# North and South: Naming practices and the hidden dimension of global disparities in knowledge production

Andrés F. Castro Torres[a,1] and Diego Alburez-Gutierrez[b]

[a]Laboratory of Fertility and Well-Being, Max Planck Institute for Demographic Research, 18057 Rostock, Germany; and [b]Laboratory of Digital and Computational Demography, Max Planck Institute for Demographic Research, 18057 Rostock, Germany

Edited by Susan Fiske, Princeton University, Princeton, NJ; received October 27, 2021; accepted January 23, 2022

**The legacy of Eurocentrism continues to affect knowledge production in the social sciences. Evidence produced in and about the global North is assumed to be more "universal," whereas evidence from or produced in the global South is considered valid only for specific contexts (i.e., "localized"). We argue that these dynamics are evident in the phrasing of articles' titles based on the examination of more than half a million social science research articles indexed by Scopus (1996 to 2020). We find that empirical articles written by authors affiliated to institutions of the global North, using data from these countries, are less likely to include a concrete geographical reference in their titles. When authors are affiliated to global South institutions, and use evidence from global South countries, the names of these countries are more likely to be part of the article's title. We confirm this overarching pattern by looking at 1) differences between world regions, 2) differences within world regions, and 3) patterns in 23 social science subfields. These gaps are large and consistent, yet article naming conventions are merely the "tip of the iceberg" of the imbalances in knowledge production between the global North and South.**

knowledge production | global inequalities | Eurocentrism

The development of institutionalized modern social sciences was closely linked to the European colonial projects that spanned from the 15th to the 20th century (1, 2). As a consequence, the production of academic knowledge is embedded in power structures that can be characterized in terms of center–periphery relations (3, 4). The imprint of these centuries of economic and political subordination on knowledge production has neither been fully understood nor overcome (5). Particularly, Eurocentrism, understood as a worldview that considers Western thought as culturally and intellectually superior, continues to shape the global production of social sciences, including its questions, methods, and approaches (6).

One problematic aspect of this perspective is that it glosses over the historical contingencies and structural violence that produced and sustain Western hegemony, including the imposition of metrics that makes the West the "default case" and the search for universal, timeless, and context- and value-free knowledge in science. This might result in societal processes observed in countries of the global North, such as market-based economic growth and rising human development being considered the "default" cases toward which other nations and societies ought to converge in the mid- or long term (7, 8). Multiple calls to decolonize social science research have been made (1, 9, 10); yet much needs to be done before we can claim that the social sciences have overcome their colonial past and the consequent Eurocentric view of social processes.

This article examines one aspect in which the Eurocentric view of empirical social science research manifests itself, namely the degree to which articles studying peoples in the global North are explicit about which populations they study. Failing to disclose the geographical provenance of the empirical evidence in the title, an article's most visible section, might be misleading as to the generalizability of the findings (11). Readers may interpret these "delocalized" titles as describing universal processes, as is the case with generic statements (12). There have been multiple challenges to the unwarranted generalizability of psychology studies on Western, Educated, Industrialized, Rich and Democratic (WEIRD) samples (13, 14) and multiple, local, epistemological alternatives developed by scholars in the global South. Yet despite this growing awareness, the production of mainstream narratives about contemporary social processes remains largely Eurocentric because of the economic, political, and cultural hegemony of the global North (3, 10, 15).

**Scientometric Studies of Titles' Characteristics.** The title is the most visible part of an academic article, summarizing its content and aiming to attract potential readers. Indeed, a title's phrasing is a crucial component of science communication. This is particularly true for empirical studies in the social sciences, which, unlike studies in the natural sciences, are highly contingent on contextual social factors. There is evidence to suggest that authors choose the wording of their titles strategically (16). Short, generic, and amusing titles are more likely to capture researchers' attention at first glance (17, 18). However, long, specific, and detailed titles are informative and may therefore be preferred if authors want to convey specific ideas and attract specialized audiences (19).

## Significance

**Contemporary social sciences aim to be diverse and inclusive, but traces of the historical dominance of Western European and North American academic institutions persist in scientific practices. One such practice is the phrasing of article titles. Our analysis shows that articles studying the global North are systematically less likely to mention the name of the country they study in their title compared to articles on the global South. This constitutes, potentially, an unwarranted claim on universality and may lead to lesser recognition of global South studies. Social and behavioral scientists must reflect on the phrasing of their article titles to avoid reproducing harmful relations of intellectual domination which limit inclusivity and constitute a barrier to the generalizability of scientific knowledge.**

Author contributions: A.F.C.T. and D.A.-G. designed research; A.F.C.T. and D.A.-G. performed research; A.F.C.T. analyzed data; and A.F.C.T. and D.A.-G. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.



[1]To whom correspondence may be addressed. Email: castro@demogr.mpg.de.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2119373119/-/DCSupplemental.

Published March 1, 2022.

Several studies have documented the variability in title characteristics and how they affect an article's readership (20). Previous studies have examined titles' length and syntactic structures (21–24), the use of question marks and semicolons (25), the prevalence of generic expressions (12), and the inclusion of country names (11). Our focus is on exploring the propensity to include geographical references in titles across geographies and subdisciplines of the social sciences.

**Geographical References in Titles and Generic Language.** We distinguish "localization" from "delocalization" as two distinct strategies for the naming of articles. In the former, authors include a concrete geographical reference to the context or population being studied (i.e., the title is "marked"). In the latter, this concrete reference is omitted (i.e., the title is "unmarked"). Therefore, the degree to which titles are localized or not likely reflects how scientists think about their data—its scope, validity, and generalizability—and their audience—who should read their work and for what purposes (14, 26).

