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

Andrés F. Castro Torres ι castro@demogr.mpg.de
Diego Alburez-Gutierrez ι alburezgutierrez@demogr.mpg.de

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

Abstract

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 (i) differences between world regions, (ii) differences within world regions, and (iii) patterns in 23 social science subfields. These gaps are large and consistent, yet paper naming conventions are merely the "tip of the iceberg" of the imbalances in knowledge production between the global North and South.

Introduction

The development of institutionalized social sciences, as we know them today, was closely linked to the European colonial projects that spanned from the 15th to the 20th century (Bhambra and Holmwood 2021; Grosfoguel 2013). As a consequence, the production of academic knowledge is embedded in power structures that can be characterized in terms of center-periphery relations (Quijano 2000; Wallerstein 1976). The imprint of these centuries of economic and political subordination on knowledge production has neither been fully understood, nor overcome (Go 2020). 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 (Mignolo 2014).

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, rising human development, liberal democracy, market/trade integration, and globalization, being considered the "default" cases towards which other nations and societies ought to converge in the mid- or long term (Coronil 1996; Krause 2016). Multiple calls to decolonize university curriculums and research projects have been made in sociology (Bhambra 2014), global health (Abimbola and Pai 2020), and economics (Kvangraven and Kesar 2021), to mention a few cases; 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 paper examines one aspect in which the Eurocentric view of empirical social science research manifests itself, namely the degree to which papers 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. Readers may tend to interpret these "delocalized" titles as describing universal processes, as is the case with generic statements (DeJesus et al. 2019; Kahalon

et al. 2021). Despite the growing awareness about the pitfalls of such a perspective and the multiple local epistemological alternatives developed by scholars in the global South, the production of mainstream narratives about contemporary social processes remains largely Eurocentric due to the economic, political, and cultural hegemony of the global North (Quijano 2000; de Sousa Santos 2016).

We use the categories global North and global South to group countries according to their hegemonic positions in social science production. The former category includes countries in Europe, North America, Australia, and New Zealand, but excludes Russia. The latter includes all other world countries. This categorization aims to reflect the histories of colonialism and neocolonialism, as well as the geopolitical oppositions and socioeconomic disparities that emerged after the Cold War (Chant and Mcllwaine 2009).

Our study makes three contributions. First, we provide a novel bibliometric measure of a paper's "localization," which can be used for comparative analysis. Second, we interpret the spatial and temporal trends of this measure in terms of the epistemological dominance of global North social sciences. This interpretation extends previous theoretical research on the influence of the Eurocentric view to the everyday practices of social sciences worldwide. Third, our cross-national evidence provides a macro-level perspective on knowledge production inequalities. We show that peripheral locations are more likely to declare the specificity (lack of generalizability) of their results compared to centers of academic production. We document this trend globally and regionally, analyzing more than half a million articles from 27 subfields of the Social Sciences and Humanities, published from 1995 to 2020 and indexed by Scopus.

Background

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. Because titles are an up-front statement of a paper's significance,

multiple factors affect their phrasing, and there is evidence to suggest that authors choose the wording strategically (Rosner 1990). Short, generic, and amusing titles are more likely to capture researchers' attention at first glance (Subotic and Mukherjee 2014). However, long, specific, and detailed titles are informative and may therefore be preferred if authors want to convey specific ideas and attract specialized audiences (Haggan 2004; Paiva, da Silveira Nogueira Lima, and Ribeiro Paiva 2012).

Several studies have documented the variability in title characteristics and how they affect an article's readership (Sagi and Yechiam 2008). Previous studies have examined titles' length and syntactic structures (Cheng, Kuo, and Kuo 2012; Hartley 2005; Moody 2006; Rath 2010), the use of question marks and semicolons (Hyland 2002), the prevalence of generic expressions (DeJesus et al. 2019), and the inclusion of country names (Kahalon et al. 2021). Our focus is on exploring, for the first time, the propensity to include geographical references in titles across geographies and sub-disciplines of the social sciences.

Geographical references in titles and generic language. We distinguish "localization" from "delocalization" as two distinct strategies for the naming of papers. In the former, authors include a concrete geographical reference to the context or population being studied. In the latter, this concrete reference is omitted. The regular use of either strategy 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. The deliberate choice to omit a concrete geographical reference to the context of study may reflect the authors' "pose and gaze" (Abimbola 2019). Regardless of authorial intent, the lack of a concrete geographical reference in a research article's title may mislead readers by detaching results from the specific context where they were observed, implying universal conclusions.

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 (DeJesus et al. 2019). Generic statements are prevalent in psychology research, to a great extent due to the type of research questions that have historically interested psychologists. Despite research showing a cultural context's relevance to basic cognitive processes (Nisbett and Masuda 2003), the view that research in psychology should produce knowledge that is applicable to all of humanity, is still widely held (Rozin 2006). 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 non-generic 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," "immigrants"), they can reinforce stereotypes and essentialist understandings of people's behaviors (Gelman and Roberts 2017). This is neither desirable nor an accurate description of social phenomena (Hacking 1996; Wodak, Leslie, and Rhodes 2015).

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" (Bongaarts 1987), are more generic than localized titles, such as "Modelling the proximate determinants of fertility for Brazil: the advent of competing preferences" (Coutinho and Golgher 2018). The inclusion of a concrete geographical reference is an important part of a title's phrasing, as it contextualizes the results: it indicates the spatial context that the results refer to.

Measures and hypotheses. Our initial sample includes 1,256,554 social science English-language *publication records* indexed by Scopus between 1995 and 2020. We focus on the 560,893 publications 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). We are interested in cross-national and cross-regional differences in the proportion of articles that include the name of the country studied or the country's demonym in their title (herein "localized

papers"). We call these "*localization rates*" to emphasize the fact that we are interested in articles that mention a country name or demonym in their corresponding abstracts (i.e., articles at risk of being localized in their titles).

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. Our overarching hypothesis is that countries in hegemonic positions in terms of social science production will display lower *localization rates* compared to peripheral countries. Evidence from and about enclaves that dominate social science production are implicitly taken as more generalizable than evidence from elsewhere. Specifically, we expect to observe low *localization rates* among studies of global North countries and high *localization rates* in studies about non-European and global South countries. We expect to observe relatively lower rates among "regional hegemons," i.e., countries with hegemonic positions in social science production within regions in the global South, compared to their neighboring countries. Center-periphery relations highlight the centrality of hegemony as the logic behind aggregated trends in social sciences production, though 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 (six regional categories adapted from the United Nations Sustainable Development Goals; details in the Supplementary Materials). 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 within the six world regions that we study. These analyses 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 number of countries studied, number of authors, year of publication, title length, and subfield of study.

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 (Figure 1, panel A), but the *localization rate* of these articles is the lowest, hovering around 42% for papers published between 1996 and 2020 (panel B). This percentage contrasts with the *localization rates* in other regions, particularly in Eastern and South-Eastern Asia and Sub-Saharan Africa. The ratio of the *localization rates* between these regions and Europe and North America ranges from 1.5 to 1.8 throughout the period of analysis. The other regions also display persistently higher *localization rates* compared to Europe and North America, with ratios around 1.4. Numerical dominance and low *localization rates* signal the hegemonic position of research on European and North American countries in the Scopus data.

