Paper Tiger? Chinese Science and Home Bias in Citations*

Shumin Qiu
East China University of Science and Technology, School of Business
Shanghai 200237
P.R. China

Claudia Steinwender
Ludwig Maximilian University
Department of Economics
Ludwigstr. 33
D-80539 Munich

Pierre Azoulay
MIT & NBER
Sloan School of Management
100 Main Street – E62-487
Cambridge, MA 02142

May 16, 2024

Abstract

We investigate the phenomenon of home bias in scientific citations, where researchers disproportionately cite work from their own country. We develop a benchmark for expected citations based on the relative size of countries, defining home bias as deviations from this norm. Our findings reveal that China exhibits the largest home bias across all major countries and in nearly all scientific fields studied. This stands in contrast to the pattern of home bias for China's trade in goods and services, where China does not stand out from most industrialized countries. After adjusting citation counts for home bias, we demonstrate that China's apparent rise in citation rankings is overstated. Our adjusted ranking places China fourth globally, behind the US, the UK, and Germany, tempering the perception of China's scientific dominance.

*All authors contributed equally. Address correspondence to claudia.steinwender@econ.lmu.de.

1 Introduction

China has emerged as the leading global producer of scientific articles (Tollefson, 2018; Xie and Freeman, 2019), but does this surge corresponds to an equivalent increase in the *quality* of scientific output? Citations are widely used to assess research quality, but this metric is known to come with important limitations (Teplitskiy et al., 2022; Aksnes et al., 2019). We highlight a novel source of bias in citations that is particularly relevant for cross-national comparisons: home bias, i.e., the tendency of researchers to excessively cite researchers from their own country.

One challenge in estimating the extent of home bias in citations is specifying a benchmark for the geographic distribution of citations that would be observed in a no-bias world. A country's self-citations may appear large merely by virtue of the fact that larger countries receive more citations than small countries. We contribute to the literature by providing a yardstick that allows us to systematically identify home bias in citations, building on the literature on home bias in international trade (Santamaría et al., 2023). Specifically, we construct a benchmark based on the number of citations one would expect to observe if citations were distributed solely according to the sizes of all citing and cited countries. We define home bias as the deviation of actual citations from this benchmark.

We find that China exhibits by far the largest home bias among all countries. This is not a recent phenomenon. While China's home bias has been steadily increasing over the past twenty years, Chinese citations were already strongly home biased in 2000, the start of our observation period. In addition, China's home bias is not driven by any particular research field. Rather, China exhibits the strongest home bias in 18 out of 20 broad scientific fields. We also compare China's home bias in citations to its home bias in trade. Only for the former is China's home bias not in line with that of other countries.

Finally, we find that home-bias has exaggerated the rise of China in science. While China ranks second behind the US in terms of raw citations, it falls back to the fourth position behind the US, the UK and Germany once we use our de-biased metric. Home-debiasing citation counts might be seen as especially informative if one believes that home citations are especially prone to reflect political or strategic considerations, rather than the acknowledgement of scientists cumulatively building on the ideas contained in the articles they choose to cite.

2 Data Description

We select cited articles from top international scientific journals published in the English language.[1] We classify a journal in this category if it belongs to the top decile of journals within a specific research field, based on both its Impact Factor (IF) and its Eigenfactor (EF), which we draw from Clarivate's *Journal Citation Report* (JCR). This definition yields a sample of 461 journals covering 20 broad fields in the physical sciences, engineering, and biomedical research.[2]

We consider all original publications in top journals that were published between 2000 and 2021, covering the period of China's rise in science. This yields a total number of 3.75 million articles. Table A.1 of the online appendix shows that 17% of these articles are in clinical medicine, 13% in chemistry, and 11% in engineering. We assign each article to one or more countries depending on the addresses that appear in the list of unique affiliations for each article. On average, an article is affiliated with 1.6 countries (median: 1; minimum: 1; maximum: 112). We assign a fraction of the article to a country based on the share of a country's addresses in all listed addresses.[3]

During this period, China's share of top journal articles increased from 1.1% to 25.8%. China experienced the most rapid increases in the fields of materials science (from 3.9% to 45.2%), computer science (from 2.4% to 42.5%), agricultural science (from 1.2% to 41.0%) and engineering (from 2.0% to 41.6%). By 2021, China had become the world's largest publisher of top papers in 7 out of our 20 research fields, while the United States remained the world's leading publisher in the other 13 fields.

To analyze citation patterns, we consider all articles in *Web of Science* published between 2000 and 2021 without restricting citing papers to top journals. However, since we want to link citations to countries, we remove citing articles lacking country information (0.3% of citations). This yields a total of 202,142,130 unique citations (i.e., citing-cited article pairs). On average, each article is cited 53.93 times. We assign all citing papers to countries based on their addresses using the same method as described for top papers. We aggregate citations to the country-pair-field level to generate a complete count of citations from a country to

[1] One might object that our focus on English-language journals is perhaps inherently flattering to English-speaking countries. However, English is the *lingua franca* of international scientific communication, and the inclusion of journals published in other languages would certainly contribute to *increasing* home bias for countries—such as China—whose scientists frequently publish in domestic journals (He, 2003).

[2] The *Web of Science* assigns articles published in journals that cover the spectrum of major scientific disciplines (such as *Science* or *Nature*) to a separate field called "Multidisciplinary".

[3] Appendix A provides more details about data construction.

other countries using fractional counts, i.e., if $count_{pi}$ is the share of country i's addresses in paper p, and $count_{rj}$ is the share of country j's addresses in paper r that cites paper p, the bilateral citations of country j citing country i, $citations_{ij}$, are defined to be

$$citations_{ij} = \sum_p \sum_r count_{pi} \cdot count_{rj}.$$

Since articles are uniquely allocated to a research field based on the journal they are published in, we can construct the bilateral citation count separately for each field f (based on the field of the cited paper, not the citing paper), $citations_{fij}$.

3 Measuring Home Bias

Figure 1 displays the share of home citations in a country's overall citations. China receives the largest share of their citations from China itself, 57.2%. The US also receives a large share of their citations from home, 37.1%. China and the US stand out: Other countries have a lower share of home citations, such as India with 26.8% and Iran with 24.8%. However, the US and China are also countries with a large scientific workforce. If Chinese researchers are heavily cited by other Chinese researchers, this may just reflect that there are a larger number of potential Chinese citers relative to citers from other countries. It would not be legitimate to interpret this mechanical effect of country size as "bias".

We propose to measure home bias as deviations from a benchmark model, which we base on a "dartboard approach" by asking: How many citations would we expect country i to receive from country j, if world citations were distributed only according to the size of the citing and cited countries? Formally, we express this as:

$$citations_{ij}^{BM} := world\_cites \times \frac{\sum_j citations_{ij}}{world\_cites} \times \frac{pubs_j}{\sum_j pubs_j} \tag{1}$$

where i and j index the cited and citing countries, respectively, and $world\_cites$ denotes the total citations in the world,

$$world\_cites = \sum_i \sum_j citations_{ij}.$$

The benchmark citation flows in equation (1) distribute world citations according to the total share of citations the cited country i received from the world, and the share of the size of the potential citing country j based on its world publication share. We now define home bias of a country i as the deviation of its actual citation shares from the benchmark citation share:

$$\begin{aligned} home\_bias_i &:= \frac{citations_{ii}}{\sum_j citations_{ij}} - \frac{citations_{ii}^{BM}}{\sum_j citations_{ij}} \\ &= \frac{citations_{ii}}{\sum_j citations_{ij}} - \frac{pubs_i}{\sum_j pubs_j} \end{aligned} \tag{2}$$

4 Home Bias Across Countries

In Figure 2 we calculate the home bias across countries implied by equation (2). All countries in our sample exhibit home bias to some extent. However, China is a clear outlier with a home bias of 42.3%, followed by Iran (23.2%), and India (23.2%). The home bias of the US is on the larger side of the remaining countries, at 15.9%. In Figure 3 we plot home bias over time, according to the publication year of the cited article. China's home bias was already large (at 28.5%) at the beginning of our sample period. However, it has been steadily increasing ever since, in contrast to the home bias in other countries, which has remained fairly stable over the entire time period.

