The WEIRD problem in a “non-WEIRD” context: A meta-research on the representativeness of human subjects in Chinese psychological research

Lei Yue [1, 2], Xi-Nian Zuo [3], Hu Chuan-Peng [1*]

1. School of Psychology, Nanjing Normal University, Nanjing, China.
2. School of Psychology, Central China Normal University, Wuhan, China.
3. State Key Laboratory of Cognitive Neuroscience and Learning, International Data Group/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China.

***Corresponding author:** Hu Chuan-Peng, School of Psychology, Nanjing Normal University, No. 122 Ninghai Road, Gulou District, Nanjing, China;

Email: hu.chuan-peng@nnu.edu.cn, or, hcp4715@hotmail.com

Abstract

Psychological science aims at understanding human mind and behavior, but it primarily relies on subjects from Western, Educated, Industrialized, Rich, and Democratic regions, i.e., the WEIRD problem. This lack of diversity and representativeness of subjects compromised the generalizability of psychological science. To address this issue, large-scale international collaborative projects were initiated, and more data are collected from non-WEIRD regions. However, it is unknown whether subjects from “non-WEIRD” regions can represent their local population. In this meta-research, we plan to survey the characteristics of Chinese subjects reported in empirical studies published in five mainstream Chinese psychological journals and in large-scale international collaborations. The results will provide a realistic picture of Chinese participants in psychology, and we will discuss potential solutions to the issue of representativeness in both China and worldwide.

Keywords: Meta-science; Population psychology; Representativeness; WEIRD; Generalizability

1 Introduction

Psychological science aims at understanding human mind and behaviour. However, it largely relies on unrepresentative human samples: most human participants in published psychological studies are undergraduate students who take psychology courses from "Western, Educated, Industrialized, Rich, and Democratic" (WEIRD) regions (Henrich et al., 2010; Henry, 2008; Sears, 1986). For example, Arnett (2008) analysed articles in six premier American Psychology Association (APA) journals and found that 96% relied on samples drawn from Western industrialized nations (Europe, North America, Australia, or Israel). More recent surveys found little change in the past decade (Nielsen et al., 2017; Pollet & Saxton, 2019; Rad et al., 2018). However, the population in WEIRD regions is only consistent less than ¼ of the global population (Henrich et al., 2010). The lack of representativeness in psychological science and related fields (such as cognitive neuroscience, (Zuo et al., 2019)) limits our understanding of the whole picture of human mind and behavior (Apicella et al., 2020; Barrett, 2020; Jones, 2010)[1] and may lead to incorrect policies (Arnett, 2008). This issue, combined with other methodological issues, created a generalizability crisis in psychology (Yarkoni, 2020).

As a starting point to solve this problem, researchers in the field started to include more diverse data. Many international collaborative projects have been initiated (Gordon et al., 2020; Moshontz et al., 2018). Typically, these projects invite collaborators globally, especially those from non-WEIRD regions, such as Asia, Middle East, Latin America, and Africa. These efforts are applaudable and indeed increased the geographical diversity and sample size of psychological science. These projects, however, have not examined whether

[1] While it is generally accepted that samples should be representative to the target population, Mook (1983) argued that generalization may be misplaced in some cases where showing some effects do exist, even in rare and artificial settings, is valuable. This argument is invalid because most psychological research aims higher than mere existence of certain effects (e.g., guide the polices, IJzerman et al., 2020). Also, focusing on a narrow sub-population, we may miss phenomena that are outside that sub-population and the consequence of these missed phenomena is unknown. Finally, the selection of samples reflect the fact that researchers themselves are from narrow sub-population, they may priorities the phenomena are importance to that sub-population and thus distort the whole picture of psychology.

data collected from non-WEIRD regions are representative of the local population. Left this issue unaddressed, these large collaboration projects may create an illusion that the representativeness problem can be solved by involving more researchers from non-WEIRD regions, ignoring the fact that data collected from non-WEIRD regions may suffer a problem of representativeness (see also Forscher et al., 2021). In fact, there are great variations within non-WEIRD regions (Ghai, 2021). However, the convenient sampling method employed by psychologists will cause the problem of unrepresentativeness in both WEIRD and non-WEIRD regions.

To understand how representative is the sample in psychological research from a typical non-WEIRD, China, we propose to survey the studies conducted by Chinese psychological researchers. China is the second-largest economy and has the largest population in the world, yet with a very different history and cultural tradition from the West. In recent years, Chinese researchers have actively participated in international collaborations (e.g., Human Penguin Projects, Many Labs 5, Psychological Science Accelerator). However, it is unknown whether Chinese psychological participants represent the Chinese population. By the word "represent", we mean the sample in a study (or studies) should be a miniature of the targeted population without selection biases, or theoretically can be a miniature of the targeted population and without selection bias (Kruskal & Mosteller, 1979a, 1979b, 1979c, 1980; Kukull & Ganguli, 2012).

In the current study, we will explore the representativeness of Chinese psychological participants by examining three issues (see Figure 1). Firstly, whether the characteristics of Chinese samples reported in large-scale international collaborations are similar to those reported in Chinese psychological journals. Secondly, to what extent the Chinese participants in psychological science represent the Chinese population, as compared with the census data from the National Bureau of Statistics of China and data from a large-scale social survey, Chinese Family Panel Study (CFPS). Lastly, we will explore the shared and distinct patterns of Chinese samples and samples from other regions. In addition, we will provide preliminary evidence about the "default participant" in Chinese psychological research. And we will compare the similarities and differences between the keywords of the big team projects and Chinese journals' articles based on bibliometric methods.

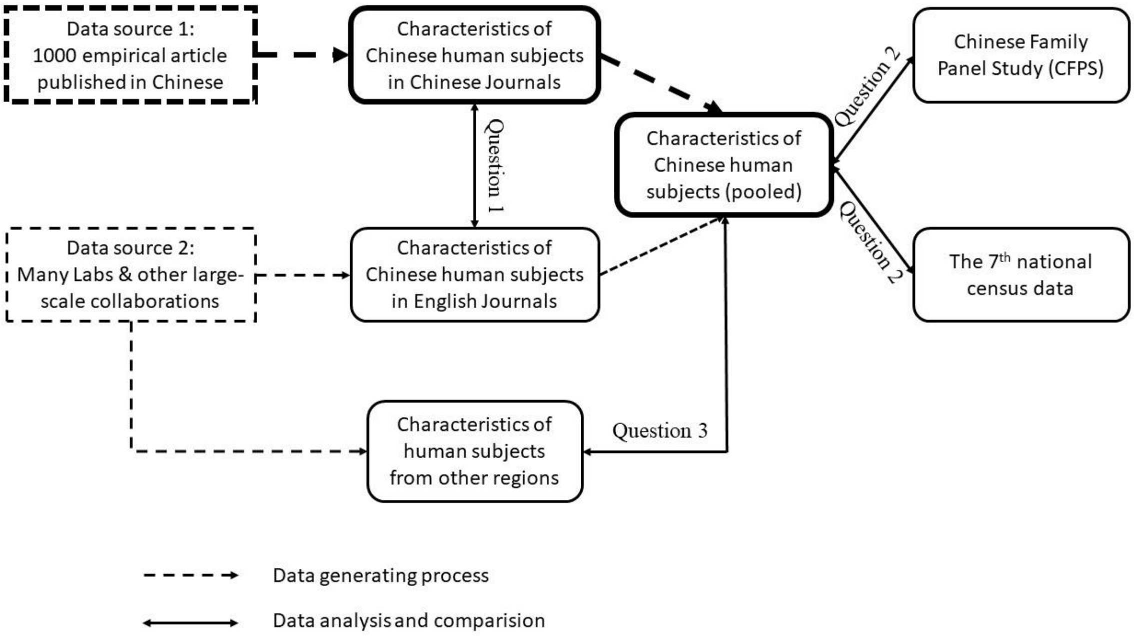


Figure 1. Schema of the current meta-research. Question 1: Whether characteristics of Chinese participants reported in large-scale international collaboration are similar to those reported in Chinese psychological journals, given they have the same target population; Question 2: To what extent the Chinese participants in psychological science can represent their target population, e.g., the whole population of census data from the National Bureau of Statistics of China and from a large-scale social survey, Chinese Family Panel Study (CFPS); Question3: What are the shared and distinct patterns of Chinese participants and participants from other regions?

