Check for updates

The MIT Press

RESEARCH ARTICLE

Consistency pays off in science

Şirag Erkol[1] iD, **Satyaki Sikdar[1]** iD, **Filippo Radicchi[1]** iD, and **Santo Fortunato[1,2]** iD

[1]Center for Complex Networks and Systems Research, Luddy School of Informatics, Computing, and Engineering, Indiana University Bloomington, Bloomington, IN, USA
[2]Indiana University Network Science Institute (IUNI), Indiana University Bloomington, Bloomington, IN, USA

ABSTRACT

The exponentially growing number of scientific papers stimulates a discussion on the interplay between quantity and quality in science. In particular, one may wonder which publication strategy may offer more chances of success: publishing lots of papers, producing a few hit papers, or something in between. Here we tackle this question by studying the scientific portfolios of Nobel Prize laureates. A comparative analysis of different citation-based indicators of individual impact suggests that the best path to success may rely on consistently producing high-quality work. Such a pattern is especially rewarded by a new metric, the *E*-index, which identifies excellence better than state-of-the-art measures.

1. INTRODUCTION

The number of scientific papers has been growing exponentially for over a century (Dong, Ma et al., 2017; Fortunato, Bergstrom et al., 2018). The number of papers per author has been relatively stable for a long time, but it has been increasing over the past decades (Dong et al., 2017), favored by the growing tendency of scientists to work in teams (Wuchty, Jones, & Uzzi, 2007).

Such increased productivity is incentivized by career evaluation criteria that typically reward large outputs, making scientists less risk averse when choosing research directions (Franzoni & Rossi-Lamastra, 2017). This, however, may come at the expense of the quality of research outcomes (Bornmann & Tekles, 2019; Sunahara, Perc, & Ribeiro, 2021). Indeed, it has been shown that the exponential growth of the number of publications corresponds to a much slower increase in the number of new or disruptive ideas (Chu & Evans, 2021; Milojević, 2015).

However, although scholars should focus on quality, it is unclear whether it is more rewarding to pursue rare hit papers, have a consistent track record of valuable outputs, or be in between these scenarios. Analyzing the careers of arguably the most successful class of scientists, Nobel Prize laureates, may help address this issue. In particular, we would like to check if there is a dominant path to success in the careers of such illustrious scholars.

To that effect, we consider a broad range of evaluation metrics that reward one-hit wonders alongside those that favor a consistent production of high-quality research and investigate their effectiveness in identifying Nobelists from within a more extensive set of similarly productive scientists. We find that the best-performing metrics are indeed the ones that prioritize a consistent stream of high-quality research.

The rest of this article is organized as follows. We first describe the data collection and curation in Section 2. Then, we briefly review some popularly adopted impact metrics and

introduce two new ones. In Section 3, we describe and discuss the two sets of experiments we used to check which of the two competing scenarios is more common. Finally, we give our conclusions in Section 4.

2. METHODS

2.1. Data

We consider three fields in which the Nobel Prize is awarded: Physics, Chemistry, and Physiology or Medicine (abbreviated henceforth as Medicine).

The publication records of scientists are obtained from two sources. For Nobelists, we use the hand-curated data set with explicit annotations for prize-winning papers (Li, Yin et al., 2019). As a baseline, we consider scientists with verified Google Scholar (GS) profiles tagged with Physics, Chemistry, Physiology, or Medicine as of May 2021.

We use the 2017 version of the Web of Science (WoS) database to compile the citation statistics of the articles. We rely on gathering data from different sources on purpose, as WoS and GS complement each other well. GS offers the possibility of obtaining accurate publication records of individual scientists without the need to perform name disambiguation (Radicchi & Castellano, 2013). WoS lets us reconstruct the citation history of individual papers. Both ingredients are necessary for the type of analysis that we perform in this paper.

We adopt a similar methodology to that of Sinatra, Wang et al. (2016) to match papers across databases. Given a paper $\hat{p}$ written by author a in GS, we list the papers P_a in WoS authored by people with the same last name as a. From P_a, we select the paper p with the highest normalized Levenshtein similarity between the corresponding paper titles (Levenshtein, 1966). We consider it a successful match only if the similarity exceeds 90%. Otherwise, we discard $\hat{p}$ from further analysis. Following this procedure, we could match 78.1% of papers by Nobelists and 49.6% of papers by baseline scientists, respectively. For our analysis, we only consider scientists who published their first paper after 1960 and have a portfolio with at least 10 papers. Detailed statistics are provided in Table 1.

