Who Stands on the Shoulders of Chinese (Scientific) Giants? Evidence from Chemistry*

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

Claudia Steinwender
MIT & NBER
Sloan School of Management
100 Main Street – E62-521
Cambridge, MA 02142

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

April 5, 2021

Abstract

In recent decades, Chinese researchers have become preeminent contributors to the scientific enterprise, as reflected by the citation impact of publications originating from Chinese research institutions. But are there frictions specific to the diffusion of scientific knowledge originating from China? Focusing on elite chemistry researchers, we assemble a sample of articles by Chinese and non-Chinese PIs, carefully matched on "quality." We find that relative to non-Chinese, non-US PIs, Chinese PIs' articles receive 32% fewer citations from US researchers on average. This discount vanishes for the articles of scientists who received the entirety of their training in the US, and is twice as large in subfields afflicted by a large number of retraction scandals. Our results imply that US researchers do not build as readily on the work of Chinese researchers, relative to the work of other foreign scientists, even in a setting where Chinese scientists have long excelled.

*All authors contributed equally. Address all correspondence to pazoulay@mit.edu.

1 Introduction

In recent decades, China has become a preeminent contributor to the scientific enterprise, as reflected in the number of publications that originate from China, and the citation impact of these publications (Xie and Freeman, 2019). Even with the acknowledgement that (scientific) quantity might have a quality of its own, interpreting this dramatic increase is difficult. From the standpoint of its impact on the global economy, an important question is whether, beyond its undeniable quantitative importance, Chinese research contributes to pushing the world scientific frontier outward.

Recent empirical findings lend credence to the view that the quality of Chinese research has improved in concert with the number of articles emanating from Chinese research institutions (and researchers). For instance, the incidence of Chinese addresses (and Chinese names) in world-leading journals such as *Science* and *Nature* has more than doubled between 2000 and 2016 (Xie and Freeman, 2020). The average number of citations per article, and China's overall share of citations has also risen markedly.

These stylized facts notwithstanding, the extent to which Chinese scientific knowledge offers "broad shoulders" for follow-on researchers to stand on remains an open question. In particular, how are citations to Chinese research geographically distributed? The last twenty years have seen a 2.5 fold increase in the number of Chinese academic scientists (PRC National Bureau of Statistics, various years), many of them working in relatively new, less research-intensive institutions. Because of this increase in scientific labor supply, the rising impact of Chinese research could merely reflect an elevated propensity on the part of Chinese researchers to cite research "made in China" (Qiu et al., 2021). This could arise because of the localized nature of knowledge spillovers, or because of other frictions, such as lower communication costs for researchers who share the same language. Conversely, foreign scientists might discount the importance of Chinese scholarship, compared to research produced in their home countries or elsewhere in the world.

Contrasting Chinese and non-Chinese (and non-US) researchers, we study the extent to which articles of similar observable quality are differentially likely to be cited by researchers based in the US. Our preferred specifications point to a "China citation discount" equal to 32% of the baseline probability of citation. This discount vanishes completely for Chinese researchers who received the entirety of their scientific training in the United States, and affects disproportionately articles in the middle of the impact distribution. We also find evidence that this discount is not a mere reflection of clustering of Chinese researchers in particular subfields. Nor is it likely to reflect ethnic animus, since we do not observe a similar discount for researchers with Chinese names located outside China. In contrast, we do find that the discount is even higher in subfields of chemistry that suffer from abnormally high retraction rates.

These results are notable because our choice of setting—elite scientists, in a domain where China has a long tradition of excellence—would seem to be one without particular impediments to the diffusion of knowledge across borders. Yet this appears to be far from the case. Together, our results imply that China's pronounced citation "home bias" reflects, at least in part, missing citations from non-Chinese authors, perhaps offset by a surplus of citations attributable to the vastly expanded pool of Chinese potential citers when aggregating citation data at the level of a field, a journal, or an entire country.

The paper proceeds as follows. We begin by a brief history of Chemistry research in China. Section 3 describes our data sources and sample. Section 4 reports the results of the analysis on the matched sample of articles. Section 5 concludes.

2 Chinese Research in Chemistry

For thousands of years ancient China led the world with remarkable inventions and achievements in the chemical arts (Agnew, 1997). Many important empirical discoveries and knowledge of metallurgical arts originated from ancient Chinese alchemy and medicinal chemistry; their translation into Western languages had a pronounced influence on modern chemical science (Leicester, 1971). For example, many historians believe that gunpowder technology, one of the most influential inventions in human history, had its origins in China (581-681 A.D.), and then spread to the Middle East and Europe along the Silk Road (Needham et al., 1986). Chinese pre-modern "scientists" also pioneered the manufacturing processes for salt, wine, paper, and porcelain (Li, 1948; Needham and Tsuen-Hsuin, 1985).

Although historians suggest that modern chemistry grew, at least in part, out of the work of Chinese alchemists (Leicester, 1971), chemistry as a modern science was absent in China until the 19^{th} century, when European science were introduced through missions, trade, and wars (Li, 1948). In the late Ch'ing Dynasty (mid-to-late 19^{th} century), during which the rulers adopted a closed-door policy with very limited communication with the outside world, Chinese chemistry (as well as other sciences) lagged far behind western countries. After wars with European countries had broken China's door open, modern chemistry started to develop with the purpose of "learning from foreigners to compete with them" as China became integrated into the global "Republic of Science" (Bai, 2000). Research by western chemists were intensively translated into Chinese and disseminated in China. According to bibliographic statistics (Beijing Library, 1986), between 1912 and 1949, 41% of chemistry articles and textbooks were translated from English, while the remainder (very few of them were original scientific research) were written by Chinese chemists.

In order to further acquire frontier knowledge, a first wave of Chinese students were sent to the United States for scientific training under the aegis of the Boxer Indemnity Scholarship Program. The first generation of returnee students had a lasting influence on Chinese modern science and some of them became pioneers and academic leaders in the field of Chemistry after coming back to China.[1]

The rapid development of modern Chinese chemistry took place after the founding of the People's Republic of China, especially after the deep opening policy begun in 1978 (Bai, 2000). Between 2000 and 2017, the number of Chinese universities increased by 140% (from 753 to 1,805), and correspondingly the number of chemistry departments rose by 182% (from 243 to 686). Research faculty in Chinese universities increased by 69% during the same period, and the number of chemistry researchers tripled. Public research funding invested in Chemistry also shows a 14-fold increase between 2000 and 2017, higher than the ten-fold increase observed for other fields on average.[2]

Meanwhile, China has continuously expanded global collaboration and communication by funding students' graduate studies abroad, and facilitating Chinese scholars' participation in international collaboration through the funding of shorter-term stays in frontier countries. The number of state-financed students studying abroad increased five-fold, from 7,564 in 2000 to 46,347 in 2017, whereas attendance of international conferences increased almost eight fold during the same period. Between 1978 and 2018, a total of 5.86 million students studied abroad, 82% of whom returned to China. The flow of transnational human capital, particularly the return of elite scientists, has helped create a solid foundation for Chinese scientific research. Scientists holding overseas degrees account for 37% of the total number of members of Chinese Academies of Sciences and Engineering elected between 1955 to 2009.

[1]In 1908, US and China reached an agreement to use the excess funds from the Boxer Indemnity to establish a scholarship program for Chinese students to study in the US. During 1909 to 1929, this program sent around 1,300 Chinese students to the US, studying in several selected fields that serve to the urgent need for Chinese development, such as Science, Engineering, Medicine, and Agriculture. Celebrated alumni of this program include Hou Debang [BA, MIT, MS, Columbia], Chen Hwang [BS, MS, MIT], Chang Tsun [BA, MIT], Hsu Paul Hwang [BS, MIT], and Chien Shih-Liang [MS, PhD, UIUC]).

[2]The source for these figures, and those mentioned below is the *Compilation of University Science and Technology Statistics* produced by the Chinese Ministry of Education.

