PNAS RESEARCH ARTICLE | SOCIAL SCIENCES

Ranking mobility and impact inequality in early academic careers

Ye Sun[a], Fabio Caccioli[a,b,c], and Giacomo Livan[a,b,1]

Edited by Yu Xie, Princeton University, Princeton, NJ; received March 30, 2023; accepted July 17, 2023

How difficult is it for an early career academic to climb the ranks of their discipline? We tackle this question with a comprehensive bibliometric analysis of 57 disciplines, examining the publications of more than 5 million authors whose careers started between 1986 and 2008. We calibrate a simple random walk model over historical data of ranking mobility, which we use to 1) identify which strata of academic impact rankings are the most/least mobile and 2) study the temporal evolution of mobility. By focusing our analysis on cohorts of authors starting their careers in the same year, we find that ranking mobility is remarkably low for the top- and bottom-ranked authors and that this excess of stability persists throughout the entire period of our analysis. We further observe that mobility of impact rankings has increased over time, and that such rise has been accompanied by a decline of impact inequality, which is consistent with the negative correlation that we observe between such two quantities. These findings provide clarity on the opportunities of new scholars entering the academic community, with implications for academic policymaking.

impact ranking mobility | inequality | science of science

Significance

The mobility of researchers in impact rankings is widely considered an important indicator of opportunity because it describes upward movements within or across academic stratification. We find that the top- and bottom-ranked authors have the lowest impact ranking mobility, and that the overall mobility has increased in most disciplines over the last few decades, making it easier for a newcomer to establish themselves in the field. Furthermore, we find that the higher the impact inequality of a discipline, the lower its ranking mobility, a finding that is reminiscent of empirical observations made in the Social Sciences that wealth inequality and social mobility tend to be negatively correlated, and which has numerous implications for academic policymaking.

Recognition and rewards in modern academia are highly stratified (1). The citations received by authors and their work (2–4), the amount of funding allocated to research projects (5, 6), and the number of prizes awarded to scientists (7) are all very unevenly distributed quantities.

The Matthew effect explains uneven outcomes in terms of self-reinforcing dynamics (8–10), suggesting that an author's early success or fortune translate—through a process of cumulative advantage—into higher chances of further future success. There are several notable manifestations of the Matthew effect in academia. For instance, faculty at US universities are significantly more likely to have at least one parent with a PhD compared to the general population (11) and are very likely to have been trained in a small group of elite institutions (12, 13). In a similar fashion, an author's career impact is often significantly correlated with the visibility and prestige of their early career mentors (14) and/or coauthors (15), and the likelihood of publishing as senior author in top interdisciplinary venues often boils down to being "chaperoned" to such venues by well-established scientists (16). Quite naturally, this has resulted in increasing inequality, with already highly cited authors receiving a rising share of citations (17), leading to the formation of "rich clubs" (18). These dynamics are responsible for an effective narrowing of the literature, i.e., a small minority of papers capturing most of the attention (19, 20), and for a more unequal allocation of funding across research institutions (5, 21).

The above results would suggest that a researcher's future impact overly depends on the initial conditions of their career. This, in turn, would imply that mobility within the impact ranking of a given discipline is highly restricted, which may stifle scientific innovation and fairness.

On the other hand, a few studies have illustrated somewhat opposite mechanisms and trends, showing that sustained career impact can emerge from early career failures and challenges, e.g., near-miss grant applications (22) or working on an interdisciplinary subject with low recognition (23), which facilitate mobility in impact rankings.

Motivated by the above findings, in this paper, we examine the dynamics of academic impact rankings, that is the ranking of authors in a given discipline based on the citations they accrue over a period of time—we are fully aware that citations are not a comprehensive measure of academic impact, yet citation-based indicators are the standard in the literature as they are easily quantifiable (24–26). We operationalize our analysis in terms of ranking mobility—the extent to which authors can move across the ranks of their discipline—and impact inequality—the concentration of the citations received by authors in a discipline. In analogy to what has been reported in studies of wealth dynamics in the Social Sciences (27), we intuitively expect such two

Author affiliations: [a]Department of Computer Science, University College London, London WC1E 6EA, United Kingdom; [b]Systemic Risk Centre, London School of Economics and Political Science, London WC2A 2AE, United Kingdom; and [c]London Mathematical Laboratory, 8 Margravine Gardens, London WC 8RH, United Kingdom

Author contributions: Y.S., F.C., and G.L. designed research; Y.S. performed research; Y.S., F.C., and G.L. analyzed data; and Y.S., F.C., and G.L. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: g.livan@ucl.ac.uk.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2305196120/-/DCSupplemental.

Published August 14, 2023.

quantities to display a negative relationship. This is because, due to the Matthew effect and similar self-reinforcing dynamics, it should be hard to move across the ranks of a discipline characterized by a highly uneven distribution of citations. At the same time, high mobility should prevent a concentration of citations at the very top of a discipline's ranks.

Our analysis will focus on cohorts of authors whose careers began in the same year. Each author will be characterized by their position in the rank of their cohort, which we will monitor over the first ten career years. This allows us to compare authors that began their careers under similar "environmental" conditions (in terms, e.g., of funding availability, volume of publications, maturity of their discipline, etc.) and competing for the same attention pool (in terms of potential readership of their work), therefore discounting the heterogeneity that would arise from the simultaneous presence of authors from different "generations" and focusing only on within-cohort inequality and ranking mobility.

