Assessing diversity and representativeness of big team science in psychology [Stage 1 RR Protocol]

Weibiao Liu [1], Sakshi Ghai [2], Flavio Azevedo[3], Patrick S. Forscher [4], Hu Chuan-Peng [1*]

1. School of Psychology, Nanjing Normal University, Nanjing, China

2. Department of Psychological and Behavioral Science, London School of Economics and Political Science

3. Department of Interdisciplinary Social Science, Utrecht University, the Netherlands

4. Busara Center for Behavioral Economics, Nairobi, Kenya

***Corresponding author:** Hu Chuan-Peng (Email: hu.chuan-peng@nnu.edu.cn, or, hcp4715@hotmail.com)

Abstract

Psychological science has a persistent diversity problem: samples, authors, and journal editors are predominantly from the Global North. Big team science, characterized by collaborative projects involving researchers worldwide, has emerged as a potential solution to these diversity and generalizability concerns. Despite its widespread adoption and claims of increased generalizability, the actual impact of big team science on improving diversity and representation in psychological studies remains unknown. Here, we plan to systematically examine the diversity and representativeness of big team science by comparing big team science to (1) traditional psychological studies and (2) its target population (the world population or a specific population). Specifically, we will compare these two reference datasets to the demographics of participants and authors affiliations from big team science. In addition, we will examine how socioeconomic/cultural factors might contribute to the presence of researchers and participants in big team science and traditional psychological studies in a country/region.

Keywords: Meta-science; Population psychology; Representativeness; Big team science; Generalizability

1 Introduction

Psychological science faces a generalizability crisis (Bauer, 2023; Yarkoni, 2022). One critical reason is a well-known problem: Samples in psychological science are notoriously lacking in diversity and representativeness (Arnett, 2008; Henrich et al., 2010; Liu et al., 2024; Nielsen et al., 2017; Pollet & Saxton, 2019; Rad et al., 2018). For example, Arnett (2008) pointed out that most of the samples published in the six premier APA journals come from western countries, and this pattern has not changed in the past decades (Thalmayer et al., 2021). However, this issue is much broader than the WEIRD problem. While psychological research has historically overrepresented participants from Western, Educated, Industrialized, Rich, and Democratic regions (Henrich et al., 2010), within country diversity has been routinely overlooked (Ghai, 2021; Ghai et al., 2024). Simply incorporating participants from more countries does not necessarily ensure greater representativeness, as there is a greater extent of diversity within regions.

The lack of diverse and representative samples in psychological science leads to several severe consequences for the field. Theories often fail to generalize beyond Western undergraduate students (e.g., Tindle, 2021), and findings relevant to clinical settings (e.g., Chen et al., 2023) and public policy (e.g., Ijzerman et al., 2020) are biased. Moreover, excluding non-Western populations creates a situation where we are unaware of what we do not know about the human mind and behavior (i.e., the 'unknown unknown'; Adetula et al., 2022), thereby limiting the scope of psychological science.

A closely related issue, often overlooked by researchers, is the lack of diversity among authors publishing in psychology journals, which may contribute to the limited diversity and representativeness of participants (Arnett, 2008; Lin & Li, 2022; Medin et al., 2017; Thalmayer et al., 2021). Arnett (2008) found that, on average, over 70% of first authors in six APA journals are from the United States, and this trend has persisted. Lin and Li (2022) reported that 86% of authors in 68 leading psychology journals were primarily from North America and Western Europe. Similarly, IJzerman et al. (2021) noted that 90% of awards from the Society for Personality and Social Psychology were given to researchers affiliated with United States institutions. In the Open Science movement, the majority of researchers and grassroots networks are also from North America and Western Europe (Jin et al., 2023). This lack of author diversity shapes research questions, methodologies, and sampling decisions, often reinforcing biases in study designs and limiting generalizability (Bauer, 2023; Yang et al., 2023). Beyond methodology, it also affects whose perspectives are represented in psychological science, as researchers from underrepresented backgrounds may prioritize different topics, theoretical frameworks, and interpretations that remain overlooked in the field (Arnett, 2008; Azevedo et al., 2022; Corral-Frias et al., 2023; Jeftic et al., 2024; Lin & Li, 2022).

Big team science has emerged as a potential solution to the generalizability issue in psychological science (Tiokhin et al., 2019; Moshontz et al., 2018; Uhlmann et al., 2019). Big team science[1] is loosely defined as projects which involve a relatively large number of collaborators spread across different labs, cultures, and regions, all of whom pool resources to complete larger, more ambitious projects than they could individually (Baumgartner et al., 2023; Forscher et al., 2023). Because researchers from multiple countries participate in local data collection, big team science often involves a large number of participants from dozens of countries/regions, producing stronger impacts than studies conducted by single teams (Nogrady, 2023). Networks for big team science have rapidly expanded in psychology, such as the Psychological Science Accelerator (Moshontz et al., 2018), ManyLabs (Ebersole et al., 2016; Klein et al., 2018), ManyBabies (Byers-Heinlein et al., 2020), EEGManyLabs (Pavlov et al., 2021), EEGManyPipelines (Trübutschek, 2022), COSN (e.g., Wang et al., 2021), FORRT (e.g., Röseler et al., 2024), and ABRIR (https://abrirpsy.org). These big team science networks either enabled projects like the Open Science Collaboration (2015), which provided compelling evidence for the reproducibility problem in psychological science, or collected extensive data to test the effects of emotional regulation, nudging, and the communication of health information (Dorison et al., 2020; Psychological Science Accelerator Self-Determination Theory Collaboration, 2022; Wang et al., 2021), or diversified the topics in psychological research (*e.g.*, Bertolo et al., 2023; Cologna et al., 2025; Tybur et al., 2016 Vlasceanu et al., 2024).

Focusing on the number of countries/regions without looking into the heterogeneity within each region (Ghai, 2021), findings from big team science might create an "illusion of generalizability" (Ghai et al., 2024). Indeed, published big team science often claim diversity of their data and generalizability or globalization of their findings. For example, the abstract of the Many Labs 2 report states that there is "little heterogeneity between Western, educated, industrialized, rich, and democratic (WEIRD) cultures and less WEIRD cultures". Similarly, based on data collected in 61 countries/regions, Ruggeri et al. (2022) claimed that "temporal discounting is a robust, global, and generalizable effect", Buchanan et al (2023) wrote that the Psychological Science Accelerator's COVID-19 rapid-response dataset include "a diverse, global sample". These claims about diversity and generalizability, however, may not be supported by the data. A re-analysis of the data from Many-Lab 2 showed that its samples

[1] It should be noted that big team science exists not only in psychology, but also in other disciplines such as ManyBirds (Miller et al., 2022) and ManyDogs (Espinosa et al., 2023). In this study, we focus only on big team science in the field of psychology.

were mostly from WEIRD people worldwide (Schimmelpfennig et al., 2024). Similarly, Ghai et al (2024) re-analyzed data from Ruggeri et al. (2022) and found that the samples from different countries were not representative of their own country's population, especially in terms of age and education level. A meta-science focused on Chinese psychological samples also found that Chinese samples in big team science are not representative of the Chinese adult population when using the census data as a reference (Yue et al., 2023). As for the author diversity, although big team science includes authors from different countries, the leading authors (first authors and senior authors) are dominated by teams from "WEIRD" countries, and researchers from culturally diverse contexts only help to collect data (Adetula et al., 2022; Coles et al., 2022a; Forscher et al., 2023; Ghai et al., 2024). As researchers tend to investigate the topics that are familiar or of interest to them (Medin et al., 2017), and the topics that of interest to researcher in culturally diverse contexts may systematically differ from those in WEIRD countries (Adetula et al., 2022; Atari et al., 2025), big team science led by researchers from WEIRD countries focuses mainly on topics of interest to researchers from WEIRD countries (Azevedo et al., 2022; Corral-Frias et al., 2023; Jeftic et al., 2024). This bias in research topics is unfair These studies suggested that big team science, with its big team of authors and large sample size, may not have solved the diversity and representativeness problem yet as it promised.

More importantly, we should not expect this issue to be resolved without substantial changes in research investment. As Forscher et al. (2023) note, the common cause of many challenges in psychology is the “insufficient investment of intellectual and nonintellectual resources into the typical psychology study”. Although big team science involves more researchers per project, the investment structure has not changed: big team science is often crowdsourced, with little additional financial support for each participating team (e.g., Coles et al., 2022a). In other words, the scarcity of non-intellectual resources faced by researchers in developing countries/regions remains unchanged (UNESCO, 2021a). Big team science often transfers financial burdens onto “data collectors”, offering authorship in exchange for their contributions, which may exclude many researchers from the Global South who lack established data collection capacity (Korbmacher et al., 2023). For instance, capacity surveys conducted by the Psychological Science Accelerator show that most participating labs are based in the Global North (Kijilian et al., 2022; Paris et al., 2020). Thus, factors such as GDP per capita, research and development (R&D) investment, number of universities, number of psychology researchers, urbanization, and average years of schooling may influence participation in big team science. Additionally, since most big team science recruit collaborators and collect data via the internet—particularly using English-language social media—it is likely that cultural factors, such as English proficiency, cultural proximity to

WEIRD regions, internet access, and the level of economic globalization in a given country/region, also play a significant role. In summary, although big team science has mobilized greater non-intellectual resources, the costs of such investments vary depending on local socioeconomic and cultural factors, which remain unchanged.

However, there is no comprehensive assessment of the diversity and representativeness of samples and author affiliations[2] in big team science. The relationships between socioeconomic and cultural factors and the engagement in big team science across regions are also unknown. To address this gap, we plan to examine both the samples and author affiliations in big team science by comparing them to two reference data sources: samples from traditional psychological studies and the target population for big team science (either the global population or a specific sub-population). The first reference dataset will consist of traditional psychological studies that match big team science in terms of keywords and target population. The second reference dataset will be the world population data from the United Nations Population Division or the World Bank. Specifically, we will focus on the following four questions (the corresponding hypotheses and analytical plan are shown in Hypotheses Table):

Question 1: Are the demographics of samples in big team science more representative than those in traditional psychological studies?

Question 2: To what extent do big team science samples reflect their intended target populations (e.g., global or sub-population demographics)?

Question 3: Are the author affiliations, including leading author affiliations, of big team science more geographically diverse than those in traditional psychological studies?

Question 4: Are the author affiliations, including leading author affiliations, of big team science geographically representative for the global population?

Additionally, we explored several socioeconomic/cultural factors that might be related to the lack of presence for some countries/regions. The results of the current study will provide insights for big team science to move toward a fairer science that contributes to all, regardless of region.

2 Method

[2] Note that author affiliations do full represent the nationality or home country of researchers who are studying abroad or have immigrated, we will discuss this at Stage 2.

2.1 Data sources

The data for big team science will be drawn from all available studies in psychology that involve large teams. Big team science is defined as "a method involving a relatively large number of collaborators who may be dispersed across labs, institutions, disciplines, cultures, and continents" (Forscher et al., 2023). However, there is no operationalization for how large is "a relatively large number". Also, big team science may be used interchangeably with "crowdsourced studies" and sometimes with "big-sample" studies (Baumgartner et al., 2023). Crowdsourced studies refer to large-scale collaborations throughout the research process, from idea generation to peer review (Uhlmann et al., 2019). Big-sample studies refer to traditional psychological studies with relatively large sample sizes.

We distinguish three types of studies based on six key aspects (see Table 1). In most cases, crowdsourced science can be used interchangeably with big team science, except when participating teams are involved only in data analysis, such as NARPS (Botvinik-Nezer et al., 2020), the Crowdsourced Replication Initiative (Breznau et al., 2019; Breznau et al., 2022), and EEGManyPipelines (Trübutschek, 2022). On the other hand, while big team science and big-sample studies differ, they share two key features: a large number of participants and the participants' countries or regions. Big-sample studies may be led and completed by one or two teams with the support of their funding and resources, which distinguishes them from big team science (e.g., Coles et al., 2022a).

Here, we define big team science as studies that meet key criteria based on prior meta-research: crowdsourced data collection, a large author team ($N \geq 25$), multiple institutional affiliations (≥ 5), and a substantial participant pool ($N \geq 2000$)[3]. Typical big team science including Many Labs projects (Ebersole et al., 2016; Ebersole et al., 2020; Klein et al., 2014; Klein et al., 2018; Klein et al., 2022), ManyBabies (Byers-Heinlein et al., 2020; Byers-Heinlein et al., 2021; Steffan et al., 2023), Trust in Science and Science-Related Populism (Cologna et al., 2025), International Collaboration on Social Moral Psychology of COVID-19 project (Azevedo et al., 2023), Climate ManyLabs (Vlasceanu et al., 2024), The International Study of Metanorms (Eriksson et al., 2021), ManySmiles (Coles et al., 2022b), Emerging

[3] The number of authors was based on meta-research by Lin & Lu (2023), who analyzed the number of authors in psychology journals. Our re-analysis found that the number of articles with 20 or more authors is 1.2% and the number of articles with 25 or more authors is 0.6%. We consider 0.6% represent a rare case of psychological studies. Thus, we chose 25 authors as the criteria for big team science in terms of number of authors. The number of author affiliations was based on meta-research by Burghardt et al. (2021), which reported that most articles had 1 to 3 author affiliations, but author affiliations between more than four teams are rare. Therefore, we chose five author affiliations as the threshold for big team science in terms of the number of affiliated institutions. The number of participants was based on meta-research by Kühberger et al. (2014), who analyzed sample sizes in psychology journals. Our re-analysis found that only 5% of studies had sample sizes of 2000 or more participants. We consider this 5% to represent the proportion of big sample psychological studies. Therefore, we set a sample size of $N \geq 2000$ as the criterion for this aspect of big team science.