Figure 1 – Distribution of the region of focus of articles in the analytical sample (Panel A) and *localization rate* by region of study (Panel B), 1996-2020 (n = 560,893).
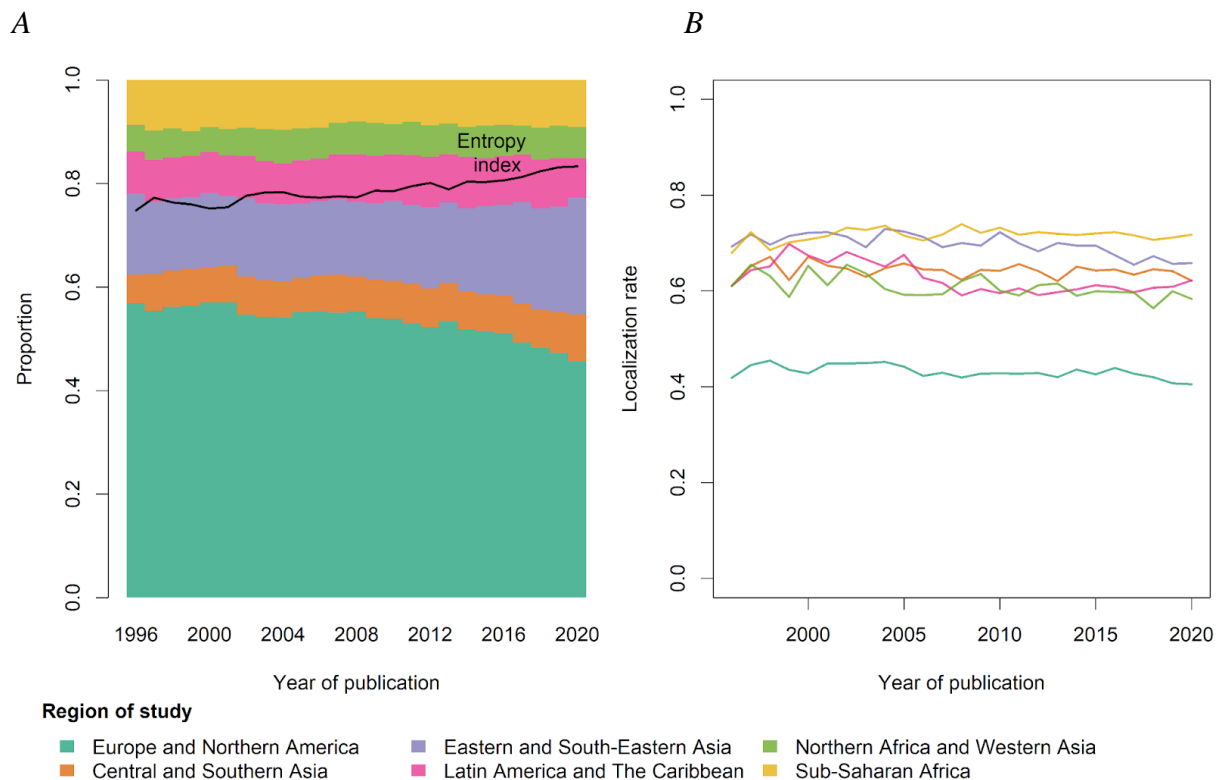
Since the 2010s, there has been a slight increase in the proportion of papers about Eastern and South-Eastern Asian countries, mostly driven by studies about China. As a consequence, the proportion of articles about European and North American countries declined 10 p.p. between 1996 and 2020. By contrast, the share of research articles about all other regions has remained relatively stable. The yearly entropy index - a measure of uncertainty in the distribution of a categorical variable - for the distribution of articles across regions displays a very small positive slope over the period of analysis, but the literature in our sample continues to be dominated by studies on the global North.

We visualize these two results on a world map, pooling the data for the entire period (1996-2020) and grouping countries based on the number of times they are mentioned in abstracts (panel A) and their corresponding *localization rate* (panel B). Groups are determined using Jenk's algorithm (Jenks 1967). Panel A of Figure 2 highlights the fact that most studies in our sample studied either the US, the UK, 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).

The joint examination of panels A and B suggests the existence of regional hegemons. 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).

Figure 2 - Global disparities in knowledge production and papers' localization. *Panel A:* Five-group classification of the total number of mentions of country names and demonyms in the analytical sample. *Panel B:* Five-group classification of countries based on the *localization rate* of studies that mention a country name or demonym in their abstract.