Some authors may select a bold title that glosses over the context, variability, uncertainty, and limitations of their results, trusting that a careful reading of the entire research will make these points clear. Other authors may prefer accuracy and informativeness when phrasing the titles of their studies. Crucially, authors might worry that mentioning a country name in the title may discourage potential readers, who might not consider research on this particular country to be relevant to their work.

Previous research has shown that the use of generic statements in abstracts, highlights, and titles can be misleading (12). Generic statements are prevalent in psychology research to a great extent because of the type of research questions that have historically interested psychologists. Despite research showing a cultural context's relevance to basic cognitive processes (27), the view that research in psychology should produce knowledge that is applicable to all of humanity is still widely held (14, 28). The implications are crucial because when research results are presented in the form of generic statements (e.g., "boys are different than girls") people consider them to be of greater validity compared to results presented using nongeneric expressions (e.g., "the boys in our sample displayed a different behavior compared to girls"). Moreover, because generic statements typically refer to social categories (e.g., "women," "adolescents," and "immigrants"), they can reinforce stereotypes and essentialist understandings of people's behaviors (29). This is neither desirable nor an accurate description of social phenomena (30, 31).

Titles with no geographical references offer less information than titles that name a country, city, region, or continent (provided that the study has an empirical component). In this sense, delocalized titles, such as "The proximate determinants of fertility" (32), are more generic than localized titles, such as "Modelling the proximate determinants of fertility for Brazil: The advent of competing preferences" (33).

**Measures and Hypotheses.** Our initial sample includes 1,256,554 social science English language publication records indexed by Scopus between 1996 and 2020. We focus on the 560,893 English language articles that mention at least one country name or demonym ("country name" hereafter) in their abstracts (9.4% of the total sample did not include an abstract, see *Materials and Methods*). This may exclude empirical articles that do not include a country name in the abstract. We expect this to render our results conservative, given the overrepresentation of studies on global North countries in the Scopus data (34, 35) and the fact that bias toward not mentioning WEIRD countries also exists at the abstract level (11).

We are interested in cross-national and cross-regional differences in the proportion of articles that include the country name of study in their title (herein "localized articles"). We call these "localization rates." A first set of analyses examines localization rates by country. The numerator of these rates is the number of mentions of a country name in titles, and the denominator is the number of mentions of the same country name in the corresponding abstracts. We expect to observe low-localization rates among studies of global North countries (center of knowledge production) and high-localization rates in studies about non-European and global South countries (periphery). We expect to observe relatively lower rates among "regional hegemons" (i.e., countries with the largest share of articles within regions) compared to their neighboring countries. We expect these regional gaps to be smaller than global North–South disparities.

A second set of analyses uses three specifications of a multivariate Poisson linear model to predict a binary variable, stating whether an article is localized (Y = 1) or not (Y = 0). In the first specification (MS-1), our main predictors of interest are the location of first authors' institutions and the geographical focus of the study. The second and third specifications (MS-2 and MS-3, respectively) include dummy variables for each of the top 10 most studied countries and the top one most studied country in each region. These three model specifications capture the direction and magnitude of cross-regional and cross-national gaps in the localization rate after accounting for basic titles and articles' characteristics, including the number of countries studied, number of authors, year of publication, title length, and subfield of study.

We group authors' locations and countries of study by UN Sustainable Goals Regions (UN-SDG). We combine the SDG categories "Europe and Northern America" and "Australia and New Zealand" into a single category. This region represents the global North and is equivalent to the "More developed regions" category in the human development index–based UN classification scheme (excluding Japan). There is a strong consistency between the global North/South and WEIRD/non-WEIRD categories. However, we consider the former to be more adequate to describe the potential causes and consequences of cross-national and cross-regional localization disparities. Whereas the WEIRD/non-WEIRD categories refer to samples' compositions (mostly in psychology studies), the global North and South refer to macrolevel patterns of economic, political, and developmental inequalities that have influenced the historical development of the social sciences as a whole. As a robustness check, we replicated our analyses using the UN-M49 standard country grouping and ran the analysis separately by academic discipline (*SI Appendix*, Figs. S4 and S5).

## Results

**Sustained Global Hegemony of Europe and North America over Time.** We find that, although articles about Europe and North America dominate our sample, they have the lowest localization rates. The vast majority of the research articles we study focuses on countries in the global North—more than 60% of the total articles mention a European or North American country in their abstract Fig. 1, *Left*, but the localization rate of these articles is the lowest, hovering around 0.42 for articles published between 1996 and 2020 Fig. 1, *Right*. This percentage contrasts with the localization rates in other regions, particularly in Eastern and South-Eastern Asia and Sub-Saharan Africa. Numerical dominance and low-localization rates signal the hegemonic position of research on European and North American countries in the Scopus data.

We visualize these two results on a world map, pooling the data for the entire period (1996 to 2020) and grouping countries based on the number of times they are mentioned in abstracts (Fig. 1, *Left*) and their corresponding localization rate (Fig. 1, *Right*). Groups are determined using Jenk's algorithm