Adulthood Measured at Multiple Institutions (Grahe et al., 2018), the Human Penguin Project (Hu et al., 2019; IJzerman et al., 2018), and all completed projects from PSA (e.g., Chen et al., 2023, Jones et al., 2021; Wang et al., 2021). These studies are also included in our list of big-team science, even if they did not meet all the criteria in the operational definition above. We also include other cross-cultural studies with large teams in our definition of big team science. However, due to the absence of specific, common keywords for big team science, it is challenging to compile a comprehensive list via keywords search. Therefore, we have used a self-compiled method to gather a list of big team science. A complete list of these studies can be found here: https://docs.google.com/spreadsheets/d/18lF1KohchjKR5nM2utvlnd9A-zc3iKyf4EXLJ_jMPgk/edit?gid=1741341307#gid=1741341307[4]. To create this list, we employed methods such as searching for big team science projects, monitoring posts on Twitter/X or Bluesky, and tracking relevant literature.

Table 1. Aspects of studies that may distinguish three types of "big" sciences.

	Big # of authors	Big # of author countries	Big # of participants	Crowdsourcing data collection	Other crowdsourcing (e.g. data analysis)
Big Team Science	√	√	√	√	-
Crowdsourcing science	√	-	-	√	√
Big-Sample/data science	-	-	√	-	-

The first reference dataset will be extracted from traditional psychological studies that align with big team science published in 2023, based on research topics and target populations. The matching process will proceed as follows: First, we selected nine journals from different fields of psychology, including journals for all fields of psychology (*Nature Human Behaviour*, *Psychological Science*), clinical psychology (*Journal of Psychopathology and Clinical Science*), developmental psychology (*Developmental Psychology*), experimental psychology (*Journal of Experimental Psychology: General*), applied psychology (*Journal of Applied Psychology*), biological psychology (*Journal of Cognitive Neuroscience*), educational psychology (*Journal of Educational Psychology*), and social and personality psychology (*Journal of Personality and Social Psychology*). These nine journals are representative and have relatively high impact factors in their respective fields.

Next, we will use a customized text-matching algorithm to select traditional psychological studies that focus on similar research topics as each big team science study. This approach ensures that our comparisons are based on semantic meaning rather than just

[4] This list was compiled independently for our purpose, the list by Dwayne Lieck and Daniel Lakens https://osf.io/wx4zd/ was used for internal validation and cross-checking.

exact word matches, allowing for more accurate identification of relevant studies. This process involves the following steps:

(1) **Extracting Keywords**: If a big team science study or a traditional psychological study does not contain predefined keywords, two coders will independently generate five keywords per article based on its full text. Any discrepancies will be resolved through discussion.

(2) **Computing Text Similarity**: We will use the BERT language model (via the R package *PsychWordVec*, Bao, 2023) to generate dynamic word vector embeddings for each set of keywords. These embeddings capture semantic relationships between words.

(3) **Matching Process**: The similarity between big team science studies and traditional studies will be calculated using cosine similarity scores. This metric quantifies how similar two sets of keywords are in a multi-dimensional space—higher scores indicate greater conceptual overlap.

(4) **Selecting the Best Matches**: Up to five traditional psychological studies with the highest cosine similarity scores will be selected for each big team science study.

For details see https://osf.io/mh6ta.

After selecting five articles for each big team science, we will screen these based on their target population. We will code the target populations for both the big team science studies and the five selected traditional psychological studies. Specifically, we will read the titles, abstracts, and discussions (or conclusions) of each study to code its target population (see the code manual in the Supplementary Materials for details). To improve accuracy, we will use a large language model (e.g., ChatGPT-4o) as a third coder to assist with coding the target population, with any discrepancies resolved through discussion between the two human coders. The traditional psychological studies will be coded sequentially based on their similarity to the corresponding big team science studies. Coding will stop once two matches are identified. If two matches cannot be found in the five selected traditional psychological studies, we will return to the keywords matching step and select the next five studies with the highest similarity scores.

To avoid duplication, matching across different big team science articles was done alphabetically by the first author's surname. Once a traditional psychological study was matched to a big team science article, it was excluded from further matching.

The second reference dataset, the world population data, will be sourced from several sources. We will primarily use data from the United Nations Population Division, which is based on census data for each country (United Nations Population Division, 2022), and The World Bank (https://data.worldbank.org.cn/). If the data is outdated (prior to 2012), we will seek more recent data from other reliable sources, such as national government agencies for census or statistical data. The sources of all data will be recorded in detail in our R scripts.

2.2 Data extraction

2.2.1 Data extraction for traditional psychological studies

We will extract data from traditional psychological studies. As some papers from traditional psychological studies provide raw data with demographic variables while others do not, we will distinguish between these two types when extracting data.

For studies with available raw data (either publicly accessible or obtainable after emailing the authors) that includes demographic information, we will extract data directly from the raw dataset. For studies without raw data, we will extract data from the full text and supplementary materials. The extraction process will be carried out in three stages: pre-encoding, encoding, and proofreading.

In the pre-coding stage, we will first develop the initial version of the coding manual (see https://osf.io/w32t9) based on previous research (Arnett, 2008; Nielsen et al., 2017; Pollet & Saxton, 2019; Rad et al., 2018; Yue et al., 2023). At least two coders will then code five randomly selected studies, compare their results, resolve any discrepancies, and revise the manual. They will then code another five randomly selected studies to further refine the manual. This process will be repeated until the differences between the coders no longer affect the quantitative indices or subsequent analyses. The final version of the coding manual will be used for coding the remaining articles.

Once the formal coding manual is established, we will re-code the previous articles using the formal coding manual and code all remaining empirical studies in traditional psychological studies. All trained coders will code approximately 200 articles, with each article being coded by at least two coders.

The sample information we will extract includes the article number, journal source, article title, number of studies, study type, sample type, sample size, recruitment method, and reporting of participant information in the abstract. Additionally, we will collect all available sample information, such as sex, age, socioeconomic status, education level, race, occupation, religion, nationality, urban-rural status, and the location where the subjects were recruited. For race/ethnicity classification, based on previous meta-research (Kissel & Friedman, 2023; Singh et al., 2023; Wilson, 2024), most psychological studies use classification standards similar to those of the U.S. census (American Psychological Association, 2020; U.S. Census Bureau, 2021). However, given that U.S.-based racial categories may not be universally applicable, we will employ two classification methods: (a) a region-based classification—Western Europe, Eastern Europe, North America, Oceania, Northeast Asia, Central Asia, South Asia, Southeast Asia, Sub-Saharan Africa, Latin America, and West Asia/North Africa (Khan et al., 2022; Saucier et al., 2014); and (b) U.S.-based categories commonly used in prior research—White, Asian, Black/African American, Hispanic/Latino, multiracial, and other races. In addition, we will read the titles, abstracts, and discussions (or conclusions) to extract statements on constraints to generalizability (Hoekstra & Vazire, 2021). We will also code whether the article compares the sample with the target population.

To ensure the accuracy and objectivity of the coding process, two independent coders will discuss any discrepancies after the initial coding to address errors where the coding deviated from the manual. A third coder will rate the consistency of each article's coding, rating it on a scale from 0 to 1 (0 represents completely different, 1 represents identical). These consistency scores will be used to calculate the internal reliability of the coding using the R package *irr* (Gamer et al., 2019). If inconsistencies persist between the two coders, they will be resolved through collective discussion.

2.2.2 Data extraction for big team science

We will use the same strategy to code big team science studies as we do for empirical studies published in traditional psychological studies. For big team science studies that include demographic information in the shared raw data, we will directly use this data. For studies without demographic information in the shared raw data, or when no raw data is shared, we will extract the relevant data from the article and supplementary materials using the same coding manual applied to traditional psychological studies.

2.3 Data analyses

2.3.1 Data preprocessing and visualization

We will use R 4.4.1 (R Development Core Team 2024) for data preprocessing, visualization, and data analyses.

2.3.2 Hypotheses testing

We will use the Bayes factor to compare whether participants from two data sources differ on certain dimensions (e.g., sex ratio, age distribution, education level [if available], and country distribution). Bayes factors can provide evidence for both the null and alternative hypotheses (Dienes, 2016; Dienes & Mclatchie, 2018; Hu et al., 2018; Wagenmakers et al., 2018). We interpret $BF_{10} \geqslant 10$ or $log(BF_{10}) \geqslant 2.30$ as strong evidence for the alternative hypothesis and as strong evidence against it (Hu et al., 2018; Wagenmakers et al., 2018). However, we avoid strict dichotomization and instead report Bayes Factors as a continuous measure of evidence strength. Specifically, we will use the χ^2 value-based Bayes factors proposed by Johnson (2013), which first applies a χ^2 goodness-of-fit test to compare proportions and then calculates Bayes factors based on the χ^2 value and degrees of freedom using uniformly most powerful Bayesian tests (Nikooienejad & Johnson, 2021). We will first calculate the χ^2 goodness-of-fit test by comparing sample characteristics (e.g., sex ratio, age distribution, education level [if available], and country distribution) between two data sources. Then, we will calculate Bayes factors based on the χ^2 value and degrees of freedom using uniformly most powerful Bayesian tests, which will automatically determine the optimal prior based on the Bayes factor threshold (Nikooienejad & Johnson, 2021). Due to the presence of zero categories in our data, which can invalidate the χ^2 goodness-of-fit test, we will replace all zeros with 1e-09 to facilitate the computation (Brzezińska, 2015; Clogg & Eliason, 1987).

For our first question, *are the demographics of samples in big team science more representative than those in traditional psychological studies?* We will compare the samples from big team science with those from traditional psychological studies, focusing on the same target population. Specifically, we will categorize the target population type for each study based on the coded target populations of both big team science and traditional psychological studies (see Table 2), and then compare studies from different data sources that focus on the same population. For example, we will compare samples from big-team science studies and traditional psychological studies when both types of research target the global adult population. We will provide visual comparisons of participant characteristics, including sex ratio, age distribution, education level (if available), and country distribution. In particular, following Yue et al. (2023), we will create age bins for the age distribution based on developmental stages (0–17, 18–25, 26–40, 41–59, and ≥60) and census data (0–9, 10–19, 20–29, 30–39, 40–49, 50–59, and ≥60)[5]. We will report statistical results using the first set of

[5] For a better visualization, smaller age bins (with an interval of 5 years) are used in the pyramid plot, but not in the statistical analysis.

age bins in the main text and the second set in the supplementary results. Finally, we will use the χ^2 value-based Bayes factor for each sample dimension.

Table 2. Comparison table based on target population

Reference population	Traditional psychological studies	Big team science	The world population
General population	Samples from matched study 1, matched study 2	Samples from big team science 1	Global population
Global adult population (age >=18)	Samples from matched study 3, matched study 4	Samples from big team science 2	Global adult population (age >= 18)
...	...	...	...

For our second question, *to what extent do big team science samples reflect their intended target populations (e.g., global or sub- population demographics)?* We will compare the participants in big team science with its target population (as shown in Table 2). For example, we will compare samples from traditional psychological studies targeting adults with the adult population from world census data. We will present visual comparisons between big team science participants and world population data across sex ratio, age distribution, education level (if available), and country distribution. We will then use the χ^2 value-based Bayes factor for each dimension of the samples, as described above.

For our third question, *are the author affiliations, including leading author affiliations, of big team science more geographically diverse than those in traditional psychological studies?* We will analyze the geographical diversity of all authors, as well as leading authors. Leading authors are defined as the first author(s), corresponding author(s), and last author(s)[6]. Because big team science challenges traditional authorship structures due to its large number of authors and potential variations in contributions (Coles et al., 2023), we will also identify the leading author(s) in the authors' notes (e.g., project administration listed in CREDIT). For both types of authors, we will visualize and compare the national diversity of all or leading authors between empirical research in traditional psychological studies and big team science. We will combine visualization and χ^2 value-based Bayes factors (Nikooienejad & Johnson, 2021) to compare the proportions of high-income, upper-middle-income, lower-middle-

[6] Last author(s) are included only when the author list is not arranged alphabetically by surname.

income, and low-income countries[7] (The World Bank, 2024) represented by leading author affiliations of traditional psychological studies and big team science.

For the fourth question, *are the author affiliations, including leading author affiliations, of big team science geographically representative for the global population?* We will combine visualization and the χ^2 value-based Bayes factor to compare the proportions of high-income, upper-middle-income, lower-middle-income, and low-income countries (The World Bank, 2024) for all or leading author affiliations of big team science and the world population.

2.3.3 Exploratory analyses

To examine the impact of socioeconomic and cultural factors on the representation of countries in psychological research, we will collect country-level data on these factors and explore their relationship with the proportion of participants and authors in traditional psychological studies and big team science. Specifically, for socioeconomic factors, we will assess the following: GDP per capita, R&D investment, number of universities per 100,000 people, number of psychology researchers per 100,000 people, urbanization, average years of schooling, globalization, and internet penetration rate. For cultural factors, we will consider cultural distance, linguistic distance, and language barriers (see Table 3).

Since the dependent variable consists of count data (number of participants per country) that is overdispersed and contains many zeros, we will use a Bayesian zero-inflated negative binomial regression model for analysis (Douma & Weedon, 2019), implemented through the R package brms (Bürkner, 2017). Our analysis addresses four primary research questions: (1) How do socioeconomic and cultural factors influence the likelihood of a country being included in the sampling process? and (2) How do those factors affect the sample size in countries where sampling occurred? and (3) How do these factors influence the likelihood of researcher(s) from a country's research institutions being included? and (4) How do those factors influence the number of resarcher(s) included in the country?

For the first and third question, we will assess the effects by examining whether the 95% highest density interval (HDI) of the regression coefficients in the zero component of the

[7] We will use the World Bank's classification criteria based on national income levels: countries with a GNI per capita of $1,145 or less in 2023 are classified as low-income countries; countries with a GNI per capita between $1,146 and $4,515 are classified as lower-middle-income countries; countries with a GNI per capita between $4,516 and $14,005 are classified as upper-middle-income countries; and countries with a GNI per capita above $14,005 are classified as high-income countries (The World Bank, 2024). Additionally, because a small proportion of the world's countries and regions are not included in these classifications, we will present results excluding unclassified countries and regions in the main text, and results including these countries and regions (with a new classification, "other countries") in the supplementary material.

posterior distribution excludes zero (Kruschke, 2018). Effects will be interpreted using the posterior median. For the second and fourth question, we will interpret and infer the effects based on the posterior median and 95% HDI of the regression coefficients in the non-zero component of the model. The complete Bayesian workflow (Gelman et al., 2020), using GDP per capita as an example, is provided in the supplementary material. This workflow includes prior predictive checks to select an appropriate prior. All other socioeconomic and cultural factors will be analyzed using the same approach.