A
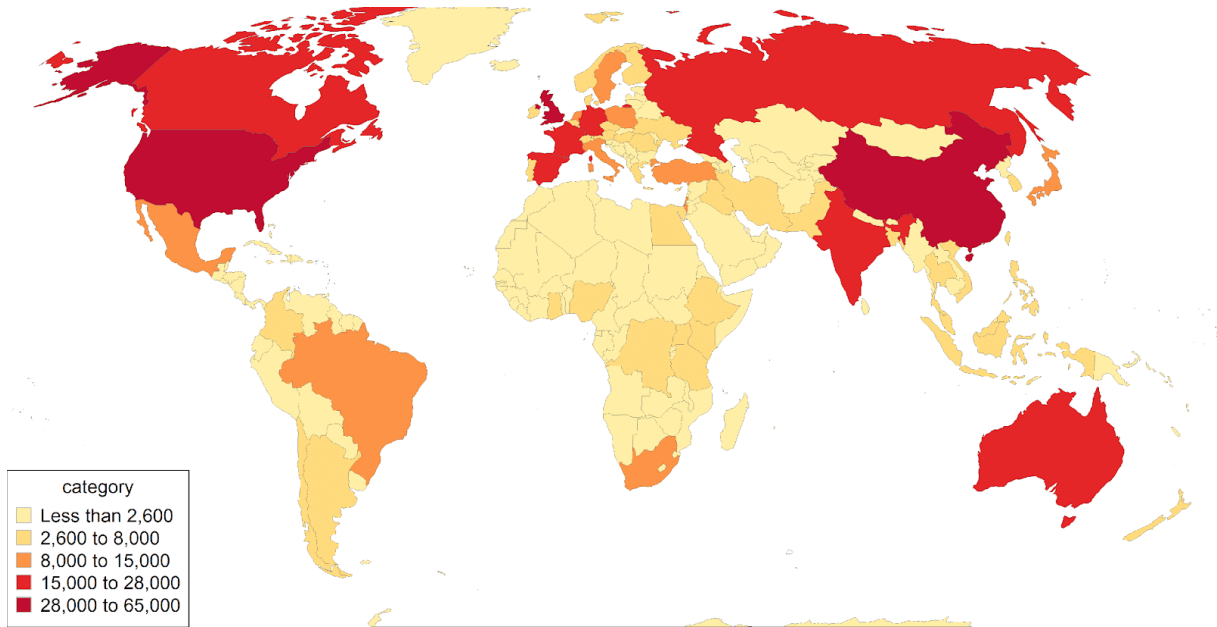
B
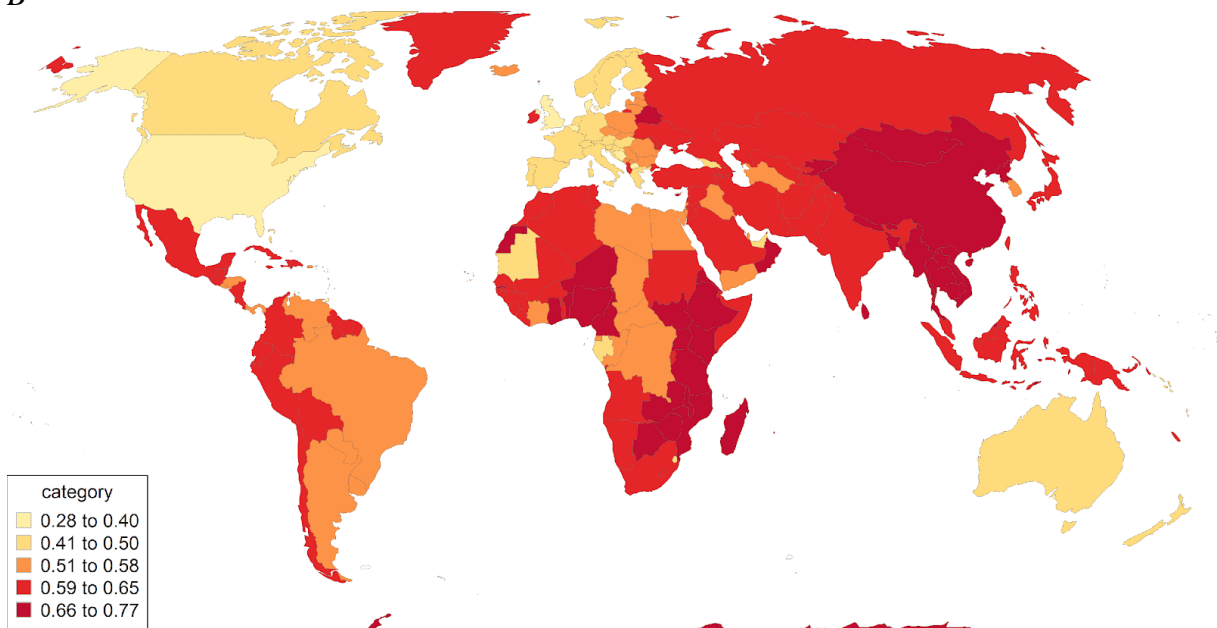
Note: For both panels, the five categories are obtained by applying Jenk's algorithm to the country-level number of mentions and localization rates, respectively. Jenk's algorithm minimizes the variance within categories and maximizes the variance between categories.

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 Figures 1 and 2 after accounting for the titles' basic characteristics. Figure 3 displays the regression coefficients of a Poisson model (link function = logarithm) that predicts the localization of articles (Y = 1 if a country is mentioned in the title, Y = 0 otherwise) based on the two geographical variables of interest and the four controls (MS-1). The reference categories are written in parentheses beneath the variables' names, and 95% confidence intervals are represented by red boxes.

Figure 3 - Regression coefficients for a multivariate Poisson model predicting the localization (Y = 1) of papers. The rectangles comprise the 95% confidence intervals.
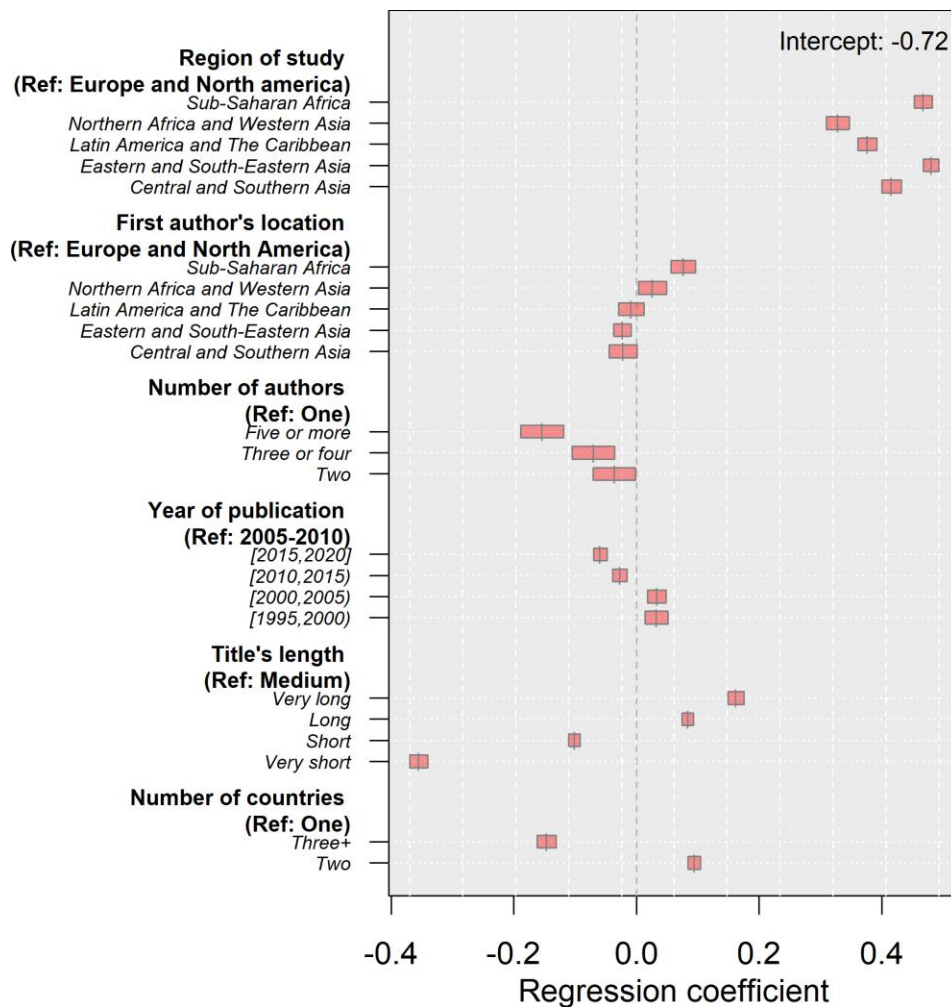
Note: Confidence intervals are obtained by multiplying the standard error of each coefficient by the 97.5[th] percentile of a normal distribution with mean zero and standard deviation one. Results are robust to the use of a normal model with the identity function as link ($f(x) = x$). The category "Europe and North America" includes Australia and New Zealand.

The intercept indicates that the predicted *localization rate* for articles in the reference categories is exp(-0.72) = 0.49, meaning that, according to this model, slightly fewer than half of the papers in the reference categories are localized. 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 Figure 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 normal model (link function=identity).

According to the MS-1, the location of the first author matters for the *localization rate* of papers, 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.