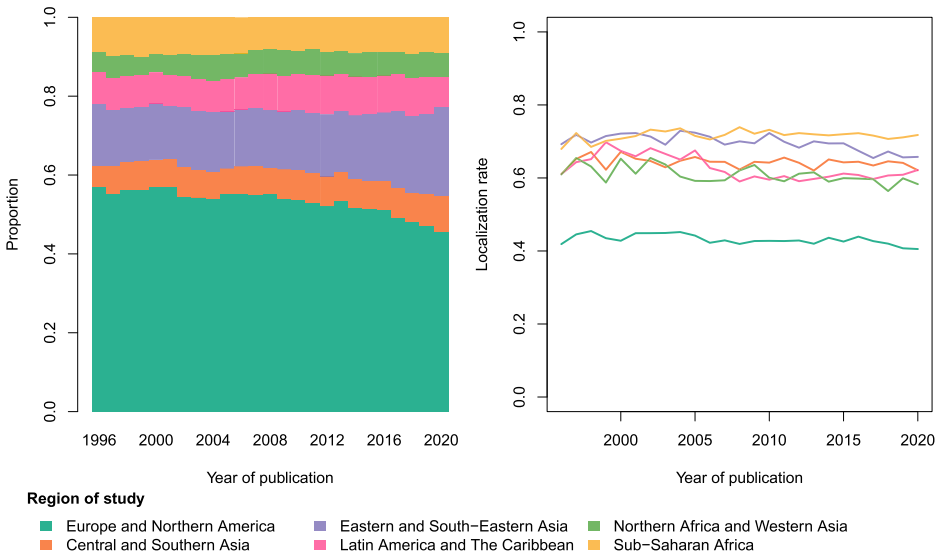


**Fig. 1.** Distribution of the region of focus of articles in the analytical sample (*Left*) and localization rate by region of study (*Right*), 1996 to 2020 ($n$ = 560,893). Note that the regional classification of countries is taken from the UN Sustainable Development Goals (https://unstats.un.org/sdgs/indicators/regional-groups/). We merge the category "Australia and New Zealand" into one category, "Europe and Northern America," and excluded papers about Oceania because of the small sample size ($n$ = 1,583).

(36). Fig. 2, *Left* highlights the fact that most studies in our sample studied either the United States, the United Kingdom, or China (the only three countries with more than 28,000 mentions in abstracts). Five out of the seven countries that are mentioned between 15,000 and 28,000 times in abstracts are in the global North: Australia, Canada, France, Germany, and Spain (the other two are India and Russia).

Fig. 2, *Top* suggests the existence of regional hegemons. These are countries that produce the most scholarship in a region and often have lower-localization rates than their neighbors. This is the case with Brazil in Latin America and the Caribbean, Israel in Northern Africa and Western Asia, Russia in Eastern Europe, India in South-Eastern Asia, and South Africa in Sub-Saharan Africa. China is particular because it has high-localization rates (0.69), despite being widely studied (7.1% of our sample are articles on China).

**Global Hegemony in a Multivariate Framework.** We test the robustness of these results in a multivariate framework. Initially, we want to assess the magnitude of the regional gaps shown in Figs. 1 and 2 after accounting for the titles' basic characteristics. Fig. 3 displays the regression coefficients obtained with our first model specification (MS-1). The reference categories are written in parentheses beneath the variables' names, and 95% CI are represented by red boxes.

According to this model, slightly fewer than half of the articles in the reference categories are localized [$\exp(-0.72) = 0.49$]. The coefficients for the variable "Region of study" are the largest in absolute terms, and they confirm the robustness of the regional gaps depicted in Fig. 1. All regression coefficients for the region of study are positive, statistically different from zero, and larger than 0.3. This means that, compared to articles about Europe and North America, research in other regions of the world is between $\exp(0.33) = 1.39$ and $\exp(0.48) = 1.62$ times more likely to be localized, all other things being equal. The two largest coefficients are those for Eastern and South-Eastern Asia (0.48) and Sub-Saharan Africa (0.47). The results are virtually identical when using a binomial (link function = logit) model and a normal (link function = identity) model. We report results from the Poisson (link function = log) model because of the simplicity for calculating relative risks [i.e., exp(coefficients)].

According to the MS-1, the location of the first author matters for the localization rate of articles but less than the relevance of the region of study. For example, when the first author is affiliated with a Sub-Saharan African institution (largest coefficient in absolute terms across this variable's categories), the localization rate is $\exp(0.08) = 1.08$ times higher than when the first author is affiliated with a European or North American institution.

Localization rates display a slight decrease over time: The coefficient for the last period indicates that, compared to articles published between 2005 and 2010, those published in the last 5 y are $\exp(-0.06) = 0.94$ times as likely to be localized. The results of the MS-1 also confirmed that the localization rate behaves as expected vis-a-vis basic characteristics, such as title length, number of countries studied, and number of authors. However, the explicative power of these covariates, all related to space constraints, is minimal compared to the variable for the region of study. Authors of articles examining two countries may want to emphasize the comparative nature of their research, whereas articles researching three or more countries may face space constraints. Space constraints seem to be more relevant than the emphasis on two-country comparisons, as the coefficient for articles with more than three countries is substantially larger, in absolute terms, than the one for articles studying two countries (−0.15 versus 0.09, respectively).

**Regional Hegemonies.** Center–periphery dynamics are replicated at the regional level between local hegemons and their neighbors. These disparities, however, are less pronounced than those between the global North and South. We first consider the differences between the top 10 most studied countries, according to the number of times they are mentioned in abstracts, and the rest of the countries (MS-2). Next, we consider the differences within regions between regional hegemons and their neighboring countries (MS-3).

Fig. 4 summarizes the results for these two specifications. Fig. 4, *Left* compares the regression coefficients for the top 10 countries of study, including the United States as the reference category and the rest of the countries grouped into regions. Fig. 4, *Right* compares regional hegemons and the rest of the countries against the United States. The 95% CI, depicted as