In the following, we examine the ranking mobility of authors in different strata of their cohort, and we find that mobility is the lowest for the top- and bottom-ranked authors. We study the evolution of ranking mobility by implementing a simple model akin to a random walk, which allows us to quantify mobility in different disciplines and epochs in terms of the model's "diffusion" coefficient. We observe that over time author cohorts have experienced, on average, increasing mobility and decreasing inequality in their discipline's impact rankings, as well as the negative relationship between mobility and inequality that we hypothesized.

Results

Exploring Authors' Impact Ranking Mobility. We begin our analysis by quantifying an author's mobility in the scientific impact ranking of their discipline. To this end, we create author publication profiles by employing a high-precision name disambiguation algorithm on Web of Science data (WoS, *Materials and Methods*), yielding a total of 5,194,173 authors active in 57 different disciplines spanning four macroareas of WoS data, i.e., Life Sciences & Medicine, Physical Sciences, Social Sciences, and Technology (*SI Appendix*, Table S1). In line with numerous studies on scientific careers (e.g., ref. 28), we quantify the impact of each publication with the number of citations received within 5 y after publication (c_5). We quantify an author's aggregate impact over a period of time as the sum of the c_5 scores of their publications during such period.

We group the authors in our dataset into cohorts based on their career starting year, which we identify with the year of their first publication on record, and examine their impact mobility between the first 5 y and second 5 y (i.e., years 6 to 10) of their careers. Namely, we rank the authors in each cohort based on their aggregate impact over their first five and second five career years and divide both rankings into deciles. In line with the literature (see, e.g., refs. 15 and 28), note that only authors who have published at least one paper in each 5 y window are considered in our analysis. Our motivation to do this is to compute mobility as the relative position in the ranking of authors with an active publication record over an extended period of time. In *SI Appendix*, Fig. S1 we report the time evolution of the retention rate, i.e., the fraction of authors who remain active during their second five career years, showing that around 20% of authors in our dataset match the above criteria on publication activity. Throughout the rest of the paper, the rankings we will refer to will always be restricted to such fraction of authors.

We then characterize impact mobility between the two time windows by computing 10×10 column-stochastic transition matrices, with entries given by the empirically estimated probability of an author moving from one decile to another. Fig. 1*A* shows the heat-maps of the transition matrices for author cohorts (with careers staring in 2000) in four disciplines, i.e., Biotechnology, Materials Sciences, Business and Economics, and Chemistry. Consistently with those examples, most disciplines are characterized by transition matrices displaying concentration around the diagonal, indicating that most authors typically remain in the same decile or move to an adjacent one. Notably, in most disciplines, we observe comparably large probabilities at each end of the diagonal in transition matrices, signaling a particularly high stability of the Top and Bottom of the impact ranking over time.

To check whether these phenomena are common among authors entering the academic workforce at different times, we compute the average mobility of impact ranking for authors starting their careers in 1998, 2003, and 2008, respectively. We quantify impact mobility as the absolute difference $|\Delta Q|$ between an author's decile rank in the second and first five career years. We compare these results with a null model obtained by randomly reshuffling the authors' impact in the second five career years, so as to completely decorrelate it with respect to impact over the first 5 y. Fig. 1*B* shows the average absolute mobility of impact ranking $|\Delta\overline{Q}|$ as a function of an author's decile rank Q in their first five career years, as obtained both from the empirical data and from the null model. As one can see, due to the boundaries at the Top and Bottom of the ranking, the null model leads to an apparent U-shaped relationship between $|\Delta\overline{Q}|$ and Q, with larger moves taking place for authors at the Top and Bottom of the rankings and smaller moves for those in the Middle of the ranking. In contrast, the corresponding relationship is much more flat in the empirical data, and, on average, authors in each decile have lower mobility than in the null model baseline. In particular, the largest mobility gap between the empirical data and the null model occurs for Top- and Bottom-ranked authors, further highlighting the stability of such portions of impact rankings.

To further characterize author impact ranking mobility, we develop a simple model akin to a random walk aimed at capturing the aforementioned observed tendency of most authors to remain in the same decile or move to an adjacent one. Specifically, we assume the probability of an author moving from decile j to decile i to be given by

$$P_{ij} = e^{-\Delta_{ij}^2/D} / \sum_{\ell} e^{-\Delta_{\ell j}^2/D}, \qquad [1]$$

where Δ_{ij} is the distance between the two deciles, with $\Delta_{ij} = |i-j|,\ \forall i, j$. In the random walk analogy, the parameter D in Eq. **1** plays the role of the diffusion coefficient, controlling to which degree authors are able to "diffuse" toward higher/lower deciles of impact rankings. The higher the value of D, the more uniform the probabilities in a discipline's transition matrix, resulting in higher mobility (*SI Appendix*, Fig. S2). With a given value of D, we are able to compute all the entries of the transition matrix successively. To capture and compare the overall mobility of author cohorts, we fit the optimal value of D for authors in each discipline and year by minimizing the Frobenius norm of the difference between the transition matrices of the empirical data and those of the random walk model.

