Global citation inequality is on the rise

Mathias Wullum Nielsen[a,1,2] and Jens Peter Andersen[b,1]

[a]Department of Sociology, University of Copenhagen, 1353 Copenhagen, Denmark; and [b]Danish Centre for Studies in Research and Research Policy, Department of Political Science, Aarhus University, 8000 Aarhus C, Denmark

Citations are important building blocks for status and success in science. We used a linked dataset of more than 4 million authors and 26 million scientific papers to quantify trends in cumulative citation inequality and concentration at the author level. Our analysis, which spans 15 y and 118 scientific disciplines, suggests that a small stratum of elite scientists accrues increasing citation shares and that citation inequality is on the rise across the natural sciences, medical sciences, and agricultural sciences. The rise in citation concentration has coincided with a general inclination toward more collaboration. While increasing collaboration and full-count publication rates go hand in hand for the top 1% most cited, ordinary scientists are engaging in more and larger collaborations over time, but publishing slightly less. Moreover, fractionalized publication rates are generally on the decline, but the top 1% most cited have seen larger increases in coauthored papers and smaller relative decreases in fractional-count publication rates than scientists in the lower percentiles of the citation distribution. Taken together, these trends have enabled the top 1% to extend its share of fractional- and full-count publications and citations. Further analysis shows that top-cited scientists increasingly reside in high-ranking universities in western Europe and Australasia, while the United States has seen a slight decline in elite concentration. Our findings align with recent evidence suggesting intensified international competition and widening author-level disparities in science.

scientific elites | citations | inequality | science | sociology of science

Science is a highly stratified social system. The distribution of scientific rewards is remarkably uneven, and a relatively small stratum of elite scientists enjoys exceptional privileges in terms of funding, research facilities, professional reputation, and influence (1–5). The so-called Matthew effect, well-documented in science (6–15), implies that accomplished scientists receive more rewards than their research alone merits, and recent evidence indicates a widening gap between the "haves" and the "have nots" of science in terms of salary levels (5), research funding (16), and accumulation of scientific awards (17).

Inequality may foster creative competition in the science system (18, 19). However, it can also lead to a dense concentration of resources with diminishing returns on investment (intellectual and fiscal) (16, 20), and to monopolies in the marketplace of ideas (21, 22).

The social processes that sort scientists into more or less prestigious strata are complex and multifaceted (1, 10, 23) and may be changing in response to external pressures such as globalization, the advent of new information technologies, and shifts in university governance models (3). However, a few common characteristics have always separated elite scientists from the rest of us, most notably their scientific output and visibility. Publications and citations are critical building blocks for status and success in science (23, 24), and the scientific elite accounts for a large share of what is published and cited.

In 1926, Lotka observed that the publication frequencies of chemists followed an inverse-square distribution, where the number of authors publishing N papers would be $\sim 1/N^2$ of the number of authors publishing one paper (25). Building on Lotka's work, de Solla Price later went on to suggest that 50% of all publications

were produced by a mere 6% of all scientists (26). More recent research demonstrates even larger disparities in citation distributions at the author level (2, 6, 11, 27, 28), but variations in citation concentration across disciplinary, institutional, and national boundaries remain uncertain. Further, it is unclear whether the observed inequalities in citation shares have intensified over time.

Advances in author-disambiguation methods (29) allow us to investigate these questions on a global scale. We used a linked dataset of 4,042,612 authors and 25,986,133 articles to examine temporal trends in the concentration of citations at the author level, and differences in the degree of concentration across fields, countries, and institutions.

Publication and citation data were retrieved from Clarivate's Web of Science (WoS). We limited our focus to disciplines within the medical and health sciences, natural sciences, and agricultural sciences, where journal publication is the primary form of scholarly communication (*Materials and Methods*). We used a disambiguation algorithm to create publication profiles for all authors with five or more publication entries in WoS. The disambiguated dataset allowed us to measure developments in citation concentration from 2000 onward.

Per-author citation impact was measured using field-normalized citation scores (*ncs*). *ncs* is calculated by dividing the raw per-paper citation scores with the average citation counts of comparable papers published in the same year and subfield. *ncs* was rescaled to account for citation inflation, represented here as *nics*. We report per-author cumulative citation impact based on a full and fractional counting. The full counting gives the sum of *nics* for all papers published by a scientist. The fractional counting also gives the sum of citations accrued by a scientist in all her papers, but divides the per-article citation scores with the number of contributors per paper. We use citation density plots and Gini coefficients to gauge trends in citation imbalance and concentration.

Significance

By analyzing a global sample of 4 million authors and 26 million scientific papers, this study finds that the top 1% most-cited scientists have increased their cumulative citation shares from 14 to 21% between 2000 and 2015 and that the Gini coefficient for citation imbalance has risen from 0.65 to 0.70. The growing citation concentration should be understood in the context of diverging trends in publication and collaboration activities for the top 1% compared to the "ordinary scientist." Our study raises intriguing questions about how rising inequalities will shape the evolution of science.

SOCIAL SCIENCES

Results

Our analyses suggest that the concentration of citations within a small stratum of elite authors has increased during the past two decades. Fig. 1*A* shows the general trend since 2000 of the share of total citations accrued by the top 1% most cited authors (i.e., the "citation elite"). The steep rise in full-count citation concentration from 2011 and onward coincides with a sharp increase in collaboration. This trend is likely driven by the launch of large-scale collaborative experiments in physics and astronomy, such as the Large Hadron Collider, generating "mega-papers" with abnormally high citation rates and coauthor numbers. Fig. 1*B* and *C* juxtapose the trends in citation concentration in physics and astronomy with all other disciplines and confirm that the steep incline since 2011 is specific to physics and astronomy. Moreover, while we observe diverging trends for full-count and fractionalized citation concentration in physics and astronomy, this is not the case for the remaining disciplines. In what follows, we therefore restrict the focus to the health

sciences, agricultural sciences, and natural sciences, excluding physics and astronomy. As shown in Fig. 1*C*, citation concentration has grown incrementally in these fields. From 2000 to 2015, the top 1% extended its citation share from 14.7 to 19.6% in fractional counts (relative increase = 33%) and from 14.1 to 21% in full counts (relative increase = 49%). This concentration trend is not sensitive to the general growth in publishing scientists or the expanding number journals in WoS. Complementary analyses with a fixed number of authors and a fixed set of journals across years produce near-identical results (*SI Appendix,* Fig. S1).