Table 3. Country level socioeconomic and cultural variables and corresponding data source

Country-level variable	**Data**
GDP per capita	GDP per capita (from The World Bank, 2022a)
R&D investment	Research expenditure as a share of GDP (from UNESCO, 2021b)
Globalization level	Globalization index (from KOF Swiss Economic Institute, 2022)
Number of universities per capita	Number of universities per 100,000 people (from Webometrics Ranking of World Universities, 2023)
Number of psychology researchers per capita	Number of psychology researchers per 100,000 people (from International Union of Psychological consists, 2023)
Urbanization level	Proportion of urban population (from The World Bank, 2022b)
Average years of schooling	Average years of formal education for individuals aged 15-64 (from Our World in Data, 2023)
Internet penetration rate	Proportion of population using the internet (from The World Bank, 2022c)
Cultural distance from the United States	Cultural distance from the United States (from Muthkrishna et al., 2020)
Linguistic distance from the United States	Linguistic distance from the United States (from Melitz & Toubal, 2014)
English proficiency	English proficiency rank (from Education First, 2022)

3 Open Science Practices

We will provide all details of data extraction, data exclusion (if applicable), and data analysis decisions in the manuscript or supplementary materials. All data and code will be made publicly available on the Open Science Framework (https://osf.io/c7hta/) and GitHub (https://github.com/Chuan-Peng-Lab/BTS_Sample_Stage1_RR).

4 Results

4.1 Overview of participants

For XXX papers traditional psychological studies, there are XXX participants from XXX countries (Fig 1 A), with XXX authors from XXX institutes. There are N = XXX in XXX big team science, from XXX countries/regions (Fig 1 B), with XXX authors from XXX institutes.

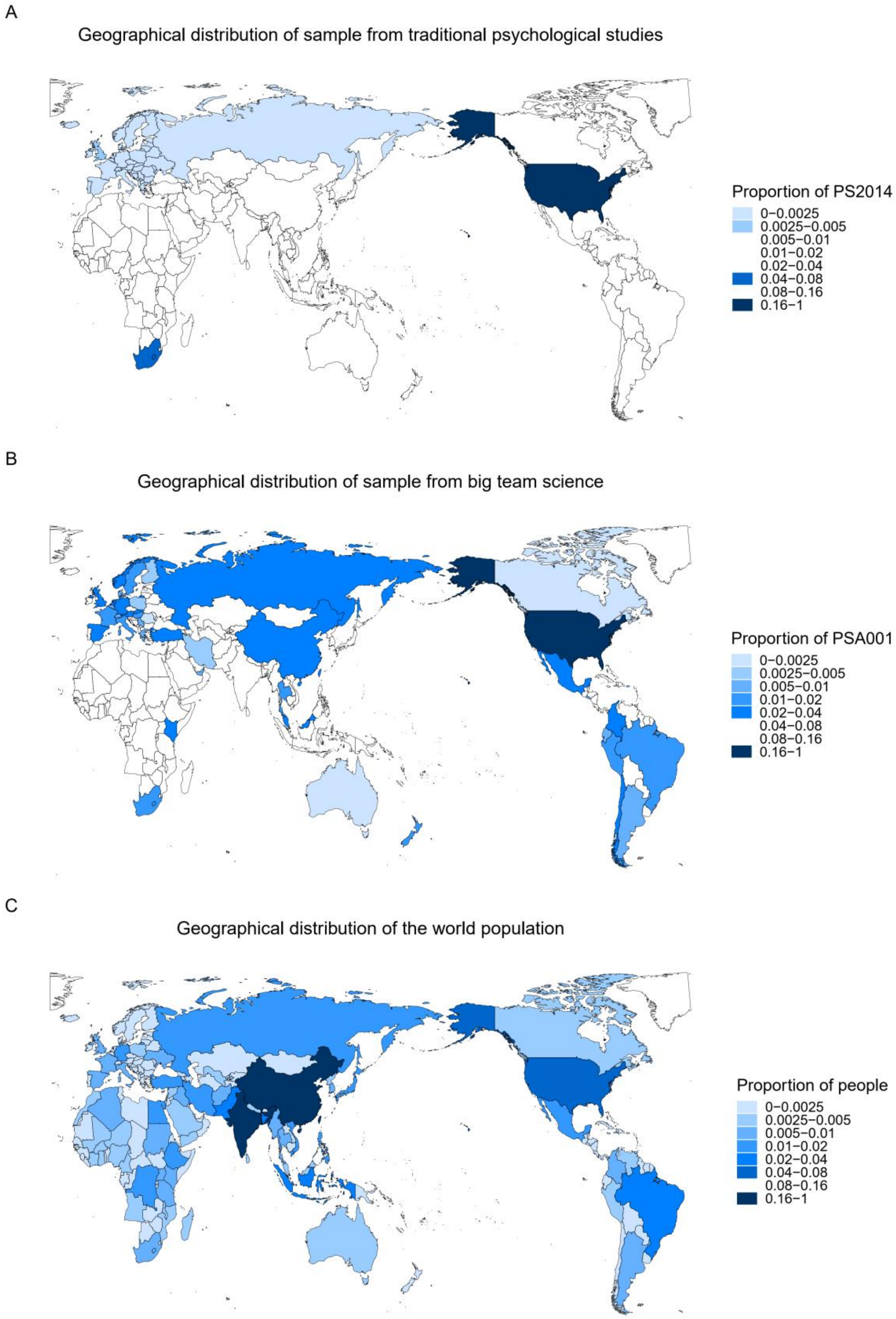

Figure 1. Country distribution of samples from traditional psychological studies (A), big team science (B), and the world population data (C). [*Note panel A will be replaced with data from*

traditional psychological studies after collecting data; panel B will be replaced with data from all big team science after collecting data.]

4.2 Sample diversity and representativeness

4.2.1 Comparing with big team science and traditional psychological studies

4.2.1 Comparing with big team science and traditional psychological studies

[*We will compare the sample characteristics between big team science and traditional psychological studies in the following dimension: sex ratio, age distribution, education level, and country distribution. We will also explore the potential differences in racial/ethnicity and SES if possible. Below are the preliminary results, we used data from Rad et al., (2018) for illustrating traditional psychological studies and PSA001*[8] *(Jones et al., 2021) for illustrating big team science.*

More specifically, for data from Rad et al., (2018), we used data from 35 studies reported in 21 papers in the supplementary material. Data were coded by Yue et al. (2023) and reported in their replies to reviewers, including sample size, sex ratio, mean age, and standard deviation of age. We estimated the number of participants in each age bin based on mean and standard deviation of age (as in Yue et al., 2023) from each paper (In addition to a paper with original data on demographic variables).

First, we compare the sample characteristics from Rad et al (2018) and big team science in sex ratio and age distribution (Figure 2).

Second, we compare the sample characteristics from Rad et al (2018) and big team science in country distribution (Figure 1A & 1B) and population distribution (Figure 3A).]

[8] To avoid the double dipping issue, **we will exclude the data of PSA001(Jones et al., 2021) in stage 2**, but will additionally report the data that include these two datasets in the supplementary materials.

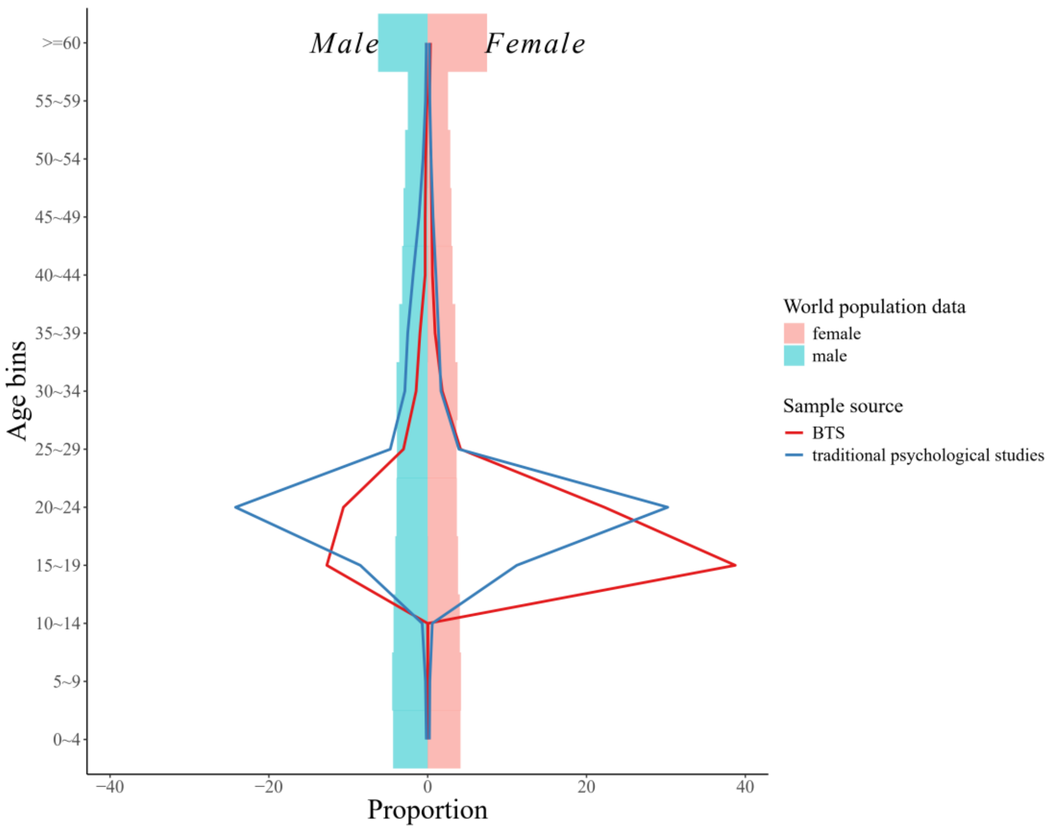

Figure 2. Preliminary results of sample' sex ratio and age distributions from traditional psychological studies (using Rad et al 2018's data), and the big team science (using PSA001's data), and the world population data. The Y-axis is the age bins, and the X-axis is the proportion. The blue lines represent the sample of traditional psychological studies (using Rad et al 2018's data); the red lines represent the big team science samples (using PSA001's data), and the pyramid chart represents the world population data. The left side of the picture is the male and the right is the female.

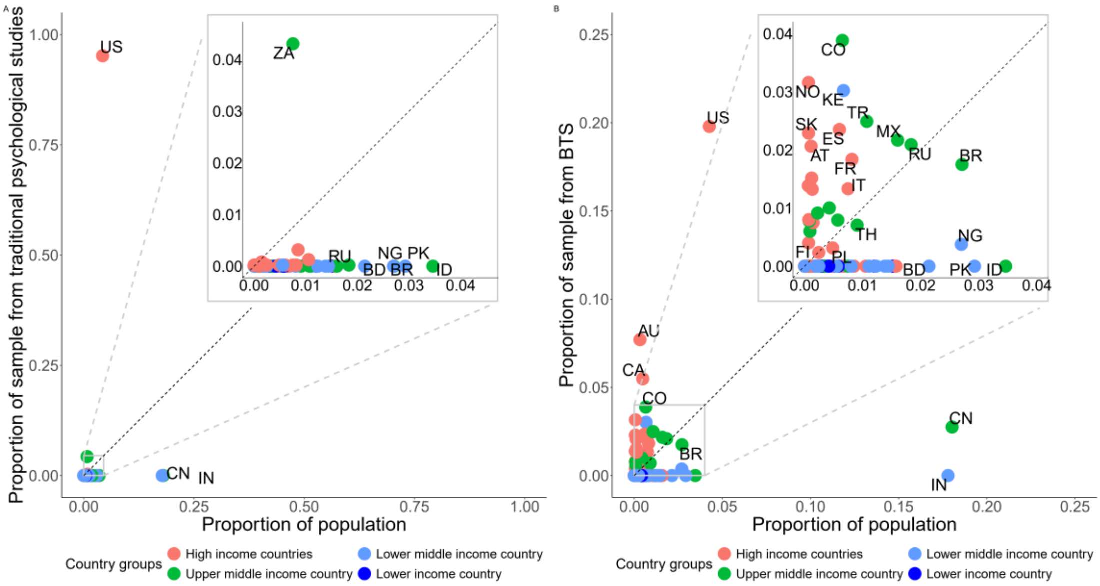

Figure 3. Preliminary results of sample' country distributions from the traditional psychological studies (using Rad et al 2018's data), and the big team science (using PSA001's data), and the world population data. The x-axis represents the proportion of each country's population in the world's population, and the y-axis represents the proportion of each country's sample. (A) Comparing the samples' country distributions from the traditional psychological studies (using Rad et al 2018's data) to the world population data (year 2022). (B) Comparing the samples' country distributions from the big team science (using PSA001's data) to the world population data (year 2022).

4.2.2 Comparing with big team science and the world population data

[*We will compare the sample characteristics from big team science to the world population in the following dimension: sex ratio, age distribution, education level, and country distribution. We will also explore the potential differences in racial/ethnicity and SES if data is available. Below are the preliminary results, we used data from PSA001 for big team science and World Population Prospects for the world population data.*

First, we compare the characteristics of big team science samples and the world population data in sex ratio and age distribution (Figure 2).

Second, we compare the characteristics of big team science samples and the world population data in country distribution (Figure 1B & 1C)/population distribution (Figure 3B) by illustrating the proportion of each country's participants in big team science and the corresponding proportion of each country's population in the world population.]