In Figure 4 we ask whether the home bias of China and other countries is more pronounced in particular research fields. We compute the home bias for each country as given by equation (2) separately by research field. China's home bias is larger than that of all other countries in 18 out of 20 broad research fields.[4]

It has long been established that trade in goods and services exhibits home bias as well, where countries disproportionately engage in economic exchanges with their nearby neighbors.[5] How does our home bias in 'knowledge exports', i.e., citations, compare to the home bias in trade? On one hand, we would expect knowledge flows to show a smaller home bias than goods flows, as physical trade barriers such as transport cost and customs are irrelevant. On the other hand, citations do not only reflect knowledge flows, but also strategic considerations, which may yield a larger measured home bias for citations compared to goods and services. Past research has also highlighted frictions in the diffusion of scientific knowledge (e.g., Qiu et al. [2023]), so the extent to which home bias in trade and home bias in citations correlate is an empirical question.

[4] Iran in space science and Brazil in pharmacology are the only country-field combinations with larger home bias than China's. Other countries tend to be home biased in a single field, such as in the case of India and geoscience.

[5] So well documented is this feature of international trade patterns that it features as one of "The Six Major Puzzles in International Macroeconomics" highlighted by Obstfeld and Rogoff (2000).

Our approach is most closely related to Santamaría et al. (2023) who define home bias as deviations of trade from a benchmark model that they label 'naive gravity,' in which trade between two countries is determined by the size (measured as total spending) of the origin and destination country. In this spirit, we define the home bias in exports in an analogous way to home bias in citations. First, as in equation (1), we express benchmark exports from country i to country j by distributing world GDP according to the relative size of the exporting country i, given by $\sum_j exports_{ij} = GDP_i$, and the relative size of the importing country j, given by $\sum_i exports_{ij} = GDP_j$, as follows:

$$exports_{ij}^{BM} := world\_GDP \times \frac{\sum_j exports_{ij}}{world\_GDP} \times \frac{GDP_j}{\sum_j GDP_j} \tag{3}$$

Then, we define the home bias of a country i as the deviation of its actual export share from the benchmark export share:

$$\begin{aligned} home\_bias\_exports_i &:= \frac{exports_{ii}}{\sum_j exports_{ij}} - \frac{exports_{ii}^{BM}}{\sum_j exports_{ij}} \\ &= \frac{exports_{ii}}{GDP_i} - \frac{GDP_i}{\sum_j GDP_j} \end{aligned} \tag{4}$$

Empirically, we follow the literature and derive a country i's exports to itself, $exports_{ii}$, i.e., domestic absorption, as the residual of a country's total production minus net exports to all other countries, $exports_{ii} = GDP_i - (\sum_{j \neq i} exports_{ij})$. We use data on bilateral net exports for 190 exporting countries and 250 importing countries between 2000 and 2021 from the United Nations *Comtrade* database and data on countries' GDP from the World Bank, averaging the data across time.

Figure 5 compares the home bias in citations (y-axis) to the home bias in the exports of goods and services (x-axis). Similar to the patterns in citations and consistent with the literature in trade, all countries exhibit some form of home bias. However, unlike citations, no country stands out as having a particularly large home bias in trade. In particular, China's home bias in trade is in line with that of other countries. In general, the home bias in citations and trade are only weakly positively correlated.

These findings suggests that China's exceptionally large home bias in citations is driven by the Chinese scientific research environment. While uncovering the behavioral mechanisms that lead to home bias is beyond the scope of this article, a possible cause is the large degree of power-orientation in Chinese society. Power relationships have been shown to influence resource allocation to Chinese researchers (Jia et al., 2019). In the same vein, the selection of Fellows at the Chinese Academies of Sciences and Engineering reflects a culture of favor exchange (Fisman et al., 2018). Focusing on the field of nanotechnology, Tang et al. (2015) show that a higher rate of citing takes place within individual, institutional, and national networks in China, relative to other countries. These authors speculate that *"the norms of interpersonal relationships (guanxi) in China may lead Chinese scholars to cite the work of their colleagues in the same institute, who they meet frequently, or leading scholars in their own country, who have an influence in proposal review and external evaluation for promotion"* (p. 1930).[6]

5 Implications of home bias for rankings

Since home citations can be influenced by factors beyond scholarly merit, the inclusion of home citations may bias citation counts as a measure of a country's research achievements.[7] This effect is particularly salient in countries with a large home bias, such as China, Iran and India. To correct for possibly inflated home citations and derive an adjusted citation measure that better reflects research quality, we suggest a revised citation count that adjusts the abnormally large number of home citations by using a country-specific debiasing factor δ_i, which is given by the ratio of benchmark home citations to actual home citations:

$$\delta_i := \frac{citations_{ii}^{BM}}{citations_{ii}}$$

Using the definition of benchmark citations from equation (1) we derive:

$$\begin{aligned} \delta_i &= \frac{world\_cites \times \frac{\sum_j citations_{ij}}{world\_cites} \times \frac{pubs_i}{\sum_j pubs_j}}{citations_{ii}} \\ &= \frac{\frac{pubs_i}{\sum_j pubs_j}}{\frac{citations_{ii}}{\sum_j citations_{ij}}} \end{aligned} \tag{5}$$

Applying this country-specific debiasing discount factor δ_i to home citations yields our adjusted measure of aggregate citations to assess the research quality of a country:

$$citations\_adjusted_i = \sum_{j \neq i} citations_{ij} + \delta_i \cdot citations_{ii} \tag{6}$$

[6]Note that the rapid increase of the university sector in China over the past twenty years (Xie et al., 2014) is likely not a culprit for the patterns we document, since our benchmark incorporates the surge in Chinese scientific labor supply.

[7]Appendix C provides a plausibility check for this conjecture by demonstrating that, for home-biased countries, home citations are less likely to be topically-related to the articles they cite, relative to non-home citations.

We compare the rankings of countries with and without de-biasing adjustment in Figure 6. The dark blue bars represent country rankings based on the total number of citations between 2000 and 2021. The US is first, with a total of over 81 million citations. China is second with 16.9 million citations, which is approximately one-fifth of the citations that US publications generate. Following are the UK (13.8 mn), Germany (12.2 mn), and Japan (9.8 mn).

The light blue bars in Figure 6 reflect the ordering of countries when we use the adjusted citation counts as given by equation (6). The US retains its top position, even though a large share of citations received by US scientists is discounted. China, however, falls two ranks (from second to fourth behind Germany and the UK), while the ranking of other countries is largely unchanged.[8]

6 Conclusion

China's recent emergence as the world's largest producer of scientific research (Tollefson, 2018; Zhou and Leydesdorff, 2006), with an attending dramatic rise in citations (Wagner et al., 2022), has led to an ongoing debate about the quality of research originating from the Middle Kingdom (Huang, 2018).

We investigate whether China's share of home citations is 'abnormal,' by comparing the geographic composition of Chinese citations to a no-bias benchmark. We find a large and increasing home bias in Chinese research that is almost twice as large as that of other countries, and prevalent across all fields of research. This stark home bias in citations contrasts with China's average home bias in the export of goods and services.

Other scholars have documented the increase in publications originating from China, the citations received by these publications, as well as the high share of overall citations accounted by domestic Chinese researchers (Xie and Freeman, 2020; Baccini and Petrovich, 2023; Larivière et al., 2018). However, a country's rate of self-citation may appear large merely because the country itself is large (as measured by its share of the world's publications and pool of potential citers). We contribute to the literature by providing a yardstick that

[8] In Figure B.2 of the online appendix we rank countries based on citations per paper rather than aggregate citations, which is one way of distinguishing between quantity (number of published papers) from quality (citations per published paper). The US also ranks first in citations per paper (with 71.4 citations per paper), whereas China falls behind to 14^{th} place, as it receives far less citations per paper than the US (39.7 citations per paper). Our adjusted citation measure pushes China even further behind to the 20^{th} rank (23.0 citations per paper).

allows us to systematically quantify home bias in citations, building on similar approaches in the international trade literature.

Home bias can distort rankings based on citation counts. We propose an adjusted citation measure that accounts for this type of bias, and find that the prominence of China in research is diminished when employing this method. While we only demonstrate the relevance of home bias for cross-country comparisons, the correction could also be applied to the assessment of articles, researchers, or institutions.