We find no clear temporal trend. Even though *localization rates* decreased over time, they did so very slowly (e.g., the coefficient for the last period indicates that, compared to articles published between 2005 and 2010, those published in the last five years 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).

Figure 4 summarizes the results for these two specifications. The left panel compares the regression coefficients for the top 10 countries of study, including the US as the reference category, and the rest of the countries grouped into regions. The right panel compares regional hegemons and the rest of the countries against the US. The 95% confidence intervals, depicted as boxes, allow us to compare *localization rates* between individual countries and regions.

Figure 4. *Panel A*: Regression coefficients for the top 10 most studied countries and other countries, grouped by regions. *Panel B*: 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% confidence intervals.
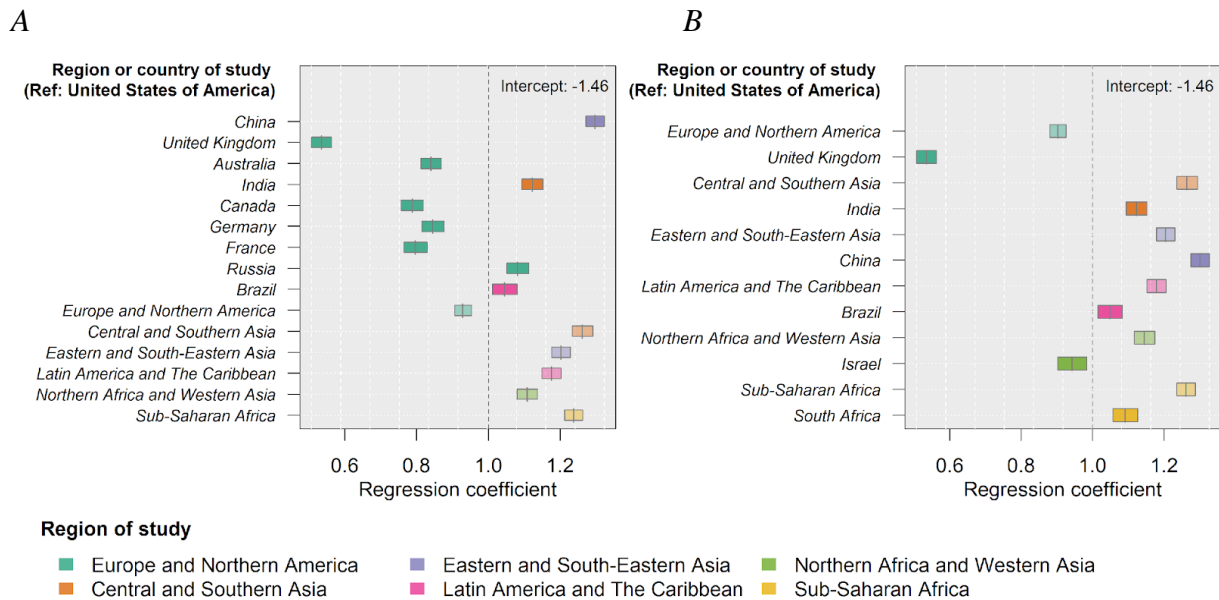
Note: Confidence intervals are obtained by multiplying the standard error of each coefficient by the 97.5th percentile of a normal distribution with mean zero and standard deviation one. Results are robust to the use of a normal model with the identity function as link (*f(x) = x*). The category "Europe and North America" includes Australia and New Zealand.

The intercept in both panels is -1.46, meaning that the *localization rate* for articles in the reference categories, including the US as a country of study, is exp(-1.46) = 0.23. Less than one-fourth of papers about the US (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 papers about the US 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 papers are localized.

According to panel A, compared to the US, the other top 10 countries of study display higher *localization rates*. The lowest coefficient among these countries pertains to the UK (0.54), implying that papers about the UK are 1.72 times more likely to be localized than papers about the US. This is a very significant gap, and it suggests that, even at the top of numerical dominance, the lack of localization is not explained by the use of the English language alone. By contrast, the largest coefficient pertains to articles about China. This coefficient implies a relative risk of 3.67, meaning that, while two-thirds of US-focused papers are not localized, more than two-thirds of China-focused articles are.

Consistent with these previous results, the coefficients for the top 10 European and other European countries (except Russia) are below one, indicating gaps in the *localization rate* below exp(1.0) = 2.7, compared to the US. In contrast, the coefficients for the top 10 countries of study outside of Europe and 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 US.

Panel B in Figure 4 further supports the notion that hegemony could be a driving factor of regional gaps in the *localization rate*. When compared to the US, all regions and countries in the right panel display positive coefficients, i.e., higher *localization rates*. Moreover, within all regions (except Eastern and South-Eastern Asia), the country with the largest share of articles (the most studied one, or the regional hegemon) displays a larger coefficient than all other countries in their region. Non-overlapping confidence intervals 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 vs. Brazil, to 0.20 between Northern Africa and 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 rate*s 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 resist the call to specify the geographic provenance of their sample (Buyalskaya, Gallo, and Camerer 2021; Kahalon et al. 2021).

Our analysis of the *localization rate* across 27 subfields of the social sciences confirmed the existence of these disciplinary regularities. The *localization rate* ranges from 0.33 among papers classified as "Applied Psychology" (n = 3,422) to 0.66 among papers classified as "Development" (n = 69,724). Articles on "Political Science and International Relations" (n = 52,961) and "Demography" (n = 21,328) display *localization rates* very similar to "Development," namely 0.64, whereas articles on "Psychology (miscellaneous)" (n = 247), "Experimental and Cognitive Psychology" (n = 1,416), and "General Psychology" (n = 3,771) display *localization rates* below 0.4 (see Material and Methods for further details).

Nevertheless, the geographical patterns described above hold within each subdiscipline. We replicate the analyses displayed in Figures 3 and 4 for 23 of our 27 subfields and find that our results are valid at the subfield level. Due to the small sample size, we cannot replicate these analyses for "Clinical psychology" (n = 2,219), "Industrial relations" (n = 1,181), "Experimental and cognitive psychology" (n = 1,416), and "Psychology (miscellaneous)" (n = 247). For all other subfields, results are presented in the Supplementary Materials for global hegemony (Figure A1, MS-1), top 10 countries of study (Figure A2, MS-2), and regional hegemony (Figure A3, 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 papers 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 sub-regional level. Empirical research in development studies and psychology and theoretical essays on the social sciences were instrumental in revealing these patterns (Brohman 1995; Kahalon et al. 2021; Mignolo 2014); our work has extended this finding to several other subfields of the social sciences and provides 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, 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. A similar argument has been made about social theory, whereby theory produced outside of Europe is considered ethno-theory (e.g., Latin American dependency theory), whereas European social theory is simply labeled social theory (Alvares 2011; Boatcă 2020; Grosfoguel 2013). This is in line with rising concerns about the lack of diversity in the so-called WEIRD samples in psychology studies (Arnett 2008). Researchers studying the global South are more prone to (consciously or unconsciously) declare their geographical focus, signaling, by extension, the specificity and non-universality 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. Finally, authors might be required by editors or reviewers to mention or remove a country name from the title. We could not evaluate this empirically, given data limitations. However, even if this were the case, it supports our argument that the claim to universality goes unchallenged more often in studies on populations from the global North.

