included 657,335 participants from 39 countries (see Supplementary Information for exclusion criteria). b, Gender distribution of participants by country after exclusions. Across countries, there tended to be more female participants relative to male participants (M = 0.64 proportion females; SD = 0.06). c, Age distribution of participants by country after exclusions. Ranges correspond to 95% CIs. Red points show median age by country. |
| **Extended Data Fig. 2** | Geographic distribution of IAT scores | ED2.jpg | a, Residualized implicit career-gender association (IAT score) shown by country. IAT scores are residualized for participant age, gender, and task order. Larger values (blue) indicate a larger bias to associate men with the concept of career and women with the concept of family. Countries in white correspond to countries for which there was insufficient data to estimate the country-level career-gender association. Inset shows IAT scores for European countries only. Note that while Hindi is identified as the most frequently spoken language in India, India is highly multilingual and so Hindi embeddings may be a poor representation of the linguistic statistics for speakers in India as a group. b, Distribution of raw implicit career-gender association (IAT D-score) across countries. All countries in our sample showed a tendency to associate men with career and women with family. |
| **Extended Data Fig. 3** | Models predicting IAT effect size at the participant level | ED3.jpg | a, Median country age predicts IAT effect size over and above participant age at the participant level: Countries with older populations tend to have individuals with stronger career-gender associations, even after controlling for participant age. The table presents an additive mixed-effect regression predicting IAT D-score at the participant level with participant age and median country age, controlling for participant sex and trial order. The model includes by-country random intercepts. b, The relationship between median country age and IAT effect size holds, even after controlling for the percentage women in STEM. The table presents an additive mixed effect model predicting IAT D-score at the participant level with participant age, median country age and percentage women in STEM in country, controlling for participant sex and trial order. The model includes by-country random intercepts. |
| **Extended Data Fig. 4** | Replication of Caliskan et al. (2017) with our corpora | ED4.jpg | We replicate the original set of Caliskan, Bryson, and Narayanan (2017; CBN) findings using the English-trained versions of the models used in our main analyses (models trained on the Wikipedia and Subtitles corpora). For each model, we calculate an effect size for each of the 10 IAT types reported in CBN: flowers/insects–pleasant/unpleasant, instruments/weapons–pleasant/unpleasant, European-American/Afro-American–pleasant/unpleasant, males/females–career/family, math/arts–male/female, science/arts–male/female, mental-disease/physical-disease–permanent/temporary, and young/old–pleasant/unpleasant (labelled as Word-Embedding Association Test (WEAT) 1-10 in CBN). We calculate the bias using the same effect size metric described in CBN, a standardized difference score of the relative similarity of the target words to the target attributes (i.e. relative similarity of male to career vs. relative similarity of female to career). This measure is analogous to the behavioural effect size measure where larger values indicate larger bias. The figure shows the effect size measure derived from the English Wikipedia corpus (a) and the English Subtitle corpus (b) plotted against effect size estimates reported by CBN from two different models (trained on Common Crawl and Google News corpora). Point color corresponds to bias type, and point shape corresponds to the two CBN models. With the exception of biases related to race and age, effect sizes from our corpora are comparable to those reported by CBN. In particular, for the gender-career IAT—the bias relevant to our current purposes—we estimate the effect size to be 1.78 (Wikipedia)/1.65 (Subtitle), while CBN estimates it to be 1.81 (Common Crawl)/1.89 (Google News). |
| **Extended Data Fig. 5** | Pairwise Correlations partialing out the effect of median country age | ED5.jpg | Partial correlations (Pearson’s *r*) for all measures in Study 1b and 2 using language as the unit of analysis, controlling for median country age. 95% CIs are given in brackets followed by the corresponding *p*-value. Implicit and explicit male-career association measures are residualized for participant age, gender, and task order. "Assoc." = association; "Lang."= language; "Subt."/ "Wiki." = Subtitle/Wikipedia corpora; "Prop. Gendered Occup. Terms." = proportion of occupation terms that are gendered. "Occup. Genderness" = degree to which occupation terms in a language tend to be associated with a particular gender in the language statistics. |