2.2. Metrics

Let us consider a portfolio $\mathcal{P} = \{c_1, ..., c_N\}$ of $N = |\mathcal{P}|$ papers that collectively receive C_{tot} citations (i.e., $C_{tot} = \sum_i^N c_i$). We consider the following metrics:

N: total number of papers.

C_{tot}: total number of citations.

C_{avg}: average number of citations (i.e., $C_{avg}(\mathcal{P}) = \frac{C_{tot}}{N}$).

C_{max}: citations received by the most cited paper (i.e., $C_{max}(\mathcal{P}) = \max\{c_1, \cdots, c_N\}$).

H: H-index (i.e., the largest number H of the top-cited papers with at least H citations; Hirsch, 2005).

G: G-index (i.e., the largest number G of the top-cited papers with at least G^2 combined citations; Egghe, 2006).

Table 1. Number of scientists in each category and field

Category	Physics	Chemistry	Medicine
Nobelists	55	51	56
Baseline scientists	4,081	3,330	2,715

Q: Q-index, proposed by Sinatra et al. (2016), $Q(\mathcal{P}) = \exp\left(\frac{1}{\sum_{i=1}^{N}\Theta(c_{10,i})}\sum_{i=1}^{N}\Theta(c_{10,i})\log c_{10,i}\right)$, up to a constant factor, where Θ is the Heaviside function (i.e., $\Theta(x) = 1$ if $x > 0$ and 0 otherwise), and $c_{10,i}$ is the citations gained by paper i within 10 years of publication. We normalize $c_{10,i}$ by dividing it with the average c_{10} of all papers published in the same discipline and year as paper i (Sinatra et al., 2016).

$\widetilde{Q}$: a variant of the unnormalized Q-index, where we use the total number of citations c_i instead of $c_{10,i}$.

We observe that these measures have their unique preferences for ranking portfolios. Some, like $C_{\max}$, appear to reward one-hit wonders, and others, like H, reward consistency. One of the goals of this work is to identify and differentiate Nobelists from baseline scientists. Therefore, we argue that we need a new, simple, yet interpretable metric covering the whole portfolio spectrum.

2.3. Citation Moment and *E*-Index

Given a publication portfolio $\mathcal{P}$, one may consider the following extreme scenarios:

- Citations are equally distributed among the papers, with each paper having C_{tot}/N citations.
- A single paper accounts for all citations.

In the first case, there is a sustained production of work of similar quality, while the second represents a one-hit-wonder situation.

2.3.1. Citation moment

We propose the *citation moment* M_α, a new parametric measure that can reward both scenarios, as well as the ones in between, depending on the value of the parameter α. It is defined as

$$M_\alpha(\mathcal{P}) = \frac{1}{N}\sum_{i=1}^{N}c_i^\alpha, \tag{1}$$

where α is a real positive number. We remark that M_α is essentially an average of the citation scores of the papers, where the weight of each score is modulated by the exponent α. We can make the following observations of the behavior of our metric for different values of α.

$\alpha \to 0$: M_α behaves like $\widetilde{Q}$ as $c^\alpha \approx \log c$, but unlike $\widetilde{Q}$, it accounts for uncited papers.
$0 < \alpha < 1$: M_α is higher for balanced portfolios (i.e., ones with a more uniform distribution of citations).
$\alpha = 1$: M_α becomes identical to C_{avg}.
$\alpha > 1$: M_α is higher for unbalanced portfolios.
$\alpha \to \infty$: M_α closely imitates $C_{\max}$.