2 Method

2.1 Data sources

Data will come from two sources. The first data source is 1000 empirical studies published in five mainstream Chinese journals: *Acta Psychological Sinica*, *Journal of Psychological Science*, *Chinese Journal of Clinical Psychology*, *Psychological Development and Education*, *Psychological and Behavioral Studies*. These journals are chosen because the following reasons. First, these five journals are indexed by CSSCI (Chinese Social Sciences Citation Index), which is regarded as authoritative and comprehensive database for bibliometric studies of China's social sciences (e.g., Gong & Cheng, 2022). Thus, all these five journals are selected as of high-quality among all Chinese psychological journals. Second, these five journals cover most fields of psychology. Among them, *Acta Psychological Sinica*, *Journal of Psychological Science*, and *Psychological and Behavioral Studies* are comprehensive journals, studies from all sub-fields of psychology are included; *Psychological Development and Education* is the only journal for developmental and educational psychology in China; *Chinese Journal of Clinical Psychology* focuses studies in clinical psychology and mental health. Of 1000 empirical papers, 500 were published between 2017 to 2018 that were selected by Wang et al. (2021). Another 500 papers will be selected from the same five journals but published at different time points. To be specific, we will select 250 articles published in 2008 and 250 articles in 2020~2021. The criteria and procedure of article selection, same as Wang et al. (2021), are described as below:

Step 1: Assigning identifiers to articles. For papers published in different periods, we will obtain information of all the papers published in those journals in three different periods and assign a unique identifier to each article. Each article ID has 8 digits. The first four number represents the selected period (we will use 2008, 2018, 2021 to represent articles from three different periods), the fifth number represents the journal ID, from 1 to 5, and the last three number represent the order of the paper in the journal. For example, the first article in *Acta Psychologica Sinica* from 2008 is coded as 20081001, the second article in *Acta Psychologica Sinica* at the same year is coded as 20081002. All articles and their ID can be found at https://osf.io/avb7t/?view_only=a7e4610491374093851fc2b7da57e85c.

Step 2: Random sampling from all articles. We will use the `*sample*` function of R base to randomly select a certain number of papers from all the papers in each journal (see the code at https://osf.io/avb7t/?view_only=a7e4610491374093851fc2b7da57e85c). The number of papers sampled from each journal will be weighted by the total number of papers published in that year (the total papers each period and the number of sampled papers, see Table 1). After getting the identifiers of the selected papers, two independent researchers will check each article to make sure that it is an empirical study. If not, we will replace the article with the empirical article, which has the smallest distance to the article whose identifier is sampled.

The second source of data will come from large-scale international collaborations that are aimed at addressing the WEIRD problem. More specifically, we will check the data from all Many Labs projects (especially Man Lab 2 (Klein et al., 2018)), the Human Penguin Project (Hu et al., 2019; IJzerman et al., 2018), and all finished projects from PSA (Jones et al., 2021; Wang et al., 2021). These projects were chosen because they opened raw data. If possible, we will also include data from other large-scale collaborations which contain samples from China. We will search and extract demographical characteristics of Chinese samples and other samples reported in those studies.

Table 1 The number of articles in five Chinese Psychological Journals

Journal	2008	2017~2018	2020~2021
Acta Psychological Sinica	138 (39)	246 (95)	91 (28)
Journal of Psychological Science	379 (107)	299 (115)	203 (61)
Chinese Journal of Clinical Psychology	227 (64)	379 (146)	310 (94)
Psychological Development and Education	87 (24)	162 (62)	95 (29)
Psychological and Behavioral Studies	57 (16)	213 (82)	125 (38)

Note: Each column includes the total number of articles published in each journal at that time interval and the number of articles selected (inside parentheses)

2.2 Articles code

We will extract the data from the first source. The data extraction procedure has three stages: pre-coding, coding, and proofreading.

In the pre-coding stage, we first developed the initial version code manual based on the previous study (Arnett, 2008; Nielsen et al., 2017; Pollet & Saxton, 2019; Rad et al., 2018). Then, at least two coders will code ten random articles independently, they will compare the results, resolve the differences and revise the manual. After that, they will code another ten articles and compare the results and revise the coding manual again. This procedure will iterate until the disagreement between two coders is negligible.

When the formal coding manual is established, we will start to code all 1000 papers. In this stage, we will randomly divide the 1000 papers into several parts. For each part, there will be two coders who independently extract data from papers based on the coding manual. Each coder will go through the methods section and further inspect the data used in those studies. Note that studies used secondary data, data from web or app scraping, large-scale databases or using animals, or case studies will be excluded. For the remaining studies, we will extract the following information of the study: articles IDs, article title, study number, participants’ group, study type, sample type, sample size, sampling method, and methods for participant recruitment. More importantly, we will extract all information, if available, about participants: sex, age, socio-economic status, educational attainment, ethnicity, occupation, religion, and region for participants' recruitment.

Given that representativeness depends on the target population, we will also read the conclusion or other parts of articles to extract statements about the target population to which studies were intended to be generalized. However, representativeness has not been taken seriously in the field (see, Thalmayer et al., 2021). This situation is similar to causality in psychology (Grosz et al., 2020): researchers may use vague statements about representativeness or generality. Our initial coding, based on a few papers from the same 5 journals but not the final sample of papers, suggested that it is difficult to code the target population. To make the coding task more doable, we added the two items related to the target population (or generality) in our coding manual. The first item is the target population to which the studies intended to generalize. We will divide all the subjects into four categories: stated specific population; inferred specific population; inferred general population; stated general population. The second item is the exact sentences/words excerpted from the full text of the paper that is associated with the statement about the target population. The coder will be instructed to search sentences related to the target population in the conclusion section of the papers. If no related information was found, they will search for information in other parts of the articles (firstly introduction, and then, results or other parts). These two items will code both the target population but also keep the transparency of the coding process. See the supplemental document "Code_Manual_Chin_Subj_V2" for more details (https://osf.io/avb7t/?view_only=a7e4610491374093851fc2b7da57e85c).

Additionally, we will export the keywords of Chinese journals’ articles and the big team projects through CNKI (China National Knowledge Infrastructure) and Web of Science, and then, if feasible, try to use bibliometric methods to compare the similarities and differences between the keywords of the big team projects and Chinese journals’ articles. Also, we will record when did researchers do not report details of the sample to provide preliminary evidence about the “default participant” in Chinese psychological research. Specifically, we will first recode whether the articles mention participants' information in the abstract. And then we will code distinguish studies that relied only on college students' samples and studies used other samples. We will then compare the percentage of mentioned and not mentioned in two groups of studies (see below for the template of our table 2).

Table 2. Different study types and their sample mentions.

Study type	Samples mentioned	Sample not mentioned	Total
Only college students			
College students & other populations			
Only sample outside colleges			
Total			

To ensure the accuracy of the coding content, the results from the two coders are compared after completing the initial coding. Two coders rate the consistency of each article from 0 to 1, with 0 represents completely different and 1 represents identical. This consistency score will be then used for calculating the inter-rater reliability. We will use the R package *irr* for this index (Gamer et al., 2019).

2.3 Data analysis

We will use R 4.1.1 to pre-process and visualise data (R Core Team, 2021) and Bayes factor analyses. Bayes factor was chosen because it can provide evidence for both null hypothesis and the alternative hypothesis (Dienes, 2016; Dienes & Mclatchie, 2018; Hu et al., 2018; Wagenmakers et al., 2018). Given that the final data we use are percentage data, we will use Bayesian multinomial test (corresponding to frequentists' goodness-of-fit test or χ^2) to test whether percentage data from two sources differ from each other on certain dimension (e.g., sex, age, education attainment).