The counting exercise presented above may serve to temper the discourse surrounding China's purported eclipse of the West, at least as far as scientific prowess is concerned. Should it emerge that the quality of Chinese scientific output still lags even as its quantity leaps ahead, there may be an opportunity to encourage Chinese leadership to recalibrate their emphasis from the latter to the former. Conversely, in the Western context, alleviating the existential dread of losing a scientific showdown with China could stimulate innovations in peer review processes and foster international collaboration, rather than contribute to emulating a potentially overstated Chinese model of "success."

References

Aksnes, Dag W., Liv Langfeldt, and Paul Wouters. 2019. "Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories." *SAGE Open* 9:2158244019829575.

Baccini, Alberto and Eugenio Petrovich. 2023. "A Global Exploratory Comparison of Country Self-citations 1996-2019." *PLoS ONE* 18(12):e0294669.

Fisman, Raymond, Jing Shi, Yongxiang Wang, and Rong Xu. 2018. "Social Ties and Favoritism in Chinese Science." *Journal of Political Economy* 126:1134–1171.

He, Tianwei. 2003. "Difficulties and Challenges of Chinese Scientific Journals: Statistical Analysis of Chinese Literatures using Chinese Science Bulletin as Example." *Scientometrics* 57:127–139.

Huang, Futao. 2018. "Quality Deficit Belies the Hype." *Nature* 564:S70–S71.

Jia, Ruixue, Huihua Nie, and Wei Xiao. 2019. "Power and Publications in Chinese Academia." *Journal of Comparative Economics* 47:792–805.

Larivière, Vincent, Kaile Gong, and Cassidy R. Sugimoto. 2018. "Citations Strength Begins at Home." *Nature* 564:S70–S71.

Obstfeld, Maurice and Kenneth Rogoff. 2000. "The Six Major Puzzles in International Macroeconomics: Is there a Common Cause?" In *NBER Macroeconomics Annual*, edited by Ben S. Bernanke and Kenneth Rogoff, volume 15, pp. 339–390. Chicago, IL: University of Chicago Press.

Qiu, Shumin, Claudia Steinwender, and Pierre Azoulay. 2023. "Who Stands on the Shoulders of Chinese (Scientific) Giants? Evidence from Chemistry." NBER Working Paper #30772.

Santamaría, Marta, Jaume Ventura, and Uğur Yeşilbayraktar. 2023. "Exploring European Regional Trade." *Journal of International Economics* 146:103747.

Tang, Li, Philip Shapira, and Jan Youtie. 2015. "Is There a Clubbing Effect Underlying Chinese Research Citation Increases?" *Journal of the Association for Information Science and Technology* 66:1923–1932.

Teplitskiy, Misha, Eamon Duede, Michael Menietti, and Karim R. Lakhani. 2022. "How Status of Research Papers Affects the Way They Are Read and Cited." *Research Policy* 51:104484.

Tollefson, Jeff. 2018. "China Declared World's Largest Producer of Scientific Articles." *Nature* 553:390.

Wagner, Caroline S., Lin Zhang, and Loet Leydesdorff. 2022. "A Discussion of Measuring the Top-1Publications: Quality and Impact of Chinese Papers." *Scientometrics* 127:1825–1839.

Xie, Qingnan and Richard B. Freeman. 2019. "Bigger Than You Thought: China's Contribution to Scientific Publications and Its Impact on the Global Economy." *China & World Economy* 27:1–27.

Xie, Qingnan and Richard B. Freeman. 2020. "The Contribution of Chinese Diaspora Researchers to Scientific Publications and China's 'Great Leap Forward' in Global Science." NBER Working Paper #27169.

Xie, Yu, Chunni Zhang, and Qing Lai. 2014. "China's Rise as a Major Contributor to Science and Technology." *Proceedings of the National Academy of Sciences* 111:9437–9442.

Zhou, Ping and Loet Leydesdorff. 2006. "The Emergence of China as A Leading Nation in Science." *Research Policy* 35:83–104.

Figures

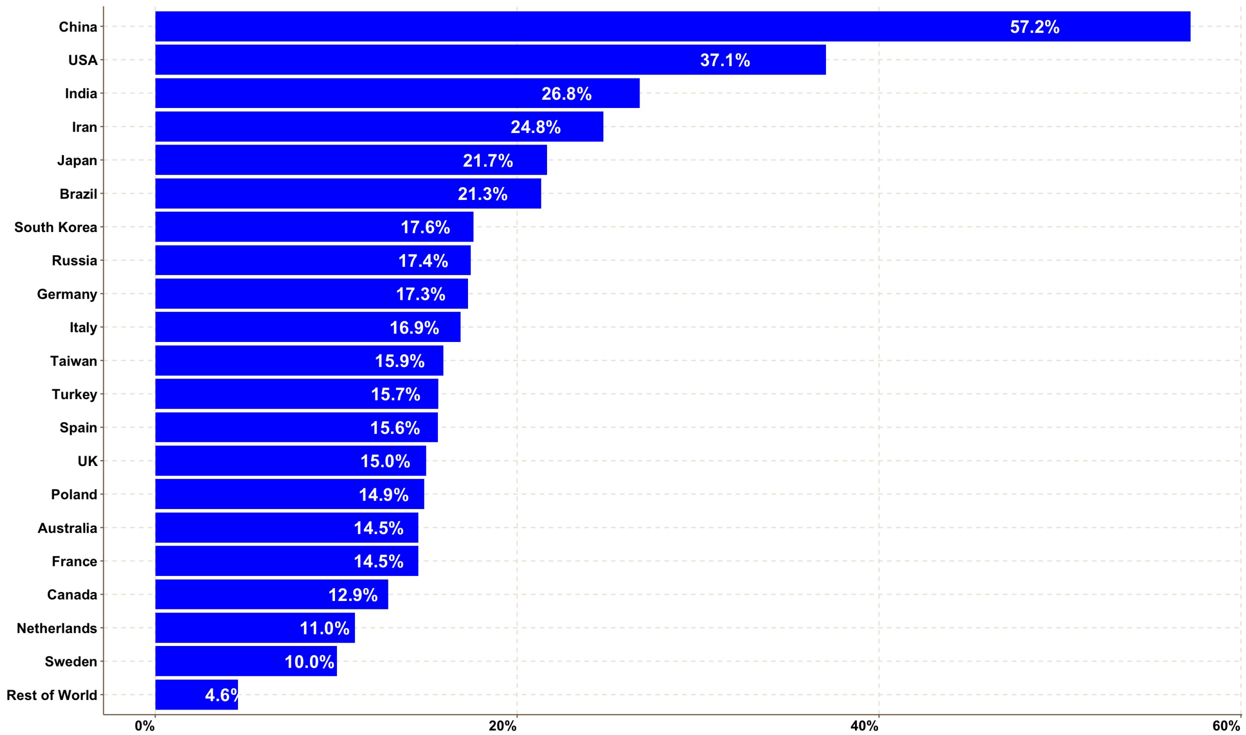

Figure 1: Share of Home Citations

Notes: This figure computes the raw share of home citations to top journal articles for each country/region. This figure is based on the citations of articles in 20 research fields published in 461 top journals between 2000 and 2021. For the figure, we selected the 20 countries with the largest number of total publications.