During this period, 300 US-trained academics returned to China, a figure to be compared with 160 Soviet-trained and 80 UK-trained academics who returned during the same period.[3] In Chemistry specifically, there is evidence that students receiving graduate training in the United States are among the best and brightest. Gaulé and Piacentini (2013) document that Chinese students perform about as well as the awardees of the prestigious NSF doctoral fellowship program, and far better than other foreign students.

While China has been a rising star across a broad cross-section of scientific domains, its status as a producer of frontier scientific knowledge has stood out in a narrower set of fields, Chemistry preeminent among them. According to *Nature Index*, a database consisting of research articles published in an independently selected group of high-quality science journals, China's fractional count of articles grew by 84% between 2012 and 2017, making the country second to only the United States. In some Chemistry subfields, such as organic chemistry, China even surpassed the United States in recent years to become the world's top producer of publications.

Table 1 demonstrates the importance of Chinese elite researchers. According to the annual *Highly Cited Researchers (HCRs)* rankings published by *Clarivate Analytics* between 2014 and 2018, Chemistry ranks highest among scientific fields in terms of highly cited researchers (column 1). These 211 researchers account for 19.27% of the world's HCRs in Chemistry (column 2).

Table 2 shows that Chinese chemists have become world-leading contributors compared to other countries. During 2000 to 2015, China's share of publications in Chemistry was 14.96%, ranking it second only after the United States. Japan is a distant third with 7.66%, followed by Germany, India, and the United Kingdom. The ranking with respect to HCRs is similar, with the United States accounting for the largest share of the world's elite chemists (43.01%), followed by China (19.27%).

China's strong position in Chemistry becomes particularly striking when we compare it to China's ranking with respect to all fields. Across the sciences, the United States are clearly the most dominant nation with 8,306 HCRs amounting to almost half (46.48%) of the world's top scientists. Second by a large margin is the UK with 1,701 HCRs (9.52% of the world), and China is third with 1,104 HCRs (6.18% of the world). This makes China's HCR share in Chemistry more than three times larger than the average across fields.

[3]Source: *Survey Report of Academicians of Chinese Academy of Sciences and Chinese Academy of Sciences.*

3 Description of Data Sources

The goal of this paper is to investigate how research undertaken in China disseminates compared to research undertaken by other countries. To do this, we focus on the publications of the world's best researchers, as defined by the lists of *Highly Cited Researchers* in Chemistry.[4] HCRs are defined as the top 1% of researchers of the field with respect to the citations from all of their publications in *Web of Science*. We focus on HCRs from all countries, excluding the United States.[5]

Researcher level data. In order to obtain individual characteristics for the HCRs, we collected the curriculum vitae (CV) of each scientist through either their laboratory or institution faculty page, their *Who's Who* profile, Google searches, or e-mail requests if no online information was available. This search yielded CV information for 381 HCRs. From the CVs, we extract information about demographics (birth year, gender), PhD education (university, country, completion year), post-doctoral experience (organization and time period), as well as employment spells since post-doc (organization, country, and time period). We define the "year of independence" of each researcher as the year of their first faculty employment after post-doctoral education. We also use the country of this first non post-doctoral affiliation to assign each HCR to a unique country. Note that this may not be the nationality of the HCR; instead, this should be interpreted as the country in which the first independent research is undertaken.[6] Overall, the majority of HCRs are male (97%), and 73% of HCRs have some postdoctoral experience. Their average doctoral degree year is 1987, while the average year of independence is 1990 and the average number of post-doctoral years is 2.13. 12% of HCRs hold a PhD degree from universities located in the United States, and 45% of them spent their post-doc years at institutions in the United States.

[4]These lists were compiled by *Clarivate Analytics*, an information services provider who owns and operates the citation database *Web of Science*. The HCR list was first published in 2001; then again in 2014, and annually updated since then.

[5]As explained in more detail below, we will consider the US a "neutral territory" whose researchers are at risk of citing articles written by Chinese and non-Chinese scientists.

[6]HCRs may move to different countries over the course of their career, and we worry about the selectivity bias that may arise from assigning each publication to the country in which it was produced, since opportunities to move could plausibly be related to productivity. In practice, this distinction does not matter very much: For 95% of publications the country of the first faculty employment is the same as the country of the affiliation at the time of publication, and our results are robust to this alternative definition.

Publication data. We compile the full publication list for all 381 HCRs in our sample published between the years 2000 and 2015.[7] To ensure that we capture only knowledge creation that was influenced to a significant extent by the HCR, we restrict the publications in two ways. First, we focus on publications which list the HCR as last author, which indicates the principal investigator according to publication norms in the field of Chemistry. Second, we consider only articles that were published after the HCR became an "independent" researcher, according to the definition above. Overall, our sample comprises 40,906 scientific articles in Chemistry. On average, each HCR published 107.5 articles as last author in the time period we consider (*std. dev* = 92.86), ranging from a minimum of 3 to a maximum of 527 publications.

Citation data. We compile a list of citations of the publications to our HCR sample from Web of Science. Since we want to link citations to countries, we remove citing articles lacking country information (4.2% of citations), which results in our database comprising of 1,905,490 citation records from 2000 to 2015 for the 40,906 HCR last-authored articles.[8] Each article in our data set received on average 46.58 cites.

4 HCR Publications Cited by US Researchers

To uncover the causes of differences in cross-country citation behavior, we focus on the propensity of US researchers to cite articles that originate from China versus other countries. We single out the US as a "neutral territory" for two reasons. First, the US is undoubtedly a frontier country in Chemistry research, and a pole that attracts collaborations and trainees from the world at large. Second, its large size implies that citation linkages between the US and other countries are frequent enough to make the statistical analysis tractable. In order to isolate the role of cited papers' country of origin, we match each publication by a Chinese HCR to a similar publication by a non-Chinese, non-US HCR.

[7] Section **??** in the Online Appendix provides more details about how we link scientists to journal articles to ensure accurate matches. We restrict our sample to publications after 2000, because there were very few Chinese HCRs active before 2000.

[8] Note that *Web of Science* does not always assign affiliations to specific authors, which is why we cannot focus on the address of a specific author, e.g., the last one mentioned in the list, as we do for the HCR publications. 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.

4.1 Matched Sample of Articles

We refer to publications of Chinese HCRs in our sample as the treatment group.[9] For each of these articles, we are looking for at least one similar article among non-Chinese, non-US HCRs, the control group. We implement a "Coarsened Exact Matching" (CEM) procedure (Blackwell et al., 2009). The first step is to select a relatively small set of covariates on which we need to guarantee balance *ex ante*. This choice entails judgement, but is strongly guided by our desire to hold the "fertility" of cited papers approximately equal across the treatment and control groups. The second step is to create a large number of strata to cover the entire support of the joint distribution of the covariates selected in the previous step. In a third step, each observation is allocated to a unique strata, and for each observation in the treated group, control observations are selected from the same strata.

The procedure is coarse because we do not attempt to precisely match on covariate values; rather, we coarsen the support of the joint distribution of the covariates into a finite number of strata, and we match a treated observation if and only if a control observation can be recruited from this strata. An important advantage of CEM is that the analyst can guarantee the degree of covariate balance *ex ante*, but this comes at a cost: the more fine-grained the partition of the support for the joint distribution (i.e., the higher the number of strata), the larger the number of unmatched treated observations.

Our list of matching covariates includes a single researcher-level variable: an indicator variable for receipt of post-doctoral training in the US. At the article level, we match exactly on the journal, the publication year, and coarsely on the number of authors (1-3; 4-6; 7-9; 10 or more coauthors) and the number of citations from articles originating outside the US ($0-25^{th}$ percentile; $25-50^{th}$ percentile; $50-75^{th}$ percentile; $75-95^{th}$ percentile; $95-99^{th}$ percentile; and top percentile).[10] The union of all matching criteria defines a strata. Within each strata, papers are indistinguishable from the perspective of the CEM algorithm, and the matching is performed at the level of the strata.[11]

This procedure yields 1,855 treated articles written by 89 Chinese HCRs, and 2,902 control articles written by 219 non-Chinese HCRs. Table 3 provides descriptive statistics at the individual level. There is a small share of female researchers (4.5% in the treatment

[9]We use HCR publications between 2000 and 2012 for this exercise, to give each HCR article at least three years to be cited (our publication database ends in 2015).