In Fig. 1*C*, we show the transition matrices for Biotechnology, Materials Sciences, Business and Economics, and Chemistry at

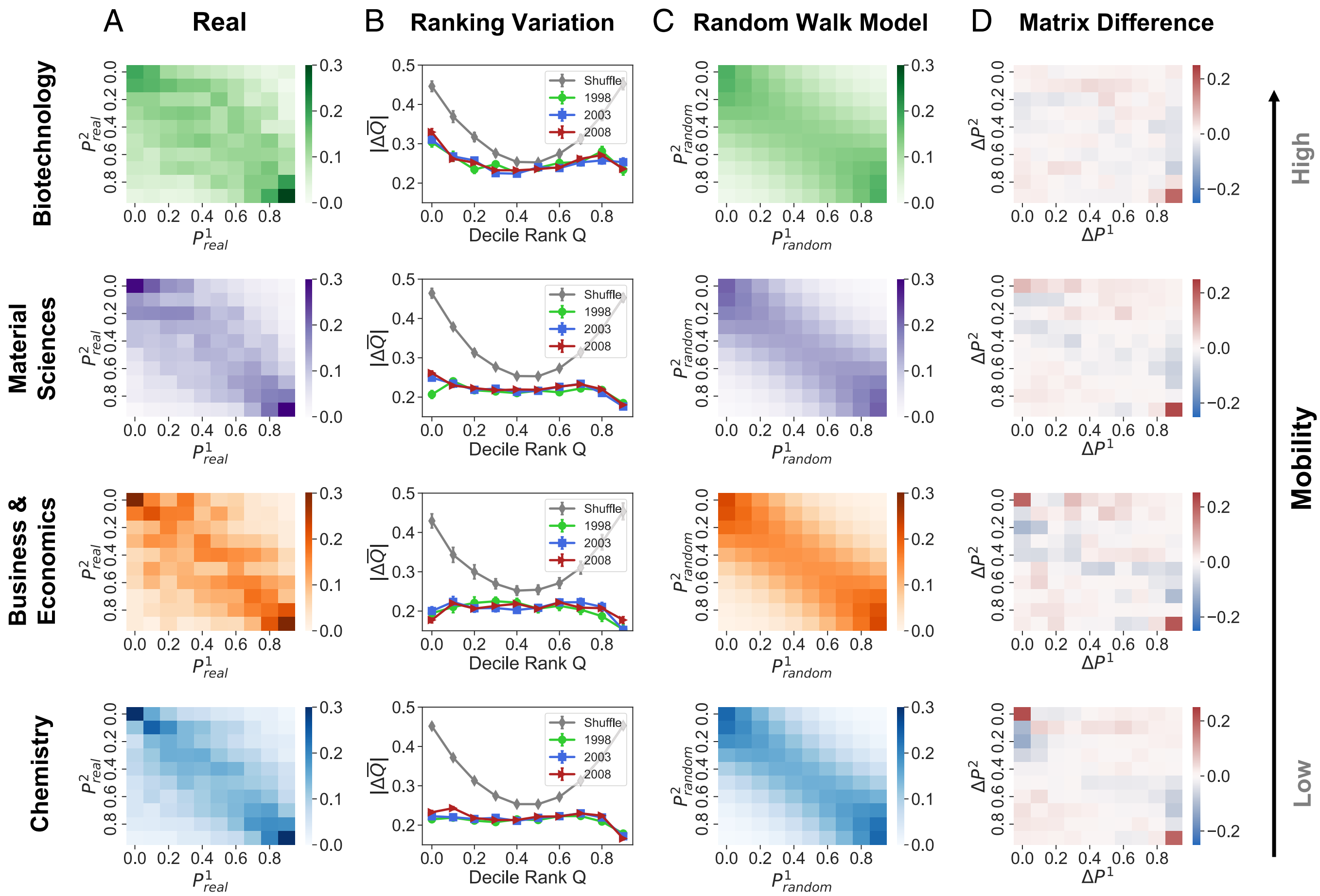


Fig. 1. Author impact ranking mobility in Biotechnology, Materials Sciences, Business and Economics, and Chemistry. (*A*) Transition matrix of author impact rankings between the first and second 5 y of their careers. Here, we consider the cohorts of authors who started their careers in the year 2000, i.e., with first and second five career years covering 2000 to 2004 and 2005 to 2009, respectively. The scientific impact of an author over a period of time is calculated as the total number of citations received (within 5 y of publication) by their papers published over that period. Based on that we compute the decile rank of an author in their discipline during the first and second five career years, which we indicate with P^1_{real} and P^2_{real}. (*B*) Average absolute variation of impact ranking mobility $|\Delta\overline{Q}|$ for each author percentile rank group Q. Results are shown for three different cohorts of authors, starting their careers in 1998, 2003, and 2008, respectively. These are compared with the results from a null model obtained by randomly reshuffling the impact of authors during their second five career years. Error bars represent the SE of the mean. (*C*) Transition probability matrix of authors in our random walk model (Eq. **1**). The optimal values of D for Biotechnology, Materials Sciences, Business and Economics, and Chemistry are equal to 0.35, 0.23, 0.20, and 0.18 respectively, implying that the mobility of authors in these disciplines decreases from Top to Bottom in the figure. (*D*) Differences between the empirical transition matrices (column *A*) and those obtained from the random walk model (column *C*).

the optimal values of D, equal to 0.35, 0.23, 0.20, and 0.18, respectively. From Top to Bottom, we can observe that—as expected based on such values of D—transition matrices become less uniform, with more probability concentrated around the diagonal, signaling a lower mobility. Fig. 1*D* depicts the element-wise differences ΔP between empirical transition matrices and those obtained from optimal values of D. Generally speaking, we find that the random walk model captures well the overall characteristics of author ranking mobility (with most values of ΔP nearly equal to 0). In line with the results shown in Fig. 1*B*, the model underestimates the probabilities for authors in Top and Bottom deciles to remain in such groups.