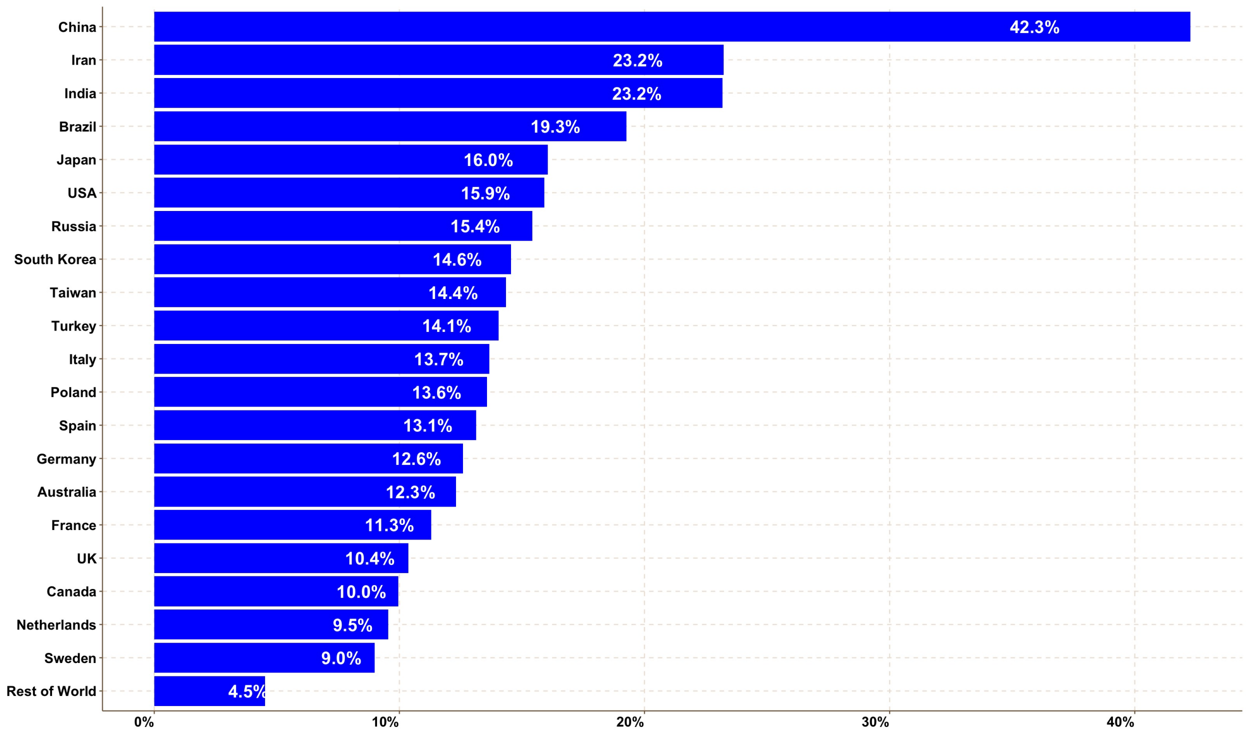

Figure 2: Home Bias in Citations

Notes: This figure computes home bias for each country as defined in equation (2). It is based on the citations of articles in 20 research fields published in 461 top journals between 2000 and 2021. For the figure, we selected the 20 countries/regions with the largest number of total publications.

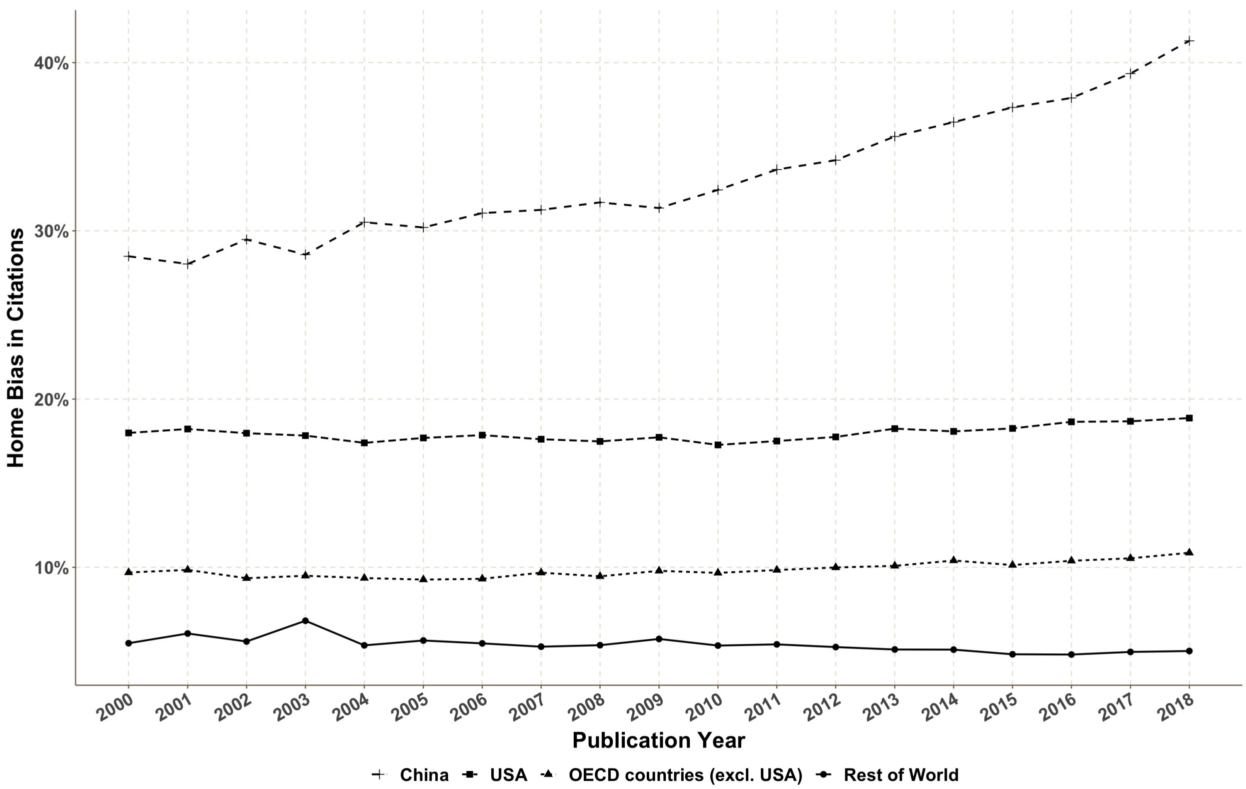

Figure 3: Home Bias in Citations over Time

Notes: This figure computes home bias for each country/region as defined in equation (2), separately by cited article publication year. It is based on articles that were published in 461 top journals in 20 research fields during 2000-2018, and citations to those articles during 2000-2021, in order to ensure that each article has at least a 3-year window to be cited.

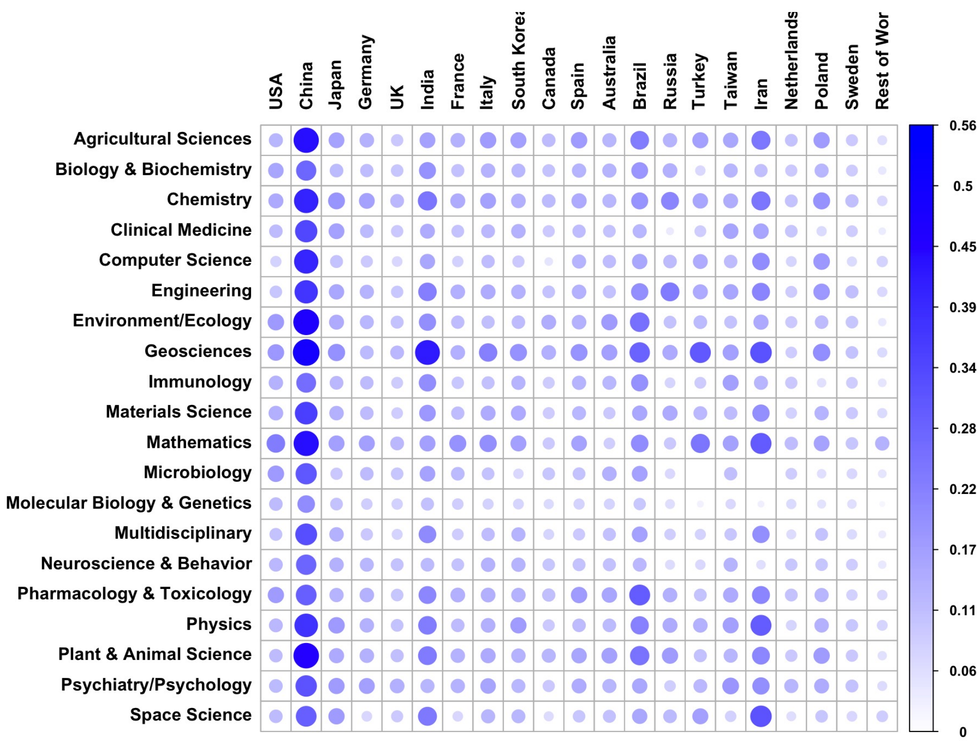

Figure 4: Home Bias in Citations by Field

Notes: This figure computes home bias for each country/region as defined in equation (2), separately by the research field of the cited article. This figure is based on the citations of articles in 20 research fields published in 461 top journals during 2000 to 2021.

