Scientific elite revisited:
Patterns of productivity, collaboration, authorship and impact

Jichao Li[1,2,3,4], Yian Yin[2,3,5], Santo Fortunato[6,7] and Dashun Wang[2,3,4,5*]

[1] *College of Systems Engineering, National University of Defense Technology, Changsha, China*

[2] *Center for Science of Science and Innovation, Northwestern University, Evanston, IL, USA.*

[3] *Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL, USA.*

[4] *Kellogg School of Management, Northwestern University, Evanston, IL, USA*

[5] *McCormick School of Engineering, Northwestern University, Evanston, IL, USA*

[6] *School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, USA*

[7] *Indiana University Network Science Institute (IUNI), Indiana University, Bloomington, IN, USA*

* e-mail: dashun.wang@northwestern.edu

ABSTRACT

Throughout history, a relatively small number of individuals have made a profound and lasting impact on science and society. Despite long-standing, multi-disciplinary interests in understanding careers of elite scientists, there have been limited attempts for a quantitative, career-level analysis. Here, we leverage a comprehensive dataset we assembled, allowing us to trace the entire career histories of nearly all Nobel laureates in physics, chemistry, and physiology or medicine over the past century. We find that, although Nobel laureates were energetic producers from the outset, producing works that garner unusually high impact, their careers before winning the prize follow relatively similar patterns as ordinary scientists, being characterized by hot streaks and increasing reliance on collaborations. We also uncovered notable variations along their careers, often associated with the Nobel prize, including shifting coauthorship structure in the prize-winning work, and a significant but temporary dip in the impact of work they produce after winning the Nobel. Together, these results document quantitative patterns governing the careers of scientific elites, offering an empirical basis for a deeper understanding of the hallmarks of exceptional careers in science.

Introduction

According to Harriet Zuckerman[1], scientific elites "are worthy of our attention not merely because they have prestige and influence in science, but because their collective contributions have made a difference in the advance of scientific knowledge". Indeed, across the broad spectrum of sciences, scientific elites are often pathbreakers and pacesetters in the science of their time[2-7]. Understanding patterns governing the careers of scientific elites helps us uncover insightful markers for exceptional scientific careers, useful for scientists and decision makers who hope to identify and develop individual careers and institutions[8].

The Nobel Prize, widely regarded as the most prestigious award in science, offers a unique opportunity to systematically identify and trace many of the world's greatest scientists[1,3,8-15]. These scientific elites have attracted interest from a wide range of disciplines[1,3,8,11,12,15-27], spanning sociology, economics, psychology, and physics. On the one hand, quantitative studies analyzing publication and citation records have mainly focused on the prize-winning work alone, helping uncover a set of highly reproducible patterns ranging from understanding the link between age and creativity[3,16,17,28-30], to allocating credits and recognition[4,15,19,21]. On the other hand, Zuckerman's canonical work[1] probes into the *entire* career histories of Nobel laureates through qualitative methods[13,14,16,31-35]. The rich patterns articulated by Zuckerman vividly highlight the need to go beyond their prize-winning works, and put them in the context of the entire careers of laureates. Together, the two strands of research call for a quantitative, career-level analysis relying on large-scale datasets to study patterns of productivity, collaboration, authorship, and impact governing the careers of scientific elites.

Despite recent surge of interest in the science of science[3,19,28,29,36-43] and efforts in constructing large-scale datasets of scholarly activities[3,44-46], large-scale studies of the career histories of Nobel laureates remained limited, largely owing to the difficulty in collecting systematic data for their scientific contributions. Here, by combining information collected from the Nobel Prize official websites, laureates' university websites, Wikipedia entries, publication and citation records from the Microsoft Academic Graph (MAG)[47], and extensive manual curations, we constructed a unique dataset capturing career histories of nearly all Nobel laureates in physics, chemistry, and physiology or medicine from 1900 to 2016 (545 out of 590, 92.4%)[48]. We cross-validated this dataset with four different approaches to ensure the reliability of our results. We deposited the derived dataset in a public data repository[49], and describe our data collection and validation procedures in a data descriptor with great detail[48].

We further constructed a comparison dataset of scientific careers using data from the Web of Science (WOS) and Google Scholar (GS)[46], representing the kinds of "ordinary" careers that tend to be studied in the science of science literature[29,50]. For each laureate who published the first paper after 1960, we randomly selected 20 scientists in the same discipline who started their careers in the same year (Supplementary Information S1). Note that the goal here is not to create a matching sample of Nobel-caliber scientists, but a comparison group consisting of scientists who are more similar to typical scientists in the field. One advantage of this comparison approach is that, by selecting individuals with long careers and well-maintained GS profiles, it covers scientists with relatively higher visibility and impact than typical scientists, indicating that our comparisons offer a conservative estimate of the difference between Nobel laureates and their contemporary peers.