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 one of the most comprehensive archives of academic literature available to researchers (Falagas et al. 2008). 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 focus on papers that mention at least one country in their abstract. This ignores national-level studies that do not disclose their geographic focus in the abstract. Finally, we do not evaluate the impact of localized versus non-localized papers (e.g., are non-localized papers more likely to be cited because they are regarded as more universal?). This is an exciting prospect 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, in particular when evidence-based policy is involved. Recognizing these gaps and quantifying their magnitude is a first step towards understanding and addressing the unseen

disparities in global knowledge production. We are not in a position to prescribe solutions, such as forcing all social science research products to be explicitly localized (or explicitly delocalized) in their titles; this work should be read as an invitation to reflect on our own disciplinary practices and take the actions we consider pertinent to avoid the perpetuation of epistemic inequalities. 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 (Wu and Dunning 2020). 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 (Falagas et al. 2008; Mongeon and Paul-Hus 2016). 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," "erratum").

We restrict our sample to publications that were coded as belonging to a selection of 27 subfields of the "Social Sciences & Humanities" category: Religious Studies, Industrial Relations, General Economics, Econometrics and Finance Economics, Econometrics and Finance (miscellaneous), Economics and Econometrics, General Psychology, Psychology (miscellaneous), Applied Psychology, Clinical Psychology, Developmental and Educational Psychology, Experimental and Cognitive Psychology, Social Psychology, General Social Sciences, Social Sciences (miscellaneous), Development, Education, Geography, Planning and Development Health (social science), Sociology and Political Science, Anthropology, Cultural Studies, Demography, Gender Studies, Life-span and Life-course Studies, Political Science and International Relations, Public Administration, and Urban Studies. 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. For example, in the subfield of development, our algorithm flags the following title as localized: "Xi Jinping's 'major country diplomacy': The impacts of China's growing capacity" (Lin 2019). In the same subfield, our algorithm classifies the following title as delocalized: "Green areas, the most significant indicator of the sustainability of cities: Research on their utility for urban planning" (Gómez et al. 2011). According to its abstract, the latter paper uses data on the city of Valencia in Spain. Examples of localized and delocalized titles from applied psychology include "Screening for mood difficulties in men in Italy and Australia using the Edinburgh Postnatal Depression Scale and the Matthey Generic Mood Questionnaire" (Matthey and Della Vedova 2020) and "Financial management, coping and debt in households under financial strain" (Walker 1996). This latter paper uses data collected at health facility centers in England. 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 an ISO 3166-1 code, as implemented in the R "countrycode" package (Arel-Bundock, Enevoldsen, and Yetman 2018). Our algorithms did not capture sub-national or supra-national entities (e.g. cities or continent names), but excluding these had a negligible effect on our findings, as shown in the following section.

Assessing the accuracy of the country extraction algorithms. In order to evaluate the accuracy of our algorithmic approach, we hand-coded a stratified random sample of the Scopus data (n = 2,510). We use the geographical focus of a paper (as identified by the algorithm) as the stratifying

variable to make our results representative in terms of our main variable of interest. The data was 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 sub-national entity (e.g., a region or city within a country), or a supra-national entity (e.g., a continent) that referred to the geographical context of the study.

This analysis showed that our algorithms successfully identified 94.3% (standard error = 0.39) of the localized papers. As expected, the algorithms missed 5.2% (s.e. = 0.42) and 4.5% (s.e. = 0.39) cases where only a sub-national or supra-national unit was mentioned, respectively (e.g., "Manchester" and "Sub-Saharan Africa"). The research assistants were also asked to confirm that the papers in our sample were indeed social science papers. The analysis showed that only 9.3% (s.e. = 0.53) of the papers was misclassified as being social science publications. We are thus confident that our algorithmic approach was able to locate country names, or their derivations, and that the proportions of missed articles due to sub- and supra-national levels are of no consequence to our results.

Multivariate analysis, variables of interest, and controls. We used three specifications of a multivariate Poisson linear model to predict a binary variable stating whether a publication record 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 study's geographical focus. We assign countries to regions using a modified version of the United Nations Sustainable Development Goals regions (Europe and North America, Central and Southern Asia, Eastern and South-Eastern Asia, Latin America and the Caribbean, Northern Africa and Western Asia, and Sub-Saharan Africa). Our regional grouping differs from the United Nations in that we add Australia and New Zealand to "Europe and North America." The region "Europe and North America" groups all countries of the global North, creating a useful reference category for model estimation. This is the group of countries against which we want to compare all the rest. All other categories include countries in the global South.

For the second and third model specifications (MS-2 and MS-3), we separate the top 10 countries with the highest share of articles globally (i.e., the most studied countries) and the top one country

with the highest share of research regionally (regional hegemons), respectively. The reference category in these two model specifications is the US, the most researched country in the sample.

Our control variables include the number of countries mentioned in the abstract (one, two, three or more), the length of the title (very short, short, medium, long, and very long), the year of publication grouped into five-year periods (1995-1999, 2000-2004, 2005-2009, 2010-2014, and 2015-2020), the number of authors (one, two, three or four, five or more). As a robustness check, we run the analysis in our two hand-coded random sub-samples of the data. Results are highly consistent. Finally, we replicate the analyses for each subfield.

References

Abimbola, Seye. 2019. "The Foreign Gaze: Authorship in Academic Global Health." *BMJ Global Health* 4(5):e002068. doi: 10.1136/bmjgh-2019-002068.

Abimbola, Seye, and Madhukar Pai. 2020. "Will Global Health Survive Its Decolonisation?" *The Lancet* 396(10263):1627–28. doi: 10.1016/S0140-6736(20)32417-X.

Alvares, Claude. 2011. "A Critique of Eurocentric Social Science and the Question of Alternatives." *Economic and Political Weekly* 46(22):72–81.

Arel-Bundock, Vincent, Nils Enevoldsen, and Cj Yetman. 2018. "Countrycode: An R Package to Convert Country Names and Country Codes." *Journal of Open Source Software* 3(28):848. doi: 10.21105/joss.00848.

Arnett, Jeffrey J. 2008. "The Neglected 95%: Why American Psychology Needs to Become Less American." *American Psychologist* 63(7):602–14. doi: 10.1037/0003-066X.63.7.602.

Bhambra, Gurminder K. 2014. *Connected Sociologies*. London: Bloomsbury Academic.

Bhambra, Gurminder K., and John Holmwood. 2021. *Colonialism and Modern Social Theory*. Cambridge Medford, MA: Polity.

Boatcă, Manuela. 2020. "Thinking Europe Otherwise: Lessons from the Caribbean." *Current Sociology* 26. doi: 10.1177/0011392120931139.

Bongaarts, John. 1987. "The Proximate Determinants of Fertility." *Technology in Society* 9(3–4):243–60. doi: 10.1016/0160-791X(87)90003-0.