We then proceed to test whether the excess of stability in the Top and Bottom 10% is persistent over time. In order to do that, we report the difference between transition probabilities ΔP in empirical data and in the random walk model across all disciplines at the aggregate level in Fig. 2 *A* and *B*. Overall, the average values of ΔP remain positive and roughly constant throughout our entire period of observation, indicating that the stability at the Top and Bottom of impact rankings has been consistently higher, and that the gap has not increased over time. However, we find a higher degree of stability at the top of impact rankings, as testified by a narrower distribution of ΔP with a higher average value (0.19) with respect to the Bottom of impact rankings (average ΔP equal to 0.10). We further validate these results at the level of WoS macroareas, arriving at the same conclusions, but the extent of the gaps between the data and the random walk model is different between the macroresearch areas (*SI Appendix*, Figs. S3 and S4). Notably, for authors in the Top 10%, the average value of ΔP in Physical Sciences is statistically higher than that of the other three macroareas, while for the Bottom 10% authors, the average value of ΔP in Physical Sciences is the lowest ($P < 0.001$, two tailed t-tests). This suggests that authors in Physical Sciences generally have a very stable Top portion of the impact ranking, but a relatively flexible Bottom portion compared with other macroareas.

Although the overall degree of differences remains roughly constant over time when aggregating over all disciplines or at the level of macroareas, we do observe some specific temporal trends at the level of individual disciplines (*SI Appendix*, Figs. S5–S12). For example, the gap ΔP for the Top 10% ranked authors in Chemistry has been steadily increasing from 1986 to 2008

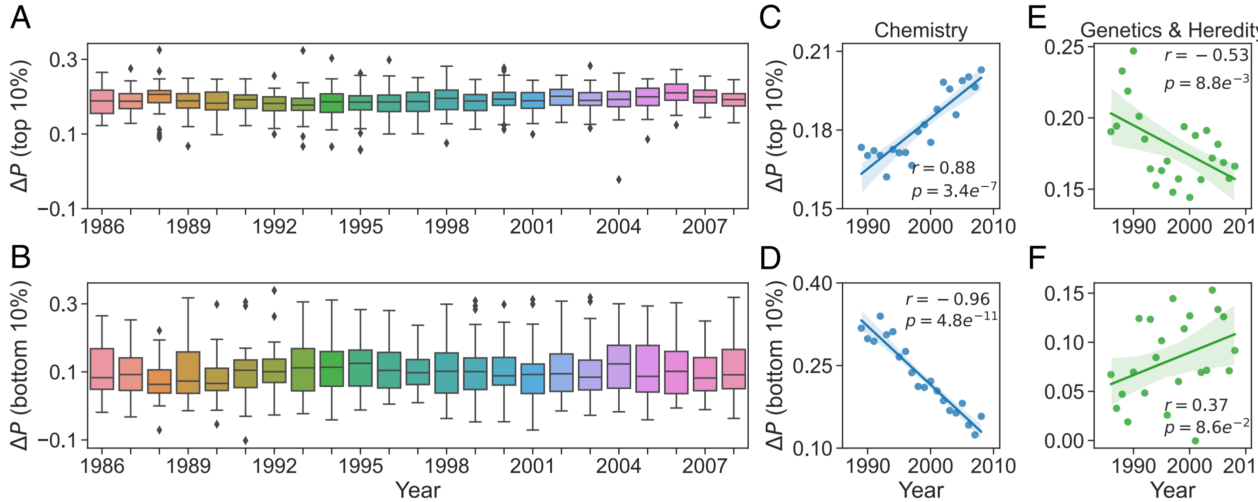


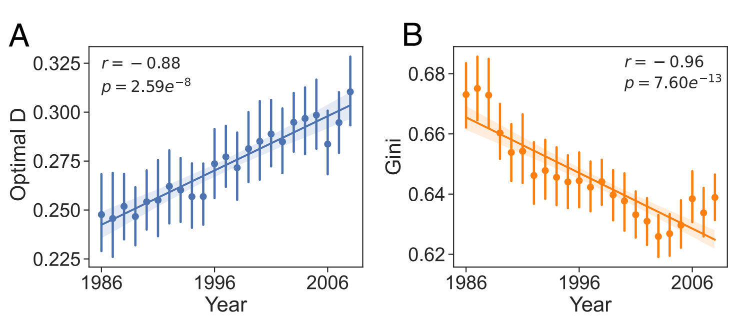
Fig. 2. Evolution of mobility for authors in the Top and Bottom 10% of impact rankings of their discipline. We show box plots of the overall evolution of ΔP for authors in the (*A*) Top 10% and (*B*) Bottom 10% of impact rankings obtained by aggregating over all disciplines. ΔP measures the difference of probability between empirical transition matrices and those obtained from the random walk model in Eq. **1**. We then report examples of different temporal evolutions of ΔP for different disciplines. Compared to the random walk model, the Top 10% of the impact ranking in Chemistry has become more stable over time (*C*), whereas the Bottom 10% has become more mobile (*D*). The opposite holds in the case of Genetics and Heredity, where stability has decreased in the Top 10% (*E*) and increased in the Bottom 10% (*F*). In figures (*C*)–(*F*) the solid lines and shaded areas represent regression lines and 95% confidence level intervals, respectively. The Pearson coefficients—and the corresponding *P*-values—obtained from the regressions are shown in the figures.