The growing citation shares accrued by the top 1% should be understood in the context of diverging trends in publication rates and collaboration activities. Fig. 2*A* plots the trend since 2000 in the citation elite's share of all published papers and shows an increase from 5 to 12% in fractionalized output and from 4 to 11% in full-count output. Fig. 2*B* displays developments in the mean, annual publication rates per author for the 50th, 75th, and 99th percentile of most-cited scientists and indicates an upward
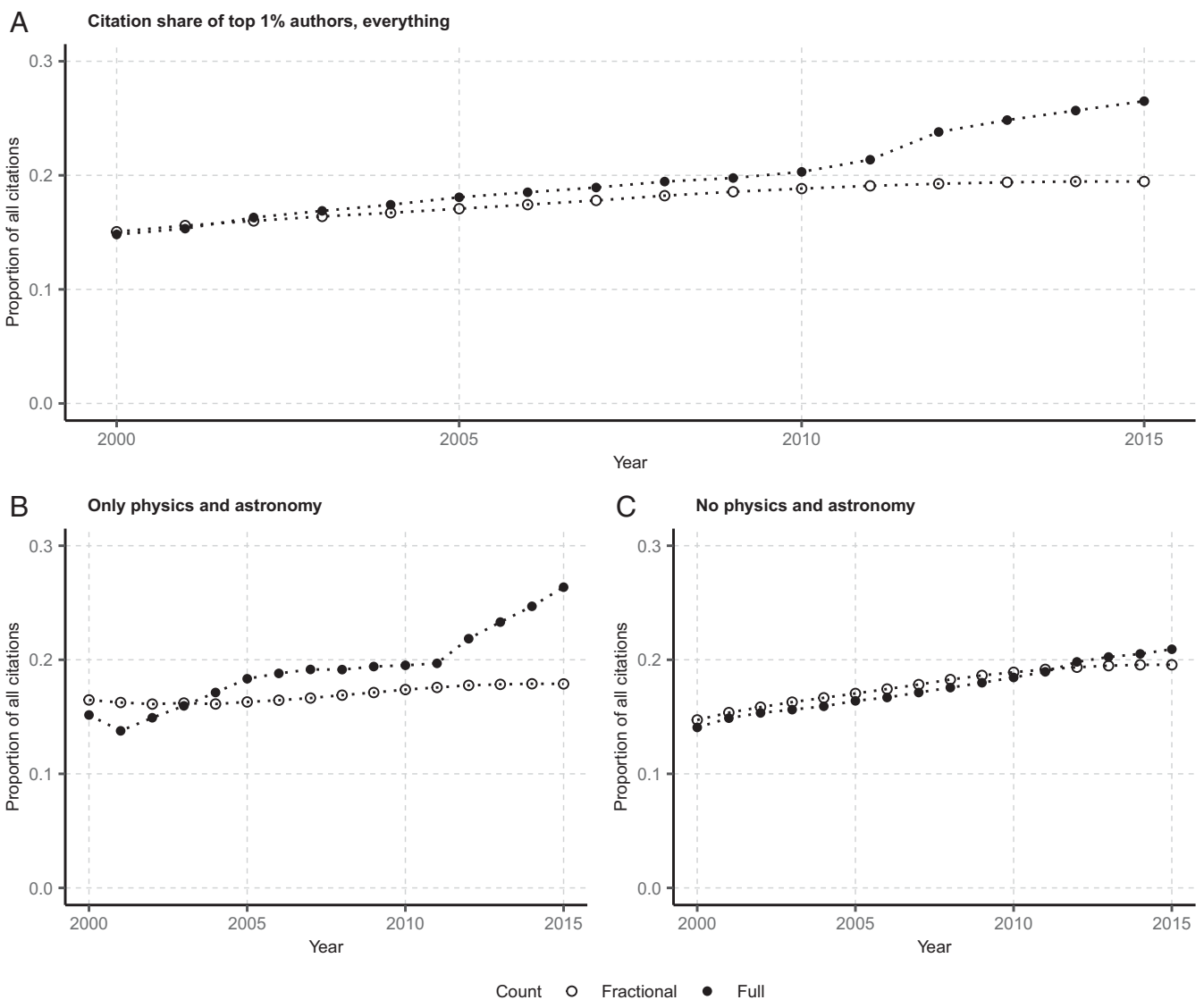

Fig. 1. The rise in citation concentration for the 1% most cited as total, inflation-adjusted citations (*nics*) cumulated per year from 2000 to 2015. Solid dots show the proportion based on full-count citation rates per article. The hollow dots show the fractional-count citation proportion in which the score of each article is divided by the number of authors. (*A*) Proportion for all authors in the set. (*B*) Sample restricted to authors in physics and astronomy. (*C*) All authors in *A*, excluding authors in physics and astronomy.
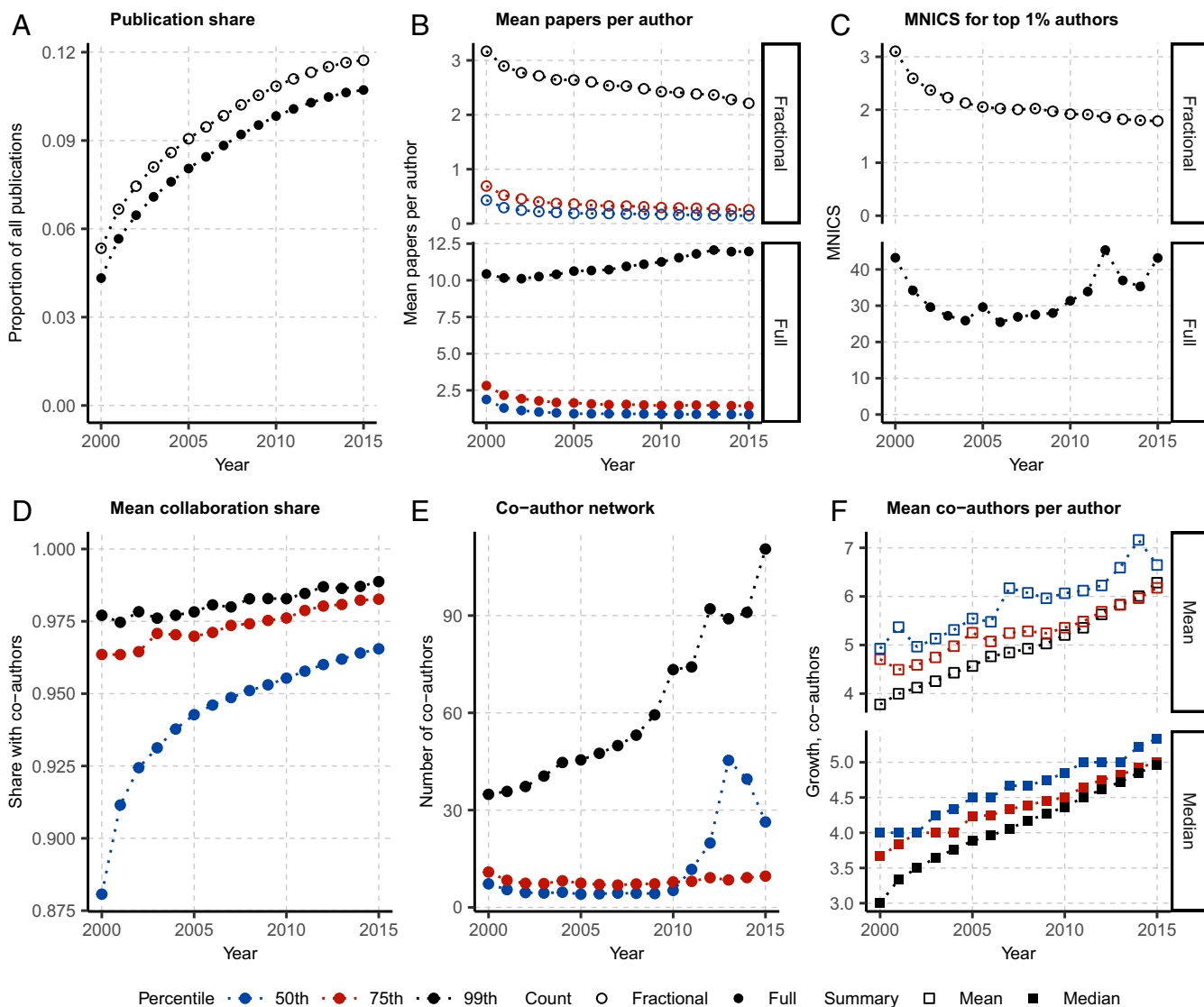
www.manaraa.com

Fig. 2. Publication, citation, and collaboration trends. (*A*) Changes over time in the share of total papers accrued by the top 1% (99th percentile), 75th, and 50th percentile from 2000 to 2015. (*B*) Developments in average publication output per year (per author) for the top 99th, 75th, and 50th percentile based on a full and fractional counting of papers. (*C*) Mean citation rate per paper (per year) for the 99th percentile. (*D*) Mean proportion of papers with at least one coauthor for the 99th, 75th, and 50th percentile. (*E*) Average number of coauthors per year for the 99th, 75th, and 50th percentile. (*F*) Annual mean and median number of coauthors per paper for authors in the three percentile bins. In all panels, the black lines and dots show the 99th percentile, red shows the 75th percentile, and blue shows the 50th percentile. Solid dots show the scores by full count and hollow dots show fractional counts. Solid squares show the median and hollow squares show the mean. *B*, *C*, and *F* are split into two facets with common *x* axes and individual *y* axes.

trend for the 99th percentile (i.e., the top 1%) in full counts, from 10.1 to 12.4, and a downward trend in fractional counts from 3.1 to 2.1. In comparison, both full-count and fractional-count outputs have declined for authors in the 50th and 75th percentile, and this decline may help explain why the citation elite has increased its total share of fractionalized publications (Fig. 2*A*), while decreasing its annual rate of fractional-count publications. In relative terms, the fractional-count publication output has decreased less for the top 1% (relative change = −31%) than for the ordinary scientist (50th percentile, relative decrease = −67%; 75th percentile, relative decrease = −64%) (Fig. 2*B*).