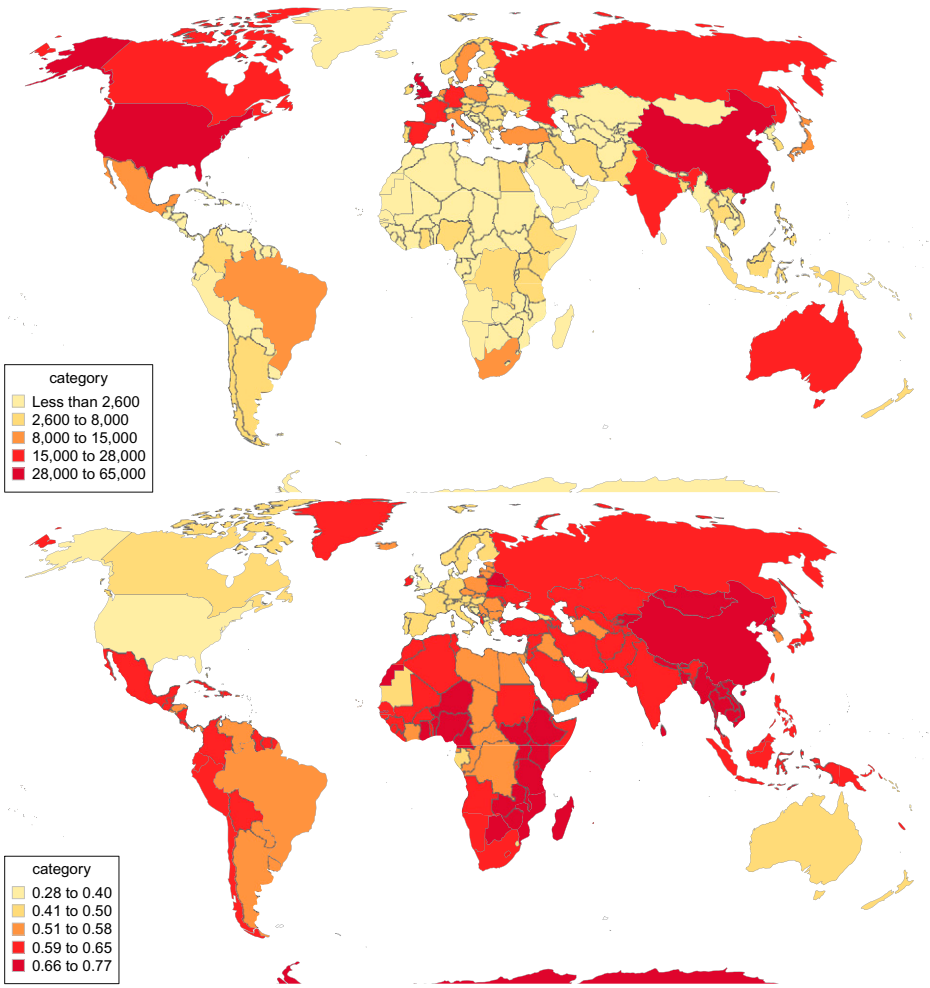


**Fig. 2.** Global disparities in knowledge production and papers' localization. (*Top*) Five-group classification of the total number of mentions of country names and demonyms in the analytical sample. (*Bottom*) Five-group classification of countries based on the localization rate of studies that mention a country name or demonym in their abstract. For both panels, the five categories are obtained by applying Jenk's algorithm to the country-level number of mentions and localization rates, respectively. Jenk's algorithm minimizes the variance within categories and maximizes the variance between categories.

boxes, allow us to compare localization rates between individual countries and regions.

The United States, the global hegemon, displays the lowest localization rates in both models. The intercept in both panels is −1.46, meaning that the localization rate for articles in the reference categories, including the United States as a country of study, is exp(−1.46) = 0.23. Less than one-fourth of articles about the United States (in all reference categories) are localized. This localization rate implies that the vast majority (two-thirds) of the US-focused research uses delocalized titles. The predicted localization rate for articles about the United States in the categories that are positively associated with the localization rate (e.g., very long titles) does not surpass 0.3 (i.e., at best, less than one-third of the US-focused articles are localized).

According to Fig. 4, *Left*, compared to the United States, the other top 10 countries of study display higher-localization rates, particularly those that do not belong to the global North (China, India, Russia, and Brazil). Instead, the lowest coefficient among top 10 countries pertains to the United Kingdom (0.54), implying that articles about the United Kingdom are 1.72 times more likely to be localized than articles about the United States. Even though this is a significant gap, the United Kingdom is still among the least localized of all countries [predicted localization rate according to this model = exp(−1.46 + 0.53) = 0.40]. Notably, gaps between global North and global South country persist even after accounting for the differences among top 10 producers. Indeed, the largest coefficient pertains to articles about China (top two country of study). This coefficient implies a relative risk of 3.67, meaning that, while two-thirds of US-focused articles are not localized, more than two-thirds of China-focused articles are. Russia is the only European country among the top 10 countries of study with considerably higher localization rate than the United States [predicted localization rate = exp(−1.46 + 1.08) = 0.68].

Consistent with these previous results, the coefficients for the top 10 European (except Russia) and other European countries are below one, indicating gaps in the localization rate below exp(1.0) = 2.7, compared to the United States. In contrast, the coefficients for the top 10 countries of study outside of Europe, North America, and Russia (i.e., China, India, and Brazil) are all above one. Finally, coefficients for regions are also large and significant (>1.11), implying gaps in the localization rate above exp(1.11) = 3.0, compared to the United States.