4.3 Author diversity and representativeness

[*We will compare the country distribution of all or leading author affiliations in big team science, traditional psychological studies, and the world population. Below are the preliminary results, we used non-big-team science articles published in Psychological Science in 2014 data from Rad et al., (2018) for illustrating traditional psychological studies and PSA001 for illustrating big team science.*]

4.2.1 Comparing with all or leading author affiliations in big team science and traditional psychological studies

First, we will compare the country distribution of all author affiliations from traditional psychological studies (Figure 4A) and big team science (Figure 4B), and compare the country distribution of leading author affiliations from traditional psychological studies (Figure 5A) and big team science (Figure 5B).

Second, we will use the χ^2 value-based Bayes factor to compare with all author affiliations from traditional psychological studies and big team science. The results indicate that there is strong evidence that all author affiliations are greater in big team science than in traditional psychological studies ($\log(BF_{10}) = 9.61$).

Third, we will use the χ^2 value-based Bayes factor to compare with leading author affiliations from traditional psychological studies and big team science. The results indicate that there is strong evidence that leading author affiliations are greater in big team science than in traditional psychological studies ($\log(BF_{10}) = 7.23$).

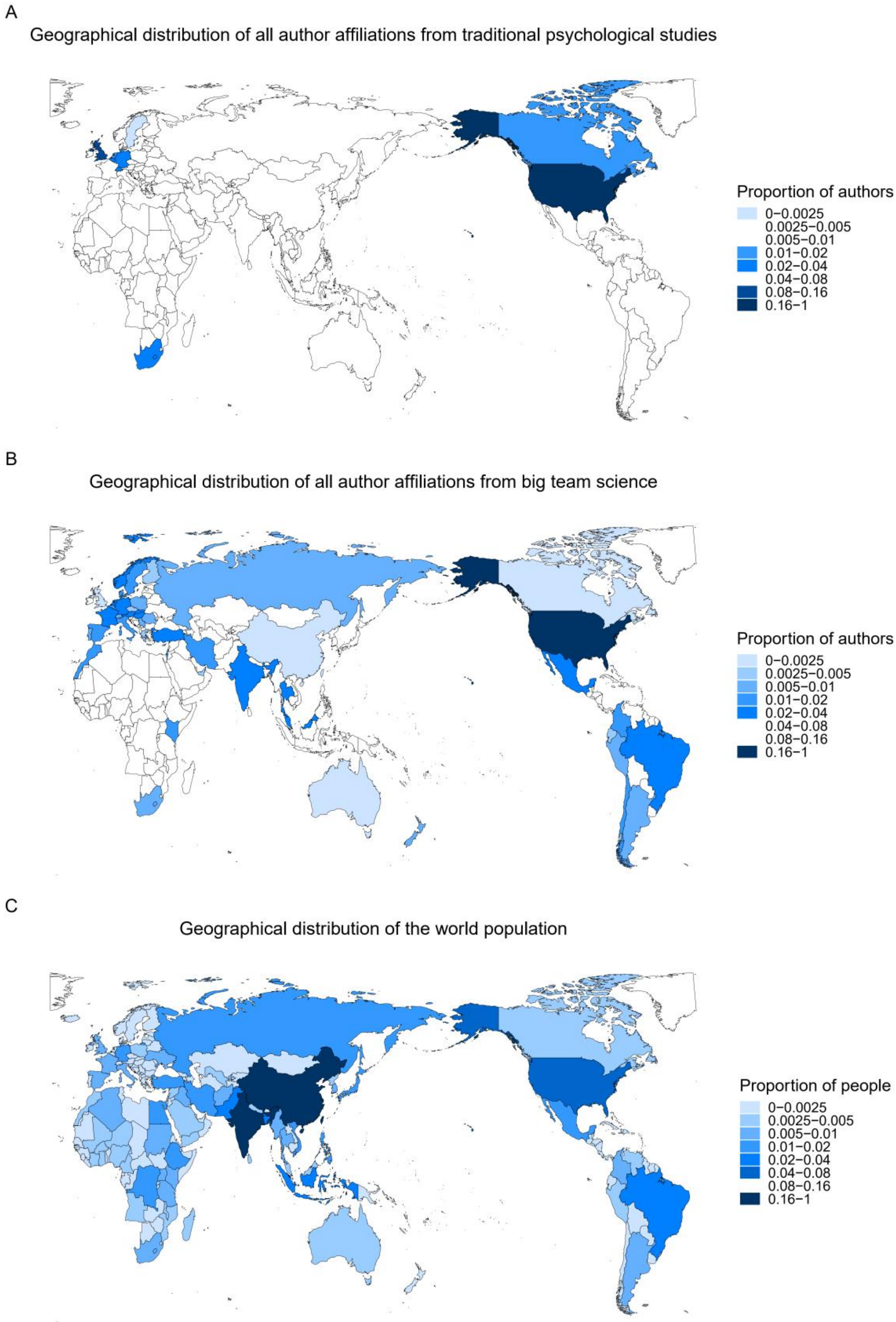

Figure 4. Country distribution of all author affiliations from traditional psychological studies (A), all author affiliations from big team science (B), and the world population data (C). [*Note*

panel A will be replaced with data from traditional psychological studies after collecting data; panel B will be replaced with data from all big team science after collecting data.]

4.2.2 Comparing all or leading author affiliations of big team science with the world population data

[*First, we will compare the country distribution for both all author affiliations from big team science (Figure 4B) and the world population data (Figure 4C), and compare the country distribution for both leading author affiliations from big team science (Figure 5B) and the world population data (Figure 5C).*

Second, we will use the χ^2 value-based Bayes factor to compare all author affiliations of big team science with the world population. The results indicate that there is strong evidence that the national diversity of the world population is greater in big team science than the diversity of all author affiliations in big team science ($\log(BF_{10}) = 33.40$).

Third, we will use the χ^2 value-based Bayes factor to compare leading author affiliations of big team science with the world population. The results indicate that there is strong evidence that the national diversity of the world population is greater than the national diversity of leading author affiliations in big team science ($\log(BF_{10}) = 41.84$).

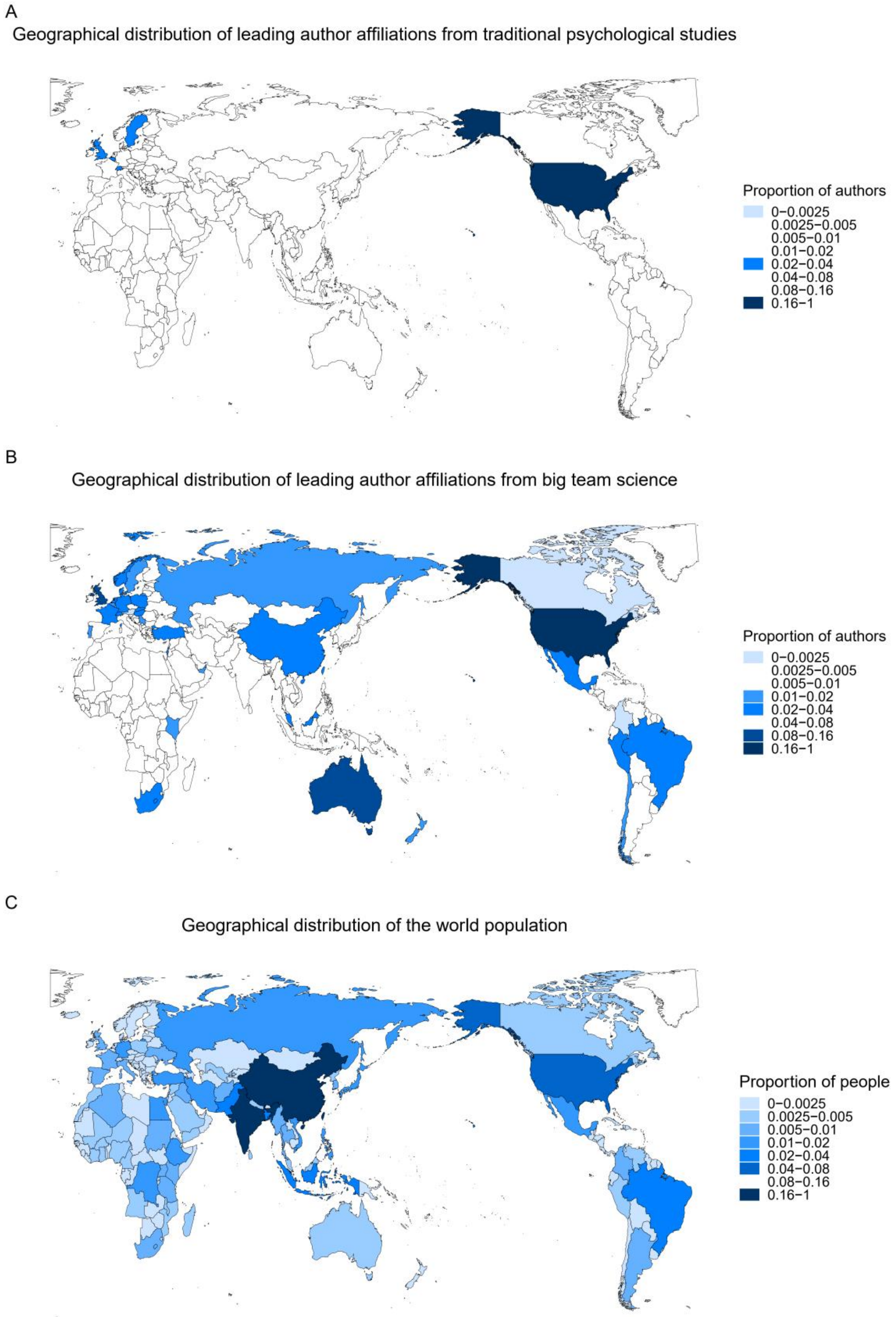

Figure 5. Country distribution of leading author affiliations from traditional psychological studies (A), leading author affiliations from big team science (B), and the world population

data (C). [*Note panel A will be replaced with data from traditional psychological studies after collecting data; panel B will be replaced with data from all big team science after collecting data.*]

4.4 Exploratory Analysis

[*We will report the results of exploratory analysis, including the relationship between the proportion of participants/authors and GDP per capita, R&D investment, number of universities per 100,000 people, number of psychology researchers per 100,000 people, Average years of schooling, Urbanization, Globalization, Internet penetration rate, cultural distance from the United States, linguistic distance from the United States, and English Proficiency of each country/region. Taking GDP per capita as an example, the results (see Figure 6 and Table 4) show that GDP per capita negatively predicted whether a country was included in the big team science sample (median = -0.27, 95% HDI = [-0.49, -0.03]). This suggests that for every $1000 increase in GDP per capita, the proportion of countries not included in the sample decreased by 0.27%. However, GDP per capita had no effect on the sample size of countries included in the big team science (median = 0.14, 95% HDI = [-0.02, 0.30]).*]

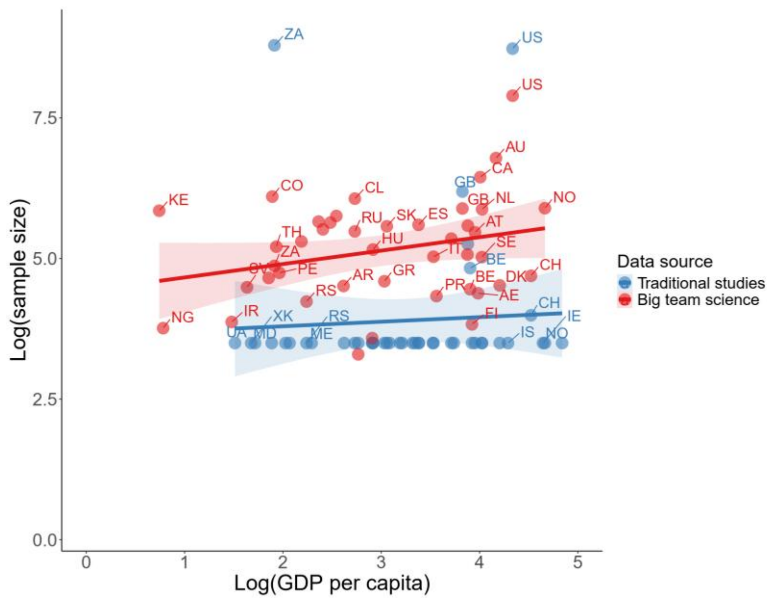

Figure 6. Country-level relationship between proportion of sample and socioeconomic/cultural factors (GDP per capital as an example) for traditional psychological studies (blue, using data from Rad et al., (2018) as an example) and big team science (red, using PSA 001 as an example).

Table 4. Regression analysis between country level socioeconomic/cultural factors (GDP per capital as an example) and sample sizes in traditional psychological studies and big team science.

	Big team science		Traditional psychological studies	
	Median	*95% HDI*	*Median*	*95% HDI*
Intercept (GDP per capita)	*5.15*	[4.57, 5.75]		
GDP per capita	0.14	[-0.02, 0.30]		
Intercept_zi (GDP per capita)	1.75	[1.27, 2.24]		
Zi_GDP per capita	-0.27	[-0.42, -0.11]		

Note. HDI, highest density interval.