Results

Early Performance. Widely held is the belief that the great minds do their critical work early in their careers[3,16,17], prompting us to ask if there is any early signal that distinguishes Nobel laureates. Here we focus on the first five years since their first publication and measure their productivity and impact at this early stage of their careers. Consistent with Zuckerman's observation[1], we find that Nobel laureates were energetic producers from the outset, publishing almost twice as many papers as scientists in our comparison group (Fig. 1a). Yet, compared with this productivity difference, more impressive is the gap in impact. Indeed, the future laureates had a more than six-fold increase over the comparison group in terms of the rate of publishing hit papers, defined as the papers with top 1% of rescaled 10-year citations (Eq. (1)) in the same year and field (Supplementary Information S3.1) (Fig. 1b). This difference is not simply driven by the early onset of prize-winning works. Indeed, we repeated our measurements by omitting the careers of laureates who published their prize-winning work in this period, finding that a substantial gap remained (Fig. S1).

To conceptualize the observed difference in productivity and impact, we separated team- and solo-authored papers, finding that both types of work boost early performance, but they do so in different ways: Most of the difference in early productivity is accounted for by team-authored papers, as solo-authored papers show meager productivity difference between the laureates and their comparison group (Fig. 1c), documenting a greater propensity toward collaborations for scientific elites in their early careers[1]. The only exception is physics laureates who published slightly more solo-authored papers than their comparison group (1.73 vs 1.07, Student's t-test, p-value=0.07). Yet interestingly, solo-authored papers in early careers turned out to disproportionally more likely to be prize-winning papers than team-authored ones. Indeed, comparing the fractions of prize-winning papers within solo- and team-authored papers, we find that the former is about twice as high as the latter on average (Chi-squared test, p-value$<10^{-11}$, Fig. 1d).

Career before the prize. Figure 1 documents outstanding early performance of future laureates. This is consistent with the innovation literature, which shows that the most important works tend to occur early in the lifecycle[3,16,51], speaking to the idea that great, young minds disproportionally break through. Yet on the other hand, growing evidence shows that ordinary scientific careers are governed by the random impact rule[28], predicting that the highest impact work occurs randomly within the sequence of works. To reconcile these two schools of thought, we focus on the career of laureates before they were awarded the Nobel and measure the positions of the prize-winning work and highest impact work within the sequence of works one produced. Here the paper impact is measured by rescaled 10-year citation (Methods). We find both types of works tend to occur early within the sequence of papers (Fig. 2a), a result that contradicts the random impact rule governing typical scientific careers[28,46]. Yet, our earlier analysis suggests that a selection effect may offer a potential explanation for this observation[52]—since the Nobel Prize in science has never been awarded posthumously, those who produced ground-breaking works early were more likely to wait long enough to be recognized[20,22]. Indeed, we removed prize-winning papers and calculate among the remaining ones the position of the highest impact papers. We find that the timing of each of the three remaining highest impact works for Nobel laureates all follow clearly uniform patterns[52] (Fig. 2b). This means, apart from the prize-winning work, all other important works in Nobel careers closely follow the random impact rule: They could be, with equal likelihood, the very first work, the last, or any one in between. This observation is in line with the recent discovery of hot streaks that occur at random within individual careers[46], and therefore raises an important next question: Are these high-impact works clustered together in time?

To answer this question, we quantify the relative timing between the two most cited papers (N^* and N^{**}) within each career by calculating the joint probability $P(N^*, N^{**})$ with a null model in which the two papers each follow their independent temporal patterns. We uncovered clear diagonal patterns across all three domains (Figs. 2c-e), showing that high-impact papers are more likely to cluster together than expected by chance. The diagonal pattern disappeared when we shuffle the order of the works, while preserving the random impact rule (Figs. 2f-h). We also measured the distribution of the longest streak within a career L, finding that $P(L)$ follows a broader distribution compared with that in shuffled careers across all three disciplines (Figs. 2i-k) (Supplementary Information S4.3-4.5). We further find that their hot streaks occur randomly within the sequence of works (Fig. 2l), and are not associated with any detectable change in the overall productivity (Fig. 2m, Kolmogorov-Smirnov test, *p*-value=0.18). Together, these results demonstrate a remarkable resemblance between the career histories of Nobel laureates and ordinary scientists[46].