As shown in Fig. 2*C*, the rise in citation concentration does not reflect an average increase in the elite's impact per paper. From 2000 to 2015, the top 1% decreased its average, fractional-count citation rate per paper from 3.10 to 1.79, while the average, full-count citation rate started and ended around 43 (Fig. 2*C*).

A closer look at trends in coauthorships shows that today's scientists are engaged in more collaborations than in the past and have increased their average and total number of collaborators per year. Fig. 2*D* indicates a slight increase in the proportion of papers with at least one collaborator for the 75th and 99th percentile, and a more striking upsurge for the 50th percentile. Compared to the 50th and 75th percentile, the citation elite has notably higher total (mean) rates of annual collaborators, and this number has escalated from 35 in 2000 to 111 in 2015 (Fig. 2*E*). However, the mean and median numbers of collaborators per paper remain lower for the citation elite than for the 50th and 75th percentiles of most-cited authors throughout the observation period (Fig. 2*F*).

Taken together, these trends suggest that the citation elite has extended its shares of all fractional-count publications and citations without increasing its general productivity level (in fractional
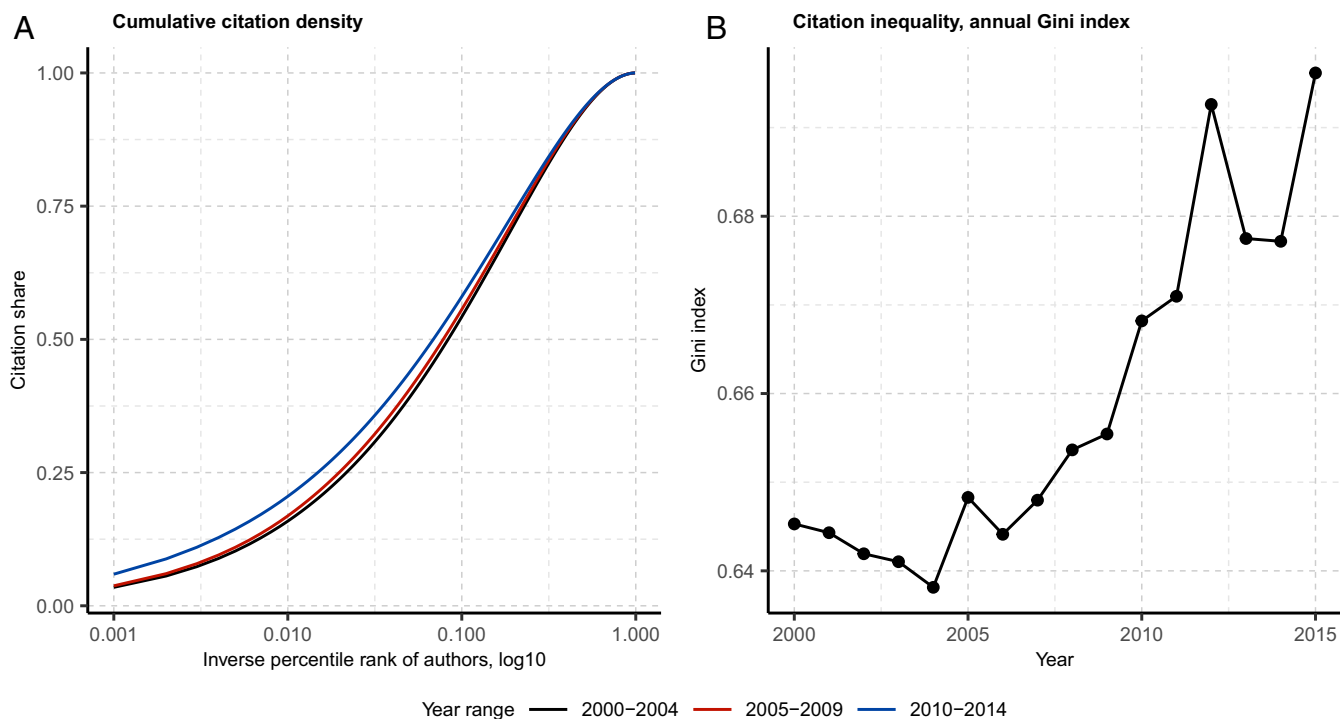
Nielsen and Andersen
Global citation inequality is on the rise

www.manaraa.com

Fig. 3. Citation gap. (*A*) Citation density (full count) by fraction of authors, stratified by 5-y intervals. The *x* axis is on a logarithmic scale. (*B*) Gini coefficients of citation density (full count) by percentile rank, per year, from 2000 to 2015.

counts) or impact per paper. While the number of coauthors per paper has generally gone up, the fractional-publication output per author (per year) has decreased. However, the citation elite has seen smaller relative decreases in fractionalized productivity than authors in lower percentiles of the citation distribution. Moreover, while increasing collaboration and full-count publication rates go hand in hand for the top 1%, scientists in the 50th and 75th percentiles are engaging in more and larger (average) collaborations but publishing slightly less over time. Results for physics and astronomy are reported in *SI Appendix*, Fig. S2.

Fig. 3*A* plots the distribution of cumulative citation shares (full count) by percentile ranks of authors, stratified by 5-y intervals, and for instance shows that the 0.1% most-cited authors earned larger citation shares than the 38% least cited in the period 2000–2004 (black line) and the 47% least cited in the period 2011–2014 (teal line). Gini coefficients for the citation imbalance have increased from 0.65 to 0.70 in the period 2000–2015 (Fig. 3*B*).

The generality of the trend toward citation concentration across fields and disciplinary domains is reported in Fig. 4 and Table 1. Plots of citation density curves and annual Gini coefficients show that citation inequality and concentration are on the rise in all three Organisation for Economic Co-operation and Development (OECD) major fields (Fig. 4 *A* and *B*). Table 1 juxtaposes the periods 2000–2004 and 2011–2015 and shows that the top 1% most cited per discipline has garnered >10% relative increases in citation shares in 58% of agricultural-science subfields, 58% of medical and health-science subfields, and 27% percent of natural-science subfields. Some of the highest increases in citation concentration are observed in clinical medicine (28.9%) and the health sciences (25%), and the largest decrease in computer and information sciences (−7%).

