nature research

Corresponding author(s): Lei Wang

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

   Describe how sample size was determined.

   For the purpose of multilevel analysis, we ensured that there were at least 30 participants nested within each city or ZIP-code.

2. Data exclusions

   Describe any data exclusions.

   Study 1: To preclude reverse causality, where certain personalities may cause individuals to migrate to cities with certain temperatures, we limited our sample to students who had spent the entirety of their pre-college youth in their birthplace. To rule out another alternative explanation—that parents with certain personalities chose to migrate to a certain city and then gave birth to children who resemble their personalities—we further limited the sample to participants whose birthplace matched their ancestral home (i.e., jiguan, the home of their patrilineal ancestors). Importantly, all results remained substantively unchanged without these exclusion criteria.
   Study 2: We applied the following criteria. First, consistent with the Chinese data, participants must have chosen the United States as the country in which they spent most of their youth. Second, participants must have provided a meaningful U.S. ZIP code for the question "What is the ZIP-code/postal code of the place where you spent most of your youth?" Third, the ZIP code provided must have correctly matched their answer to the question "What state did you spend most of your youth?" Fourth, for the purpose of multilevel modeling15, we only included youth ZIP codes that had at least 30 participants (as in the Chinese data). Fifth, in line with prior work28, we only included participants between 16 to 60 years old due to the concern that older participants might be particularly susceptible to self-selection bias. Sixth, we only included participants who completed the study in English (as opposed to Dutch, German, or Spanish). Seventh, participants must have responded with "yes" to the question "Did you answer truthfully on all of these questions?" Lastly, participants must have responded with "no" to the question "Have you ever previously filled out this particular questionnaire on this site?"

3. Replication

   Describe whether the experimental findings were reliably reproduced.

   The results were reliably reproduced across the two samples.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   N/A

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   N/A

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☒	☐ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	We used R and Stata to analyze the data.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	No unique materials are used.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	N/A

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	N/A
b. Describe the method of cell line authentication used.	N/A
c. Report whether the cell lines were tested for mycoplasma contamination.	N/A
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	N/A

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	N/A

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	For study 1, we had 5,587 participants from 59 Chinese cities, with 42.4% females, Mage = 22.07, SDage = 2.05; and for study 2, we had 1,660,638 participants from 12,499 ZIP codes of the United States (65.3% female; Mage = 27.05 years, SDage = 11.00; 17.0% with a college degree, 9.44% with a graduate degree).