[10]When creating these bins, we compute a separate empirical distribution of citations for each year between 2000 and 2012.

[11]As there may be different numbers of treated and control articles in different strata, CEM assigns a weight to each matched article to adjust for strata size, and we use this weight in all regressions models.

group, 1.8% in the control group); Chinese and non-Chinese scholars spent a similar time of their pre- and post-doctoral education in the United States. However, Chinese HCRs are on average ten years younger (possibly reflecting the relatively recent "rise" of Chinese science), and therefore have published less publications and garnered less overall citations than their non-Chinese counterparts.

Table 4 compares the characteristics of control and treated articles. On average, there are 1.56 control articles per treated article. By construction, the two sets of articles were published in the same year, have on average about 5.6 authors, and the same number of citations from outside the US. At the researcher level, differences subsist in spite of the fact that at the article-level, the distribution of citations received from non-US countries are balanced across treated and control articles, as Figure 1 indicates. In order to not shrink the size of the article sample any further, we do not match on individual-level differences in observed achievement (career publications or citations). However, our regression specifications will include these individual-level covariates as controls. The combination of matching and covariate inclusion results in comparisons that flexibly and plausibly hold "fertility" constant across Chinese and non-Chinese publications.

4.2 Potential Citers from US

To test whether articles in the control or treatment group are cited differentially by US authors, we first need to determine which US articles are *at risk* of citing the articles by the HCRs, not just the articles that correspond to *actual* citations. Moreover, since we would like to evaluate how social or geographic proximity shapes the propensity to cite, it is crucial that participation in the risk set not be mechanically influenced by such factors. We deemed an article eligible to be part of the citation risk set if it fulfills the following three criteria: (i) all authors are affiliated with a US institution; (ii) it was published after the cited HCR paper; and (iii) it is topically related to the HCR paper.

Of these three criteria, the last one is the most challenging to implement in practice. We rely on the fact that most of the articles in the HCR sample (and most of the citations to these articles) appear in journals indexed by *PubMed* in addition to the *Web of Science*.[12] We then use the "Related Articles" function in *PubMed* to harvest journal articles that are intellectually proximate to the HCRs' own papers. This functionality is based on a topic-

[12] *PubMed* is an online resource from the National Library of Medicine that provides free and comprehensive access to the biomedical research literature, indexing more than 40,000 journals within the life sciences, including almost all the journals in which HCRs routinely publish.

based content similarity model called *PubMed Related Citations Algorithm* or *PMRA* (Lin and Wilbur, 2007). This algorithm yields relatedness rankings and scores between any two articles based on the extent to which two articles are similar with respect to titles, abstracts, and keywords.[13] The *PMRA* algorithm is designed to estimate the conditional probability that a researcher would be interested in another article, given her interest in a given article. For each HCR article, its citation risk set includes every PMRA neighbor whose authors work in US institutions and appeared after the focal article was published. Of the 23,551 US articles actually citing the 4,757 HCR articles in the matched sample, only 1,877 (8.0%) correspond to related records in the sense of PMRA. Importantly, the risk set does not include actual citations that are PMRA-unrelated.[14]

The combined risk set of the 4,757 HCR articles in the matched sample comprises 72,550 citable/potentially citing article pairs, with each HCR article having on average 15.25 potentially citing articles in its risk set.

4.3 Model specification

We model the probability that HCR article i is cited by each paper $j \in J_i$, the risk set of article i, as a function of the characteristics of article i and article pair ij, using the following linear probability model:

$$\mathbb{1}_{(j\,\text{cites}\,i)} = \beta_0 + \beta_1 China_i + \beta_2 X_{ij} + \beta_3 China_i \times X_{ij} + \varphi(i,j) + \varepsilon_{ij} \quad (1)$$

The dependent variable is an indicator variable that takes on value 1 if paper j *actually* cites paper i, and 0 otherwise. Our main regressor of interest, $China_i$, is an indicator variable for whether i's last author is Chinese, whereas X is a vector of control covariates, and $\varphi(i,j)$ correponds to a large set of fixed effects for i, j, and $i \times j$ characteristics. These include fixed effects for the interaction of i and j publication years; fixed effects for each strata defined in the coarsened exact matching algorithm; fixed effects for the HCRs' highest degree years, and an HCR gender indicator variable. The specifications also include PMRA rank bins for

[13] To facilitate the harvesting of *PubMed*-related records on a large scale, we have developed an open-source software tool that queries *PubMed* and PMRA and stores the retrieved data in a MySQL database. The software is available for download at `http://www.stellman-greene.com/FindRelated/`. Prior research leveraging the intellectual linkages between articles generated by PMRA include Azoulay et al.(2015), Azoulay et al. (2019), and Myers (2020). Appendix C in Azoulay et al. (2019) describes the algorithm in detail.

[14] In a robustness test, we have verified that leaving these unrelated citations in the risk set does not alter our substantive conclusions.

each ij article pair.[15] We do not report coefficient estimates for these covariates, but they are always included.

HCR scientist-level covariates. The covariates defined at the level of the individual HCR include: (i) an indicator variable for HCR countries (as defined in section 3) that list English among their official languages; (ii) two indicator variables denoting whether HCRs have obtained their PhD degree in the US, or have both a PhD degree as well as post-doctoral experience in the US; (iii) the cumulative number of publications (respectively cumulative citations received) for each HCR up to the year *before* article i was published.[16] Recall that the coarsened exact matching procedure did not yield balance on many investigator-level covariates. Including them in the specifications alleviates the concern that unobserved differences in "quality" explain the results we present below.

HCR article covariates. The covariates defined at the source article level i include: (i) the number of unique countries (excluding the US) that appear among the affiliation lists of all coauthors (this variable is a proxy for the cosmopolitan character of the research team); (ii) an indicator variable equal to 1 whenever the reprint or first author has a US affiliation;[17] (iii) An indicator variable for whether any middle author has a US affiliation.[18]

Subfield covariates. To the extent that researchers in certain countries concentrate in different subfields, it is important for the analysis to control for these country-level specialization patterns.[19] Rather than assigning each source article to a subfield arbitrarily, we rely on the PMRA tool described earlier and define the subfield of each source article as the

[15] We group similarity ranks between articles into 26 bins, with finer bins for rankings below 150 (10 ranks per bin), and gradually create larger bins for higher ranks (151-170, 171-190, 191-210, 211-230, 231-260, 261-300, 301-350, 351-450, 451-650, 651-1000, above 1000).

[16] To avoid making functional form assumptions, we create $13 \times 2 = 26$ indicator variables each capturing whether the cumulative number of publications (respectively citations) to date belongs to a particular quantile bin. The corresponding coefficients are not reported.

[17] Recall that only the last author is constrained to be a non-US HCR researcher. First- and middle-authors in articles published by non-US HCRs can be affiliated with US research institutions, and it is plausible that such coauthorships elevate the propensity of citation by other US-based researchers. In order to assign countries to specific authors, we need to link the address lines listed on papers to authors; this is possible for 80% of the articles in our database. For these articles we find that in most cases (95%), the first authors are linked to the first address record. Therefore, for the remaining 20% of articles for which we cannot accurately assign countries to specific authors, we use the first address line to define the first author's country for this exercise.

[18] According to publication conventions in Chemistry, first or reprint authors have contributed significantly to the research undertaken.