Further analysis shows that authors in the citation elite increasingly reside in western Europe and Australasia, while the United States has decreased its concentration of top-cited scientists (Fig. 5). Among the 10 countries with the highest concentration of

citation elites, all, except for the United States, have extended their concentration of authors in the top 1% (Fig. 5*A*). The United States had the highest elite concentration between 2000 and 2004, but has since been surpassed by The Netherlands, England, Switzerland, and Belgium (Fig. 5*A*). Fig. 5*B* displays developments for the 10 countries with the largest numerical representation of citation elites, and shows that elite concentration has also declined in China and Japan, while Italy, Germany, and France have seen slight increases in elite concentration. Seven of the 10 countries with the largest growth in elite concentration are located in Europe, but South Africa, Australia, and New Zealand have also seen notable inclines (Fig. 5*C*). The countries exhibiting the smallest growth and largest declines are mostly large economies, such as the United States, China, India, Russia, and Brazil. As shown in *SI Appendix*, Fig. S3, country-level imbalances in citation accumulation (measured by the Gini coefficient) have remained relatively stable since 2000, although concentration has decreased in the highest percentiles of the citation distribution.

Fig. 6 depicts similar trends at the university level. While top-ranking universities in the United States and United Kingdom still hold the highest proportions (Fig. 6*A*) and numbers (Fig. 6*B*) of scientists in the top 1%, some of them, including Harvard University, Stanford University, Massachusetts Institutes of Technology, and University of California San Francisco have decreased their elite concentration (Fig. 6*D*). Meanwhile, European and Australian Universities such as Leiden University, Kings College London, University of Amsterdam, University of Melbourne, and University of Sydney have increased their concentration of elites (Fig. 6*C*). Institution-level citation imbalances (measured by the Gini coefficient) have remained stable over three consecutive periods of 5 y, while citation concentration has slightly increased at the upper end of the citation distribution (*SI Appendix*, Fig. S3).

Discussion

Existing evidence on trends in citation inequality is mixed. Some studies indicate a rise in citation concentration at the article level
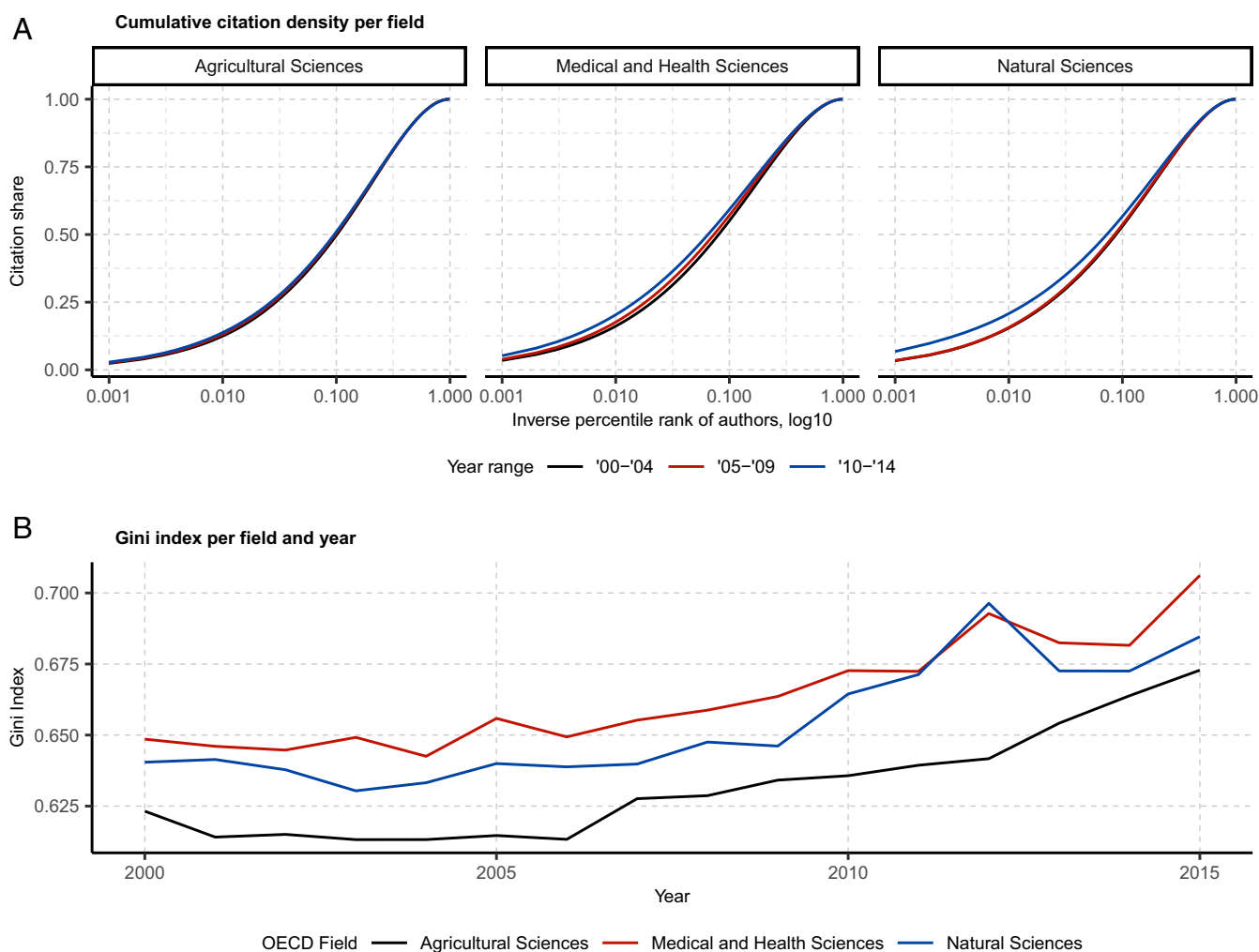
www.manaraa.com

A

Cumulative citation density per field

B

Gini index per field and year

Fig. 4. Citation concentration and inequality by field. (*A*) Citation density (full counting) by fraction of authors, stratified by 5-y intervals for agricultural sciences, medical and health sciences, and natural sciences. The *x* axes are on a logarithmic scale. (*B*) Field-specific Gini coefficients of citation density (full count) by percentile rank, per year from 2000 to 2015.

(30–32), while others suggest a decline (33–35). These inconsistencies may largely result from variations in the examined time periods, data, disciplinary focus, and research design (including attention—or lack thereof—to citation inflation, database effects, and field normalization). This study adds to the literature by focusing on developments in citation concentration at the author level, which, according to our results, have been on the rise.

In a study of the publication trajectories of elite scientists within a select set of 23 high-impact economics and multidisciplinary-science journals, Petersen and Penner (11) found decreasing citation inequalities at the author level, with Gini coefficients declining from 0.83 (economics) and 0.74 (multidisciplinary) in the 1970s, to 0.74 (economics) and 0.63 (multidisciplinary) in the 1990s. Our analysis, which draws on a larger and more diverse sample of authors and journals, suggests a different development. From 2000 to 2015, the share of total citations accrued by the top 1% most-cited scientists increased from 14 to 21% in full counts and the annual Gini coefficient for citation imbalance rose from 0.65 to 0.70. These main results account for citation inflation and do not appear to be driven by the general growth in publishing scientists nor the expanding list of journals covered by WoS.

Our findings align with prior work suggesting increasing team sizes and declining fractionalized productivity rates (36, 37). We add to this work by showing diverging trends for the top 1% most cited compared to the ordinary scientist, with top-cited researchers

seeing larger increases in coauthored papers and smaller relative declines in fractional-count publication rates. Moreover, while increasing collaboration and full-count publication rates go hand in hand for the top 1%, the productivity levels (in full counts) for authors in lower percentiles of the citation distribution have slightly declined. Taken together, these trends have allowed the top 1% percent to extend its share of both fractional- and full-count citations. Note here that team-based research is known to accrue higher citation rates than solo-authored work, which may partially explain our findings (36, 38, 39).

The rise in citation concentration should be understood in the context of changing dynamics in the career trajectories of science, where scientists increasingly enter the enterprise to support research projects led by others, without embarking on a long-term, publication-active research career of their own. Indeed, recent research documents a rise in the global proportion of "transient scientists" with short careers in academic science and low publication outputs (40).