2.3.2. E-index

We also propose an additional parameter-free measure that, like M_α, is sensitive to the distribution of citations. We call this metric *E-index*, defined as

$$E(\mathcal{P}) = -\frac{1}{N}\sum_{i=1}^{N}c_i\log\frac{c_i}{C_{\text{tot}}}, \tag{2}$$

Table 2. Values of metrics for portfolios with N papers with C_{tot} citations, of which n are equally cited and $N - n$ are uncited

Metric	Value
H	$\min\{\lfloor C_{tot}/n \rfloor, n\}$
G	$\min\{\lfloor \sqrt{C_{tot}} \rfloor, \lfloor C_{tot}/n \rfloor, N\}$
$\widetilde{Q}$	C_{tot}/n
M_α	$\frac{C_{tot}^\alpha}{N \ n^{\alpha-1}}$
E	$\frac{C_{tot}}{N} \log n$

which reaches its maximum $C_{avg} \log N$ when citations are distributed equally among papers, favoring authors with large average numbers of citations. In fact, $E(\mathcal{P})$ is just the product of the average number of citations C_{avg} and of the Shannon entropy of the citation distribution.

2.4. Behavior of Metrics on Stylized Portfolios

To better understand the behavior of the different metrics in our analysis, we consider a portfolio with n cited papers with C_{tot}/n citations each and $N - n$ uncited papers. In Table 2, we show the values that several key metrics take in this case.

We see that the citation moment M_α (for $\alpha \neq 0, 1$), E-index, and G-index depend on n, N, and C_{tot}. The H-index and the $\widetilde{Q}$ depend only on the cited papers. So, for example, two portfolios with identical values of C_{tot} and n would have the same H-index, regardless of the number of uncited papers. Furthermore, even though the G-index depends on all three parameters, it depends on them in a somehow undesirable way. For example, a portfolio with more uncited papers may have a G-index value greater than or equal to the G-index of another portfolio with identical C_{tot} and n values. Instead, ranking the portfolio with fewer uncited works higher (lower $N - n$), as M_α and E would, seems more intuitive.

3. RESULTS

In Figure 1, we plot Nobelists and baseline scientists according to their number of papers and the total number of citations. As expected, most Nobelists lie in the top right region, indicating high levels of both productivity and impact. However, there appear to be a few Nobelists in the top left, indicating that they only produced a handful of high-impact papers. To further illustrate this difference, we consider two Nobelists in Physics, David J. Gross (2004) and John M. Kosterlitz (2016), and plot their publication timelines in Figure 2. Gross has a consistent production of high-impact works, but Kosterlitz stands out for having a single big paper.
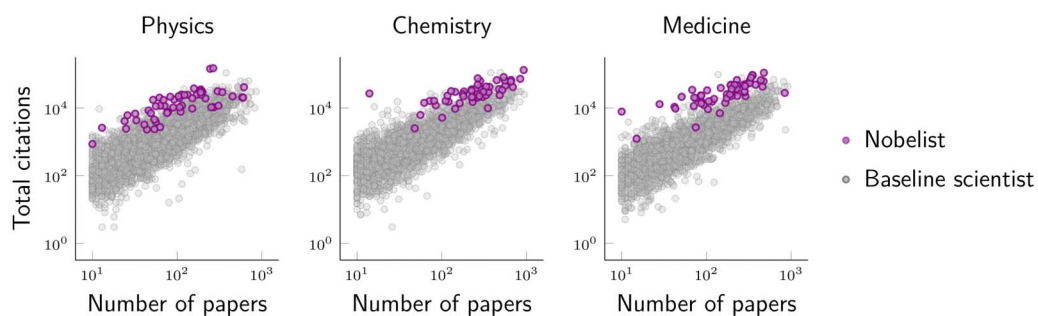