[19] The existence of such patterns is not mere speculation on our part. For instance, Borjas and Doran (2015) document the persistence of Russian influence in certain mathematical subfields even after the dissolution of the Soviet Union.

set of its PMRA-neighbors, counting only the neighbors whose similarity score is above 0.5 and appeared before the source article. Using these PMRA-derived subfields, we construct three subfield-level covariates: (i) the subfield's *home-research intensity* corresponds to the sum of the PMRA-relatedness scores for the articles in the subfield whose researchers are from the HCR's country; (ii) the subfield's *foreign-research intensity* corresponds to the sum of the PMRA-relatedness scores for the articles in the subfield whose researchers are not from the HCR's country; and (iii) the subfield's *retraction intensity* captures the number of neighbors that have been retracted, have been the object of an "expression of concern," or are associated with an erratum.[20]

Article-pair covariates. The covariates defined at the level of the cited-citing article pairs ij include: (i) the geographic distance between the city of the HCR, and the city of the affiliations of the US authors.[21] (ii) an indicator variable that switches to 1 if both articles were published in the same journal; (iii) an indicator variable for the presence of a common author on the authorship rosters of articles i and j; (iv) an indicator variable for the presence of a past coauthor of the HCR on article j's authorship roster; and (v) an indicator variable that captures shared ethnicity between the HCR and at least one author from article j.[22]

We cluster standard errors simultaneously at the level of the individual HCR—to allow for arbitrary correlation of citation patterns across publications within an individual HCR—and the level of a strata—to allow for correlation of citation patterns across publications within a strata (Cameron and Miller, 2015).

4.4 Empirical Results

Table 5 reports the estimation results corresponding to equation (1). Column 1 only includes the Chinese HCR indicator variable, column 2 adds controls for characteristics that

[20]Because these events are rare, in this case we do not apply the 0.5 cutoff for these articles' relatedness score. In a robustness check, we found that imposing the cutoff weakens the precision of the corresponding coefficient estimate, but does not change its magnitude. Although we only count retractions of articles that were published before the source, the retraction event can occur either before, in the same year, or after the publication of the source.

[21]Specifically, we compute the log average geographical distance between the HCR city of affiliation and all cities listed in the citing article j.

[22]To identify the ethnic origin of US-based scholars, we map a scholar's last name to its ethnic origins based on the algorithm developed by Nguyen (2019). Her novel algorithm computes the probability that a last name corresponds to a particular ethnic origin based on the de-anonymized full population samples of US Censii between 1910 and 1940. A last name may be mapped to multiple countries with different probabilities based on their relative frequencies. We use the country with the maximum probability to identify the ethnic origin of each last name. More details regarding the name-ethnicity mapping can be consulted in Nguyen (2019).

vary at the article, subfield, and investigator level, and column 3 adds control variables at the article-pair level. Across these specifications, we observe a statistically significant and negative "China effect": articles written by Chinese HCRs receive significantly fewer citations from US scientists than articles written by non-Chinese HCRs. The magnitude of the effect is empirically meaningful: Since the baseline probability of being cited by a US paper is low in our sample (2.5%), the probability of a Chinese-authored article being cited is 32% lower than the baseline probability (based on the estimates from column 3).

This effect is not explained by language. The coefficient on the English-speaking indicator variable for the HCR country is positive in columns 2 and 3, but tiny compared to the China effect and imprecisely estimated. Nor does it reflect education in the US (recall that we already matched articles on the basis of US postdoctoral training for the treated and control HCRs), country-subfield intellectual specialization patterns, or the presence of US coauthors on the cited paper. We distinguish between US coauthors on the cited paper who share the HCR's ethnic background, versus US coauthors from a different ethnic background (inferred from last names using the technique proposed by Nguyen [2019]). While the signs of these effects are positive, they are both small and imprecisely estimated.

In column (3), we find that articles are significantly more likely to be cited by a publication that appeared in the same journal, one that lists a coauthor common to the cited and citing article, one that lists a past coauthor of the HCR on the authorship roster, or one that lists an author with the same ethnic background as the HCR (once again inferred from last names). These effects, however, attenuate only slightly the magnitude of the China citation discount, and the effect remains statistically significant.

In columns 4 through 7 we explore whether various types of social, spatial, or intellectual connections differentially affect articles by Chinese and non-Chinese HCRs. We do so by including in the specification interaction terms between these covariates and the Chinese HCR indicator variable.

Column 4 shows that HCR researchers' US citations benefit from having a US coauthor, especially one that share their ethnicity, but perversely this result does not seem to apply to Chinese HCR researchers. The publications of these scholars receive increased attention when they also include a US coauthor from a different ethnicity, but the effect is small and not statistically significant. One possibility is that potential citers discount such instances of collaboration because they suspect the presence of the US author on the authorship roster to reflect scientifically "impure" motives, such as the need to curry favor with editors of leading journals.[23]

Column 5 examines whether US-based researchers who have ethnic roots in China help diffuse Chinese research to the US, as suggested by recent research (Xie and Freeman, 2020). Although the main effect of citations from authors of the same ethnicity is positive and marginally significant in column 3, we do not find this effect to be particularly more pronounced for Chinese HCRs.

Column 6 shows that the discount experienced by Chinese scholars disappears entirely for those who received doctoral and postdoctoral training in the US (but not either one alone) and returned to China. In fact, their citations from the US are larger compared to HCRs from other countries who spent their PhD and post-doctoral education in the US and then returned home.

Because of the relatively high frequency of retraction scandals that have afflicted Chinese scientific teams (Liao et al., 2018; Huang, 2017), we speculate that non-Chinese scientists could deem knowledge and ideas that originate in China to be less reliable than those originating in other countries, leading researchers to cite Chinese research less heavily even when it would appear equally fruitful based on observable covariates. In column 7, we test this conjecture by interacting the Chinese HCR indicator with our measure of subfield retraction intensity. We do find evidence of a large additional citation discount imposed on Chinese articles that belong to subfields that are relatively more "retraction-heavy." And yet the magnitude of the Chinese HCR effect barely changes when adding to the specification this additional interaction term.

Column 8 allows for all interaction effects to enter the specification simultaneously, with similar results. Overall, these specifications point towards an obdurate citation discount experienced by articles published by elite Chinese chemists.

Finally, we ask whether the magnitude of the discount is modulated by the underlying quality of article, as assessed by the citations it received from non-US sources. We interact the China HCR indicator with indicator variables for the position of the HCR paper in the citation distribution of all other papers with the same publication year. We create 12

[23]This interpretation is quite speculative. Perhaps the most infamous case of "ghost authorship"—a fraudulent paper in the field of stem cell research—embroiled a South Korean team, not a Chinese one (Hwang et al., 2005).

percentile bins, allowing for more heterogeneity at the top of the distribution.[24] To allow for comparability of the estimates across percentile bins—which have a different baseline probability of being cited by US authors—we plot the coefficients measured in units of standard deviations. In Figure 2, we find a pronounced U-shape pattern: citations to the lowest-quality papers are unaffected by Chinese authorship. The China citation discount is predominantly driven by publications of middling impact, i.e., those that lie between the 40^{th} and 90^{th} percentile of the impact distribution. At the very top (above the 99^{th} percentile), the "China discount" turns into a "China premium."

4.5 Robustness checks

We document the existence of a discount in the rate of US citations received by Chinese researchers, relative to non-US researchers locating in other countries. However the choice of of China to define the treated group of articles is arbitrary. Would we find similar evidence of a discount if we chose to make researchers from other countries with a storied legacy of chemistry research pivotal?

To probe the extent to which the citation discount is specific to China, we replicate our analysis by making the articles from HCR researchers located in eight other countries the treated group (the articles of Chinese HCRs are eligible to participate in the matching that helps construct the control group). These eight countries are Germany (N=1,725 treated articles), Japan (N=1,152 treated articles), The Netherlands (N=376 treated articles), Canada (N=352 treated articles), France (N=323 treated articles), Singapore (N=291 treated articles), Switzerland (N=1,265 treated articles), and the United Kingdom (N=234 articles).