(Fig. 2*C*), while the gap for the Bottom 10% has been steadily decreasing (Fig. 2*D*). Conversely, we observe opposite trends in Genetics and Heredity (Fig. 2 *E* and *F*), with a decline in stability for Top-ranked authors, and an increase for Bottom-ranked ones. Taken together, these findings demonstrate that, on average, Top- and Bottom-ranked authors consistently experience higher stability, but with notable differences in terms of temporal patterns across different disciplines.

Evolution of Authors' Mobility and Inequality. As anticipated in the introduction, we hypothesize that Higher (Lower) levels of mobility in impact rankings should be accompanied by a Lower (Higher) level of concentration in the distribution of citations received by authors. If this is the case, we should observe opposite trends over time for these two quantities, as well as an overall negative relationship between them.

We first proceed to examine how mobility in author impact rankings has evolved over time on an aggregate level across all disciplines we consider. We estimate the mobility of different author cohorts by estimating the optimal value of the parameter D in Eq. **1** for each discipline and year. As can be seen in Fig. 3*A*, there has been a significant increase in mobility over the two decades in our analysis. In other words, the later an author started their academic career the lower—on average—the likelihood that their future impact depends on their past impact.

We then gauge the concentration of author impact distributions by calculating the Gini coefficient (one of the most widely used indicators of inequality) for each author cohort and each discipline. Here, an author's impact is defined as the total number of citations received for all the papers published in the first 5 y of their careers within 5 y of publication. Indeed, aggregating over all disciplines we observe that the overall concentration of impact at the level of cohorts has steadily decreased for authors who started their careers between 1986 and 2003, albeit with a rebound in impact inequality from 2003 onward (Fig. 3*B*).



C

Macro areas	N	$N_D^{\uparrow}$	$N_G^{\downarrow}$	N^-
1. Life Sciences & Biomedicine	30	25	22	18
2. Physical Sciences	12	4	7	6
3. Social Sciences	5	1	5	1
4. Technology	10	4	8	4

Fig. 3. Evolution of mobility in impact rankings and inequality in impact distributions. (*A*) Mobility of author impact rankings (measured in terms of optimal D, see Eq. **1**) for cohorts of authors starting their careers between 1986 and 2008. (*B*) Gini coefficients of impact distributions for cohorts of authors starting their careers between 1986 and 2008. The Gini coefficient is calculated based on the cumulative number of citations authors have received from the papers published during their first five career years, within 5 y of publication. In subplots *A* and *B*, the solid line and the shaded area represent regression lines and 95% confidence level intervals, respectively. The error bars denote the 95% confidence intervals of the distribution of optimal D. Each regression has also been annotated with the corresponding Pearson correlation coefficient r and its *P*-value. (*C*) Number of disciplines with statistically significant trends over time within each macroarea (with Pearson correlation coefficient P value < 0.1). N denotes the number of disciplines in each macroarea. $N_D^{\uparrow}$ and $N_G^{\downarrow}$ respectively denote the number of disciplines with a statistically significant increasing trend in mobility and a decreasing trend in inequality. N^- counts the number of disciplines with a statistically significant negative correlation between mobility and inequality.