What seems to distinguish the Nobel laureates from ordinary scientists, however, is that they are disproportionately more likely to have more than one hot streak. Indeed, while hot streak is usually unique for typical scientists[46], Nobel laureates are characterized by 1.93 hot streaks on average (Fig. 2n). Furthermore, their hot streaks also tend to sustain for longer. We measured the duration distribution of hot streaks for Nobel laureates, finding that it peaks around 5.2 years (Fig. 2o), compared with 3.7 years for typical scientists[46]. The longer duration of laureates' hot streaks is also captured by its proportion over career length (Fig. 2p). We also find that prize-winning works are disproportionately more likely to be produced during hot streaks (Fig. 2q). Overall the vast majority of all Nobel winning works (88%) occurred within hot streaks.

Collaboration patterns. One of the most fundamental shifts in science over the past century is the flourishing of large teams across all areas of science[29,39,53,54]. Compared with the overall rate of this shift, Nobel laureates' papers are produced by an even higher proportion of large teams

(Fig. 3a). One possible factor that may explain this team-size difference is impact, as larger teams tend to produce papers with higher impacts[37]. To control for this factor, we created a matching sample for each paper published by the laureates by selecting 20 papers from the same field and year but with the most similar number of citations. We find that, after controlling for impact, the Nobel laureates' papers are still more likely to be produced by larger teams in all times across the last century (Fig. 3b).

Figure 3ab thus underscore another similarity between Nobel and ordinary careers, highlighting the increasing reliance of team work across all types of scientific careers. Yet the ubiquitous increase in team size can be in tension with the fact that the Nobel Prize can only be awarded to at most three recipients for each subject every year[1], prompting us to compare the team size of all prize-winning papers with those published immediately before and after them by the same laureates[52] (Supplementary Information S5.1). We find a greater propensity for the prize-winning papers to be written by less than three authors[52] (61.43% vs 53.28%, Chi-squared test, $p\text{-value}<10^{-4}$, Fig. 3c). We further examine the authorship structure of the prize-winning papers, finding that they are substantially more likely to have the laureates as the first author than other joint papers published by them (45.04% vs 30.64%, Chi-squared test, $p\text{-value}<10^{-7}$) (Fig. 3d). We also calculated the probability of being the last author, finding no statistical difference (Chi-squared test, p-value=0.41).

To test if these phenomena are unique to the prize-winning work, we removed the prize-winning papers and repeated the same analysis for the most cited paper among the remaining papers. We find that there is no statistical difference in their likelihood of being written by small teams[52] (60.18% vs 56.17%, Chi-squared test, p-value=0.1193, Fig. 3e). While the difference in the likelihood of being the first author still exists for chemistry laureates, there is no statistical difference for laureates in physics or medicine (Fig. 3f). Together, these results show that prize-winning papers are more likely to be authored by fewer than three authors, with an intriguing tendency for laureates to claim the first authorship in the prize-winning works. While these observations are consistent with the finding that works produced by small teams tend to disrupt science and technology[37], they are also consistent with Zuckerman's argument that "the future laureates were especially concerned to have the record clear for their most significant work, and particularly in their prize-winning research papers"[1].

After the prize. How does winning the Nobel impact one's subsequent career? The Matthew effect[4,55] tells us that winning begets more winnings. Hence one may expect that works produced after the Nobel garner more impact than those produced before, given their substantially elevated reputation and visibility[15]. Here we find that, to the contrary, when comparing the average impact of papers (defined in Eq. (4)) published by the laureates in each of the four years before and after winning the Nobel, the average impact per paper shows a significant *drop* in the two years following the Nobel. The effect is most significant in the year immediately after, where impact dropped by 11.1% on average compared with the year before. Furthermore, the effect is not permanent, with impact quickly bouncing back by year four to a similar level as the year of the Nobel (Fig. 4a). The "Nobel dip" is most pronounced for physics laureates, as the impacts of their papers were reduced by 18.1%, compared with 4.8% for chemistry and 13.4% for medicine (Supplementary Information S6.2, Fig. S17). Interestingly, in contrast to the common perception of decreased productivity following the Nobel[1,21], possibly due to "the disruptive consequences of abrupt upward social mobility"[1], we find that the average number of papers by the laureates shows no significant change (Fig. 4b), indicating that the uncovered Nobel dip mainly pertains to impact rather than productivity. Note that winning the Nobel may introduce citation boosts to prior papers by the laureate[15,26]. To understand if the observed dip in impact may be explained by this factor, we alter the observation window to exclude post-Prize citations to pre-Prize works, finding that the "dip and bounce back" pattern remains robust (Fig. 4c, Supplementary Information S6.4, Fig. S20). We also find that the number of solo-authored papers decreased precipitately after the Nobel (Student's t-test, p-value=0.004, Fig. 4d), whereas the fraction of team-authored papers increased (Student's t-test, p-value=0.008, Fig. 4e), suggesting that collaboration and teamwork carry an increasing importance for the laureates after winning the Nobel.