The percentage data from one source is treated as the observed and the other is treated as expected. The null hypothesis (H_0) is that the observed percentage data are sampled from a multinomial distribution with parameters as defined by the expected percentage, the alternative hypothesis (H_a) is that the observed proportion data are sample from a multinomial distribution with equal probability for each cell. The multinomial distribution is a generalization of the binomial distribution to variables that can take values in $K \geq 2$ categories. The parameters of multinomial is a vector of probabilities, $\theta = (\theta_1, \theta_2, \dots, \theta_K)$, with which N observations are distributed across K categories. The distribution of the parameters of multinomial distribution follows a Dirichlet distribution with concentration parameters $(\alpha_1, \alpha_2, \ldots, \alpha_K)$, where each α is larger than zero. In Bayesian multinomial test, the null hypothesis (H_0) is a point hypothesis that the parameters, θ_s, of the observed percentage data equal to the expected, which is a point in the Dirichlet distribution, see Sarafoglou et al. (2020) for details. Usually, the H_0 is tested against the encompassing hypotheses, or alternative hypothesis (H_a), that all category proportions are free to vary. For example, when testing whether the Chinese participants in PSA001 data (Jones et al., 2021) represent the Chinese population in terms of age, we first calculate the percentage of participants in each of seven age bins (0 ~ 9, 10 ~ 19, 20 ~ 29, 30 ~ 39, 40 ~ 49, 50 ~ 59, >= 60) for PSA001 data. The result, [0, 21, 71, 1, 5, 2, 0], is treated as the observed and compared to the expected percentage data, [12, 11, 12, 16, 14, 16, 19], which is from the 7th census data of China. In this case, the H_0 and H_a are specified as below:

H_0: [0, 21, 71, 1, 5, 2, 0] is sampled from a multinomial distribution with $P = Pr(x_1, x_2, \ldots, x_7 \mid n = 100, p_1 = 0.12, p_2 = 0.11, p_3 = 0.12, p_4 = 0.16, p_5 = 0.14, p_6 = 0.16, p_7 = 0.19)$, which is a point in a Dirichlet distribution.

H_a: [0, 21, 71, 1, 5, 2, 0] is sampled from a multinomial distribution with $P = Pr(x_1, x_2, \ldots, x_7 \mid n = 100, p_1, p_2, \ldots, p_7)$, where $(p_1, p_2, \ldots, p_7)$ is distributed as a Dirichlet distribution with concentration parameter $(\alpha_1, \alpha_2, \ldots, \alpha_7)$.

An non-informative prior $(\alpha_1 = 1, \alpha_2 = 1, \ldots, \alpha_K = 1)$ of Dirichlet distribution is chosen because it is relatively diffused. We use non-informative prior for testing all hypotheses[2]. Note that the expected percentage data may vary in length, as we will describe below. Bayes factor will be interpreted as recommended in (Wagenmakers et al., 2018): BF_{10} >= 10 or $\log(BF_{10})$ >= 2.303 means strong evidence for H_a, and 6 <= BF_{10} < 10 or 1.792 <= $\log(BF_{10})$ < 2.303 means moderate evidence for H_a, BF_{10} <= 1/10 or $\log(BF_{10})$ <= -2.303

[2] We conducted simulation analyses to test to what extend the Bayesian factor analysis with non-informative prior and 3 as the criterion detect an difference. For Bayesian factor analysis for sex ratio, the results revealed that we can have detect deviation from $p = 0.5$ with more 80% of the times if the difference is greater than 0.17. For Bayesian factor analysis for age distribution with five bins, the results revealed that for 93.8% of the multinomial probabilities generated by an “uniform” Dirichlet distribution, the current setting can provide evidence that the probability is different from null with in 80% of the case. See supplementary and our R Notebook for the details of simulation.

means strong evidence for H_a, and 1/10 <= BF_{10} < 1/6 or -2.303 <= log(BF_{10}) < -1.792 means moderate evidence for H_0. We implement Bayesian multinomial test using R code based on JASP 0.16.4 (see the R script in R Notebook for more details).

For our first question, whether there are differences between Chinese human subjects reported in Chinese journals and in large-scale international collaborative projects, we will first visualize the proportion of the reported information of subjects and, then, we will compare subjects from Chinese psychology journals and Chinese subjects from the international collaborative projects with regard to sex, age, and, if possible, geographical distribution. Given that studies from Chinese psychological journals may have different target populations as compared to international collaborative projects, we compare samples from studies that share the same target population. More specifically, only if articles from Chinese psychological journals and international collaborative projects targeted the general population (inferred or stated Chinese population or humans), we will compare their sample characteristics. In the same vein, samples from other shared target populations by both Chinese psychological journal articles and international collaborative projects, e.g., adolescents, will also be compared.

We will also use Bayes factor for statistical inference. The data from international collaborative projects will be as observed and the data from Chinese psychological journals will be used as expected. More specifically, for the sex distribution, we will test whether the sex ratio of subjects from international collaborative projects is sampled from a distribution with parameters equal to the proportions of data extracted from Chinese psychology journals. The null hypothesis (H_0) is that the parameters of observed data are equal to the Chinese samples from Chinese psychology journals. The H_a is that the parameters of the observed data are free to vary instead of a fixed point[3].

For the age distribution, we will use two different approaches. The first approach is similar to census data's age bins: 0 ~9, 10 ~ 19, 20 ~ 29, 30 ~ 39, 40 ~ 49, 50 ~ 59,

[3] When H_0 is a point null hypothesis and the H_a is a distribution, the Bayes factor is usually calculated by Savage-Dickey ratio or similar approach (e.g., Wagenmakers et al., 2010; Sarafoglou et al., 2020).

and >=60[4]. The second one is based on the developmental stage, which is more important in psychological research than age itself. We created five age bins based on developmental stages: 0 ~ 17 (children and adolescents), 18 ~ 25 (early adulthood), 26 ~ 40 (middle adulthood), 41 ~ 59 (later adulthood), and >= 60 (elders). Unless stated, we report statistical results based on the second set of age bins in the main text and results based on the first set of age bins in the supplementary results. The H_0 is that the age percentage of Chinese subjects from international collaborative projects are sampled from a multinomial distribution with parameters same as the percentage data of Chinese subjects from Chinese psychology journals, and the H_a is the parameters of Chinese subjects from international collaborative projects are free to vary. For data extracted from Chinese journal articles, we will estimate the number of participants in each age bin using Monte Carlo simulation, based on the reported age information (i.e., mean and SD of age as reported in articles). For example, an article reported 30 participants, with age = 23.3 ± 3.5, we estimate the approximate number of participants under 20 is 5 (r code: `round((pnorm(20, mean = 23.3, sd =3.5) * 30))`), the number of participants aged between 21 ~ 30 is 24 (r code: `round((pnorm(30, mean = 23.3, sd =3.5) * 30)) - 5`), and participant aged between 30 ~ 40 will be 1.

Format of the data for this question is illustrated by fake data, see "figure2a_sex_template.jasp" for the test for sex distribution and "figure2b_age_template.jasp" for the test for age distribution at https://osf.io/y9hwq/.

The second question of this study is whether all Chinese samples data available, regardless of the sources of the data (see Figure 1), come from a very narrow slice of the Chinese population. Given sample representativeness indeed depend on the target population, we will further distinguish two types of analyses. For studies that targeted the general population, inferred or stated, we will compare their sample characteristics to the whole census data from the National Bureau of Statistics. For studies that targeted a specific population, we will compare the sample characteristics to that specific population selected from census data. If the information of that specific population is not available in census data,

[4] Note that a more fine-grained age bins (with 5-year intervals) were only used for pyramid plot but not for statistical analysis.

we will search for other reliable data sources as the reference data. We will also use Bayesian multinomial test for statistical inference. For those targeted at the general population, we will use the pooled Chinese subjects' data from both sources as the "observed" and the census data as the "expected". The age bins will be the same as we testing the first hypothesis. The prior setting of Bayes factor analyses and the criteria for interpreting Bayes factor will be the same as in testing our first question.

For the third question, whether Chinese psychological samples differ from other countries' samples, we will visualize the distributions of the participants from different countries and visually compare other countries with the Chinese samples. We will also compare each country's data with Chinese data using Bayesian multinomial test, where the other countries' data are treated as "observed" and Chinese data are "expected. The null hypothesis (H_0) is that the parameters of the observed sex and age distributions equal the expected proportions (i.e., Chinese psychological samples). The alternative (H_a) is that those parameters are free to vary. The prior setting of Bayes factor analyses and the criteria for interpreting Bayes factor will be the same as in testing our first question.

It should be noted that the graphics mentioned above are not fixed, and we will choose the graphics that can best illustrate the data characteristics according to the actual situation. The specific analysis code will be updated in OSF (https://osf.io/avb7t/?view_only=a7e4610491374093851fc2b7da57e85c) or Gitee (https://gitee.com/hcp4715/chin-subj).

