

## 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 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

Data analysis

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 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

# Behavioural & social sciences study design

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

|                   |  |
|-------------------|--|
| Study description | Study 1 was an archival quantitative cross-sectional study designed to quantify authentic self-expression on social media and estimate its correlation with subjective well-being. Study 2 was a quantitative experiment designed to examine if the relationship between authenticity and subjective well-being observed in Study 1 was causal such that authenticity increases subjective well-being.   |
| Research sample   | <p>The sample for Study 1 was a non-representative sample of Facebook users who used the MyPersonality application and agreed to submit their Facebook data for academic research. We were able to ascertain gender and age for a subset of participants; this included 6,648 participants in the Likes-based model (61.24% female, mean age = 24.97 years, SD age = 8.42 years) and 2,943 participants in the Language-based model (58.27% female, mean age = 25.97 years, SD age = 10.42 years). Sample size was determined based on data availability.</p> <p>Study 2 was a non-representative sample of college students who indicated they were active social media users (mean age = 22.98 years old, SD age = 4.17 years, 73.86% female). We aimed to recruit 200 students based on financial constraints, however data collection was halted at the onset of the COVID-19 pandemic and the closure of the campus research lab.</p> |
| Sampling strategy | Study 1 was an archival, cross-sectional, survey-based study limited based on data availability. Study 2 was an experiment, where participants were randomly assigned to one of two conditions. The sample size for Study 2 was originally determined based on financial limitations; data collection was halted at the onset of the COVID-19 pandemic and the closure of the campus research lab.   |
| Data collection   | In both studies, data were collected at a time and place of the participants choosing on an electronic device. In Study 1, data was collected using the MyPersonality application. The original data collection was not related to the present research question. In Study 2, data was collected using the online platform Qualtrics. The researcher was not blind to the condition or hypotheses.   |
| Timing            | <p>Data for Study 1 were collected between 2007-2012.</p> <p>Data for Study 2 were collected in February 2020.</p>   |
| Data exclusions   | 39 participants were excluded from Study 2. As defined in our per-registration plan, we excluded participants who indicated at the end of the two weeks that they did not take the study seriously.  |
| Non-participation | In Study 2, 71 participants dropped out (did not complete all check-in surveys).   |
| Randomization     | Study 1 did not involve randomization. Therefore to estimate the robustness of our findings, we included a number of control variables including the participant's self-reported personality traits and their most extreme trait. In a subsample, presented in the Supplementary Information, we were able to obtain gender and age which were used as additional controls. Our findings were largely robust to these control variables. In Study 2, participants were randomly assigned to one of two 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 type="checkbox"/>            | <input checked="" type="checkbox"/> Human research participants |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data                          |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern           |

### 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 |

## Human research participants

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

|                            |  |
|----------------------------|--|
| Population characteristics | See above  |
| Recruitment                | <p>In Study 1, participants were users of a Facebook application which involved personality assessments, therefore participants self-selected into the study. The purpose of the app was to give participants information about their personality, thus incentivizing honest responses. Additionally, participant personality expression was obtained from their Facebook profiles, providing ecological validity. This data set is limited to social media users and therefore cannot be generalized to those who do not use social media.</p> <p>In Study 2, participants were recruited through a University subject pool, and pre-screened for social media use. They self-selected into a study about social media, however were not told the explicit hypotheses of the study. This self-selection and sourcing of participants means that the observed sample may be younger and have a greater interest in social media related research than the average population. It may be that the effects of the treatments are overstated in the younger sample size. However, they may also be understated, given the importance and usage rates of social media among younger individuals.</p> |
| Ethics oversight           | Both studies were reviewed by the Columbia University Human Research Protection Office and Institutional Review Board (IRB). They determined that Study 1 did not qualify as Human Subjects Research because the data was archival. The procedures of Study 2 were approved by Columbia's IRB.   |

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