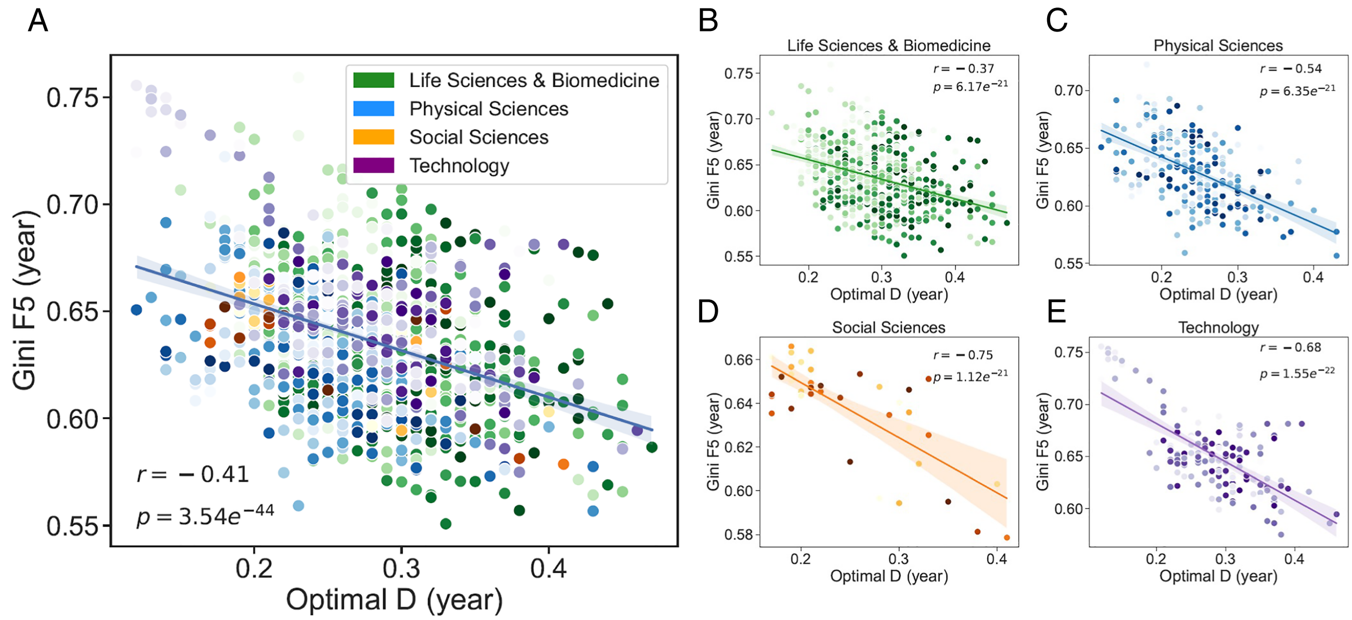


Fig. 4. Correlation between mobility in impact rankings and inequality in impact distributions for all author cohorts across all disciplines (*A*), and for author cohorts in Life science and Biomedicine (*B*), Physics Sciences (*C*), Social Sciences (*D*) and Technology (*E*). The Gini coefficient is calculated based on the cumulative number of citations authors have received from the papers published during their first five career years, within 5 y of publication. In each subplot, each circle represents an author cohort in a given discipline. The shading of the circles, from light to dark, indicates the year in which a cohort of authors started their careers, from 1986 to 2008. The solid line and the shaded area indicate regression lines and 95% confidence level interval, respectively. Each regression has also been annotated with the corresponding Pearson correlation coefficient *r* and its *P*-value.

We validate these results by repeating the analysis at the level of individual disciplines (see *SI Appendix*, Figs. S13–S16 for trends in mobility and in *SI Appendix*, Figs. S17–S20 for trends in inequality). Fig. 3*C* summarizes the number of disciplines with statistically significant increasing trends of mobility and decreasing trends of inequality, confirming the generality of the trends observed above. In Fig. 4*A*, we provide an overall representation of the negative relationship between mobility and inequality, by depicting each discipline and author cohort as a point. The negative correlation is still found at the level of four macroareas (Fig. 4 *B*–*E*) and in most individual disciplines (Fig. 3*C* and *SI Appendix*, Figs. S21–S24). These results remain qualitatively the same when computing inequality over the full first ten career years of authors (*SI Appendix*, Fig. S25).

The observed differences in mobility and inequality across disciplines may reflect the different underlying opportunities for researchers entering an academic discipline. To rank the disciplines in terms of their overall impact mobility and inequality, we first compute a single discipline-specific value of the parameter D by calibrating our random walk model on the transition matrices of a given discipline for all years in our analysis. We do that by finding the value of the parameter which minimizes the sum of Frobenius norms of the differences between empirical transition matrices and those generated by the model. Similarly, we can characterize the overall inequality of a discipline by computing its average annual Gini coefficient. This naturally leads to a ranking of the disciplines in terms of their overall ranking mobility and inequality. In Table 1, we report the top/bottom 5 disciplines in both dimensions. We find that the discipline characterized by the highest (lowest) overall mobility is Biophysics (Astronomy & Astrophysics), whereas the discipline characterized by the highest (lowest) overall inequality is Research & Experimental Medicine (Microbiology).

Discussion

In this paper, we have studied the careers of a large pool of authors across 57 disciplines, focusing on impact ranking mobility and inequality at the level of cohorts, that is we only compare authors

Table 1. Disciplines with highest or lowest mobility and inequality

Top 5 mobility	Bottom 5 mobility	Top 5 inequality	Bottom 5 inequality
1. Biophysics	1. Astronomy & Astrophysics	1. Research & Experimental Medicine	1. Microbiology
2. Pediatrics	2. Science & Technology Other Topics	2. Cell Biology	2. Physiology
3. Nuclear Science & Technology	3. Business & Economics	3. Science & Technology Other Topics	3. Water Resources
4. Respiratory System	4. Chemistry	4. Gastroenterology & Hepatology	4. Geochemistry & Geophysics
5. Public Administration	5. Physics	5. Cardiovascular System & Cardiology	5. Food Science & Technology

with peers starting their careers in the same year. We document the existence of a statistically significant negative relationship between these quantities. This is reminiscent of observations made in the Social Sciences about the negative correlation between metrics of wealth inequality and metrics of social mobility, a phenomenon often referred to as the "Great Gatsby Curve" (27). From the perspective of this analogy, one could be tempted to identify ranking mobility in a discipline as the academic equivalent of social mobility in a country, and citations as the academic equivalent of wealth. Although academic impact is certainly a multifaceted concept, it is in fact undeniable that citations and citation-based indicators have become the currency of academic progression in a number of countries and academic systems (29), sometimes leading to unintended consequences on citation behaviors and patterns (18, 30). To push this analogy even further, a number of studies have shown that academic impact, as measured via citations, is partially inherited from mentors and/or senior collaborators (15, 16, 31, 32), quite similarly to family wealth which is passed on through generations. We wish to stress that here we are not endorsing an uncritical use of citations and citation-based indicators as a measure of academic impact. Quite the contrary, in light of our results, it is even clearer that such metrics should be properly contextualized.