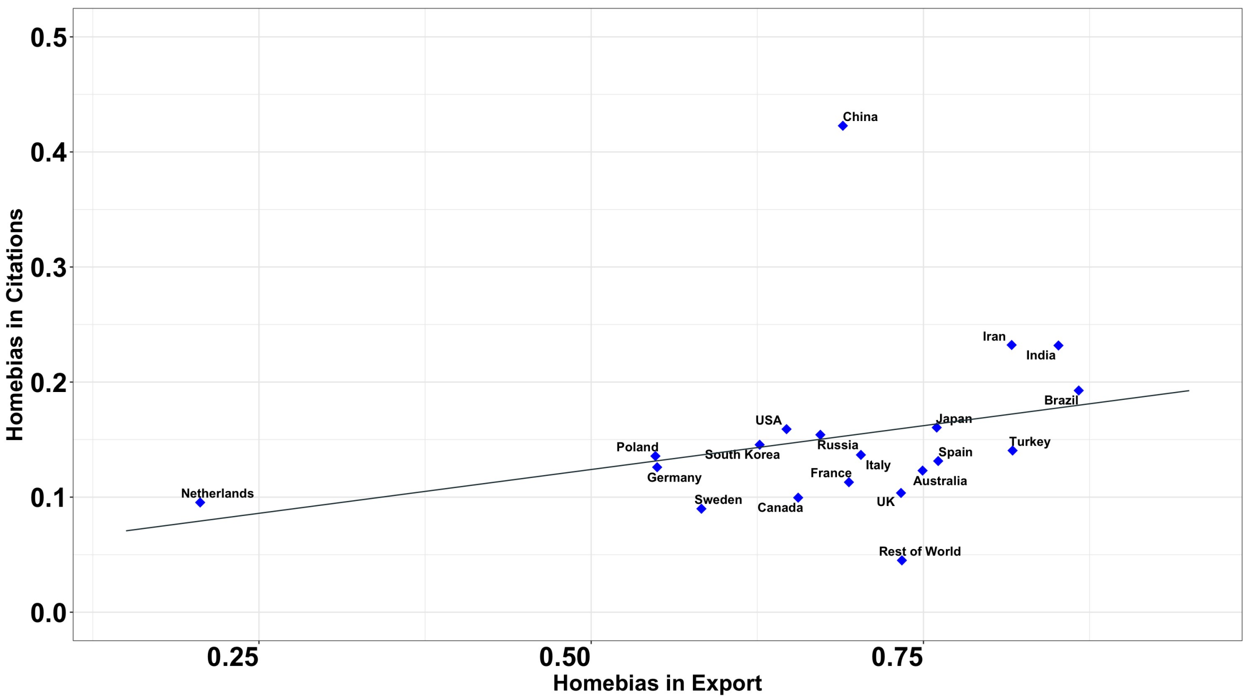

Figure 5: Home Bias in Knowledge Flows vs. Trade Flows

Notes: Home bias in citations (y-axis) is computed according to equation (2), and home bias in exports (x-axis) is computed in an analogous way according to equation (4).

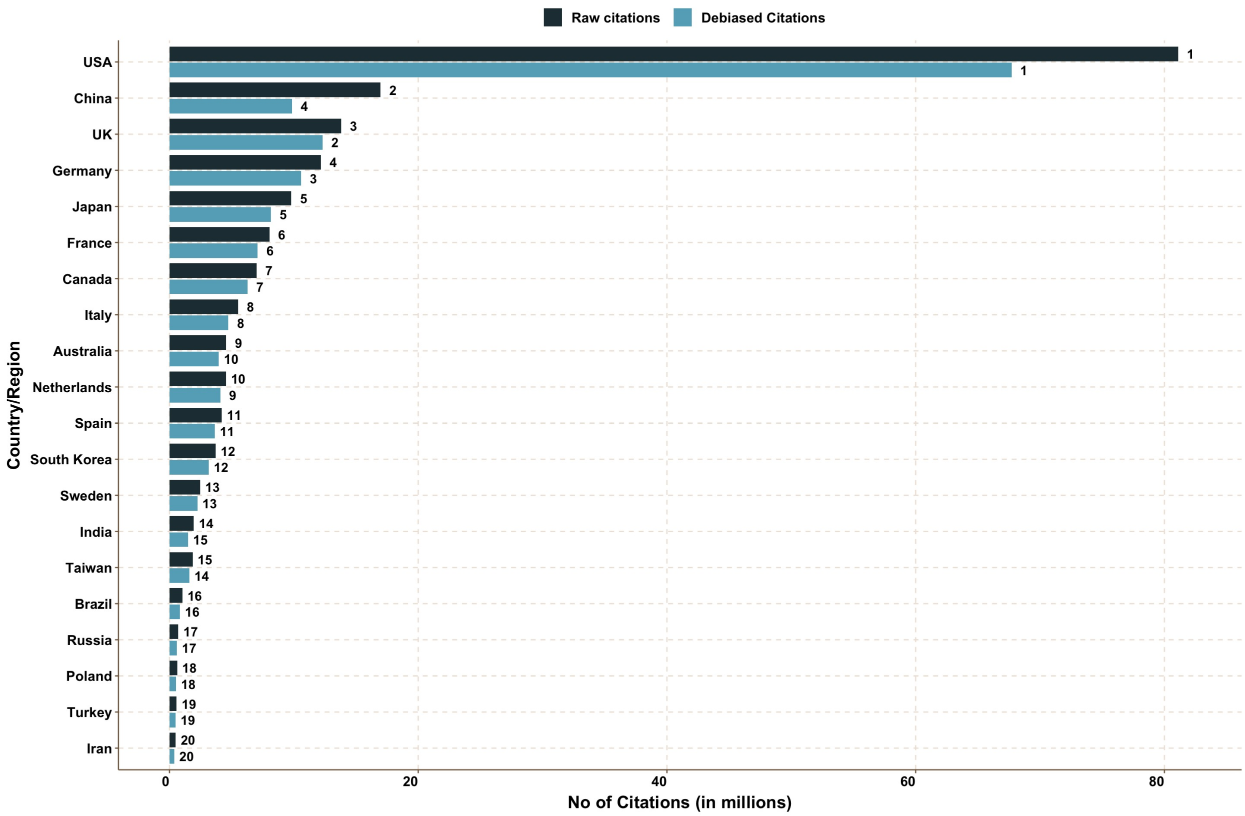

Figure 6: Adjusted Rankings of Countries/Regions by Total Citations Received

Notes: This figure adjusted the rankings for each country/region by equation (6). This figure is based on articles that were published in 461 top journals in 20 research fields during 2000-2021, and citations to those articles during the same period.

Online Appendices

Appendix A
Dataset Construction

We use both Impact Factor and Eigenfactor to ensure that our selection of elite journals does not depend on a single indicator. The Impact Factor (IF) measures the average number of citations to articles published in a journal over a two-year period. Despite its popularity, the IF is often criticized for putting more emphasis on journal popularity than prestige (Bollen et al., 2006). In addition, editorial policies that encourage journal self-citations can be used to deliberately manipulate the IF. The more recently developed Eigenfactor (EF) has been introduced as a supplement to the IF. Rather than just counting the number of citations a journal receives, these citations are weighted by the popularity (based on citations) of the citing journal. All else equal, journals that receive more citations from top-tier journals will have a higher EF, indicating a higher prestige of the journal (Yin 2011; see also `www.eigenfactor.org`).

The *Journal Citation Reports* by *Clarivate Web of Science* provide both IF and EF annually since 2007. We calculate the average IF and EF for each journal between 2007 and 2020, and select journals that belong to the top 10% of journals in a research field according to both IF and EF. The distribution of top journals across fields is similar to the distribution of all journals from *Web of Science* across fields, as can be seen in Table A.1 in the online appendix.

We restrict articles to English articles in the *Web of Science Core Collection—Science Citation Index Expanded and Social Sciences Citation Index.* Non-English articles contribute only 0.07% of the observations in the sample. We exclude books, editorials, interviews, commentaries, reviews, conference proceedings, corrections, meeting abstracts, letters, book chapters and book reviews. Journal articles account for more than 90% of all English publications. We also exclude two research fields ("Economics and Business"and "Social Sciences, General") to keep our focus on the physical sciences, the life sciences, and engineering.

The share of the article allocated to a specific country is equal to the number of unique affiliations from that country divided by all unique affiliations listed. Note that we have to drop 1.67% of articles that do not have any address information. Also note that *Web of Science* does not always assign affiliations to specific authors (only in 41% of papers), which is why each unique affiliation gets an equal weight, regardless of whether there is a coauthor who has several affiliations or there are several coauthors who are affiliated to the same research institution.