3 Results

3.1 Overview of participants

[Here we will insert the graph info with China's map; relative density of participants in different dimensions]

3.2 Comparing Chinese papers and international collaborations

We predict that there will be moderate to strong evidence that the Chinese samples in Chinese papers and in international collaborations have similar sex ratio, age distribution, and distribution along other dimensions (if data are available), $BF_{01}s \geq 6$.

[*Here we will insert Figure 2 to visualize the comparisons between two data sources. Also, Bayes factors will also be reported here*]

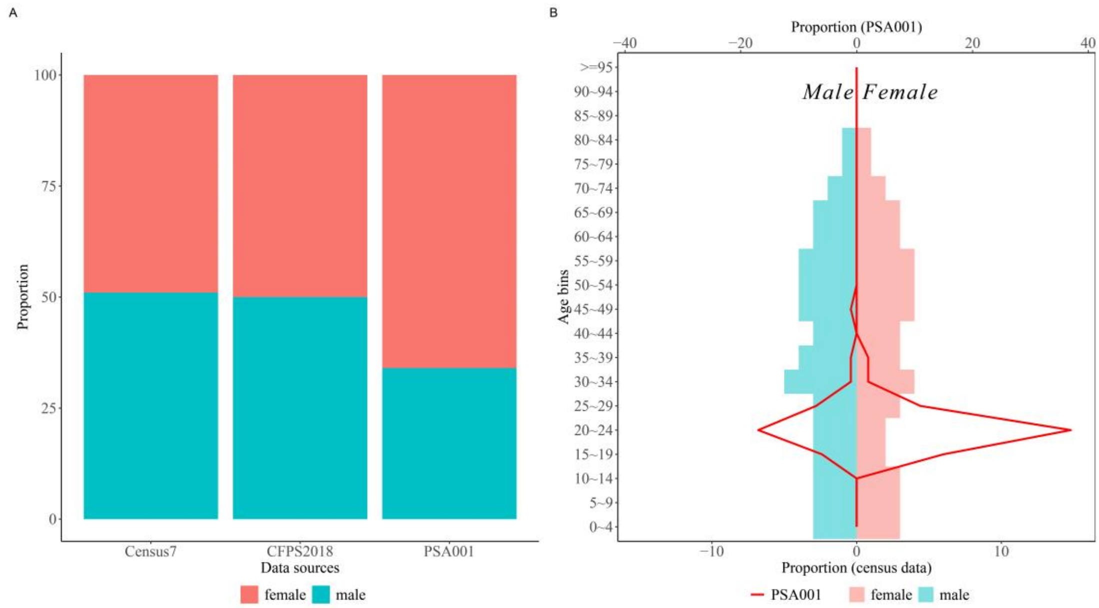
3.3 Comparing Chinese samples and samples from census data and CFPS

We will pool data from both data sources and compared with CFPS and census data. [The results will be reported as below, using data from international collaboration data]

We used PSA 001 data (Jones et al., 2021) as the Chinese sample to demonstrate the analyses and visualization, see below (all code is available at osf.io/y9hwq/). Note that these results will be replaced by the final results after data collected and analyses carried out.

First, we compared the Chinese samples (Jones et al., 2021) with the CFPS data in 2018 with China's censuses data. We tested whether the sex ratio in psychological sample is different from that of the census data using Bayesian multinomial test (Bayesian version of Goodness-of-fit). The results revealed strong evidence that the psychological sample data (Jones et al., 2021) is different from the census data, $log(BF_{10}) = 3.73$. *In contrast to psychological sample data, data from sociology, CFPS 2018 data, is not different from census data,* $log(BF_{10}) = -2.06$. *As we can see from Figure 3 A, Chinese psychological science sample included more female participants, while the CFPS data has a similar pattern as the census data.*

For the age distribution, we found that the psychological samples' age distribution is different from that of the census data, with strong evidence from Bayesian multinomial test, $log(BF_{10}) = 162.67$. *This difference is further revealed by the demographic pyramid (See figure 3 B), which showed that the Chinese psychological samples consist of females aged 15~24 years.*



Figures 3. Preliminary results of the sex and age distribution from different data sources. (A) Sex ratio from the 7th census data, CFPS 2018 data, and psychological science sample (PSA 001's data is used as an example); (B) Age distribution of the 7th census data the transparent bar plot, and psychological science samples (PSA 001's data as an example), the y-axis is age bins, the x-axis on the top is for the line plot of PSA 001 and x-axis on the bottom is for the pyramid plot of the 7th census data.

3.4 Comparing Chinese samples and samples from other countries

We also explore the common and distinct pattern between Chinese psychological samples and psychological samples from other regions.

The preliminary results from available data illustrate how the final results will look like. These results will be replaced by the final results after data collection. For sex ratio, the pairwise Bayesian multinomial test with non-informative revealed that data from 14 countries have different sex ratios as compared to Chinese psychological samples (see Figure 4A). Similar pattern was obtained for Bayesian multinomial test with informative prior but overall evidence for H_0 is weaker. However, only one of them (Indian samples) has lower proportion of females than Chinese samples, all other 13 countries have higher proportion of female participants than Chinese samples (see Figure 4C). For age distribution, the pairwise Bayesian multinomial test with non-informative revealed strong evidence that samples from *twenty-five countries are the same as Chinese psychological samples (see Figure 4B, 4D). These preliminary results indicated that the psychological samples from many regions are similar, probably most of them are college students or communities around university campuses (Arnett, 2008), but also there is variability in both sex ratio and age distribution.*