When monitoring the temporal evolution of ranking mobility and inequality, we find that—in the majority of disciplines—the former has been on the rise while the latter has decreased over most of our window of analysis. When aggregating over all disciplines, we find the same result roughly over the first 15 y of our analysis. After that (i.e., around the year 2003), we observe an uptick in impact inequality (Fig. 3*B*). This is consistent with other findings in the literature (see, for instance, ref. 17). It should be noted that, in our case, such observation relies only on a handful of data points, and therefore should be monitored as more data become available over time. Also, we stress that our units of analysis are different from those typically considered in the literature, where each author is usually compared with all other authors in their discipline that are active at the same time, regardless of their seniority. We compare each author with those that started their careers in the same year. In other words, our measures of ranking mobility and inequality are based on conditional probabilities over cohorts rather than the entire population of active authors.

Our findings suggest that, over time, it has become easier for new authors to climb the ranks of their own cohorts. Intracohort mobility mostly takes place in the middle strata of the cohort, with the Lowest and Highest deciles being characterized by a notable lack of mobility (as well captured by the discrepancies with respect to our random walk model of mobility). It should be reminded here that these results hold on the subpopulation of authors obtained after removing those who did not publish at least one paper after their first five career years, i.e., our notion of mobility does not account for attrition, which could potentially lead to an underestimation of downward ranking mobility, particularly in the Lower deciles. Coming back to the previous analogy with the social sciences, this is reminiscent of a long-standing observation first made by Pareto (33) that most social mobility takes place in the Middle classes (34).

Materials and Methods

Dataset. The bibliometric data for this study are provided by the Web of Science (WoS). Publications are classified into 153 disciplines, which constitute a subject categorization scheme that is shared by all Web of Science product databases. One publication can be assigned to more than one discipline. Disciplines are further grouped into five broad macroareas: 1) Arts & Humanities; 2) Life Sciences & Biomedicine; 3) Social Sciences; 4) Physical Sciences; and 5) Technology. For the analysis, we select several of the largest disciplines within four macroareas: 30 disciplines in Life Sciences, 12 in Physical Sciences, 5 in Social sciences, and 8 in Technology. The disciplines in Arts & Humanities are excluded from our analysis because a large portion of articles in this macroarea do not receive any citations after publication. Note that the publications with more than 20 listed authors are not considered in our analysis.

Name Disambiguation Algorithm. The main focus of our work is to examine the mobility of authors' scientific impact rankings in their early career. Therefore, adequate disambiguation of author names in bibliometric databases is pivotal for any reliable analysis at the author level. The WoS dataset does not maintain unique author identifiers. To associate an author with all of their publications, we apply a state-of-the-art algorithm proposed by Caron and van Eck (35) to disambiguate all names of authors starting their academic career after 1986 in the entire WoS database. This approach has been validated by Tekles and Bornmann, who showed that it outperforms four other unsupervised disambiguation methods (36) in large-scale bibliometric analysis.

Specifically, this method quantifies the similarity between two author mentions using rule-based scoring and clustering. A set of criteria that rely on several paper-level and author-level attributes have been considered, including ORCID identifiers, names, affiliations, email addresses, coauthors, grant numbers, subject categories, journals, self-citations, bibliographic coupling, and cocitations. Each criterion is assigned a specific score, and the scores of all matching criteria are summed to give an overall similarity score for the two author mentions (*SI Appendix*, Table S2). The higher the similarity scores of the two author mentions, the more likely they are to be considered as the same author in real world. The threshold to decide whether two author mentions are sufficiently similar depends on the size of the corresponding name block (i.e., a group of authors sharing the same family name and first name initials). In general, the employed threshold is expected to increase with the block size class, as larger name blocks carry higher false positive rates, i.e., a higher probability of erroneously connecting different author mentions. To obtain the optimal threshold that maximizes the $F1$ score (a well-known metric used to assess the performance of classification algorithms) in each block size class, we apply this approach to a subset of author mentions annotated with ORCIDs as ground truth to determine whether author mentions belong to the same identity. In *SI Appendix*, Table S3 we report information about block size classes, the size of subsets with ground truth information, and the corresponding thresholds employed by the approach. For each size block class, the performance of the approach has been shown to exceed 90% in terms of $F1$ score. Based on these scoring rules and block-size dependent thresholds, the approach has then been applied to disambiguate the complete WoS bibliographic databases. See *SI Appendix* for more details on the process of name disambiguation and threshold selection.

Data, Materials, and Software Availability. All study data are included in the article and/or *SI Appendix*. The Web of Science data used in this study were obtained by one of the authors as part of the 2020 Eugene Garfield Award by Clarivate. The same data can be obtained through a subscription to Web of Science.

ACKNOWLEDGMENTS. Y.S. and G.L. acknowledge support from a Leverhulme Trust research project grant (RPG-2021-282).