The Nobel dip signals that scientific community's attention is not driven by status but the quality of work. To unearth potential mechanisms underlying the "dip and bounce back" dynamics, we trace topic changes before and after Nobel as reflected in their publications. We use an established method[43] that detects research topics based on communities in the co-citing network of papers published by a scientist, offering a discipline-independent method to identify and trace research topics across a career (Supplementary Information S6.5). As an illustrative example, Fig. 4f shows the constructed co-citing network and topic communities for the career of Jean-Marie Lehn, who was awarded the 1987 chemistry Nobel together with Donald Cram and Charles Pedersen for the synthesis of cryptands. In his remarkable career, Lehn published more than 700 papers. Figure 4f visualizes his publication history by topic, showing that his research agenda was almost exclusively focused on cryptands related research, until he was awarded the Nobel in 1987. Yet, just as this line of research was officially recognized, we observed a clear shift in topic right after winning the Nobel (Fig. 4g). In the next 10 years, his research was primary focused on self-assembly and self-organization. Most interestingly, this is a topic that he had never published on before winning the Nobel.

The intriguing example of Lehn's career prompts us to ask if laureates disproportionately shift research topic after winning the Nobel. We randomly selected two papers, within 4 years before and after the Nobel, respectively, and measured the probability of two papers belonging to the same topic, finding only 36.8% of the two papers cover the same topic before and after winning the Prize. We then build a null model by randomly choosing a year as the pretended prize-winning year for comparison, finding that the probability is significant higher (45.2% vs 36.8%, *p*-value=0.004, Fig. 4h), which suggests the laureates have a higher likelihood of shifting research topics after winning the Nobel. We further measured the likelihood of laureates studying a new topic after winning Prize, and compare it with a null model where we shuffled the topic of the works, finding that the laureates are much more likely to study a new topic after winning Prize than expected (14.2% vs 1.8%, p-value$<10^{-14}$, Fig. 4i). To ensure that these results are not affected by specific community detection methods used to detect topics, we repeated our analyses with another well-known algorithm (Infomap[56]), obtaining the same conclusions (Supplementary Information S6.6).

To understand potential forces behind the uncovered change in research agenda following the Nobel, we examined several different factors, including the popularity of research topics before and after the Prize (Supplementary Information S6.7), changes in collaborators (Supplementary Information S6.8), and funding opportunities (Supplementary Information S6.9). We find that the topic studied after the Nobel tends to be less popular at the time. The number of new collaborators does not increase after the Nobel, but these collaborators tend to more established in terms of productivity and impact. And somewhat surprisingly, the overall funding to each laureate remains mostly constant around the time of the award. Although none of these factors can directly explain the observed topic change and the associated citation dip (Figs. 4j-m, Supplementary Information S6.7-S6.9, Figs. S24-S26), they appear consistent with an endogenous shift in the laureate's interest to explore new directions. Note that, although the uncovered dip-bounce-back dynamics and topic shifting behavior both occur around the same time (when awarded the Nobel), it does not imply that the two are causally related. On the other hand, while one may be better at anticipating which work will be recognized by the Nobel eventually[57], it remains difficult to precisely predict the year of winning, indicating that the award year can be viewed as a largely exogenous variation in a career[58], which then coincides with topic shifting behavior that is largely endogenous to the individual. Regardless, these results highlight the unwavering scientific efforts by the laureates, actively pursuing new lines of enquiry while undeterred by the extra burdens imposed by growing duties and responsibilities[1].