Using column (3) in Table 5 as a benchmark, we display the estimates and their associated confidence intervals in Figure 3.[25] We find that the country effects are not statistically significant from zero in all other cases. However, except for Germany and Japan (the only countries in this set who accounts for a broadly similar number of treated articles relative to China), the effects are quite imprecise owing to the small size of the set of treated articles.

[24]More specifically, we create percentile bins at the following cutoffs: 10^{th}, 20^{th}, 30^{th}, 40^{th}, 50^{th}, 60^{th}, 70^{th}, 80^{th}, 90^{th}, 95^{th}, and 99^{th} percentile.

[25]Each estimate stems from a separate regression, rather than a pooled regression, in order replicate faithfully the empirical design exploited in the Chinese case.

5 Conclusion

The inclusion of Chinese scientists in the global "Republic of Science" has gathered pace for the last two decades. An increasing body of evidence points to a gradual bridging of the gap that long existed between the impact of Chinese published scientific output and that of frontier countries (Xie and Freeman, 2019). Observers note—with a mix of awe and trepidation—that Chinese scientists are about to overtake US scientists in at least one domain: Artificial Intelligence (O'Meara, 2019).

Whereas the "quality view" stresses the broader shoulders provided by Chinese researchers for follow-on scientific developments (wherever they come from), the "spillover view" emphasizes the localized nature of much of the citations accruing to Chinese articles. The last twenty years have seen a 2.5 fold increase in the number of academic scientists (PRC National Bureau of Statistics, various years), many of them working in relatively new, less research-intensive institutions. Because of this increase in scientific labor supply, the rising impact of Chinese research could merely reflect an elevated propensity on the part of Chinese researchers to cite research "made in China." This could arise because of the localized nature of knowledge spillovers, or because of other frictions, such as lower communication costs for researchers who share the same language (Xie and Freeman, 2020).

Our study purposefully sidesteps this debate to shed light on the propensity to cite research emanating from Chinese scientists *holding quality constant*, by pairing Chinese and non-Chinese articles well matched on attributes that plausibly capture the scientific "fertility" of each publication. Focusing on elite researchers in a single domain, Chemistry, we uncover the existence of a sizable citation discount for Chinese articles, relative to non-Chinese articles. What explains the relative underciting of Chinese science by US scientists?

One possibility is that in spite of our best efforts, systematic differences in citation potential subsist between treated and control articles in our sample, even after carefully matching on journal and citations received from non-US sources. Another possibility is that US scientists are simply less aware of Chinese research, perhaps because Chinese scientists, even if they belong to the elite, have less access to the networks that provide broad exposure to research findings. Our evidence does not lend much credence to this hypothesis, however. For instance, the discount exists relative to other non-English speaking countries, and is not alleviated by the presence of an author with a Chinese name on the potentially citing paper.

Instead, the totality of the results we present point to differences in *perceived* quality between Chinese and non-Chinese articles. Further research is needed to unpack the precise mechanisms that lie behind these differences in citation behavior. One must entertain that it might reflect animus directed at Chinese scientists, but this hypothesis does not sit well with the evidence that the citation discount vanishes in the case of returnees who completed their scientific training in the US. More plausibly, the discount might reflect perceptions of lower reliability for Chinese-produced knowledge (Liao et al., 2018). These perceptions might arise due to the number of well-publicized cases of scientific misconduct in China (Huang, 2017). Azoulay et al. (2015) find that areas of science tainted by retraction scandals leads to an exodus from scientists in that field which likely corresponds to an "overcorrection." Accurate or inaccurate beliefs regarding the lower reliability of science produced in China is a plausible mechanism for the findings we report.

Is the China citation discount likely to be a transitory phenomenon? On the one hand, institutional efforts are under way in China to root out various aspects of corruption in scientific institutions, from fake journals (Mallapaty, 2020), to authorship-for-sale schemes (Hvistendahl, 2013), to endemic plagiarism (Zhang, 2010). On the other hand, the violation of scientific norms appears entrenched in the upper echelons of Chinese scientific institutions (e.g., Fisman et al. (2018)), and we might expect reform in this domain to be a slow process. Current US-China tensions, as well as the disruption of scientific travel induced by the COVID-19 pandemic, might further solidify the negative perceptions of foreign citers vis-à-vis research produced in China.

References

Agnew, Neville, *Conservation of Ancient Sites on the Silk Road*, Los Angeles, CA: MGetty Publications,

Azoulay, Pierre, Christian Fons-Rosen, and Joshua S Graff Zivin, "Does science advance one funeral at a time?," *American Economic Review*, 2019, *109* (8), 2889–2920.

_ , Jeffrey L Furman, Joshua L Krieger, and Fiona Murray, "Retractions," *Review of Economics and Statistics*, 2015, *97* (5), 1118–1136.

Bai, Chunli, "Development and Prospect of Chinese chemistry," *University Chemistry (in Chinese)*, 2000, *15* (2), 1–10.

Blackwell, Matthew, Stefano Iacus, Gary King, and Giuseppe Porro, "cem: Coarsened Exact Matching in Stata," *The Stata Journal*, 2009, *9* (4), 524–546.

Borjas, George J and Kirk B Doran, "Cognitive mobility: Labor market responses to supply shocks in the space of ideas," *Journal of Labor Economics*, 2015, *33* (S1), S109–S145.

Cameron, A Colin and Douglas L Miller, "A practitioner's guide to cluster-robust inference," *Journal of human resources*, 2015, *50* (2), 317–372.

Fisman, Raymond, Jing Shi, Yongxiang Wang, and Rong Xu, "Social ties and favoritism in Chinese science," *Journal of Political Economy*, 2018, *126* (3), 1134–1171.

Gaule, Patrick and Mario Piacentini, "Chinese graduate students and US scientific productivity," *Review of Economics and Statistics*, 2013, *95* (2), 698–701.

Huang, Echo, "China publishes more science research with fabricated peer-review than everyone else put together'," *Quartz, May*, 2017, *8.*

Hvistendahl, Mara, "China's publication bazaar," 2013.

Hwang, Woo Suk, Sung Il Roh, Byeong Chun Lee, Sung Keun Kang, Dae Kee Kwon, Sue Kim, Sun Jong Kim, Sun Woo Park, Hee Sun Kwon, Chang Kyu Lee et al., "Patient-specific embryonic stem cells derived from human SCNT blastocysts," *Science*, 2005, *308* (5729), 1777–1783.

Leicester, Henry Marshall, *The Historical Background of Chemistry*, North Chelmsford, MA: Courier Corporation,

Li, Chiao-Ping, *The Chemical Arts of Old China*, Easton, PA: Journal of Chemical Education,

Liao, Qing-Jiao, Yuan-Yuan Zhang, Yu-Chen Fan, Ming-Hua Zheng, Yu Bai, Guy D Eslick, Xing-Xiang He, Shi-Bing Zhang, Harry Hua-Xiang Xia, and Hua He, "Perceptions of Chinese biomedical researchers towards academic misconduct: a comparison between 2015 and 2010," *Science and Engineering Ethics*, 2018, *24* (2), 629–645.

Lin, Jimmy and W John Wilbur, "PubMed related articles: a probabilistic topic-based model for content similarity," *BMC Bioinformatics*, 2007, *8* (423), 1–14.

Mallapaty, Smriti, "China's research-misconduct rules target'paper mills' that churn out fake studies.," *Nature*, 2020.

Myers, Kyle, "The elasticity of science," *American Economic Journal: Applied Economics*, 2020, *12* (4), 103–34.

Needham, Joseph and Tsien Tsuen-Hsuin, "Volume 5.1 - Chemistry and Chemical Technology: Paper and Printing," in Joseph Needham, ed., *Science and Civilization in China*, Vol. 5.1 of *Science and Civilization in China*, Cambridge University Press, England, 1985, pp. 132 – 184.