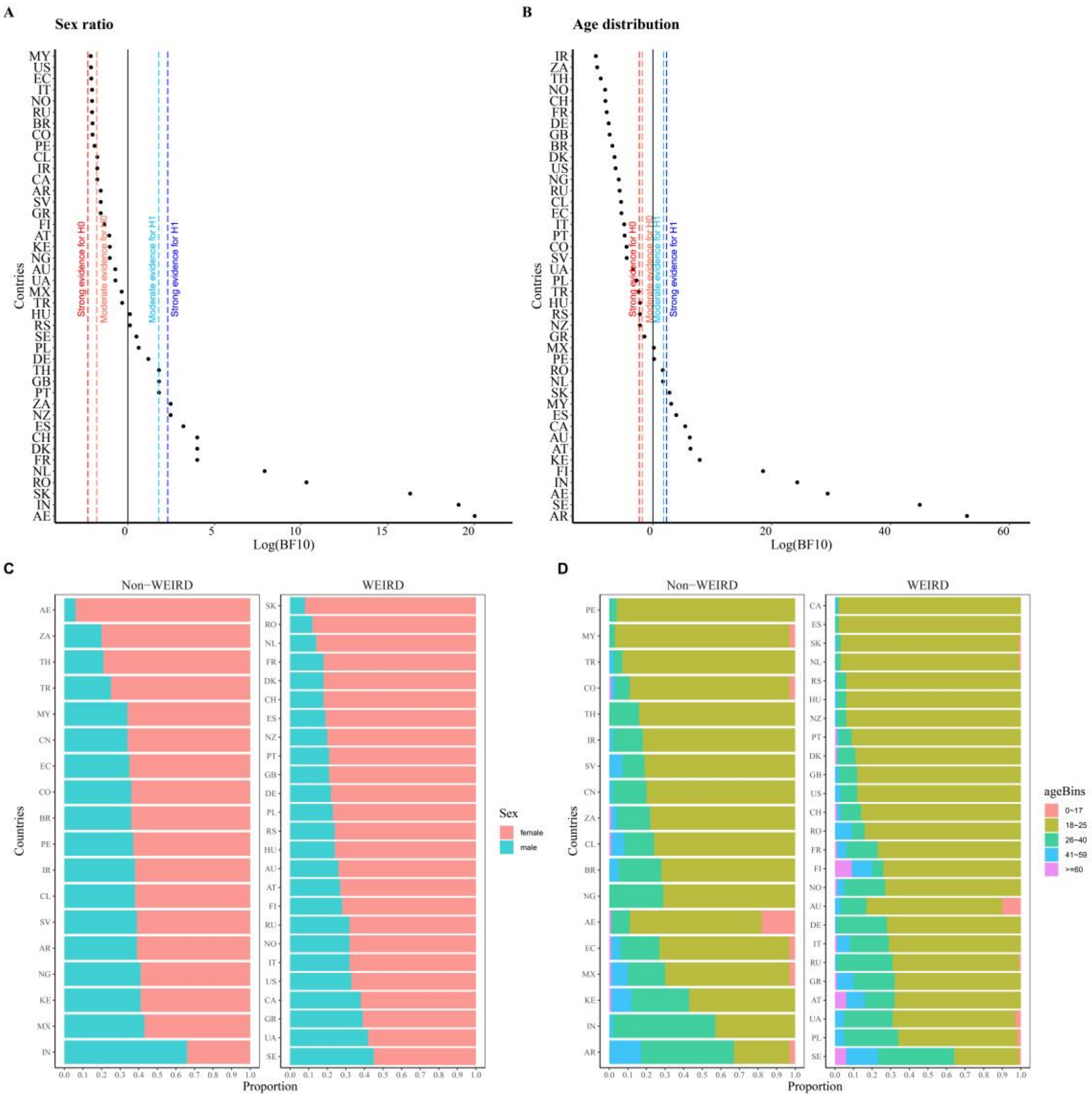


Figure 4. The sex and age distribution from PSA 001. (A) Pairwise comparisons of sex ratio between Chinese psychological sample and available data from other countries; (B) Pairwise comparisons of age distribution between Chinese psychological sample and available data from other countries; (C) Sex ratio of all data; (D) Age proportion of all data. Country code: United Arab Emirates (AE), South Africa (ZA), Thailand (TH), Turkey (TR), Malaysia (MY), China (CN), Ecuador (EC), Colombia (CO), Brazil (BR), Peru (PE), Iran (IR), Chile (CL), El Salvador (SV), Argentina (AR), Nigeria (NG), Kenya (KE), Mexico (MX), India (IN), Slovakia (SK), Romania (RO), Netherlands (NL), France (FR), Denmark (DK), Switzerland (CH), Spain (ES), New Zealand (NZ), Portugal (PT), United Kingdom (GB), Germany (DE), Poland (PL), Serbia (RS), Hungary (HU), Australia (AU), Austria (AT), Finland (FI), Russia

(RU), Norway (NO), Italy (IT), United States (US), Canada (CA), Greece (GR), Ukraine (UA), Sweden (SE).

4 Discussion

[recap of the results]

…..

Overview of the research question, hypotheses, analytical plan, and interpretations of the current study

Question	Hypothesis	Analysis Plan	Interpretation given different outcomes
1. Whether characteristics of Chinese participants reported in large-scale international collaborative projects are similar to those reported in Chinese psychological journals?	H1: There is no difference between Chinese human subjects reported in Chinese journals and in large-scale international collaborative projects.	We will visualize and compare the sample characteristics from Chinese psychological journals and international collaboration projects based on their targeted population. Also, we will use Bayesian multinomial test to inspect their similarity.	H1 is supported if the background information of the subjects in the Chinese psychological articles is similar to that in the international collaborative projects (visual inspection). If $BF_{01} > = 6$, we infer there is relatively strong evidence that this hypothesis is supported.
2. To what extent the Chinese participants in psychological science can represent Chinese population, as compared with the census data from the National Bureau of Statistics of China and from a large-scale social survey, Chinese Family Panel Study (CFPS).	H2: As WEIRD sample only represents a narrow slice of human beings, the Chinese samples also come from a very narrow slice of the Chinese population.	We will visualize the Chinese sample (Chinese journal articles and international collaboration projects) to inspect their representativeness relative to their targeted population from reliable data sources (e.g., census data). For example, when the articles are targeting at the general population, we will compare the sample to the whole population of census data. We will also use Bayesian multinomial test to compare the Chinese sample with census data (or CFPS or other reliable data).	H2 is supported if the subject in Chinese psychology is from a narrow slice of the entire population in China, which is estimated by the census data from the National Bureau of Statistics of China and CFPS (visual inspection). If $BF_{10} > = 6$, we infer there is relatively strong evidence that this hypothesis is supported.
3. What are the shared and distinct patterns of Chinese participants and participants from other regions?	H3: Chinese human subjects share many characteristics as most other non-WEIRD and WEIRD samples.	We will visualize the distributions of the participants from different countries and visually compare others countries with the Chinese samples. We will also compare the distributions of participants with that of the Chinese participants, with the data from Chinese participants as expected and data from other countries as observed.	H3 is supported if the characteristics of subjects in Chinese psychology articles share characteristics with more than half of other non-WEIRD regions or WEIRD regions in describing statistical results and visualizations. If $BF_{01} > = 6$ for more than half of the other countries, we infer there is relatively strong evidence that this hypothesis is supported.

Reference

Apicella, C., Norenzayan, A., & Henrich, J. (2020). Beyond WEIRD: A review of the last decade and a look ahead to the global laboratory of the future. *Evolution and Human Behavior*, *41*(5), 319–329. https://doi.org/10.1016/j.evolhumbehav.2020.07.015

Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist*, *63*(7), 602–614. https://doi.org/10.1037/0003-066X.63.7.602

Barrett, H. C. (2020). Towards a cognitive science of the human: Cross-cultural approaches and their urgency. *Trends in Cognitive Sciences*, *24*(8), 620–638. https://doi.org/10.1016/j.tics.2020.05.007

Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour*, *5*, 980–989. https://doi.org/10.1038/s41562-021-01143-3

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, *72*, 78–89. https://doi.org/10.1016/j.jmp.2015.10.003

Dienes, Z. (2021). How to use and report Bayesian hypothesis tests. *Psychology of Consciousness: Theory, Research, and Practice*, *8*(1), 9–26. https://doi.org/10.1037/cns0000258

Dienes, Z., & Mclatchie, N. (2018). Four reasons to prefer Bayesian analyses over significance testing. *Psychonomic Bulletin & Review*, *25*(1), 207–218. https://doi.org/10.3758/s13423-017-1266-z

Forscher, P. S., Basnight-Brown, D. M., Dutra, N., Adetula, A., Silan, M., & IJzerman, H. (2021). *Psychological science needs the entire globe, part 3: Does psychological science deserve brilliant researchers from outside north America and Europe?* Association For Psychological Science. https://www.psychologicalscience.org/observer/entire-globe-part-3

Gamer, M., Lemon, J., & Singh, I. F. P. (2019). *irr: Various coefficients of interrater reliability and agreement (R package version 0.84.1) [Computer software]*. https://CRAN.R-project.org/package=irr

Ghai, S. (2021). It’s time to reimagine sample diversity and retire the WEIRD dichotomy. *Nature Human Behaviour*, *5*, 971–972. https://doi.org/10.1038/s41562-021-01175-9

Gong, K., & Cheng, Y. (2022). Patterns and impact of collaboration in China’s social sciences: cross-database comparisons between CSSCI and SSCI. *Scientometrics*. https://doi.org/10.1007/s11192-022-04483-7

Gordon, M., Viganola, D., Bishop, M., Chen, Y., Dreber, A., Goldfedder, B., Holzmeister, F., Johannesson, M., Liu, Y., Twardy, C., Wang, J., & Pfeiffer, T. (2020). Are replication rates the same across academic fields? Community forecasts from the DARPA SCORE programme. *Royal Society Open Science*, *7*(7), 200566. https://doi.org/10.1098/rsos.200566

Grosz, M. P., Rohrer, J. M., & Thoemmes, F. (2020). The taboo against explicit causal inference in nonexperimental psychology. *Perspectives on Psychological Science*, *15*(5), 1243–1255. https://doi.org/10.1177/1745691620921521

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Henry, P. J. (2008). College sophomores in the laboratory redux: Influences of a narrow data base on social psychology's view of the nature of prejudice. *Psychological Inquiry*, *19*(2), 49–71. https://doi.org/10.1080/10478400802049936

Hoijtink, H., Mulder, J., van Lissa, C., & Gu, X. (2019). A tutorial on testing hypotheses using the Bayes factor. *Psychological Methods*, *24*(5), 539–556. https://doi.org/10.1037/met0000201