Discussion

In summary, building on Zuckerman's canonical work on scientific elites[1], here we present a systematic empirical investigation of the careers of Nobel laureates by studying patterns of productivity, collaboration, authorship, and impact. This analysis is now possible thanks to a novel dataset we curated—both algorithmically and manually—which links several disparate biographical and bibliographical data sources, offering a unique opportunity to quantitatively study the scientific contributions and recognitions of scientific elites. Despite the clear difference between the Nobel laureates and "ordinary" scientists, we find universal career patterns that are applicable to both ordinary and elite scientists. Indeed, we find the careers of the laureates before winning the prize are governed by remarkably similar patterns as ordinary scientists, characterized by hot streaks and increasing reliance on team work. Hence these results help advance the canonical innovation literature by offering new empirical evidence from large-scale datasets. At the same time, we also uncovered notable but previously unknown variations along their careers associated with the Nobel prize, including shifting coauthorship structure in the prize-winning work, and a temporary but significant dip in the impact of work they produce after winning the

Nobel. Overall, these results represent new empirical patterns that further enrich our understanding of careers of the scientific elite.

Together, this paper takes an initial but crucial step probing our quantitative understanding of career patterns of scientific elite, which not only offer an empirical basis for future studies of individual careers and creativity in broader domains[16,51], but also deepen our quantitative understanding of patterns governing exceptional careers in science.

Methods

Rescaled number of citations. To approximate the scientific impact of each paper, we calculate the number of citations the paper received after 10 years, C_{10}, and use it as a proxy for the paper's impact. Previous studies[29,37,44] have shown that the average number of citations per paper changes over time. To be able to compare the impact of papers published at different times and to adjust for temporal effects, the rescaled number of citations a paper receives after 10 years, $\hat{C}_{10,i}$, is suggested as a good proxy for publication impact. According to the reference[29], given a paper i, $\hat{C}_{10,i}$, is defined as follows:

$$\hat{C}_{10,i} = 10 \cdot \frac{C_{10,i}}{\langle C_{10} \rangle}, \quad (1)$$

where $C_{10,i}$ is the raw number of 10-year citations for paper i, and $\langle C_{10} \rangle$ is the average C_{10} calculated over all publications published in the same year and field.

Definition of hit paper rate. In Fig. 1b, we compare the "hit" paper rate—defined as the probability of publishing papers in the top 1% of rescaled 10-year citations in the same year and field—for Nobel laureates and typical authors. Our collected Nobel laureate dataset is based on information provided by the MAG, which assigns the field of subject for each paper. It is worth noting that the field of subject is a hierarchal structure with six levels. The first level contains 19 main fields, such as "physics," "chemistry," "medicine," and "biology." The second level contains 295 subfields, such as "astrophysics," "biophysics," and "geophysics." In this paper, we choose the second level fields in calculating the hit paper rate for Nobel laureates. The GS typical scientist dataset is based on information from the WOS, and it is almost impossible to precisely match the career histories of 3540 GS scientists from the WOS to the MAG. Thus, the hit rate analysis of the GS scientists is based on the WOS database itself. Papers in the WOS are also assigned to one of 234 specific field categories, such as "astronomy & astrophysics," "biophysics," and "geochemistry & geophysics." The hit paper rate for typical scientists is calculated using these 234 specific fields from the WOS.

Selecting matching papers. In Fig. 3b, we created a matching sample for each paper published by Nobel laureates. The procedure for selecting matching papers is introduced here in detail. For each Nobel prize-winner's work, we first determine its year of publication, total citation number, and subject categories based on the MAG dataset. Next, all the MAG papers with the same publishing year and specific field are obtained and sorted according to their number of citations. It is worth noting that when a laureate's paper spans multiple subjects, we deem MAG papers appropriate matches if they share at least one common subject with the laureate's. We then select the 20 papers with citation counts that are most similar to the laureate's paper and use these as matching papers.

Quantifying impact. In Fig. 4a, we compare the average impact of papers published by the laureates in each of the four years before and after winning the Nobel. We propose a measure to quantify average impact of papers: We first calculate the average impact within all papers in specific years, and then we take the individual heterogeneity of Nobel laureates into consideration when quantifying average impact of papers.

The impact of paper i, is quantified by $\Gamma_{\mathrm{i}} = \log(\hat{C}_{10,\mathrm{i}} + 1)$ where $\hat{C}_{10,\mathrm{i}}$ measures the rescaled number of citations within 10 years of publication. We denote $\Delta \mathrm{y} = \mathrm{y_i} - \mathrm{y_{Prize}}$ as a laureate's relatively publishing time after winning the Nobel Prize, where $\mathrm{y_i}$ is the publication year of paper i. Assuming there are $\mathrm{N}_{\Delta \mathrm{y}}$ papers publishing in the $\Delta \mathrm{y}$ year after winning the Prize, we define the average impact of papers as follows:

$$\langle \Gamma_{\mathrm{P}} \rangle_{\Delta \mathrm{y}} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}_{\Delta \mathrm{y}}} \Gamma_{\mathrm{i}}}{\mathrm{N}_{\Delta \mathrm{y}}}. \tag{2}$$