5 Discussion

[Recap of the results]

[Limitations of the current study]

Hypotheses Table

Question	Hypothesis	Sampling plan	Analysis Plan	Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis	Interpretation given different outcomes	Theory that could be shown wrong by the outcomes
1. Are the demographics of samples in big team science more representative than those in traditional psychological studies?	$H1a_1$: The BTS have a more equal sex ratio than the traditional psychological studies. $H1a_0$: The BTS and the traditional psychological studies have the same sex ratio. $H1b_1$: The BTS has a more diverse age distribution than the traditional psychological studies. $H1b_0$: The BTS and the traditional psychological studies have the same age distribution. $H1c_1$: The BTS has a more diverse education level than the traditional psychological studies. $H1c_0$: The BTS and the traditional psychological studies have the same education level diversity. $H1d_1$: The BTS has a more diverse country distribution than traditional psychological studies. $H1d_0$: The BTS and the traditional psychological studies have the same country distribution.	Demographic variable data will be collected from traditional psychological studies and from all available big-team science (BTS), including preprints.	For H1a: We will use the χ^2 value-based Bayes factors to test the hypothesis by comparing the sex ratio of samples of the traditional psychological studies and the BTS. For H1b: We will use the χ^2 value-based Bayes factors to test the hypothesis by comparing age distribution of samples of the traditional psychological studies and the BTS. For H1c: We will use the χ^2 value-based Bayes factors to test the hypothesis by comparing the education level of samples of the traditional psychological studies and the BTS. For H1d: We will use the χ^2 value-based Bayes factors to test the hypothesis by comparing the country distribution of samples of the traditional psychological studies and the BTS.		H1a is supported if, compared to the traditional psychological studies, the BTS have a more equal sex ratio (visual inspection). H1b is supported if, compared to the traditional psychological studies, the BTS have a more diverse age distribution (visual inspection). H1c is supported if, compared to the traditional psychological studies, the BTS have a more diverse education level (visual inspection). H1d is supported if, compared to the traditional psychological studies, the BTS have a more diverse country distribution (visual inspection). If BF_{10} > = 10, we infer there is relatively strong evidence that this hypothesis is supported.	This is an observational study, we have no specific theory to falsify.

			We will also visualize the generalizability reports of the two sets of data.			
2. To what extent do big team science samples reflect their intended target populations (e.g., global or sub- population demographics)?	H2a$_1$: The world population data has a more equal sex ratio than the BTS. H2a$_0$: The BTS and the world population data have the same sex ratio. H2b$_1$: The world population data has a more diverse age distribution compared to the BTS. H2b$_0$: Age distribution of the BTS and the world population data are the same. H2c$_1$: The world population data has a more diverse education level than the BTS. H2c$_0$: Education level of both the BTS and the world population data are the same. H2d$_1$: The world population data has a more diverse country distribution than the BTS. H2d$_0$: The country distribution of both the BTS and the world population data are the same.	Demographic variable data will be collected from all available BTS, including preprints.	For H2a: We will use the χ^2 value-based Bayes factors to test the hypothesis by comparing the sex ratio of the BTS sample and the world population data. For H2b: We will combine use the χ^2 value-based Bayes factors to test the hypothesis by comparing the age distribution of the BTS sample and the world population data. For H2c: We will use the χ^2 value-based Bayes factors to test the hypothesis by comparing the education level of the BTS sample and the world population data. For H2d: We will use the χ^2 value-based Bayes factors to test the hypothesis by comparing the country distribution of the BTS sample and the world population data.		H2a is supported if, compared to the BTS, the world population data has a more equal sex ratio (visual inspection). H2b is supported if, compared to the BTS, the world population data has a more diverse age distribution (visual inspection). H2c is supported if, compared to the BTS, the world population data has a more diverse education level (visual inspection). H2d is supported if, compared to the BTS, the world population data has a more diverse country distribution (visual inspection). If $BF_{10} > = 10$, we infer there is relatively strong evidence that this hypothesis is supported.	N/A
3. Are the authors affiliations, including leading authors affiliations, of big team science more geographically diverse than those in	H3a$_1$: All authors affiliations in traditional psychological studies and BTS have different proportions from high-income, upper-middle-income, lower-middle-income, and low-income countries. H3a$_0$: All authors affiliations in traditional psychological studies and BTS have the same proportion from high-income, upper-middle-income, lower-middle-income, and low-income countries. H3b$_1$: Leading authors affiliations in traditional psychological studies and BTS have different proportions	Collect authors affiliations' national data from traditional psychological studies; collect authors affiliations' national data from all	For H3a: We will use the χ^2 value-based Bayes factors to test the hypothesis by comparing the proportion of upper-middle-income country/lower-middle-income country/lower-income country to which traditional psychological studies and BTS all author affiliationss belong. For H3b: We will use the χ^2 value-based Bayes factors to test the hypothesis by comparing the proportion of upper-middle-		H3a is supported if, leading author affiliations in traditional psychological studies and BTS have different proportions from high-income, upper-middle-income, lower-middle-income, and low-income countries. (visual inspection). H3b is supported if, leading author affiliations in traditional psychological studies and BTS have different proportions from high-income, upper-middle-income, lower-middle-income, and low-income countries. (visual inspection).	N/A

traditional psychological studies?	from high-income, upper-middle-income, lower-middle-income, and low-income countries. H3b$_0$: Leading author affiliations in traditional psychological studies and BTS have the same proportion from high-income, upper-middle-income, lower-middle-income, and low-income countries.	published BTS, including preprints.	income country/lower-middle-income country/lower-income country to which traditional psychological studies and BTS leading author affiliations belong.		If BF_{10} > = 10, we infer there is relatively strong evidence that this hypothesis is supported.	
4. Are the author affiliations, including leading author affiliations, of big team science geographically representative for the global population?	H4a$_1$: The proportion of all author affiliationss in BTS from high-income, upper-middle-income, lower-middle-income, and low-income countries is different from that of the world population. H4a$_0$: The proportion of all author affiliationss in BTS from high-income, upper-middle-income, lower-middle-income, and low-income countries is the same as that of the world population. H4b$_1$: The proportion of leading author affiliations in BTS from high-income, upper-middle-income, lower-middle-income, and low-income countries is different from that of the world population. H4b$_0$: The proportion of leading author affiliations in BTS from high-income, upper-middle-income, lower-middle-income, and low-income countries is the same as that of the world population.	Collect author affiliations' national data from all published BTS, including preprints.	For H4a: We will use the χ^2 value-based Bayes factors to test the hypothesis by comparing the proportion of high-income, upper-middle-income, lower-middle-income, and low-income countries to which BTS' all author affiliationss and the world population belong. For H4b: We will use the χ^2 value-based Bayes factors to test the hypothesis by comparing the proportion of high-income, upper-middle-income, lower-middle-income, and low-income countries to which BTS' leading author affiliations and the world population belong.		H4a is supported if, the proportion of all author affiliationss in BTS from high-income, upper-middle-income, lower-middle-income, and low-income countries is different from that of the world population. (visual inspection). H4b is supported if, the proportion of leading author affiliations in BTS from high-income, upper-middle-income, lower-middle-income, and low-income countries is different from that of the world population. (visual inspection). If BF_{10} > = 10, we infer there is relatively strong evidence that this hypothesis is supported.	N/A

Note. BTS, big team science.

Acknowledgements

P.S.F. is supported by the Transformation Fund grant from the Cultural Evolution Society. H.C.-P. is supported by the National Natural Science Foundation of China (No. 32471097). The funders have/had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

Weibiao Liu: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Visualization, Writing-Original Draft, Writing-Review & Editing; **Sakshi Ghai**: Conceptualization, Investigation, Writing-Original Draft, Writing-Review & Editing; **Flavio Azevedo**: Investigation, Writing-Review & Editing; **Patrick S. Forscher**: Conceptualization, Investigation, Writing-Original Draft, Writing-Review & Editing; and **Hu Chuan-Peng**: Conceptualization, Formal Analysis, Investigation, Methodology, Project Administration, Supervision, Validation, Resources, Visualization, Writing-Original draft, Writing-Review & Editing.

Competing interests

The authors declare no competing interests.

Reference

Adetula, A., Forscher, P. S., Basnight-Brown, D., Azouaghe, S., & IJzerman, H. (2022). Psychology should generalize from—not just to—Africa. *Nature Reviews Psychology, 1*(7), 370-371. https://doi.org/10.1038/s44159-022-00070-y

American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). American Psychological Association.

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

Atari, M., Kroupin, I., Morhayim, L., Blasi, D. E., Schulz, J., & Henrich, J. (2025). *WEIRD Questions: Diversifying Conceptual Sampling*. OSF. https://doi.org/10.31234/osf.io/vxcrj_v1

Azevedo, F., F., Middleton, S., Phan, J. M., Kapp, S., Gourdon-Kanhukamwe, A., Iley, B., Elsherif, M., & Shaw, J. J. (2022). Navigating academia as neurodivergent researchers: promoting neurodiversity within Open Scholarship. *APS Observer*, *35*.

Azevedo, F., Tomislav Pavlović, Rêgo, G. G., F Ceren Ay, Biljana Gjoneska, Etienne, T. W., Ross, R. M., Philipp Schönegger, Riaño-Moreno, J. C., Aleksandra Cichocka, Capraro, V., Luca Cian, Chiara Longoni, Ho Fai Chan, Joseph, J., Hallgeir Sjåstad, Nezlek, J. B., Alfano, M., Gelfand, M. J., & Birtel, M. D. (2023). Social and moral psychology of COVID-19 across 69 countries. *Scientific Data*, *10*(1). https://doi.org/10.1038/s41597-023-02080-8

Bao, H.-W.-S. (2023). *PsychWordVec: Word embedding research framework for psychological science* (Version 2023.9) [Computer software]. https://CRAN.R-project.org/package=PsychWordVec

Bauer, P. J. (2023). Generalizations: The grail and the gremlins. *Journal of Applied Research in Memory and Cognition*, *12*(2), 159–175. https://doi.org/10.1037/mac0000106

Baumgartner, H. A., Alessandroni, N., Byers-Heinlein, K., Frank, M. C., Hamlin, J. K., Soderstrom, M., Voelkel, J. G., Willer, R., Yuen, F., & Coles, N. A. (2023). How to build up big team science: a practical guide for large-scale collaborations. *Royal Society Open Science*, *10*(6). https://doi.org/10.1098/rsos.230235

Bertolo, M., Müllensiefen, D., Peretz, I., Woolley, S. C., Sakata, J., & Mehr, S. (2023). Human music perception ability is not a sexually dimorphic trait. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *45*(45).

Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J. A., Adcock, R. A., Avesani, P., Baczkowski, B. M., Bajracharya, A., Bakst, L., Ball, S., Barilari, M., Bault, N., Beaton, D., Beitner, J., . . . Kong, X. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature, 582*(7810), 84–88. https://doi.org/10.1038/s41586-020-2314-9

Breznau, N., Rinke, E. M., Wuttke, A., Adem, M., Adriaans, J., Alvarez-Benjumea, A., Andersen, H. K., Auer, D., Azevedo, F., Bahnsen, O., Balzer, D., Bauer, G., Bauer, P., Baumann, M., Baute, S., Benoit, V., Berning, C., Bernauer, J., Berthold, A., . . . Nguyen, H. H. V. (2019). The Crowdsourced Replication Initiative: Investigating Immigration and Social Policy Preferences. Executive Report. *SocArXiv*. https://doi.org/10.31235/osf.io/6j9qb

Breznau, N., Rinke, E. M., Wuttke, A., Nguyen, H. H. V., Adem, M., Adriaans, J., Alvarez-Benjumea, A., Andersen, H. K., Auer, D., Azevedo, F., Bahnsen, O., Balzer, D., Bauer, G., Bauer, P. C., Baumann, M., Baute, S., Benoit, V., Bernauer, J., Berning, C., . . . Meyer, D. (2022). Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. *Proceedings of the National Academy of Sciences, 119*(44). https://doi.org/10.1073/pnas.2203150119

Brzezińska, J. (2015). The Problem of Zero Cells in the Analysis of Contingency Tables. *Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie, 941*(05), 49-61. https://doi.org/10.15678/ZNUEK.2015.0941.0504

Buchanan, E. M., Lewis, S. C., Paris, B., Forscher, P. S., Pavlacic, J. M., Beshears, J. E., Shira Meir Drexler, Amélie Gourdon-Kanhukamwe, Mallik, P. R., Alejandro, M., Miller, J. K., IJzerman, H., Moshontz, H., Beaudry, J. L., Suchow, J. W., Chartier, C. R., Coles, N. A., MohammadHasan Sharifian, Anna Louise Todsen, & Levitan, C. A. (2023). The Psychological Science Accelerator's COVID-19 rapid-response dataset. *Scientific Data, 10*(1). https://doi.org/10.1038/s41597-022-01811-7

Burghardt, K. A., He, Z., Percus, A. G., & Lerman, K. (2021). The emergence of heterogeneous scaling in research institutions. *Communications Physics, 4*(1), 189. https://doi.org/10.1038/s42005-021-00693-2

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software, 80*(1). https://doi.org/10.18637/jss.v080.i01

Byers-Heinlein, K., Bergmann, C., Davies, C., Frank, M. C., Hamlin, J. K., Kline, M., Kominsky, J. F., Kosie, J. E., Lew-Williams, C., Liu, L., Mastroberardino, M., Singh, L., Waddell, C. P. G., Zettersten, M., & Soderstrom, M. (2020). Building a collaborative psychological science: Lessons learned from ManyBabies 1. *Canadian Psychology/Psychologie Canadienne, 61*(4), 349. https://doi.org/10.1037/cap0000216

Byers-Heinlein, K., Tsui, A. S. M., Bergmann, C., Black, A. K., Brown, A., Carbajal, M. J., Durrant, S., Fennell, C. T., Fiévet, A.-C., Frank, M. C., Gampe, A., Gervain, J., Gonzalez-Gomez, N., Hamlin, J. K., Havron, N., Hernik, M., Kerr, S., Killam, H., Klassen, K., & Kosie, J. E. (2021). A Multilab Study of Bilingual Infants: Exploring the Preference for Infant-Directed Speech. *Advances in Methods and Practices in Psychological Science*, *4*(1), 251524592097462. https://doi.org/10.1177/2515245920974622

Chen, Z., Hu, B., Liu, X., Becker, B., Eickhoff, S. B., Miao, K., Gu, X., Tang, Y., Dai, X., Bhattarai, R., Leonov, A., Xiao, Z., Feng, Z., Chen, J., & Hu, C.-P. (2023). Sampling inequalities affect generalization of neuroimaging-based diagnostic classifiers in psychiatry. *BMC Medicine, 21*(1). https://doi.org/10.1186/s12916-023-02941-4