Hu, C.-P., Kong, X.-Z., Wagenmakers, E.-J., Ly, A., & Peng, K. (2018). Bayes Factor and Its Implementation in JASP: A Practical Primer (in Chinese). *Advances in Psychological Science*, *26*(6), 951–965. https://doi.org/10.3724/SP.J.1042.2018.00951

Hu, C.-P., Yin, J.-X., Lindenberg, S., Dalğar, İ., Weissgerber, S. C., Vergara, R. C., Cairo, A. H., Čolić, M. V., Dursun, P., Frankowska, N., Hadi, R., Hall, C. J., Hong, Y., Joy-Gaba, J., Lazarević, D., Lazarević, L. B., Parzuchowski, M., Ratner, K. G., Rothman, D., Sim, S., Simão, C., Song, M., Stojilović, D., Blomster, J. K., Brito, R., Hennecke, M., Jaume-Guazzini, F., Schubert, T. W., Schütz, A., Seibt, B., Zickfeld, J. H., & Ijzerman, H. (2019). Data from the Human Penguin Project, a cross-national dataset testing social thermoregulation principles. *Scientific Data*, *6*, 32. https://doi.org/10.1038/s41597-019-0029-2

Ijzerman, H., Lewis, N. A., Przybylski, A. K., Weinstein, N., DeBruine, L., Ritchie, S. J., Vazire, S., Forscher, P. S., Morey, R. D., Ivory, J. D., & Anvari, F. (2020). Use caution when applying behavioural science to policy. *Nature Human Behaviour*, *4*(11), 1092-1094. https://doi.org/10.1038/s41562-020-00990-w

Ijzerman, H., Lindenberg, S., Dalğar, İ., Weissgerber, S. S. C., Vergara, R. C., Cairo, A. H., Čolić, M. V., Dursun, P., Frankowska, N., Hadi, R., Hall, C. J., Hong, Y., Hu, C.-P., Joy-Gaba, J., Lazarević, D., Lazarević, L. B., Parzuchowski, M., Ratner, K. G., Rothman, D., Sim, S., Simão, C., Song, M., Stojilović, D., Blomster, J. K., Brito, R., Hennecke, M., Jaume-Guazzini, F., Schubert, T. W., Schütz, A., Seibt, B., & Zickfeld, J. H. (2018). The Human Penguin Project: Climate, social integration, and core body temperature. *Collabra: Psychology*, *4*(1), 37. https://doi.org/10.1525/collabra.165

Jamil, T., Ly, A., Morey, R. D., Love, J., Marsman, M., & Wagenmakers, E.-J. (2017). Default “Gunel and Dickey” Bayes factors for contingency tables. *Behavior Research Methods*, *49*(2), 638–652. https://doi.org/10.3758/s13428-016-0739-8

Jones, B. C., DeBruine, L. M., Flake, J. K., Liuzza, M. T., Antfolk, J., Arinze, N. C., Ndukaihe, I. L. G., Bloxsom, N. G., Lewis, S. C., Foroni, F., Willis, M. L., Cubillas, C. P., Vadillo, M. A., Turiegano, E., Gilead, M., Simchon, A., Saribay, S. A., Owsley, N. C., Jang, C., Mburu, G., Calvillo, D. P., Wlodarczyk, A., Qi, Y., Ariyabuddhiphongs, K., Jarukasemthawee, S., Manley, H., Suavansri, P., Taephant, N., Stolier, R. M., Evans, T. R., Bonick, J., Lindemans, J. W., Ashworth, L. F., Hahn, A. C., Chevallier, C., Kapucu, A., Karaaslan, A., Leongómez, J. D., Sánchez, O. R., Valderrama, E., Vásquez-Amézquita, M., Hajdu, N., Aczel, B., Szecsi, P., Andreychik, M., Musser, E. D., Batres, C., Hu, C.-P., Liu, Q.-L., Legate, N., Vaughn, L. A., Barzykowski, K., Golik, K., Schmid, I., Stieger, S., Artner, R., Mues, C., Vanpaemel, W., Jiang, Z., Wu, Q., Marcu,

G. M., Stephen, I. D., Lu, J. G., Philipp, M. C., Arnal, J. D., Hehman, E., Xie, S. Y., Chopik, W. J., Seehuus, M., Azouaghe, S., Belhaj, A., Elouafa, J., Wilson, J. P., Kruse, E., Papadatou-Pastou, M., De La Rosa-Gómez, A., Barba-Sánchez, A. E., González-Santoyo, I., Hsu, T., Kung, C.-C., Wang, H.-H., Freeman, J. B., Oh, D. W., Schei, V., Sverdrup, T. E., Levitan, C. A., Cook, C. L., Chandel, P., Kujur, P., Parganiha, A., Parveen, N., Pati, A. K., Pradhan, S., Singh, M. M., Pande, B., Bavolar, J., Kačmár, P., Zakharov, I., Álvarez-Solas, S., Baskin, E., Thirkettle, M., Schmidt, K., Christopherson, C. D., Leonis, T., Suchow, J. W., Olofsson, J. K., Jernsäther, T., Lee, A.-S., Beaudry, J. L., Gogan, T. D., Oldmeadow, J. A., Balas, B., Stevens, L. M., Colloff, M. F., Flowe, H. D., Gülgöz, S., Brandt, M. J., Hoyer, K., Jaeger, B., Ren, D., Sleegers, W. W. A., Wissink, J., Kaminski, G., Floerke, V. A., Urry, H. L., Chen, S.-C., Pfuhl, G., Vally, Z., Basnight-Brown, D. M., Jzerman, H. I., Sarda, E., Neyroud, L., Badidi, T., Van der Linden, N., Tan, C. B. Y., Kovic, V., Sampaio, W., Ferreira, P., Santos, D., Burin, D. I., Gardiner, G., Protzko, J., Schild, C., Ścigała, K. A., Zettler, I., O'Mara Kunz, E. M., Storage, D., Wagemans, F. M. A., Saunders, B., Sirota, M., Sloane, G. V., Lima, T. J. S., Uittenhove, K., Vergauwe, E., Jaworska, K., Stern, J., Ask, K., van Zyl, C. J. J., Körner, A., Weissgerber, S. C., Boudesseul, J., Ruiz-Dodobara, F., Ritchie, K. L., Michalak, N. M., Blake, K. R., White, D., Gordon-Finlayson, A. R., Anne, M., Janssen, S. M. J., Lee, K. M., Nielsen, T. K., Tamnes, C. K., Zickfeld, J. H., Rosa, A. D., Vianello, M., Kocsor, F., Kozma, L., Putz, Á., Tressoldi, P., Irrazabal, N., Chatard, A., Lins, S., Pinto, I. R., Lutz, J., Adamkovic, M., Babincak, P., Baník, G., Ropovik, I., Coetzee, V., Dixson, B. J. W., Ribeiro, G., Peters, K., Steffens, N. K., Tan, K. W., Thorstenson, C. A., Fernandez, A. M., Hsu, R. M. C. S., Valentova, J. V., Varella, M. A. C., Corral-Frías, N. S., Frías-Armenta, M., Hatami, J., Monajem, A., Sharifian, M., Frohlich, B., Lin, H., Inzlicht, M., Alaei, R., Rule, N. O., Lamm, C., Pronizius, E., Voracek, M., Olsen, J., Giolla, E. M., Akgoz, A., Özdoğru, A. A., Crawford, M. T., Bennett-Day, B., Koehn, M. A., Okan, C., Gill, T., Miller, J. K., Dunham, Y., Yang, X., Alper, S., Borras-Guevara, M. L., Cai, S. J., Tiantian, D., Danvers, A. F., Feinberg, D. R., Armstrong, M. M., Gilboa-Schechtman, E., McCarthy, R. J., Muñoz-Reyes, J. A., Polo, P., Shiramazu, V. K. M., Yan, W.-J., Carvalho, L., Forscher, P. S., Chartier, C. R., & Coles, N. A. (2021). To which world regions does the valence–dominance model of social perception apply? *Nature Human Behaviour*, *5*, 159–169. https://doi.org/10.1038/s41562-020-01007-2

Jones, D. (2010). A WEIRD view of human nature skews psychologists' studies. *Science*, *328*(5986), 1627–1627. https://doi.org/10.1126/science.328.5986.1627