Figure 1. Total number of citations vs. total number of papers for Nobelists (purple dots) and baseline scientists (gray dots).
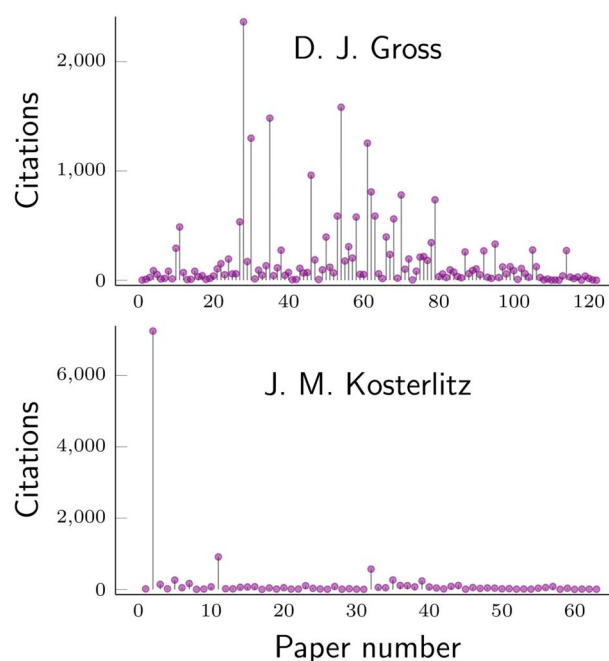
Figure 2. Consistency versus single-hit scenario. On the *x*-axis, we indicate the temporal sequence of papers, and on the *y*-axis the citations accrued by each paper. The two panels show the profiles of D. J. Gross (top) and J. M. Kosterlitz (bottom). The former has a portfolio with multiple highly cited papers, and the latter has one highly cited paper. D. J. Gross: $N = 122$, $C_{tot} = 24{,}144$, $C_{avg} = 197.9$, $E = 768.6$. J. M. Kosterlitz: $N = 63$, $C_{tot} = 11{,}688$, $C_{avg} = 185.5$, $E = 348.8$.

We now focus on two tasks: portfolio classification and future Nobelist identification.

3.1. Portfolio Classification

We test the performance of the metrics in distinguishing the portfolios of Nobelists from those of the baseline scientists. We consider two subtasks which we describe below. We use the area under the precision-recall curve (AUC-PR) in each task as the performance metric. This curve shows the trade-off between precision and recall at different thresholds. Bounded between 0 and 1, higher AUC-PR values indicate better classification performance. For random predictions, AUC-PR is the fraction of positive samples. AUC-PR is better suited for imbalanced data sets than the area under the receiver operating characteristic curve (ROC-AUC) (Saito & Rehmsmeier, 2015). Results for the ROC-AUC are reported in the Supplementary material and are consistent with the analysis done using AUC-PR.

- **Full**. We use the entire portfolio of the scientists described in Section 2.1.
- **Preaward**. We construct the preaward portfolio of Nobelists (i.e., the set of papers published until the year of the prize-winning paper), discarding those with fewer than 10 papers. We find that 15 (27%), 28 (55%), and 22 (39%) of Nobelists in Physics, Chemistry, and Medicine, respectively, satisfy the above criteria.

    Specifically, for a Nobelist who published their first paper in year y_0 and wrote their prize-winning article in year y_p, we consider the papers published and citations accrued between years y_0 and $y_p - 1$. We then pair the Nobelist with 20 baseline scientists who published their first papers around the year y_0 and wrote at least 10 papers in their careers' first $y_p - y_0$ years.
- **Optimal α selection**. Recall that, unlike other measures, M_α has a tunable parameter α. Therefore, for each task, we record the performance of M_α across a range of α values
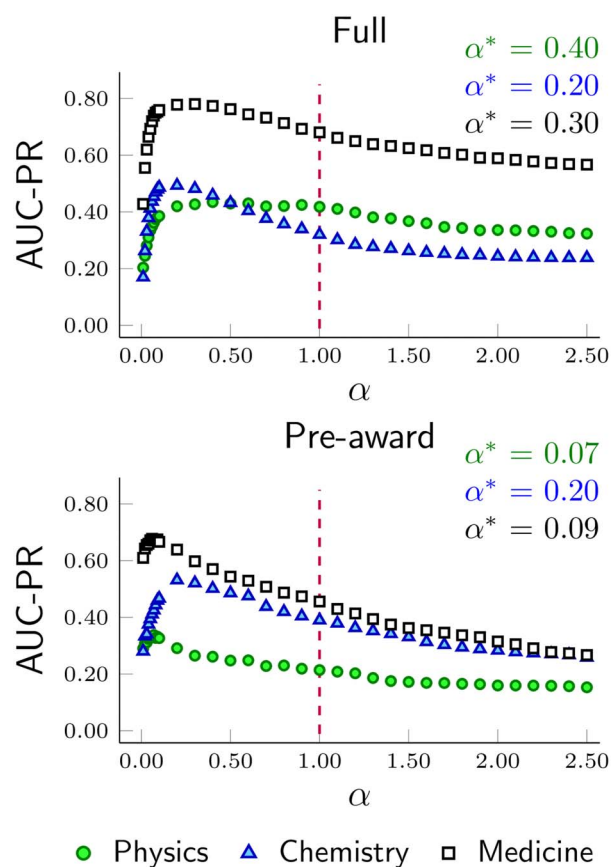
Figure 3. Classification performance of M_α for varying α. Different symbols denote different fields. The dashed line $\alpha = 1$ separates the two regimes. We use the optimal values of α (α^*) in our analyses.