The observed differences in inequality trends across subfields may reflect underlying variations in the magnitude and concentration of funding streams, but may also be attributed to dissimilarities in citation and publication practices (including publication frequencies, author-group sizes, paper lengths, the number of references per paper, and the cited half-life of published papers) (34). Future research should examine how these and other variations in

Table 1. Trends in citation concentration by field

OECD major field	Minor field	WoS categories			
		N	G_{nics} (Δ %)	G_n (Δ %)	P
Agricultural sciences	Agriculture, forestry, fisheries	6	4 (13.1)	1 (4.1)	470,781
	Animal and dairy science	1	0 (6.4)	1 (18.7)	180,786
	Other agricultural science	4	3 (10.9)	4 (12.1)	729,668
	Veterinary science	1	0 (8.9)	0 (9.5)	382,829
Medical and health sciences	Basic medical research	10	4 (7.7)	4 (6.1)	3,226,140
	Clinical medicine	32	24 (28.9)	19 (12.4)	9,591,374
	Health sciences	17	8 (25.0)	6 (9.0)	1,557,064
Natural sciences	Biological sciences	21	5 (12.5)	9 (0.8)	4,636,995
	Chemical sciences	9	2 (1.9)	4 (6.9)	3,650,044
	Computer and information sciences	8	2 (−7.0)	1 (−0.2)	15,087
	Earth and related environmental sciences	10	3 (12.4)	6 (9.9)	1,544,402
	Mathematics	7	0 (1.3)	4 (23.7)	361,530
	Other natural sciences	1	1 (40.8)	0 (−9.8)	674,780

The table reports the number of WoS subject categories (N) grouped under each OECD major and minor field of science. G_{nics} is the number of WoS subject categories with >10% increase in citation concentration for the top 1% most-cited authors in the period 2011–2015 compared to the period 2000–2004. The percentages in parentheses report the total, weighted change in citation concentration. G_n is the corresponding change in publication concentration, using full count, for the 1% most-cited authors. P is the total number of papers written by authors classified under each OECD minor field.

the social, material, and intellectual organization of science contribute to explain the diverging patterns of inequality across subfields.

Our finding that citation elites increasingly reside in western Europe and Australasia, while the United States has slightly decreased its concentration of top-cited scientists, aligns with recent research suggesting changing power relations in global science (41–43). The intensified international competition is also reflected in the growing concentration of elite scientists within reputed research institutions outside of North America, while world-leading universities in the United States have decreased their concentration of top-cited scientists. Despite these developments, top schools in the United States and United Kingdom still employ the largest numbers and concentrations of elite researchers, and the stark inequalities in institution-level citation distributions observed in this study (*SI Appendix*, Fig. S3) reflect the continuous dominance of elite universities in the global production of scientific knowledge (44).

There are certain caveats and limitations to the conclusions that can be drawn from this study. First, our author-disambiguation method introduces some level of uncertainty, especially for scientists with low publication outputs. To reduce this uncertainty, we have limited our focus to authors with >4 WoS publications (*Materials and Methods*). However, because most scientists with a publication output ≤4 have low cumulative citation rates, our results should be interpreted as conservative estimates of the actual citation disparities in a global science system, where short-term, academic research careers are becoming increasingly prevalent (41).

Second, outstanding citation impact is not in itself a sufficient condition for elite membership in the scientific community, which is also determined by a scientist's access to resources and research facilities, the strength of her social ties to other elite members, and her influence over the activities of others (1, 18). In the future, more detailed studies should examine how these factors interact to determine status and success across institutional contexts, disciplines, and countries.

Third, our dataset likely includes spuriously highly cited authors whose impact has been generated through extreme self-citations, citation farms, and ghost authorships (45). While these authors are indeed a part of the citation elite, their characteristics may differ considerably from those of the typical top-cited scientist.

Fourth, because our analysis focuses on per-author cumulative citation rates, citation-elite membership is strongly correlated with age. Future research should add to our work by examining variations in the degree of citation concentration within and

across age cohorts. Another promising avenue for future research concerns developments in the authorship positions of elite scientists that are working in areas where first and last authorships are specifically important.

Fifth, while our study suggests that citation concentration should be understood in the context of diverging trends in publication and collaboration activities for the elite compared to the ordinary scientist, it does not offer any evidence on the underlying drivers of this development. A promising avenue for future research would be to explore how external factors such as the fusion of fields, the emergence of interdisciplinary research, the increasing availability of big data, improved measurement and computational techniques, and the requirement of huge teams for large-scale and long-term experiments, may have shaped the observed trend toward citation concentration at the author level.

In science, like in other social systems, hierarchies of status tend to be self-reinforcing. Our finding that a general decline in productivity rates and a "collaborative advantage" enjoyed by the most-cited scientists have coincided with increasing citation inequality, aligns with evidence suggesting accelerating trends toward funding concentration at the individual level (16, 20). Indeed, these processes may be closely interlinked. Highly cited authors have better odds of winning grants, allowing them to expand their research laboratories and collaborative networks, ultimately leading to increased (full-count) publication and citation rates (46). In an evaluative system, where funding and hiring decisions are heavily informed by bibliometric indices, we may expect to see further increases in both funding, publication and citation concentration.

Our results raise intriguing questions about how rising inequalities will shape the evolution of science. Is the widening gulf between the haves and the have nots inhibiting the creative competition between divergent ideas, paradigms, theories, and approaches and narrowing the opportunities for scientific fields to develop in new directions (21)? Or is the concentration trend a symptom of an overcrowded intellectual market, where the number of scientists could be sharply reduced without slowing the rate of advance (2)? Answers to these questions could help improve the science system's capacity to efficiently use its global talent pool.