All country shares add up to one for a given article. For example, if an article lists two Chinese addresses, one address in the US, and one in the UK, we assign China a fractional count of 0.5, and 0.25 each to the US and the UK. Overall, the publications in our sample have been written by authors residing in 196 different countries.[i]

By 2021, US scientists rank first in terms of top journal publication share in 13 fields: mathematics, neuroscience & behavior, molecular biology & genetics, biology & biochemistry, psychiatry/psychology, space science, immunology, geosciences, pharmacology & toxicology, microbiology, clinical medicine, environment/ecology, and multidisciplinary sciences. China ranks first in the remaining 7 fields (followed by the US), which are materials science, chemistry, computer science, engineering, physics, agricultural sciences, and plant and animal science.

[i] We treat Hong Kong, Taiwan and China as separate regions for the purpose of this article.

Aggregating the data across the entire time period, the US accounts for the largest share of publications in our data (31.18%), followed by China (12.25%), UK (6.12%), Germany (6.02%) and Japan (5.35%). The US ranks first in 17 of our 20 research fields, with publication shares ranging from 23.14% in plant and animal science to 50.86% in immunology. There are three fields in which China has a larger publication share than the US: agriculture; engineering; and materials science. China ranks second after the US in computer science (19.87%), chemistry (18.33%), plant and animal science (12.52%), multidisciplinary sciences (12.41%), physics (11.25%), mathematics (10.48%), geosciences (9.41%), and environment/ecology (8.67%), However, China's presence is comparatively less pronounced in psychiatry/psychology (2.47%), immunology (2.60%), and neuroscience and behavior (2.61%).

Table A.1: Sample of fields and top journals

Fields	No. of Journals in WoS	Top Journals in WoS	English journal articles in top journals (2000-2021)
Clinical Medicine	2,042	108	623,577
Chemistry	546	24	502,982
Engineering	898	40	401,066
Physics	320	11	342,348
Multidisciplinary	64	5	298,706
Materials Science	404	20	253,920
Plant & Animal Science	816	41	172,509
Biology & Biochemistry	440	17	158,464
Geosciences	432	22	147,719
Space Science	56	5	132,053
Neuroscience & Behavior	346	18	117,750
Agricultural Sciences	358	15	112,360
Environment/Ecology	374	19	107,127
Psychiatry/Psychology	655	41	102,358
Molecular Biology & Genetics	305	18	72,544
Computer Science	402	14	59,228
Immunology	170	9	59,165
Mathematics	511	19	36,708
Pharmacology & Toxicology	284	10	34,704
Microbiology	131	5	12,917
All fields	9,554	461	3,748,205

Notes: We use the research area definitions that are used by *Essential Science Indicators (ESI)* to define fields. ESI covers 11,728 journals indexed in the *Web of Science Core Collection* and each journal is assigned to only one of 22 research fields. The field "Multidisciplinary" includes journals that publish research in many different fields (such as *Science*, *Nature*, or *Proceedings of the National Academy of Sciences*), journals that use a multidisciplinary approach, as well as interdisciplinary journals that aim to build connections between fields. For more details see `https://esi.help.clarivate.com/Content/scope-notes.htm`. We define top journals as journals that belong to the top 10% of journals in a research field according to either Impact Factor (IF) or Eigenfactor (EF).

Appendix B
Additional Results

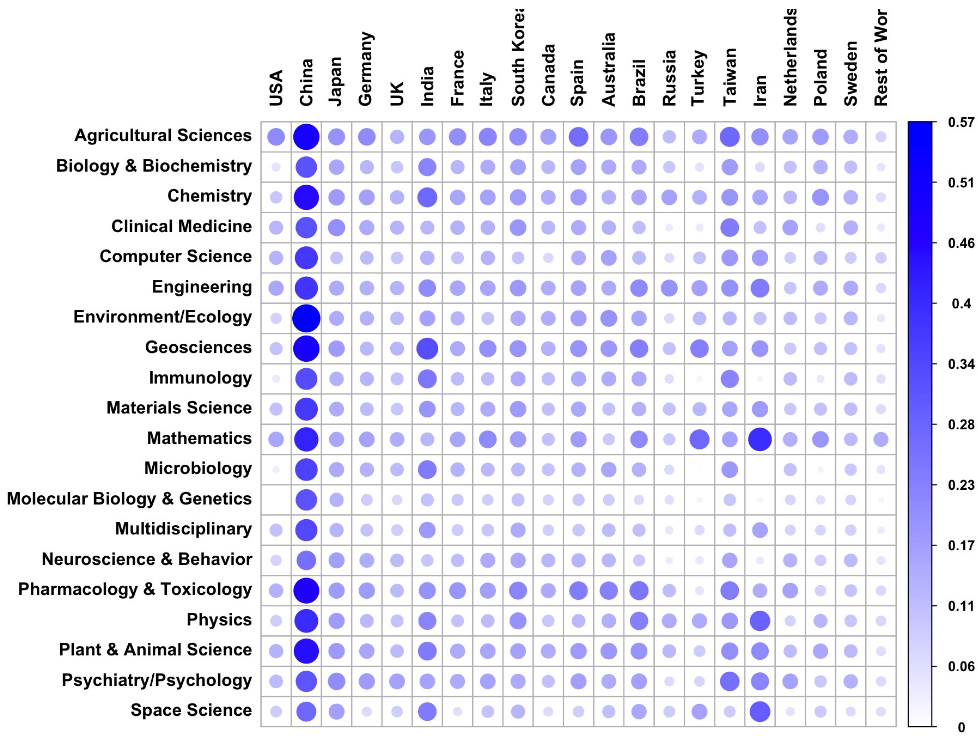

Figure B.1: Home Bias in Citations: Citing Article Restricted to Top Journals

Notes: This figure replicates Figure 4, except that it restricts citing articles to be from top journals.

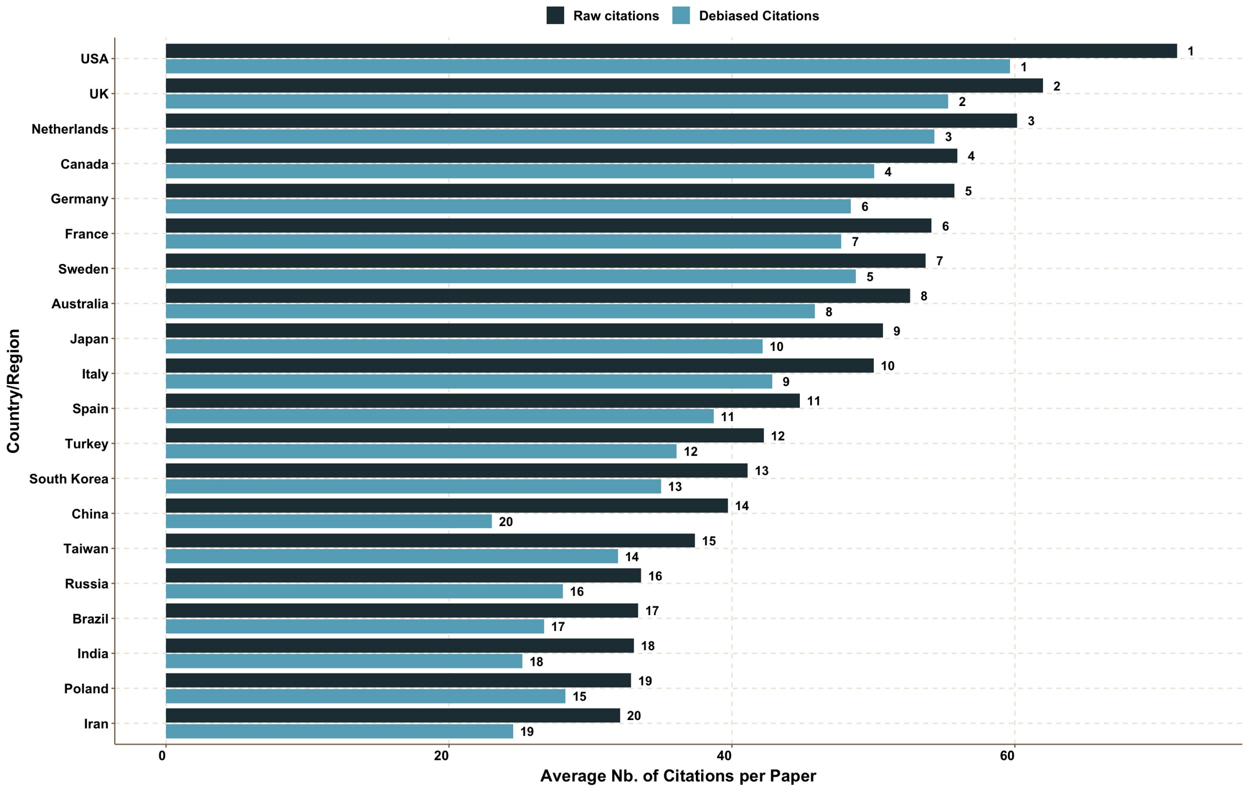

Figure B.2: Adjusted Rankings of Countries/Regions by Average Citations per Paper

Notes: This figure replicates Figure 6, except that it ranks countries/regions by the number of average adjusted citations per paper.