Table 3. AUC-PR values for the Full and Preaward (PA) portfolio classification tasks. The best-performing metrics for each field are marked in bold type. M_α and E are the standout performers. Note that values across columns are not comparable as the baseline values are determined by the respective class imbalance ratios

Metric	Physics		Chemistry		Medicine	
	Full	PA	Full	PA	Full	PA
N	0.03	0.07	0.13	0.12	0.06	0.06
C_{tot}	0.21	0.15	0.43	0.34	0.52	0.24
C_{avg}	0.42	0.19	0.32	0.39	0.68	0.46
C_{max}	0.24	0.12	0.25	0.21	0.49	0.18
H	0.12	0.16	0.44	0.36	0.50	0.24
G	0.15	0.15	0.41	0.33	0.48	0.17
$\widetilde{Q}$	0.30	0.19	0.32	0.41	0.67	0.48
Q	0.08	0.15	0.13	0.20	0.26	0.45
M_α	0.43	**0.34**	0.49	**0.53**	**0.78**	**0.68**
E	**0.44**	0.23	**0.53**	0.45	0.75	0.44

and plot the results in Figure 3. We observe a slight dependence of the optimal α-value (α^*) on the task and the field. We use the corresponding α^* values while comparing the performance of M_α with other metrics. In each case, however, we find $\alpha^* < 1$, which indicates that portfolios are most separable when the metric prioritizes consistent impact.

We record the metrics' performance in Table 3. In the Supplementary material, we report the classification results on the American Physical Society (APS) bibliographic data set.

Metrics agnostic to the distribution of citations appear to perform worse than their counterparts across either task. This includes the total number of papers N, as well as total citations C_{tot}, and maximum citations C_{max}. We highlight the performance of three metrics: N, C_{avg}, and C_{max}. N is consistently the worst performer because it does not account for the impact, only volume. C_{avg} is among the top performers considering the whole portfolio. We believe that is partly due to the nature of the distributions observed in Figure 1, where the

Table 4. Top 20 baseline scholars with the largest E-index in each discipline. The ones marked in bold type received the Nobel Prize between 2018 and 2022. Some authors are assigned multiple labels, so they may appear in multiple lists

Rank	Physics	Chemistry	Medicine
1	H. Dai	H. Dai	S. Kumar
2	A. L. Barabási	J. Godwin	R. A. Larson
3	D. Finkbeiner	R. Ruoff	A. L. Barabási
4	P. McEuen	K. L. Kelly	**G. L. Semenza**
5	I. Bloch	H. Wang	A. S. Levey
6	**A. Ashkin**	M. Egholm	**S. Paabo**
7	U. Seljak	L. Umayam	R. A. North
8	S. Inouye	L. Zhang	**A. Patapoutian**
9	**S. Manabe**	R. Freeman	J. Goldberger
10	M. Tegmark	P. Cieplak	M. Snyder
11	J. R. Heath	G. Church	J. Magee
12	L. Verde	**D. Macmillan**	**M. Houghton**
13	S. G. Louie	**G. Winter**	G. Loewenstein
14	D. I. Schuster	J. Kuriyan	S. Via
15	N. D. Lang	J. R. Heath	R. Jaeschke
16	B. Hammer	E. H. Schroeter	G. Hollopeter
17	D. Holmgren	W. Lin	S. J. Wagner
18	M. Lazzeri	W. L. Jorgensen	V. V. Fokin
19	L. P. Kouwenhoven	J. Clardy	**J. Allison**
20	M. Buttiker	D. Zhao	B. Moss

Nobelists are likely to accumulate higher than average citations over their careers. However, performance for the preaward portfolios is a bit worse, probably because we only consider the preaward period of their careers. Winning the prize has been shown to provide a tangible boost to the overall visibility of a scientist, resulting in more citations (Inhaber & Przednowek, 1976). The number of citations of the most cited paper C_{max} is among the worst performers, which suggests that the one big-hit portfolio is not typical among Nobelists. This finding supports the idea that scientists win the Nobel Prize after years of consistent, high-quality work.