The phenomenon we document is more complex than an "Americanization" of academia. Fig. 4, *Right* further supports the notion that hegemony could be a driving factor of regional gaps in the localization rate. It is true that all regions and countries in the Fig. 4, *Right* display positive coefficients when compared to the United States. However, within all regions (except Eastern and South-Eastern Asia), the country with the largest share of

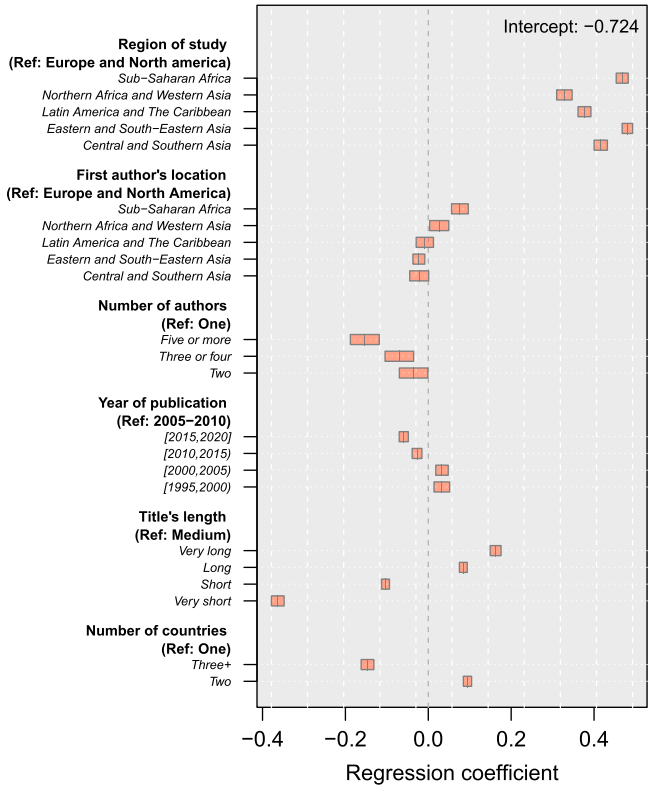


**Fig. 3.** Regression coefficients for a multivariate Poisson model predicting the localization (Y = 1) of papers. The rectangles comprise the 95% CI. CI are obtained by adding and subtracting the corresponding SE of each coefficient multiplied by the 97.5th percentile of a standard normal distribution. The category "Europe and North America" includes Australia and New Zealand. The deviance of this specification is 342,443, which is substantially lower than the degrees of freedom (i.e., 547,784), suggesting that the model fits the data adequately.

articles (i.e., the regional hegemon) displays a larger coefficient than all other countries in their region. Nonoverlapping CI suggest that the differences between regional hegemons and neighboring countries are statistically significant. In addition, the differences between the coefficients are also substantial, ranging from 0.13 between Latin American and Caribbean countries versus Brazil to 0.20 between Northern Africa, Western Asia, and Israel. These results suggest that hegemony operates on different levels, and it does not depend solely on numerical dominance.

**Heterogeneity and Consistency across Subfields.** It could be argued that differential disciplinary conventions affect the localization rates of articles. Social science subfields, such as development studies, demography, and political science, may be more likely to study geographically bound problems (e.g., population dynamics or pension systems) and therefore more likely to use a country's names in the title than other subfields, such as psychology. In particular, researchers claiming to study universal problems might not see the need to specify the geographic provenance of their sample (11, 37).

Our analysis of the localization rate across 23 subfields of the social sciences confirmed the existence of these disciplinary regularities. The localization rate ranges from below 0.4 (e.g., "Applied Psychology" and "General Psychology") to above 0.6 (e.g., "Development" and "Political Science and International Relations"). The geographical patterns described in the *Regional Hegemonies* section hold within each subdiscipline. We replicate the analyses displayed in Figs. 3 and 4 for 23 of our 27 subfields and find that our results are valid at the subfield level. Results are presented in *SI Appendix* for global hegemony (*SI Appendix*, Fig. S1 and MS-1), top 10 countries of study (*SI Appendix*, Fig. S2 and MS-2), and regional hegemony (*SI Appendix*, Fig. S3 and MS-3).

## Discussion

We show that global power imbalances are reflected in the temporal and spatial trends in the localization rates of more than half a million scientific articles across several social science subfields. Our main conclusion is that localization practices, the degree to which the regional focus of a study is declared in the title, follow a power-based logic between centers of academic production and the periphery. This conclusion is supported by four main findings. The gaps in localization rates between regions and countries of study are 1) enormous, 2) persistent over time, 3) robust across subfields, and 4) valid at the global and subregional level. We extend existing findings in psychology and development studies (6, 11, 14, 38) to several other subfields of the social sciences and provide a quantitative measure of its epistemological scope and global reach, including its prevalence across world subregions.

We propose that, at a global level, these center–periphery relations stem from the widespread view of the global North's superiority (i.e., Eurocentrism), which translates into an implicit belief that knowledge produced by authors in the global North about societies and individuals in the global North is more generalizable than knowledge produced by, in, and about peoples in the global South. It is worth emphasizing that Eurocentrism does not refer to a discrete geographical entity such as the European continent. Rather, it refers to the hegemony of former colonial powers and Western countries, including the United States (1, 3, 5, 6). A similar argument has been made about social theory, whereby theory produced outside of Europe and the United States is considered ethnotheory (e.g., Latin American dependency theory), whereas European social theory is simply labeled social theory (2, 10, 39). This conclusion is also in line with rising concerns about the lack of diversity and presumed universality of WEIRD samples in psychology studies (13, 14, 40). Our study shows that these conclusions apply to other fields in the social sciences. Researchers studying the global South are more prone to (consciously or unconsciously) declare their geographical focus, signaling by extension, the specificity and nonuniversality of their work. The epistemic hegemony of the global North, proxied by its localization rate, cannot be fully explained by the high share of studies about these countries and populations. The cases of China, India, and Russia, who, despite being large producers of knowledge and widely studied countries, display high-localization rates, illustrate this point.

Factors beyond these center–periphery unbalanced, epistemological relations may affect naming conventions. In cases where evidence from a given country is scarce in the English literature (e.g., if survey data has not been widely available), authors might choose to include the country name in the title to emphasize the "novelty" of the data. China is the best example of this trend. Alternatively, authors facing restrictive word limits or studies focusing on multiple countries might be less likely to mention country names in their title. While this is true, we showed that our main finding (the gap in localization rates between the global North and South) still holds after controlling for these factors in a multivariate regression setting. Disciplinary conventions and differences in socialization practices might also explain part of the difference. Whereas it is true that some disciplines are characterized by altogether lower localization rates (e.g., psychology as opposed to demography), we were able to replicate our main findings within each subfield.