Appendix C
The Intellectual Content of Home Citations

Citations are often considered the "coin of the realm" in academic science because they are supposed to reflect scholars' acknowledgement of the ideas they build on in their own research. Although the "standing on shoulders" interpretation of the citing act is legitimate, scientists cite one another's research for numerous reasons. Citations can be "negative" in the sense that the ideas cited are being criticized (Catalini et al., 2015), but they may also be ceremonial in nature, or reflect a desire to curry factors with editors or referees (Teplitskiy et al., 2022; Aksnes et al., 2019).

Below, we attempt to compare the "intellectual content" of citations from one's home country with that of citations coming from other countries. We do so by leveraging the PubMed Related Citations Algorithm (PMRA), which delineates (for any publication appearing in a journal indexed by PubMed) a set of intellectual neighbors on the basis of shared keywords, title words, and abstract words (Lin and Wilbur, 2007). Importantly, neither the identity or locations of authors, nor citation or collaboration linkages, past or present, has any bearing on whether PMRA identifies two articles as belonging to the same intellectual neighborhood.

Our approach is to model the likelihood that a citing article is intellectually related to its source as a function of covariates of each citing/cited pair, including whether the citing and cited authorship teams hail from the same country (i.e., a "home" citation). If one could show that home citations in home-biased countries are less likely to be related (in the sense of PMRA) than non-home citations, this would lend credibility to the claim that home citations are more likely to reflect strategic or status considerations.

Because harvesting intellectual relatedness information from PMRA is costly, we limit our analysis to a subset of the data considered in the main body of the manuscript. The focus is on a set of 74,216 articles published by 751 elite chemists (see Qiu et al. [2023] for more details). Each source article is identified with a country by the location of its senior author, which occupies the last position on the authorship roster. To form citing/cited article pairs, we focus on 350,720 unique citing articles (from any journal) where all addresses belong to a single country, so that distinguishing between home and non-home citations is unambiguous, for a total of 955,029 citing/cited article pairs.[ii]

With these data available, we model the probability that the citation from article j to article i is topically related in the sense of PMRA by estimating the coefficients of the following linear probability model:

$$\mathbb{1}_{(j \text{ PMRA rel. to } i)} = \beta_1 \cdot Home_{ij} + \beta_2 \cdot HomeBias_i + \beta_3 \cdot Home_{ij} \times HomeBias_i + \varphi(i,j) + \varepsilon_{ij} \qquad \text{(C.1)}$$

The dependent variable is an indicator variable that takes on value 1 if citing article j is PMRA-related to cited article i, and 0 otherwise. Our regressors of interest include $Home_{ij}$, an indicator variable for whether i's last author and the authors of j are located in the same country, and $HomeBias_i$, the home bias of the country of cited article i as defined in equation (2) in Section 3 in the main body of the manuscript (re-computed based on the sample of papers used in this exercise), as well as the interaction between these two variables. $\varphi(i,j)$ corresponds to a large set of fixed effects for i and $i \times j$ characteristics, including fixed effects for i's publication year, the number of years between the citing and cited publication years, the citing journals, and the cited journals. To assess the robustness of the effects, we even include citing *article* fixed effects, as well as citing journal by year fixed effects and citing country by year fixed effects in some specifications. We cluster standard errors simultaneously at the level of the cited and citing article.

[ii] Having a well-characterized population of PIs allows us to assign their work to countries unambiguously. Because we have carefully disambiguated the bibliome of these 751 chemists using individually-collected CVs, we can also purge the data from self-citations—usually a challenging proposition for Asian names.

We do not take a strong stand on the sign (or magnitude) of β_1. The main effect of home citation will reflect a mix of influences, of which depth of intellectual engagement with the cited article is but a single one. But the home effect could also reflect differential density of coauthorship networks at home and abroad, or the clustering of researchers in narrow subfields within specific countries or regions. In contrast, we are interested in the sign (and magnitude) of β_3: If home citations in home-biased countries are more likely to stem from strategic—as opposed to intellectual—considerations, we would expect β_3 to be negative.

Tables C.1 and C.2 present descriptive statistics for the sample, at the level of the cited article and the level of the cited/citing article pair, respectively. The cited papers in the sample were published between 2000 and 2018 and garnered 12.9 citations on average between the time of their publication and the year 2021 (this entails that we have a citation window of at least three years for every article in the sample). The citations come from a range of journals whose impact factors range between 0 and 43. The cited articles are from countries with a varying degree of home bias, from Denmark, the country with the lowest home bias of 0.018, to China, the country with the largest home bias of 0.376. Chinese PIs (that is, investigators located in mainland China) account for almost a fourth of the sample. Other large countries include Germany (17.2%), Japan (14.2%), UK (7.6%) and Switzerland (5.3%). We also highlight Danish PIs (1.2%), because Denmark is the only country which simultaneously contributes a non-trivial number of articles to the sample and exhibits an especially low home bias (home bias of 0.018).[iii]

Interesting descriptive patterns also emerge at the cited/citing pair level in Table C.2. 14.1% of citations are topically related to their source articles in the sense of PMRA, while 22% of citations in the sample come from the cited PI's home country.[iv] The average degree of home bias for the countries of the cited articles is 14.8%, while a little over 5 years separate the publication dates of the source and citing articles, on average.

Turning to column 1 in Table C.3, we find that home citations in a country with no home bias tend to be more PMRA-related than foreign citations—this effect is worth approximately 0.24 of a standard deviation of the outcome variable. However, relatedness is smaller for home-citations in home-biased countries, as the interaction term is negative and significant. To give a sense of the magnitude of this effect (since home bias is a continuous variable), the results imply that endowing China (the most home biased country in our data, $HomeBias_{CN} = 0.376$) with Denmark's level of home-bias (Denmark being the least home-biased country in our data, $HomeBias_{DK} = 0.018$) would roughly double the likelihood that a home citation is related. In column 2 we run an even more demanding specification by including cited article fixed effects (which causes journal effects, cited publication year effects, as well as the home bias indicator variable to drop). Comparing citations to the same article, the results hold and magnitudes are only slightly attenuated. In addition to cited article effects, the next three columns add to the specification citing journal by year interaction effects (column 3), citing country by year interaction effects (column 4), and finally both citing journal by year and country by year interaction effects (column 5). The magnitude and sign of $\hat{\beta}_3$ remains stable as we saturate the regression model with more and more granular fixed effects.

In Figure C.1, we modify the specification in equation (C.1) by replacing home bias with a full set of country fixed effects, and a full set of interaction terms between the country effects and the home citation indicator variable. We omit from the specification the main effect and interaction effect for France, which is a large country with the median level of home bias in our sample of 28 countries ($HomeBias_{FR} = 0.065$). We graph the interaction effects between Home and the country effects for five countries: China, Japan, Germany, the UK, Switzerland, and Denmark.

For this set of countries and sample, China is the only country for which the interaction effect is negative (the sum of the main home citation effect and the interaction effect would still be positive, although not statistically different from zero). In short, our results lend plausibility to the conjecture that home citations

[iii]It should be noted that our data excludes US-based PI articles, because they were collected for a project where citations *from* US scientists was the outcome of interest (Qiu et al., 2023).