Materials and Methods

Publication metadata for this study were retrieved from the Web of Science: Science Citation Index-Expanded, Social Science Citation Index and Arts &
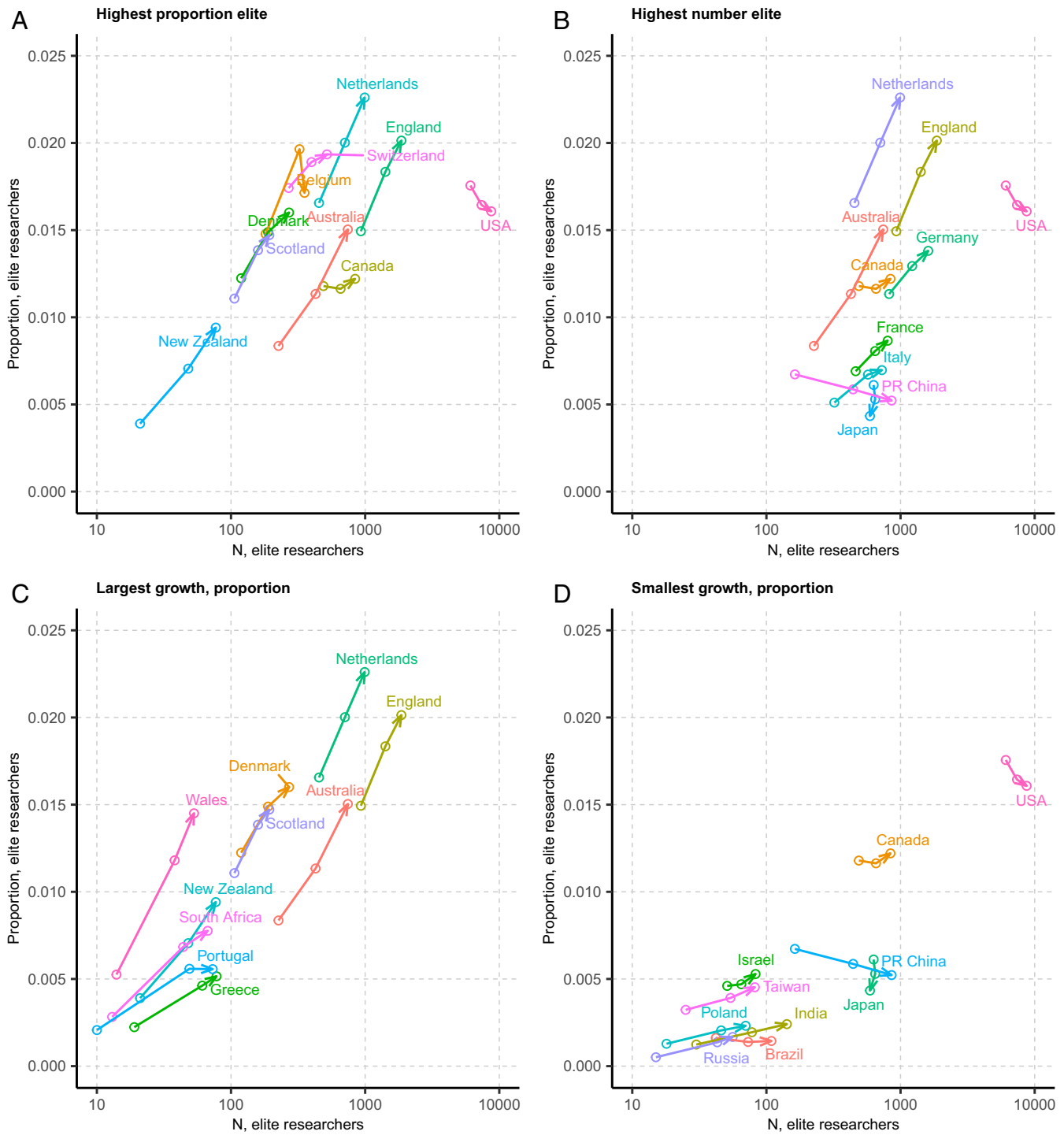
www.manaraa.com

Fig. 5. Country-level developments in the representation and concentration of elite researchers (i.e., the top 1% most cited) from 2000 to 2014, by 5-y increments. (*A*) Changes in elite concentration within the 10 countries with the highest proportion of top-cited scientists overall. (*B*) Changes in elite concentration within the 10 countries with the highest number of top-cited scientists overall. (*C*) Developments in elite concentration within the 10 countries with the largest growth in proportions of elite scientists. (*D*) Developments in elite concentration within the 10 countries with the smallest growth or largest declines in proportions of elite scientists. Trends are plotted by 5-y increments in all panels (2000–2004, 2005–2009, 2010–2014). The arrows indicate the direction of the development over time. Only countries with at least 2,000 disambiguated authors and at least 30 elite researchers are included in the figure.

Humanities Citation Index produced by Clarivate Analytics (henceforth WoS). The analysis relies on the structured database access to these data provided through the WoS installation at Center for Science and Technology Studies, Leiden University. The data are essentially the same as in WoS, but with some improvements with respect to author disambiguation and cleaning of missing and faulty reference data. The analysis relies on two further

additions to the data available through this installation: an algorithmic disambiguation of author names (29) and a highly detailed field-classification system.

A high-precision approach to name-based author disambiguation is essential for any analysis of researcher-specific impact and productivity. Author disambiguation can be challenging for two reasons: 1) individual authors may use

Nielsen and Andersen
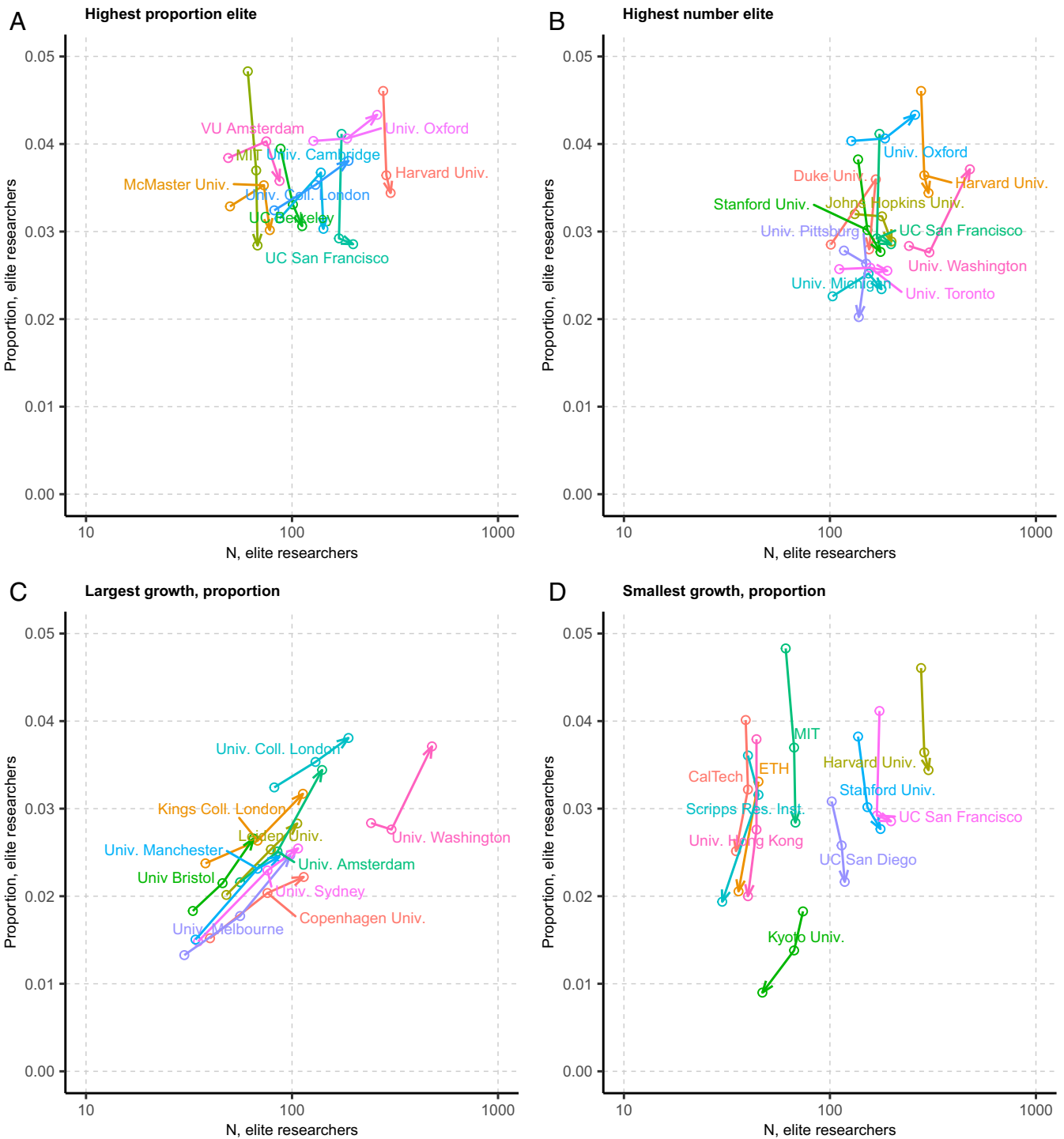Global citation inequality is on the rise

Fig. 6. University-level developments in the representation and concentration of elite researchers (i.e., the top 1% most cited) from 2000 to 2014. (*A*) Changes in elite concentration within the 10 institutions with the highest proportion of top-cited scientists overall. (*B*) Developments in elite concentration within the 10 institutions with the highest number of top-cited scientists overall. (*C*) Changes in elite concentration within the 10 institutions with the largest growth in proportions of elite scientists. (*D*) Changes in elite concentration within the 10 institutions with the largest decline in proportions of elite scientists. Trends are plotted by 5-y increments in all panels (2000–2004, 2005–2009, 2010–2014). The arrows indicate the direction of the development over time. Only institutions with at least 2,000 disambiguated authors and at least 30 elite researchers are included in the figure. Hospitals and medical centers have been omitted from the visualization.

more than one version of their name in the author byline (for instance, the same author may decide to include her middle name in one publication, while leaving it out in another, or she may be inconsistent in her use of hyphenated family names across publications); 2) many authors have identical author names. Caron and van Eck (29) have approached this problem computationally

by clustering papers based on names, affiliations, emails, coauthors, grant numbers, subject categories, journals, self-citations, bibliographic coupling, and cocitations. This method has been used to disambiguate all authors in the entire WoS database. The disambiguation is not perfect. Caron and van Eck report 95% precision and 90% recall. In a recent study, Tekles and Bornmann

found that Caron's and Eck's method was superior to four other unsupervised disambiguation approaches (47).