_ , Ping-Yü Ho, Gwei djen Lu, and Ling Wang, *Military technology: the gunpowder epic*, Cambridge University Press, 1986.

Nguyen, Kieu-Trang, "Trust and Innovation within the Firm: Evidence from Matched CEO-Firm Data," 2019. Working Paper.

O'Meara, Sarah, "Will China lead the world in AI by 2030," *Nature*, 2019, *572* (7770), 427–428.

Qiu, Shumin, Claudia Steinwender, and Pierre Azoulay, "Home Bias: China's Trade in Science," 2021. Working Paper.

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

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

Zhang, Yuehong, "Chinese journal finds 31% of submissions plagiarized," *Nature*, 2010, *467* (7312), 153–153.

Figures & Tables

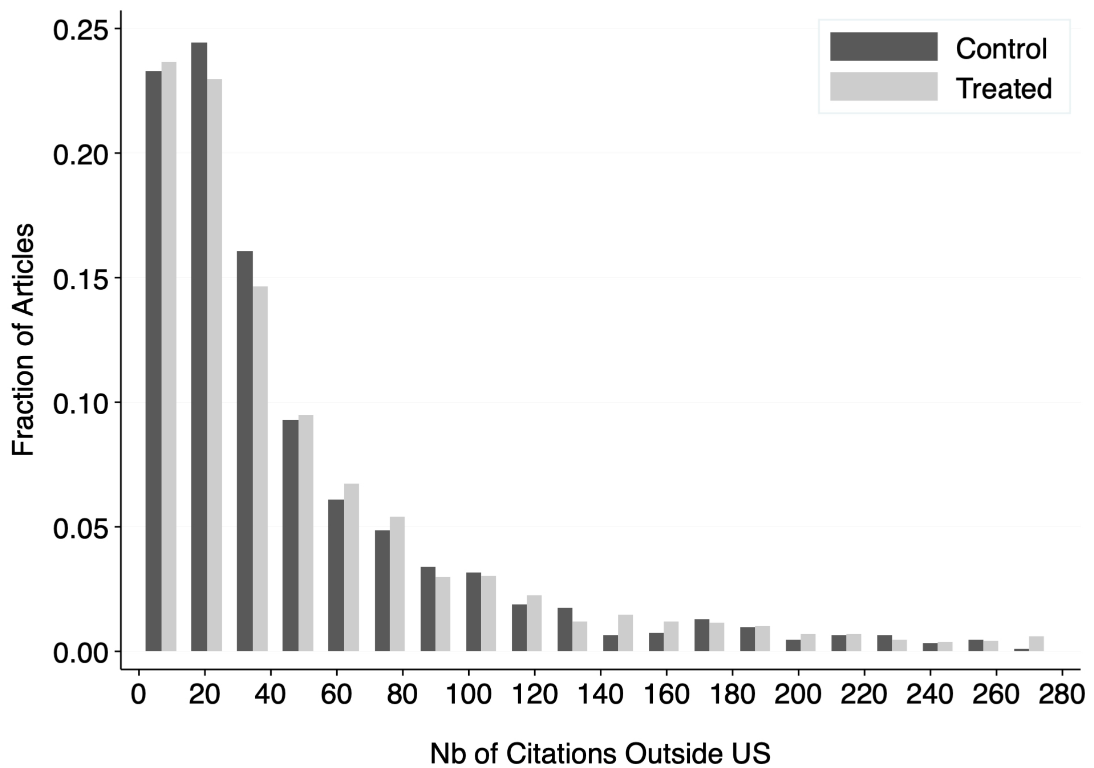

Figure 1: Histogram of Citation Outsider US for Control and Treated Articles

Note: The histogram excludes publications with 280 or more citations outside the US (approx. 1% of the sample)

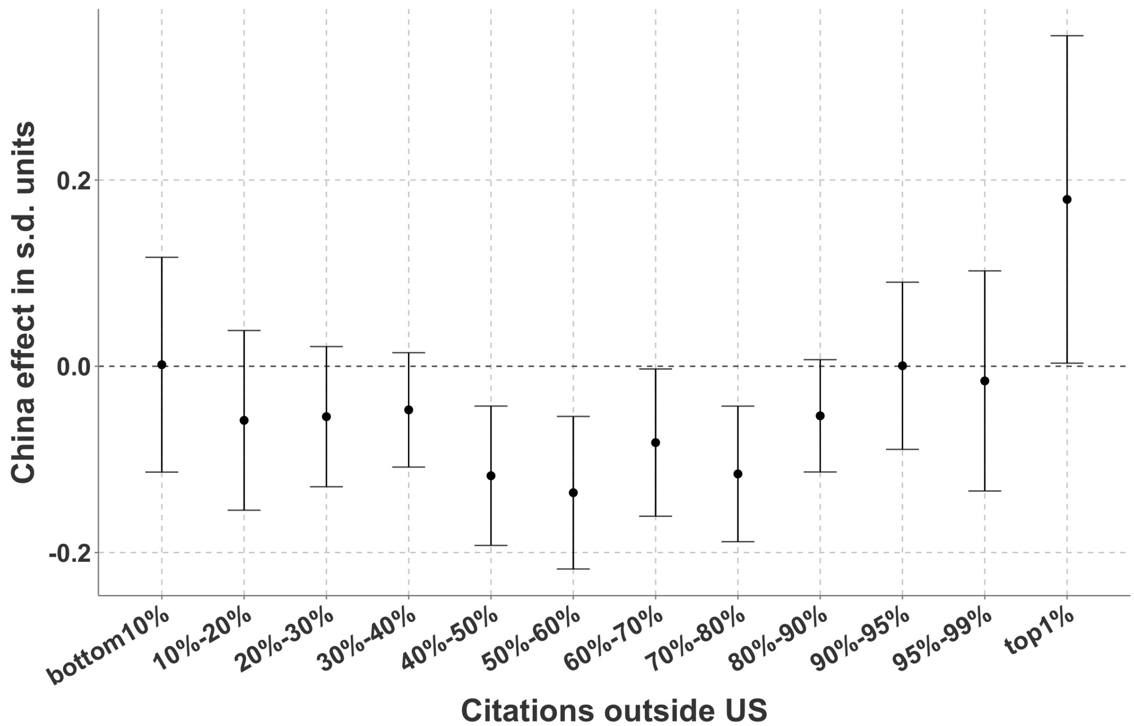

Figure 2: Heterogeneous effect of Chinese HCR on US citations, by HCR article citation impact

Note: The dark dots in the above plots correspond to coefficient estimates stemming from a Linear Probability Model in which the dependent variable is an indicator variable that equals 1 if the related paper cited the HCR paper, and 0 otherwise. The covariates of interest are 12 interaction terms between the China indicator variable and indicator variables for various quantiles of the distribution of non-US citations received. The corresponding specification also includes all the covariates included in column(3) of Table 5. The 95 percent confidence interval (the corresponding standard errors are two-way clustered at the investigator and matching strata levels) around these estimates is plotted with vertical lines.