We identify four main limitations of our study. First, our analysis is limited to articles published in English in journals indexed by Scopus. This limited our global reach, even though Scopus is

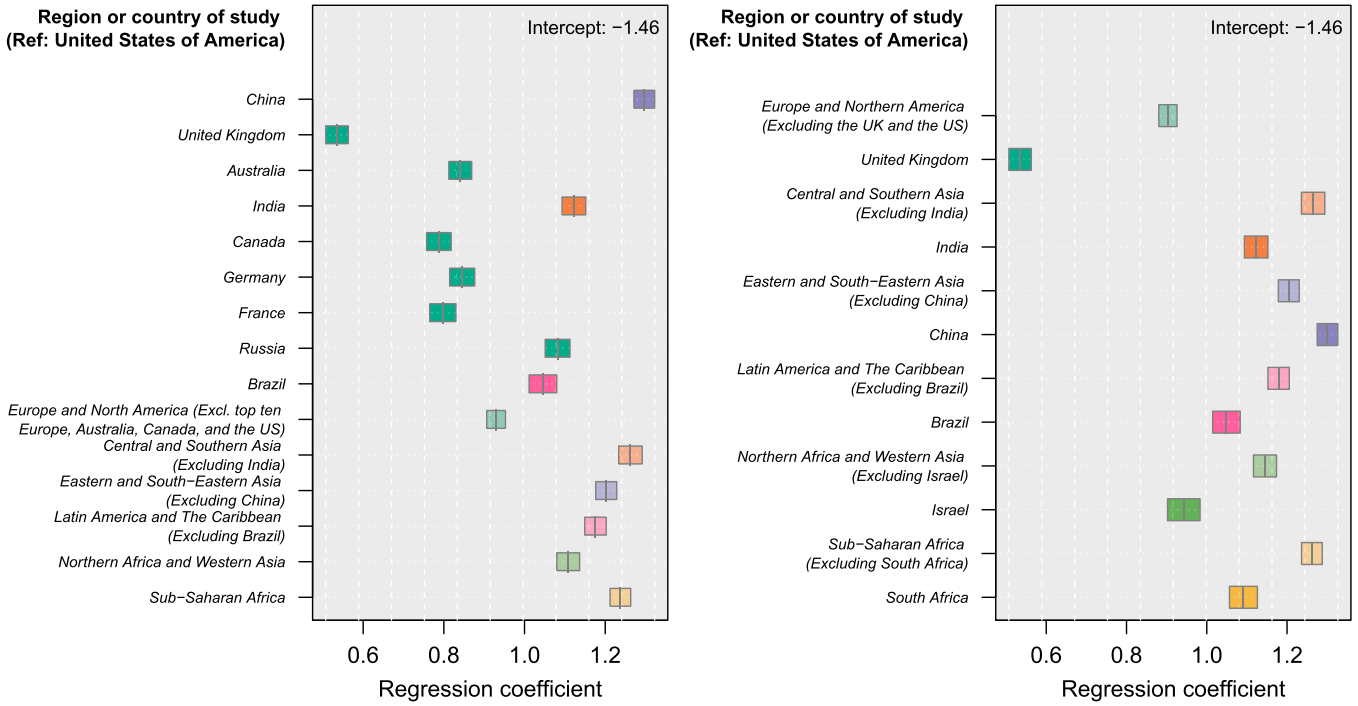


**Fig. 4.** (*Left*) Regression coefficients for the top 10 most studied countries and other countries, grouped by regions. (*Right*) Regression coefficients for the most studied country within region and other countries, grouped by region. Single-country coefficients are plotted with a darker shade than countries grouped in regions. Rectangles comprise the 95% CI. CI are obtained by adding and subtracting the corresponding SE of each coefficient multiplied by the 97.5th percentile of a standard normal distribution. The category "Europe and North America" includes Australia and New Zealand. The deviances for model specifications in *A* and *B* (332,926 and 333,182, respectively) are substantially lower than the degrees of freedom, indicating an adequate fit.

one of the most comprehensive archives of academic literature available to researchers (34). Second, we lack information on the peer review process, which makes it impossible for us to determine the degree to which country names were included or excluded from the titles throughout an article's life cycle. Third, we exclude articles without country names in their abstracts. It is unlikely that including these articles would invalidate our conclusions given that articles about global North countries are overrepresented in the Scopus data (35) and that the bias toward not mentioning WEIRD countries exists also at the abstract level (11). Finally, we do not evaluate the impact of localized versus nonlocalized articles (e.g., are nonlocalized articles more likely to be cited because they are regarded as more universal?) and the potential, country-level correlates of localization rates including, for instance, measures of countries' cultural distances (41). These are exciting prospects for future research.

We conclude by emphasizing the pitfalls of deeming evidence from the global North to be more general or universal than evidence from the global South. This practice can be misleading, if not outright harmful. Recognizing these gaps and quantifying their magnitude is a first step toward understanding and addressing the unseen disparities in global knowledge production. The lack of recognition of this bias among researchers requires further examination and may be explained by the concept of hypocognition [i.e., the notion that privilege is invisible to those who have and benefit from it (42)]. Indeed, in their attempt to achieve context- and value-free explanations of the social world, researchers might rely on their privilege to perpetuate the inequalities they study.

## Materials and Methods

**Bibliometric Data on Publications from Social Sciences and Humanities.** We use data from Scopus, a database of scientific publications, most of them written in English (34, 35). Our unit of observation is a publication record, which contains the title and abstract of a unique publication. We include empirical, peer-reviewed publications in English and exclude other types of publication (e.g., "chapter," "articles in press," "editorial," "book review," and "erratum").