| **Extended Data Fig. 6** | Replication of Study 1b on Wikipedia corpus excluding translations | ED6.jpg | Both the Subtitle and Wikipedia corpora likely contain some documents that are translated from other languages (e.g., the Wikipedia article on “Paris” is written in French and then translated into English). The parallel content across languages allows us to estimate the gender bias in language statistics, while holding content constant across languages. Nevertheless, content may itself be a driver of gender bias (e.g. one language may have more articles about male politicians relative to another). To understand the contribution of language-specific content on gender bias, we constructed a corpus of Wikipedia articles in each language that were originally written in the target language (i.e., untranslated). We identified untranslated articles by examining the interlanguage links on a Wikipedia article page. These links are pointers to the same article content in other languages (e.g. the “Paris” article in French contains a link to the “Paris” article in English). Since the original source language of an article could not be inferred, we excluded all articles that contained one or more interlanguage links. This ensured that all remaining articles contained only text originally written in the target language. We constructed a corpus for each language containing all untranslated articles. There were a median of 168,326 articles per language (range: 10,307-14,676,484). We trained fastText42 on each language corpus with default parameters and a dimension size of 200. We then used these models trained on native text to calculate by-language IAT bias scores and by-language occupation bias scores, using the same procedure as with the models described in the Main Text (Studies 1b and 2). One language was excluded following the same exclusion criteria as in the main analyses (>= 20% missing words in model; Mandarin), but the results remain the same when this language is included. Using models trained on the untranslated corpora, we replicate the key finding from Study 1b showing a positive correlation between the bias measured behaviorally with the IAT and measured in language (r = .60; p = .002). Notably, the effect size is somewhat larger relative to the other two corpora types, presumably because additional bias is introduced by allowing the corpus content to vary across languages. |
| **Extended Data Fig. 7** | Models examining UK-US bias difference in AIID dataset (Study 1c) | ED7.jpg, ED7a.jpg | a, The exact pre-registered analysis of Study 1c is presented. Pairwise correlations between all variables (language bias, behavioral bias, and UK-US difference measures) are shown, averaging across estimates of language bias from the 5 model runs. Error bars are 95% CIs. As stated in the pre-registration, the key test of our hypothesis is that the correlation between the UK - US linguistic difference (``Language Bias Difference'') and the UK - US behavioral difference (``Behavioral Bias Difference'') is greater than 0 (shown in red). That data are consistent with this prediction. The confirmatory dataset is shown on the right, along with the smaller exploratory dataset on the left for reference. b, The full results to the mixed-effect model described in the paper are presented. |
| **Extended Data Fig. 8** | Models predicting implicit male-career association with proportion gender distinct labels and language career-gender association (Study 2) | ED8.jpg | We predict the magnitude of implicit male-career association by language with an additive linear model. Predictors are proportion of occupation terms that are gendered ("Prop. Gendered Occup. Terms") and language male-career association as measured by word embeddings of the IAT words ("Male-Career Assoc."). Model coefficients are shown for two models using estimates of language career-gender association from embedding models trained on Subtitle (a) and Wikipedia (b) corpora. The linear models account for 40.63% (Subtitle) and 45.32% (Wikipedia) of the variance in implicit male-career association. "Subt."/ "Wiki." = Subtitle/Wikipedia corpora. |
| **Extended Data Fig. 9** | Gender associations in language and other psychological measures | ED9.jpg | Several recent studies6, 35 have presented novel theories to account for cases of structural inequality related to gender. Both of these studies argue that psychological differences play a causal role in the emergence of structural inequality. Here, we show that degree of gender bias in language is correlated with these psychological differences at the country level, consistent with the idea that language experience could be playing a causal role in the emergence of psychological differences. a, Gender differences in preferences35 (composite score of “six fundamental preferences with regard to social and nonsocial domains: willingness to take risks; patience, which captures preferences over the intertemporal timing of rewards; altruism; trust; and positive and negative reciprocity, which capture the costly willingness to reward kind actions or to punish unkind actions, respectively.”) as a function of language male-career association measured in the Subtitle corpus. These two measures are correlated (r(25) = 0.48 [0.12, 0.73],p= 0.01): Countries with greater differences in gender preferences also have greater gender bias present in their languages. We also find that per capita GDP49 is correlated with language gender male-career association measured in both corpora (Wikipedia: r(35) = 0.64 [0.4, 0.8],p< .0001; Subtitle: r(31) = 0.58 [0.29, 0.77],p< .001). However, the magnitude of the male-career association in the language spoken in a country predicts the magnitude of the male-career association measured via the behavioral IAT, controlling for both national GDP and median country age, in an additive mixed-effect model. b, Gender difference in STEM Self Efficacy6 (“The sex difference in self efficacy (boys – girls)”) as a function of male-career association measured in the Subtitle corpus. These two measures are correlated (r(28) = 0.59 [0.3, 0.79], p< .001): Countries with greater gender differences in self-efficacy also have greater gender bias present in their languages. Further, self-efficacy mediated the effect of language statistics on percentage of women in stem (path-ab = -0.33, p= 0.01), suggesting that language statistics could be critical causal factor underlying gender differences in STEM participation. |
| **Extended Data Fig. 10** |  |  |  |