Chen, S., Buchanan, E. M., Kekecs, Z., Miller, J. K., Szabelska, A., Aczel, B., Bernabeu, P., Forscher, P., Szuts, A., Vally, Z., Al-Hoorie, A. H., Helmy, M., da Silva, C. S. A., da Silva, L. O., de Moraes, Y. L., Hsu, R. M. C. S., Mafra, A. L., Dixson, B., Peters, K., … Zickfeld, J. (2023, May 31). Investigating object orientation effects across 18 languages. *PsyArXiv*. https://doi.org/10.31219/osf.io/2qf6w

Clogg, C. C., & Eliason, S. R. (1987). Some common problems in log-linear analysis. *Sociological Methods & Research*, *16*(1), 8-44. https://doi.org/10.1177/0049124187016001002

Coles, N. A., Hamlin, J. K., Sullivan, L. L., Parker, T. H., & Altschul, D. (2022a). Build up big-team science. *Nature, 601*(7894), 505–507. https://doi.org/10.1038/d41586-022-00150-2

Coles, N. A., March, D. S., Marmolejo-Ramos, F., Larsen, J. T., Arinze, N. C., Ndukaihe, I. L. G., Willis, M. L., Foroni, F., Reggev, N., Mokady, A., Forscher, P. S., Hunter, J. F., Kaminski, G., Yüvrük, E., Kapucu, A., Nagy, T., Hajdu, N., Tejada, J., Freitag, R. M. K., & Zambrano, D. (2022b). A multi-lab test of the facial feedback hypothesis by the Many Smiles Collaboration. *Nature Human Behaviour*, *6*(6), 1–12. https://doi.org/10.1038/s41562-022-01458-9

Coles, N. A., DeBruine, L. M., Azevedo, F., Baumgartner, H. A., & Frank, M. C. (2023). 'Big team' science challenges us to reconsider authorship. *Nature Human Behaviour, 7*(5), 665-667. https://doi.org/10.1038/s41562-023-01572-2

Cologna, V., Mede, N. G., Berger, S., Besley, J., Brick, C., Joubert, M., Maibach, E. W., Mihelj, S., Oreskes, N., Schäfer, M. S., Van Der Linden, S., Aziz, N. I. A., Abdulsalam, S., Shamsi, N. A., Aczel, B., Adinugroho, I., Alabrese, E., Aldoh, A., Alfano, M., . . . Zwaan, R. A. (2025). Trust in scientists and their role in society across 68 countries. Nature Human Behaviour. https://doi.org/10.1038/s41562-024-02090-5

Corral-Frías, N. S., Castillo, E. L., Lucas, M. Y., Armenta, M. F., Rodriguez, Y. V., Dutra, N., Micheli, L., & Azevedo, F. (2023). Latin American Psychological Science: Will the Global North Make Room?. *APS Observer, 36*.

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology, 7*2, 78–89. https://doi.org/10.1016/j.jmp.2015.10.003

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

Dorison, C., Lerner, J. S., Heller, B., Rothman, A. J., Kawachi, I., Wang, K., Rees, V. W., Gill, B., Gibbs, N., Ebersole, C. R., Vally, Z., Tajchman, Z., Zsido, A. N., Zrimsek, M., Chen, Z., Ziano, I., Gialitaki, Z., Ceary, C. D., Lin, Y., . . . Ren, D. (2022). In COVID-19 Health Messaging, Loss Framing Increases Anxiety with Little-to-No Concomitant Benefits: Experimental Evidence from 84 Countries. *Affective Science, 3*(3), 577–602. https://doi.org/10.1007/s42761-022-00128-3

Douma, J. C., & Weedon, J. T. (2019). Analysing continuous proportions in ecology and evolution: A practical introduction to beta and Dirichlet regression. *Methods in Ecology and Evolution*, *10*(9), 1412–1430. https://doi.org/10.1111/2041-210x.13234

Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J., Banks, J., Baranski, E., Bernstein, M. J., Bonfiglio, D. B. V., Boucher, L., Brown, E. R., Budiman, N. I., Cairo, A. H., Capaldi, C. A., Chartier, C. R., Chung, J. M., Cicero, D. C., Coleman, J. A., Conway, J., . . . Nosek, B. A. (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology, 67*, 68–82. https://doi.org/10.1016/j.jesp.2015.10.012

Ebersole, C. R., Mathur, M. B., Baranski, E., Bart-Plange, D., Buttrick, N. R., Chartier, C. R., Corker, K. S., Corley, M., Hartshorne, J. K., IJzerman, H., Lazarević, L. B., Rabagliati, H., Ropovik, I., Aczel, B., Aeschbach, L. F., Andrighetto, L., Arnal, J., Arrow, H., Babincak, P., . . . Montealegre, A. (2020). Many labs 5: Testing Pre-Data-Collection Peer review as an intervention to increase replicability. *Advances in Methods and Practices in Psychological Science, 3*(3), 309–331. https://doi.org/10.1177/2515245920958687

Education First (2022). *EF EPI English Proficiency*. https://www.ef.com/wwen/epi/.

Eriksson, K., Strimling, P., Gelfand, M., Wu, J., Abernathy, J., Akotia, C. S., Aldashev, A., Andersson, P. A., Andrighetto, G., Anum, A., Arikan, G., Aycan, Z., Bagherian, F., Barrera, D., Basnight-Brown, D., Batkeyev, B., Belaus, A., Berezina, E., Björnstjerna, M., & Blumen, S. (2021). Perceptions of the appropriate response to norm violation in 57 societies. *Nature Communications, 12*(1), 1481. https://doi.org/10.1038/s41467-021-21602-9

Espinosa, J., Stevens, J.R., Alberghina, D., Alway, H.E.E., Barela, J.D., Bogese, M., Bray, E.E., Buchsbaum, D. Byosiere, S.-E., Byrne, M., Cavalli, C. M., Chaudoir, L. M., Collins-Pisano, C., DeBoer, H.J., Douglas L.E.L.C., Dror, S., Dzik, M.V., Ferguson, B., Fisher, L., . . . Zylberfuden, S.G. (2023). ManyDogs 1: A multi-lab replication study of dogs' pointing comprehension. *Animal Behavior & Cognition, 10*(3), 232-28. https://doi.org/10.26451/abc.10.03.03.2023

Forscher, P. S., Wagenmakers, E., Coles, N. A., Silan, M. a. A., Dutra, N. B., Basnight-Brown, D., & IJzerman, H. (2023). The benefits, barriers, and risks of Big-Team science. *Perspectives on Psychological Science*, *18*(3), 607–623. https://doi.org/10.1177/17456916221082970

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

Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, P.-C., & Modrák, M. (2020). Bayesian Workflow. https://doi.org/10.48550/arXiv.2011.01808

Ghai, S. (2021). It's time to reimagine sample diversity and retire the WEIRD dichotomy. *Nature Human Behaviour*, *5*(8), 971-972. https://doi.org/10.1038/s41562-021-01175-9

Ghai, S., Forscher, P. S., & Chuan-Peng, H. (2024). Big-team science does not guarantee generalizability. *Nature Human Behaviour*, 8, 1053–1056. https://doi.org/10.1038/s41562-024-01902-y

Grahe, J. E., Chalk, H. M., Cramblet Alvarez, L. D., Faas, C. S., Hermann, A. D., & McFall, J. P. (2018). Emerging adulthood measured at multiple institutions 2: The data. *Journal of Open Psychology Data*, *6*(1). https://doi.org/10.5334/jopd.38

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

Hoekstra, R., & Vazire, S. (2021). Aspiring to greater intellectual humility in science. *Nature Human Behaviour*, *5*(12), 1602-1607. https://doi.org/10.1038/s41562-021-01203-8

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., Lindenberg, S., Dalgar, I., Weissgerber, S. C., Vergara, R. C., Cairo, A. H., Čolić, M., Dursun, P., Frankowska, N., Hadi, R., Hall, C. J., Hong, Y., Joy-Gaba, J. A., Lazarević, D., Lazarević, L. B., Parzuchowski, M., Ratner, K. G., Rothman, D. J., . . . IJzerman, H. (2019). Data from the Human Penguin Project, a cross-national dataset testing social thermoregulation principles. *Scientific Data*, *6*(1). https://doi.org/10.1038/s41597-019-0029-2

IJzerman, H., Dutra, N., Silan, M., Adetula, A., Brown, D. M. B., & Forscher, P. (2021). Psychological science needs the entire globe, Part 1. *APS Observer*, *34*.

IJzerman, H., Lewis, N. A., Przybylski, A. K., Weinstein, N., DeBruine, L. M., Ritchie, S. J., Vazire, S., Forscher, P. S., Morey, R. D., Ivory, J., & 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., Dalgar, I., Weissgerber, S. C., Vergara, R. C., Cairo, A. H., Čolić, M., Dursun, P., Frankowska, N., Hadi, R., Hall, C. J., Hong, Y., Chuan-Peng, H., Joy-Gaba, J. A., Lazarević, D., Lazarević, L. B., Parzuchowski, M., Ratner, K. G., Rothman, D. J., . . . Zickfeld, J. (2018). The Human Penguin Project: Climate, social integration, and core Body temperature. *Collabra: Psychology, 4*(1), 37. https://doi.org/10.1525/collabra.165

International Union of Psychological consists. (2023). *National Members*. https://www.iupsys.net/about/members/national-members/

Jeftic, A., Lucas, M. Y., Corral-Frias, N., & Azevedo, F. (2024). Bridging the majority and minority worlds: Liminal researchers as catalysts for inclusive open and big-team science. *A better how: Notes on developmental meta-research.* https://doi.org/10.62372/ISCI6112.

Jin, H., Wang, Q., Yang, Y., Zhang, H., Gao, M., Jin, S. J., Chen, Y., Xu, T., Zheng, Y., Chen, J., Xiao, Q., Yang, J., Wang, X., Geng, H., Ge, J., Wang, W., Chen, X., Zhang, L., Zuo, X., & Chuan-Peng, H. (2023). The Chinese Open Science Network (COSN): Building an open science community from scratch. *Advances in Methods and Practices in Psychological Science*, *6*(1), 251524592211449. https://doi.org/10.1177/25152459221144986

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., Foroni, F., Willis, M., Cubillas, C. P., Vadillo, M. A., Turiegano, E., Gilead, M., Simchon, A., Saribay, S. A., Owsley, N., Jang, C., . . . Baskin, E. (2021). To which world regions does the valence–dominance model of social perception apply? *Nature Human Behaviour, 5*(1), 159–169. https://doi.org/10.1038/s41562-020-01007-2

Khan, T., Abimbola, S., Kyobutungi, C., & Pai, M. (2022). How we classify countries and people—and why it matters. BMJ Global Health, 7(6), e009704. https://doi.org/10.1136/bmjgh-2022-009704

Kijilian, C., Sharifian, M., Paris, B., Iyer, A., & Silan, M. A. (2022, September 2). PSA Study Capacity Report 2022-2023. https://doi.org/10.31234/osf.io/7krhn

Kissel, H. A., & Friedman, B. H. (2023). Participant diversity in Psychophysiology. *Psychophysiology*, *60*(11), e14369. https://doi.org/10.1111/psyp.14369

Klein, R. A., Cook, C. L., Ebersole, C. R., Vitiello, C., Nosek, B. A., Hilgard, J., Ahn, P. H., Brady, A. J., Chartier, C. R., Christopherson, C. D., Clay, S. L., Collisson, B., Crawford, J. T., Cromar, R., Gardiner, G., Gosnell, C. L., Grahe, J., Hall, C. J., Howard, I., . . . Ratliff, K. A (2022). Many Labs 4: Failure to replicate mortality salience effect with and without original author involvement. *Collabra: Psychology*, *8*(1), 35271.https://doi.org/10.1525/collabra.35271

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., . . . Nosek, B. A. (2014). Investigating variation in replicability. *Social Psychology*, *45*(3), 142–152. https://doi.org/10.1027/1864-9335/a000178

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., Aveyard, M., Axt, J., 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., . . . Lazarević, L. B. (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

KOF Swiss Economic Institute. (2022) *KOF Globalisation Index.* https://kof.ethz.ch/en/forecasts-and-indicators/indicators/kof-globalisation-index.html

Korbmacher, M., Azevedo, F., Pennington, C. R., Hartmann, H., Pownall, M., Schmidt, K., Elsherif, M., Breznau, N., Robertson, O., Kalandadze, T., Yu, S., Baker, B. J., O'Mahony, A., Olsnes, J. Ø.-S., Shaw, J. J., Gjoneska, B., Yamada, Y., Röer, J. P., Murphy, J., Alzahawi, S., Grinschgl, S., Oliveira, C. M., Wingen, T., Yeung, S. K., Liu, M., König, L. M., Albayrak-Aydemir, N., Lecuona, O., Micheli, L., & Evans, T. (2023). The replication crisis has led to positive structural, procedural, and community changes. *Communications Psychology, 1*(1), 3 https://doi.org/10.1038/s44271-023-00003-2

Kruschke, J. K. (2018). Rejecting or accepting parameter values in Bayesian estimation. *Advances in methods and practices in psychological science*, *1*(2), 270-280. https://doi.org/10.1177/2515245918771304

Kühberger, A., Fritz, A., & Scherndl, T. (2014). Publication bias in psychology: A diagnosis based on the correlation between effect size and sample size. *PLoS ONE, 9*(9). https://doi.org/10.1371/journal.pone.0105825

Lin, Z., & Li, N. (2022). Global diversity of authors, editors, and journal ownership across subdisciplines of psychology: current state and policy implications. *Perspectives on Psychological Science*, *18*(2), 358–377. https://doi.org/10.1177/17456916221091831

Lin, Z., & Lu, S. (2023). Exponential authorship inflation in neuroscience and psychology from the 1950s to the 2020s. *American Psychologist.* Advance online publication. https://doi.org/10.1037/amp0001216

Liu, W., Chen, Z., Chuan-Peng, Hu. (2024). Sample representativeness in psychological and brain science research (in Chinese). *Chinese Science Bulletin. 69*(24), 3515–3531. https://doi.org/10.1360/TB-2024-0322