However, the above measure did not consider the individual heterogeneity of Nobel laureates. For example, average impact may be driven by those laureates with high productivity as well as high paper quality. Thus, we first measure the average impact of papers for each laureate and then calculate the average for all Nobel laureates. For laureate j, the average impact of papers published in the $\Delta \mathrm{y}$ year after winning the Prize is defined as:

$$\langle \Gamma_{\mathrm{P}} \rangle_{\Delta \mathrm{y},j} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}_{\Delta \mathrm{y},j}} \Gamma_{\mathrm{i}}}{\mathrm{N}_{\Delta \mathrm{y,j}}}, \tag{3}$$

where $\mathrm{N}_{\Delta \mathrm{y},j}$ is the number of papers published in the $\Delta \mathrm{y}$ year of laureate j. Factoring in individual heterogeneity, the average impact of papers is defined as follows:

$$\langle \Gamma_{\mathrm{N}} \rangle_{\Delta \mathrm{y}} = \frac{\sum_{\mathrm{j}=1}^{\mathrm{M}_{\Delta \mathrm{y}}} \langle \Gamma_{\mathrm{P}} \rangle_{\Delta \mathrm{y,j}}}{\mathrm{M}_{\Delta \mathrm{y}}}, \tag{4}$$

where $\mathrm{M}_{\Delta \mathrm{y}}$ denotes the number of laureates who still publish papers in the $\Delta \mathrm{y}$ year after winning the Nobel Prize. In the main text (Fig. 4a), we use $\langle \Gamma_{\mathrm{N}} \rangle$ to measure the average impact of papers.

Topic changing after winning the Nobel Prize. To quantify the topic of a paper, we adopt a recent method based on community structure of the co-citing network of a scientist's papers[51]. To ensure meaningful community detection results, we consider all Nobel laureates who have published at least 50 papers. We also excluded Nobel laureates who published fewer than five papers after winning the Prize. Finally, we selected 283 Nobel laureates (74 for physics, 96 for Chemistry, 113 for medicine) who satisfied these requirements.

In Fig. 4g, we measure the probability of two papers belonging to the same topic within 4 years before and after the reception of the Prize and a random year. To measure the probability of changing topics of Nobel laureates after winning Prize, we randomly selected two papers, within 4 years before and after the Nobel, respectively, and measured the probability of two papers belonging to the same topic. We then build a null model by randomly choosing a year as the pretended prize-winning year for comparison.

To test if Nobel laureates tend to study a new topic after winning Prize, we measure the chance of Nobel laureates shifting to a new topic after winning the Nobel, $\frac{\#\text{new topics after winning Prize}}{\#\text{topics}}$. We also shuffled the topic of the works and repeated the measurement as a null model for comparison.

Author Contributions D.W. and S.F. conceived the project, D. W. designed the experiments; J.L. and Y.Y. collected data and performed empirical analyses with help from S.F. and D.W.; all authors discussed and interpreted results; D.W., J.L. and Y.Y. wrote the manuscript; all authors edited the manuscript.

Acknowledgements The authors thank L. Liu, Y. Wang, Y. Ma, and all members of Northwestern Institute on Complex Systems (NICO) for invaluable comments. Funding data sourced from Dimensions, an inter-linked research information system provided by Digital Science (https://www.dimensions.ai).

Data Accessibility The main data that support the findings of this study is freely available. Deposited in public repositories with detailed descriptions in Harvard Dataverse (https://doi.org/10.7910/DVN/6NJ5RN).

Funding Statement This work is supported by the Air Force Office of Scientific Research under award number FA9550-15-1-0162, FA9550-17-1-0089 and FA9550-19-1-0354, National Science Foundation grant SBE 1829344 and Northwestern University's Data Science Initiative.

Reference

1 Zuckerman, H. *Scientific Elite: Nobel Laureates in the United States*. (Free Press, 1977).

2 Barabasi, A. L., Song, C. M. & Wang, D. S. Handful of papers dominates citation. *Nature* **491**, 40-40 (2012).

3 Jones, B. F. & Weinberg, B. A. Age dynamics in scientific creativity. *P Natl Acad Sci USA* **108**, 18910-18914 (2011).

4 Merton, R. K. The Matthew effect in science. *Science* **159**, 56-63, https://doi.org/10.1126/science.159.3810.56 (1968).