Brohman, John. 1995. "Universalism, Eurocentrism, and Ideological Bias in Development Studies: From Modernisation to Neoliberalism." *Third World Quarterly* 16(1):121–40.

Buyalskaya, Anastasia, Marcos Gallo, and Colin F. Camerer. 2021. "The Golden Age of Social Science." *Proceedings of the National Academy of Sciences* 118(5):e2002923118. doi: 10.1073/pnas.2002923118.

Chant, Sylvia, and Cathy McIlwaine. 2009. *Geographies of Development in the 21st Century. An Introduction to the Global South*. Fist. Cheltenham: Edward Elgar Publishing Limited.

Cheng, Stephanie W., Chih-Wei Kuo, and Chih-Hua Kuo. 2012. "Research Article Titles in Applied Linguistics." *Journal of Academic Language & Learning* 6(1):A1–14.

Coronil, Fernando. 1996. "Beyond Occidentalism: Toward Nonimperial Geohistorical Categories." *Cultural Anthropology* 11(1):51–87.

Coutinho, Raquel Zanatta, and Andre Braz Golgher. 2018. "Modelling the Proximate Determinants of Fertility for Brazil: The Advent of Competing Preferences." *Revista Brasileira de Estudos de População* 35(1):1–28. doi: 10.20947/S0102-3098a0041.

DeJesus, Jasmine M., Maureen A. Callanan, Graciela Solis, and Susan A. Gelman. 2019. "Generic Language in Scientific Communication." *Proceedings of the National Academy of Sciences* 116(37):18370–77. doi: 10.1073/pnas.1817706116.

Falagas, Matthew E., Eleni I. Pitsouni, George A. Malietzis, and Georgios Pappas. 2008. "Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and Weaknesses." *The FASEB Journal* 22(2):338–42. doi: 10.1096/fj.07-9492LSF.

Gelman, Susan A., and Steven O. Roberts. 2017. "How Language Shapes the Cultural Inheritance of Categories." *Proceedings of the National Academy of Sciences* 114(30):8. doi: 10.1073/pnas.1621073114.

Go, Julian. 2020. "Race, Empire, and Epistemic Exclusion: Or the Structures of Sociological Thought." *Sociological Theory* 38(2):79–100. doi: 10.1177/0735275120926213.

Gómez, Francisco, José Jabaloyes, Luis Montero, Vicente De Vicente, and Manuel Valcuende. 2011. "Green Areas, the Most Significant Indicator of the Sustainability of Cities: Research on Their Utility for Urban Planning." *Journal of Urban Planning and Development* 137(3):311–28. doi: 10.1061/(ASCE)UP.1943-5444.0000060.

Grosfoguel, Ramón. 2013. "Racismo/sexismo epistémico, universidades occidentalizadas y los cuatro genocidios/epistemicidios del largo siglo XVI." *Tabula Rasa* (19):31–58. doi: 10.25058/20112742.153.

Hacking, Ian. 1996. "The Looping Effects of Human Kinds." Pp. 351–83 in *Causal Cognition*, edited by D. Sperber, D. Premack, and A. J. Premack. Oxford University Press.

Haggan, Madeline. 2004. "Research Paper Titles in Literature, Linguistics and Science: Dimensions of Attraction." *Journal of Pragmatics* 36(2):293–317. doi: 10.1016/S0378-2166(03)00090-0.

Hartley, James. 2005. "To Attract or to Inform: What Are Titles For?" *Journal of Technical Writing and Communication* 35(2):203–13.

Hyland, Ken. 2002. "What Do They Mean? Questions in Academic Writing." *Text - Interdisciplinary Journal for the Study of Discourse* 22(4). doi: 10.1515/text.2002.021.

Jenks, George. 1967. "The Data Model Concept in Statistical Mapping." *International Yearbook of Cartography* 7.

Kahalon, Rotem, Verena Klein, Inna Ksenofontov, Johannes Ullrich, and Stephen C. Wright. 2021. "Mentioning the Sample's Country in the Article's Title Leads to Bias in Research Evaluation." *Social Psychological and Personality Science* 194855062110240. doi: 10.1177/19485506211024036.

Krause, Monika. 2016. "'Western Hegemony' in the Social Sciences: Fields and Model Systems: 'Western Hegemony' in the Social Sciences: Fields and Model Systems." *The Sociological Review Monographs* 64(2):194–211. doi: 10.1002/2059-7932.12008.

Kvangraven, Ingrid Harvold, and Surbhi Kesar. 2021. "Standing in the Way of Rigor? Economics' Meeting with the Decolonizing Agenda."

Lin, Zhimin. 2019. "Xi Jinping's 'Major Country Diplomacy': The Impacts of China's Growing Capacity." *Journal of Contemporary China* 28(115):31–46. doi: 10.1080/10670564.2018.1497909.

Matthey, Stephen, and Anna Maria Della Vedova. 2020. "Screening for Mood Difficulties in Men in Italy and Australia Using the Edinburgh Postnatal Depression Scale and the Matthey Generic Mood Questionnaire." *Psychology of Men & Masculinities* 21(2):278–87. doi: 10.1037/men0000227.

Mignolo, Walter D. 2014. "Spirit out of Bounds Returns to the East: The Closing of the Social Sciences and the Opening of Independent Thoughts." *Current Sociology* 62(4):584–602. doi: 10.1177/0011392114524513.

Mongeon, Philippe, and Adèle Paul-Hus. 2016. "The Journal Coverage of Web of Science and Scopus: A Comparative Analysis." *Scientometrics* 106(1):213–28. doi: 10.1007/s11192-015-1765-5.

Moody, James. 2006. "Trends in Sociology Titles." *The American Sociologist* 37(1):77–80. doi: 10.1007/s12108-006-1016-6.

Nisbett, R. E., and T. Masuda. 2003. "Culture and Point of View." *Proceedings of the National Academy of Sciences* 100(19):11163–70. doi: 10.1073/pnas.1934527100.

Paiva, Carlos Eduardo, Joao Paulo da Silveira Nogueira Lima, and Bianca Sakamoto Ribeiro Paiva. 2012. "Articles with Short Titles Describing the Results Are Cited More Often." *Clinics* 67(5):5. doi: 10.6061/clinics/2012(05)17.

Quijano, Aníbal. 2000. "Coloniality of Power and Eurocentrism in Latin America." *International Sociology* 15(2):215–32. doi: 10.1177/0268580900015002005.

Rath, Alex. 2010. "Dual Function of First Position Nominal Groups in Research Article Titles: Describing Methods and Structuring Summary." *Journal of Applied Language Studies* 1(2):19.

Rosner, Judah L. 1990. "Reflections of Science as a Product." *Nature* 345(6271):108–108. doi: 10.1038/345108a0.

Rozin, Paul. 2006. "Domain Denigration and Process Preference in Academic Psychology." *Perspectives on Psychological Science* 1(4):365–76. doi: 10.1111/j.1745-6916.2006.00021.x.