To our knowledge, Caron and van Eck's method represents the state-of-the-art approach to author disambiguation in large-scale bibliometric analysis. The method's main source of error stems from authors with very few publications. For this reason, we restrict our analysis to authors with at least five publications listed in WoS. This gives us a starting point of 4,995,039 disambiguated author names. For these, we identify the most common WoS subject category that they have published in. These subject categories are tied to the journals listed in WoS. When journals are assigned to more than one category, we count the reciprocal of the number of categories. The journal subject categories can be further tied to the OECD fields of science classification (48). Clarivate Analytics offers a conversion table between the WoS subject categories and the OECD fields. We use this conversion table to identify and remove all authors with a majority of publications published within the humanities, social sciences, and engineering and technology, leaving only those from medicine and health sciences, natural sciences, and agricultural sciences. We removed 825,378 authors by using these exclusion criteria. An additional 125,690 papers were removed because their subfield had a coverage of less than 66.7% of their references in WoS. This number was found by averaging the ratio of covered references and total references in a given subfield. This leaves 4,042,612 authors, which we consider to be the relevant population of all researchers in the three selected main fields.

Linked to these authors are 25,986,133 distinct papers, limited to full research papers and reviews published in journals. Letters, editorials, comments, and other document types are not included here. The papers are linked to the authors through 92,856,807 authorships, meaning that the average paper has 3.57 authors and the average author has published 22.97 papers.

Author-disambiguation methods are known to perform less well for East Asian names. Errors may occur, where bundles of authors with similar names are merged. As a robustness check we juxtaposed developments in the average number of papers per author for select countries (China, Denmark, Germany, Japan, South Korea, Taiwan, and the United States) using fractional paper counts. As displayed in *SI Appendix*, Fig. S4 the average paper rates for Chinese and Taiwanese authors remain relatively stable over time at levels comparable to authors in Denmark, Germany, and the United States. If the algorithm were systematically more likely to bundle East Asian authors with similar names, we would expect to see higher (and increasing) mean number of papers per author in China and Taiwan compared to the reference countries, but this is not the case.

Another possible bias concerns variations in the disambiguation method's precision across the observed time period. If the algorithm's precision varies for early authors in our dataset compared to more recent ones, this may indirectly increase the observed trend toward citation concentration. While we cannot categorically exclude the possibility of such bias, we expect its influence to be marginal for the following reasons: 1) we have excluded all authors with ≤4 WoS publications, hereby removing the less-reliable author profiles and (expectedly) improving the (average) precision for the remaining set; 2) extant research suggests a 95% precision of the algorithm (47), and since authors with few publications make up a large proportion of the full-author sample, this considerably restricts the possible variation that one would expect to see across strata (e.g., in the precision for the elite compared to other strata, or between younger and older researchers).

Citation Inflation and Field Normalization. WoS is growing annually, both due to a general increase in the number of published articles and to the inclusion of new journals in the database. *SI Appendix*, Fig. S5 displays the growth in N papers and references for the entire WoS, and for the specific subset included in our study.

Given the increase in the production of papers and references over time, the general value of a citation will depend on when it was produced. In the scientometric literature, this phenomenon is known as citation inflation (49). While field normalization and especially normalization by publication year should remove a large part of the temporal effect on a citation's value (e.g., papers published in 2000 are weighted according to other papers published only in the same year), it is unclear if an additional element of inflation may be at play for certain types of papers. Therefore, we compute a field-normalized, and inflation-corrected citation score, $nics$. First, we define the inflated citation value of any citing reference in year y as $\rho(y)$, the reciprocal of the sum of references, r, given in the same year:

$$\rho(y) = \left(\sum r(y) \right)^{-1}.$$

Thus, any reference given is weighted by the sum of all references given the same year, meaning that newer references are worth less than older ones, as there are more references in play. The inflation-corrected citation score, ics, then becomes the sum of all inflated citation values of articles, j, citing it:

$$ics = \sum_{j=1}^{n} \rho_j.$$

This is by definition a very small number, as every ρ is in the magnitude 10^{-7}. However, we also normalize these scores by field. To do this, we first calculate the mean ics for all fields, f, for all years y, individually. We denote this simple arithmetic mean as $mics$, in line with the mean citation score, mcs, used by Waltman and van Eck (50), defining the field-normalization technique used here. The $nics$ of a paper published in field f and year y is then

$$nics(f, y) = ics(f, y)/mics(f, y).$$

This corresponds to the state-of-the art approach to field normalization (50), ncs, but uses inflation-corrected citations instead of raw citations in all places. When using this normalization, all values are rescaled to be centered around 1, with a lower bound at 0 (uncited) but no upper bound.

We use a field delineation based on the bibliographic properties of articles, as proposed by ref. 50. The same fields are used in the Leiden ranking (51). The algorithm generates a little more than 4,000 clusters of articles with similar reference and citation patterns.

Inequality Measures. We use the Gini coefficient to measure inequality in citation distributions. The coefficient is calculated with the *ineq* package in R (52). Since we rely on inflation-adjusted, field-normalized citation indices, our use of Gini coefficients should be robust to "marginal biases" resulting from increases in the total number of authors and articles (53). *SI Appendix*, Fig. S6 replicates the plot of Gini coefficients presented in Fig. 3*B* with a fixed number of authors across years and produce near-identical results.

Data Availability. Bibliometric data have been deposited in GitHub.com, https://github.com/ipoga/elite_citations (54).