We now shift our focus to the other category of indicators (i.e., ones sensitive to the citation distributions). We find that H records mediocre performance despite rewarding consistency. Its dependence on productivity likely fails to account for the Nobelists with a few highly cited papers. The Q-index performs poorly. However, its variant, $\widetilde{Q}$, fares considerably better, which is consistent with the fact that it is similar to M_α for small α.

M_α and E consistently rank in the top two positions. This further supports the hypothesis that Nobelists set themselves apart by producing a steady stream of high-impact work.

3.2. Identifying Future Nobelists

As a test of the predictive power of the metrics, we check whether we can identify scholars who received the Nobel Prize from 2018 to 2022 (i.e., the period not covered by our WoS data set). First, we note that our set of baseline scientists may be missing some of these new Nobelists, in which case we add them manually, provided they have a GS profile.

Then, for each metric, we construct a top 20 list of baseline scientists by ranking them in descending order and highlighting the Nobelists. We report the table for the E-index in the main text (Table 4), while the remaining lists can be found in the Supplementary material.

In Table 5, we show how many Nobelists appeared in the top 20 lists for each metric. E-index outperforms all other indicators, proving particularly effective for Medicine.

Table 5. Count of Nobelists awarded in the period [2018, 2022] identified in the top 20 lists of various metrics. The numbers in parentheses indicate how many such Nobelists have a GS profile

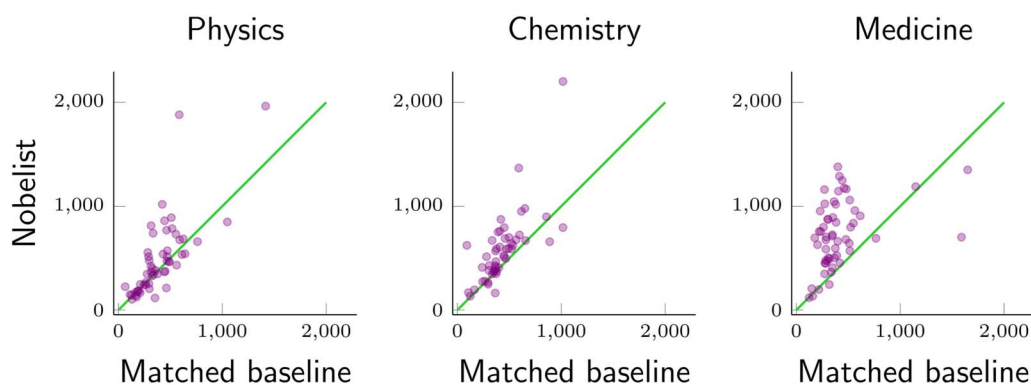| Metric	Physics (9)	Chemistry (8)	Medicine (5)
N	1	0	0
C_{tot}	**2**	0	3
C_{avg}	1	0	3
C_{max}	0	0	2
H	**2**	1	2
G	**2**	0	3
$\widetilde{Q}$	1	1	3
Q	0	1	1
M_α	1	0	2
E	**2**	**2**	**5**

Figure 4. *E*-index of Nobelists versus baseline scientists with comparable numbers of papers and citations. We see a prevalent trend towards larger *E* values for Nobelists. Some 58.2% (Physics), 86.3% (Chemistry), and 87.5% (Medicine) of Nobelists have larger *E* values than their counterparts.

To further corroborate this conclusion, we matched each Nobelist with a baseline scientist with (nearly) identical N and C_{tot} values. In Figure 4, we plot the *E*-index of each Nobelist and matched baseline pair. We find that the *E*-index of Nobelists usually exceeds that of their matches. Some exceptions correspond to Nobelists with a low number of highly cited papers. Other outliers might be prominent scholars who have not yet received the award but might receive it in the future.