***Delete rows as needed to accommodate the number of figures (10 is the maximum allowed).***

1. **Supplementary Information:**
2. **Flat Files**

**Complete the Inventory below for all additional textual information and any additional Supplementary Figures, which should be supplied in one combined PDF file.**

* **Row 1:** A combined, flat PDF containing any **Supplementary Methods, Supplementary Results, Supplementary Discussion, Supplementary Figures, Supplementary Tables (if simple), and Supplementary References**. Only one such file is permitted and all information must be organized under the above headings [no custom headings permitted, but there are no restrictions regarding subheadings].
* **Row 2:** Nature Research’s Reporting Summary; please provide an updated Summary, fully completed, without any mark-ups or comments. **Please note that this is a required document.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Item** | **Present?** | **Filename**  This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. The extension must be .pdf | **A brief, numerical description of file contents.**  i.e.: *Supplementary Discussion, Supplementary Figures 1-4, Supplementary Tables 1-4, Supplementary References.* |
| **Supplementary Information** | Choose an item. |  |  |
| **Reporting Summary** | Choose an item. |  |

1. **Additional Supplementary Files**

**Complete the Inventory below for all additional Supplementary Files that cannot be submitted as part of the Combined PDF.**

* Do not list Supplementary Figures in this table (see section 2A)
* Where possible, include the title and description within the file itself
* Spreadsheet-based tables and data should be combined into a workbook with multiple tabs, not submitted as individual files.
* Please note that the *ONLY* allowable types of additional Supplementary Files are:

|  |  |  |  |
| --- | --- | --- | --- |
| * Supplementary Tables | * Supplementary Audio | Suppl. Software | * Computational Data |
| * Supplementary Videos | * Supplementary Data |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Type** | **Number**  If there are multiple files of the same type this should be the numerical indicator. i.e. “1” for Video 1, “2” for Video 2, etc. | **Filename**  This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. i.e.: *Smith\_ Supplementary Video 1.mov* | **Legend or Descriptive Caption**  Describe the contents of the file |
| Choose an item. |  |  |  |
| Choose an item. |  |  |  |
| Choose an item. |  |  |  |
| Choose an item. |  |  |  |
| Choose an item. |  |  |  |
| Choose an item. |  |  |  |

***Add rows as needed to accommodate the number of files.***

1. **Source Data**

**Complete the Inventory below for all Source Data files.**

* Acceptable types of Source Data are:
  + Statistical Source Data
    - Excel format only
    - One file for each relevant Figure, containing all source data

|  |  |  |
| --- | --- | --- |
| **Figure** | **Filename**  This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. i.e.: *Smith\_Source Data Fig1.xls* | **Data description**  i.e.: Statistical Source Data, etc. |
| **Source Data Fig. 1** |  | Statistical source data |
| **Source Data Fig. 2** |  |  |
| **Source Data Fig. 3** |  |  |
| **Source Data Fig. 4** |  |  |
| **Source Data Fig. 5** |  |  |
| **Source Data Fig. 6** |  |  |
| **Source Data Fig. 7** |  |  |
| **Source Data Fig. 8** |  |  |
| **Source Data Extended Data Fig. 1** | ED1\_data.xls | Statistical source data |
| **Source Data Extended Data Fig. 2** | ED2\_data.xls | Statistical source data |
| **Source Data Extended Data Fig. 3** | NA | NA |
| **Source Data Extended Data Fig. 4** | ED4\_data.xls | Statistical source data |
| **Source Data Extended Data Fig. 5** | NA | NA |
| **Source Data Extended Data Fig. 6** | ED6\_data.xls | Statistical source data |
| **Source Data Extended Data Fig. 7** | ED7\_data\_1.xls | Statistical source data |
| **Source Data Extended Data Fig. 8** | ED8\_data.xls | Statistical source data |
| **Source Data Extended Data Fig. 9** | ED9\_data.xls | Statistical source data |
| **Source Data Extended Data Fig. 10** |  |  |