Sagi, Itay, and Eldad Yechiam. 2008. "Amusing Titles in Scientific Journals and Article Citation." *Journal of Information Science* 34(5):680–87. doi: 10.1177/0165551507086261.

de Sousa Santos, Boaventura. 2016. *Epistemologies of the South: Justice against Epistemicide*. London New York: Routledge.

Subotic, Sinisa, and Bhaskar Mukherjee. 2014. "Short and Amusing: The Relationship between Title Characteristics, Downloads, and Citations in Psychology Articles." *Journal of Information Science* 10.

Walker, Catherine M. 1996. "Financial Management, Coping and Debt in Households under Financial Strain." *Journal of Economic Psychology* 17(6):789–807. doi: 10.1016/S0167-4870(96)00036-0.

Wallerstein, Immanuel. 1976. "A World-System Perspective on the Social Sciences." *The British Journal of Sociology* 27(3):11.

Wodak, Daniel, Sarah-Jane Leslie, and Marjorie Rhodes. 2015. "What a Loaded Generalization: Generics and Social Cognition: What a Loaded Generalization." *Philosophy Compass* 10(9):625–35. doi: 10.1111/phc3.12250.

Wu, Kaidi, and David Dunning. 2020. "Hypocognition and the Invisibility of Social Privilege." Pp. 1–23 in *Advances in Group Processes*, edited by S. R. Thye and E. J. Lawler. Emerald Publishing Limited.

Supplementary material

Figure A1. Regression coefficients for a multivariate Poisson model predicting the localization (Y = 1) of papers for each of the 23 social sciences and humanities subfields. The top- and bottom-three subfields with the highest and lowest localization rates are highlighted. The solid circles are statistically significant coefficients.
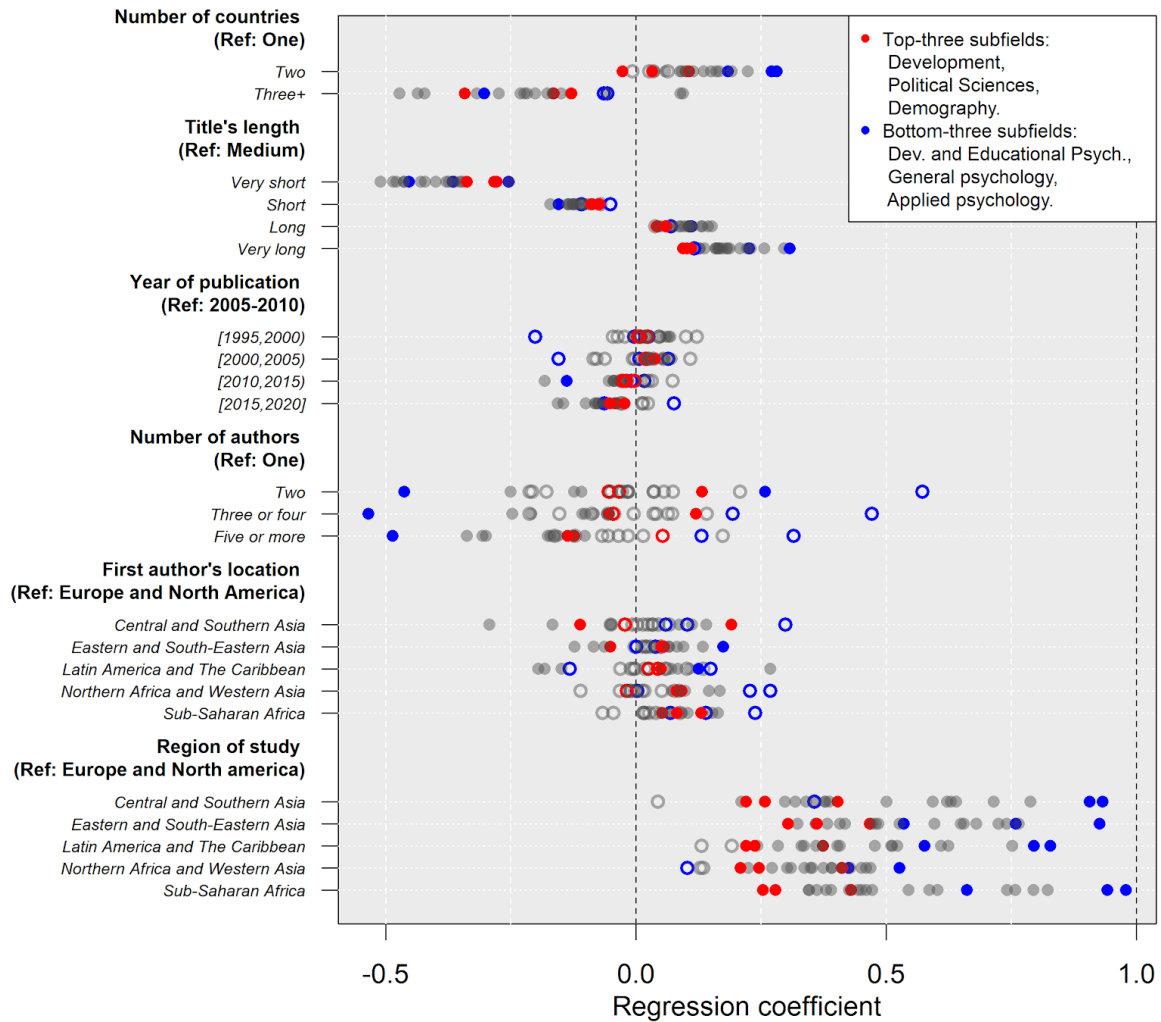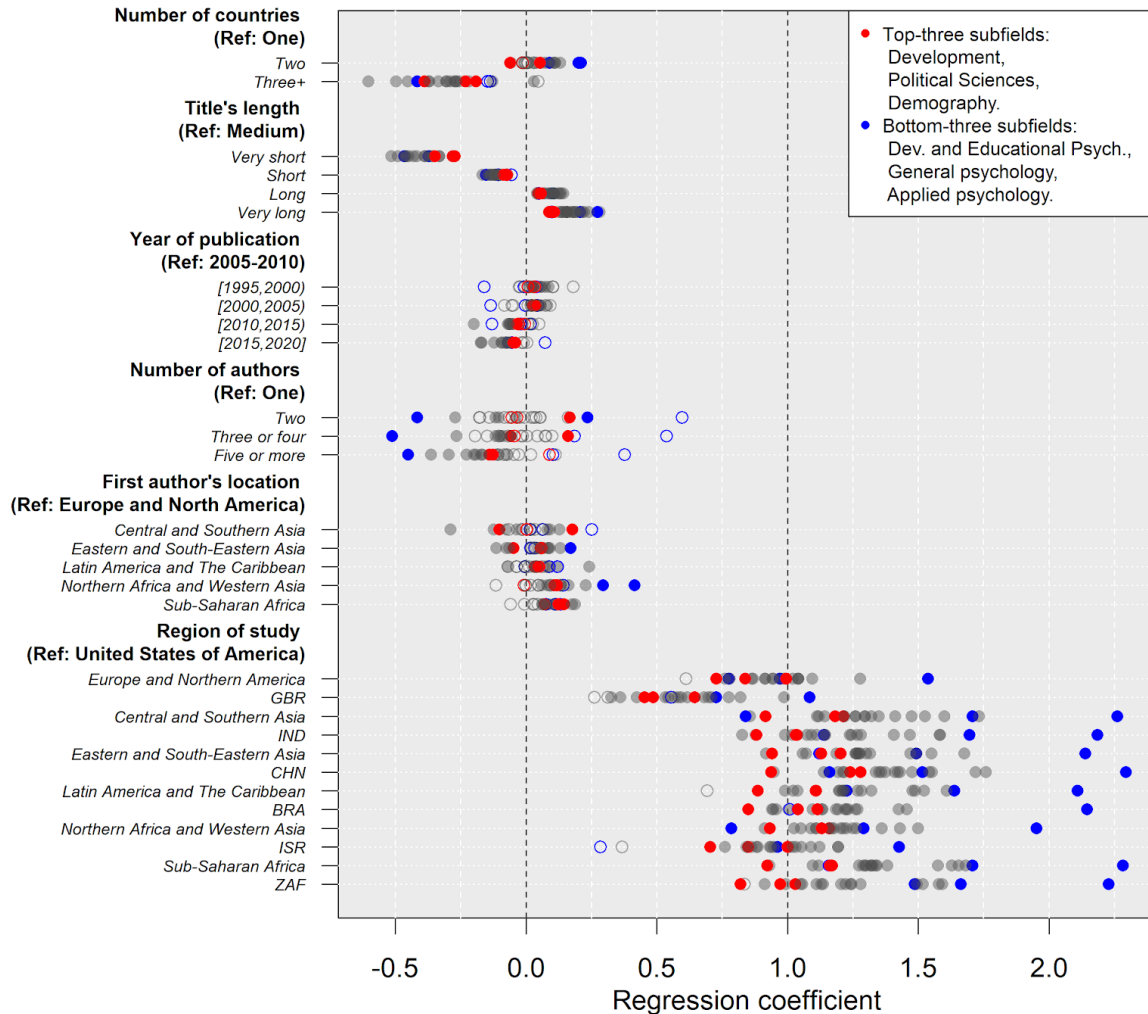