4. CONCLUSION

In this work, we searched for productivity patterns in excellent scientific careers. Specifically, we aimed to assess whether the output of high-profile scientists is more likely to be characterized by a low number of hit papers or by a consistent production of high-quality work. To address this question, we have examined the scientific portfolios of Nobel Prize winners in Physics, Chemistry, and Medicine and checked which citation-based metrics are most suitable to recognize them among a much larger number of baseline scholars. In addition, we introduced two new metrics, the *E*-index and M_α, that reward both consistency and high average impact (when $\alpha < 1$).

We found that the best-performing metrics are the ones that peak when citations are distributed among a considerable number of works rather than being concentrated on a few hit papers. The *E*-index, in particular, proves especially effective in identifying future Nobelists. A portal for the calculation of *E*-index and other scores of individual performance can be found at e-index.net.

While there are Nobelists whose success relied on isolated hit papers, the most successful scientists usually stayed on top of their game for most of their careers.

COMPETING INTERESTS

The authors have no competing interests.

FUNDING INFORMATION

DATA AVAILABILITY

The data for Nobel laureates is available at Li et al. (2019). The disambiguated APS data set is available at Sinatra et al. (2016). The raw data set for the APS can be requested at https://journals.aps.org/datasets. The code is available at https://github.com/siragerkol/Consistency-pays-off-in-science. WoS data are not publicly available.

REFERENCES

Bornmann, L., & Tekles, A. (2019). Productivity does not equal usefulness. *Scientometrics*, *118*(2), 705–707. https://doi.org/10.1007/s11192-018-2982-5

Chu, J. S., & Evans, J. A. (2021). Slowed canonical progress in large fields of science. *Proceedings of the National Academy of Sciences*, *118*(41), e2021636118. https://doi.org/10.1073/pnas.2021636118, PubMed: 34607941

Dong, Y., Ma, H., Shen, Z., & Wang, K. (2017). A century of science: Globalization of scientific collaborations, citations, and innovations. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1437–1446). https://doi.org/10.1145/3097983.3098016

Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, *69*(1), 131–152. https://doi.org/10.1007/s11192-006-0144-7

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., ... Barabási, A.-L. (2018). Science of science. *Science*, *359*(6379), eaao0185. https://doi.org/10.1126/science.aao0185, PubMed: 29496846

Franzoni, C., & Rossi-Lamastra, C. (2017). Academic tenure, risk-taking and the diversification of scientific research. *Industry and Innovation*, *24*(7), 691–712. https://doi.org/10.1080/13662716.2016.1264067

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, *102*(46), 16569–16572. https://doi.org/10.1073/pnas.0507655102, PubMed: 16275915

Inhaber, H., & Przednowek, K. (1976). Quality of research and the Nobel Prizes. *Social Studies of Science*, *6*(1), 33–50. https://doi.org/10.1177/030631277600600102

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady*, *10*(8), 707–710.

Li, J., Yin, Y., Fortunato, S., & Wang, D. (2019). A dataset of publication records for Nobel laureates. *Scientific Data*, *6*(1), 33. https://doi.org/10.1038/s41597-019-0033-6, PubMed: 31000709

Milojević, S. (2015). Quantifying the cognitive extent of science. *Journal of Informetrics*, *9*(4), 962–973. https://doi.org/10.1016/j.joi.2015.10.005

Radicchi, F., & Castellano, C. (2013). Analysis of bibliometric indicators for individual scholars in a large data set. *Scientometrics*, *97*(3), 627–637. https://doi.org/10.1007/s11192-013-1027-3

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, *10*(3), e0118432. https://doi.org/10.1371/journal.pone.0118432, PubMed: 25738806

Sinatra, R., Wang, D., Deville, P., Song, C., & Barabási, A.-L. (2016). Quantifying the evolution of individual scientific impact. *Science*, *354*(6312), aaf5239. https://doi.org/10.1126/science.aaf5239, PubMed: 27811240

Sunahara, A. S., Perc, M., & Ribeiro, H. V. (2021). Association between productivity and journal impact across disciplines and career age. *Physical Review Research*, *3*(3), 033158. https://doi.org/10.1103/PhysRevResearch.3.033158

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036–1039. https://doi.org/10.1126/science.1136099, PubMed: 17431139