[iv]Recall that we constrain all the citing addresses on a citing paper to belong to a single country so that every citation pair is either foreign or domestic.

reflect a less deep intellectual engagement with the subject matter of the cited work, relative to foreign citations, especially for research originating from countries exhibiting a large home bias.

Table C.1: Summary Statistics, Elite PIs (N= 74,216 Cited Articles)

Variable	Mean	Median	SD	Min	Max
Cited Year	2011	2011	4.715	2000	2018
No. of Citations	12.868	7	26.19	1	2,414
Journal Impact Factor	5.271	4.777	3.814	0	43.28
Home Bias	0.156	0.104	0.127	0.018	0.376
China	0.241	0	0.427	0	1
Germany	0.172	0	0.378	0	1
Japan	0.142	0	0.349	0	1
UK	0.076	0	0.265	0	1
Switzerland	0.053	0	0.225	0	1
Denmark	0.007	0	0.075	0	1

Notes: Home bias corresponds to the home bias for the cited article's country. A cited article is assigned to unique country through the address and institutional affiliation of its last author, who is in all cases an elite chemist part of a sample of 751 well-published scientists.

Table C.2: Summary Statistics, Elite PIs' (N=955,029 Citation Pairs)

Variable	Mean	Median	SD	Min	Max
Related Citation	0.141	0	0.348	0	1
Home Citation	0.220	0	0.414	0	1
Home Bias	0.148	0.104	0.125	0.018	0.376
Cited Year	2009	2010	4.625	2000	2018
Citing Year	2015	2015	4.470	2000	2022
Cited-Citing Year Lag	5.253	4	3.880	0	22

Notes: Home bias corresponds to the home bias for the cited article's country. A cited article is assigned to unique country through the address and institutional affiliation of its last author, who is in all cases an elite chemist part of a sample of 751 well-published scientists.

Table C.3: Home Citations and Topical Relatedness (Linear Probability Model)

	(1)	(2)	(3)	(4)	(5)
Home Citation	0.0844**	0.0729**	0.0703**	0.0677**	0.0667**
	(0.0032)	(0.0028)	(0.0027)	(0.0030)	(0.0029)
Home Bias	0.0304**				
	(0.0077)				
Home Citation × Home Bias	-0.1213**	-0.0928**	-0.0886**	-0.0733**	-0.0818**
	(0.0119)	(0.0106)	(0.0105)	(0.0118)	(0.0117)
Mean of Dependent Variable	0.141	0.141	0.142	0.141	0.142
s.d. of Dependent Variable	0.348	0.348	0.349	0.348	0.349
Adjusted R^2	0.045	0.144	0.149	0.142	0.150
No. of Investigators	751	751	751	751	751
No. of Articles	74,201	66,790	66,666	66,798	66,652
Cited Journal FE	yes				
Cited Year FE	yes				
Cited Article FE		yes	yes	yes	yes
Citing Journal FE	yes	yes			
Cited-Citing Year Lag FE	yes	yes	yes	yes	yes
Citing Journal × Citing Year FE			yes		yes
Citing Country × Citing Year FE				yes	yes

Notes: The dependent variable is an indicator variable that equals 1 if the citation is a related citation, and 0 otherwise. Standard errors in parentheses are two-way clustered at the cited article and citing article level. Home bias corresponds to the home bias for the cited article's country. A cited article is assigned to unique country through the address and institutional affiliation of its last author, who is in all cases an elite chemist part of a sample of 751 well-published scientists.

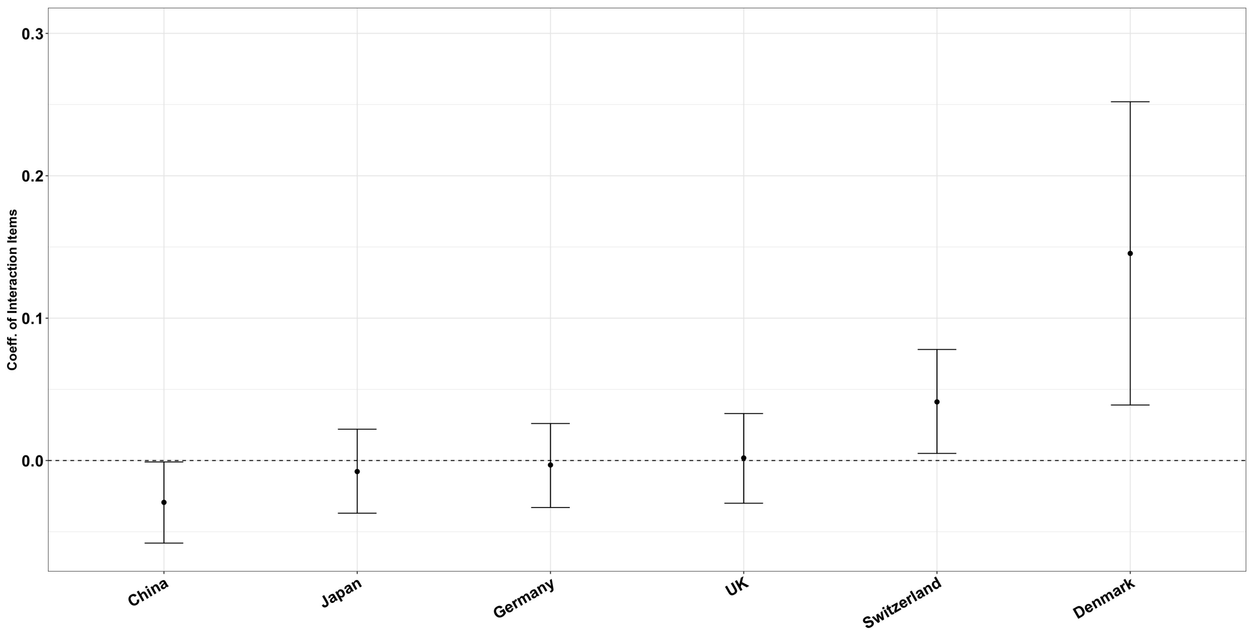

Figure C.1: Home Effect on Related Citations, by Country

Notes: Dots correspond to the coefficients for the interaction terms between the home citation indicator variable and a set of country fixed effects (where each cited article's PI is assigned to a single country). The estimates stem from linear probability model specifications in which the related citation indicator is regressed onto the home citation indicator variable, cited journal fixed effects, citing journal fixed effects, cited year and cited-citing year lag fixed effect, PI country fixed effect, as well as 27 interaction terms between the home indicator and PI country indicators. As France's home bias is the median (0.065) in the country sample, we took out France and its interaction as the reference category. We create a ROW category for small countries including Norway, South Africa, Russia and Finland. We plot the interaction coefficients for China, Japan, Germany, the UK, Switzerland, and Denmark. China (with the largest home bias 0.376), Japan (0.114), Germany (75 percentile of home bias 0.104), the UK (0.070), and Switzerland (0.047) are the top 5 countries in terms of cited article size. Denmark is the country with the smallest home bias (0.018). The 95% confidence interval (with robust standard errors, two-way clustered at the cited article and citing article level) around these estimates is depicted by the horizontal bars.

References

Aksnes, Dag W., Liv Langfeldt, and Paul Wouters. 2019. "Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories." *SAGE Open* 9:2158244019829575.

Bollen, Johan, Marko A. Rodriquez, and Herbert Van de Sompel. 2006. "Journal Status." *Scientometrics* 69:669–687.

Catalini, Christian, Nicola Lacetera, and Alexander Oettl. 2015. "The Incidence and Role of Negative Citations in Science." *Proceedings of the National Academy of Sciences* 112:13823–13826.

Lin, Jimmy and W. John Wilbur. 2007. "PubMed Related Articles: A Probabilistic Topic-based Model for Content Similarity." *BMC Bioinformatics* 8:1–14.

Qiu, Shumin, Claudia Steinwender, and Pierre Azoulay. 2023. "Who Stands on the Shoulders of Chinese (Scientific) Giants? Evidence from Chemistry." NBER Working Paper #30772.

Teplitskiy, Misha, Eamon Duede, Michael Menietti, and Karim R. Lakhani. 2022. "How Status of Research Papers Affects the Way They Are Read and Cited." *Research Policy* 51:104484.

Yin, Chun-Yang. 2011. "Do Impact Factor, h-index and Eigenfactor of Chemical Engineering Journals Correlate Well with Each Other and Indicate the Journals' Influence and Prestige?" *Current Science* 100:648–653.