5 Newman, M. E. J. The structure of scientific collaboration networks. *P Natl Acad Sci USA* **98**, 404-409, https://doi.org/10.1073/pnas.021544898 (2001).

6 Price, D. J. D. General Theory of Bibliometric and Other Cumulative Advantage Processes. *J Am Soc Inform Sci* **27**, 292-306, https://doi.org/10.1002/asi.4630270505 (1976).

7 Petersen, A. M., Jung, W. S., Yang, J. S. & Stanley, H. E. Quantitative and empirical demonstration of the Matthew effect in a study of career longevity. *P Natl Acad Sci USA* **108**, 18-23, https://doi.org/10.1073/pnas.1016733108 (2011).

8 Zuckerman, H. The sociology of the Nobel prizes. *Sci Am* **217**, 25-+, https://doi.org/10.1038/scientificamerican1167-25 (1967).

9 Zuckerman, H. Interviewing an Ultra-Elite. *Public Opin Quart* **36**, 159-+, Doi 10.1086/267989 (1972).

10 Moulin, L. Sociology of Nobel-Prize Winners for Science, 1901-1960, with Special-Reference to Nationality. *Cah Int Sociol* **31**, 145-163 (1961).

11 Zuckerman, H. Nobel laureates in science: Patterns of productivity, collaboration, and authorship. *Am Sociol Rev* **32**, 391-403, https://doi.org/10.2307/2091086 (1967).

12 Hansson, N., Halling, T. & Fangerau, H. Nobel nomination letters point to a winning formula. *Nature* **555**, 311, https://doi.org/10.1038/d41586-018-03057-z (2018).

13 Garfield, E. Do Nobel-Prize Winners Write Citation-Classics. *Curr Contents*, 3-8 (1986).

14 Garfield, E. & Welljams-Dorof, A. Of Nobel class: a citation perspective on high impact research authors. *Theor Med* **13**, 117-135 (1992).

15 Mazloumian, A., Eom, Y. H., Helbing, D., Lozano, S. & Fortunato, S. How citation boosts promote scientific paradigm shifts and nobel prizes. *Plos One* **6**, https://doi.org/10.1371/journal.pone.0018975 (2011).

16 Simonton, D. K. *Genius, Creativity, and Leadership*. (Harvard University Press, 1984).

17 Simonton, D. K. Creative productivity: A predictive and explanatory model of career trajectories and landmarks. *Psychol Rev* **104**, 66-89, https://doi.org/10.1037/0033-295x.104.1.66 (1997).

18 Moreira, J. A. G., Zeng, X. H. T. & Amaral, L. A. N. The distribution of the asymptotic number of citations to sets of publications by a researcher or from an academic department are consistent with a discrete lognormal model. *Plos One* **10**, https://doi.org/10.1371/journal.pone.0143108 (2015).

19 Shen, H. W. & Barabasi, A. L. Collective credit allocation in science. *P Natl Acad Sci USA* **111**, 12325-12330, https://doi.org/10.1073/pnas.1401992111 (2014).

20 Fortunato, S. Growing time lag threatens Nobels. *Nature* **508**, 186-186 (2014).

21 Nobel reactions. *Nat Phys* **13**, 921-921, https://doi.org/10.1038/nphys4296 (2017).

22 Chan, H. F. & Torgler, B. Time-lapsed awards for excellence. *Nature* **500**, 29-29, https://doi.org/10.1038/500029c (2013).

23 Seeman, J. I. Synthesis and the Nobel prize in chemistry. *Nat Chem* **9**, 925-929, https://doi.org/10.1038/nchem.2864 (2017).

24 Fleming, L., Mingo, S. & Chen, D. Collaborative brokerage, generative creativity, and creative success. *Admin Sci Quart* **52**, 443-475, https://doi.org/10.2189/asqu.52.3.443 (2007).