Medin, D., Ojalehto, B., Marin, A., & Bang, M. (2017). Systems of (non-) diversity. *Nature Human Behaviour*, *1*(5), 0088. https://doi.org/10.1038/s41562-017-0088

Melitz, J., & Toubal, F. (2014). Native language, spoken language, translation and trade. *Journal of International Economics, 93*(2), 351-363. https://doi.org/10.1016/j.jinteco.2014.04.004

Miller, R., Lambert, M. L., Frohnwieser, A., Brecht, K. F., Bugnyar, T., Crampton, I., ... & Clayton, N. S. (2022). Socio-ecological correlates of neophobia in corvids. *Current Biology, 32*(1), 74-85. https://doi.org/10.1016/j.cub.2021.10.045

Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., ... & 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

Muthukrishna, M., Bell, A. V., Henrich, J., Curtin, C. M., Gedranovich, A., McInerney, J., & Thue, B. (2020). Beyond Western, Educated, Industrial, Rich, and Democratic (WEIRD) psychology: Measuring and mapping scales of cultural and psychological distance. *Psychological science, 31*(6), 678-701. https://doi.org/10.1177/0956797620916782

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

Nikooienejad, A., & Johnson, V. E. (2021). On the existence of uniformly most powerful bayesian tests with application to non-central chi-squared tests. *Bayesian analysis, 16*(1), 93. https://doi.org/10.1214/19-ba1194

Nogrady, B. (2023). Hyperauthorship: the publishing challenges for ‘big team’ science. *Nature, 615*, 175-177. https://doi.org/10.1038/d41586-023-00575-3

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Our World in Data. (2023 July 17). *Average years of schooling.* https://ourworldindata.org/grapher/mean-years-of-schooling-long-run

Paris, B., IJzerman, H., & Forscher, P. S. (2020, December 2). PSA 2020-2021 study capacity report. https://doi.org/10.31234/osf.io/v9zma

Pavlov, Y. G., Adamian, N., Appelhoff, S., Arvaneh, M., Benwell, C. S. Y., Beste, C., Bland, A. R., Bradford, D. E., Bublatzky, F., Busch, N. A., Clayson, P. E., Cruse, D., Czeszumski, A., Dreber, A., Dumas, G., Ehinger, B., Ganis, G., He, X., Hinojosa, J. A., . . . Mushtaq, F. (2021). #EEGManyLabs: Investigating the replicability of influential EEG experiments. *Cortex, 144*, 213–229. https://doi.org/10/gjx8qt

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

Psychological Science Accelerator Self-Determination Theory Collaboration. (2022). A global experiment on motivating social distancing during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences, 119*(22), e2111091119. https://doi.org/10.1073/pnas.2111091119

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

R Core Team (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Röseler, L., Kaiser, L., Doetsch, C., Klett, N., Seida, C., Schütz, A., Aczel, B., Adelina, N., Agostini, V., Alarie, S., Albayrak-Aydemir, N., Aldoh, A., Al-Hoorie, A. H., Azevedo, F., Baker, B. J., Barth, C. L., Beitner, J., Brick, C., Brohmer, H., . . . Zhang, Y. (2024). The Replication Database: Documenting the replicability of Psychological Science. *Journal of Open Psychology Data*, *12*(1), 8. https://doi.org/10.5334/jopd.101

Ruggeri, K., Panin, A., Vdovic, M., Većkalov, B., Abdul-Salaam, N., Achterberg, J., Akil, C., Amatya, J., Amatya, K., Andersen, T. L., Aquino, S., Arunasalam, A., Ashcroft-Jones, S., Askelund, A. D., Ayacaxli, N., Sheshdeh, A. B., Bailey, A., Arroyo, P. B., Mejía, G. B., . . . Meyer, S. A. (2022). The globalizability of temporal discounting. *Nature Human Behaviour*, *6*(10), 1386–1397. https://doi.org/10.1038/s41562-022-01392-w

Saucier, G., Kenner, J., Iurino, K., Malham, P. B., Chen, Z., Thalmayer, A. G., Kemmelmeier, M., Tov, W., Boutti, R., Metaferia, H., Çankaya, B., Mastor, K. A., Hsu, K., Wu, R., Maniruzzaman, M., Rugira, J., Tsaousis, I., Sosnyuk, O., Adhikary, J. R., . . . Altschul, C. (2014). Cross-Cultural differences in a global "Survey of World Views." *Journal of Cross-Cultural Psychology*, *46*(1), 53–70. https://doi.org/10.1177/0022022114551791

Schimmelpfennig, R., Spicer, R., White, C., Gervais, W. M., Norenzayan, A., Heine, S., Henrich, J., & Muthukrishna, M. (2024). The Moderating Role of Culture in the Generalizability of Psychological Phenomena. *Advances in Methods and Practices in Psychological Science*, *7*(1). https://doi.org/10.1177/25152459231225163

Singh, L., Cristia, A., Karasik, L. B., Rajendra, S. J., & Oakes, L. M. (2023). Diversity and representation in infant research: Barriers and bridges toward a globalized science of infant development. *Infancy, 28*(4), 708-737. https://doi.org/10.1111/infa.12545

Steffan, A., Zimmer, L., Arias-Trejo, N., Bohn, M., Rodrigo Dal Ben, Flores-Coronado, M. A., Franchin, L., Garbisch, I., Charlotte Grosse Wiesmann, J Kiley Hamlin, Havron, N., Hay, J. F., Hermansen, T. K., Jakobsen, K. V., Kalinke, S., Ko, E.-S., Kulke, L., Mayor, J., Marek Meristo, & Moreau, D. (2023). Validation of an open source, remote web-based eye-tracking method (WebGazer) for research in early childhood. *Infancy*, *29*(1), 31–55. https://doi.org/10.1111/infa.12564

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

The World Bank. (2022a). *World development indicators: GDP per capita.* https://data.worldbank.org.cn/indicator/NY.GDP.PCAP.CN

The World Bank. (2022b). *World development indicators: Urban population (proportion in total population).* https://data.worldbank.org.cn/indicator/SP.URB.TOTL.IN.ZS?end=2022&start=1960&view=chart

The World Bank. (2022c). *World development indicators: Individuals using the Internet (% of population).* https://data.worldbank.org.cn/indicator/IT.NET.USER.ZS

The World Bank (2024). *World Bank Country and Lending Groups*. https://datahelpdesk.worldbank.org/knowledgebase/articles/906519

Tindle, R. (2021). Improving the global reach of psychological research. *Discover Psychology, 1*, 1-4. https://doi.org/10.1007/s44202-021-00004-4

Tiokhin, L., Hackman, J., Munira, S., Jesmin, K., & Hruschka, D. (2019). Generalizability is not optional: Insights from a cross-cultural study of social discounting. *Royal Society open science, 6*(2), 181386. https://doi.org/10.1098/rsos.181386

Trübutschek, D., Yang, Y., Gianelli, C., Cesnaite, E., Fischer, N. L., Vinding, M. C., Marshall, T.R., Algermissen, J., Pascarella, A., Puoliväli, T., Vitale, A., Busch, N. A., Nilsonne, G., Nilsonne, G. (2022). EEGManyPipelines: A large-scale, grass-root multi-analyst study of EEG analysis practices in the wild. *MetaArXiv*. https://doi.org/10.31222/osf.io/jq342

Tybur, J. M., Inbar, Y., Aarøe, L., Barclay, P., Barlow, F. K., De Barra, M., Becker, D. V., Borovoi, L., Choi, I., Choi, J. A., Consedine, N. S., Conway, A., Conway, J. R., Conway, P., Adoric, V. C., Demirci, D. E., Fernández, A. M., Ferreira, D. C. S., Ishii, K., . . . Žeželj, I. (2016). Parasite stress and pathogen avoidance relate to distinct dimensions of political ideology across 30 nations. *Proceedings of the National Academy of Sciences*, *113*(44), 12408–12413. https://doi.org/10.1073/pnas.1607398113

Uhlmann, E. L., Ebersole, C. R., Chartier, C. R., Errington, T., Kidwell, M. C., Lai, C. K., McCarthy, R. J., Riegelman, A., Silberzahn, R., & Nosek, B. A. (2019). Scientific Utopia III: Crowdsourcing Science. *Perspectives on Psychological Science*, *14*(5), 711–733. https://doi.org/10.1177/1745691619850561

UNESCO. (2021a). *UNESCO Science Report: The race against time for smarter development; executive summary* (SC-2021/WS/7). https://unesdoc.unesco.org/ark:/48223/pf0000377250

UNESCO. (2021b). *UNESCO Science Report: The Race Against Time for Smarter Development. S. Schneegans, T. Straza and J. Lewis (eds).* UNESCO Publishing: Paris. https://www.unesco.org/reports/science/2021/en/statistics

United Nations Population Division. (2022). *World population prospects*. https://population.un.org/wpp

U.S. Census Bureau. (2021). *Quick facts.* https://www.census.gov/quickfacts/fact/table/US/PST045221

Vlasceanu, M., Doell, K. C., Bak-Coleman, J. B., Todorova, B., Berkebile-Weinberg, M. M., Grayson, S. J., Patel, Y., Goldwert, D., Pei, Y., Chakroff, A., Pronizius, E., Van Den Broek, K. L., Vlasceanu, D., Constantino, S., Morais, M. J., Schumann, P., Rathje, S., Fang, K., Aglioti, S. M., . . . Van Bavel, J. J. (2024). Addressing climate change with behavioral science: A global intervention tournament in 63 countries. *Science Advances*, *10*(6). https://doi.org/10.1126/sciadv.adj5778

Wagenmakers, E., 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., Van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., . . . 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, K., Goldenberg, A., Dorison, C. A., Miller, J. K., Uusberg, A., Lerner, J. S., ... & Štrukelj, E. (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

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

Webometrics Ranking of World Universities. (2023). *Countries arranged by Number of Universities in Top Ranks*. https://www.webometrics.info/en/distribution_by_country

Wilson, S. (2024). Sociodemographic reporting and sample composition over 3 decades of psychopathology research: A systematic review and quantitative synthesis. *Journal of Psychopathology and Clinical Science, 133*(1), 20 -36. https://doi.org/10.1037/abn0000871

Yang, Q., Zhang, W., Liu, S., Gong, W., Han, Y., Lu, J., Jiang, D., Nie, J., Lyu, X., Liu, R., Jiao, M., Qu, C., Zhang, M., Sun, Y., Zhou, X., & Zhang, Q. (2023). Unraveling controversies over civic honesty measurement: An extended field replication in China. *Proceedings of the National Academy of Sciences of the United States of America, 120*(29). https://doi.org/10.1073/pnas.2213824120

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences, 45*. https://doi.org/10.1017/S0140525X20001685

Yue, L., Zuo, X.-N., & Hu, C.-P. (2023) The WEIRD problem in a “non-WEIRD” context: A meta-research on the representativeness of human subjects in Chinese psychological research, in principle acceptance of Version 7 by Peer Community in Registered Reports. https://osf.io/ehw54

Supplementary materials

1. Bayesian workflow of Bayesian zero-inflated negative binomial regression model

Step 1: check data

GDP per capita is the independent variable and is continuous data, however, it is not normally distributed, as evidenced by visual inspection (see Figure S1) and the Shapiro-Wilk test ($p < .001$). To address this, we applied a logarithmic transformation to the GDP per capita data.

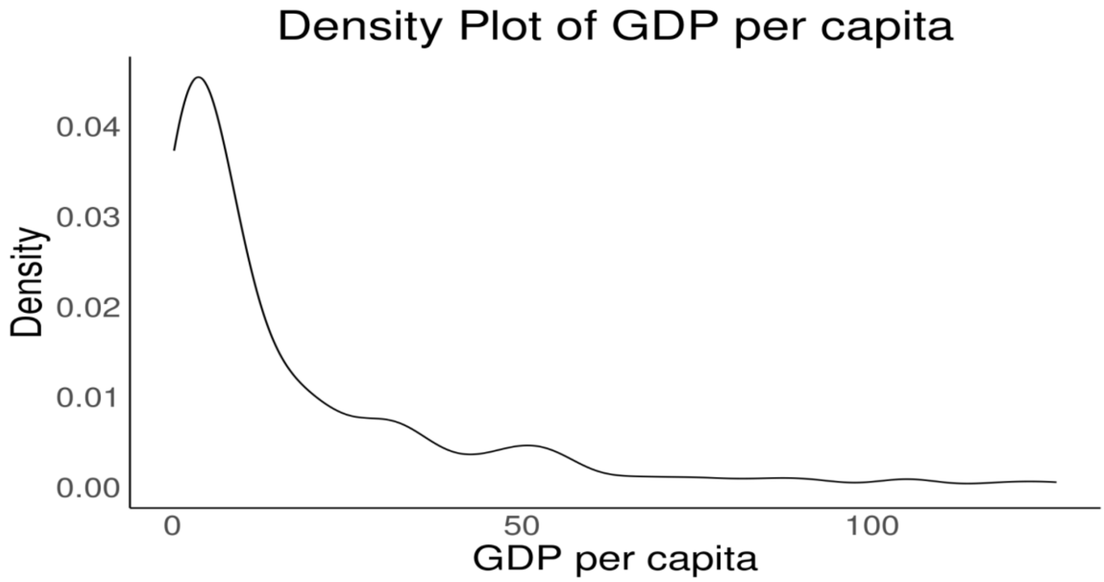

Figure S1. Density plot of GDP per capita

The sample size is the dependent variable, which is a simple count data that contains many zeros, and non-zero data have over-dispersion. To diagnose overdispersion of count data, we used the model comparison leaf one out cross validation (LOO-CV) method to compare the zero-inflation Poisson model with the zero-inflation negative binomial regression (Winter & Bürkner, 2021). The results show that the zero inflation Poisson model fits worse (*elpd-diff* = -5872.9, *se-diff* = 3133.1), indicating that the data are overdispersion. Therefore, Bayesian zero-inflated negative binomial regression is appropriate to process the data.