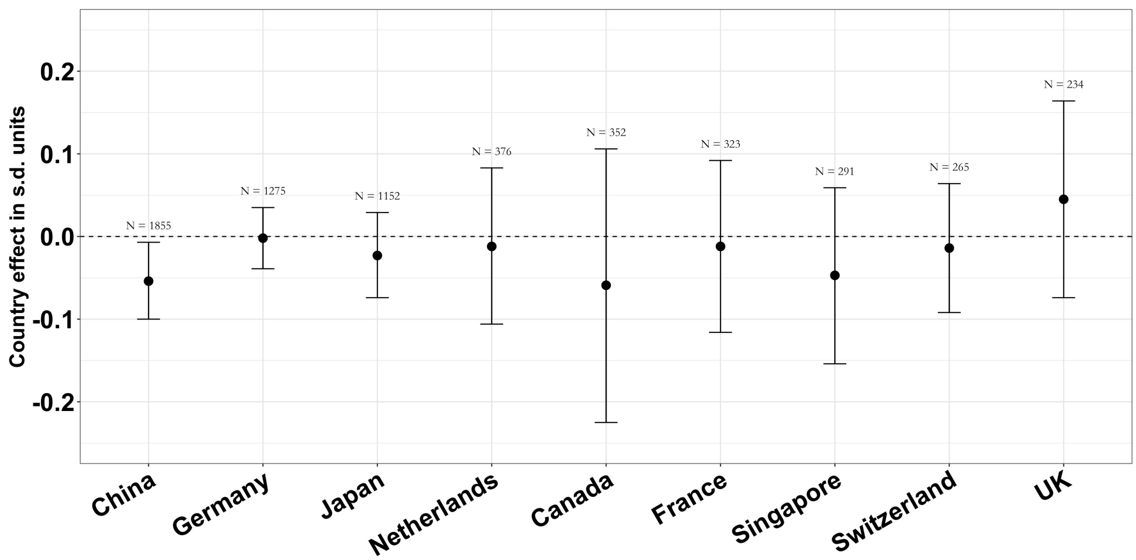

Figure 3: The Comparison of Country Effect

Note: We replace China with Germany, Japan, Netherlands, Switzerland, Canada, France, Singapore and Switzerland respectively to generate new treated and control groups, and estimate the country effect for each treated country with the same specification as column (3) in Table 5. The dark dots in the above plots correspond to country effect in s.d. units for each treated country. The 95 percent confidence interval (the corresponding standard errors are two-way clustered at the investigator and matching strata levels) around these estimates is plotted with vertical lines. The number of treated articles for each country is indicated above the corresponding coefficient estimate.

Table 1: China's Highly Cited Researchers (HCRs) and China's Publications across Fields

Field	Nb of China's HCRs in field	China's share of world HCRs in field (%)	China's share in world publications in field (%)
Chemistry	211	19.27	14.96
Materials Science	205	26.12	19.99
Engineering	164	18.94	14.78
Computer Science	59	9.83	13.13
Physics	57	7.33	14.78
Mathematics	49	10.06	15.93
Geosciences	45	5.77	10.97
Molecular Biology/Biochemistry/Genetics	29	1.36	10.00
Plant/Animal Science	20	2.03	9.79
Agricultural Sciences	18	2.62	9.34
Pharmacology/Toxicology	10	1.44	10.42
Environment/Ecology	9	1.18	10.34
Neuroscience/Behavior	9	1.09	4.88
Microbiology	5	0.88	6.65
Immunology	5	0.83	6.27
Clinical Medicine	2	0.10	5.33
Psychiatry/Psychology	1	0.16	2.24

Notes: (1) Highly Cited Researchers (HCRs) are selected based on their production of multiple highly cited papers that rank in the top 1% by citations in a field and year (in the Web of Science database); (2) We count the number of HCRs of each country without dropping duplicates (i.e., the same person on the HCR list of different years is counted repeatedly); for researchers who are affiliated to more than one institution, we defined their affiliation (country) based on their primary institution in the year when the HCR report was issued; (3) The share is computed based on English-language research articles published in each field between 2000 and 2015. Articles are attributed to countries on the basis of the share of institutional addresses located in the country, relative to the total number of institutional addresses.

Table 2: China's Research in Chemistry Compared to other Countries

Country/Region	Share of Chemistry Articles	Nb. of HCRs in Chemistry	Share of HCRs in Chemistry	Share of Articles in all fields	Nb. of HCRs in all fields	Share of HCRs in all fields
United States	19.31	471	43.01	25.17	8,306	46.48
China	14.96	211	19.27	9.20	1,104	6.18
Japan	7.66	30	2.74	6.32	420	2.35
Germany	5.55	77	7.03	5.06	1,023	5.73
India	4.68	4	0.37	2.95	27	0.15
United Kingdom	4.20	31	2.83	5.70	1,701	9.52
France	3.94	27	2.47	3.57	474	2.65
South Korea	3.28	33	3.01	2.83	163	0.91
Italy	2.96	3	0.27	3.32	274	1.53
Spain	2.95	26	2.37	2.55	309	1.73
Canada	2.47	20	1.83	3.34	451	2.52
Taiwan	1.49	2	0.18	1.74	86	0.48
Australia	1.48	22	1.83	2.44	581	3.25
Iran	1.37	1	0.09	1.09	46	0.26
Switzerland	1.15	33	3.01	1.08	420	2.35
Netherlands	1.06	7	0.64	1.75	465	2.60
Sweden	0.97	1	0.09	1.23	163	0.91
Belgium	0.74	3	0.27	0.88	197	1.10
Czech	0.67	7	0.64	0.49	26	0.15
Israel	0.59	10	0.91	0.80	58	0.33
Singapore	0.56	19	1.74	0.53	164	0.92
Denmark	0.53	7	0.64	0.68	163	0.91
Hong Kong	0.43	12	1.10	0.57	144	0.81
South Africa	0.29	3	0.27	0.45	32	0.18
Ireland	0.26	5	0.46	0.33	79	0.44
Saudi Arabia	0.25	32	2.92	0.23	272	1.52
Rest of world	16.22	0	0.00	15.71	722	4.04

Notes: (1) We list 26 countries and a residual "rest of the world" category that jointly include all HCRs in Chemistry, ranked by the share of Chemistry articles they produce. (2) The share of Chemistry articles is computed based on English-language original research articles in Chemistry during the period 2000-2015. Articles are attributed to countries on the basis of the share of institutional addresses located in the country, relative to the total number of institutional addresses. (3) The figures corespond to *Highly Cited Researchers* reports during the 2014-2018 period, and excludes the social sciences and business categories when tallying the number of HCRs in a country. (4) We count the number of HCRs of each country without dropping duplicates (i.e., the same person on the HCR list of different years is counted repeatedly); for researchers who are affiliated to more than one institution, we defined their affiliation (country) based on their primary institution in the year when the *Highly Cited Researchers* report was issued.

Table 3: Summary Statistics of Control and Treated Highly Cited Researchers

	HCRs from Rest of the World without US (N=219)					HCRs from China (N=89)				
	Mean	Median	Std Dev	Min	Max	Mean	Median	Std Dev	Min	Max
HCR Highest Degree Year	1984.374	1986	13.147	1951	2008	1993.876	1997	11.739	1952	2008
HCR Career Independent Year	1987.068	1989	13.504	1951	2010	1996.865	2000	12.391	1952	2010
HCR Female	0.018	0	0.134	0	1	0.045	0	0.208	0	1
HCR from English-speaking Country	0.342	0	0.476	0	1	0	0	0	0	0
HCR Earned US PhD Degree	0.027	0	0.164	0	1	0.011	0	0.106	0	1
HCR with US Post-doc Experience	0.370	0	0.484	0	1	0.348	0	0.479	0	1
HCR Did the Entire Training in the US (PhD & Post-doc)	0.078	0	0.268	0	1	0.079	0	0.271	0	1
HCR Career Nb. of Publications	475	395	315	57	2,207	355	306	249	58	1,226
HCR Career Nb. of Cites	24,790	21,367	18,000	1,082	166,540	14,347	12,060	9,516	1,697	52,380
HCR Career Nb. of Publications in Last Authorship Position	205	167	168	8	1,250	193	138	173	5	862
HCR Career Nb. of Cites for Papers in Last Authorship Position	9,870	7,929	9,617	151	100,552	7,057	5,356	6,480	76	29,108

Notes: Independent career starts when the scientist gets his/her first faculty job, typically a after postdoctoral fellowship, or more rarely, immediately after receipt of his/her PhD. Career publications (respectively cites) are computed from the year of independent career start until 2015.