We restrict our sample to publications that were coded as belonging to a selection of 27 subfields of the "Social Sciences and Humanities" category (*SI Appendix*). This selection criteria produces a database of 1,256,554 unique publication records published between 1995 and 2020. We excluded from this sample the 9.4% of the publication records (118,125) that were missing an abstract, leaving a final sample of 1,138,429 English language publication records with complete title and abstract data.

**Country Name Extraction.** We use regular expression matching algorithms to extract references to country names from our publication records. Our algorithms identified whether a given title or abstract includes one or more English country names, including abbreviations, alternative spellings (e.g., "U.S.A"), and demonyms (e.g., "Colombian"). Among the 1,138,429 publication records, 560,893 (49.2%) mention at least one country or country demonym in their abstract. This set of publications constitutes our analytical sample.

We assume that, when present, a country name or demonym refers to the geographical area of focus or the population under study [e.g., "Xi Jinping's 'major country diplomacy': The impacts of China's growing capacity" (43)]. We excluded country names mentioned in the context of a copyright statement in the abstract (e.g., "© Akadémiai Kiadó, Budapest, Hungary 2014"). We considered all countries with a country code according to the International Organization for Standardization list (ISO 3166–1), as implemented in the R "countrycode" package (44). Our algorithms did not capture subnational or supranational entities (e.g., cities or continent names) but excluding these had a negligible effect on our findings, as shown in the *Assessing the Accuracy of the Country Extraction Algorithms* section.

**Assessing the Accuracy of the Country Extraction Algorithms.** We hand coded a stratified random sample of the Scopus data ($n$ = 2,510). We use the geographical focus of an article (as identified by the algorithm) as the stratifying variable to make our results representative in terms of our main variable of interest. The data were coded independently by two research assistants who identified whether the research items' titles included a reference to a national entity (i.e., a country name or demonym), a subnational entity (e.g., a region or city within a country), or a supranational entity (e.g., a continent) that referred to the geographical context of the study.

Our algorithms successfully identified 94.3% (SE = 0.39) of the localized articles. The algorithms missed 5.2% (SE = 0.42) and 4.5% (SE = 0.39) cases in which only a subnational or supranational unit was mentioned, respectively (e.g., "Manchester" and "Sub-Saharan Africa"). Only 9.3% (s.e. = 0.53) of the articles were misclassified as being social science publications. This reaffirms our confidence in the performance of the chosen algorithmic approach.

We assessed the role of sample size on effect size detection by replicating our main analysis using 1,000 random samples of varying size from 800 to 20,000, extracted from the full Scopus data. We find that samples of sizes between 2,500 and 5,000 are sufficient to detect statistically significant differences for almost all regions except for Northern Africa, Western Asia, Latin America, and the Caribbean. Samples of 5,000 to 7,500 articles detect statistically significant differences virtually for all regions except Northern Africa and Western Asia. Statistically significant and consistent results for this latter region are observed for samples of 7,500 articles or more (*SI Appendix*, Fig. S6).

**Data Availability.** We accessed the Scopus data provided by Elsevier thanks to the Competence Centre for Bibliometrics (Kompetenzzentrum für Bibliometrie) (45). Our contractual agreement precludes us from redistributing the raw unprocessed data (i.e., the individual-level publication records). Users interested in accessing the microdata should contact Elsevier or the Kompetenzzentrum für Bibliometrie directly to enquire about the conditions of access and use (https://www.bibliometrie.info/index.php?id=kontakt). Script (in R language) and plain text .RDS format data have been deposited in Open Science Framework repository (DOI: 10.17605/OSF.IO/JT7AP). These materials constitute the evidential base for our claims and can be used to replicate our figures. All other study data are included in the article and/or *SI Appendix*.

**ACKNOWLEDGMENTS.** This study has received access to the bibliometric data through the project "Kompetenzzentrum Bibliometrie" and we acknowledge their funder Bundesministerium für Bildung und Forschung (funding identification number 01PQ17001). We thank our colleagues Tom Theile and Aliakbar Akbaritabar for the technical support and comments, respectively. We thank our colleagues at the Max Planck Institute for Demographic Research who, always eager to engage in critical dialogue, helped shape this paper. We thank Research Assistants Saskia Morwinsky and Antonia Busch for their manual coding work. This work also benefited from discussions with Dr. Seye Abimbola.