1. J. R. Cole, S. Cole, Social stratification in science. *Am. J. Phys.* **42**, 923–924 (1974).
2. D. W. Aksnes, A. Rip, Researchers' perceptions of citations. *Res. Policy* **38**, 895–905 (2009).
3. J. Ruiz-Castillo, R. Costas, The skewness of scientific productivity. *J. Inf.* **8**, 917–934 (2014).
4. J. Ruiz-Castillo, R. Costas, Individual and field citation distributions in 29 broad scientific fields. *J. Inf.* **12**, 868–892 (2018).
5. A. Ma, R. J. Mondragón, V. Latora, Anatomy of funded research in science. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 14760–14765 (2015).

6. C. Bloch, M. P. Sørensen, The size of research funding: Trends and implications. *Sci. Public Policy* **42**, 30–43 (2015).
7. 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).
8. R. K. Merton, The Matthew effect in science: The reward and communication systems of science are considered. *Science* **159**, 56–63 (1968).
9. A. M. Petersen *et al*., Reputation and impact in academic careers. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15316–15321 (2014).
10. 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).
11. A. C. Morgan *et al*., Socioeconomic roots of academic faculty. *Nat. Hum. Behav.*, 1–9 (2022).
12. A. Clauset, S. Arbesman, D. B. Larremore, Systematic inequality and hierarchy in faculty hiring networks. *Sci. Adv.* **1**, e1400005 (2015).
13. K. H. Wapman, S. Zhang, A. Clauset, D. B. Larremore, Quantifying hierarchy and dynamics in US faculty hiring and retention. *Nature* **610**, 120–127 (2022).
14. Y. Ma, S. Mukherjee, B. Uzzi, Mentorship and protégé success in stem fields. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 14077–14083 (2020).
15. W. Li, T. Aste, F. Caccioli, G. Livan, Early coauthorship with top scientists predicts success in academic careers. *Nat. Commun.* **10**, 5170 (2019).
16. V. Sekara *et al*., The chaperone effect in scientific publishing. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 12603–12607 (2018).
17. M. W. Nielsen, J. P. Andersen, Global citation inequality is on the rise. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2012208118 (2021).
18. W. Li, T. Aste, F. Caccioli, G. Livan, Reciprocity and impact in academic careers. *EPJ Data Sci.* **8**, 20 (2019).
19. J. A. Evans, Electronic publication and the narrowing of science and scholarship. *Science* **321**, 395–399 (2008).
20. A. Varga, The narrowing of literature use and the restricted mobility of papers in the sciences. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2117488119 (2022).
21. K. Aagaard, A. Kladakis, M. W. Nielsen, Concentration or dispersal of research funding? *Quant. Sci. Stud.* **1**, 117–149 (2020).
22. Y. Wang, B. F. Jones, D. Wang, Early-career setback and future career impact. *Nat. Commun.* **10**, 1–10 (2019).
23. Y. Sun, G. Livan, A. Ma, V. Latora, Interdisciplinary researchers attain better long-term funding performance. *Commun. Phys.* **4**, 263 (2021).
24. E. Garfield, Citation indexes for science: A new dimension in documentation through association of ideas. *Science* **122**, 108–111 (1955).
25. J. K. Vanclay, Impact factor: Outdated artefact or stepping-stone to journal certification? *Scientometrics* **92**, 211–238 (2012).
26. S. Fortunato *et al*., Science of science. *Science* **359**, eaao0185 (2018).
27. S. N. Durlauf, A. Kourtellos, C. M. Tan, The great Gatsby curve. *Annu. Rev. Econ.* **14**, 571–605 (2022).
28. R. Sinatra, D. Wang, P. Deville, C. Song, A. L. Barabási, Quantifying the evolution of individual scientific impact. *Science* **354**, aaf5239 (2016).
29. D. Moher *et al*., Assessing scientists for hiring, promotion, and tenure. *PLoS Biol.* **16**, e2004089 (2018).
30. A. Baccini, G. De Nicolao, E. Petrovich, Citation gaming induced by bibliometric evaluation: A country-level comparative analysis. *PLoS One* **14**, e0221212 (2019).
31. S. Kelty *et al*., Don't follow the leader: Independent thinkers create scientific innovation. arXiv [Preprint] (2023). http://arxiv.org/abs/2301.02396 (Accessed 1 August 2023).
32. A. Yadav, J. McHale, S. O'Neill, How does co-authoring with a star affect scientists' productivity? Evidence from small open economies *Res. Policy* **52**, 104660 (2023).
33. V. Pareto, *Oeuvres complètes* (Droz, Geneva, 1964–1989).
34. M. Bardoscia, G. De Luca, G. Livan, M. Marsili, C. J. Tessone, The social climbing game. *J. Stat. Phys.* **151**, 440–457 (2013).
35. E. Caron, N. J. van Eck, "Large scale author name disambiguation using rule-based scoring and clustering" in *Proceedings of the 19th International Conference on Science and Technology Indicators*, E. Noyons, Eds. (CWTS-Leiden University, Leiden, 2014), pp. 79–86.
36. A. Tekles, L. Bornmann, Author name disambiguation of bibliometric data: A comparison of several unsupervised approaches. *Quant. Sci. Stud.* **1**, 1510–1528 (2020).