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., Aveyard, M., Axt, J. R., Babalola, M. T., Bahník, Š., Batra, R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Binan, E. D., Bocian, K., Brandt, M. J., Busching, R., Rédei, A. C., Cai, H., Cambier, F., Cantarero, K., Carmichael, C. L., Ceric, F., Chandler, J., Chang, J.-H., Chatard, A., Chen, E. E., Cheong, W., Cicero, D. C., Coen, S., Coleman, J. A., Collisson, B., Conway, M. A., Corker, K. S., Curran, P. G., Cushman, F., Dagona, Z. K., Dalgar, I., Dalla Rosa, A., Davis, W. E., de Bruijn, M., De Schutter, L., Devos, T., de Vries, M., Doğulu, C., Dozo, N., Dukes, K. N., Dunham, Y., Durrheim, K., Ebersole, C. R., Edlund, J. E., Eller, A., English, A. S., Finck, C., Frankowska, N., Freyre, M.-Á., Friedman, M., Galliani, E. M., Gandi, J. C., Ghoshal, T., Giessner, S. R., Gill, T., Gnambs, T., Gómez, Á., González, R., Graham, J., Grahe, J. E., Grahek, I., Green, E. G. T., Hai, K., Haigh, M., Haines, E. L., Hall, M. P., Heffernan, M. E., Hicks, J. A., Houdek, P., Huntsinger, J. R., Huynh, H. P., Ijzerman, H., Inbar, Y., Innes-Ker, Å. H., Jiménez-Leal, W., John, M.-S., Joy-Gaba, J. A., Kamiloğlu, R. G., Kappes, H. B., Karabati, S., Karick, H., Keller, V. N., Kende, A., Kervyn, N., Knežević, G., Kovacs, C., Krueger, L. E., Kurapov, G., Kurtz, J., Lakens, D., Lazarević, L. B., Levitan, C. A., Lewis, N. A., Lins, S., Lipsey, N. P., Losee, J. E., Maassen, E., Maitner, A. T., Malingumu, W., Mallett, R. K., Marotta, S. A., Međedović, J., Mena-Pacheco, F., Milfont, T. L., Morris, W. L., Murphy, S. C., Myachykov, A., Neave, N., Neijenhuijs, K., Nelson, A. J., Neto, F., Lee Nichols, A., Ocampo, A., O'Donnell, S. L., Oikawa, H., Oikawa, M., Ong, E., Orosz, G., Osowiecka, M., Packard, G., Pérez-Sánchez, R., Petrović, B., Pilati, R., Pinter, B., Podesta, L., Pogge, G., Pollmann, M. M. H., Rutchick, A. M., Saavedra, P., Saeri, A. K., Salomon, E., Schmidt, K., Schönbrodt, F. D., Sekerdej, M. B., Sirlopú, D., Skorinko, J. L. M., Smith, M. A., Smith-Castro, V., Smolders, K. C. H. J., Sobkow, A., Sowden, W., Spachtholz, P., Srivastava, M., Steiner, T. G., Stouten, J., Street, C. N. H., Sundfelt, O. K., Szeto, S., Szumowska, E., Tang, A. C. W., Tanzer, N., Tear, M. J., Theriault, J., Thomae, M., Torres, D., Traczyk, J., Tybur, J. M., Ujhelyi, A., van Aert, R. C. M., van Assen, M. A. L. M., van der Hulst, M., van Lange, P. A. M., van 't Veer, A. E., Vásquez- Echeverría, A., Ann Vaughn, L., Vázquez, A., Vega, L. D., Verniers, C., Verschoor, M., Voermans, I. P. J., Vranka, M. A., Welch, C., Wichman, A. L., Williams, L. A., Wood, M., Woodzicka, J. A., Wronska, M. K., Young, L., Zelenski, J. M., Zhijia, Z., & Nosek, B. A. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, *1*(4), 443–490. https://doi.org/10.1177/2515245918810225

Kruskal, W., & Mosteller, F. (1979a). Representative sampling, I: Non-scientific literature. *International Statistical Review / Revue Internationale de Statistique*, *47*(1), 13–24. https://doi.org/10.2307/1403202

Kruskal, W., & Mosteller, F. (1979b). Representative sampling, II: Scientific literature, excluding statistics. *International Statistical Review / Revue Internationale de Statistique*, *47*(2), 111–127. https://doi.org/10.2307/1402564

Kruskal, W., & Mosteller, F. (1979c). Representative sampling, III: The current statistical literature. *International Statistical Review / Revue Internationale de Statistique*, *47*(3), 245–265. https://doi.org/10.2307/1402647

Kruskal, W., & Mosteller, F. (1980). Representative sampling, IV: The history of the concept in statistics, 1895-1939. *International Statistical Review / Revue Internationale de Statistique*, *48*(2), 169–195. https://doi.org/10.2307/1403151

Kukull, W. A., & Ganguli, M. (2012). Generalizability: The trees, the forest, and the low-hanging fruit. *Neurology*, *78*(23), 1886. https://doi.org/10.1212/WNL.0b013e318258f812

Mook, D. G. (1983). In defense of external invalidity. *American Psychologist*, *38*(4), 379–387. https://doi.org/10.1037/0003-066X.38.4.379

Moshontz, H., Campbell, L., Ebersole, C. R., Ijzerman, H., Urry, H. L., Forscher, P. S., Grahe, J. E., McCarthy, R. J., Musser, E. D., Antfolk, J., Castille, C. M., Evans, T. R., Fiedler, S., Flake, J. K., Forero, D. A., Janssen, S. M. J., Keene, J. R., Protzko, J., Aczel, B., Solas, S. Á., Ansari, D., Awlia, D., Baskin, E., Batres, C., Borras-Guevara, M. L., Brick, C., Chandel, P., Chatard, A., Chopik, W. J., Clarance, D., Coles, N. A., Corker, K. S., Dixson, B. J. W., Dranseika, V., Dunham, Y., Fox, N. W., Gardiner, G., Garrison, S. M., Gill, T., Hahn, A. C., Jaeger, B., Kačmár, P., Kaminski, G., Kanske, P., Kekecs, Z., Kline, M., Koehn, M. A., Kujur, P., Levitan, C. A., Miller, J. K., Okan, C., Olsen, J., Oviedo-Trespalacios, O., Özdoğru, A. A., Pande, B., Parganiha, A., Parveen, N., Pfuhl, G., Pradhan, S., Ropovik, I., Rule, N. O., Saunders, B., Schei, V., Schmidt, K., Singh, M. M., Sirota, M., Steltenpohl, C. N., Stieger, S., Storage, D., Sullivan, G. B., Szabelska, A., Tamnes, C. K., Vadillo, M. A., Valentova, J. V., Vanpaemel, W., Varella, M. A. C., Vergauwe, E., Verschoor, M., Vianello, M., Voracek, M., Williams, G. P., Wilson, J. P., Zickfeld, J. H., Arnal, J. D., Aydin, B., Chen, S.-C., DeBruine, L. M., Fernandez, A. M., Horstmann, K. T., Isager, P. M., Jones, B., Kapucu, A., Lin, H., Mensink, M. C., Navarrete, G., Silan, M. A., & Chartier, C. R. (2018). The Psychological Science Accelerator: Advancing psychology through a distributed collaborative network. *Advances in Methods and Practices in Psychological Science*, *1*(4), 501–515. https://doi.org/10.1177/2515245918797607

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, *162*, 31–38. https://doi.org/10.1016/j.jecp.2017.04.017

Pollet, T. V., & Saxton, T. K. (2019). How diverse are the samples used in the journals 'Evolution & Human Behavior' and 'Evolutionary Psychology'? *Evolutionary Psychological Science*, *5*(3), 357–368. https://doi.org/10.1007/s40806-019-00192-2

R Core Team (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rad, M. S., Martingano, A. J., & Ginges, J. (2018). Toward a psychology of Homo sapiens: Making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences*, *115*(45), 11401–11405. https://doi.org/10.1073/pnas.1721165115

Sarafoglou, A., Haaf, J. M., Ly, A., Gronau, Q. F., Wagenmakers, E.-J., & Marsman, M. (2020). Evaluating Multinomial Order Restrictions with Bridge Sampling. PsyArXiv. https://doi.org/10.31234/osf.io/bux7p

Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, *51*(3), 515–530. https://doi.org/10.1037/0022-3514.51.3.515