1. G. K. Bhambra, J. Holmwood, *Colonialism and Modern Social Theory* (Polity Press, 2021).
2. R. Grosfoguel, Racismo/sexismo epistémico, universidades occidentalizadas y los cuatro genocidios/epistemicidios del largo siglo XVI. *Tabula Rasa* **19**, 31–58 (2013).
3. A. Quijano, Coloniality of power and Eurocentrism in Latin America. *Int. Sociol.* **15**, 215–232 (2000).
4. I. Wallerstein, A world-system perspective on the social sciences. 1974. *Br. J. Sociol.* **61**, 167–176 (2010).
5. J. Go, Race, empire, and epistemic exclusion: Or the structures of sociological thought. *Sociol. Theory* **38**, 79–100 (2020).
6. W. D. Mignolo, Spirit out of bounds returns to the East: The closing of the social sciences and the opening of independent thoughts. *Curr. Sociol.* **62**, 584–602 (2014).
7. F. Coronil, Beyond occidentalism: Toward nonimperial geohistorical categories. *Cult. Anthropol.* **11**, 51–87 (1996).
8. M. Krause, 'Western hegemony' in the social sciences: Fields and model systems. *Sociol. Rev. [Monogr]* **64**, 194–211 (2016).
9. S. Abimbola, M. Pai, Will global health survive its decolonisation? *Lancet* **396**, 1627–1628 (2020).
10. M. Boatcă, Thinking Europe otherwise: Lessons from the Caribbean. *Curr. Sociol.* **69**, 389–414 (2021).
11. R. Kahalon, V. Klein, I. Ksenofontov, J. Ullrich, S. C. Wright, Mentioning the sample's country in the article's title leads to bias in research evaluation. *Soc. Psychol. Personal. Sci.*, 194855062110240 (2021).
12. J. M. DeJesus, M. A. Callanan, G. Solis, S. A. Gelman, Generic language in scientific communication. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 18370–18377 (2019).
13. J. Henrich, S. J. Heine, A. Norenzayan, The weirdest people in the world? *Behav. Brain Sci.* **33**, 61–83 (2010).
14. B. K. Cheon, I. Melani, Y.-y. Hong, How USA-centric is psychology? An archival study of implicit assumptions of generalizability of findings to human nature based on origins of study samples. *Soc. Psychol. Personal. Sci.* **11**, 928–937 (2020).
15. B. de Sousa Santos, *Epistemologies of the South: Justice Against Epistemicide* (Routledge, 2016).
16. J. L. Rosner, Reflections of science as a product. *Nature* **345**, 108 (1990).
17. C. E. Paiva, J. P. S. N. Lima, B. S. R. Paiva, Articles with short titles describing the results are cited more often. *Clinics (São Paulo)* **67**, 509–513 (2012).
18. S. Subotic, B. Mukherjee, Short and amusing: The relationship between title characteristics, downloads, and citations in psychology articles. *J. Inf. Sci.* **40**, 115–124 (2014).
19. M. Haggan, Research paper titles in literature, linguistics and science: Dimensions of attraction. *J. Pragmat.* **36**, 293–317 (2004).
20. I. Sagi, E. Yechiam, Amusing titles in scientific journals and article citation. *J. Inf. Sci.* **34**, 680–687 (2008).
21. S. W. Cheng, C.-W. Kuo, C.-H. Kuo, Research article titles in applied linguistics. *J. Acad. Lang. Learn.* **6**, A1–A14 (2012).
22. J. Hartley, To attract or to inform: What are titles for? *J. Tech. Writ. Commun.* **35**, 203–213 (2005).
23. J. Moody, Trends in sociology titles. *Am. Sociol.* **37**, 77–80 (2006).
24. A. Rath, Dual function of first position nominal groups in research article titles: Describing methods and structuring summary. *J. Appl. Lang. Stud.* **1**, 5–23 (2010).
25. K. Hyland, What do they mean? Questions in academic writing. *Text Talk* **22**, 529–557 (2002).
26. S. Abimbola, The foreign gaze: Authorship in academic global health. *BMJ Glob. Health* **4**, e002068 (2019).
27. R. E. Nisbett, T. Masuda, Culture and point of view. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 11163–11170 (2003).
28. P. Rozin, Domain denigration and process preference in academic psychology. *Perspect. Psychol. Sci.* **1**, 365–376 (2006).
29. S. A. Gelman, S. O. Roberts, How language shapes the cultural inheritance of categories. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 7900–7907 (2017).
30. I. Hacking, "The looping effects of human kinds" in *Causal Cognition*, D. Sperber, D. Premack, A. J. Premack, Eds. (Oxford University Press, 1996), pp. 351–383.
31. D. Wodak, S.-J. Leslie, M. Rhodes, What a loaded generalization: Generics and social cognition. *Philos. Compass* **10**, 625–635 (2015).
32. J. Bongaarts, The proximate determinants of fertility. *Technol. Soc.* **9**, 243–260 (1987).
33. R. Z. Coutinho, A. B. Golgher, Modelling the proximate determinants of fertility for Brazil: The advent of competing preferences. *Rev. Bras. Estud. Popul.* **35**, 1–28 (2018).
34. M. E. Falagas, E. I. Pitsouni, G. A. Malietzis, G. Pappas, Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. *FASEB J.* **22**, 338–342 (2008).
35. P. Mongeon, A. Paul-Hus, The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics* **106**, 213–228 (2016).
36. G. Jenks, The data model concept in statistical mapping. *Int. Yearb. Cartogr.* **7**, 186–190 (1967).
37. A. Buyalskaya, M. Gallo, C. F. Camerer, The golden age of social science. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2002923118 (2021).
38. J. Brohman, Universalism, Eurocentrism, and ideological bias in development studies: From modernisation to neoliberalism. *Third World Q.* **16**, 121–140 (1995).
39. C. Alvares, A critique of Eurocentric social science and the question of alternatives. *Econ. Polit. Wkly.* **46**, 72–81 (2011).
40. J. J. Arnett, The neglected 95%: Why American psychology needs to become less American. *Am. Psychol.* **63**, 602–614 (2008).
41. M. Muthukrishna *et al.*, Beyond Western, educated, industrial, rich, and democratic (WEIRD) psychology: Measuring and mapping scales of cultural and psychological distance. *Psychol. Sci.* **31**, 678–701 (2020).
42. K. Wu, D. Dunning, "Hypocognition and the invisibility of social privilege" in *Advances in Group Processes*, S. R. Thye, E. J. Lawler, Eds. (Emerald Publishing Limited, 2020), pp. 1–23.
43. Z. Lin, Xi Jinping's 'major country diplomacy': The impacts of China's growing capacity. *J. Contemp. China* **28**, 31–46 (2019).
44. V. Arel-Bundock, N. Enevoldsen, C. Yetman, countrycode: An R package to convert country names and country codes. *J. Open Source Softw.* **3**, 848 (2018).
45. Kompetenzzentrums für Bibliometrie, Competence Centre for Bibliometrics. Bibliometrics Database (WoS & Scopus). https://www.bibliometrie.info/index.php?id=infrastruktur. Accessed 8 October 2020.