Step 2: Check model parameters

For zero-inflated negative binomial Regression:

$$\mathbf{y_i \sim ZINB(\mu_i, \phi, p_i)}$$
$$\mathbf{\log(\mu_i) = \beta_0 + \beta_1 \times \chi_i}$$
$$\mathbf{Logit(p_i) = \gamma_0 + \gamma_1 \times \chi_i}$$

Including 3 parameters:

μ is the mean of the negative binomial regression part;

ϕ is the shape parameter of the negative binomial regression part;

p is the probability of zero.

Step 3: Set priors

(1) The shape parameter ϕ

For the shape parameters in Bayesian negative binomial regression, an inv_gamma(0.4, 0.3) prior is appropriate, as it can cover a broader range (Bürkner, 2024; https://github.com/paul-buerkner/brms/issues/1614). Therefore, we will also use an inv_gamma (0.4, 0.3) prior for the shape parameters in our Bayesian zero-inflated negative binomial regression model.

(2) The intercept β_0 of μ_i

Based on the self-compiled big team science data, the total sample size, after filtering according to our operational definition, ranges from 1,100,000 to 1,200,000. To ensure that the intercept β_0 can cover all possible ranges from 1 to 1,200,000, we consider two priors: a more diffuse prior, $\beta_0 \sim$ Normal (0, 4.67), and a more concentrated prior, $\beta_0 \sim$ Normal (6, 2.67).

Using the properties of the log-normal distribution, the $\beta_0 \sim$ Normal (0, 4.67) prior covers a range from about 8.23e-07 ($e^{0-4.67\times3}$) to 1,214,691 ($e^{0+4.67\times3}$), while the $\beta_0 \sim$ Normal (6, 2.67) prior ranges from about 0.13 ($e^{6-2.67\times3}$) to 1,214,691 ($e^{6+2.67\times3}$). Both priors effectively cover the target range of 1 to 1,200,000.

We have made a visual comparison between the intercept priors $\beta_0 \sim$ Normal (6, 2.67) and $\beta_0 \sim$ Normal (0, 4.67) (see figure below). Notably, the prior $\beta_0 \sim$ Normal (0, 4.67) is more extreme, with a median of $e^0 = 1$, whereas the median for $\beta_0 \sim$ Normal (6, 2.67) is $e^6 = 403$.

Given that one big team science study sampled from 187 countries, and that there are approximately 230 countries and regions worldwide, it is reasonable to conclude that all big team science samples we ultimately collected represent the majority of countries worldwide. As the participants in most big team science studies come from dozens of countries, countries representing the median sample size in the final data are likely to appear less frequently in big team science. For this median, sample sizes in the range of a few hundred are appropriate. Therefore, $\beta_0 \sim$ Normal (6, 2.67) is the more appropriate prior.

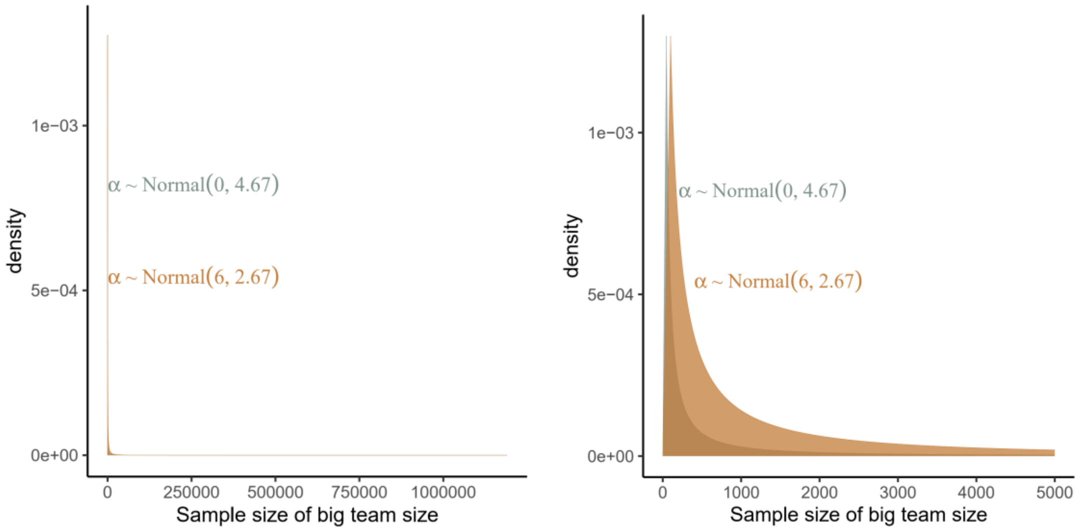

Figure S2. Prior prediction checking of β_0. The right panel shows the same data as the left panel, but with the range of the x-axis restricted to better highlight the differences between the two distributions.

(3) The intercept γ_0 of p_i

The probability of 0 in our data may be mostly concentrated between 0.1-0.9, and after logit transformation it is (-2.2, 2.2). We set $\gamma_0 \sim$ Normal (0, 1), which corresponds exactly to this range.

(4) The slope β_1 of μ_i and the slope γ_1 of p_i

For the slope parameters β_1 and γ_1, we use the changing slope to plot the prior prediction distribution and then select the appropriate prior (For detailed information, please refer to https://osf.io/y8twn).

The prior selection can meet two criteria (as shown below): (1) it is within the possible range of values (there are values on the x-axis, and the sample size can take the larger value of 1,200,000); (2) the change in the data is not extreme.

Based on the above criteria, the priors we finally chose were β_1 ~ Normal (0, 0.1) & γ_1 ~ Normal (0, 0.1).

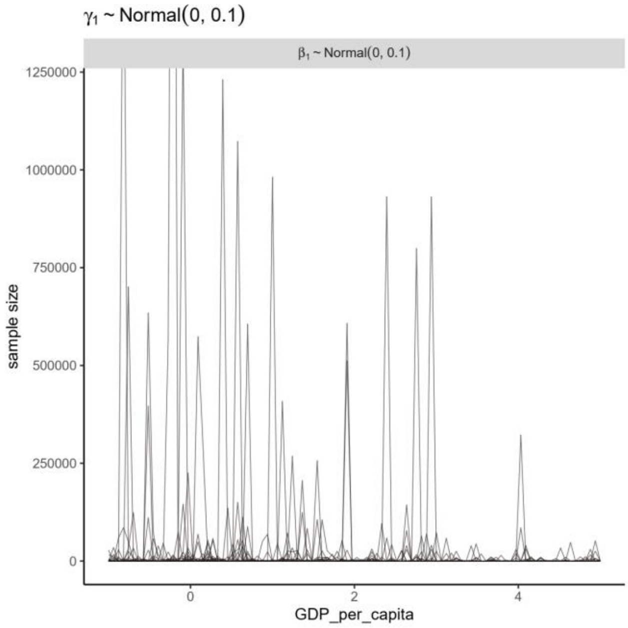

Figure S3. Prior prediction checking of β_0 & γ_1.

In summary, the priors we use are as follows:

```
prior = set_prior("normal(0, 0.1 )", class = "b") +
        set_prior("normal(6, 2.67)", class = "Intercept")+
        set_prior("inv_gamma(0.4, 0.3)", class = "shape")+
        set_prior("normal(0, 0.1)", class = "b", dpar= "zi") +
        set_prior("normal(0, 1)", class = "Intercept", dpar= "zi")
```

Next, we will use this prior for further prior prediction checks.

Step 4: Prior prediction checking

The visualization of prior predictive checks is presented in Figure S4. In the left panel we can see that the sample size can accommodate values up to 1,200,000. In the right panel, the actual data fall within the range of the prior distribution, demonstrating the appropriateness of the prior.

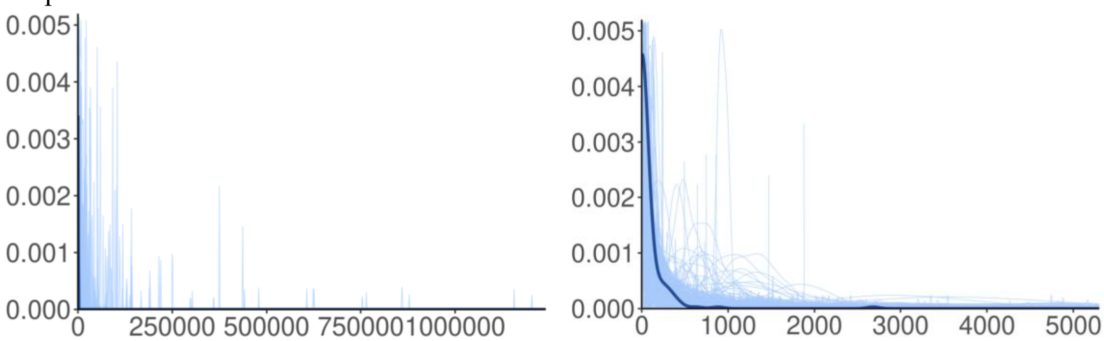

Figure S4. Prior prediction checking that includes all parameters. The right panel shows the same data as the left panel, but with the range of the x-axis restricted to better highlight actual data.

Step 5: Model operation results

Interpreting results:

(1) Intercept = 5.43 means that when log (GDP per capita) = 0, i.e. GDP per capita = 1 ($1000), the model predicts a sample size of exp (5.15) = 95.58.

(2) log_GDP_per_capita = 0.34 means that when GDP per capital increases by 1%, the model's predicted sample size increases by 0.34%.

(3) Zi_intercept = 1.38 means that when log (GDP per capita) = 0, i.e. GDP per capita=1 ($1000), the probability of the model predicting a sample size of 0 is plogis (1.38) = 79.90%.

(4) Zi_log_GDP_per_capita = -0.09 means that when GDP per capita increases by 1%, the probability of the model predicting a sample size of 0 decreases by 0.08%.

(5) Shape = 1.33 indicates excessive dispersion of the sample size data. Based on the variance formula below, the closer the shape parameter is to positive infinity, the smaller the data dispersion, and the closer the shape parameter is to 0, the larger the data dispersion.

$$E(Y|\mu, r) = \mu \qquad Var(Y|\mu, r) = \mu + \frac{\mu^2}{r}$$

Step 6: Posterior prediction checking

First, as shown in panel (a) of Figure S5, the actual data are well contained within the posterior predictive distribution. In addition, panel (b) shows that the minimum value of the actual data, zero, is exactly at the center of the posterior predictive distribution. Panel (d) shows that although the maximum value of 2,682 is an outlier (since all other values are below 1,000) and does not fall entirely at the center of the posterior predictive distribution, the distribution sufficiently covers this extreme value, making it acceptable. In other words, the posterior predictive distribution adequately covers the range of values in the actual data. Panel (c) shows that the mean of the actual data is approximately at the center of the posterior predictive distribution, indicating that the distribution adequately represents the central tendency of the actual data. Overall, this suggests that our model effectively describes the actual data.

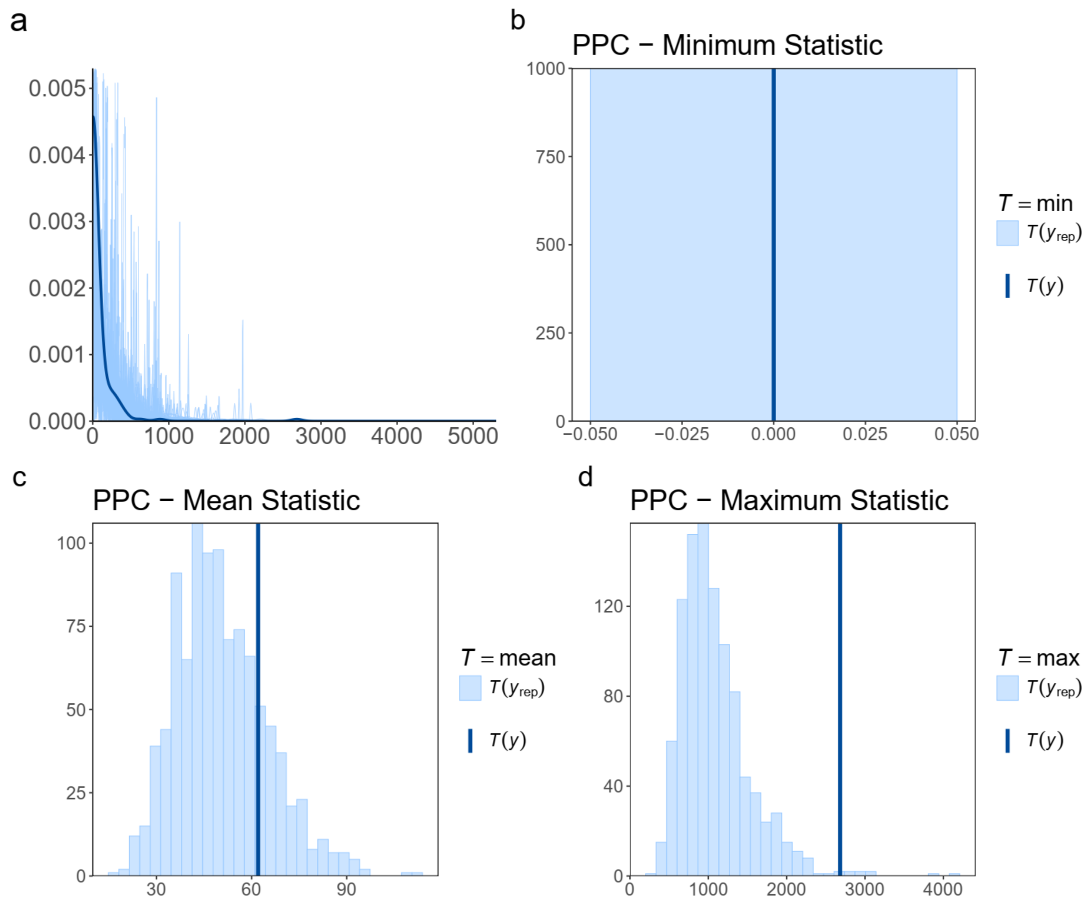

Figure S5. Prior prediction checking

In summary, the prior we have set is appropriate based on current knowledge.