Note: We applied the Bonferroni correction to the significance of the coefficients for the regions of study. We obtained 108 statistically significant coefficients out of 115 (5 regression coefficients across 23 subfields).

Figure A2. Regression coefficients for the top 10 countries of study and other countries, grouped by regions and control variables. The top- and bottom-three subfields with the highest and lowest localization rates are highlighted. The solid circles are statistically significant coefficients.
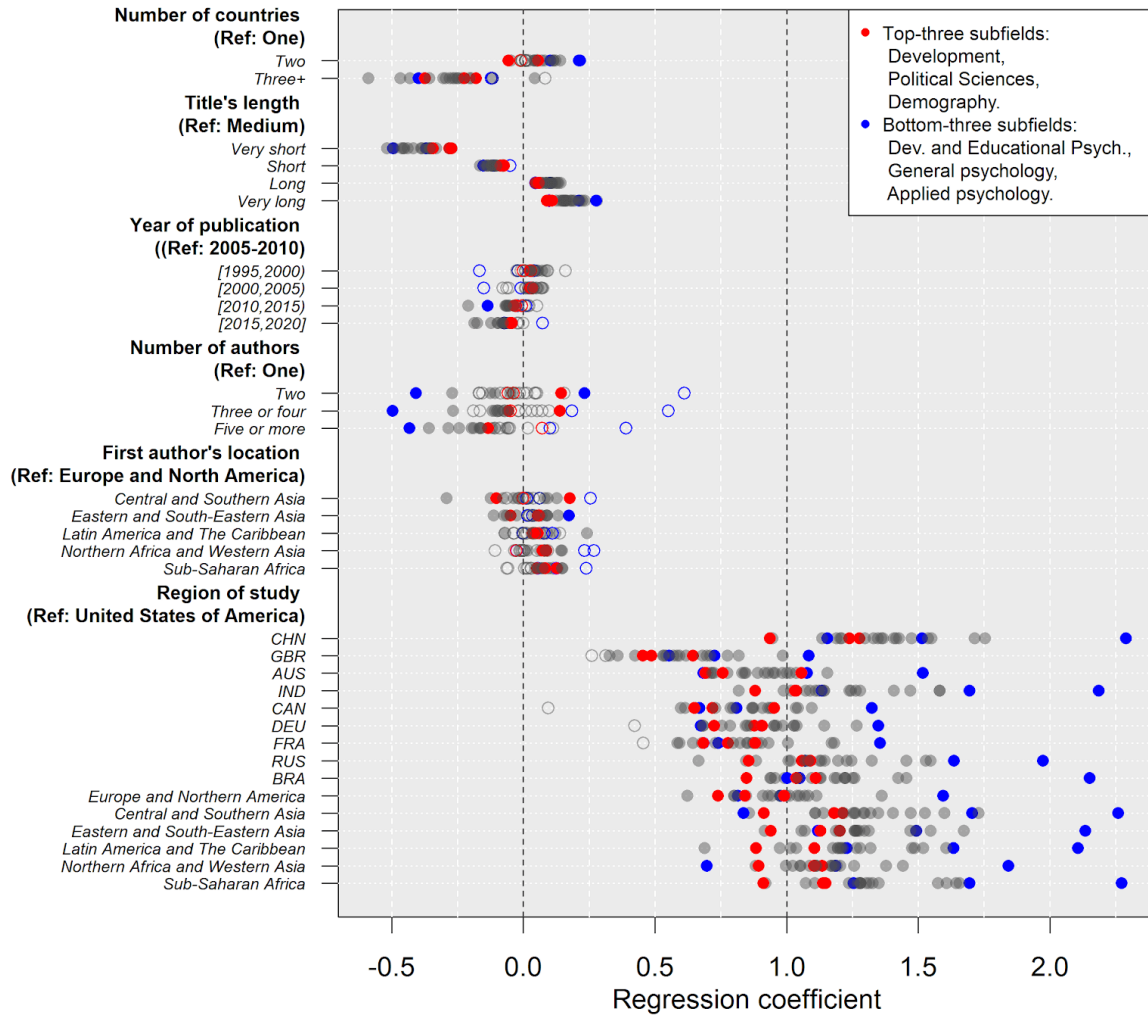
Note: We applied the Bonferroni correction to the significance of the coefficients for the top 10 countries and the regions of study. We obtained 332 statistically significant coefficients out of 345 (15 regression coefficients across 23 subfields).

Figure A3. Regression coefficients for within-region most studied country and other countries, grouped by region and control variables. The top- and bottom-three subfields with the highest and lowest localization rates are highlighted. The solid circles are statistically significant coefficients.

Note: We applied the Bonferroni correction to the significance of the coefficients for the top 10 countries and the regions of study. We obtained 267 statistically significant coefficients out of 276 (12 regression coefficients across 23 subfields).