Thalmayer, A. G., Toscanelli, C., & Arnett, J. J. (2021). The neglected 95% revisited: Is American psychology becoming less American? *American Psychologist*, *76*(1), 116–129. https://doi.org/10.1037/amp0000622

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., Kesteren, E.-J. v., Doorn, J. v., Smira, M., Epskamp, S., Etz, A., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian Inference for Psychology. Part II: Example Applications with JASP. *Psychonomic Bulletin & Review*, *25*(1), 58–76. https://doi.org/10.3758/s13423-017-1323-7

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E.-J., van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., de Jong, T., van den Bergh, D., Sarafoglou, A., Steingroever, H., Derks, K., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*(1), 58–76. https://doi.org/10.3758/s13423-017-1323-7

Wang, J., Song, Q.-Y., Xu, Y.-P., Jia, B.-B., Lu, C.-L., Chen, X., Dai, Z.-X., Huang, Z.-Y., Z-J, L., Lin, J.-X., Luo, W.-Y., S-N, S., Zhang, Y.-Y., Zang, Y.-F., Zuo, X.-N., & Hu, C.-P. (2021). Interpreting nonsignificant results: A quantitative investigation based on 500 Chinese psychological research. *Advances in Psychological Science*, *29*(3), 381–393. https://doi.org/10.3724/SP.J.1042.2021.00381

Wang, K., Goldenberg, A., Dorison, C. A., Miller, J. K., Uusberg, A., Lerner, J. S., Gross, J. J., Agesin, B. B., Bernardo, M., Campos, O., Eudave, L., Grzech, K., Ozery, D. H., Jackson, E. A., Garcia, E. O. L., Drexler, S. M., Jurković, A. P., Rana, K., Wilson, J. P., Antoniadi, M., Desai, K., Gialitaki, Z., Kushnir, E., Nadif, K., Bravo, O. N., Nauman, R., Oosterlinck, M., Pantazi, M., Pilecka, N., Szabelska, A., van Steenkiste, I. M. M., Filip, K., Bozdoc, A. I., Marcu, G. M., Agadullina, E., Adamkovič, M., Roczniewska, M., Reyna, C., Kassianos, A. P., Westerlund, M., Ahlgren, L., Pöntinen, S., Adetula, G. A., Dursun, P., Arinze, A. I., Arinze, N. C., Ogbonnaya, C. E., Ndukaihe, I. L. G., Dalgar, I., Akkas, H., Macapagal, P. M., Lewis, S., Metin-Orta, I., Foroni, F., Willis, M., Santos, A. C., Mokady, A., Reggev, N., Kurfali, M. A., Vasilev, M. R., Nock, N. L., Parzuchowski, M., Espinoza Barría, M. F., Vranka, M., Kohlová, M. B., Ropovik, I., Harutyunyan, M., Wang, C., Yao, E., Becker, M., Manunta, E., Kaminski, G., Marko, D., Evans, K., Lewis, D. M. G., Findor, A., Landry, A. T., Aruta, J. J. B., Ortiz, M. S., Vally, Z., Pronizius, E., Voracek, M., Lamm, C., Grinberg, M., Li, R., Valentova, J. V., Mioni, G., Cellini, N., Chen, S.-C., Zickfeld, J., Moon, K., Azab, H., Levy, N., Karababa, A., Beaudry, J. L., Boucher, L., Collins, W. M., Todsen, A. L., van Schie, K., Vintr, J., Bavolar, J., Kaliska, L., Križanić, V., Samojlenko, L., Pourafshari, R., Geiger, S. J., Beitner, J., Warmelink, L., Ross, R. M., Stephen, I. D., Hostler, T. J., Azouaghe, S., McCarthy, R., Szala, A., Grano, C., Solorzano, C. S., Anjum, G., Jimenez-Leal, W., Bradford, M., Pérez, L. C., Cruz Vásquez, J. E., Galindo-Caballero, O. J., Vargas-Nieto, J. C., Kácha, O., Arvanitis, A., Xiao, Q., Cárcamo, R., Zorjan, S., Tajchman, Z., Vilares, I., Pavlacic, J. M., Kunst, J. R., Tamnes, C. K., von Bastian, C. C., Atari, M., Sharifian, M., Hricova, M., Kačmár, P., Schrötter, J., Rahal, R.-M., Cohen, N., FatahModarres, S., Zrimsek, M., Zakharov, I., Koehn, M. A., Esteban-Serna, C., Calin-Jageman, R. J., Krafnick, A. J., Štrukelj, E., Isager, P. M., Urban, J., Silva, J. R., Martončik, M., Očovaj, S. B., Šakan, D., Kuzminska, A. O., Djordjevic, J. M., Almeida, I. A. T., Ferreira, A., Lazarevic, L. B., Manley, H., Ricaurte, D. Z., Monteiro, R. P., Etabari, Z., Musser, E., Dunleavy, D., Chou, W., Godbersen, H., Ruiz-Fernández, S., Reeck, C., Batres, C., Kirgizova, K., Muminov, A., Azevedo, F., Alvarez, D. S., Butt, M. M., Lee, J. M., Chen, Z., Verbruggen, F., Ziano, I., Tümer, M., Charyate, A. C. A., Dubrov, D., Tejada Rivera, M. d. C. M. C., Aberson, C., Pálfi, B., Maldonado, M. A., Hubena, B., Sacakli, A., Ceary, C. D., Richard, K. L., Singer, G., Perillo, J. T., Ballantyne, T., Cyrus-Lai, W., Fedotov, M., Du, H., Wielgus, M., Pit, I. L., Hruška, M., Sousa, D., Aczel, B., Szaszi, B., Adamus, S., Barzykowski, K., Micheli, L., Schmidt, N.-D., Zsido, A. N., Paruzel-Czachura, M., Bialek, M., Kowal, M., Sorokowska, A., Misiak, M., Mola, D., Ortiz, M. V., Correa, P. S., Belaus, A., Muchembled, F., Ribeiro, R. R., Arriaga, P., Oliveira, R., Vaughn, L. A., Szwed, P., Kossowska, M., Czarnek, G., Kielińska, J., Antazo, B., Betlehem, R., Stieger, S., Nilsonne, G., Simonovic, N., Taber, J., Gourdon-Kanhukamwe, A., Domurat, A., Ihaya, K., Yamada, Y., Urooj, A., Gill, T., Čadek, M., Bylinina, L., Messerschmidt, J., Kurfalı, M., Adetula, A., Baklanova, E., Albayrak-Aydemir, N., Kappes, H. B., Gjoneska, B., House, T., Jones, M. V., Berkessel, J. B., Chopik, W. J., Çoksan, S., Seehuus, M., Khaoudi, A., Bokkour, A., El Arabi, K. A., Djamai, I., Iyer, A., Parashar, N., Adiguzel, A., Kocalar, H. E., Bundt, C., Norton, J. O., Papadatou-Pastou, M., De la Rosa-Gomez, A., Ankushev, V., Bogatyreva, N., Grigoryev, D., Ivanov, A., Prusova, I., Romanova, M., Sarieva, I., Terskova, M., Hristova, E., Kadreva, V. H., Janak, A., Schei, V., Sverdrup, T. E., Askelund, A. D., Pineda, L. M. S., Krupić, D., Levitan, C. A., Johannes, N., Ouherrou, N., Say, N., Sinkolova, S., Janjić, K., Stojanovska, M., Stojanovska, D., Khosla, M., Thomas, A. G., Kung, F. Y. H., Bijlstra, G., Mosannenzadeh, F., Balci, B. B., Reips, U.-D., Baskin, E., Ishkhanyan, B., Czamanski-Cohen, J., & Dixson, B. J. W. (2021). A multi-country test of brief reappraisal interventions on emotions during the COVID-19 pandemic. *Nature Human Behaviour*, *5*(8), 1089–1110. https://doi.org/10.1038/s41562-021-01173-x

Yarkoni, T. (2020). The generalizability crisis. *Behavioral and Brain Sciences*, 1–37. https://doi.org/10.1017/S0140525X20001685

Zuo, X.-N., Xu, T., & Milham, M. P. (2019). Harnessing reliability for neuroscience research. *Nature Human Behaviour*, *3*, 768–771. https://doi.org/10.1038/s41562-019-0655-x