1. M. Mulkay, The mediating role of the scientific elite. *Soc. Stud. Sci.* **6**, 445–470 (1976).
2. J. R. Cole, S. Cole, The Ortega Hypothesis: Citation analysis suggests that only a few scientists contribute to scientific progress. *Science* **178**, 368–375 (1972).
3. Y. Xie, Sociology of science. "Undemocracy": Inequalities in science. *Science* **344**, 809–810 (2014).
4. R. G. Ehrenberg, Studying ourselves: The academic labor market. *J. Labor Econ.* **21**, 267–287 (2002).
5. P. E. Stephan, *How Economics Shapes Science* (Harvard University Press, Cambridge, MA, 2012).
6. P. D. Allison, J. A. Stewart, Productivity differences among scientists: Evidence for accumulative advantage. *Am. Sociol. Rev.* **39**, 596–606 (1974).
7. P. D. Allison, J. S. Long, T. K. Krauze, Cumulative advantage and inequality in science. *Am. Sociol. Rev.* **47**, 615–625 (1982).
8. P. Azoulay, T. Stuart, Y. Wang, Matthew: Effect or fable? *Manage. Sci.* **60**, 92–109 (2012).
9. T. Bol, M. de Vaan, A. van de Rijt, The Matthew effect in science funding. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4887–4890 (2018).
10. R. K. Merton, The Matthew effect in science. The reward and communication systems of science are considered. *Science* **159**, 56–63 (1968).
11. A. M. Petersen, O. Penner, Inequality and cumulative advantage in science careers: A case study of high-impact journals. *EPJ Data Sci.* **3**, 1–25 (2014).
12. A. M. Petersen et al., Reputation and impact in academic careers. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15316–15321 (2014).
13. A. M. Petersen, W. S. Jung, J. S. Yang, H. E. Stanley, Quantitative and empirical demonstration of the Matthew effect in a study of career longevity. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 18–23 (2011).
14. D. J. D. S. Price, A general theory of bibliometric and other cumulative advantage processes. *J. Am. Soc. Inf. Sci.* **27**, 292–306 (1976).
15. H. Zuckerman, *Scientific Elite: Nobel Laureates in the United States* (Transaction Publishers, 1977).
16. K. Aagaard, A. Kladakis, M. W. Nielsen, Concentration or dispersal of research funding? *Quantitative Sci. Studies* **1**, 117–149 (2020).
17. Y. Ma, B. Uzzi, Scientific prize network predicts who pushes the boundaries of science. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 12608–12615 (2018).

Nielsen and Andersen
Global citation inequality is on the rise

18. P. Bourdieu, The specificity of the scientific field and the social conditions of the progress of reason. *Soc. Sci. Inf. (Paris)* **14**, 19–47 (1975).

19. T. S. Kuhn, *The Structure of Scientific Revolutions* (University of Chicago Press, 1970).

20. E. B. Madsen, K. Aagaard, Concentration of Danish research funding on individual researchers and research topics: Patterns and potential drivers. https://osf.io/preprints/socarxiv/j874c/ (27 April 2020).

21. P. Azoulay, C. Fons-Rosen, J. S. G. Zivin, Does science advance one funeral at a time? *Am. Econ. Rev.* **109**, 2889–2920 (2019).

22. M. Planck, *Scientific Autobiography and Other Papers* (Philosophical Library, New York, 1950).

23. J. R. Cole, S. Cole, *Social Stratification in Science* (University of Chicago Press, Chicago, 1973).

24. S. S. Blume, R. Sinclair, Chemists in British universities : A study of the reward system in science. *Am. Sociol. Rev.* **38**, 126–138 (1973).

25. A. J. Lotka, The frequency distribution of scientific productivity. *J. Wash. Acad. Sci.* **16**, 317–323 (1926).

26. D. J. de Solla Price, *Little Science, Big Science* (Columbia University Press, New York, 1963).

27. V. Larivière, B. Macaluso, É. Archambault, Y. Gingras, Which scientific elites? On the concentration of research funds, publications and citations. *Res. Eval.* **19**, 45–53 (2010).

28. P. D. Allison, Inequality and scientific productivity. *Soc. Stud. Sci.* **10**, 163–179 (1980).

29. E. Caron, N. J. van Eck, "Large scale author name disambiguation using rule-based scoring and clustering" in Proceedings of the Science and Technology Indicators conference, E. Noyons, Ed. (CWTS, Leiden University, Leiden, Netherlands, 2014), pp. 79–86.

30. J. A. Evans, Electronic publication and the narrowing of science and scholarship. *Science* **321**, 395–399 (2008).

31. A. Varga, Shorter distances between papers over time are due to more cross-field references and increased citation rate to higher-impact papers. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 22094–22099 (2019).

32. A. L. Barabási, C. Song, D. Wang, Publishing: Handful of papers dominates citation. *Nature* **491**, 40 (2012).

33. R. K. Pan, A. M. Petersen, F. Pammolli, S. Fortunato, The memory of science: Inflation, myopia, and the knowledge network. *J. Informetrics* **12**, 656–678 (2018).

34. V. Larivière, Y. Gingras, É. Archambault, The decline in the concentration of citations, 1900-2007. *J. Am. Soc. Inf. Sci. Technol.* **60**, 858–862 (2009).

35. S. Yang, F. Ma, Y. Song, J. Qiu, A longitudinal analysis of citation distribution breadth for Chinese scholars. *Scientometrics* **85**, 755–765 (2010).

36. S. Wuchty, B. F. Jones, B. Uzzi, The increasing dominance of teams in production of knowledge. *Science* **316**, 1036–1039 (2007).

37. D. Fanelli, V. Larivière, Researchers' individual publication rate has not increased in a century. *PLoS One* **11**, e0149504 (2016).

38. W. D. Figg *et al.*, Scientific collaboration results in higher citation rates of published articles. *Pharmacotherapy* **26**, 759–767 (2006).

39. E. S. Vieira, J. A. Gomes, Citations to scientific articles: Its distribution and dependence on the article features. *J. Informetrics* **4**, 1–13 (2010).

40. S. Milojević, F. Radicchi, J. P. Walsh, Changing demographics of scientific careers: The rise of the temporary workforce. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 12616–12623 (2018).

41. T. Heinze, A. Jappe, D. Pithan, From North American hegemony to global competition for scientific leadership? Insights from the Nobel population. *PLoS One* **14**, e0213916 (2019).

42. C. J. Gomez, A. C. Herman, P. Parigi, Moving more, but closer: Mapping the growing regionalization of global scientific mobility using ORCID. *J. Informetrics* **14**, 101044 (2020).

43. Y. Xie, A. A. Killewald, *Is American Science in Decline?* (Harvard University Press, Cambridge, MA, 2012).

44. B. F. Jones, S. Wuchty, B. Uzzi, Multi-university research teams: Shifting impact, geography, and stratification in science. *Science* **322**, 1259–1262 (2008).

45. J. P. A. Ioannidis, J. Baas, R. Klavans, K. W. Boyack, A standardized citation metrics author database annotated for scientific field. *PLoS Biol.* **17**, e3000384 (2019).

46. Y. Katz, U. Matter, Metrics of inequality: The concentration of resources in the U.S. biomedical elite. *Sci. Cult. (Lond)* **0**, 1–28 (2019).

47. A. Tekles, L. Bornmann, Author name disambiguation of bibliometric data: A comparison of several unsupervised approaches. https://arxiv.org/abs/1904.12746 (29 April 2019).

48. OECD Working Party of National Experts on Science and Technology Indicators, "Revised fields of science and technology (FOS) in the Frascati Manual" (Tech. Rep. JT03222603, DSTI/EAS/STP/NESTI(2006)19/FINAL, OECD Publishing, Paris, 2007).

49. A. M. Petersen, R. K. Pan, F. Pammolli, S. Fortunato, Methods to account for citation inflation in research evaluation. *Res. Policy* **48**, 1855–1865 (2019).

50. L. Waltman, N. J. van Eck, A new methodology for constructing a publication-level classification system of science. *J. Am. Soc. Inf. Sci. Technol.* **63**, 2378–2392 (2012).

51. L. Waltman *et al.*, The Leiden ranking 2011/2012: Data collection, indicators, and interpretation. *J. Am. Soc. Inf. Sci. Technol.* **63**, 2419–2432 (2012).

52. A. Zeileis, ineq: Measuring inequality, concentration, and poverty" (Version 02-13, R Package, (2014). https://cran.r-project.org/web/packages/ineq/index.html.

53. L. Kim, C. Adolph, J. D. West, K. Stovel, The influence of changing marginals on measures of inequality in scholarly citations: Evidence of bias and a resampling correction. *Sociol. Sci.* **7**, 314–341 (2020).

54. J. P. Andersen, M. W. Nielsen. Citation concentration and inequality. GitHub. https://github.com/ipoga/elite_citations. Deposited 21 October 2020.

www.manaraa.com