Table 4: Summary Statistics of Control and Treated Articles

	Control articles (N=2,902)					Treated articles (N=1,855)				
	Mean	Median	Std Dev	Min	Max	Mean	Median	Std Dev	Min	Max
Source Article Publication Year	2008.909	2010	2.722	2000	2012	2008.909	2010	2.722	2000	2012
Source Article Nb. of Authors	5.551	5	2.099	1	19	5.623	5	1.960	2	18
Source Article Nb of Citations outside US	53	33	60	0	538	56	33	65	0	548
Source Article Nb. of Countries Represented (excl. US)	1.415	1	0.638	1	6	1.094	1	0.318	1	4
Source Article has US Author, Co-ethnic	0.013	0	0.112	0	1	0.033	0	0.178	0	1
Source Article has US Author, not Co-ethnic	0.035	0	0.184	0	1	0.010	0	0.098	0	1
Subfield Retraction Intensity	0.027	0	0.149	0	2.962	0.016	0	0.103	0	1.479
Subfield Home Research Intensity	1.436	1	2.071	0	32	4.519	3	5.120	0	58
Subfield Foreign Research Intensity	15.117	13	13.090	0	140	11.386	10	11.191	0	149
HCR Highest Degree Year	1985.460	1987	11.119	1951	2008	1989.032	1994	13.977	1952	2008
HCR Female	0.021	0	0.142	0	1	0.078	0	0.268	0	1
HCR from English-speaking Country	0.336	0	0.473	0	1	0	0	0	0	0
HCR earned US PhD Degree	0.036	0	0.187	0	1	0.031	0	0.174	0	1
HCR with US Post-doc Experience	0.258	0	0.438	0	1	0.267	0	0.442	0	1
HCR completed PhD & Post-doc in the US	0.061	0	0.239	0	1	0.052	0	0.223	0	1
HCR Cumulative Publications	294	233	244	4	1,934	224	170	194	0	997
HCR Cumulative Citations	8,744	6,048	9,071	1	73,033	3,950	2,535	4,172	0	25,884

Notes: Cumulative publications and citations are assessed at the end of the year prior the year of publication for each source article.

Table 5: Estimating the China location discount (or premium) on the rate of US citations [Linear Probability Model]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Chinese HCR	-0.0107**	-0.0095*	-0.0084*	-0.0073*	-0.0086*	-0.0100*	-0.0081*	-0.0093*
	(0.0029)	(0.0038)	(0.0037)	(0.0037)	(0.0037)	(0.0042)	(0.0037)	(0.0041)
Source Article Level Controls								
Source Article Nb. of Countries Represented (excl. US)		-0.0041†	-0.0037†	-0.0035	-0.0037†	-0.0043*	-0.0037†	-0.0041†
		(0.0022)	(0.0022)	(0.0022)	(0.0022)	(0.0021)	(0.0022)	(0.0022)
Source Article has US Author, Co-ethnic		0.0177	0.0099	0.0434*	0.0099	0.0072	0.0103	0.0446*
		(0.0116)	(0.0107)	(0.0202)	(0.0107)	(0.0102)	(0.0107)	(0.0203)
Source Article has US Author, not Co-ethnic		0.0065	0.0030	0.0046	0.0030	0.0014	0.0031	0.0044
		(0.0065)	(0.0064)	(0.0069)	(0.0064)	(0.0067)	(0.0064)	(0.0069)
Subfield Level Controls								
Subfield Retraction Intensity		-0.0060	-0.0064	-0.0064	-0.0063	-0.0054	-0.0027	-0.0020
		(0.0056)	(0.0056)	(0.0056)	(0.0056)	(0.0056)	(0.0061)	(0.0060)
Subfield Home Research Intensity		-0.0004	-0.0004	-0.0004	-0.0004	-0.0004	-0.0004	-0.0004
		(0.0005)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Subfield Foreign Research Intensity		-0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Investigator Level Controls								
HCR from English-speaking Country		-0.0010	0.0000	0.0002	-0.0000	0.0000	0.0001	0.0004
		(0.0038)	(0.0037)	(0.0036)	(0.0037)	(0.0036)	(0.0037)	(0.0035)
HCR earned US PhD Degree		-0.0063	-0.0050	-0.0082	-0.0050	-0.0071	-0.0053	-0.0133
		(0.0114)	(0.0108)	(0.0100)	(0.0108)	(0.0162)	(0.0108)	(0.0142)
HCR completed both PhD & Post-doc in the US		0.0067	0.0063	0.0068	0.0063	-0.0100	0.0063	-0.0099
		(0.0071)	(0.0069)	(0.0071)	(0.0069)	(0.0082)	(0.0069)	(0.0081)
Citing-cited Level Controls								
Log(Avg. Distance)			-0.0059†	-0.0058†	-0.0059†	-0.0053	-0.0059†	-0.0052
			(0.0033)	(0.0033)	(0.0033)	(0.0034)	(0.0033)	(0.0034)
Same Journal			0.0129**	0.0129**	0.0129**	0.0129**	0.0129**	0.0128**
			(0.0025)	(0.0025)	(0.0025)	(0.0025)	(0.0025)	(0.0025)
Common Coauthor			0.2118**	0.2080**	0.2117**	0.2123**	0.2119**	0.2080**
			(0.0509)	(0.0508)	(0.0508)	(0.0509)	(0.0509)	(0.0507)
Citing Coauthor is HCR's Past Collaborator			0.0242**	0.0241**	0.0242**	0.0243**	0.0243**	0.0242**
			(0.0064)	(0.0064)	(0.0064)	(0.0064)	(0.0064)	(0.0064)
Citing Author and HCR are Co-ethnics			0.0052†	0.0051†	0.0038	0.0052†	0.0051†	0.0031
			(0.0031)	(0.0031)	(0.0064)	(0.0031)	(0.0031)	(0.0064)
Interactions								
Chinese HCR × US Cited Author, Co-ethnic				-0.0482*				-0.0532*
				(0.0239)				(0.0230)
Chinese HCR × US Cited Author, not Co-ethnic				-0.0097				-0.0218
				(0.0162)				(0.0192)
Chinese HCR × Citing Author and HCR are Co-ethnics					0.0022			0.0032
					(0.0071)			(0.0071)
Chinese HCR × HCR earned US PhD Degree						0.0045		0.0104
						(0.0189)		(0.0171)
Chinese HCR × HCR with US Post-doc Experience						-0.0054		-0.0041
						(0.0062)		(0.0062)
Chinese HCR × HCR completed both PhD & Post-doc in the US						0.0425**		0.0464**
						(0.0071)		(0.0063)
Chinese HCR × Subfield Retraction Intensity							-0.0189*	-0.0168†
							(0.0084)	(0.0095)
Mean of Dependent Variable	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025
Std. Dev. of Dependent Variable	0.156	0.156	0.156	0.156	0.156	0.156	0.156	0.156
China effect in s.d. units	-0.068	-0.061	-0.054	-0.047	-0.055	-0.064	-0.052	-0.059
Adjusted R^2	0.059	0.060	0.065	0.066	0.065	0.066	0.066	0.066
Investigators	299	299	299	299	299	299	299	299
HCR Papers	4,757	4,757	4,757	4,757	4,757	4,757	4,757	4,757
Citing Papers	36,197	36,197	36,197	36,197	36,197	36,197	36,197	36,197
Citing-Cited Pairs	72,550	72,550	72,550	72,550	72,550	72,550	72,550	72,550

Notes: The dependent variable is a dummy variable that equals 1 if the related paper cites the HCR paper, and 0 otherwise. All regressions include fixed effects for rank bins of each citing paper j with respect to its topic similarity to paper i; fixed effects for the interaction of citing and cited paper publication year; fixed effects for each CEM strata; fixed effects for the HCR's highest degree year and a HCR gender indicator variable (coefficients not reported). Standard errors in parentheses are two-way clustered at the investigator and strata level. $^{\dagger}p < 0.10$, * $p < 0.05$, ** $p < 0.01$