25 Singh, J. & Fleming, L. Lone inventors as sources of breakthroughs: Myth or reality? *Manage Sci* **56**, 41-56, https://doi.org/10.1287/mnsc.1090.1072 (2010).
26 Azoulay, P., Stuart, T. & Wang, Y. B. Matthew: Effect or fable? *Manage Sci* **60**, 92-109, https://doi.org/10.1287/mnsc.2013.1755 (2014).
27 Weinberg, B. A. & Galenson, D. W. Creative careers: The life cycles of Nobel laureates in economics. *De Economist* **pp**, 1-18 (2019).
28 Sinatra, R., Wang, D., Deville, P., Song, C. M. & Barabasi, A. L. Quantifying the evolution of individual scientific impact. *Science* **354** (2016).
29 Fortunato, S. *et al.* Science of science. *Science* **359**, eaao0185, https://doi.org/10.1126/science.aao0185 (2018).
30 Jones, B. F., Kim, N. J. D. & Miranda, M. J. Age and High-Growth Entrepreneurship. (2018).
31 Runco, M. A. *Creativity: Theories and themes: Research, development, and practice*. (Elsevier, 2014).
32 Taleb, N. N. *The black swan: The impact of the highly improbable*. Vol. 2 (Random house, 2007).
33 Latour, B. & Woolgar, S. *Laboratory life: The construction of scientific facts*. (Princeton University Press, 2013).
34 Taubes, G. *Nobel dreams: Power, deceit and the ultimate experiment*. (Microsoft Press, 1988).
35 Cole, S. & Cole, J. R. Scientific output and recognition: A study in the operation of the reward system in science. *Am Sociol Rev*, 377-390 (1967).
36 Zeng, A. *et al.* The science of science: From the perspective of complex systems. *Phys Rep* **714**, 1-73, https://doi.org/10.1016/j.physrep.2017.10.001 (2017).
37 Wu, L., Wang, D. & Evans, J. A. Large teams develop and small teams disrupt science and technology. *Nature* **566**, 378–382 (2019).
38 Clauset, A., Arbesman, S. & Larremore, D. B. Systematic inequality and hierarchy in faculty hiring networks. *Sci Adv* **1** (2015).
39 Milojevic, S. Principles of scientific research team formation and evolution. *P Natl Acad Sci USA* **111**, 3984-3989, https://doi.org/10.1073/pnas.1309723111 (2014).
40 Petersen, A. M. Multiscale impact of researcher mobility. *J R Soc Interface* **15**, 20180580 (2018).
41 Pfeiffer, T., Tran, L., Krumme, C. & Rand, D. G. The value of reputation. *J R Soc Interface* **9**, 2791-2797 (2012).
42 Youn, H., Strumsky, D., Bettencourt, L. M. & Lobo, J. Invention as a combinatorial process: evidence from US patents. *J R Soc Interface* **12**, 20150272 (2015).
43 Zeng, A. *et al.* Increasing trend of scientists to switch between topics. *Nature Communications* **10**, 3439 (2019).
44 Ioannidis, J. P., Baas, J., Klavans, R. & Boyack, K. W. A standardized citation metrics author database annotated for scientific field. *PLoS biology* **17**, e3000384 (2019).
45 Vuong, Q. H. *et al.* An open database of productivity in Vietnam's social sciences and humanities for public use. *Sci Data* **5**, https://doi.org/10.1038/sdata.2018.188 (2018).
46 Liu, L. *et al.* Hot streaks in artistic, cultural, and scientific careers. *Nature* **559**, 396–399 (2018).
47 https://www.microsoft.com/en-us/research/project/microsoft-academic-graph/.
48 Li, J., Yin, Y., Fortunato, S. & Wang, D. A dataset of publication records for Nobel laureates. *Sci Data* **6**, 33 (2019).
49 Li, J., Yin, Y., Fortunato, S. & Wang, D. A dataset of publication records for Nobel laureates. *Harvard Dataverse*, https://doi.org/10.7910/DVN/6NJ5RN (2018).
50 Azoulay, P. *et al.* Toward a more scientific science. *Science* **361**, 1194-1197, https://doi.org/10.1126/science.aav2484 (2018).
51 Simonton, D. K. *Scientific genius: A psychology of science*. (Cambridge University Press, 1988).
52 Li, J., Yin, Y., Fortunato, S. & Wang, D. Nobel laureates are almost the same as us. *Nature Reviews Physics* **1**, 301–303 (2019).
53 Wuchty, S., Jones, B. F. & Uzzi, B. The increasing dominance of teams in production of knowledge. *Science* **316**, 1036-1039 (2007).

54 Petersen, A. M. Quantifying the impact of weak, strong, and super ties in scientific careers. *P Natl Acad Sci USA* **112**, E4671-E4680, https://doi.org/10.1073/pnas.1501444112 (2015).
55 Perc, M. The Matthew effect in empirical data. *J R Soc Interface* **11**, 20140378 (2014).
56 Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences* **105**, 1118-1123 (2008).
57 https://en.wikipedia.org/wiki/Clarivate_Citation_Laureates.
58 Adams, D. *Mostly Harmless*. (Del Rey, 1993).
59 Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* **2008**, P10008 (2008).